The Journal of

Economic Perspectives

Winter 2021

The Journal of
Economic Perspectives

A journal of the American Economic Association

The Journal of
Economic Perspectives

Contents   *Volume 35 • Number 1 • Winter 2021*

Symposia

Articles

Feature

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

The Elusive Employment Effect of the Minimum Wage

Alan Manning

Thirty years ago, Charles Brown (1988) invited readers of the *Journal of Economic Perspectives* to consider the question "Minimum Wage Laws: Are They Overrated?" and concluded that "the minimum wage is overrated by its critics as well as its supporters" (p. 144). At that time, there was a strong academic consensus that the minimum wage caused job losses and was not well-targeted on those it set out to help, and that as a result, it was dominated by other policies to help the working poor like the Earned Income Tax Credit. Although the minimum wage still commanded wide support among the general population, policymakers seemed to be paying attention to the economic consensus of the time: for example, in 1988 the US federal minimum wage had not been raised for almost a decade and only 10 states had higher minima. Minimum wages seemed to be withering away in other countries too. In 1993, the United Kingdom abolished the Wages Councils that had set minimum wages in some low-paying industries since they were established by Winston Churchill in 1909, leaving only a minimum wage in agriculture. In 1994, the OECD published its view on desirable labor market policies in a prominent *Jobs Study* report, recommending that countries "reassess the role of statutory minimum wages as an instrument to achieve redistributive goals and switch to more direct instruments" (OECD 1994).

The landscape looks very different today. There is pressure to make more use of minimum wages almost everywhere. In the United States, the current logjam in Congress means no change in the federal minimum wage is immediately likely.

■ *Alan Manning is Professor of Economics and Director of the Community Programme at the Centre for Economic Performance, both at the London School of Economics, London, United Kingdom. His email address is a.manning@lse.ac.uk.*

However, 29 states plus Washington, DC have a higher minimum wage. A number of cities are also going their own way, passing legislation to raise the minimum wage to levels (in relation to average earnings) not seen for more than a generation (in this issue, Dube discusses these sub-state level minimum wages). Outside the United States, countries are introducing minimum wages (for example, Hong Kong in 2011 and Germany in 2015) or raising them (for example, the introduction of the United Kingdom's National Living Wage in 2016, a higher minimum wage for those over the age of 25). Professional advice to policymakers has changed too. A joint report from the IMF, World Bank, OECD, and ILO in 2012 wrote "a statutory minimum wage set at an appropriate level may raise labour force participation at the margin, without adversely affecting demand, thus having a net positive impact especially for workers weakly attached to the labour market" (ILO 2012). The IMF (2014) recommended to the United States that "given its current low level (compared both to US history and international standards), the minimum wage should be increased." The updated OECD (2018) *Job Strategy* report recommended that "minimum wages can help ensure that work is rewarding for everyone" (p. 9) and that "when minimum wages are moderate and well designed, adverse employment effects can be avoided" (p 72).

Central to this change in view is what has sometimes been called the New Minimum Wage Research (perhaps best exemplified by Card and Krueger 1995) which, starting in the early 1990s, cast doubt on the conventional wisdom that the minimum wage inevitably destroyed jobs with the only interesting question being the size of the loss. Disagreement among economists remains: 25 years after this research began, there is no consensus on the employment effects of the minimum wage.[1] However, it does seem fair to say that clear negative impacts on employment of minimum wages are "elusive," a phrase used in Kennan's (1995) review of Card and Krueger (1995).

In the large number of papers that try to estimate the effect of minimum wages on employment, there is a danger of losing sight of the "why" question: specifically, why is it so hard to find negative employment effects of the minimum wage? Perhaps certain economic factors can explain the small and often ambiguous effects of the minimum wage on employment? Or perhaps labor markets are fundamentally different from other markets? These are the issues discussed by this paper. The conclusion is that the employment effect of the minimum wage is elusive, but that economists should not be surprised by this, given the way labor markets operate in which deviations from perfect competition are much larger than in many other markets. Indeed, perhaps it is time for the literature to move on to try to address the question of how high the minimum wage can be raised without significant employment effects appearing.

The Elusive Employment Effect: US Evidence

The employment effect has always constituted the bulk of research on the impact of the minimum wage and continues to do so, though there is a growing

---

[1] O'Neill (2015) shows that there are systematic differences in the characteristics of economists who signed the competing petitions for and against the 2013 US Fair Minimum Wage Act.

literature about the impacts on other outcomes (as discussed by Clemens in this issue). In many parts of the labor market, the minimum wage is set so low relative to prevailing wages that it is likely to have no detectable effect on any labor market outcome. Thus, it makes sense to confine the search for the employment effect of the minimum wage to segments where one can detect a significant wage effect. For this reason, most studies of the employment effect of the minimum wage focus on groups where the minimum is high in relation to average earnings, including teenagers, low-wage industries, or workers with low education levels.

While this approach is sensible from one perspective, it does create a sample selection problem. There is probably no economist who does not believe that there is some point at which higher minimum wages reduce employment, so it seems as if one should be able to choose samples in which this effect can be found. Even if one discovers such samples, it would not mean that the impact of the minimum wage is everywhere negative. However, the practical problem is that it is very hard to find sub-samples which demonstrate a robust negative effect of minimum wages on employment: the employment effect is elusive.

This essay is not a literature review and will not attempt an encyclopedic survey of the vast literature on the employment effect of the minimum wage: excellent reviews are available in Neumark and Wascher (2008); Belman and Wolfson (2014); and Schmitt (2015) as well as a number of meta-studies (Doucouliagos and Stanley 2009; Leonard, Stanley, and Doucouliagos 2014; Chletsos and Giotis 2015). In addition, Andrews and Kasy (2019) report some evidence of modest publication bias: studies with significant negative employment effects seem more likely to be published.

Instead, to illustrate the elusive employment effect, this paper focuses on the most studied group, the American teenager. This is also the group where it has been argued that a negative employment effect is most commonly found—although studies on fast food restaurants or retail, for example, tend to be less conclusive (for example Addison, Blackburn, and Cotti 2009, 2013). Teenagers may hate having adults poking their noses into their business, but their employment and wages have been a fertile field for generations of labor economists. The early literature (reviewed in Brown, Gilroy, and Kohen 1982) largely focused on time series analysis, but for more than 25 years (starting with Neumark and Wascher 1992), state-level panel data has been the focus of attention.

There is, of course, some reason for the focus on teens. They are the age group most affected by the minimum wage: Table 1 shows that in both 1979 and 2019, more than 25 percent of teens had a reported hourly wage at or below the minimum—defined here as the maximum of the federal or state-level minimum wage—although the fraction was somewhat lower in the intervening years. In several states in 2019, the minimum wage is above the median hourly wage for teens. Reported wages may be below the minimum because of measurement error, the use of the youth sub-minimum, because they are tipped workers or because the worker is not covered by the minimum wage as coverage is not universal. Thus, while the minimum wage in the United States as a whole is at quite a low level relative to median earnings (directly affecting less than 5 percent of workers), it is often binding in the teen labor market.

Table 1
The Diminishing Importance of the Teen Worker

	Age 16–19	Age 20–24	Age 25–29	Age 30–49	Age 50–64
Percent at or below minimum wage					
1979	37.7	13.7	7.5	7.4	8.9
1990	23.5	8.1	4.4	3.1	3.6
2000	18.1	8.4	4.1	2.9	3.0
2019	25.0	12.7	6.4	4.3	4.4
Percent of total hours					
1979	10.7	20.0	15.4	36.7	17.2
1990	4.2	12.1	15.9	52.3	15.5
2000	3.8	10.1	12.5	54.0	19.6
2019	2.3	9.1	13.1	47.0	28.4
Percent of total hours of minimum wage workers					
1979	32.5	22.4	9.9	22.4	12.8
1990	19.5	19.6	14.2	34.2	12.4
2000	15.6	19.4	12.1	38.6	14.3
2019	9.6	19.3	14.3	35.5	21.3
Percent in HS/college last week					
1990	61.9	23.8			
2000	68.0	28.6			
2019	75.6	35.6			

Source: Author's computations from the Merged Outgoing Rotation Groups of the Current Population Survey (CPS-MORG). See the online Appendix available at the *JEP* website for details.

But there is also something slightly odd about a research focus on teens, because this group is a small and declining share of total employment, representing only 2 percent of total hours worked in 2019, as also shown in Table 1. The teen share of minimum wage workers has been declining as well. In 1979, almost one in three minimum wage workers was a teen, this is now one in ten. In 1979, the number of minimum wage workers who were teens was much bigger than the number of minimum wage workers aged 20–24—but now the 20–24 group of minimum wage workers is over twice as large as the teen group. Finally, Table 1 shows that 75 percent of teens were in education last week (averaged across the whole year) and students' labor supply could plausibly be different from that of the population as a whole. For example, a finding that hours and employment fall when minimum wages rise could be a backward-bending labor supply curve as teenagers seek to earn a certain amount of income to finance their education, but otherwise want to spend time on their studies. Clearly, there is some risk in extrapolating the evidence on teens (or other low-paid groups) to the whole labor market, yet this practice is quite common (for example, Congressional Budget Office 2014, 2019) in producing estimates of the impact of proposed rises in the federal minimum wage on total employment.

In spite of all this, the effect of a minimum wage on teen employment is elusive. To illustrate this point, we present some estimates of the impact of minimum wages

Table 2
Explanatory Variables in the Seven Specifications

	Specification						
Explanatory variables	1	2	3	4	5	6	7
Constant term	Y	Y	Y	Y	Y	Y	Y
Log of minimum wage	Y	Y	Y	Y	Y	Y	Y
Prime-age unemployment rate	Y	Y	Y	Y	Y	Y	Y
Percent of teens in population	Y	Y	Y	Y	Y	Y	Y
State fixed effects	Y	Y	Y	Y	Y	Y	Y
Time fixed effects	Y	Y	Y	Y	Y	Y	Y
State time trend	N	linear	N	N	quadratic	cubic	quartic
Census division × time fixed effects	N	N	Y	Y	N	N	N

Note: In this paper, the dependent variable for these seven specifications will in some cases be wages; in other cases, it will be the employment/population ratio.

on the teenage labor market using quarterly state-level panel for 1979–2019. We do not argue for a "best" specification; instead, we present seven estimates based on those used in Neumark, Salas, and Wascher (2014) who are responding in part to Allegretto, Dube, and Reich (2011). In some cases, the left-hand dependent variable will be log wages; in other cases, it will be the log employment/population ratio. In all seven models, the right-hand side explanatory variables include a constant term, the log of the minimum wage defined as the maximum of the state and federal minimum (from Vaghul and Zipperer 2019) and other control variables including state and time fixed-effects (to account for permanent differences across states and aggregate labor market conditions), the prime-age unemployment rate (to account for state differences in business cycles), and the proportion of teens in the population (to account for the fact that labor market outcomes for teens may be affected by how many of them there are). We call these the baseline regressors (other factors like the changing demographics of teenagers are not commonly controlled for).

The estimated models differ according to whether other fixed effects and trends are included, as laid out in Table 2. Model 1 is just the baseline regressors. Model 2 allows for the possibility that states have different trends in the teen labor market by adding state-specific linear trends. Model 3 allows for the possibility that different regions of the US have different aggregate labor market conditions in every year by adding the census division interacted with time fixed effects. Model 4 includes both the census division x fixed time effects and the linear trends. Model 5 has a state-specific quadratic trend, model 6 a cubic trend; and model 7 a quartic trend, all intended to capture the possibility that the evolution of labor market conditions may vary across states in a complicated way. Again, these specifications are common in this literature. In the text of this paper, we will summarize results and confidence intervals for the coefficient on the log minimum wage in these regressions; full regression results are available in the online Appendix available with this paper at the *JEP* website.

Preliminaries: The Impact of Minimum Wages on Wages

Before considering the impact of the minimum wage on teen employment, it is important to verify that there is a detectable impact of the minimum wage on actual wages received by workers; failure to demonstrate this in any paper means there should be skepticism about any subsequent findings that the minimum wage has an employment effect. Figure 1 shows the estimated impact of the log minimum wage on mean log hourly wages for the seven different specifications, together with the 95 percent confidence intervals. This figure shows that for teen wages, there is a clear robust and significant impact of minimum wages on the mean log hourly wage for teens. Although the estimates vary with specification, the range is not very large, between 0.20 and 0.25. The size of the elasticity is slightly larger than the average fraction of teens reporting hourly wages at or below the minimum, which is consistent with modest spillover effects from the minimum wage (Autor, Manning, and Smith 2016).

One problem with the first specification is that it implies that the elasticity of teen wages with respect to the minimum wage is a constant, whatever the level of the minimum wage. This is implausible as a universal model, because it predicts that a rise in the minimum wage from $1 to $1.10 has the same impact as a rise from $10 to $11 dollars. We would expect the marginal effect of changes in the minimum wage on wages to be increasing in the minimum wage—a very low minimum wage relative to the prevailing level of wages (what Lee 1999 terms the effective minimum) will have little impact and a higher minimum wage a larger impact. The simplest way to investigate non-linearity is to include a quadratic term in the effective minimum. Details of the procedure (which follows Autor, Manning, and Smith 2016) and the estimates are in the online Appendix, but for the wages of teenagers, one can detect a non-linear effect. The estimated elasticity of log wages with respect to the minimum wage is about 0.25 when the minimum wage is 40 percent of the average wage, but about 0.57 when the minimum wage is 60 percent of the average wage.

The bottom line is that the estimated impacts of the minimum wage on the wages of teens is in line with what would be expected and reasonably robust to different specifications. The wage effect from a higher minimum wage is not elusive and estimates seem plausible.

Although most research focuses on teens, Figure 2 reports results for those aged 20–24. Remember, Table 1 showed that this group now accounts for a higher fraction of minimum wage workers than do teens. The estimated impacts of the minimum on wages are significantly different from zero in all models in the range 0.06–0.15, though lower than the estimated effect for teens as should be expected given the bite of the minimum wage is lower for the older age group. Similar regressions for the age group 25–29 (provided in the online Appendix) show no significant detectable impact on wages.

The Impact of Minimum Wages on Employment

Figure 3 presents estimates for the effect of the log minimum wage on the log employment rate of teenagers; in other words, these are the same seven specifications, and all that is changing is the dependent variable. Model 1 (which includes

Figure 1

The Sensitivity of Mean Log Hourly Wages with Respect to the Log Minimum Wage for Those Aged 16–19
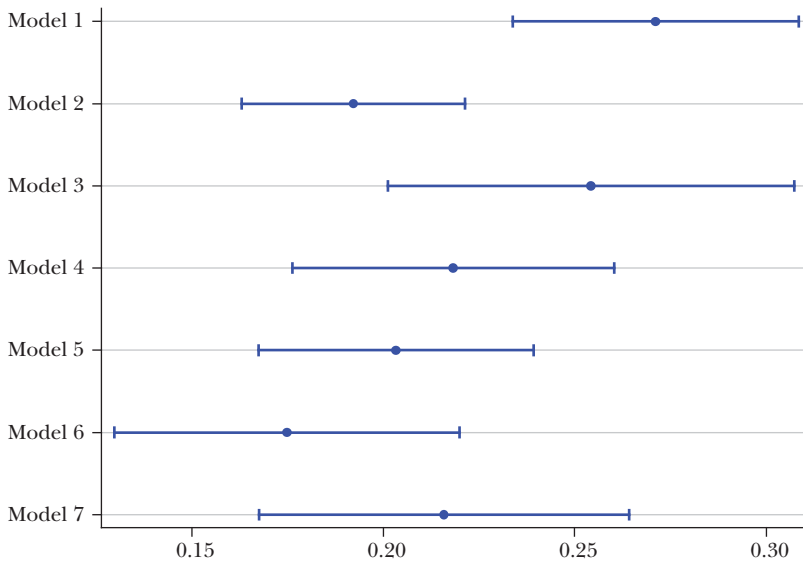
Note: This is the estimated coefficient from a regression of the mean log hourly wage on the log of minimum wage and other controls detailed in the text. The seven models are described in the text. Ninety-five percent confidence intervals are marked. Actual estimates reported in the online Appendix.

only the baseline regressors) produces a significant negative effect with an elasticity of –0.28, a common finding in the literature.[2] But unlike the wage effect, this is not at all robust. One other model (Model 3 with census division interacted with time effects, but no state trends) has a significant negative coefficient, but for all the others the estimated effects are positive, though not significantly different from zero.[3] Figure 4 reports the parallel results for the age group 20–24, which shows a similar pattern of results, although, as expected, the estimated elasticities are smaller.

This—in a nutshell—is the elusive employment effect. Even for groups where one can estimate a sizeable, robust wage effect, the employment effect is hard to find and not robust to specification. It seems unlikely that state-level panel data on teens is going to deliver clear evidence on a non-zero employment effect of the minimum wage (if it exists) in the near future. This literature has come to resemble trench warfare (complete with six feet of mud) in which the two sides are fighting over a small patch of ground which is not of much strategic importance in any case

---

[2] Cengiz, Lindner, and Zipperer (2019) argue that this result is sensitive to the inclusion in the data of the 1980s and 1990s when there was little change in the minimum wage.
[3] This is slightly different from Neumark, Salas, and Wascher (2014) who do estimate a negative effect in some of the later columns for the shorter sample period 1991–2011.

Figure 2

The Sensitivity of Mean Log Hourly Wages with Respect to the Log Minimum Wage for Those Aged 20–24

Note: This is the estimated coefficient from a regression of the mean log hourly wage on the log of minimum wage and other controls detailed in the text. The seven models are described in the text. Ninety-five percent confidence intervals are marked. Actual estimates reported in the online Appendix.
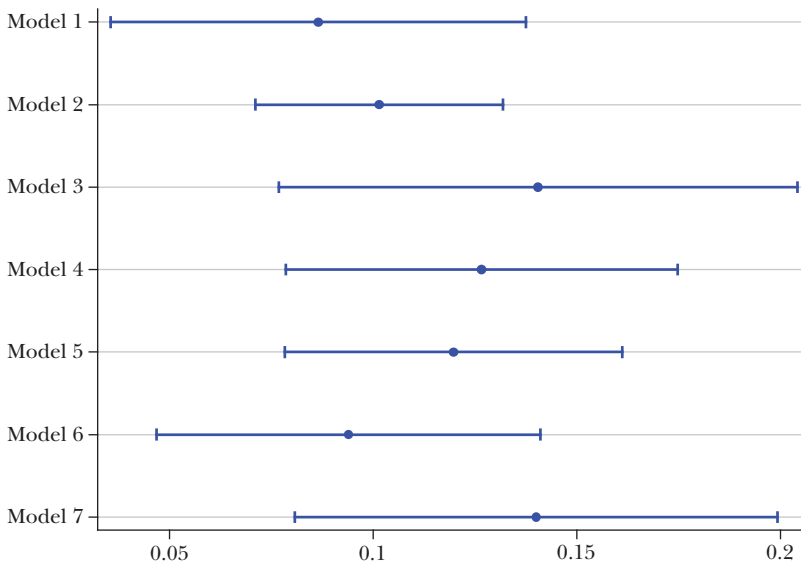
given the small size of the teen labor market. It is time to call a truce, acknowledge that no clear evidence of a negative effect on employment has been found, and redeploy the energy expended on this issue to other areas.

Although the specifications reported here are quite standard in the literature, there are a number of empirical criticisms that could be argued. Perhaps an employment effect can be identified if there is a better specification, a better statistical methodology, and/or better data. I consider these possibilities in the next few sections.

Better Specification?

The estimates reported so far are of a contemporaneous relationship between the level of the minimum wage and the level of wages and/or employment. Meer and West (2016) suggest it is better to estimate the model in first differences. However, when using this approach in this data, the elusive employment effect remains (detailed results in the online Appendix). Meer and West also suggest there may be lags in the impact of the minimum wage. This idea dates back to at least Baker, Benjamin, and Stanger's (1999) study of the minimum wage in Canada, but more recently, Sorkin (2015) argues that lags are likely to result from sluggish adjustment of capital that has a "putty-clay" technology (that is, a technology where producers

Figure 3

The Sensitivity of Log Employment Rate with Respect to the Log Minimum Wage for Those Aged 16–19
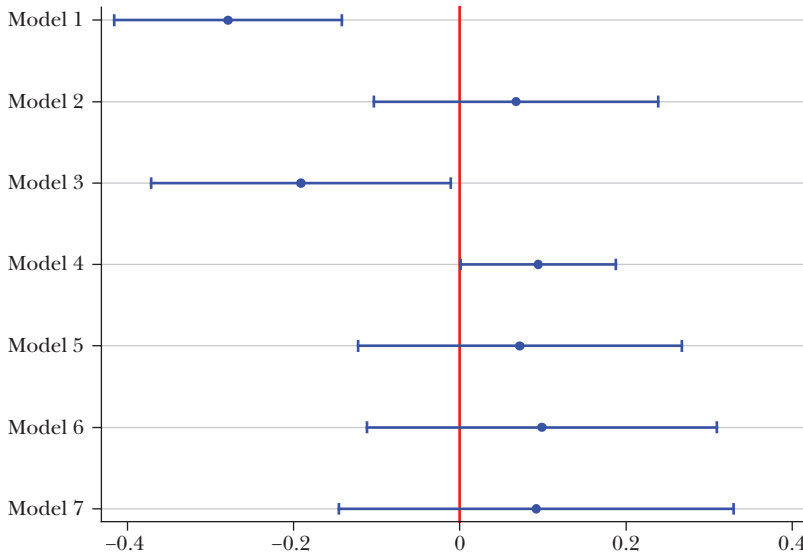
Note: This is the estimated coefficient from a regression of the mean log employment rate on the log of minimum wage and other controls detailed in the text. The seven models are described in the text. Ninety-five percent confidence intervals are marked. Actual estimates reported in the online Appendix.

can pick a level of capital and labor at the start, but then once the level of capital is chosen, it cannot be changed). Aaronson et al. (2018) present some evidence on restaurant entry, exit, and employment growth consistent with this insight, and Meer and West (2016) report finding a negative employment effect using long lags in aggregate employment data (though they do not demonstrate a wage effect). In the context of the teen employment, dynamics do not seem a major issue. Again, the online Appendix shows some estimates of the long-run impact of the minimum wage on teen employment rates allowing for a lagged impact. The elusive employment effect remains.

Another criticism is of the use of the log of the employment-population rate as the dependent variable. The theory of labor demand might be taken to mean that it is the level of employment that should be affected by the minimum wage. In this case, one might prefer simply the log of employment as the dependent variable: for example, Monras (2019) argues for this in part because he claims interstate migration responds to the minimum wage in which case the employment-population rate might not fall when the minimum wage rises because the population falls in line with employment. The online Appendix shows some estimates of the long-run impact of the minimum wage on teen employment levels. The elusive employment effect remains.

Figure 4

The Sensitivity of Log Employment Rate with Respect to the Log Minimum Wage for Those Aged 20–24
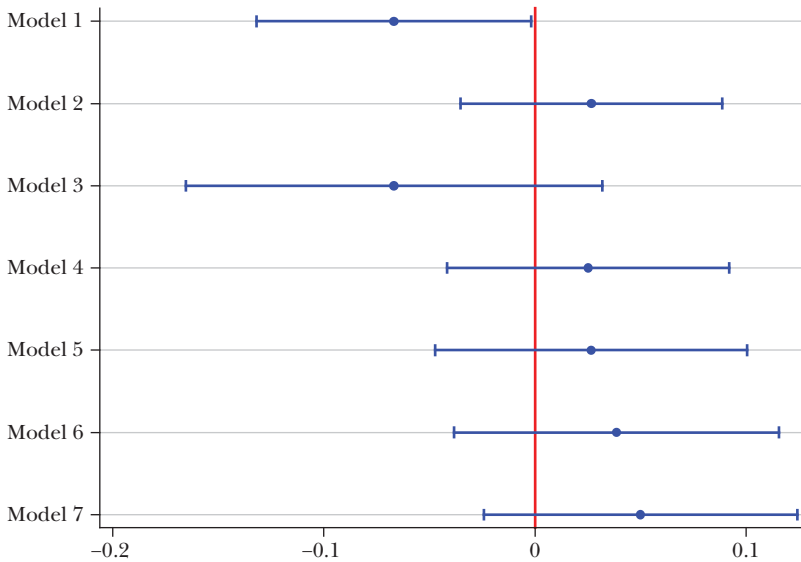
Note: This is the estimated coefficient from a regression of the mean log employment rate on the log of minimum wage and other controls detailed in the text. The seven models are described in the text. Ninety-five percent confidence intervals are marked. Actual estimates reported in the online Appendix.

Better Methodology?

A central concern in the estimates reported here is whether one has controlled appropriately for economic conditions affecting employment other than the minimum wage. Failure to do so effectively will lead to bias if the minimum wage is correlated with the omitted economic conditions. The state time trend and census division time effects in the models reported here are common ways of seeking to address this concern, but there might be better approaches.

As an alternative approach, Neumark, Salas, and Wascher (2014) use a synthetic control effect, arguing that a negative employment effect emerges when that is used. This conclusion is, however, contested by Dube and Zipperer (2015). It seems unlikely a synthetic control approach will deliver clear evidence on a negative employment effect, and the method is a bit of a black box. The synthetic controls for a state often vary implausibly with the sample period used, and even with the outcome being studied (for example, wages or employment) for the same sample. Dube, Lester, and Reich (2010) argue that a better way to control for other economic conditions is to use counties that border each other but are in different states and sometimes have different minimum wages as a result (an approach first used in Card and Krueger 1994). They find clear evidence of wage effects from the minimum wage but not evidence of disemployment effects.

Further support for the conclusion that there are no significant disemployment effects from the existing levels of minimum wages comes from Cengiz, Lindner, and Zipperer (2019). Instead of looking at how minimum wages affects the overall level of employment, this study looks at where in the wage distribution changes in employment occur. They show that a rise in the minimum wage causes jobs paid below the new minimum to disappear (unsurprising if there is compliance with the minimum wage) but are almost exactly compensated with higher employment at slightly higher wages. Overall, they find little effect on employment but do report some disemployment in tradeable sectors (although few minimum wage workers are to be found there).

Better Data?

If state-level panel data does not provide clear evidence for the disemployment effect of the minimum wage, perhaps better data is the solution. Studies that focus on aggregate data from low-wage industries or low-educated workers tend to have the same problem of an elusive employment effect as the analysis of the teen employment rates. For example, Bailey, DiNardo, and Stuart (2020) study the large rise in the minimum wage to a high level associated with the 1966 amendment to the Fair Labor Standards Act, finding it increased wages dramatically but reduced aggregate employment only modestly. They do report a much larger impact on the employment of African-Americans (40 percent of whom were affected by the rise), though another study by Derenoncourt and Montialoux (2020) of the same fails to find any employment effect.

The use of longitudinal individual data offers another approach. Using data prior to an increase in the minimum wage, one can identify workers likely to be affected by it and examine their subsequent employment outcomes. Care needs to be taken to find an appropriate control group, as it is well-known that low-wage workers are more likely to leave employment even absent any minimum wage. Using this approach, Clemens and Wither (2019) estimate large negative employment effects, but Zipperer (2016) argues that this conclusion is not robust to reasonable changes in model specification.

In many parts of labor economics, administrative data from both government and firms has been very useful. Social Security data is often hard to use in research on the impact of the minimum wage because this data often contains only earnings, not hours, making it difficult to identify minimum wage workers. The use of payroll data from firms (pioneered by Neumark and Wascher 2000) may be of more use. Such studies are rare, but some recent studies (Giuliano 2013; Hirsch, Kaufman, and Zelenska 2015; Coviello, Deserranno, and Persico 2019) all find clear wage effects but no negative employment effects.

The Elusive Employment Effect: International Evidence

The elusive employment effect of the minimum wage in the United States may be the result of the fact that minimum wages are quite low in relation to prevailing

wages. Perhaps evidence from other countries with higher minimum wages is more conclusive.

Table 3 presents OECD statistics on the level of the minimum wage in member countries in relation to full-time median earnings.[4] The United States is an outlier in having a relatively low level of minimum wages in relation to median earnings. In some middle-income countries, the statutory minimum wage is extremely high, though enforcement is probably weak in many cases. However, some high-income countries have minimum wages in the region of 50–60 percent of median earnings, so we might hope to look to these other countries where the minimum wage is higher to search for the employment effect. We will briefly consider a number of cases, selected to focus on where there is a large body of empirical research, and the minimum wage seems much higher than in the United States.

The United Kingdom

After the United States, the country with the most research on the employment impact of the minimum wage is the United Kingdom. The UK example is particularly interesting because in the period 1993–1999[5] it had no minimum wage, except in agriculture, leading to some very low job offers—as one example, a security officer who must provide their own dog to be paid £2 per hour (an example given in Low Pay Commission 1998). In 1999, the UK National Minimum Wage was introduced. Unlike in the United States, there is age variation in the form of lower minimum wage rates for workers aged less than 21 and for apprentices. In 2016, the UK government introduced what was termed the National Living Wage—essentially a higher minimum wage for the over-25 age group.[6]

The UK National Minimum Wage was initially set at a low level (about 46 percent of median hourly earnings), but subsequently increased to now be about 58 percent for the over-25s (Dube 2019). In 2019, the Conservative government (in what may come as a surprise to some American readers given the attitudes of many Republicans to raising the minimum wage) announced the ambition to raise this to 60 percent, economic conditions allowing.

The body responsible for making recommendations on the level of the minimum wage to the government, the Low Pay Commission, has always commissioned

---

[4]These figures are often lower than those from other sources because they exclude part-time workers who tend to be lower-paid than full-time workers. This obviously has a larger effect on countries with a higher proportion of part-time workers. One should also exercise some caution in making comparisons across countries: for example, Askenazy (2014) discusses how the treatment of tips in the minimum wage is very different in the United States, United Kingdom, and France. Also, many countries have lower rates for younger workers in a way that the United States does not. Finally, employer payroll taxes may also be important: for example, the French minimum wage does not appear as high in relation to average earnings when earnings are computed as total labor costs for the employer because payroll taxes are high in France but higher at the median than for a minimum wage worker (the introduction of this differential was studied by Kramarz and Philippon 2001).

[5]Prior to 1993, the United Kingdom had Wages Councils, which set minimum wages in a number of low-paid sectors, although some large low-paid sectors were excluded.

[6]This announcement was unexpected. Bell and Machin (2018) study the impact of the announcement on the share price of firms likely to be affected.

Table 3

The Level of Minimum Wages in OECD Countries, 2018

Country	Minimum wage/median full-time earnings
United States	33%
Spain	41%
Czech Republic	42%
Japan	42%
Mexico	42%
Estonia	43%
Germany	46%
Belgium	46%
Netherlands	47%
Ireland	48%
Greece	48%
Slovak Republic	49%
Latvia	50%
Lithuania	51%
Canada	51%
Hungary	52%
Poland	53%
Luxembourg	54%
Australia	54%
United Kingdom	54%
Romania	58%
Korea	59%
Slovenia	59%
Israel	59%
New Zealand	61%
Portugal	61%
France	62%
Chile	69%
Costa Rica	69%
Turkey	71%
Colombia	89%

Source: OECD.

Notes: These are national minimum wage as a percent of median hourly earnings for full-time workers. Because part-time workers are more likely to be low-paid, these figures are typically lower than those that report the minimum wage as a percent of median hourly earnings for all workers.

research on the employment effects (for its latest report, see Low Pay Commission 2019), and apart from a few studies in specific sectors (as in Machin, Manning, and Rahman 2003, which studied a sector where 30 percent of workers were paid the minimum), has found little or nothing. The National Minimum Wage is what it says, a national minimum wage, so investigation of the employment effect has concentrated on comparing groups where the minimum wage differs in its impact. For example, the National Minimum Wage has had more impact on younger than older workers, on women than men, and in lower-wage regions of the United Kingdom.

A useful review of the evidence was commissioned by the UK Treasury (Dube 2019). For example, the report shows (chart 4.I) that the introduction of the National Living Wage in 2016 raised wages in those segments of the labor market

most affected, but had no effect on employment. The effect of the minimum wage on employment in the United Kingdom has also been elusive.

France

France is often held up as the model of what happens when the minimum wage, known as the SMIC (for *salaire minimum de croissance*), is set too high. France has had high unemployment in general, and high youth unemployment in particular, for a very long time. The SMIC is currently €10.15 per hour (about $11.30) and 13.4 percent of workers directly benefitted from the 2019 increase (Direction Générale du Trésor 2019, 36), much higher than the proportions directly affected by minimum wage increases in the United States or the United Kingdom. This suggests that the French minimum wage is at a relatively high level. However, the cost of a minimum wage worker relative to a worker on median earnings is not that high because France has very progressive payroll taxes: The Générale du Trésor (2019) expert group reports that once this factor is taken into account, the French minimum wage relative to average labor costs is quite similar to the international median.

Since the inception of the French Générale du Trésor expert committee in 2008, it has never recommended increasing the minimum faster than the minimum required by the law, reflecting the view of economists that the minimum wage in France is too high. However, the fact that the minimum wage is increased every year according to a formula, with only small variation relative to average earnings, means that it is very difficult, if not impossible, to identify clearly the impact of the minimum wage in France. One study that does consider a large change is Kramarz and Philippon (2001), who consider a change in 1996 that reduced the payroll tax on minimum wage workers at the same time that the minimum wage itself was increased. That study found an increase in teen employment, but because the cost of labor to the employer was reduced at the same time as the incentive to work was raised, one would expect both the demand and the supply for labor to increase (as explained below) so employment would be expected to rise. This is not the same as a general cut or rise in the minimum wage.

It may be that the minimum wage is set in France at a level where it results in sizeable job losses, but we don't have the studies to back up such an assertion.

Australia

Minimum wages in Australia are complicated. Although Australia does have a Federal Minimum Wage (currently set at AUS$19.49, approximately US$12.50), it also has a system of "Modern Awards" that sets minimum wages by industry, occupation, and seniority. As a result, there are over 1,500 different minimum wages in total, and the federal minimum wage is just an absolute floor. But even the federal minimum is at a relatively high level relative to average earnings—around 16 percent of workers are currently paid the minimum wage (Plunkett and Borland 2014). Unlike France, the Australian labor market has not been plagued by persistently high unemployment rates (in the year or so leading up to the pandemic, its monthly unemployment rate was a bit above 5 percent).

Like in France, the nature of the minimum wage variation in Australia does not lend itself to a high-quality research design when it comes to investigating the impact of minimum wages on employment. However, Bishop (2018) concludes that there iss an impact on wages, but none on employment or hours. The Fair Wage Commission, an expert body that advises on the appropriate level of awards, is also of the view that the prevailing level of minimum wages is not harmful to employment. The Australian experience is a useful counterpoint to the argument that all the countries with the highest minimum wages have a clear unemployment problem.

Germany

Germany introduced a National Minimum Wage in 2015, replacing a previous system that had set minimum wages in a small number of sectors. It is currently at €9.35 per hour (about $10.40). An increasing number of papers have sought to estimate the effect on wages and employment: for example, see the summary of the expert commission in Mindestlohn Kommission (2018), or the surveys by Caliendo, Schröder, and Wittbrodt (2019), Bossler and Garner (2019), and Dustmann et al. (2020). Because Germany has a national minimum wage, these studies have to compare groups where the minimum wage has small or large effects, as in the UK studies. While all studies agree there has been an impact on wages, there is less agreement about the impact on employment, reminiscent of the debates in the United States and the United Kingdom. The employment effect of the minimum wage in Germany also seems elusive.

Other Countries

There are many studies of minimum wages in other countries, far too many to summarize here. Consider two recent papers that do report evidence that rises in minimum wages have reduced employment.

Looking at a large rise in minimum wages in Hungary in 2001 from 35 percent of median earnings to 50 percent, Harasztosi and Lindner (2019) find evidence that wages rose faster in more affected firms, but these firms had slower employment growth, though the implied elasticity was small (somewhat larger in tradeable sectors). Most of the impact occurred soon after the rise, so lags did not seem important. The estimated employment effects are based on the identifying assumption that there is no impact of the minimum wage on unaffected firms, which assumes no reallocation effect of the kind that Dustmann et al. (2020) find for Germany.

Looking at the 40 percent rise in Danish minimum wages when workers turn 18 and become eligible for the adult rate (which is set at a level comparable to a $15 minimum in California) and using a regression discontinuity design and administrative data, Kreiner, Reck, and Skov (2020) document a 33 percent fall in the employment rate and a 45 percent fall in hours. The size of this effect could represent the high elasticity of substitution between workers around their 18th birthday.

It may be tempting to conclude that studies like these show that minimum wages clearly harm employment when minimum wages are in the region of 50–60 percent of median earnings. However, minimum wages are this high relative to median earnings in the US teen labor market. Also, studies from many other countries with

similarly high levels of minimum wages often fail to find a clear employment effect: for example, studies of UK jumps in minimum wage at different ages have failed to uncover anything like the clear effects found in the Danish study. One could also point to studies that fail to find clear evidence of negative employment effects of large changes in minimum wages in countries where the level of minimum wages are relatively high: for example, in New Zealand (Hyslop and Stillman 2007, 2011), Ireland (O'Neill, Nolan, and Williams 2006), and Portugal (Portugal and Cardoso 2006).

The conclusion is that the employment effect remains—for the moment—elusive. But it may be that we are about to learn more about how high one can push the minimum wage before there are clear adverse employment effects as a number of jurisdictions are pushing minimum wages higher and higher (as discussed by Dube in this issue).

Why Is the Employment Effect So Elusive?

The difficulty of establishing a clear employment effect at moderate levels of the minimum wage now seems relatively well-established. What are some possible reasons for this pattern?

Start with a very simple model based on a competitive labor market for the analysis of the wage and employment effects of the minimum wage. In this model, when a minimum wage is enacted, some workers who had been receiving a lower wage now get a higher wage, while other workers who had been receiving the lower wage become unemployed. As noted earlier, a common pattern in the empirical literature is that the minimum wage does seem to result in a wage increase but not to a fall in employment.

One possibility is that in econometric terms, it may be harder to find an equal-sized impact on employment than wages, perhaps because employment has more residual variation leading to larger standard errors. Looking at Figures 1–4 one can see that, for teens, the standard errors on the employment estimates are considerably larger than those on the wage estimates—meaning that the employment effect is less precisely estimated. On the other hand, for the age group 20–24, the standard errors in the employment equations are only slightly larger than those in the wage equations.

Alternatively, it may be that the impact of the minimum wage on employment is weaker than on wages: one possibility is that for some reason the pass-through from minimum wages to the cost of labor to employers is low; another possibility is that the labor demand curve has a low elasticity. We consider arguments for these possibilities in turn.

Low Pass-Through

Why might the link between wages and employer labor costs be weak? Perhaps employers will react to higher minimum wages by being less generous with other aspects of the employment contract—what Brown (1988) called "offsets"—such as meal breaks, fringe benefits, health benefits, or training (for more discussion, see Clemens in this issue). Such offsets would mean that the overall gain to workers from

a higher minimum wage is lower than the wage gain. Depending on the value that workers place on the offsets, it is theoretically possible that workers could be made worse-off with a combination of a higher wage and additional offsets, although the fact that minimum wage workers support raises in the minimum wage suggests we are not at that extreme. But although offsets are a theoretical possibility, evidence for such offsets is decidedly weak (see the review in Neumark and Wascher 2008; or see Belman and Wolfson 2014).

But there are other reasons why labor costs may not rise as fast as wages which do not depend on workers being made worse-off: specifically, turnover and monitoring costs as in the Rebitzer and Taylor (1995) application of the Shapiro and Stiglitz (1984) efficiency wage models. Say that an employer faces turnover costs in the form of a fixed cost to hire and train a new worker. In addition, say that the higher the wage, the lower the rate at which workers leave the firm (a view supported by a lot of evidence) so the less hiring the firm needs to do to maintain employment at its desired level. Indeed, we have accumulating evidence that increases in the minimum wage are associated with lower labor turnover even if total employment is unchanged (Portugal and Cardoso 2006; Brochu and Green 2013; Dube, Lester, and Reich 2016; Dube, Giuliano, and Leonard 2019).

In this model, a firm faces a trade-off in deciding on the appropriate level of wages. Higher wages means higher direct labor costs but lower turnover and associated costs. An employer unconstrained by any minimum wage will set the wage to minimize total labor costs.

Now consider how a firm with turnover costs will adjust its hiring with the imposition of a minimum wage. The minimum wage raises the wage that the firm needs to pay but also reduces turnover costs for the firm. The effect of a higher minimum wage on total labor costs will be less—and perhaps substantially less—than the rise of the minimum wage itself. For an employer for which the minimum wage is only slightly higher than the wage it would choose when constrained, one can show that the rise in total labor costs would be zero. Because of turnover costs, a small change in the minimum wage can have a zero effect on total labor costs and hence, on employment.

Higher wages may also raise productivity, implying that a rise in the minimum wage has a smaller effect on the cost of an efficiency unit of labor than one might expect. Coviello, Deserranno, and Persico (2019) find for a sample of workers from a large US retailer with outlets across a number of states who are paid a mixture of base pay and commission that higher minimum wages lead to higher productivity. Their proposed model is that a minimum wage increases the value of the job, and workers increase effort in order to avoid bearing the costs of being fired. Consistent with this, they find that the effect is concentrated on the least productive workers who might be at greater risk of job loss. Although they find clear impacts on wages and productivity, they find no effect on employment or terminations—another example of the elusive employment effect.

The Elasticity of the Labor Demand Curve

Another possibility for why the employment effect is elusive is that the elasticity of the labor demand curve is low. Competitive market theory would predict that

a higher minimum wage will lead to a move away from the use of minimum wage labor in the production of a given level of output and thus to a fall in the level of output produced. The labor elasticity of demand for minimum wage workers depends on several underlying parameters: the share of minimum wage workers in total costs for the firm, the elasticity of substitution of minimum wage workers with other inputs, and the price elasticity of demand for the final product.

For most firms, the share of minimum wages in total costs is small. Teulings (2000) provides a theoretical explanation for a very low value for the elasticity of substitution. Haratoszi and Lindner (2019) distinguish between the elasticity between minimum wage labor on capital and materials. They suggest that materials are the most important non-labor input and they are complementary to labor. The literature on the impact of minimum wages on prices (for example, Aaronson 2001) typically does find some pass-through at a level consistent with the impact on total costs. Some have inferred from price rises that output and employment must fall (for example, Aaronson and French 2007), but in the sectors in which minimum wages have the biggest impact, quality of service is obviously variable as well as breaking any simple link between prices, output, and employment. Taking these factors together, it seems plausible that the elasticity of labor demand for low-wage labor in many sectors is not very large.

Another set of complications arises in the distinction between non-traded goods and services, often provided in local markets, and tradeable sectors of the economy. In the case of non-traded goods and services provided mostly in local markets, a change in local or state-level minimum wage laws will affect all firms in the industry. As a result, the relevant product demand elasticity is then that for the (local) industry demand curve, which may be low. Additionally, there may be positive effects on local demand if minimum wage workers have a higher propensity to consume than the owners of firms. However, for production of goods and services traded outside a given area, firms that face a higher minimum wage must compete with firms that do not face this higher minimum wage, and so their elasticity of demand for minimum wage labor may be larger. Some evidence for a more negative employment effect in tradeable sectors can be found in Haratoszi and Lindner (2019) and Dube (this issue). But most minimum wage workers in high-income countries like the United States are not in tradeable sectors, where wages are typically higher.

Imperfectly Competitive Labor Markets

The strong a priori belief held by many that a rise in the minimum wage must cost jobs ultimately derives from the assumption that the low-wage labor market is close to perfectly competitive. The basic theoretical argument is straightforward. There is a frictionless market with a downward-sloping labor demand curve and an upward-sloping labor supply curve. In a frictionless market, the equilibrium wage is where the demand for labor equals supply. In this setting, one immediately reaches the conclusion that a binding minimum wage must put the economy in the region where the quantity of labor supplied exceeds the quantity of labor demanded, so increases in the minimum reduce employment with an elasticity given by the elasticity of the labor demand curve. Of course, there are more sophisticated competitive

analyses—recognizing there are many interconnected labor markets and that there will be effects on the demand for different goods, and general equilibrium effects that might (as is their wont) conceivably be quite complicated. But in a frictionless and perfectly competitive labor market, the basic insight that a minimum wage will reduce employment for low-wage workers typically continues to hold.

Although this analysis is regarded by many as "conventional" (a word used to describe it in, for example, Congrssional Budget Office 2014) when applied to the analysis of minimum wages, it is not conventional when applied to other labor market phenomena, like unemployment. After all, the frictionless competitive market model implies that involuntary unemployment cannot exist, except when created by institutions like the minimum wage. Such a view is not "conventional" at all. Even introductory economics courses discuss reasons why unemployment exists at the macroeconomic level. The tensions between these views come up in various ways: for example, some opponents of a higher minimum wage emphasize that employment is determined by the demand-side of the labor market alone when analyzing the impact of minimum wages, but when looking at the effect of extensions to unemployment insurance, they argue that labor supply decisions affect employment. To summarize, labor economists frequently use a different model of the labor market when analyzing minimum wages from the one they use when analyzing unemployment.

Most contemporary analyses of unemployment assume that the labor market has frictions, such that—at prevailing wages—not all workers who want a job manage to get one and that not all employers who want to hire a worker manage to find one. The consequence is that unemployment and job vacancies co-exist simultaneously in the labor market. The existence of unemployment means that actual employment is below the quantity that workers would be willing to supply at the given wage. The existence of job vacancies means that actual employment is below the quantity that employers are willing to demand at the given wage. The observed level of employment can, somewhat loosely, be thought of as being influenced both by the amount of labor employers would like to hire and the amount that workers would like to supply.

In this setting, higher minimum wages lead to both a decrease in the quantity of labor demanded and a rise in the quantity of labor supplied—with the overall impact on employment depending on which effect is larger. As a result, the impact of an increase in the minimum wage on employment is theoretically ambiguous. There is no a priori reason why an imperfect labor market with frictions should necessarily be at a current equilibrium where a rise in the wage of low-skill labor above the current equilibrium necessarily reduces employment.

The treatment of imperfect labor markets with frictions here is admittedly brief and casual, and I am sure will offend purists. Those interested in digging into these types of models in more detail might usefully look at bargaining models (for example, Pissarides 2000) or monopsony models (for example, Burdett and Mortensen 1998; Manning 2003, 2011, forthcoming; Flinn 2006).[7] But for present

---

[7]It is well-known that minimum wages can theoretically raise employment even if the labor markets are frictionless if one assumes that employers have some monopsony power. Just as a profit-maximizing monopolist will want to raise its selling price, in which case a price ceiling below that level can cause the

purposes, this high degree of simplification is justified in terms of making it very clear that an empirical finding that the minimum wage does not reduce employment should not be regarded as breaking some fundamental law of economics.

Summary

The ideas discussed here are summarized in Figure 5. In the simple textbook model, a rise in the minimum wage has a direct strong link to labor costs, which is linked to reduced labor demand and employment. This is the process represented by the arrows down the middle of Figure 5. But there may be increased offsets, reduced turnover costs, and increased productivity that mitigate the impact of rises in the minimum wage on labor costs. This is represented by the box on the left-hand side of Figure 5. In imperfect labor markets, the impact of rising minimum wages on raising labor supply also needs to be taken into account in the overall impact on employment, as represented by the box on the right-hand side of Figure 5. With the framework in mind, the elusive employment effect of a higher minimum wage should not be much of a surprise.

Conclusion

Much of the literature on the employment of the minimum wage focuses on the question of "what is the employment effect of the minimum wage" using an empirical specification in which the effect is always negative, zero, or positive. This approach has reached the point of diminishing returns. A balanced view of the evidence makes it clear that existing evidence of a negative employment effect is not robust to reasonable variation in specification, even when the wage effect is robust. This might mean that the labor demand elasticity is very small (and this paper has discussed some reasons why that might be the case), but it might mean that the effect of a higher minimum wage on employment (within the existing range of minimum wages) is not negative at all. The claim that the employment effect might not be negative continues to be met with incredulity in some quarters or to be euphemistically described as "unconventional." But as soon as one acknowledges that efficiency wage effects might be important or that labor markets have frictions—ideas that appear in mainstream introductory-level textbooks and can hardly be described as unconventional—one has to acknowledge that the impact of the minimum wage on employment is theoretically ambiguous.

Of course, there is some level of the minimum wage at which employment will decline significantly. The empirical literature on the minimum wage should reorient itself towards investigating the determinants of that point. The currently observed range of minimum wages apparently does not include the turning-point, but recent initiatives suggest we may be about to observe the impact of much higher

---

monopolist to increase output, a profit-maximizing monopsonist will want to reduce its buying price, and a price floor above that level can cause the monopsonist to increase purchases. But most credible accounts of monopsony power are based on frictions, and we do not pursue those models here.

Figure 5
A Stylized Representation of the Impact of the Minimum Wage on Employment
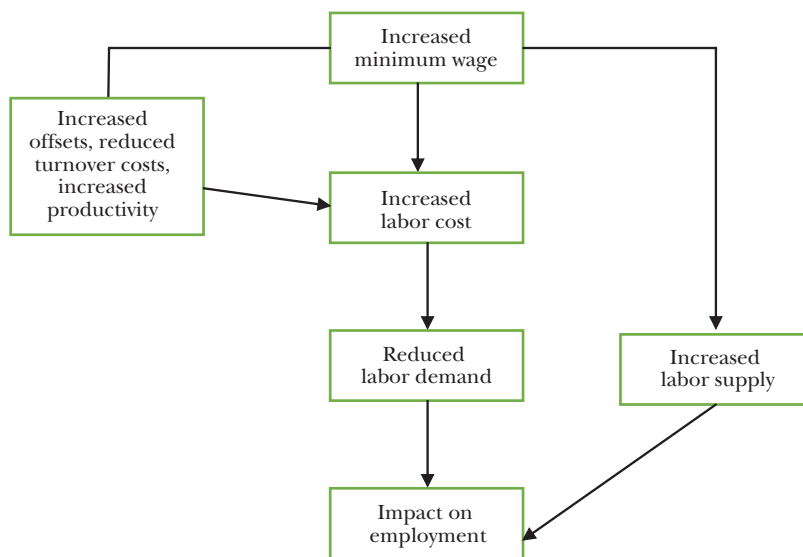
minimum wages in the near future in the United States and elsewhere (as I finish writing this, the residents of Geneva, Switzerland, have just voted for a $25 per hour minimum wage). Together with an increased use of high-quality data, we may, in this way, be entering a new chapter of minimum wage research.

References

Aaronson, Daniel. 2001. "Price Pass-Through and the Minimum Wage." *Review of Economics and Statistics* 83 (1): 158–69.

Aaronson, Daniel, and Eric French. 2007. "Product Market Evidence on the Employment Effects of the Minimum Wage." *Journal of Labor Economics* 25 (1): 167–200.

Aaronson, Daniel, Eric French, Isaac Sorkin, and Ted To. 2018. "Industry Dynamics and the Minimum

Wage: A Putty-Clay Approach." *International Economic Review* 59 (1): 51–84.

Addison, John T., McKinley Blackburn, and Chad Cotti. 2009. "Do Minimum Wages Raise Employment? Evidence from the U.S. Retail-Trade Sector." *Labour Economics* 16 (4): 397–408.

Addison, John T., McKinley L. Blackburn, and Chad D. Cotti. 2013. "Minimum Wage Increases in a Recessionary Environment." *Labour Economics* 23: 30–39.

Allegretto, Sylvia A., Arindrajit Dube, and Michael Reich. 2011. "Do Minimum Wages Really Reduce Teen Employment? Accounting for Heterogeneity and Selectivity in State Panel Data." *Industrial Relations* 50 (2): 205–40.

Andrews, Isaiah, and Maximilian Kasy. 2019. "Identification of and Correction for Publication Bias." *American Economic Review* 109 (8): 2766–94.

Askenazy, Philippe. 2014. "The Parameters of a National Minimum Hourly Wage." IZA Policy Paper 85.

Autor, David, Alan Manning, and Christopher Smith. 2016. "The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment." *American Economic Journal: Applied Economics* 8 (1): 58–99.

Bailey, Martha J., John DiNardo, and Bryan A. Stuart. 2020. "The Economic Impact of a High National Minimum Wage: Evidence From the 1966 Fair Labor Standards Act." NBER Working Paper 26926.

Baker, Michael, Dwayne Benjamin, and Shuchita Stanger. 1999. "The Highs and Lows of the Minimum Wage Effect: A Time-Series Cross-Section Study of the Canadian Law." *Journal of Labor Economics* 17 (2): 318–50.

Bell, Brian, and Stephen Machin. 2018. "Minimum Wages and Firm Value." *Journal of Labor Economics* 36 (1): 159–95.

Belman, Dale, and Paul J. Wolfson. 2014. *What Does the Minimum Wage Do?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. EQ: Maybe 2013 in text

Bishop, James .2018. "The Effect of Minimum Wage Increases on Wages, Hours Worked and Job Loss." *Bulletin of Reserve Bank of Australia*, September 20. https://www.rba.gov.au/publications/bulletin/2018/sep/pdf/the-effect-of-minimum-wage-increases-on-wages-hours-worked-and-job-loss.pdf.

Bossler, Mario, and Hans-Dieter Gerner. 2019. "Employment Effects of the New German Minimum Wage: Evidence from Establishment-Level Microdata." *Industrial and Labor Relations Review* 73 (5): 1070–94.

Brochu, Pierre, and David A. Green. 2013. "The Impact of Minimum Wages on Labour Market Transitions." *Economic Journal* 123 (573): 1203–1235.

Brown, Charles. 1988. "Minimum Wage Laws: Are They Overrated?" *Journal of Economic Perspectives* 2 (3): 133–45.

Brown, Charles, Curtis Gilroy, and Andrew Kohen. 1982. "The Effect of the Minimum Wage on Employment and Unemployment." *Journal of Economic Literature* 20 (2): 487–528.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 39 (2): 257–73.

Caliendo, Marco, Carsten Schröder, and Linda Wittbrodt. 2019. "The Causal Effects of the Minimum Wage Introduction in Germany—An Overview." *German Economic Review* 20 (3): 257–92.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.

Card, David, and Alan B. Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage.*" Princeton: Princeton University Press.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *Quarterly Journal of Economics* 134 (3): 1405–54.

Chletsos, Michael, and Georgios P. Giotis. 2015. "The Employment Effect of Minimum Wage Using 77 International Studies since 1992: A Meta-Analysis." Munich Personal RePEc Archive Paper 61321.

Clemens, Jeffrey, and Michael Wither. 2019. "The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers." *Journal of Public Economics* 170 53–67.

Congressional Budget Office. 2014. *The Effects of a Minimum-Wage Increase on Employment and Family Income.* Washington, DC: Congressional Budget Office.

Congressional Budget Office. 2019. *The Effects on Employment and Family Income of Increasing the Federal Minimum Wage.* Washington DC: Congressional Budget Office.

Coviello, Decio, Erika Desarranno, and Nicola Persico. 2019. "Minimum Wage and Individual Worker Productivity: Evidence from a Large US Retailer." https://drive.google.com/file/d/0B1g6WDc8t1zLUVpEU0ZSdFUySUE/view.

Ellora Derenoncourt, and Claire Montialoux. 2020. "Minimum Wages and Racial Inequality." https://doi.org/10.1093/qje/qjaa031.

Direction générale du Trésor. 2019. "Salaire Minimum Interprofessionnel de Croissance: Rapport du Groupe d'Experts." Paris: Direction générale du Trésor. https://www.tresor.economie.gouv.fr/Articles/2018/12/21/rapport-annuel-du-groupe-d-experts-smic (accessed November 16, 2020).

Doucouliagos, Hristos, and T.D. Stanley. 2009. "Publication Selection Bias in Minimum-Wage Research? A Meta-Regression Analysis." *British Journal of Industrial Relations* 47 (2): 406–28.

Dube, Arindrajit. 2019. *Impacts of Minimum Wages: Review of the International Evidence.* London: HM Treasury

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard. 2019. "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior." *American Economic Review* 109 (2): 620–63.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties." *Review of Economics and Statistics* 92 (4): 945–64.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2016. "Minimum Wage Shocks, Employment Flows, and Labor Market Frictions." *Journal of Labor Economics* 34 (3): 663–704.

Dube, Arindrajit, and Ben Zipperer. 2015. "Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies." IZA Discussion Paper 8944.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp vom Berge. 2020. "Reallocation Effects of the Minimum Wage: Evidence from Germany." Unpublished.

Flinn, Christopher J. 2006. "Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates." *Econometrica* 74 (4): 1013–62.

Giuliano, Laura. 2013. "Minimum Wage Effects on Employment, Substitution, and the Teenage Labor Supply: Evidence from Personnel Data," *Journal of Labor Economics* 31 (1): 155–94.

Harasztosi, Peter, and Attila Lindner. 2019. "Who Pays for the Minimum Wage?" *American Economic Review* 109 (8): 2693–2727.

Hirsch, Barry T., Bruce E. Kaufman, and Tetyana Zelenska. 2015. "Minimum Wage Channels of Adjustment." *Industrial Relations* 54 (2): 199–239.

Hyslop, Dean, and Steven Stillman. 2007. "Youth Minimum Wage Reform and the Labour Market in New Zealand." *Labour Economics* 14 (2): 201–30.

Hyslop, Dean, and Steven Stillman. 2011. *The Impact of the 2008 Youth Minimum Wage Reform.* Wellington, New Zealand: The Department of Labour.

ILO. 2012. *Boosting Jobs and Living Standards in G20 Countries: A Joint Report by the ILO, OECD, IMF and the World Bank.* Geneva: ILO.

IMF. 2014. *2014 Article IV Consultation with the United States of America Concluding Statement of the IMF Mission.* Washington, DC: IMF.

Kennan, John. 1995. "The Elusive Effects of Minimum Wages." *Journal of Economic Literature* 33 (4): 1950–65.

Kramarz, Francis, and Thomas Philippon. 2001. "The Impact of Differential Payroll Tax Subsidies on Minimum Wage Employment." *Journal of Public Economics* 82 (1): 115–46.

Kreiner, Claus Thustrup, Daniel Reck, and Peer Ebbesen Skov. 2020. "Do Lower Minimum Wages for Young Workers Raise Their Employment? Evidence from a Danish Discontinuity." *Review of Economics and Statistics* 102 (2): 339–54.

Lee, David S. 1999. "Wage Inequality in the United States during the 1980s: Rising Dispersion or Falling Minimum Wage?" *Quarterly Journal of Economics* 114 (3): 977–1023.

Leonard, Megan de Linde, T.D. Stanley, and Hristos Doucouliagos, (2014). "Does the UK Minimum Wage Reduce Employment? A Meta-Regression Analysis." *British Journal of Industrial Relations* 52 (3): 499–520.

Low Pay Comission. 1998. *The National Minimum Wage: First Report of the Low Pay Commission.* London: Low Pay Commission.

Low Pay Commission. 2019. *The Low Pay Commission Report 2019.* London: Low Pay Commission.

Machin, Stephen, Alan Manning, and Lupin Rahman. 2003. "Where the Minimum Wage Bites Hard: The Introduction of the UK National Minimum Wage to a Low Wage Sector." *Journal of the European Economic Association* 1 (1): 154–80.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets.* Princeton: Princeton University Press.

Manning, Alan. 2011. "Imperfect Competition in the Labor Markets" In *Handbook of Labor Economics*, Vol. 4, edited by Orley Ashenfelter and David Card, 973–1041. Amsterdam: Elsevier.

Manning, Alan. Forthcoming. "Monopsony in Labor Markets: A Review." *Industrial and Labor Relations*

Review.

Meer, Jonathan, and Jeremy West. 2016. "Effects of the Minimum Wage on Employment Dynamics." *Journal of Human Resources* 51 (2): 500–22.

Mindestlohn Kommission. 2018. "Zweiter Bericht zu den Auswirkungen des gesetzlichen Mindestlohns." https://www.mindestlohn-kommission.de/DE/Bericht/pdf/Bericht2018.html?nn=7916938

Monras, Joan. 2019. "Minimum Wages and Spatial Equilibrium: Theory and Evidence." *Journal of Labor Economics* 37 (3): 853–904.

Neumark, David, J.M. Ian Salas, and William Wascher. 2014. "Revisiting the Minimum Wage—Employment Debate: Throwing Out the Baby with the Bathwater?" *Industrial and Labor Relations Review* 67 (3): 608–48.

Neumark, David, and William Wascher. 1992. "Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws." *Industrial and Labor Relations Review* 46 (1): 55–81.

Neumark, David, and William Wascher. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment." *American Economic Review* 90 (5): 1362–96.

Neumark, David, and William L. Wascher. 2008. *Minimum Wages.* Cambridge, MA: MIT Press.

OECD. 1994. *The OECD Jobs Study: Facts, Analysis, Strategies.* Paris: OECD.

OECD. 2018. "Good Jobs for All in a Changing World of Work: The OECD Jobs Strategy." Paris: OECD.

O'Neill, Donal. 2015. "Divided Opinion on the Fair Minimum Wage Act of 2013: Random or Systematic Differences." *Economics Letters* 136: 175–78.

O'Neill, Donal, Brian Nolan, and James Williams. 2006. "Evaluating the Introduction of a National Minimum Wage: Evidence from a New Survey of Firms in Ireland." *Labour* 20 (1): 63–90.

Pissarides, Christopher A. 2000. *Equilibrium Unemployment Theory.* 2nd ed. Cambridge, MA: MIT Press

Portugal, Pedro, and Ana Rute Cardoso. 2006. "Disentangling the Minimum Wage Puzzle: An Analysis of Worker Accessions and Separations." *Journal of the European Economic Association* 4 (5): 988–1013.

Rebitzer, James B., and Lowell J. Taylor. 1995. "The Consequences of Minimum Wage Laws: Some New Theoretical Ideas." *Journal of Public Economics* 56 (2): 245–55.

Schmitt, John. 2015. "Explaining the Small Employment Effects of the Minimum Wage in the United States." *Industrial Relations* 54 (4): 547–81.

Shapiro, Carl, and Joseph E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74 (3): 433–44.

Sorkin, Isaac. 2015. "Are there Long-Run effects of the Minimum Wage?" *Review of Economic Dynamics* 18 (2): 306–33.

Teulings, Coen N. 2000. "Aggregation Bias in Elasticities of Substitution and the Minimum Wage Paradox." *International Economic Review* 41 (2): 359–98.

Vaghul, Kavya, and Ben Zipperer. 2019. "Historical State and Sub-State Minimum Wages." Unpublished.

Zipperer, Ben. 2016. "Did the Minimum Wage or the Great Recession Reduce Low-Wage Employment? Comments on Clemens and Wither (2016)." https://equitablegrowth.org/working-papers/comments-on-clemens-and-wither/.

City Limits: What Do Local-Area Minimum Wages Do?

Arindrajit Dube and Attila Lindner

To date, 42 cities in the United States have instituted minimum wages above the state or federal level. Of these cities, 22 have a minimum wage that is $15 per hour or more—including San Francisco, Seattle, Los Angeles, and Washington, DC—a level that seemed unthinkable just a few years ago. Moreover, advocates for city-wide minimum wages have played an important role both by changing wages in some of the largest and most dense labor markets in the United States and by reshaping the policy terrain more broadly. In these campaigns, advocates for a higher minimum wage first shifted their focus from state and federal government to city councils, mayors, and voters (via ballot initiatives). Subsequently, legislatures in states like California and Washington responded to city-wide minimum wages by passing large state-wide increases of their own. In contrast, other state legislatures passed laws preempting a city-wide minimum wage mandate.

The growing number of cities with minimum wages naturally raises the question: is local variation in minimum wage policies a good idea? Most of the extensive

■ *Arindrajit Dube is Professor of Economics, University of Massachusetts Amherst, Amherst, Massachusetts. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and a Research Fellow, IZA Institute of Labor Economics, Bonn, Germany. Attila Lindner is Assistant Professor, University College London, London, United Kingdom. He is also an Associate, Center for Economic Performance, London School of Economics and Research Associate, Institute for Fiscal Studies, both in London, United Kingdom; Research Affiliate, IZA Institute of Labor Economics, Bonn, Germany; and Research Fellow, Institute of Economics, Hungarian Academy of Sciences, Budapest, Hungary. Their email addresses are adube@umass.edu and a.lindner@ucl.ac.uk.*

minimum wage literature to date has focused on state- or federal-level changes, but city-level minimum wage changes can have potentially different implications than changes that affect a state or the whole country. For instance, city boundaries are porous, and for many businesses it might be easy to relocate to a few miles outside of the city boundaries. This distortion may also be present to some extent for state-level minimum wages, but it could be much larger for minimum wage changes that are restricted to cities. On the other hand, local variation in minimum wages can better tailor the policy to local circumstances. For example, the level of minimum wage that might raise concerns about unintended consequences in rural areas in California may not bind at all in San Francisco or Los Angeles, given the generally higher wages in those cities.

To evaluate these trade-offs, we begin with some descriptive evidence on the evolution of city-level minimum wage policies. We examine what type of cities have instituted minimum wages and discuss how these characteristics can potentially impact the effectiveness of city-level minimum wage policies. In the next part of the paper, we summarize the evolving evidence on city-level minimum wage changes and provide some new evidence of our own. By combining the existing evidence from cities with some additional insights obtained from the literature on state- and federal-level changes, we provide an overall (if tentative) evaluation on what city-level minimum wages do. The weight of evidence suggests that city mandates (especially in larger cities) have been successful in raising wages for the bottom quartile of the wage distribution, with limited impact on employment prospects for low-wage workers. But the evidence base is still limited, and for this reason we identify some key areas where further research can be particularly helpful.[1]

Some Basic Facts

The first city-level minimum wage in the United States was instituted in 1993 in Washington, DC. However, city-level minimum wages remained a rather rare phenomenon until about seven years ago. To be sure, there were some isolated attempts and even some successes starting at the turn of the century. In 2002, New Orleans attempted to raise the minimum wage by $1 above the federal standard when a majority of voters supported it on a ballot initiative, but the state of

---

[1]While we mainly focus here on US evidence, city or local minimum wages are also present in other countries. There are 15 countries (besides the US) with some type of geographical differentiation in minimum wages: Bangladesh, China, India, Indonesia, Japan, Pakistan, Philippines, Vietnam, Portugal, Switzerland, Burundi, Canada, Malawi, Tanzania, and Kenya. Among these, Bangladesh, China, India, Indonesia, Burundi, and Kenya have city-level minimum wages. Pakistan has a different minimum wage in Islamabad Capital Territory, which is federal territory (like DC). Portugal has different minimum wage for (archipelagos) Azores and Madeira. Malawi's minimum wage differentiates between urban versus rural. Tanzania's minimum wage differentiates between mainland and (archipelago) Zanzibar. The rest of countries have state/province variation. Switzerland has two cantons (Jura and Neuchatel) with minimum wages and the canton of Geneva just passed a law to introduce one. Therefore, around 6–10 countries have city minimum wages depending on the definition (Tijdens and van Klaveren 2019).

Louisiana barred local governments from setting the minimum wage. In 2004, San Francisco and Santa Fe successfully introduced local minimum wage ordinances. Both of these cities were located in states that already had minimum wages above the federal level, but these cities decided to go further. In San Francisco's case, the policy came out of a ballot initiative backed by more than 60 percent of the voters.

By 2010, only these three cities had city-level minimum wage. Yet ten years later, city-level minimum wages had spread to 42 major cities, including New York, Chicago, Los Angeles, Seattle, Denver, and Minneapolis. Today, city-level minimum wage policies cover almost 8 percent of the US workforce. However, these 42 cities with their own minimum wages are all located in only nine states, showing significant regional concentration. Moreover, of these 42 cities, 29 are located in California, and in turn, 24 of these are cities in the San Francisco Bay Area.

This shift to cities as a focus of efforts to raise the minimum wage is not the result of major changes in the legal environment. In general, cities can pass laws on specific issues for which they have explicit permission from the state, but they can also pass laws within a broader category of issues, as long as they are not specifically preempted by state or federal laws. Instead, the recent increase in city-level minimum wages seems to reflect two developments: first, the federal minimum wage has been stagnant over the past decade, and second, even when state legislatures have enacted a higher minimum wage, it has often been below the level desired in certain high-wage, high cost-of-living cities (Rapoport 2016). These two developments pushed minimum wage advocates to move their efforts to the local level. Beginning in 2012, a coalition of unions—especially the Service Employees International Union—and progressive advocates helped launch the "Fight for Fifteen" movement. In November 2012, groups of workers from many fast food chains walked off their jobs in New York City, demanding a minimum of $15 per hour and other workplace rights. The momentum spread nationally. In 2014, SeaTac and then Seattle successfully passed ordinances mandating city-wide minimums. Other cities followed, building on these early successes and from having a national-level organizing infrastructure in place.

However, it is important to remember that states ultimately have the power to decide whether cities can institute their own minimum wage policies (Briffault 2018). As mentioned earlier, preemption legislation in Louisiana barred New Orleans from setting its own minimum wage in 2002. As another example, St. Louis, Missouri, approved a minimum wage increase in 2015, which went into effect in 2017 following nearly two years of litigation. However, the state quickly passed a new law preempting cities in Missouri from setting minimum wages. As a consequence, the higher St. Louis minimum wage was in effect for only three months. By 2018, 28 US states had preemption legislations in place, banning city-level minimum wages within their jurisdiction (EPI 2018).[2]

---

[2] One of these is Oregon, which does not allow city-level minimum wages; however, in 2016, the Oregon legislature established a three-tiered minimum wage plan. The highest minimum wage tier established

There are also some counties with minimum wages above the state level, although these counties typically either contain or adjoin a city with a higher minimum wage. Examples include Cook County in Illinois, which encompasses the city of Chicago; Montgomery County and Prince George's County in Maryland, which adjoin Washington, DC; Los Angeles County in California, which includes the city of Los Angeles; and Bernalillo County in New Mexico, which includes the city of Albuquerque. The state of New York also set a separate minimum wage for "downstate" counties near New York City (Nassau, Suffolk, and Westchester counties), and Oregon has introduced a three-tiered minimum wage where the wage floor varies across rural, non-rural, and Portland metro counties. In this paper, we will focus specifically on city-level minimum wages. However, we suspect that these county-level minimum wage changes are likely to have similar implications.

In most cases, city-level minimum wages cover nearly all low-wage workers working within the city limits. A notable exception is the city of SeaTac in Washington state, where only workers in the hospitality and transportation sector are bound by the law (the law notably excludes the SeaTac international airport, the largest employer in the city). There are also some cities with separate minimum wages for tipped workers (as in New York City). Finally, in many cases the local ordinance allows for small businesses to set somewhat lower wages.

The top panel of Table 1 shows the 10 largest cities with local minimum wage ordinances. The three largest US cities—New York, Los Angeles, and Chicago—all had city-level minimum wages in place as of January 2020. Among the top ten largest cities with their own minimum wages, four have a minimum wage that is at least $15 per hour and two other large cities are scheduled to pass the $15 per hour threshold by 2022. Currently, the highest state-level minimum wage is in Washington state at $13.50 per hour; in contrast, six of the ten largest cities with minimums have levels exceeding $13.50 per hour. At the same time, large cities also tend to be areas where wages are generally higher for everyone, and this should be taken into account when assessing the level of the minimum wage.

To better gauge the bite of the city minimum wages, we calculate the median wage for each city using the latest wave of the American Community Survey and look at the ratio of the minimum wage to median wage—the so called Kaitz index. The (unweighted) average Kaitz index in the largest ten cities is around 0.58. This average is substantially higher than the average state-level Kaitz index, which is 0.48. This implies that the top ten largest cities introducing minimum wages went substantially further in their minimum wage policies than the average US state, even after accounting for differences in the overall wage levels.

The bottom panel of Table 1 reports the ten cities with the highest nominal minimum wages. The two highest nominal minimum wage cities—Seattle and SeaTac—are both located in Washington state, while the rest of cities that made the list are all in the San Francisco Bay Area. The top ten highest nominal minimum

a wage floor for the Portland Urban Growth Boundary. This is effectively a Portland city-level minimum wage, which we include in our analysis below.

Table 1
Some Cities with Minimum Wages

Cities	Population	Minimum wage in 2020	Kaitz index	Planned nominal minimum wage in 2022
Panel A. Largest Population Cities with Minimum Wages above the State Level				
1. New York City	8,398,748	15.00	0.66	15.00
2. Los Angeles	3,990,469	14.25	0.75	15.72
3. Chicago	2,705,988	13.00	0.65	13.60
4. San Jose	1,030,119	15.25	0.56	16.20
5. San Francisco	883,305	15.59	0.45	17.05
6. Seattle	744,949	16.39	0.57	17.19
7. Denver	716,492	12.85	0.58	15.87
8. Washington, DC	702,455	14.00	0.48	14.50
9. Portland	652,573	12.50	0.56	14.75
10. Albuquerque	560,234	9.35	0.55	9.60
Mean of top 10				
Unweighted	2,038,533	13.82	0.58	14.95
Population weighted		14.33	0.64	15.04
Panel B. Highest City-Level Minimum Wages (Nominal Minimum Wage in 2020)				
1. Seattle	744,949	16.39	0.57	17.19
2. SeaTac*	28,925	16.34	0.67	16.79
3. Emeryville	11,724	16.30	0.65	17.92
4. Mountain View	83,377	16.05	0.34	17.05
5. Sunnyvale	153,175	16.05	0.39	17.05
6. Berkeley	121,654	15.59	0.60	17.15
7. San Francisco	883,305	15.59	0.45	17.05
8. Los Altos	30,588	15.40	0.33	16.40
9. Palo Alto	66,655	15.40	0.33	15.85
10. Santa Clara	129,489	15.40	0.43	15.85
Mean of top 10				
Unweighted	225,384	15.85	0.48	16.83
Population weighted		15.89	0.49	16.98

Note: Kaitz index is the minimum wage divided by the median wage. The median wages of all workers are calculated from the 2018 wave of the American Community Survey and are measured in 2020 dollar value.

* Minimum wage only applies to transportation and hospitality workers within SeaTac city. We report the city-level Kaitz index, where we calculate the industry share weighted average of the minimum to median wage.

wage cities are on average small (average population is around 230,000). Among the three cities that have a long tradition of minimum wages—Washington, DC, Santa Fe, San Francisco—only San Francisco is in the top ten list for cities with the highest minimum wage.

All of the cities in Panel B of Table 1 have minimum wages exceeding $15 per hour, but these cities also have high wages, generally. As a result, in some cases the Kaitz index is rather modest: for example, the cities of Los Altos and Palo Alto in

the San Francisco Bay Area have a Kaitz index of only 0.33—which is lower than the current Kaitz index of the federal minimum wage of 0.37. The average Kaitz index among the top ten highest nominal minimum wage cities is 0.48, which is the same as the average Kaitz index among the US states. This highlights that top-line nominal minimum wage numbers can provide a misleading picture of how local minimum wage policies may affect a local economy.

Table 2 summarizes the basic characteristics of all 42 cities with minimum wages as of January 2020. We calculate city-level characteristics using the 2018 American Community Survey, the most recent data available. The first two columns report statistics for cities with minimum wages. The first column shows population weighted averages for cities with less than 100,000 residents, while the second column shows the statistics for cities with more than 100,000 residents as of 2018. For comparison, the third column reports the same statistics for all US cities with at least 100,000 residents, but no city-wide minimum wages.

As expected, the nominal minimum wage is substantially higher in the cities with a local minimum wage (columns 1–2), than in the cities where only the state or the federal minimum wage applies (column 3). The difference in the nominal level of the minimum wage is substantial—around $5 (or 50 percent). However, minimum wage cities also have around 25–75 percent larger average and median wages than other cities. As a result, the economic bite of the policy is substantially smaller than the headline nominal numbers would indicate. For smaller cities, the minimum-to-median-wage ratio is very similar to other cities without any minimum wage (0.50 in column 1 versus 0.52 in column 3). The difference in the minimum-to-median-wage ratio is sizable when we compare larger cities with and without city-wide minimums (0.63 in column 2 versus 0.52 in column 3). Furthermore, the cost of living is also much higher in minimum wage cities; using the regional price parities measured at the level of metropolitan statistical areas (MSAs), we estimate that minimum wage cities have around 16 percent higher cost of living.[3] Accounting for cost of living suggests that the real value of the minimum wages in larger cities with ordinances is around 25–30 percent larger than in cities without.

Table 2 also highlights that in cities with minimum wages, the population has higher levels of education, and workers are more likely to be employed in high-paying industries such as professional services. Furthermore, and not surprisingly, cities with minimum wages are more likely to lean Democrat: in the 2016 presidential election, for example, 76 percent of the cities with their own minimum wage voted for Democrats, while in other larger cities without minimum wages, the vote share was only 55 percent. Finally, the local unemployment rate seems to be very similar between cities with and without minimums, at least for cities with a population exceeding 100,000.

---

[3] Note that the differences in cost of living are at the level of metropolitan statistical areas (MSAs) and not at the city level. That is why the differences in median and average wages are substantially larger than the differences in cost of living. This also implies that we may be underestimating the differences in cost of living across cities.

Table 2
Basic Characteristics of Cities with and without Minimum Wages

	Cities with a minimum wage		Cities without a minimum wage
	Pop < 100,000 (1)	Pop > 100,000 (2)	Pop > 100,000 (3)
Number of cities	20	22	249
Population (in thousands)	55.2	1,034.4	266.9
Nominal minimum wage in 2020	14.57	14.27	9.44
Planned minimum wage by 2022	15.71	15.07	
Mean wage	42.31	31.42	24.58
Median wage	31.26	22.47	18.04
Cost of living index (RPP)	122.9	117.1	100.2
Minimum wage to mean wage	0.36	0.45	0.38
Minimum wage to median wage	0.50	0.63	0.52
Share Democrats	0.73	0.76	0.55
College share	0.47	0.38	0.30
Unemployment rate	3.84	5.61	5.45
Industry shares:			
Restaurants	0.06	0.07	0.08
Retail	0.09	0.09	0.11
Manufacturing	0.09	0.06	0.08
Construction	0.05	0.05	0.06
Health and social care	0.12	0.14	0.13
Professional services	0.15	0.11	0.07

Note: Own calculations based on the 2018 American Community Survey. Cost of living index is the MSA level Regional Price Parity (RPP) measured in 2017. The share of Democrats in the 2016 presidential election is obtained from McGovern (2016). Each row (except the one on population) reports population weighted averages.

City-level minimum wages are above state-level minimum wages, but the cities with higher minimum wages also tend to have median wages above the state level. Figure 1 shows the distribution of the Kaitz index (ratio of minimum to median wage) for cities with minimum wages as well as the comparable distribution of state-level minimums. For the states, we only consider minimum wages when they are above the federal level. Many cities (shown by the blue line) went beyond the highest state-level Kaitz: they have higher levels of minimum wages even after differences in the median wage across locations are taken into account. On the other hand, there are many high-wage cities where the Kaitz index is quite low even with the higher nominal minimum wages. The figure shows that the Kaitz index is more dispersed for cities than for states.

If we exclude the cities in the San Francisco Bay Area with their high levels of median wages (as shown in Table 1 above), then dispersion in the city-level Kaitz indices is more comparable to the state-level one, though the average value of the Kaitz index is considerably higher in cities. Notably, the Kaitz index exceeds 0.65 in a substantial portion of cities, which it essentially never does at the state level.

Figure 1
Distribution of Kaitz Index for US Cities and States
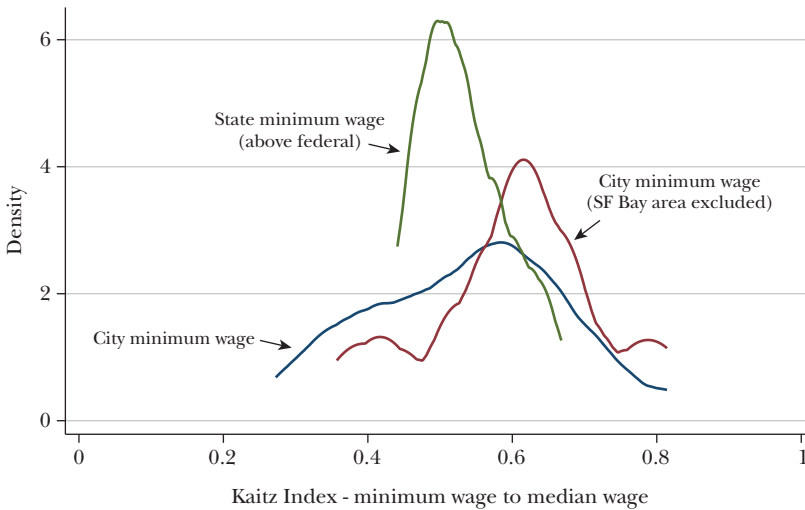
Note: The figure shows the distribution of the Kaitz index (minimum wage to median wage) for cities with minimum wages (blue and red line) and for all US states where a minimum wage above the federal one applied (green line).

To summarize, cities passing local minimum wages are typically large, with higher overall wages and cost of living. However, even after accounting for these differences, the city-wide minimum wages appear to have pushed the wage standards to be more binding than they would have from state-level policies alone.

Arguments Concerning City-Level Minimum Wages

The prevalence of city minimum wages naturally raises the question: is the growing variation in local-level minimum wages desirable? Here, we review some of the main arguments concerning minimum wages in general and consider how they apply in the context of city-level minimum wages. To do so, we assess the tradeoffs that emerge for any place-based policies (Kline and Moretti 2014). In the next section, we consider the empirical evidence for these arguments.

First, one of the oldest arguments for a minimum wage is that someone who works full-time should be able to afford the basic cost of living. However, there are large differences in costs of living across the United States (Albouy 2009). As Table 2 showed, the cities that have enacted city-wide minimum wages had, on average, a 17 percent higher cost of living as measured by the regional price parity index (at the level of metropolitan statistical areas) than other cities. A local-level minimum wage can be adjusted to take into account that workers with the same nominal wage are substantially "poorer" in locations with high costs of living.

Second, local variation in minimum wages may serve to redistribute resources from higher-income consumers to lower-wage workers. A body of empirical studies suggests that minimum wages are passed on to consumers via higher output prices (Lemos 2008; MaCurdy 2015; Harasztosi and Lindner 2019). Because most minimum wage workers are employed in local non-tradable sectors (like restaurants or retail stores), this redistribution mainly takes place among local consumers and local minimum wage workers. As shown in Table 2, cities with minimum wages have a larger share of high-educated workers, a larger share of workforce in the professional services sector, and therefore a higher-income consumer base. In these types of cities, redistribution from local high-income consumers to local low-wage workers may be more desirable.[4]

Third, another justification for minimum wages in general is that if employers have market power in the labor market, and they create a wedge between the marginal product of labor and wages, then minimum wages can potentially push wages and employment closer to the competitive equilibrium. However, the level of employer market power varies by local areas. Azar, Marinescu, and Steinbaum (2019) find that labor market concentration in the general merchandise sector varies considerably across locations and that the employment response to the minimum wages is linked to this variation.

Fourth, the minimum wage may shift the composition of local jobs. Harasztosi and Lindner (2019) and Cengiz et al. (2019) document considerable negative effects of state-level or country-level minimum wages on jobs in the tradable sector. Aaronson and Phelan (2019) find a drop in cognitive routine occupations after minimum wage hikes, while Lordan and Neumark (2018) document a drop in automatable jobs. However, in practice, such considerations do not seem to play a major role in the current context of city-level minimum wages. As shown in Table 2, the share of non-tradable sectors such as restaurants and retail or the share of tradable jobs such as manufacturing are similar in cities with and without minimum wages.

Fifth, a central concern expressed about the minimum wage is that it could reduce employment, either by causing employers to reduce the number of employees or by causing them to move out of the jurisdiction where the higher minimum wage applies. These employment and wage responses for localized minimum wage changes may differ from state or federal-level ones. After all, city boundaries are more porous than state boundaries. Businesses may simply move a few miles away to avoid minimum wage changes. Workers may seek better employment opportunities by changing their commuting patterns. In general, given the density of highways in commuting zones, labor mobility is much greater across cities than across states. As a result, it is important to assess the effect of the minimum wage directly on both employment and business reallocation across city boundaries. However, it is also important to consider whether a reallocation of businesses and

---

[4]Diamond (2016) shows that high-skilled workers do not just get higher wage premium in some cities, but they also enjoy higher amenities. This would provide an additional reason to redistribute resources from high-skilled workers to lower-skilled ones in those cities.

jobs from some high-wage cities to outside their boundaries is necessarily bad from the public perspective (Albouy 2009). If a higher minimum wage does, indeed, lead to reallocation, the creation of new jobs outside of the urban core may help relatively disadvantaged areas outside city limits.

Finally, variation in minimum wages across cities can better reflect the preferences of citizens (Tiebout 1956). Residents of some cities may prefer higher minimum wages even if the policy is accompanied by various trade-offs. Table 2 highlights considerable differences between the electorates in cities with and without minimum wages in terms of supporting major political parties. Recent survey evidence by Simonovits and Payson (2020) suggests that there is a strong correlation between city-level preferences and the prevailing minimum wage.

Evidence on the Impact of City-Level Minimum Wages

There is an extensive literature studying the impact of state- and federal-level minimum wage changes, but the existing evidence on city-level minimum wage laws is limited. Here, we consider the evidence (or sometimes the lack of evidence) on the effect of city-level minimum wages.

Estimates for Employment and Wages

Three studies provide evidence on the early wave of city minimum wages. Dube, Naidu, and Reich (2007) study the effect of introducing the minimum wage in San Francisco in 2004 using two waves of a survey of restaurants and using aggregate level data from the Quarterly Census of Employment and Wages. They use a difference-in-differences approach with a variety of control groups, including firms outside of San Francisco, smaller firms unaffected by the wage mandate within San Francisco, and higher-wage firms within San Francisco. They find that the policy increased worker pay and compressed wage inequality but did not create any detectable employment loss among affected restaurants. Potter (2006) focuses on another early example of the city-level minimum wage changes: Santa Fe, New Mexico. Based primarily on comparisons with patterns in Albuquerque (about 60 miles away), Potter shows that the 65 percent increase in the minimum wage in 2004 did not had a negative impact on employment—if anything, Santa Fe actually did better than Albuquerque.

Schmitt and Rosnick (2011) study the impact of the minimum wage in three cities using firm-level administrative data from Quarterly Census of Employment and Wages: San Francisco, Santa Fe, and Washington, DC. They find that average earnings increased in San Francisco and Santa Fe, but not in Washington, DC. They, too, use a difference-in-differences approach using alternative control groups (similar to Dube, Naidu, and Reich 2007). Their estimates on employment vary considerably across specifications, making it difficult to draw a definitive conclusion. Nonetheless, the estimates are clustered around zero—suggesting that the impact on employment was likely limited.

This early consensus on the effects of city-level minimum wage changes has been challenged recently by an influential study from Seattle. Jardim et al. (2017) study the introduction of the Seattle Minimum Wage Ordinance, which raised the minimum wage from $9.47 to $13 per hour in 2016. The study makes an important improvement relative to existing evidence as it utilizes high-quality administrative data on hourly wages. The paper documents a dramatic drop in the number of jobs below $25 per hour in Seattle relative to the other areas in Washington state. Their point estimates on employment elasticity with respect to own wage (in a competitive model, this is the elasticity of labor demand) is –2.18. Such an elasticity is outside of the range of existing estimates in the literature exploiting state- or country-level variation in the minimum wage (see Figure 4B of Dube 2019) and suggests that the policy did considerable harm to low-wage workers in Seattle.

The Seattle study received considerable attention. Jardim et al. (2017) use a credible empirical strategy that created a synthetic control for Seattle from other cities in Washington and combined it with a unique administrative data on hourly wages. Nevertheless, certain features of the Seattle experiment should lead us to a cautious interpretation of the findings. First, it turned out that the Seattle labor market evolved quite differently than the areas in the comparison group around the time of the introduction of the local ordinance, with a substantial increase in the number of jobs and wages, especially at the top of the wage distribution in Seattle. Because it is unlikely that the minimum wage has a substantial impact on jobs at the top of the wage distribution,[5] such divergence between Seattle and the comparison group suggests that other shocks also affected the Seattle labor market around the policy change. Indeed, the "Seattle boom" might have shifted the whole wage distribution in a way that led low-wage jobs to disappear at the same time as more high-wage jobs were created. While the authors are careful in constructing a control group, given the generally greater wage growth in major cities during this period, it may just not be feasible to construct a counterfactual using places in Washington state outside of Seattle (which is the data the authors are using).

Furthermore, in a follow-up paper, Jardim et al. (2018) examine the employment trajectories of workers with jobs before the introduction of the minimum wage. The employment estimates for that subgroup are substantially lower: the implied employment elasticity with respect to own wage is 0.03, and the confidence intervals rule out even moderate-sized disemployment effects.[6] While these estimates do not

---

[5] The neoclassical model does predict that low-skilled workers will be replaced by high-skilled ones in response to the minimum wage. However, because the share of minimum wage workers in total production is low, we expect limited effects on the upper tail employment under reasonable values of labor-labor substitution (for details, see Cengiz et al. 2019, Appendix B). Therefore, the overall increase in employment (relative to the synthetic control) suggests that other major shocks around the time of the reform played an important role.

[6] Jardim et al. (2018) report separate estimates on the effect of the minimum wage on total hours and on employment. We focus here on the head count estimates as those are more comparable to the existing literature. Jardim et al. (2018) find a significant drop in total hours, which amplifies the negative consequences of minimum wage changes. We discuss the change-in-hours results below.

take into account the potential for a drop in new entrants to the labor market, they are also less affected by the overall shift of the wage distribution. As a result, it is unclear whether these estimates are biased upward or downward.

In a study on the impact of city-level minimum wages on employment in the restaurant sector in six large cities, Allegretto et al. (2018) use data from the Quarterly Census of Employment and Wages aggregated at the county-by-industry level. While their analysis is based on less rich data than the Seattle study by Jardim et al. (2017), they can use all counties without minimum wages to find the best comparison group. Given that the cities with minimum wages are quite unique, it may be important to go outside of a given state (like Washington) to find a better comparison group. Allegretto et al. (2018) find considerable increases in wages and disemployment effects that are modest and statistically insignificant. Interestingly, Allegretto et al. also study the employment changes in Seattle and find no indication for negative disemployment effects in the restaurant sector.[7]

Table 3 summarizes some key estimates in the literature on the impact of the city-level minimum wages on own wages and employment. Because it is hard to interpret the findings on employment in the absence of any wage responses to the policy, we only report estimates with statistically significant effects of the minimum wage on wages. Column 5 reports the employment elasticity with respect to own wage.

Table 3 highlights that the employment elasticity estimates are centered around zero, which suggests that city-level minimum wages have no discernible effect on employment. Out of the eleven estimates, seven have positive point estimates and four have negative sign for employment. Only two point estimates have an own-wage employment elasticity less than minus one, thereby implying that the total wage bill collected by low-wage workers falls after the policy change as a result of job losses. Nevertheless, individual estimates are quite noisy even if we consider the 90 percent confidence intervals. Only two estimates can rule out that the employment is unaffected by the policy: Jardim et al. (2017) aggregate-level one finds a statistically significant negative employment effect, while Allegretto et al. (2018) in Oakland finds a statistically significant positive estimate on employment. Six estimates in the literature can rule out large negative employment effects (employment elasticity is less than –1) and three estimates can rule out medium-sized negative employment responses (employment elasticity is less than –0.4).

Overall, the evidence on city-level minimum wages is consistent with the growing body of evidence suggesting that moderate levels of minimum wage increases have a relatively small employment effect on the affected workers (Cengiz et al. 2019; Belman and Wolfson 2014). At the same time, it is important to point out that there

---

[7] Jardim et al. (2017) also provide separate estimates for restaurants. Similarly to Allegretto et al. (2018), they confirm that the overall number of jobs did not fall in the restaurant sector. At the same time, they find some drop in employment for jobs below $25 per hour, which they interpret as substitution away from low-wage workers. However, Nadler et al. (2019) show that small industry-wide employment elasticities are inconsistent with highly elastic labor demand for low-wage workers, given plausible elasticities of substitution across skill groups.

Table 3

Existing Estimates on City-level Minimum Wage Changes

Paper	City	Wage	Employment	Own-Wage Elasticity
Allegretto et al. (2018)—restaurants	Average of 6 cities	0.02 [0.01,0.03]	−0.01 [−0.02,0.01]	−0.23 [−0.78,0.32]
	Oakland	0.10 [0.06,0.14]	0.07 [0.03,0.11]	0.71 [0.20,1.22]
	San Francisco	0.06 [0.04,0.09]	0.01 [−0.05,0.07]	0.14 [−0.83,1.11]
	San Jose	0.11 [0.06,0.15]	0.00 [−0.06,0.06]	−0.02 [−0.57,0.53]
	Seattle	0.04 [0.02,0.07]	0.01 [−0.05,0.07]	0.20 [−1.16,1.57]
Dube, Naidu, and Reich (2007)—restaurants	San Francisco	0.14 [0.06,0.22]	0.04 [−0.12,0.2]	0.29 [−0.34,0.91]
Jardim et al. (2017, 2018)—jobs below $19	Seattle, worker level	0.15 [0.14,0.17]	0.01 [−0.01,0.02]	0.03 [−0.04,0.11]
	Seattle, aggregate level	0.03 [0.03,0.03]	−0.07 [−0.14,−0.01]	−2.18 [−4.14,−0.22]
Moe, Parrott, and Lathrop (2019)—full service restaurants	New York City	0.09 [0.03,0.16]	0.03 [−0.16,0.22]	0.29 [−1.74,2.32]
Schmitt and Rosnick (2011)—fast food	San Francisco	0.10 [0.05,0.14]	0.00 [−0.33,0.34]	0.03 [−3.45,3.5]
	Santa Fe	0.07 [0.02,0.12]	−0.08 [−0.29,0.13]	−1.20 [−4.36,1.96]

Note: We report the estimated impact of city-level minimum wages on wages (column 3), on employment (column 4), and on employment elasticity with respect to own wage – the labor demand elasticity in the competitive model (column 5). We only report estimates where positive wage effects have been detected in the data. When the elasticity with respect to own wage was not directly reported (Allegretto et al. 2018; Dube, Naidu, and Reich 2007, Jardim et al. 2017, 2018), we calculated the elasticity by dividing the employment effect with the wage effect. The corresponding standard errors were calculated by using the delta method. All estimates report the 90 percent confidence intervals. We calculate the standard errors in Moe, Parrott, and Lathrop (2019) using Randomization Inference.

is sizable uncertainty around the existing estimates. There is plenty of room for additional research to glean important information on this question.

Overall Impact of City-Wide Minimum Wages

Almost all studies on the impact of city-level minimum wage changes focus on a particular city and a particular minimum wage hike. However, inferences based on any particular case study is inherently difficult. Furthermore, estimates based on

any single experiment may be easily contaminated by other shocks, as is possibly the case for the Seattle study.

The Allegretto et al. (2018) study mentioned previously is an exception to the single-city approach, because it reports event-study estimates exploiting six prominent minimum wage hikes. But many more city-level minimum wage changes could be used for identification. In fact, studies on the combined evaluation of city-level minimum wage changes are noticeably missing. This is in stark contrast to the literature on state-level minimum wage changes, which moved some years ago from a case-by-case analysis (as in the work of Card and Krueger 1994) to combined evaluation of many state-level minimum wage shocks (for example, Neumark and Wascher 1992; Dube, Lester, and Reich 2010; Cengiz et al. 2019). Moreover, Allegretto et al. (2018) only look at restaurants, and therefore, cannot fully resolve the concerns raised by Jardim et al. (2017).

Here we make a first effort at filling this hole in the literature by providing an overall assessment of the city-level minimum wage changes instituted as of 2018. We have three specific objectives. First, we wish to evaluate the full set of major cities instituting or raising the city-wide minimums during the recent period. Second, we wish to consider the *overall* impact of these policies on low-wage jobs in the spirit of Cengiz et al. (2019) and Jardim et al. (2017), as opposed to focusing on particular sectors or groups. Third, having shown how selective the minimum wage cities are, we wish to address concerns about invalid counterfactuals by comparing these cities with other large cities with similar characteristics that did not raise their minimum pay standard.

For this exercise, we use the American Community Survey between 2012 and 2018 that has data on wages and employment at the city level. This dataset provides the geographic granularity needed to home in on specific cities, something not feasible with other publicly available datasets. We start with a sample of all cities with a population of at least 100,000 in 2018 (last year of our sample). This leaves us with 21 cities with minimum wage changes.

We estimate the following regression using samples from the American Community Survey from 2012, 2013, 2017 and 2018:

$$y_{ct} = \beta_0 + \beta_1 (\mathit{Treat}_c \times \mathit{Post}_t) + \beta_2 (X_{c,2012} \times \mathit{Post}_t) + \mu_c + \tau_t + e_{ct}.$$

The left-hand variable is the main outcome of interest—for us, either wage percentile or employment—at city c at time t. Treat_c is a dummy variable for cities with minimum wage as of 2018. Post_t is a dummy for years 2017 and 2018. Finally, μ_c are city fixed effects (thus adjusting for time-invariant unobserved heterogeneity) and τ_t controls for year effects.

Of course, a central issue in the minimum wage literature arises because jurisdictions which enact a higher minimum wage are not chosen exogenously or at random. Given that we expect differences between cities that introduced minimum wages and those that did not, we control for the interaction of a dummy for *Post* with a set of 2012 covariates $X_{c,2012}$ on cost of living, employment to population

ratio, average wage, the 10th, 25th, 50th, 75th, and 90th wage percentiles, shares of employment below wage cutoffs ($15, $20, $25, $30, $50), and sectoral shares measured at the one-digit level. As we will see below, controlling for pretreatment city characteristics produces much more sensible results on the upper tail of the wage distribution, which can be considered a key falsification test for both wage and employment effects (as discussed in Autor, Manning and Smith 2016; Cengiz et al. 2019). We weight the regressions by population of the city and cluster the standard errors at the state level.

Figure 2 studies the contribution of city minimum wages to inequality in spirit of Autor, Manning, and Smith (2016). We report the estimated β_1 coefficient from the regression where the outcome variable is various percentiles of income measured in log hourly wages. We compute hourly wages as (annual) salary income divided by hours per week times number of weeks worked.[8]

Panel A in Figure 2 shows the estimates without controlling for the covariates $X_{c,2012}$ in the regression. As expected, there is a clear increase in wages at the bottom of the wage distribution in cities with minimum wages relative to the cities without minimum wages. However, wages *also* increase significantly for all percentiles (including the very top) throughout the wage distribution. Because minimum wages are unlikely to have much effect on wages at the very top, the no-controls results here highlight that comparing cities with and without minimum wages can lead to misleading results. Instead, a plausible interpretation would be that cities experiencing wage growth across the income distribution may be more likely to enact their own minimum wage laws.

The second panel in Figure 2 controls for baseline differences in cities with a minimum wage, interacted with a post-treatment dummy. The results with controls show a clear change at the bottom of the wage distribution that fades out around the 30th percentile of the wage distribution. Such spillover effects are broadly similar to the estimates in Autor, Manning, and Smith (2016) who find a similar pattern for state-level minimum wage changes.

The evidence is consistent with a belief that the city-level minimum wages affect workers' pay at the bottom of the wage distribution and have compressed wage inequality. However, the magnitude seems modest. If we estimate the same regression using the log of the minimum wage as an outcome variable, we find that the minimum wages increased by 23 percent (with a standard error of 1.6 percent) more in cities with a minimum wage throughout this period. This increase is substantially larger than the roughly 4 percent increase in wages at the bottom of the distribution. This discrepancy likely reflects that the increasing tightness across all labor markets during this time period additionally led to wage growth even in cities that did not increase the minimum wage. This limited how binding these minimum

---

[8]The specific variables that we use from the American Community Survey are: "Salary income in the last 12 months"; "Usual hours worked per week in the last 12 months"; "Weeks worked during the last 12 months." Given that the weeks variable is coded using intervals (in six categories), we take the midpoint of each interval in all categories but the last one (50 to 52 weeks) where we assign a value of 52.

Figure 2
City-Level Minimum Wages and Inequality

A: Without controls

B: With controls
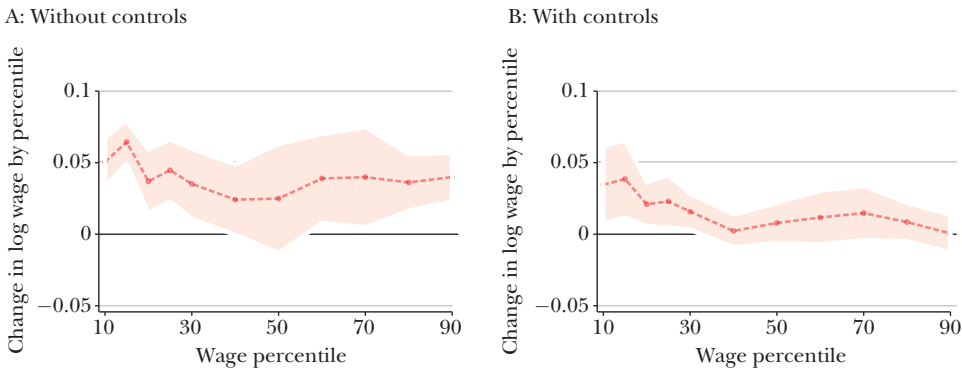
Note: This figure shows the change in log wages for each wage percentile from our regression analysis (see the equation in the text) exploiting 21 city-level minimum wage changes between 2012 and 2018. The shadowed area shows, for each percentile, the 95 percent confidence intervals around the estimate. Panel A shows the estimates with time- and city-fixed effects but without controlling for the set of 2012 covariates interacted with the post dummy. Panel B controls for 2012 values of cost of living, employment to population ratio, average wage, wage percentiles, shares of employment below wage cutoffs, and 1-digit level sectoral shares. Results are weighted by the population size of the city. For detailed regression results, see the online Appendix available at the *JEP* website with this paper.

wage changes were, and thereby, attenuated the inequality-reducing impact of the city minimum wage policies.

Did cities adopting a minimum wage experience both wage growth and weaker job growth at the bottom of the wage distribution? We assess the employment effects of the minimum wage with the distributional approach developed in Cengiz et al. (2019), which divides the wage distribution into a set of "bins." By studying the effect of the minimum wage on employment for each wage "bin" separately, we can calculate the missing number of jobs at the bottom of the wage distribution and compare it to the excess number of jobs higher up in the wage distribution.[9] This approach also allows us to study the changes in the upper part of the wage distribution. Large changes there would suggest that the estimates for the impact of the city-level minimum wages are potentially contaminated by other shocks.

Panel A in Figure 3 shows the bin-by-bin employment estimates from our earlier regression equation without controls. There is a clear drop in employment at the bottom of the wage distribution (jobs under $10) in cities with minimum wage, which is in line with a binding minimum wage policy. In addition, there is no apparent increase in the number of jobs higher up in the wage distribution, except

[9] This approach is also closely related to the Jardim et al. (2017) aggregate estimate, where they consider changes in employment below thresholds (for example, $19 per hour). This allows us to consider how similar the findings are when we pool across multiple minimum wage events and also when we use other large cities as controls (instead of rural and suburban Washington state).

Figure 3
City-Level Minimum Wages and Employment Changes

A: Without controls

B: With controls

| %Δ wage = 0.05 (s.e. 0.01) |
| %Δ employment = −0.06 (s.e. 0.03) |
| Own-wage elasticity = −1.10 (s.e. 0.55) |

| %Δ wage = 0.04 (s.e. 0.01) |
| %Δ employment = −0.00 (s.e. 0.01) |
| Own-wage elasticity = −0.12 (s.e. 0.38) |

Note: The figure shows the bin-by-bin employment changes from our regression analysis (based on the equation in the text) exploiting 21 city-level minimum wage changes between 2012 and 2018. The blue bars show, for each wage bin, the estimated average employment change in that bin relative to the total employment in the city in 2012. The error bars show the 95 percent confidence intervals. The red line shows the running sum of employment changes up to the wage bin to which it corresponds. Panel A shows the estimates with time- and city-fixed effects but without controlling for the set of 2012 covariates interacted with the post dummy. Panel B controls for 2012 values of cost of living, employment to population ratio, average wage, wage percentiles, shares of employment below wage cutoffs, and 1-digit level sectoral shares. Results are weighted by the population size of the city. For detailed regression results, see the online Appendix available at the *JEP* website with this paper.
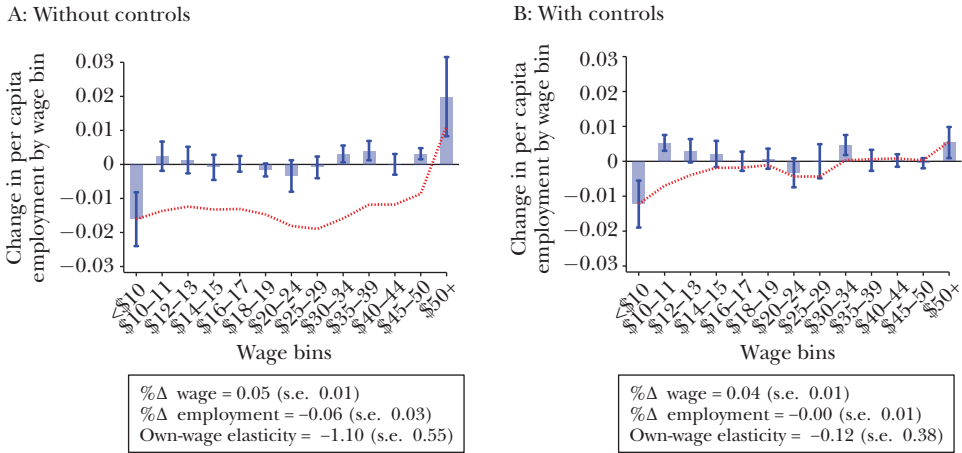
at the very top where there is a large increase in the number of jobs. The missing number of jobs under $10 per hour only recovers once jobs above $50 per hour are incorporated. In fact, overall employment increased in cities with minimum wage, even if at the bottom of the wage distribution there are large job losses. When we consider jobs up to $20 per hour, Panel A suggests that wages for this group of workers rose by around 5 percent while their employment fell by around 6 percent. The implied disemployment is quite pronounced: the estimated own-wage employment elasticity of −1.10 is statistically significant at the 95 percent confidence level.[10]

[10] We calculate the percentage change in employment and wages as in Cengiz et al. (2019). In particular, the percentage change in affected employment is the change in employment below $19 (relative to pretreatment total employment) divided by the (sample average) share of workforce below the new minimum wage. To calculate the wage changes, we use equation 2 in Cengiz et al. (2019); see the online Appendix for details. It is worth mentioning that Jardim et al. (2017) calculate the employment elasticity somewhat differently: they divide the percentage change in employment below $19 per hour by the percentage change in average wage below $19 per hour. This approach dilutes the wage effects, since the change in wages of the workers close to the $13 per hour minimum wage is compared to higher

However, the story is very different when we control for the differences in baseline characteristics across cities. In Panel B, Figure 3 shows that once we control for observable baseline differences, the dramatic change in the upper part of the wage distribution disappears. We continue to find that cities with minimum wages have some missing jobs under $10 per hour, but once we control for the baseline characteristics, we find that excess number of jobs emerge at jobs between $11 and $19 per hour. The upper part of the wage distribution is more or less stable at higher parts of the wage distribution, which is consistent with a relatively low impact of the minimum wage at that part of the wage distribution. Our estimates suggest that affected workers experienced a 4 percent additional wage gain, but the employment changes were negligible. The implied employment elasticity with respect to wage is –0.12. The 90 percent confidence interval rules out own-wage employment elasticities more negative than –0.75 (including the point estimate of –1.1 from the specification without controls).

These estimates are quite similar to the overall minimum wage literature to date. For example, the median own-wage employment elasticity in the literature is around –0.17, while it is around –0.04 when restricting attention to broad-based groups (Dube 2019). At the same time, the confidence interval here also rules out some other prominent negative estimates from the minimum wage literature. Importantly, the aggregate own-wage employment elasticity of –2.18 in the Jardim et al. (2017) study of Seattle lies far outside of our confidence interval.

Indeed, the differences between the two panels in Figure 3 can help shed light on the controversy surrounding the Seattle minimum wage studies. The findings in Panel A are strikingly similar to the aggregate-level findings in Jardim et al. (2020, see Appendix Figure 7). In Seattle, too, there was an apparent drop in jobs below the new minimum wage and those jobs did not recover if only jobs below a certain threshold (say, $20, $25, or $30 per hour) are considered. Nevertheless, similar to our results here, Jardim et al. (2020) find an overall increase in jobs in Seattle that mainly came from an unusual job creation above $50 per hour. These employment patterns are observed even though Jardim et al. (2020) are careful to construct a synthetic control; however, as we pointed out before, all of their control areas come from within Washington state. The raw-versus-control comparisons in Figures 2 and 3 document that the cities with minimum wages are often unique in terms of economic structure, costs of living, and wage and employment growth trends, and in general, it might be difficult to find comparable cities within a state with similar characteristics.

The analysis here shows that the inclusion of the full set of controls produces much better-behaved findings when it comes to the upper tail falsification tests, while also suggesting relatively modest impacts on affected employment. However, the results here suggest that the parallel trends assumption appears to hold only

---

wage workers earning just below $20 per hour. As a result, the Jardim et al. (2017) estimates overstate the employment elasticity. If we calculate the employment elasticity using their approach, we get –1.65 (with a standard error of 0.84), qualitatively similar to their estimate for Seattle of –2.18.

conditionally on covariates, highlighting the systematic differences between cities with and without minimum wages. This is different from the findings from state-level policies, where state-level estimates are not sensitive to exclusion of additional controls for time-varying heterogeneity (Cengiz et al. 2019). This naturally raises the question: how sensitive are these results to the particular set of controls included here? In the online Appendix, we show that the estimates are very similar when controlling for a small number of city characteristics chosen using a data-driven procedure, specifically the double-selection post-LASSO proposed by Belloni, Chernozhukov, and Hansen (2014). In particular—and as discussed above—treated cities had a much higher share of workers in professional services in the pretreatment period and not accounting for this particular difference seems to impart a large amount of bias.

As always in the minimum-wage literature, the key to assessing the effects of the minimum wage is to find a credible comparison group, and the selectivity of cities makes this a more difficult challenge than at the state level. However, the existing evidence, including the results presented here, does not indicate that city-wide minimum wages differ substantially from state-level ones in terms of wage and employment responses. Besides the effect of city-level minimum wages on inequality, wages, and employment, it is worth considering evidence on other aspects of these minimum wage policies—which we do in the remainder of this section.

Geographic Reallocation

Businesses may be able to avoid city-level minimum wages by shifting their production outside city boundaries. Such a shift in employment would create wage and employment spillovers in neighboring cities and counties. However, we are not aware of any studies that directly assess the presence of such spillover effects in the city minimum wage context.

If the effect of the minimum wage spills over to its own suburbs, but not to nearby cities, we would expect that the wage effects are smaller in the own-suburb estimates and the employment effects are larger than the estimates on nearby cities. There is no such tendency found in Schmitt and Rosnick (2011). Furthermore, the fact that much of the existing estimates of the effect of a city-level minimum wage on employment are centered around zero suggests that business reallocation must be limited. However, more precise documentation by future researchers on when such spillovers occur and how large they are would be useful.

Firm Entry and Exit

At least in theory, city-level minimum wage policies could affect the rates of firm entry and exit. The existing evidence on firm's closure is inconclusive. Dube, Naidu, and Reich (2007) do not detect any increase in the rate of business closure. On the other hand, Jardim and van Inwegen (2019) find that the Seattle Minimum Wage Ordinance accelerated exit of firms with a higher share of low-wage jobs. Luca and Luca (2019) exploit Yelp data to show that the exit rate of firms increased in response to the minimum wage, especially for those firms providing low quality

services (measured by low Yelp ratings on the website). Such increase in business exit rate might reflect within-city reallocation of workers from lower-paying, lower-quality firms to higher-paying, higher-quality ones—a channel that is found to be important in responding to the introduction of the minimum wage in Germany (Dustmann et al. 2020).

The evidence on firm entry in the context of city-level minimum wages is even more limited. Jardim and van Inwegen (2019) find no effect on city-level minimum wage policies on the overall rate of business entry, though they document a change in the composition of the entering firms towards less labor-intensive businesses.

Hours Worked and Other Benefits

Even if a city-level minimum wage policy does not affect the overall number of jobs, it might potentially affect hours worked or other employment benefits. In their analysis of the Seattle data, Jardim et al. (2017) find a substantial decrease in hours worked for jobs below $19 per hour. As discussed above, it is unclear whether the drop in hours reflects the shift of the wage distribution discussed above or the genuine effect of the minimum wage. Looking at earlier evidence, and contrary to the findings in Seattle, Dube, Naidu and Reich (2007) find (if anything) a positive effect of a city-level minimum wage on hours worked.

The discrepancy between these two studies may be explained, at least in part, by the different data sources used in the analysis. Jardim et al. (2017) exploit administrative data on hours, while Dube, Naidu, and Reich (2007) rely on survey data. It is possible that some firms avoid compliance to the minimum wage by under-reporting hours, a practice that was found to be important in Germany (Caliendo et al. 2018). Such underreporting might affect the results based on administrative data sets but not the results based on survey data. Dube, Naidu, and Reich (2007) also study whether the impact of a pay increase resulting from a higher city-level minimum wage is offset by cutting non-cash benefits. They find no indication for cutting health insurance benefits and document an increase in the proportion of workers receiving tips.

Output Prices

A key channel of absorption of minimum wages is passing prices through to consumers. However, if the city-level minimum wage only applies to a subset of an integrated metro-area-wide product market, price pass-through may be difficult. On the other hand, if the demand for products is tightly linked to locations within the city itself, it may be possible for prices to exhibit sharp differences near city boundaries. Additionally, as we have argued, cities raising minimum wages tend to have residents with higher incomes and these consumers may be more willing and able to absorb an increase in prices of minimum wage intensive services and goods.

The empirical findings on this front are varied. Dube, Naidu, and Reich (2007) find that output prices increase especially for the fast food sector, particularly when comparing firms within versus outside of San Francisco. On the other hand, Jardim and Van Inwegen (2019) study the effect of the minimum wage on output prices

in Seattle and find somewhat inconclusive evidence. Perhaps the most persuasive evidence on price responses comes from Allegretto and Reich (2018) who study the impact of the San Jose minimum wage change on Internet-based restaurant menus inside and outside of the city boundaries. Allegretto and Reich (2018) find clear and positive price effects in response to the minimum wage that are consistent with the large body of evidence on state-level minimum wage changes. What is particularly telling is that they document a sharp drop on output prices just a mile from the San Jose city boundary. Therefore, otherwise similar restaurants operating within a few miles of each other—but facing differential shocks to labor costs—seem able to set different consumer prices. This suggests a very sharp segmentation of certain markets, even when the boundary is as porous as it is between San Jose and adjacent smaller cities like Sunnyvale and Milpitas. Overall, the sharp reduction in prices indicates that spillover effects of a city-wide minimum wage policy are limited even in the context of the San Francisco Bay Area with highly interlinked cites.

Worker Turnover

Minimum wage policies can affect labor market flows. In models of a frictional labor market where employers have some wage-setting power, a higher minimum wage can lead to a reduction in worker turnover at the bottom of the wage distribution: essentially, a higher minimum wage improves the relative quality of the lowest-paying jobs and increases retention (Portugal and Cardoso 2006; Dube, Lester, and Reich 2016). In turn, the decrease in turnover can lead to potential cost savings that can help to explain how a higher minimum wage is absorbed by employers without a substantial drop in employment.

Such a mechanism seems to play some role in the context of city-level minimum wages. Dube, Naidu, and Reich (2007) find an increase in worker tenure for a typical worker in the context of the San Francisco minimum wage increase, though they do not detect a reduction in overall separation rate. Jardim et al. (2018) find statistically significant reductions in separations following the Seattle minimum wage changes. Overall, the evidence from city minimum wages offers a qualified similarity to the broader literature suggesting reduced worker turnover is likely to be one of the channels of adjustment.

Conclusion

A growing number of cities have recently instituted their own minimum wages above the state or the federal one recently. Local variation in minimum wages tailors the policy more closely to the local economic and political environment. At the same time, city-level minimum wages might be more distortionary, because relocating businesses outside of city boundaries may be easier than relocating outside of the state or the country. While the literature on city-level minimum wages is still at an early stage, existing evidence does not indicate that the employment and wage responses differ substantially from the responses to state-level changes. Overall, the

weight of the evidence is consistent with these policies having moderately raised wages at the bottom without a large change in employment probabilities. Much of the adjustment seems to have been similar to state- and federal-level increases: through higher consumer prices, which in this case is mostly borne by middle- and higher-income consumers, and through some reductions in labor turnover costs. Considerable uncertainty surrounds some of these estimates, and more research is needed.

We expect a growing number of case studies will emerge soon concerning the impact of the policy in large cities such as Los Angeles and Minneapolis. Nevertheless, the literature can also benefit from exploiting combined analysis of the city-level minimum wage changes.

The new minimum wage literature began with studying the impact of a single change in the minimum wage in New Jersey—comparing changes in fast food employment in that state with changes in the neighboring state of Pennsylvania (Card and Krueger 1994). After the recent rapid expansion in the number of city-level minimum wages, economic research now has the ability to exploit hundreds of minimum wage increases. Data limitations pose some real challenges, but we see much value in exploiting more than single events to identify the effect of the policy. Our own findings presented in this paper provides an initial attempt at such a synthetic analysis; we hope to see much more. One additional point merits a mention: while use of state-specific administrative data can be of great help if there are credible control groups present within the state, the costs of relying on one state may exceed the benefits if proper control groups are not available. Furthermore, there is scope to learn from analyses using widely available public-use data, like the Quarterly Census of Employment and Wages or the American Community Survey. This is similar to the conclusion reached in Cengiz et al. (2019), who show that wage data from the Current Population Survey has comparable accuracy in many cases as hourly wage data from administrative sources. We think a similar lesson may apply to the city-wide context as well, at least for some questions.

Finally, surprisingly little research has been devoted to some important aspects of city-wide minimum wages. Direct evidence on business reallocation across city boundaries seems potentially important to assess the key trade-off emerging from introducing local variation in the policy. It also would be valuable to study the welfare implications arising from the potential reallocation of business from the core of a city to the more disadvantaged areas. Additional evidence on rental and housing prices also would help us to understand the welfare implications of minimum wage policies. We hope these gaps in the literature will be filled in the next wave of research on city-level minimum wages.

References

Aaronson, Daniel, and Brian J. Phelan. 2019. "Wage Shocks and the Technological Substitution of Low-Wage Jobs." *The Economic Journal* 129 (617): 1–34.

Albouy, David. 2009. "The Unequal Geographic Burden of Federal Taxation." *Journal of Political Economy* 117 (4): 635–67.

Allegretto, Sylvia, Anna Godoey, Carl Nadler, and Michael Reich. 2018. *The New Wave of Local Minimum Wage Policies: Evidence from Six Cities.* Berkely, CA: Center on Wage and Employment Dynamics.

Allegretto, Sylvia, and Michael Reich. 2018. "Are Local Minimum Wages Absorbed by Price Increases? Estimates from Internet-Based Restaurant Menus." *ILR Review* 71 (1): 35–63.

Autor, David H., Alan Manning, and Christopher L. Smith. 2016. "The Contribution of the Minimum Wage to US Wage Inequality Over Three Decades: A Reassessment." *American Economic Journal: Applied Economics* 8 (1): 58–99.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2019. "Measuring Labor Market Power Two Ways." *AEA Papers and Proceedings* 109: 317–21.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29–50.

Belman, Dale, and Paul J. Wolfson. 2014. *What Does the Minimum Wage Do?* Kalamazoo, MI: WE Upjohn Institute.

Briffault, Richard. 2018. "The Challenge of the New Preemption." Columbia Public Law Research Paper 14-580.

Caliendo, Marco, Alexandra Fedorets, Malte Preuss, Carsten Schröder, and Linda Wittbrodt. 2018. "The Short-Run Employment Effects of the German Minimum Wage Reform." *Labour Economics* 53: 46–62.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review* 84 (4): 772–93.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages On Low-Wage Jobs." *The Quarterly Journal of Economics* 134 (3): 1405–54.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3): 479–524.

Dube, Arindrajit. 2019. *Impacts of Minimum Wages: Review of the International Evidence.* London: HM Treasury.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2010. "Minimum Wage Effects across State Borders: Estimates Using Contiguous Counties." *The Review of Economics and Statistics* 92 (4): 945–64.

Dube, Arindrajit, T. William Lester, and Michael Reich. 2016. "Minimum Wage Shocks, Employment Flows, and Labor Market Frictions." *Journal of Labor Economics* 34 (3): 663–704.

Dube, Arindrajit, Suresh Naidu, and Michael Reich. 2007. "The Economic Effects of a Citywide Minimum Wage." *ILR Review* 60 (4): 522–43.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp Vom Berge. 2020. "Reallocation Effects of the Minimum Wage: Evidence from Germany." Unpublished.

Economic Policy Institute (EPI). 2018. "Worker Rights Preemption in The U.S. A Map of the Campaign to Suppress Worker Rights in The States." https://www.epi.org/preemption-map/.

Fajgelbaum, Pablo D, Eduardo Morales, Juan Carlos Suárez Serrato, and Owen Zidar. 2019. "State Taxes

and Spatial Misallocation". *The Review of Economic Studies* 86, No. 1: 333–376.

Harasztosi, Peter, and Attila Lindner. 2019. "Who Pays for the Minimum Wage?" *American Economic Review* 109 (8): 2693–2727.

Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6: 629–62.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma Van Inwegen, Jacob Vigdor, and Hilary Wething. 2017. "Minimum Wage Increases, Wages, and Low-Wage Employment: Evidence from Seattle." NBER Working Paper 23532.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma Van Inwegen, Jacob Vigdor, and Hilary Wething. 2018. "Minimum Wage Increases and Individual Employment Trajectories." NBER Working Paper 25182.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, and Hilary Wething. 2020. "Minimum Wage Increases and Low-Wage Employment: Evidence from Seattle." Unpublished.

Jardim, Ekaterina, and Emma Van Inwegen. 2019. "Payroll, Revenue, and Labor Demand Effects of the Minimum Wage." Upjohn Institute Working Paper 19-298.

Lemos, Sara. 2008. "A Survey of the Effects of the Minimum Wage on Prices." *Journal of Economic Surveys* 22 (1): 187–212.

Lordan, Grace, and David Neumark. 2018. "People versus Machines: The Impact of Minimum Wages on Automatable Jobs." *Labour Economics* 52: 40–53.

Luca, Dara Lee, and Michael Luca. 2019. "Survival of the Fittest: The Impact of the Minimum Wage on Firm Exit." NBER Working Paper 25806.

MaCurdy, Thomas. 2015. "How Effective Is the Minimum Wage at Supporting the Poor?" *Journal of Political Economy* 123 (2): 497–545.

McGovern Tony. 2016. "United States General Election Presidential Results by County from 2008 to 2016." https://github.com/tonmcg/US_County_Level_Election_Results_08-16 (accessed May 1, 2020).

Moe, Lina, James Parrott, and Yannet Lathrop. 2019. *New York City's $15 Minimum Wage and Restaurant Employment and Earnings*. New York: New York City Affairs at the New School and the National Employment Law Project.

Nadler, Carl, Sylvia A. Allegretto, Anna Godøy, and Michael Reich. 2019. "Are Local Minimum Wages Too High?" Institute for Research on Labor and Employment Working Paper 102-19.

Neumark, David, and William Wascher. 1992. "Employment Effects of Minimum and Subminimum Wages: Panel Data on State Minimum Wage Laws." *ILR Review* 46 (1): 55–81.

Portugal, Pedro, and Ana Rute Cardoso. 2006. "Disentangling the Minimum Wage Puzzle: An Analysis of Worker Accessions and Separations." *Journal of the European Economic Association* 4 (5): 988–1013.

Potter, Nicholas. 2006. *Measuring the Employment Impacts of the Living Wage Ordinance Santa Fe, New Mexico.* Albuquerque: University of New Mexico, Bureau of Business and Economic Research.

Rapoport, Abby. 2016. "Blue Cities Battle Red States." *The American Prospect*, August 22. https://prospect.org/economy/blue-cities-battle-red-states/.

Schmitt, John, and David Rosnick. 2011. *The Wage and Employment Impact of Minimum-Wage Laws in Three Cities.* Washington, DC: Center for Economic and Policy Research.

Simonovits, Gábor, and Julia Payson. 2020. "Locally Controlled Minimum Wages Are No Closer to Public Preferences." Unpublished.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (5): 416–24.

Tijdens, Kae, and Maarten van Klaveren. 2019. "Understanding the Drivers of Minimum Wage-Setting: An Analysis of 146 countries." Paper presented at the ILO's 6th RDW Conference in Geneva, July 8–10, 2019.

How Do Firms Respond to Minimum Wage Increases? Understanding the Relevance of Non-Employment Margins

Jeffrey Clemens

When analyzing the effects of minimum wages, both researchers and policymakers have long focused their attention on wages and employment as outcomes. Research on the employment effects of the minimum wage has, by itself, supported literature reviews and meta-analyses for decades (for some recent examples, see Belman and Wolfson 2019; Clemens 2019; Neumark 2019; Congressional Budget Office 2019; Dube 2019). While debate remains contentious, a rising fraction of researchers have concluded that the employment effects of moderately sized minimum wage increases are quite close to zero (Belman and Wolfson 2019). At the same time, the public and policymakers have adopted a favorable view of historically large minimum wage increases. The public's support for raising the federal minimum wage from $7.25 to $15 per hour has recently approached levels of support that have historically applied to much smaller minimum wage increases. A recent poll from Pew Research, for example, found 67 percent of Americans in support of a $15/hour federal minimum wage (Davis and Hartig 2019), which is quite similar to the 71 percent support found for a $9/hour minimum wage in a 2013 poll from Gallup (Saad 2013). As of January 2020, policymakers in California, Connecticut, Illinois, Maryland, Massachusetts, New Jersey, and New York had enacted laws that would eventually raise their states' minimum wages to $15/hour.

Debates among economists over the employment effects of the minimum wage have been highly contentious. In a provocative essay, Leonard (2000, p. 118) suggests

■ *Jeffrey Clemens is Associate Professor of Economics, University of California at San Diego, La Jolla, California. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and a Research Network Fellow, CESifo, Munich, Germany. His email address is jeffclemens@ucsd.edu.*

a reason why. He points to the debate's perceived intellectual stakes, writing that "it is the latest chapter in a longstanding methodological dispute over whether and in what domains neoclassical price theory can be said to properly apply." That is, estimates of employment elasticities are treated as tests of the neoclassical model itself. In this paper, I argue through a series of theoretical and empirical illustrations that this high-stakes "theory-testing" approach to evidence on the effects of minimum wages has been misguided.

Instead, I contend that controversies over the economics of minimum wages stem, to a surprising degree, from a common but under-considered assumption. The assumption of interest is that when studying labor markets, output prices and nonwage aspects of jobs (which include benefits and working conditions) can be taken as fixed. In standard diagrams of the labor market, this assumption implicitly underlies each supply or demand curve. When these curves are held fixed, output prices and nonwage aspects of jobs have also, whether implicitly or explicitly, been held fixed.

Models that implicitly reduce jobs to their wages pervade textbook analyses of the minimum wage. On one level, this is not surprising because wages and employment occupy the axes along which we draw labor supply and demand curves. Indeed, the wage will tend to be a job's single most important attribute. At the same time, however, textbook models of both perfectly and imperfectly competitive labor markets sweep many factors under the rug. Benefits, including employer-provided health insurance, account for around one-third of compensation costs (US Bureau of Labor Statistics 2020b). Working conditions, including safety measures and flexible schedules, can also have value to workers while generating costs to firms. Textbook labor market models also tend to treat output prices and production technologies as fixed. Yet for reasons detailed below, it would be natural for any number of these margins to shift in response to minimum wage changes. In the jargon of undergraduate instruction, the *ceteris paribus* assumption (that is, "other things held constant") that professors invoke when we draw labor supply and demand curves is unlikely to describe the real world.

In the following discussion, I show that margins including nonwage job attributes can have first-order implications for analyses of minimum wages. In models that account for such factors, predictions for the effects of minimum wages on unemployment and worker welfare can, perhaps surprisingly, be reversed from our basic intuitions. I also show how these results can be illustrated through minor extensions to basic diagrams of labor supply and demand.

I attempt to answer three questions. First, through what non-employment margins might firms adjust in response to minimum wage increases? Second, what has empirical research taught us about the minimum wage's effects on non-employment margins? Third, how should these margins enter our assessments of the minimum wage as a tool for improving the well-being of low-wage workers? I begin by analyzing minimum wage response margins that fit nicely into standard supply and demand diagrams. These margins include output prices, nonwage compensation, and other job attributes. I then turn to additional, firm-level factors that may be in play. How might it matter, for example, if firms design uniform health insurance

benefits for both minimum wage workers and higher wage workers? How might a firm alter its production technology, perhaps to improve the skill mix of its workers or to transition from labor- to capital-intensive production? Along what time horizons might these adjustments occur? In the conclusion, I offer some thoughts about what these issues imply for nuanced assessments of the minimum wage and for classroom pedagogy on this topic.

Analyzing the Minimum Wage within a Basic Labor Supply and Demand Framework

This section begins with the textbook, introductory-level approach for teaching the economics of the minimum wage in a perfectly competitive labor market. I use this initial discussion to highlight factors that are held constant when we draw labor supply and demand curves.

The Minimum Wage in a Textbook Diagram of a Perfectly Competitive Labor Market

Figure 1 presents the standard textbook diagram for analyzing the effects of a binding minimum wage in a perfectly competitive labor market. The labor supply curve slopes upward, reflecting differences in workers' reservation wages (as driven by outside opportunities related, perhaps, to leisure, home production, and economic assistance that can be received while out of work). The labor demand curve slopes downward, tracing out the relationship between the quantity of labor employed and the marginal revenue product of that labor. This, in turn, reflects the assumption of a constant price (due, perhaps, to a perfectly competitive market for the firm's output) and a production function in which, holding capital and technology fixed, labor has diminishing marginal productivity.

In a perfectly competitive labor market, a freely set wage will adjust to equilibrate supply and demand. That is, the wage will adjust to the level (w_1) such that the opportunity cost of the last worker's time equals the value of that worker's output to firms. A binding minimum wage floor, as at $w_{min,2}$ in Figure 1, breaks this equilibrium. This wage floor would increase the quantity of labor supplied while decreasing the quantity demanded. The resulting labor surplus implies unemployment. The market is demand-constrained and the minimum wage of $w_{min,2}$ causes employment to decline from L_1 to L_2.

Pass-through of Cost Increases to Consumers through Output Prices

A natural starting point for extending the traditional textbook analysis is to consider output prices. In the supply and demand diagram, the labor demand curve traces out labor's "marginal revenue product." That is, it traces the marginal product of labor, or $MP(L)$, times the output price, P. The output price is thus a shifter of the labor demand curve. Treating the output price as fixed assumes perfect competition across product markets that extend beyond the jurisdiction in which the minimum wage binds. More concretely, it is imagined that equivalent

Figure 1

Basic Supply and Demand Diagram: Fixed Non-Wage Job Attributes and No Price Pass-Through
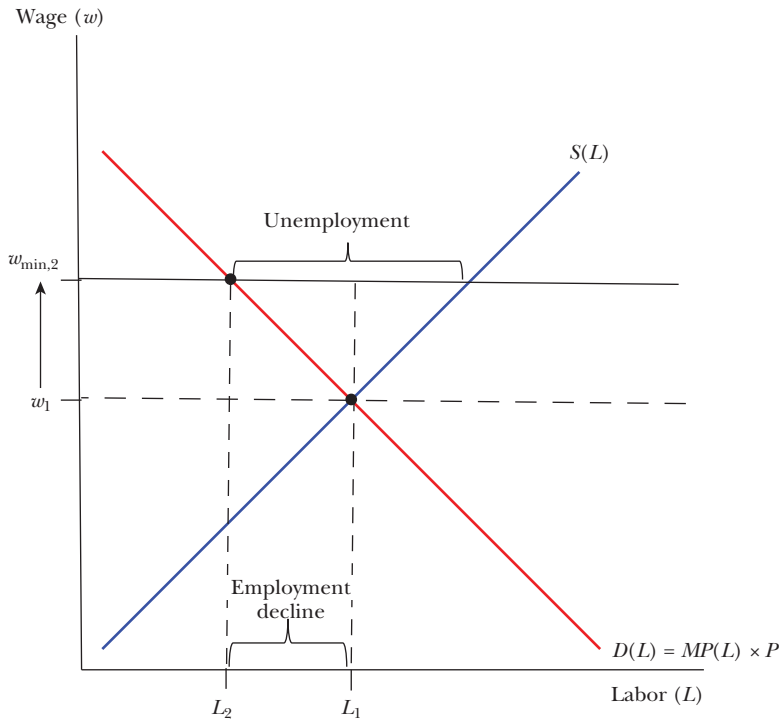

Note: The figure presents the standard textbook diagram for analyzing the effects of a binding minimum wage floor in a perfectly competitive labor market.

goods or services can be costlessly imported from producers elsewhere who are not affected by the increase in wages. As a result, any firm that raises its price will lose all of its customers.

If demand for a firm's output is not perfectly elastic, it can raise prices while losing some, but not all, of its customers. A price increase in response to a minimum wage increase is often called *pass-through*: that is, the minimum wage's cost passes through the firm to its consumers. An increase in output prices implies an outward shift of the labor demand curve, as illustrated in Figure 2. This blunts the magnitude of the decline in employment following an increase in the minimum wage, as can be seen by comparing Figure 2 with Figure 1.

A firm's ability to adjust prices depends on the elasticity of demand for its goods and/or services. This depends, in turn, on the scope of the market. Firms that produce widely traded goods or services may face large demand elasticities and thus have little capacity to raise prices. By contrast, firms that produce "non-tradable" goods and services may face smaller demand elasticities and have more

Figure 2
Partial Offset through Price Pass-Through
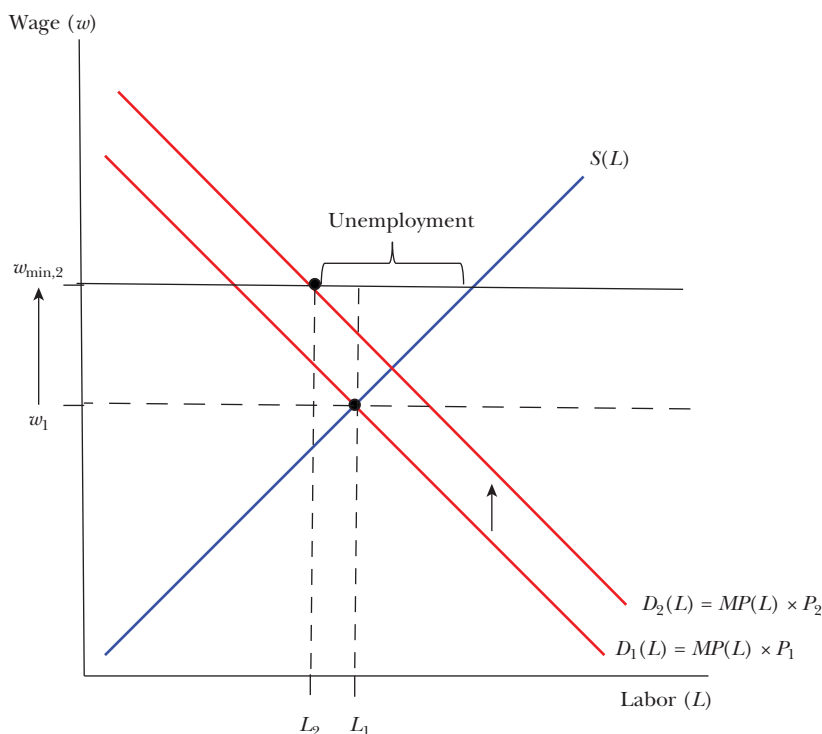
Note: The figure augments the standard textbook diagram for analyzing the effects of a binding minimum wage floor (see Figure 1) by allowing for the possibility that firms increase output prices in response to the cost increase associated with the minimum wage.

substantial scope for passing cost increases to consumers (Moretti 2010, 2011). Standard examples of non-tradables include beauty services, meals at restaurants, and home construction, which are more or less constrained to be provided where they are consumed. Pass-through may also depend on whether the minimum wage is increased at the city, state, or federal level. When a minimum wage increase is localized, there is greater scope for importing products from unaffected firms.

　　How empirically relevant is a firm's capacity to pass the cost of the minimum wage to its consumers through higher prices? Lemos (2008) reviews the literature through the mid-2000s, most of which analyzes prices in the restaurant industry. Results through that time were quite mixed. Aaronson (2001) and MacDonald and Aaronson (2000), for example, found evidence of substantial price pass-through. By contrast, analyses by Katz and Krueger (1992) and by Machin, Manning, and Rahman (2003) found no evidence of price pass-through.

　　Several researchers have recently returned to the issue of minimum wage pass-through using administrative data sources and recent developments in best-practice

empirical methods. In the US context, for example, Leung (forthcoming) analyzes pass-through using price data built up from individual-level purchases in the Nielsen Consumer Panel. Renkin, Montialoux, and Siegenthaler (forthcoming) analyze US grocery price scanner data. Harasztosi and Lindner (2019) analyze pass-through using firm-level price indices from Hungary. Each of these papers finds evidence of substantial pass-through. Harasztosi and Lindner, for example, estimate that roughly 75 percent of the cost of a large increase in Hungary's national minimum wage was passed onto consumers. In their analysis, Harasztosi and Lindner provide novel evidence of variations in pass-through across industries. Consistent with standard theory, they find that price responses were greater in non-tradable industries than in tradable industries. Conversely, they find that employment declines were greater among firms in tradable industries, which were less able to pass cost increases onto their consumers, than in non-tradable industries.

Pass-through depends on factors that can vary across industries, across markets, and over time. Harasztosi and Lindner's (2019) analysis, for example, focuses on differences across industries defined as "tradable" or "non-tradable" by earlier research. These categories may be malleable, however, as the scope of competition may vary meaningfully over time. Online retail, for example, increases the scope of the market for retail services. The rise of online retail may thus reduce the capacity for retailers to increase prices in response to localized increases in the minimum wage. A constructive path forward will be for this segment of the literature to develop additional evidence on the attributes of industries and markets that translate into lesser or greater price pass-through—which may also help to explain variations in pass-through estimates across earlier studies.

If costs are passed through to consumers, then the incidence of the minimum wage depends, in part, on which households consume products that require substantial input from minimum wage workers. MaCurdy (2015) provides evidence that this force will tend to be regressive, since minimum-wage-intensive products account for a disproportionately large fraction of the budgets of low-income households. This pattern is driven to a significant degree by the relative importance of groceries and food away from home in the budgets of low-income households relative to high-income households. Data from the 2019 Consumer Expenditure Survey, for example, reveal that food expenditures (combining food consumed both in and out of the home), account for roughly 15 percent of the expenditures of low-income households and just over 10 percent of the expenditures of high-income households (US Bureau of Labor Statistics 2020a). This point is also made by Renkin, Montialoux, and Siegenthaler (forthcoming).

In MaCurdy's (2015) analysis, the minimum wage appears far less progressive than one might expect, even in the absence of negative employment effects. In addition to the composition of their consumption bundles, this finding reflects the fact that many minimum wage workers live in middle- to high-income households (Congressional Budget Office 2019; Sabia and Burkhauser 2010). MaCurdy thus finds that income gains are diffused across the distribution of household incomes, while costs fall disproportionately on low-income consumers.

The distributional implications of minimum wage pass-through can depend on the empirical setting. In their analysis of data from Hungary, for example, Harasztosi and Lindner (2019) find that the goods and services consumed by the rich and poor have quite similar exposure to price increases due to increases in the minimum wage. Regardless of the net impact on progressivity, however, it is clear that price pass-through erodes the real income gains households might otherwise realize through minimum wage increases.

Noncash Compensation

In response to a minimum wage increase, a firm might adjust its noncash compensation offerings. Academics and their employers, for example, contract over the generosity of benefits including health insurance, pensions, research budgets, and travel budgets. Across all US workers, the US Bureau of Labor Statistics (2020b) documents that wages and salaries account for roughly 70 percent of total compensation costs. Benefits including health insurance, paid leave, and pensions account for much of the rest.

Conceptually, it is straightforward to see that offsetting changes in nonwage compensation can nullify the minimum wage's effects on compensation costs and, by extension, labor demand. While the logic is simple, however, understanding the role of noncash compensation in supply and demand diagrams requires thinking subtly about supply and demand curves. The key point, as illustrated below, is that reductions in noncash compensation can shift both the supply curve and the demand curve. The relevant theoretical points have been made in both the minimum wage literature and the literature on nonwage benefit mandates (Wessels 1980; McKenzie 1980; Summers 1989).

Figure 3 illustrates the implications of noncash compensation for labor supply and demand. A given labor supply or labor demand curve assumes a fixed level of noncash compensation (f_1 in the diagram).[1] From the perspective of firms, lower noncash compensation implies a higher labor demand curve because it increases revenues net of nonwage costs. From the perspective of workers, lower noncash compensation implies a higher labor supply curve, since a higher wage is required to make employment attractive when nonwage benefits are lower. Because f_1 is expressed as a dollar-denominated cost, it enters linearly into the labor demand curve. The marginal value of these benefits to workers may depend, however, on the quantity at which they are provided and the efficiency with which they are designed or produced by the firm. Consequently, the figure depicts the supply curve as a more general function, $S(L, f_1)$, of these benefits.

The figure illustrates the special case in which wages and nonwage benefits are perfectly substitutable on the margin, meaning their cost to firms exactly equals

---

[1] A subtle point that can easily go unnoticed is that traditional derivations of labor demand curves assume that there is no noncash compensation—that is, not simply that it is held fixed, but that it does not exist. To be more specific, it is common for textbooks to write that firms demand labor until its marginal revenue product equals the wage rate. In the presence of noncash compensation, however, the zero profit condition is that the wage equals the labor's marginal revenue product *net of* the cost of nonwage compensation.

Figure 3
Perfect Offset through Noncash Compensation (for example, "fringe benefits")
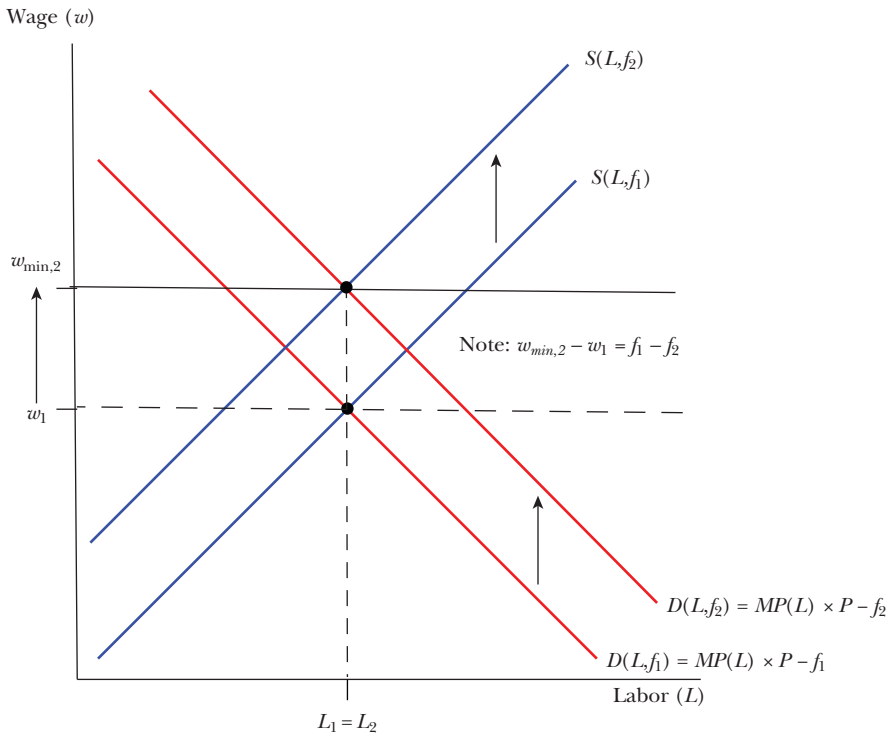
Note: The figure augments the standard textbook diagram for analyzing the effects of a binding minimum wage floor (see Figure 1) by allowing for the possibility that firms reduce the generosity of noncash compensation (for example, "fringe benefits" like employer-provided health insurance) in response to the cost increase associated with the minimum wage. A similar diagram is presented as Figure 1 in McKenzie (1980).

their value to workers. In this case, fringe benefits can be reduced so as to perfectly offset, and thus nullify, the minimum wage increase ($w_{min,2} - w_1 = f_1 - f_2$). In this special case, the net cost and value of compensation are both unchanged, as is the quantity of labor transacted. However, if the minimum wage were to distort the compensation package away from its optimal mix, then the labor supply curve would shift more than the demand curve. In this case, a lower quantity of labor would be transacted in equilibrium. Although the distortion would reduce workers' well-being, there would be no unemployment because the market would, nonetheless, be at a new equilibrium.

Research on the effects of minimum wages on nonwage compensation is relatively sparse. Because of data limitations, the primary component of nonwage compensation that can be incorporated into research on minimum wages is the margin of whether workers have employer-provided health insurance. The available data on employer-provided health insurance tends to be binary—measuring

whether a worker has such insurance or not—and thus does not capture changes in the cash value of whatever benefit might be offered.

Analyses of historical minimum wage changes in the United States tend to find weak evidence of a relationship between minimum wage increases and employer-provided health insurance (for examples, see Royalty 2001; Simon and Kaestner 2004; Marks 2011). Analyses of relatively recent minimum wage changes have tended to find negative, though modest, effects. A back of the envelope calculation by Clemens, Kahn, and Meer (2018) suggests that declines in the provision of employer-provided health insurance have offset roughly 15 percent of the cost of states' recent minimum wage increases. Additional recent work also reports negative correlations between minimum wages and employer-provided health insurance in data from both the American Community Survey and the Current Population Survey (Gooptu and Simon 2019; Eibner et al. 2017).

How do nonwage compensation margins relate to the distributional effects of minimum wage policy? Adjustments along these margins will simultaneously tend to mitigate employment effects while also blunting the wage increase's effects on a worker's well-being. Because these factors are often unmeasured, it seems appropriate to exercise humility regarding the strength of the conclusions we can draw from available data.

Other Job Attributes

In addition to changing noncash compensation, employers may adjust job attributes like effort requirements, schedule flexibility, and training opportunities in response to changes in minimum wages. Positive aspects of jobs are often referred to as "non-compensation amenities," while negatives are known as disamenities. Conceptually, a firm facing minimum wage increases might seek to offset some of the increase in costs by raising productive disamenities (like effort requirements) and reducing unproductive amenities (like the quality of office furniture). As with changes in noncash compensation, these changes will shift both the supply curve and the demand curve.

Figure 4 illustrates the relevance of productive disamenities like effort requirements, denoted e. The basic theory underlying the relevance of productive disamenities for minimum wage analyses has been laid out by Ippolito (2003). The diagrammatic analysis is quite similar to the analysis of nonwage components of compensation. Increases in productive disamenities (from e_1 to e_2 in the diagram) imply upward shifts in the labor demand curve (due to an increase in marginal product) and upward shifts in the labor supply curve (due to an increase in disutility from work effort). The figure illustrates the case in which a binding minimum wage results in suboptimal levels of productive disamenities. The increase in productive disamenities thus results in a larger shift in the labor supply curve than in the labor demand curve.

A final set of conceptual points are worth noting here. In Figure 1, where we held nonwage compensation, workplace amenities, and effort requirements fixed, the minimum wage caused unemployment to rise by more than employment declines. That is, the diagram predicts job shortages linked to both a decline in the quantity demanded and an increase in the quantity supplied. In Figure 4, however, where

Figure 4
Increase in a Productive Disamenity (for example, "effort")
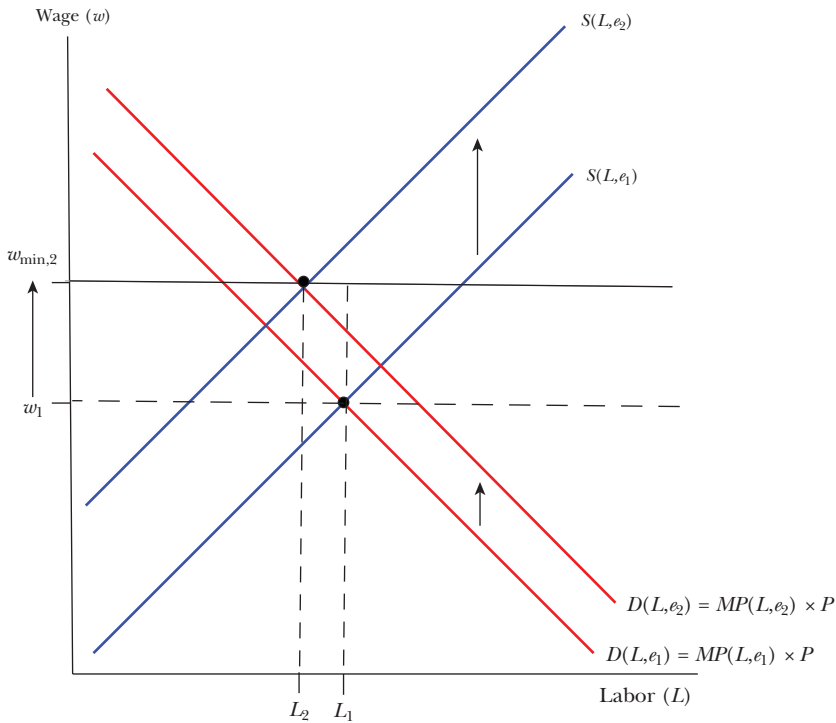
Note: The figure augments the standard textbook diagram for analyzing the effects of a binding minimum wage floor (see Figure 1) by allowing for the possibility that firms increase a productive disamenity (for example, the amount of effort required from workers) in response to the cost increase associated with the minimum wage. A similar diagram is presented as Figure 3 in Ippolito (2003).

we allowed the effort requirement to rise from e_1 to e_2 in response to the minimum wage, the labor market has settled into in a new equilibrium. In other words, while the diagram predicts that the minimum wage will lead employment to decline, it also predicts that there will not be an excess supply of workers looking for jobs. One prediction of the basic instructional diagram (in Figure 1) has thus been reversed, and another blunted, without deviating from a perfectly competitive framework. Finally, the shift to higher effort means that the welfare of the average worker may have fallen, in spite of the increase in wages. The conventional wisdom, which holds that a binding minimum wage improves worker welfare if it has minimal effects on employment, may no longer hold.

A wave of high-quality research has recently demonstrated the relevance of workplace amenities. Several recent papers, for example, highlight the value of worker-driven schedules (for example, Chen et al. 2017; Mas and Pallais 2017; He, Neumark, and Weng 2019). This set of papers include experimental studies as well as a study of workers' valuations of the flexibility enabled by Uber. A separate analysis

by Maestas et al. (2018) uses an experimental survey approach to quantify workers' willingness to pay for improved workplace conditions, and finds these valuations to be substantial.

While workplace amenities clearly matter, there is little research on the responsiveness of amenities to minimum wage changes. Clemens and Strain (2020a) provide an illustrative example, whereby minimum wage increases can result in shifts away from worker-driven schedules and towards employer-driven schedules. Employer-driven schedules can generate higher output per hour if, for example, they enable firms to dismiss workers during slack shifts. Lack of control over their schedules is costly for workers and thus requires compensation in the form of higher wages. A binding minimum wage may render low-wage, worker-friendly schedules illegal. In doing so, it can result in scheduling changes that mitigate the minimum wage's employment effects but leave workers worse off. As noted above, however, there is little, if any, empirical evidence on the minimum wage's effects on scheduling, workplace safety, workplace comfort, and other related margins.

How empirically important are effort requirements? The answer to this question is unlikely to be general, as the relevance of effort requirements may vary significantly across settings. The empirical connection between minimum wages and effort requirements can be traced back more than a century, to an analysis of minimum wage laws in Oregon. In that analysis, Obenauer and von der Nienburg (1915) described workers' experience in the wake of a minimum wage increase as follows: "Now, they said, they are under constant pressure from their supervisors to work harder; they are told the sales of their departments must increase to make up for the extra amount the firm must pay in wages." Recently, there are two high-quality papers on the effects of minimum wages on productivity. Ku (forthcoming) looks at data on temporary tomato pickers in Florida over the period surrounding a statewide minimum wage increase. Coviello, Deserranno, and Persico (2020) look at data from a sample of 10,000 US workers paid on a base-plus-commission basis by a nationally operating retailer. Both studies find that low-productivity workers increase their work effort in the wake of minimum wage increases. In both cases, effort responses appear to substantially offset the cost increases associated with minimum wage increases. These large impacts are likely connected to the specific settings the authors analyze. Both Ku (forthcoming) and Coviello, Deserranno, and Persico (2020) analyze workers and firms in settings where output is well-measured and where compensation has a strong piece-rate, or commission, component.

Firm Profits

If firms fail to respond to minimum wages, increases in their input costs will lead mechanically to declines in their profits. Draca, Machin, and Van Reenen (2011) analyze profitability in the wake of the UK's 1999 introduction of its national minimum wage. They find evidence that the cost of the UK's national minimum wage was, to a substantial degree, absorbed in the form of lower profits. Bell and Machin (2018) study firm value as measured using market capitalization. They do so following the July 8, 2015, announcement of an unexpectedly large increase in the UK's national minimum wage. As in the Draco, Machin, and Van Reenen study, the

decline in market capitalization is similar to what one would infer if the full cost of the minimum wage increase were to be absorbed in the form of lower profits.

Evidence that firm owners bear some of the cost of minimum wage increases raises the question of where these firm owners fall in the income distribution. Drucker, Mazirov, and Neumark (2019) provide a rare answer to this question using administrative records from Israel. They find that the owners of Israeli firms that employ minimum wage workers tend to be in the top half of the overall income distribution, but towards the bottom of the distribution of income across firm owners.

Reductions in profits also raise the question of firm exit, or of which firms survive. Patterns of exit can have implications for the products available to consumers, the jobs available to workers, and demand for commercial real estate. In a study using data from Yelp, for example, Luca and Luca (2019) find that minimum wage increases in San Francisco predict increases in exit among lower-rated restaurants. In an analysis of workers and firms in Germany, Dustmann et al (2020) find that the 2015 introduction of Germany's minimum wage led to a real-location of workers "from small, low paying firms to larger, higher paying firms." Firm exit also has implications for the owners of land, as a reduction in demand for commercial real estate may depress rents. The incidence of minimum wage increases on land owners has not, to the best of my knowledge, been addressed by empirical research.

The full absorption of wage increases through lower profits, as found in two studies referenced above, is particular to settings in which no other margins adjust. As discussed earlier, there are settings in which margins including prices, effort, and noncash compensation have also been found to adjust, to greater or lesser degrees. When other margins prove highly flexible, as in some of the studies mentioned earlier, profits will absorb less of the burden.

This section has discussed how traditional supply and demand analysis can provide a far richer picture of firms' responses to minimum wage changes than is widely appreciated. These straightforward extensions can quite readily be brought into the supply and demand diagrams through which labor markets are taught to undergraduates. Similar insights regarding the relevance of prices, noncash compensation, and productive disamenities can also be developed within classic monopsony models. Not all of the minimum wage's potential effects, however, can be captured through supply and demand analyses within a single standard diagram. The next section turns to several such issues.

A Broader Perspective on Personnel Policies and Production Technologies

In this section, I turn to issues that push outside the limits of traditional supply and demand diagrams: decisions by firms to evade minimum wage regulation; adjustments that apply to all of a firm's personnel, such as changes to firm-wide benefits packages; and changes to firms' production functions.

Compliance

Up to this point, I have implicitly assumed that the minimum wage is enforced as intended. Noncompliance is an empirical reality, however, making it important to understand its economic determinants. An earlier generation of papers, beginning with Ashenfelter and Smith (1979), developed theoretical models of the economics of minimum wage evasion. This research drew heavily on Allingham and Sandmo's (1972) model of tax evasion, which in turn draws on Becker's (1968) classic model of the decision to participate in crime. In such models, the decision to evade is primarily a function of the gains from successful evasion, the probability of detection, and the penalties associated with being caught. A novel point from the minimum wage litera-ture is that evasion will be more likely when it is in the joint interest of firms and workers, which can occur when a firm would not find it profitable to hire a worker at the prevailing minimum wage. In such cases, evasion of minimum wage laws can be viewed, in part, as a release valve that mitigates the wage floor's adverse effects on employment.

Empirical research on the determinants of noncompliance with the minimum wage consistently finds noncompliance to be nontrivial. Further, the extent of noncom-pliance varies in ways that are consistent with insights from the theoretical analyses discussed above. Goraus-Tańska and Lewandowski (2019), for example, find that violation rates in central and eastern Europe are highest when minimum wages are high relative to average wages. Violation rates have also been found to be high when firms are limited in their ability to pass labor costs on to consumers. For example, this pattern is observed by Weil (2005) in an analysis of US apparel companies with government contracts, and also by Bernhardt, Spiller, and Theodore (2013) in a survey of workers from Chicago, Los Angeles, and New York City. Following Germa-ny's recent introduction of a statutory minimum wage, rates of noncompliance were moderately high, at least in the short run (Caliendo, Schröder, and Wittbrodt 2019).

How much has subminimum wage payment risen following recent minimum wage increases in the United States? This question is difficult to answer due to measurement error concerns in self-reported wage data. Clemens and Strain (2020b) devote significant attention to measurement error as a potential source of the appearance of subminimum wage payment in the Current Population Survey. Their analysis of recent minimum wage changes estimates that noncompliance has averaged roughly 14 to 21 cents per $1 of realized wage gain.

The findings discussed above suggest a trade-off between enforcement and the minimum wage's effects on employment and other margins. Enforcement of minimum wage laws may reduce evasion but risks exacerbating adverse effects on workers. As Clemens and Strain (2020b) emphasize, this "trade-off between short-run economic efficiency and respect for the rule of law is not one to be taken lightly." After all, the quality of governance, or "social infrastructure," has long been understood to be an important determinant of long-run growth and economic well-being (Hall and Jones 1999).

A final point regarding estimates of noncompliance involves marginalized groups, like undocumented workers, and the role of informal economic activity. The scope for evasion will tend to be largest when informal economic activity is

common. Due to the threat of deportation, undocumented workers in informal employment arrangements may have limited recourse against an employer who pays below the minimum wage. Orrenius and Zavodny (2011) study this issue using wage and employment data from the US Current Population Survey. They report "mixed support for our hypothesis that undocumented immigrants are less likely to be affected by changes in the minimum wage." But the issue is inherently difficult to study, as the Current Population Survey does not directly identify undocumented workers. In the following section, I turn to a broader set of issues involving the effects of the minimum wage on the compensation and employment of workers who differ with respect to their productivity and credentials.

Firm-Level Decisions Regarding the Design of Jobs and Compensation

Basic supply and demand analyses make a strong assumption about the workers under consideration. Within a given supply and demand diagram, workers are implicitly assumed to be undifferentiated, meaning they have equal productivity. As employers will readily attest, however, ostensibly similar workers employed in equivalent jobs can differ dramatically in their output. In their hiring decisions, firms face choices regarding the skill mix of workers they employ.

Firms also face decisions regarding the design of benefits packages, which may involve choices that cannot be tailored to workers of different skill levels. High- and low-wage workers may have quite different preferences over the generosity of benefits on the margin: for example, high-wage individuals might prefer a more generous insurance plan and a nicer work environment, while low-wage workers might prefer higher wages. Administrative costs may make it infeasible, however, for firms to tailor insurance benefits or workplace amenities to each worker's desires. Many firms in the United States, for example, including a majority of small firms, use the same health insurance benefit for broad categories of workers.[2]

When firms offer a common benefits package to workers of multiple skill types, minimum wage increases may have the effect of altering the compensation packages offered to both minimum wage and non-minimum wage workers. As pointed out by Clemens, Kahn, and Meer (2018), changes in health insurance or other benefits can thus generate "ripple effects." That is, they create a mechanism through which minimum wage increases can impact wages for both minimum wage workers and higher wage workers. Changes in fringe benefits may thus help us to understand spillovers that have been detected in analyses of wages.

Changes in the mix of compensation can have nuanced welfare implications. Under standard assumptions (for example, that health insurance is a "normal good" over which utility increases at a diminishing rate), a benefits package that balances the interests of high- and low-wage workers will tend to include health insurance benefits that are excessive from the perspective of low-wage workers: that

---

[2]As noted in Clemens, Kahn, and Meer (2018), for example, roughly one-half of firms that provide health insurance offer only one plan (as shown in Summary Table II.a.2.d, 2016 Medical Expenditure Panel Survey at https://meps.ahrq.gov/data_stats/summ_tables/insr/state/series_2/2016/tiia2d.htm). That is, many firms design a single benefit for the entirety of their workforce, likely due to administrative costs.

is, all also equal, low-wage workers would prefer to receive additional cash and less generous health insurance. If a firm shifts from noncash towards cash compensation due to an increase in the minimum wage, the welfare of low-wage workers may thus rise, while the welfare of high-wage workers may decline. As emphasized earlier, the presence of an offset between insurance and wages implies that the welfare of low-wage workers rises less than one would infer based on the change in their wages alone.

Firms can also alter their personnel policies by substituting from low-skilled labor to higher-skilled labor. This "labor-labor substitution" can be important for assessing the welfare implications of a change in the minimum wage. Data on industry-wide wage bills and employment counts can mask changes in the types of workers that are actually employed. If the least-skilled are replaced by higher-skilled workers, industry employment counts can mask job losses for the intended beneficiaries of the minimum wage. Relatedly, employers who seek to evade the minimum wage may increase their hiring of undocumented immigrants, as discussed by Orrenius and Zavodny (2008). The empirical relevance of this evasion margin may depend crucially on immigration policy, which influences the number of undocumented workers in the labor market. The extent of informal economic activity may also be relevant in this regard.

What evidence exists on "labor-labor substitution"? One piece of evidence on how the minimum wage can affect firms' hiring decisions is from an experiment in a real-world, online labor market (Horton 2018). The experimental sample in Horton's study includes "nearly 160,000 job openings." Horton finds clear evidence that minimum wage increases lead firms to shift from low-productivity workers towards higher-productivity workers. More specifically, he finds that firms shift towards workers whose job profiles show higher past wages.

In addition to Horton's (2018) experiment, recent analyses from several settings have found evidence that firms substitute away from low-skilled workers following minimum wage increases. Clemens, Kahn, and Meer (forthcoming) find that recent increases in state minimum wages predict increases in the average age and education of workers in low-wage occupations. They also find evidence that firms altered their job vacancy postings: specifically, firms became more likely to require high school diplomas of prospective employees in the wake of recent minimum wage increases. Within food service or retail industries, Clemens and Wither (2019), Fairris and Bujanda (2008), and Giuliano (2013) all find evidence consistent with substitution across groups of low-skilled workers.[3] Finally, research on the city of Seattle's recent minimum wage increases has found disproportionately large hours reductions for those who were relatively new to their jobs relative to low-wage workers who had substantial experience in their jobs (Jardim et al. 2018).

---

[3] In an appendix analysis of food service employment, Clemens and Wither (2019) find that some of the least-skilled food service workers were displaced by slightly higher-skilled workers after the July 2009 increase in the federal minimum wage. Fairris and Bujanda (2008) find increases in employment among demographic groups for which there were skill premia after the implementation of the 1997 Los Angeles Living Wage Ordinance. At a large retail firm, Giuliano (2013) finds a shift in employment towards teenagers from affluent neighborhoods following the 1996 increase in the federal minimum wage.

How large are these adjustments to the skill mix of workers employed in low-wage occupations? In their analysis of recent minimum wage changes, Clemens, Kahn, and Meer (forthcoming) find nontrivial increases in the average age of individuals employed in particularly low-wage occupations (like food service and food preparation). This increase in age comes primarily from reductions in the probability that workers are ages 16 to 21, and is accompanied by an increase in the share of workers with at least a high school diploma. Clemens, Kahn, and Meer translate these estimates into dollar terms using coefficients from Mincer (1974)-style earnings regressions. That is, they calculate how much more one would expect the new workforce to be paid, as estimated using its demographic characteristics.[4] In their particular context, Clemens, Kahn, and Meer find that the hypothetical earnings differential for the new skill mix relative to the old skill mix was equivalent to around one-third to one-half of the wage increase made necessary by the increase in the minimum wage.

The research discussed in this section illustrates how firm-wide personnel policy can complicate evaluations of the minimum wage. First, industry-wide wage bills and employment counts can mask changes in the composition of who is employed. Second, changes in the design of compensation can have nuanced implications for the welfare of both the low- and high-skilled members of a firm's workforce. These issues highlight that extensive data on firms' personnel and compensation can be useful for generating detailed assessments of the minimum wage's incidence.

Adjustments to Firms' Production Technologies

In response to minimum wage changes, firms may alter their production functions. That is, they may increase their reliance on capital, technology, and high-skilled labor, while reducing their reliance on low-skilled labor. Firms may also change their scope, meaning the set of functions they conduct in-house, as opposed to the functions they outsource to others.

Adjustments to capital and technology can have empirically important implications for dynamics. Consider a fast-food chain, for example, which must choose between automated ordering kiosks and employee-operated cash registers. Each choice involves an initial investment that will depreciate gradually. When wages are low, labor-intensive cash registers might minimize average production costs. When minimum wages rise, new firms might minimize average costs by acquiring automated kiosks. New firm entrants to the fast-food market may thus adopt less labor-intensive technologies soon after a minimum wage increase. By contrast, consider a continuing firm that has recently acquired new employee-operated cash registers. Such firms may continue operating with cash registers until their equipment requires replacement. The response of such firms to a minimum wage increase may thus be lagged by several years. This difference in the production functions of

---

[4]One interpretation of the Mincer earnings coefficients is that they represent the returns to human capital, in which case the earnings differentials correspond with differences in productivity. Whether the differentials reflect productivity or not, however, they are useful for gauging the relative costs of employing the new versus the old mix of workers prior to increases in the minimum wage.

new entrants and continuing firms highlights that the full effect of a change to the minimum wage may unfold gradually.

Several papers have analyzed minimum wages with an emphasis on firms' choice of production technology. Sorkin (2015) shows that analyses of US minimum wage changes have historically been unable to estimate truly long-run effects, meaning effects that capture changes in firms' production technologies. This reflects the inherently short-run nature of historical variation in minimum wages across US states. Aaronson et al. (2018) develop two empirical findings that connect to Sorkin's (2015) earlier insights. Using data from the Quarterly Census of Employment and Wages, they find that restaurant exit and entry both rise in the wake of minimum wage changes and that employment is stable among restaurants that remain in business. They further show that both findings make sense within a "putty-clay" model, in which a firm's production technology "hardens" following the investments the firm makes upon entry. Again focusing on the US context, Meer and West (2016) show that the effects of minimum wages may more consistently emerge as effects on job growth than as effects on employment levels—which in turn is consistent with the idea of shifting production functions.

To date, efforts to study firms' production technologies have tended to be indirect. Aaronson and Phelan (2019) and Lordan and Neumark (2018) look at data on job characteristics that reveal which jobs are more versus less replaceable by technology. Using data from the US Current Population Survey, both papers find evidence of employment declines in occupations with tasks that are readily replaced by technology in the wake of minimum wage increases. Baek, Lee, and Park (2019) analyze establishment-level data on the wages, output per worker, and capital utilization of firms in South Korea. In an analysis of repeated cross-sectional data on new firms, they find that minimum wage increases are met with higher wages, lower employment, and more capital per worker among new firms in low-wage sectors relative to new firms in high-wage sectors.

As noted earlier, the dynamics of firms' technology choices imply that the effects of the minimum wage on employment may take time to unfold. That said, some analyses find dynamic null effects. In an analysis of 138 state-level minimum wage changes in the United States, for example, Cengiz et al. (2019) find no decline in the number of lower-wage jobs over a time horizon of several years. Several other recent studies, however, which span contexts including Germany, the city of Seattle, and recent state-level minimum wage increases in the United States, find evidence that is consistent with gradually unfolding dynamic effects. In their analysis of Seattle's minimum wage ordinance, Jardim et al. (2018) find that hours were reduced disproportionately for recently hired low-wage workers, relative to more experienced low-wage workers. In the broader US context, Gopalan et al. (forthcoming) analyze employment records from Equifax, which cover millions of employees from over 2,000 firms. Over the short run, they find that low-wage employment has declined in the wake of recent minimum wage increases. This has occurred through reductions in hiring, rather than increases in firing. In a summary of research on the 2015 introduction of Germany's minimum wage, Caliendo, Schröder, and Wittbrodt (2019) similarly observe that evidence of

employment declines has involved "reduced hiring" rather than "displacements." During the last decade's economic expansion, the evidence thus suggests that the exit margin has been muted and that employment effects have been concentrated on labor-market entry.

Research on the dynamic effects of minimum wage changes may take an interesting turn in coming years. The minimum wage increases recently enacted in a number of US states are large by historical standards, and differences across states have now been in place for a number of years. In addition, the COVID-19 pandemic led to a dramatic, temporary halt to a great deal of economic activity, which included unprecedented job losses in a number of minimum-wage-intensive industries. In the post-pandemic recovery, it will be interesting to see whether viable firms differ in their labor and capital intensity in states with high minimum wages compared to states with low minimum wages.

Conclusion

Disagreement and uncertainty over the effects of the minimum wage raise important policy questions. For academics, questions of pedagogy also arise: How should the minimum wage be taught? Pedagogical commentaries from Krueger (2001) and Dolar (2013) point out that research finding small or null employment effects create teaching opportunities. Specifically, these authors highlight an opportunity to supplement the "neoclassical" economics of the minimum wage with insights from monopsony and search models. More recently, Dolar (2020) raises the issue of a "conservative"/"liberal" divide in how the minimum wage is taught. In Dolar's survey of instructors, self-described conservatives tend to believe that the minimum wage reduces employment and is an undesirable policy. Self-described liberals, by contrast, tend to believe that the minimum wage is desirable and that it does not, at historical levels, reduce employment.

I emphasize that both of these views are incomplete. Contra the conservative view, a binding minimum wage need not reduce employment. But contra the liberal view, the minimum wage can harm its intended beneficiaries even when it has no effect on employment. This paper has highlighted the teaching opportunities offered by theory and evidence on a rich set of firms' decision margins. These margins include output prices, nonwage compensation, and job attributes like effort requirements, safety measures, and the overall quality of the working environment. My central point is that accounting for these decision margins can substantively alter the conclusions one draws when using basic theory to analyze the minimum wage's effects. As shown above, these nuances are intuitive, tractable, and readily taught through straightforward extensions to standard supply and demand diagrams.

When one thinks intuitively about the responses firms might make to minimum wages, the potential relevance of health insurance benefits, effort requirements, safety measures, and schedule flexibility are obvious. Yet these nonwage dimensions of jobs pose challenges for empirical researchers. I conclude by discussing several of these empirical challenges and their implications.

A first challenge posed by nonwage job attributes involves heterogeneity. Nonwage job attributes are numerous. This introduces a "many margins" problem, in which the relevance of various attributes may vary substantially across settings. Variations may result from characteristics of industries, the nature of specific jobs, or the preferences of the workers a firm happens to employ. This implies that heterogeneous effects should be expected. Prices, for example, may adjust readily in some settings, while effort and fringe benefits may adjust more readily in others. In advance of an empirical analysis, it may not be possible to predict which margins ought to adjust most readily in a given context. In the face of many adjustment margins, it becomes an absurdity to expect quantitatively similar estimates across settings.

Second, many nonwage job attributes pose problems of measurement. For example, the data one can use to analyze the flexibility of workers' schedules or the aesthetic qualities of their workstations range from limited to nonexistent. Margins like effort can only be measured in settings where output itself is well-measured. Employee benefits are also less fully catalogued than researchers might wish. Even employer-provided health insurance, on which substantial data exist, tends to be described by binary variables, rather than in terms of the benefit's cash equivalent value.

Third, problems of measurement translate into problems of aggregation. Aggregating the value of all nonwage job attributes is infeasible because researchers lack data on all relevant margins. This fact is humbling. At the same time, further research on nonwage attributes of jobs can help us make progress in filling gaps in the available evidence.

Estimates of how a higher minimum wage affects wages and employment are not, as is commonly presumed, sufficient to assess its desirability. It is unfortunate that standard textbook analyses convey an erroneous intuition on this important point. Fortunately, there is a constructive path forward. Straightforward extensions to basic supply and demand diagrams can readily incorporate such empirically grounded possibilities as the minimum wage's effects on prices, working conditions, and noncash compensation. These diagrams provide a practical pedagogical approach to conveying important nuances about real world labor markets.

References

Aaronson, Daniel. 2001. "Price Pass-Through and the Minimum Wage." *Review of Economics and Statistics* 83 (1): 158–169.

Aaronson, Daniel, Eric French, Isaac Sorkin, and Ted To. 2018. "Industry Dynamics and the Minimum Wage: A Putty-Clay Approach." *International Economic Review* 59 (1): 51–84.

Aaronson, Daniel, and Brian J. Phelan. 2019. "Wage Shocks and the Technological Substitution of Low-Wage Jobs." *The Economic Journal* 129 (617): 1–34.

Allingham, Michael G., and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics* 1(1972): 323–38.

Ashenfelter, Orley, and Robert S. Smith. 1979. "Compliance with the Minimum Wage Law." *Journal of Political Economy* 87 (2): 333–50.

Baek, Jisun, Changkeun Lee, and WooRam Park. 2019. "The Impact of Minimum Wage Introduction on Characteristics of New Establishments: Evidence from South Korea." KDI School Working Paper 19-11.

Becker, Gary S. 1968. "Crime and Punishment: An Economic Approach." In *The Economic Dimensions of Crime*, edited by Nigel G. Fielding, Alan Clarke, and Robert Witt, 13–68. London: Palgrave Macmillan.

Bell, Brian, and Stephen Machin. 2018. "Minimum Wages and Firm Value." *Journal of Labor Economics* 36 (1): 159–95.

Belman, Dale, and Paul Wolfson. 2019. "15 Years of Research on US Employment and the Minimum Wage." *Labour* 33 (4): 488–506.

Bernhardt, Annette, Michael W. Spiller, and Nik Theodore. 2013. "Employers Gone Rogue: Explaining Industry Variation in Violations of Workplace Laws." *ILR Review* 66 (4): 808–32.

Caliendo, Marco, Carsten Schröder, and Linda Wittbrodt. 2019. "The Causal Effects of the Minimum Wage Introduction in Germany–An Overview." *German Economic Review* 20 (3): 257–92.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics* 134 (3): 1405–54.

Chen, M. Keith, Peter E. Rossi, Judith A. Chevalier, and Emily Oehlsen. 2019. "The Value of Flexible Work: Evidence from Uber Drivers." *Journal of Political Economy* 127 (6): 2735–94.

Clemens, Jeffrey. 2019. "Making Sense of the Minimum Wage: A Roadmap for Navigating Recent Research." Cato Institute Policy Analysis Paper 867.

Clemens, Jeffrey, Lisa B. Khan, and Jonathan Meer. 2018. "The Minimum Wage, Fringe Benefits, and Worker Welfare." NBER Working Paper 24635.

Clemens, Jeffrey, Lisa Khan, and Jonathan Meer. Forthcoming. "Dropouts Need Not Apply: The Minimum Wage and Skill Upgrading." *Journal of Labor Economics.*

Clemens, Jeffrey, and Michael R. Strain. 2020a. "Implications of Schedule Irregularity as a Minimum Wage Response Margin." *Applied Economics Letters* 27 (20): 1691–94.

Clemens, Jeffrey, and Michael R. Strain. 2020b. "Understanding 'Wage Theft': Evasion and Avoidance Responses to Minimum Wage Increases." NBER Working Paper 26969.

Clemens, Jeffrey, and Michael Wither. 2019. "The Minimum Wage and the Great Recession: Evidence of Effects on the Employment and Income Trajectories of Low-Skilled Workers." *Journal of Public Economics* 170: 53–67.

Congressional Budget Office. 2019. *The Effects on Employment and Family Income of Increasing the Federal Minimum Wage.* Washington, DC: Congressional Budget Office.

Coviello, Decio, Erika Deserranno, and Nicola Persico. 2020. "Minimum Wage and Individual Worker Productivity: Evidence from a Large US Retailer." Unpublished.

Davis, Leslie, and Hannah Hartig. 2019. "Two-thirds of Americans Favor Raising Federal Minimum Wage to \$15 an Hour." *Factank in the Numbers,* July 30. https://www.pewresearch.org/fact-tank/2019/07/30/two-thirds-of-americans-favor-raising-federal-minimum-wage-to-15-an-hour/.

Dolar, Veronika. 2013. "The Treatment of Minimum Wage in Undergraduate Economics Textbooks Revisited." *International Journal of Pluralism and Economics Education* 4 (2): 157–182.

Dolar, Veronika. 2020. "Are We Teaching Outside the Box? A National Survey on Teaching the Minimum Wage in Undergraduate Economics Classes." *Forum for Social Economics* 49 (1): 51–74.

Draca, Mirko, Stephen Machin, and John Van Reenen. 2011. "Minimum Wages and Firm Profitability." *American Economic Journal: Applied Economics* 3 (1): 129–51.

Drucker, Lev, Katya Mazirov, and David Neumark. 2019 "Who Pays for and Who Benefits from Minimum Wage Increases? Evidence from Israeli Tax Data on Business Owners and Workers." NBER Working Paper 26571.

Dube, Arindrajit. 2019. "Impacts of Minimum Wages: Review of the International Evidence." https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/844350/impacts_of_minimum_wages_review_of_the_international_evidence_Arindrajit_Dube_web.pdf.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp vom Berge. 2020. "Reallocation Effects of the Minimum Wage." Centre for Research and Analysis of Migration Discussion Paper 07/20.

Eibner, Christine, Michael Dworsky, Xiaoyu Nie, and Jeffrey B. Wenger. 2017. "Do Minimum Wage Changes Affect Employer-Sponsored Insurance Coverage?" Paper presented at the 7th Annual Conference of the American Society of Health Economists.

Fairris, David, and Leon Fernandez Bujanda. 2008. "The Dissipation of Minimum Wage Gains for Workers through Labor-Labor Substitution: Evidence from the Los Angeles Living Wage Ordinance." *Southern Economic Journal* 75 (2): 473–96.

Giuliano, Laura. 2013. "Minimum Wage Effects on Employment, Substitution, and the Teenage Labor Supply: Evidence from Personnel Data." *Journal of Labor Economics* 31 (1): 155–94.

Gooptu, Angshuman, and Kosali Simon. 2019. "The Effect of Minimum Wage Laws on Employer Health Insurance: Evidence from the American Community Survey." Unpublished.

Gopalan, Radhakrishnan, Hamilton H. Barton, Ankit Kalda, and David Sovich. Forthcoming. "State Minimum Wages, Employment, and Wage Spillovers: Evidence from Administrative Payroll Data." *Journal of Labor Economics*.

Goraus-Tańska, Karolina, and Piotr Lewandowski. 2019. "Minimum Wage Violation in Central and Eastern Europe." *International Labour Review* 158 (2): 297–336.

Hall, Robert E., and Charles I. Jones. 1999. "Why Do Some Countries Produce So Much More Output Per Worker than Others?" *The Quarterly Journal of Economics* 114 (1): 83–116.

Harasztosi, Peter, and Attila Lindner. 2019. "Who Pays for the Minimum Wage?" *American Economic Review* 109 (8): 2693–2727.

He, Haoran, David Neumark, Qian Weng 2019. "Do Workers Value Flexible Jobs? A Field Experiment on Compensating Differentials." NBER Working Paper 25423.

Horton, John. 2018. "Price Floors and Employer Preferences: Evidence from a Minimum Wage Experiment." Unpublished.

Ippolito, Richard A. 2003. "The Impact of the Minimum Wage if Workers Can Adjust Effort." *The Journal of Law and Economics* 46 (1): 207–27.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, and Hilary Wething. 2018. "Minimum Wage Increases and Individual Employment Trajectories." NBER Working Paper 25182.

Katz, Lawrence F., and Alan B. Krueger. 1992. "The Effect of the Minimum Wage on the Fast-Food Industry." *ILR Review* 46 (1): 6–21.

Krueger, Alan B. 2001. "Teaching the Minimum Wage in Econ 101 in Light of the New Economics of the Minimum Wage." *The Journal of Economic Education* 32 (3): 243–58.

Ku, Hyejin. Forthcoming. "Does Minimum Wage Increase Labor Productivity? Evidence from Piece Rate Workers." *Journal of Labor Economics*.

Lemos, Sara. 2008. "A Survey of the Effects of the Minimum Wage on Prices." *Journal of Economic Surveys* 22 (1): 187–212.

Leonard, Thomas C. 2000. "The Very Idea of Applying Economics: The Modern Minimum Wage Controversy and Its Antecedents." *History of Political Economy* 32 (5): 117–44.

Leung, Justin H. Forthcoming. "Minimum Wage and Real Wage Inequality: Evidence from Pass-through to Retail Prices." *Review of Economics and Statistics.*

Lordan, Grace, and David Neumark. 2018. "People versus Machines: The Impact of Minimum Wages on Automatable Jobs." *Labour Economics* 52: 40–53.

Luca, Dara Lee, and Michael Luca. 2019. "Survival of the fittest: the impact of the minimum wage on firm exit." NBER Working Paper 25806.

MacDonald, James M., and Daniel Aaronson. 2000. "How Do Retail Prices React to Minimum Wage Increases?" Federal Reserve Bank of Chicago Working Paper 00-20.

MaCurdy, Thomas. 2015. "How Effective is the Minimum Wage at Supporting the Poor?" *Journal of*

Political Economy 123 (2): 497–545.

Machin, Stephen, Alan Manning, and Lupin Rahman. 2003. "Where the Minimum Wage Bites Hard: Introduction of Minimum Wages to a Low Wage Sector." *Journal of the European Economic Association* 1 (1): 154–80.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger. 2018. "The Value of Working Conditions in the United States and Implications for the Structure of Wages." NBER Working Paper 25204.

Marks, Mindy S. 2011. "Minimum Wages, Employer-Provided Health Insurance, and the Non-discrimination Law." *Industrial Relations: A Journal of Economy and Society* 50 (2): 241–62.

Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review* 107 (12): 3722–59.

McKenzie, Richard B. 1980. "The Labor Market Effects of Minimum Wage Laws: A New Perspective." *Journal of Labor Research* 1 (2): 255–64.

Meer, Jonathan, and Jeremy West. 2016. "Effects of the Minimum Wage on Employment Dynamics." *Journal of Human Resources* 51 (2): 500–22.

Mincer, Jacob. 1974. *Schooling, Experience, and Earnings.* New York: NBER.

Moretti, Enrico. 2010. "Local Multipliers." *American Economic Review* 100 (2): 373–77.

Moretti, Enrico. 2011. "Local Labor Markets." In *Handbook of Labor Economics*, Vol. 4b, Orley Ashenfelter and David Card, 1173–11236. Amsterdam: Elsevier.

Neumark, David. 2019. "The Econometrics and Economics of the Employment Effects of Minimum Wages: Getting from Known Unknowns to Known Knowns." *German Economic Review* 20 (3): 293–329.

Obenauer, Marie Louise, and Burtha von der Nienburg. 1915. *Effect of Minimum-Wage Determinations in Oregon.* Washington, DC: US Government Printing Office.

Orrenius, Pia M., and Madeline Zavodny. 2008. "The Effect of Minimum Wages on Immigrants' Employment and Earnings." *ILR Review* 61 (4): 544–63.

Orrenius, Pia M., and Madeline Zavodny. 2011 "The Minimum Wage and Latino Workers." In *Latinos and the Economy: Integration and Impact in Schools, Labor Markets, and Beyond*, edited by David L. Leal and Stephen J. Trejo, 169–91. New York: Springer.

Renkin, Tobias, Claire Montialoux, and Michael Siegenthaler. Forthcoming. "The Pass-Through of Minimum Wages into US Retail Prices: Evidence from Supermarket Scanner Data." *Review of Economics and Statistics.*

Royalty, Anne. 2001. "Do Minimum Wage Increases Lower the Probability that Low-Skilled Workers Will Receive Fringe Benefits?" JCPR Working Paper 222.

Saad, Lydia. 2013. "In U.S., 71% Back Raising Minimum Wage." *Gallup News,* March 6. https://news.gallup.com/poll/160913/back-raising-minimum-wage.aspx.

Sabia, Joseph J., and Richard V. Burkhauser. 2010. "Minimum Wages and Poverty: Will a $9.50 Federal Minimum Wage Really Help the Working Poor?" *Southern Economic Journal* 76 (3): 592–623.

Simon, Kosali Ilayperuma, and Robert Kaestner. 2004. "Do Minimum Wages Affect Non-Wage Job Attributes? Evidence on Fringe Benefits." *ILR Review* 58 (1): 52–70.

Sorkin, Isaac. 2015. "Are There Long-Run Effects of the Minimum Wage?" *Review of Economic Dynamics* 18 (2) 306–33.

Summers, Lawrence H. 1989. "Some Simple Economics of Mandated Benefits." *American Economic Review* 79 (2): 177–83.

US Bureau of Labor Statistics. 2020a. "Consumer Expenditure Surveys: Annual Calendar Year Tables for 2019." United States Department of Labor. https://www.bls.gov/cex/tables.htm (accessed October 5, 2020).

US Bureau of Labor Statistics, 2020b. "Employer Costs for Employee Compensation – December 2019." United States Department of Labor. https://www.bls.gov/news.release/ecec.nr0.htm (accessed April 27, 2020).

Weil, David. 2005. "Public Enforcement/Private Monitoring: Evaluating a New Approach to Regulating the Minimum Wage." *ILR Review* 58 (2): 238–57.

Wessels, Walter J. 1980. "The Effect of Minimum Wages in the Presence of Fringe Benefits: An Expanded Model." *Economic Inquiry* 18 (2): 293–313.

The Rise of American Minimum Wages, 1912–1968

Price V. Fishback and Andrew J. Seltzer

Minimum wages have been among the most controversial government interventions in labor markets for more than a century. We focus on the political economy of the rise of the American minimum wage from its beginnings as state laws for women and children in the 1910s through the years when the federal minimum wage reached its peak value in real terms in the late 1960s and the coverage of the workforce became nearly complete in the early 1970s. Figure 1 documents the trend rise in the real value of the federal minimum between 1938 and 1968 as well as its subsequent decline to levels that had been reached in the early 1950s.

In the course of the discussion, we will suggest five main themes. First, states legislated the original minimum wage laws starting in 1912 and played prominent roles in raising minimum wages in the 1940s and adding coverage in the 1950s. States have again led the way in raising minimum wages in the last 25 years.

Second, the Supreme Court limited the options of minimum wage advocates until 1937. In close votes in the 1920s, the justices struck down the state minimum wage laws, weighing freedom of contract against establishing a minimum standard for health and safety of female workers. Declining income during the Depression led the states to enact new minimum wage laws for women and children. Meanwhile, Presidents Hoover and Roosevelt pressured employers to negotiate voluntarily to

■ *Price V. Fishback is the APS Professor of Economics, University of Arizona, Tucson, Arizona. Andrew J. Seltzer is Professor of Economics and Economic History, Royal Holloway, University of London, United Kingdom. Their email addresses are pfishback@eller.arizona. edu and A.Seltzer@rhul.ac.uk.*

Figure 1
Nominal and Real Federal Minimum Wage, 1938–2019
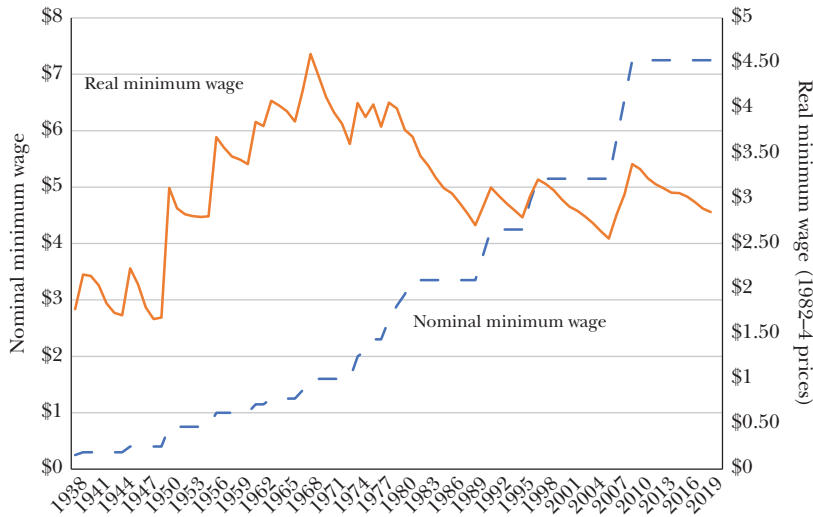
Source: Minimum wage is from Carter et al. (2006, series Ba4422, pp. 2–284). The real minimum wage is in 1982–1984 prices, using the Consumer Price from Federal Reserve Bank of St. Louis (2019, series CUUR0000SA0), downloaded in August 2019.

limit hours and set wage minimums to promote employment with earnings that covered basic needs. In 1936, the Court struck down a new type of state minimum wage law along with also striking down the National Industrial Recovery Act of 1933. However, in 1937, the Court reversed these decisions and opened the door for the federal government to set minimum wages for nearly all workers.

Third, during the New Deal of the 1930s, the federal government accepted more responsibility for regulating labor markets and combatting poverty, and the struggle over the minimum wage shifted to Congress. Proponents of the minimum called for rates that would provide a decent standard of living for workers and for an end to cutthroat wage competition that endangered the health and safety of low-wage workers. Conversely, southern leaders, who held seniority in major committees, sought to defend the low wages matched to low costs of living that had promoted the development of southern industries. The ultimate compromise in the Fair Labor Standards Act (FLSA) of 1938 called for a slow rise in the minimum over time, industry-specific minimums, and coverage of only about half of the workforce.

Fourth, each post-World War II US president until Ronald Reagan both supported and presided over a rise in the minimum wage. The compromises in the 1949 and 1955 amendments raised the minimum wage without expanding coverage, while the 1961, 1966, 1974, and 1977 amendments raised the minimum and expanded coverage to nearly all of the work force. The real value of the minimum

has followed the sawtooth pattern shown in Figure 1, with a jump following each nominal rise and then a decline as inflation eroded purchasing power.

Fifth, the debates among economists concerning the minimum wage came in waves and built on common themes debated since the Progressive Era while making use of evolving language that incorporated Joan Robinson's (1933) analysis of *labor monopsony* and Depression-era claims that minimum wages would lead to macroeconomic stimulus. An empirical debate in 1960 in the *Industrial and Labor Relations Review* set the stage for the empirical debates that followed. Earlier generations of the minimum wage debates anticipated many issues in the modern literature—causal inference, omitted variables, pre-trends, differences-in-differences differences, and discontinuities—but had to rely on cross-tabulations of aggregate data. Over the past 60 years, the empirical debates have had increasing access to disaggregated data and new econometric methods that make use of increasing computing power.

Early State-level Minimum Wage Laws for Women and Children

In the early 1900s, labor law was the responsibility of state and local governments. The primary argument given by supporters of the early state labor laws covering women and children was to end "sweated" labor—hard work for very low pay under poor conditions (Webb 1912; Women's Bureau 1928; Commons 1935)—which was often discussed as a health and safety issue. The US Supreme Court upheld state-level laws placing limits on men's daily hours in dangerous industries like mining and ore smelting refining as early as *Holden v. Hardy* (169 US 366 [1898]). However, the Court struck down a New York state law limiting the hours of male bakers in *Lochner v. New York* (198 US 45 [1905]) in a 5-4 decision on the grounds that the limit violated freedom of contract and that long hours in baking were not a threat to safety or public health. In contrast, the Court upheld a general limit on women's hours law in all industries in *Muller v. Oregon* (208 US 412 [1908]), reasoning that it was socially important for women to bear "vigorous" offspring, and that hours limitations for women could be justified on the grounds that "the physical well-being of woman becomes an object of public interest and care in order to preserve the strength and vigor of the race" (Fishback 2018).

Despite a surge in the discussion of the minimum wage for all workers in the major economics journals between 1909 and 1913, their experience with previous attempts to pass hours laws led reformers to find if more fruitful to devote their efforts to obtain a minimum wage only for women and children. Their efforts met with success when Massachusetts enacted the first minimum wage law in 1912. Oregon soon followed in 1913 and in the next decade Washington, DC, and 14 more states joined them. Table 1 lists the states that enacted minimum wage laws from 1912–1937, together with the coverage of the laws. A few states set flat minimum wages, while most copied earlier minimum wage laws in Australia, New Zealand, and the United Kingdom and created commissions to issue minimum wage orders by sector.

Table 1
State Minimum Wage Laws, 1911–37

State	Date(s)	Coverage	Exceptions (1938)
Arizona	1917–25, 1937	W, C21	A, B
Arkansas	1915–27, 1937	W	B, C, D
California	1913	W, C18	NONE
Connecticut	1933	W, C18	A, B
Colorado	1913	W, C18	NONE
District of Columbia	1918–23, 1938	W, C18	A
Illinois	1933	W, C18	A, B
Kansas	1915–27, 1938	W, C21	NONE
Kentucky	1938	W, C21	A, B, E
Louisiana	1938	W	A, B, I
Massachusetts	1912–34, 1934	W, C21	A, B
Minnesota	1913–25, 1937	W	NONE
Nebraska	1913–19	W	
Nevada	1937	W, C21	A
New Hampshire	1933	W, C21	A, B
New Jersey	1933	W, C21	A, B, F
New York	1933–36, 1937	W, C18	A, B
North Dakota	1919	W, C21	A, B
Ohio	1933	W, C21	A, B
Oklahoma	1937	All	B
Oregon	1913	W, C18	NONE
Pennsylvania	1937	W, C21	A, B, G, H
Rhode Island	1936	W, C21	A, B
South Dakota	1923	W	A, B
Texas	1919–21	W, C15	
Utah	1913–29, 1933	W, C18	NONE
Washington	1913	W, C18	NONE
Wisconsin	1913–24, 1937	W, C21	NONE

Source: United States Bureau of Labor Statistics (1933b), Phelps (1939), p. 60, and Seltzer (1994).
Note: W – women, C21 (C18) – children 21 (18) years of age or less, A – domestic service, B – agricultural labor, C – cotton factories, D – railroads covered by federal laws, E – firms regulated by the state Public Service Commission, F – hotels, G – religious or charitable organizations, H – sale and delivery of newspapers and magazines, I – municipalities with a population under 10,000.

While reformers focused on passing minimum wages for women and children, the debates in the economics profession centered on minimums for all workers. Institutional economists associated with the American Association of Labor Legislation (AALL) played significant roles in arguing for the women's minimum wage during the Progressive Era. Notables John Andrews, John R. Commons, Richard Ely, Henry Seager, and Leo Wolman drew lessons from minimum wages set in Australia, New Zealand, and the United Kingdom, where Sidney Webb (1912) and Matthew Hammond (1915) claimed minimum wages had yielded wage increases with at worst weak negative consequences.

Prasch (2000, 2007) identifies several themes in their arguments. First, women had limited opportunities and weak bargaining power. John Bates Clark (1913,

p. 292) suggested that they might be hired for less than their "worth as measured by the productivity test." Second, minimum wages would prevent payment of less than subsistence wages to "boy labor, girl labor, married women's labor, the labor of old men, of the feeble-minded, of the decrepit and broken-down invalids, and all the other alternatives to the engagement of competent male adult workers at a full Standard Rate" (Webb 1912, p. 986). Such low wages damaged the health, training, and education of their operatives in ways that would damage future generations. Third, Sidney Webb (1912, pp. 986–88) argued that low-wage employers in "sweated trades" were "parasitic." Their payment of low wages to their operatives drove wages down for more productive workers. He recognized that a minimum wage would reduce employment among these groups, but he preferred to eliminate the hidden subsidies to the sweatshop employers and have society directly pay the cost of providing the formerly sweated workers with the education, training, health care, and temporary poor relief necessary to make them successful in the longer run. In his view, introducing minimum wages would force employers to compete by paying appropriate wages to adult workers and by increasing productivity with new machinery and more efficient organization of labor. This theme was one that often led large, unionized, and more productive employers to join reformers in supporting other progressive labor legislation as well (Fishback 1998).[1]

John Bates Clark (1913), Frank Taussig (1916), A. C. Pigou (1913), and Frank Fetter (1917) wrote about the minimum wage in the marginalist tradition, which later evolved into neoclassical economics. They warned of the negative consequences of a minimum wage for employment. Bates Clark (1913, p. 294) summarized their reasoning: "What is probable, even as the result of a more modest legal increase of pay, is as follows: Of the lowest grade of workers some would be promoted to a higher rank and some would be discharged. The output of the business would be reduced." Taussig (1916) argued against views that "parasitic" industries were a cause of low wages, instead laying the blame for low wages more on a large supply of low-skilled women that was made larger by immigration. Clark and Fetter seemed willing to experiment with minimum wages, but Clark (1913, pp. 296–7) argued that it should be accompanied by emergency relief for those left unemployed. Fetter and Taussig suggested that poverty might better be prevented through improved education, training, and housing and immigration restrictions.

These early minimum wage laws applying to women and children were under continual challenge in the court system. The constitutionality of such laws was supported by state supreme courts in Arkansas, Massachusetts, Minnesota, and Oregon, and Washington (Clark 1921, p. 33). A 4–4 tie in the US Supreme Court in the case of *Stettler v. O'Hara* (243 US 629 [1917]) left the Oregon minimum wage law in place. But in 1923, the Supreme Court declared the District of Colombia law

---

[1] Leonard (2005, pp. 212–15) claims that a number of progressive economists justified the minimum wage on eugenics grounds. The arguments economists made in the prominent minimum wage papers, however, made few mentions of eugenics issues. Those who held eugenic views wrote about them in other contexts outside of the minimum wage debate.

unconstitutional by a 5–3 vote in *Adkins v. Children's Hospital* (261 US 525 [1923]).[2]
The majority affirmed the doctrine of "freedom of contract," and stated that "wages,
unlike hours affected health only 'indirectly or remotely.'" Oliver Wendell Holmes
dissented, arguing that the correct goal of a minimum wage for women and chil-
dren was to remove conditions causing "ill health, immorality, and the deterioration
of the race" (as quoted by Cushman 1998, pp. 67, 69)

Despite the series of court rulings, several states continued to revise their
minimum wages throughout the 1920s and into the 1930s, albeit without formal
sanctions (US Bureau of Labor Statistics 1933b, pp. 1344–54). The absence of
enforcement was not really much of a change. As was the case for nearly all labor
laws at the time, regulatory budgets were meager at best; therefore, enforcement
relied on the employer's willingness to abide by the law, pressure from workers,
and public opinion. Case studies of minimum wages in the 1910s and 1930s showed
increased earnings and declines in hours of employment for women subject to the
minimums, although there was an intense debate about whether other factors that
could lead to higher wages and lower hours were actually being held constant in
these analyses (as discussed in Peterson 1959, 1960; Lester 1960). More recently,
Marchingiglio and Poyker (2020) find statistically significant effects on female
employment in a triple-difference analysis of state minimum wage laws using the
full censuses from 1880 through 1930.

By the early 1930s, minimum wage advocates had not lost hope, because the
earlier votes in the minimum wage court decisions had been close and four seats
on the US Supreme Court had turned over. In the early 1930s, states passed new
minimum wage laws (as shown in Table 1). Most were based on a standard bill spon-
sored by the National Consumers' League that was designed to avoid the language
used in the unconstitutional acts (US Bureau of Labor Statistics 1933a, p. 1259).
The Bureau of Labor Statistics (1933b, p. 1346) supported the laws on grounds that
the Depression had made it "apparent that unfair wage standards not only under-
mine the health and well-being of the workers but threaten the stability of industry
itself." The laws protected "the public health, morals and welfare."

In 1936, the Supreme Court struck down the 1933 New York minimum wage
law with a 5–4 vote in *Morehead v. New York ex. Rel. Tipaldo* (298 US 587 [1936]), even
though the framers of the law had tried to differentiate it from the Washington,
DC, law that had been declared unconstitutional in 1923. Chief Justice Hughes
(p. 619) dissented: "I can find nothing in the Federal Constitution which denies
to the state the power to protect women from being exploited by overreaching
employers." In a separate dissent, Justices Stone, Brandeis, and Cardozo declared
that in the prior decade, society had learned that a wage is not always the outcome
of free bargaining; that it may be "forced upon employees by their economic neces-
sities and upon employers by the most ruthless of their competitors." In their view,

---

[2] Only eight Supreme Court justices voted in *Stettler* and *Adkins*, because Justice Louis Brandeis recused
himself after having been an advocate for these minimum wage laws in the lower courts.

insufficient wages burdened society as a whole, and the problem should be solved by the legislative branch (pp. 635–6).

In 1937, however, the Supreme Court reversed course when Justice Owen Roberts switched sides and declared the Washington minimum wage from 1913 to be constitutional in *West Coast Hotel v. Parrish* (300 US 379 [1937]). Roberts had joined the court after the 1923 *Adkins* decision and later claimed that he had voted against the New York minimum in 1936 because it was similar to the Washington, DC, law and New York's lawyers failed to challenge the *Adkins* ruling. When lawyers for Washington state directly challenged the *Adkins* decision in 1937, he chose to support their law because women were "especially liable to be overreached and exploited by unscrupulous employers," which was "not only detrimental to the health and wellbeing of the women affected, but casts a direct burden for their support upon the community."[3] A number of states quickly passed new laws for women and children and the door opened for more intense pressure for a minimum for all workers.

The Lead-up to a National Minimum Wage for All Workers

While the states focused on women's minimum wages in the early 1930s, federal government officials pressured firms to agree voluntarily to accept wage minimums and hours maximums. Between 1929 and 1931, President Hoover "jawboned" leading manufacturers to set up work-sharing arrangements with reduced weekly hours, more employment, and no reductions in hourly earnings (Rose 2010; Neumann, Taylor, and Fishback 2013). A centerpiece of President Franklin Roosevelt's New Deal was the National Industrial Recovery Act (NIRA) of 1933, which called for employers, workers, and consumers in each sector to limit excessive competition by negotiating codes that included minimum wages and maximum weekly work hours. While waiting for the codes, Roosevelt convinced thousands of firms to sign President's Reemployment Agreements (PRAs) that set maximum hours and minimum wages with the stated goal to "raise wages, create employment, and thus increase purchasing power and restore business." During the following months, over 500 sectors then set up National Recovery Administration (NRA) codes of competition that included sector-specific minimum wages (sometimes with multiple tiers) and weekly hours maximums.

The President's Reemployment Agreements and the codes differed from a statutory minimum wage because they were voluntary. Employers agreed to the constraints by signing the PRA in August 1933 or by later signing onto one of the industry codes. If firms/employers did not sign the code or agreement, however, they were not subject to the minimum wage or the maximum hours. Thus, the PRA

---

[3] Some claim that Roberts switched sides to dissuade President Roosevelt from his scheme to add justices to the Supreme Court. Cushman (1998) and Frankfurter (1955) disagree, noting that the vote was in December well before the scheme was announced.

and NRA minimums were based on bargaining, unlike statutory minimum wages. So why did firms sign up? The major gain came from being seen as patriotic, because the firm could prominently display the NRA's Blue Eagle. The federal government marketed the connection between patriotism and the Blue Eagle through parades and house-to-house drives, garnering signatures on pledges from over 20 million householders that they would favor Blue Eagle firms (Taylor 2011). Taylor (2019, chapter 4) and Meeker (1933, pp. 467–8) both suggested an undercurrent of coercion as well. The administration sought to make firms believe that noncompliance would cost them dearly with unspoken threats of boycotts. In August 1933, the mercurial NIRA head General Hugh Johnson announced, "the time is coming when someone is going to take one of those Blue Eagles off of someone's window in a clear cut case and that is going to be a sentence of economic death" (*Detroit Free Press* 1933).

Blue Eagle displays in newspaper ads and compliance with the labor restrictions declined after November 1933, and the National Recovery Administration faced a major compliance crisis in the spring of 1934 (Taylor 2019, Chapters 6 and 7, Figure 4; Bowden 1937). Meanwhile, the NRA codes were soon challenged in court. On May 27, 1935, in *L. A. Schechter Poultry Corp. v. United States* (295 US 495 [1935]), the Supreme Court unanimously struck down all of the NRA codes on grounds that the National Industrial Recovery Act of 1933 delegated the authority to make regulations to market participants and such delegation of power was unconstitutional (Taylor 2011).

In 1937, two Supreme Court decisions opened the door for a constitutional federal minimum wage law. The *West Coast Hotel* decision (mentioned earlier) established that the courts were prepared to accept arguments for a minimum wage. The case *National Labor Relations Board v Jones & Laughlin Steel Corporation* (301 US 1, 1937) established the "interstate commerce clause" as the constitutional basis for federal labor legislation. In 1937, the Roosevelt administration introduced the Fair Labor Standards Act, drafted by Secretary of Labor Frances Perkins. The original bill specified a 40-cent hourly minimum wage but appointed Industry Committees—comprising industry insiders and members of the public—who would set industry-specific rates as high as 80 cents per hour, well above both estimates of a living wage and prevailing wage rates (Congressional Record 1937, p. 439).[4]

The heated arguments in Congress for and against the Fair Labor Standards Act encompassed a broad array of reasoning. Advocates restated the arguments made by progressive economists circa 1912, which gained more power from the drops in worker income from 1929 to 1933 and again in 1937–38. Northern Democrats and some Republicans claimed a moral imperative was needed to end sweated labor. President Roosevelt spoke of the need to end "starvation wages." Senator William Borah (R-ID) proclaimed people worthy of hire are "entitled to sufficient compensation to maintain a decent standard of living" (Congressional Record 1938, p. 7793).

---

[4]Seltzer (1995, 1997) describes the Fair Labor Standards Act controversy in depth and is the source for this three-paragraph discussion.

In the January 1938 *Monthly Labor Review*, the Department of Labor described the criteria for a minimum wage: "Certain basic standards of adequacy are generally recognized as inherent in the concept of a minimum wage based on the cost of living" (Stitt and Smith 1938, p. 201). One new argument raised during the 1937–38 recession was that minimum wages would promote national recovery by increasing aggregate demand (Grossman 1978; Congressional Record 1937, p. 7745).

The US Senate passed the bill in July 1937, but when it reached the House of Representatives, the debate took on a strong North vs. South aspect. High-wage producers of shoes and textiles in New England and lumber in the Northwest supported the minimum wages in order to eliminate "excessive," "unfair" competition from the low-wage South (Congressional Record 1937, pp. 439, 505–6, 517). Most southern senators fought the bill because nearly half of all southern manufacturing workers, and higher shares in textiles and lumber, earned less than 40 cents per hour in 1937 (Seltzer 1995; Wright 1987). They opposed Industry Committees because they expected northern interests to capture them and set minimum hourly wages near 80 cents and disallow regional differentials. Some northern Republicans joined the opposition because they considered the policy to be an unwarranted federal government intrusion into private contracts. Union leaders also objected because they believed that wage bargaining was their exclusive domain.

The regional issue was likely intertwined with race, although the issue rarely was discussed in the 1937 and 1938 Congressional debates. The agriculture and domestic service sectors, which employed large numbers of Black workers, were explicitly exempted from the Fair Labor Standards Act. These exemptions were also part of the Social Security pension program, Unemployment Insurance, state workers' compensation laws, and the earlier state minimum wage laws. They also arose for other reasons: 1) considerations that farmers, small employers, and households could not afford the costs of coverage, 2) inherent problems with enforcing the law for those groups, and 3) and judicial uncertainty for federal programs about whether the sectors could be covered under the "interstate commerce clause" of the US Constitution.

After months of political wrangling, the minimum wage bill looked set to die in committee, in part because southern Democrats held outsized representation and more seniority on the key committees. The tide turned in May 1938 when ardent minimum wage supporters Senator Claude Pepper (D-FL) and Representative Lister Hill (D-AL) overwhelmingly beat opponents of the bill in Democratic primaries. The House then passed a substantially rewritten bill in late May 1938. The final Act phased in the minimum wage at 25 cents in October 1938, rising to 30 cents in October 1939, and then gave Industry Committees the discretion to raise it to 40 cents between October 1940 and October 1945. It was confined to employees who were "engaged in interstate commerce or the production of goods for interstate commerce." The entire public sector, agriculture, retail establishments, local services, domestic service, and small businesses were excluded from coverage. The constitutionality of the Fair Labor Standards Act was upheld in 1941 by a 9–0 Supreme Court vote in *United States v. Darby* (312 US 100 [1941]).

Public opinion at this time generally favored minimum wages. In May 1938, the Gallup organization asked: "Should Congress pass a law regulating wages and hours before ending this session?" A majority of 59 percent said "yes," although an August 1938 poll indicated that opposition to the law was stronger than for almost all other New Deal Legislation (American Institute of Public Opinion 1939). By January 1939, 71 percent expressed support of the newly passed act, while a 62 percent majority supported regional differences in a June 1938 poll. Support for a single national minimum would not reach as high as 46 percent until a Gallup poll in September 1948.[5]

The effects of the minimum wage in the late 1930s varied across industries and regions. Virtually all jobs in which minimum wages are binding today were exempted from coverage under the original act. A very large majority of manufacturing employees in northern and western states earned well above the prevailing minimum rates. In southern industries, like lumber and tobacco processing, which employed a significant number of Blacks, around 70 percent of workers were earning the 30-cent minimum or one penny more in 1939 and 1940–41, respectively. In 1940 in the seamless hosiery industry, which employed few Blacks, about one-third of firms paid average hourly wages within 2.5 cents of the then-prevailing minimum rate of 32.5 cents.[6]

The binding nature of the minimum wage rates established under the 1938 Act, however, were short-lived. The last Industry Committee was established in January 1942, and by July 1944, all 71 Committees had established the $0.40 minimum rate for their industry. War-driven increases in labor demand caused prevailing wages to exceed the federal minimum by the middle of the war, and federal policy shifted to setting wage ceilings. A 1947 survey in the *Monthly Labor Review* showed that at least 95 percent of sawmill workers in every state earned at least $0.45 per hour, $0.05 over the minimum.

One goal of the minimum wage was to help low-income people obtain an adequate standard of living. Budgets required to meet that fairly vague standard have been estimated by American observers and government agencies since as early as 1870 (Barrington and Fisher 2006, pp. 2–629 to 2–647; Lamale and Stotz 1960, p. 789). Prior to 1946, the US Bureau of Labor Statistics relied upon a Works Progress Administration consumption bundle for a family led by a working "unskilled manual worker," his nonworking wife, an 8-year old girl, and a 13-year old boy. The budget did not reach the level "the skilled worker hoped to attain, but [afforded] more than a 'minimum subsistence' living" (Stecker 1937, pp. xiii–xiv). Full-time work at the minimum wage in 1940 would have covered between 41 and 53 percent of the budget estimate for a four-person family, and roughly

[5] Gallup poll results from Roper Center (2020). Gallup Poll ID numbers that start with USGALLUP:060138. RA05A and 48070F. R25A.
[6] See Seltzer (1997, 2002). Wright (1987) suggests that the minimum wage played a role in integrating southern labor markets with the rest of the country, but we would put more weight on his discussion of the World War II war boom, which raised wages and generated substantial internal migration.

80 to 104 percent for a single person (as shown later in Table 3 for the cities listed).

With such realities in mind, states quickly took the lead after the passage of the Fair Labor Standards Act of 1938 in setting minimum wages above the federal minimum and also in expanding the occupations covered. In 1939, 18 states set at least one minimum higher than the 30-cent federal minimum (Women's Bureau 1939). In addition, roughly half of the states soon filled some of the large gaps in coverage under by the Fair Labor Standards Act by setting minimum wages, mostly for women, in industries such as retail stores, laundries, offices, hospitals, beauty shops, hotels, recreation, and in some cases manufacturing. Most of the state laws established boards that set minimums that varied by occupation, hours of work, age, gender, and experience. Over time, many of the states added men and covered more occupations (Women's Bureau 1942, 1951, 1953, 1958).

Post World-War II Minimum Wage Increases: The Climb to the Peak

Between 1945 and 1967, Presidents Truman, Eisenhower, Kennedy, and Johnson all supported increases in the minimum wage and expansions of coverage. The term "minimum wage" appeared in 16 presidential State of the Union addresses between 1937 and 1967 (*Complete State of the Union Addresses)* and each president signed a minimum wage bill, as shown in Table 2.

While the federal minimum did not change from its passage until 1950, the states again took the lead in raising minimums after the war. Between 1942 and 1950, when the federal minimum wage was 40 cents, state boards issued 77 orders; there were three above 70 cents, 29 between 60 and 69 cents, 25 from 50 to 59 cents, and 11 from 40 to 49 cents. Only nine were below 40 cents (Women's Bureau 1951, p. 2). The higher state minimums were almost entirely for female workers who typically earned much less than males.

The federal government finally retook the lead again when it raised the minimum to 75 cents in 1950 and $1 in 1956. In his 1955 State of the Union Address, Republican President Eisenhower (1955) cited strong economic growth over the previous five years as a reason to raise the minimum wage to 90 cents from 75 cents and to extend the coverage to more classes of workers. Liberal Democrats sought a much higher minimum of $1.25. During debates over an amendment to the minimum wage bill on June 8, 1955, Senate Leader Lyndon Johnson noticed that most of the strongest advocates from both sides had left the Senate floor and jammed a minimum wage of $1 per hour with no expansion of coverage through the Senate on a voice vote.[7] After passage in the House and adjustments in conference, Eisenhower signed the bill into law on August 12 (*New York Times* 1955).

---

[7] See Caro (2002, pp. 609–612). Caro claims that Johnson had opposed raising the minimum wage the previous time the issue came up before the Senate.

Table 2

The Fair Labor Standards Act and Amendments

Year	Minimum Wage	Coverage	Notes
1938	$0.25 (1938) $0.30 (1939) $0.40 (1945)	≈50% of private sector employees (1938)	Establishes minimum wage. Exemptions for small firms, firms outside interstate commerce, retail and service establishments, restaurants, government employees, agricultural workers, domestic service.
1939–1943	$0.40 for all by July 1944.		Under FLSA 1938 Act 71 Industry Committees, with economics professors composing about half of the members, set industry-specific minimum wages for industries with approximately 21 million workers. The goal was to establish a minimum wage that would not lead to "substantial curtailment of employment." By July 1944 all 71 had minimum rates at 40 cents.
1947			Portal to Portal Act: limits claims to back pay to 2 years, restricts definition of working time, no liquidated damages if employers made a good faith effort to comply with FSLA.
1949	$0.75	55% of private sector employees (1950)	Coverage extended to employees if they are "directly essential" to production of goods for interstate commerce, extends coverage to air transport industry, eliminates Industry Committees.
1955	$1.00	55% of private sector employees (1956)	
1961	$1.00, $1.15 (1961) $1.25 (1963)	63% of private sector employees	Coverage extended to retail and service establishments, with sales exceeding $1 million annually. Students could be employed at these establishments at 15% less than the minimum wage. Minimum for newly covered employees was $1.00.
1966	$1.40, $1.00 (1967) $1.60, $1.15 (1968) $1.60, $1.30 (1969) $1.60, $1.45 (1970)	77% of private sector employees, 40% of government employees	Coverage extended to most farm workers (with a lower minimum wage). Coverage extended to enterprises with revenue greater than $500,000 per year (1966) then $250,000 per year (1969). Automatic coverage for construction workers and employees of schools, hospitals, nursing homes, or other residential care facilities.
1974	$2.00 (1974) $2.10 (1975) $2.30 (1976)	83% of private sector employees. After 1974 coverage varies between 83% and 87%. 100% of public sector employees in 1974. Supreme Court decisions reduce public sector coverage reduced to 28% in 1976 and restore it to 100% in 1990.	Coverage extended to domestic workers and state and local government employees that were not previously covered. Minimum wage for farm workers raised to $1.60 in 1974.
1977	$2.65 (1978) $2.90 (1979) $3.10 (1980) $3.35 (1981)		Sub-minimum wage for agricultural workers eliminated. Restrictions on subminimum wage for students relaxed. Coverage test for retail trade and service enterprises increased in stages to $362,500.
1983			Migrant and Seasonal Agricultural Worker Protection Act — increases coverage for some farm workers.
1986			Allows sub-minimum wages for employees whose earning or productive capacity is impaired by age, physical or mental deficiency, or injury.
1989	$3.80 (1990) $4.25 (1991)		Enterprise coverage limit for retail and non-retail businesses increased to $500,000.
1996	$4.75 (1996) $5.15 (1997)		Established a youth sub-minimum wage of $4.25 an hour for newly hired employees under age 20 during their first 90 consecutive calendar days after being hired by their employer; revised the tip credit provisions to allow employers to pay qualifying tipped employees no less than $2.13 per hour if they received the remainder of the statutory minimum wage in tips.
2007	$ 5.85 (2007) $ 6.55 (2008) $ 7.25 (2009)		

Source: US Department of Labor (2020), *Columbia Law Review* (1948), Neumark and Wascher (2008).

Coverage under the Fair Labor Standards Act was finally extended in 1961, with newly covered employees receiving a minimum wage of $1.00. However, 14 states had a $1.00 or higher minimum rate for at least some of these workers by 1959 (Women's Bureau 1958). By the late 1960s, the federal minimum rate reached its peak in real terms and gaps in coverage were successively closed and state laws became less binding. In 1970, only Alaska ($2.10) and the District of Columbia ($2.00) had minimum wages above the federal rate of $1.60.

Congressional supporters of a higher minimum wage were typically liberal Democrats, including Senator Paul Douglas (1972, pp. 374–80), a former University of Chicago professor and president of the American Economic Association. The opposition was comprised largely of southern Democrats and some, but not all, Republicans. The primary public arguments about the minimum wage shifted again in the 1950s and 1960s. Because the federal minimum covered males, earlier Progressive arguments about the need to protect the health of prospective mothers diminished. The post-World War II boom also contributed to less emphasis on the Depression-era arguments about stimulating aggregate demand. The argument that the minimum wage reduced "unfair competition" never vanished completely, but it receded in importance because regional wage differences narrowed and gaps in coverage (mainly affecting the untraded sector) were successively closed.

The main argument for the minimum wage expressed during this period in the *Congressional Record* became the moral imperative to protect the most vulnerable workers. Indeed, the purposes as stated in the amendments themselves were the establishment of wage levels "necessary for health and welfare," or sufficient for "decent maintenance" and "adequate living." In 1960 John F. Kennedy (1960) called for expanded minimum wage coverage: "To pass them by . . . shocks the conscience of those who care."

Congressional opponents of minimum wages were careful to state their support for the overall objective of helping the working poor and reducing poverty, while arguing that minimum wages were the wrong way to achieve these goals. From 1937 through 1966, the most frequently raised objections to amendments to the Fair Labor Standards Act in the *Congressional Record* concerned: 1) the direct cost to employers and 2) the indirect costs created by the administrative requirements of complying with the law. Some opponents argued that raising business's costs would reduce employment of less skilled workers, but this argument was not central to the public case against minimum wages until the late 1960s and early 1970s, when a growing number of academic studies found negative employment effects.

The emphasis in the minimum wage debates about insuring health and safety led Congress in 1946–47 to ask the US Bureau of Labor Statistics to develop its own consumption bundle when pricing the basic standard of living for working families. The resulting budget covered "the dollar amount required to maintain the family at a level of adequate living, according to prevailing standards of what is needed for health, efficiency, the nurture of children, and for participation in social and community activities at a level of living described as 'modest but adequate'" (Lamale and Stotz 1960, p. 1). After 1954, the budgets began

appearing in the legislative debates reported in the *Congressional Record* (1955, p. 7871; 1961, pp. 5987–89; 1974, p. 7313) as did the much lower standard of the official poverty level in the 1970s (1974, pp. 7312–13; 1977, p. 32715).

Table 3 shows the income from earning the national minimum wage for 40 hours per week over 52 weeks as a percentage of the adequate budget for a four-person family as calculated by the US Bureau of Labor Statistics in various cities and years. The minimum wage covered at most 55.7 percent of the adequate budget for a four-person family in any of the years. As in 1940, the minimum only came close to covering an adequate budget for a single individual under 65.

All of those involved in the process of setting a national minimum wage realized that the cost of living varied substantially across the country, and these differences evolved over time. In 1940, workers in Mobile, Alabama—the lowest-cost city— had 28.5 percent more purchasing power than workers in New York City, the highest-cost city. The distribution of purchasing power appears to have been reduced during the 1940s war decade to a 13.9 percent gap between the lowest-cost and highest-cost cities, possibly as a consequence of war-time rent controls (Fetter 2016). Then it rose to 21.3 percent in 1959 and 38.7 percent in 1967 (25.7 percent, leaving out Honolulu). Generally, the minimum wage covered more of the budget in southern and midwestern cities and less in northeastern cities in most years. The rankings, however, did not hold steady throughout. The positions of New York and Atlanta, for example, shifted substantially across the years in Table 3.

From the 1940s through the 1970s, widespread popular support for minimum wages was expressed in Gallup polls. Gallup pollsters defined the minimum wage and highlighted the group most likely to be affected in asking a standardized question (with updated numbers each time). One version of the standard question read: "At the present time the minimum wage that can be paid to workers in every state in most businesses and industries is 40 cents an hour. This means that all persons working in such businesses, in every state, including young people who have never worked before, cannot be paid less than 40 cents an hour. Would you approve or disapprove of raising this minimum to 60 cents an hour?" The polls showed that between 61 and 78 percent supported increases from 40 to 65 or 75 cents before the 1950 change, from 75 cents to $1 in 1953 or to $1.25 in 1954, and from $1 to $1.25 in 1957. The only poll without majority support called for a raise from 40 cents to $1 in 1948.[8]

Faced with popular and presidential support for minimum wages, opponents were only able to impose some short delays in raising the minimum. However, they were more successful in delaying the expansion of coverage of the minimum wage. For example, Senator Douglas (1972, pp. 377–78) reported disappointment

---

[8] Gallup poll results from Roper Center (2020). Gallup Poll ID numbers that start with USGALLUP: 081945. QK06B, 45-356. QT06B, 040746. RT08B, 061447. RK12B, 021448. RK05B, 48-422. QK11B, 48-422. QT11B, .011249. R08B, 433T. QT08B, 062949. R08, 442T. QT08, 49-446. QKT09A, 53-520. Q18, 54-540. QK13A, and 57-577. Q025.

Table 3

Full-Time Earnings on the National Minimum Wage as a Percentage of the BLS Adequate Cost of Living Budget: 4-Person Family by Year and Location

Area	1940	Area	1950	Area	1959	Area	1967
						South urban nonmetro	55.7
						Austin	55.6
						Baton Rouge	53.9
						Orlando	53.7
		New Orleans	45.2			Nashville	52.8
Mobile	53.2	Mobile	44.5			Houston	52.5
Kansas City, MO	49.8	Kansas City, MO	44.3			Durham	52.3
New Orleans	49.3	Savannah	43.9			Atlanta	52.0
Birmingham	48.7	Scranton	43.4			Dallas	51.9
Indianapolis	48.4	Indianapolis	43.3			Cincinnati	51.1
Memphis	48.2	Portland, ME	43.1			Green Bay	50.6
Denver	48.1	Cleveland	43.0	Houston	51.7	Bakersfield	50.4
Houston	48.1	St. Louis	42.9	Scranton	49.5	Dayton	50.2
Jacksonville	48.0	New York City	42.8	Atlanta	49.4	Baltimore	50.0
Buffalo	47.7	Manchester, NH	42.6	Baltimore	49.3	Lancaster	50.0
Portland, OR	47.5	Buffalo	42.5	Philadelphia	48.1	Midwest urban nonmetro	50.0
Los Angeles	47.4	Portland, OR	42.3	New York City	47.4	Pittsburgh	49.9
Baltimore	47.2	Philadelphia	42.2	Kansas City, MO	47.0	Detroit	49.6
Atlanta	47.1	Norfolk	42.0	Cincinnati	46.3	Philadelphia	49.4
Cincinnati	47.1	Minneapolis	42.0	Minneapolis	46.3	Denver	49.3
Norfolk	47.0	Birmingham	41.9	Portland, OR	46.1	Cleveland	49.2
Richmond	46.9	Cincinnati	41.8	Washington, DC	46.0	Portland, ME	48.9
Philadelphia	46.7	Denver	41.7	Detroit	46.0	Kansas City, MO	48.9
Portland, ME	46.0	Chicago	41.7	Pittsburgh	45.4	Wichita	48.7
Scranton	45.6	Detroit	41.6	St. Louis	45.4	St. Louis	48.5
Pittsburgh	45.4	Baltimore	41.3	Cleveland	45.1	San Diego	48.5
Seattle	45.4	Jacksonville, FL	41.3	Los Angeles	44.9	New York City	48.4
St. Louis	45.1	Pittsburgh	41.3	Boston	44.8	Minneapolis	48.1
Cleveland	44.9	Memphis	41.2	San Francisco	44.8	Buffalo	47.9
Minneapolis	44.3	Los Angeles	41.2	Seattle	42.7	West urban nonmetro	47.9
Detroit	43.7	Boston	41.0	Chicago	42.6	Chicago	47.7
Boston	43.5	Seattle	41.0			Milwaukee	47.7
San Francisco	43.1	San Francisco	41.0			Indianapolis	47.6
Chicago	42.9	Atlanta	40.7			Washington, DC	47.5
Washington, DC	41.9	Houston	40.3			Cedar Rapids	46.8
New York City	41.4	Richmond	40.1			Boston	46.6
		Washington, DC	39.7			Champaign-Urbana	46.5
		Milwaukee	39.7			Los Angeles	46.2
						Hartford	45.3
						Seattle	44.7
						San Francisco	44.3
						Honolulu	40.2

Summary Statistics

	1940		1950		1959		1967
Median	47.0		41.9		46.1		49.1
Maximum	53.2		45.2		51.7		55.7
Minimum	41.4		39.7		42.6		40.2
Max/Min Ratio	1.285		1.139		1.213		1.387

Percentage of Family Budget in Median City Covered by the National Minimum Wage

	1940		1950		1959		1967
Median 1-person	92.9		82.7		91.1		97.0
Median 2-person	71.6		63.8		70.2		74.8
Median 3-person	60.1		53.6		58.9		62.8

Source: The minimum wage comes from series Ba4422 in Carter et. al (2006, p. 2-284). Budget figures are from *Monthly Labor Review*(1940, p. 1041), US Bureau of Labor Statistics (1951, p. 2), Lamale and Stotz (1960, p. 4), and Bracket (1969, p. 8).
Note: The budgets included costs of goods, rents and services, payment of personal taxes, Social Security deductions and nominal allowances for occupational expenses and life insurance. (BLS 1959, p. 1).

that the compromises in 1949 and 1955 traded away expansions of coverage for a higher minimum. The 1961 amendment finally raised the coverage of private sector employees from 55 to 63 percent by including employees in large retail and service enterprises, local transit, construction, and gasoline service (as shown in Table 2). To overcome opposition, the minimum for newly covered workers was set below the minimum for previously covered workers until 1965. Coverage was expanded to 77 percent of private workers and 40 percent of government workers with the 1966 amendments that provided lower minimum rates for government employees in hospitals, nursing homes, and schools, and also for private sector employment in agriculture, laundries, dry cleaners, large hotels, motels, and restaurants (Willis 1972). The minimum rates were not equalized for all covered workers until 1978, when roughly 87 percent of private employees were covered.

For economists, the initial passage and subsequent amendment of a federal minimum wage renewed old debates. After the experience of the Great Depression, the discussions in the 1940s often assumed less than full employment and imperfections in commodity and input markets. In addition, the language used by economists when discussing these issues had become more precise over the previous several decades. Among economists at this time, the main arguments for the minimum wage were: 1) setting minimum wages in Joan Robinson's (1933) monopsonistic labor markets would move wages closer to the marginal revenue product, raising both wages and employment; 2) a minimum wage would improve workers' health and productivity enough that the consequent increase in labor demand could more than offset the negative employment effect of the higher wage; 3) higher minimum wages would induce inventions and improvements in managerial efficiency that would raise labor productivity and increase the demand for labor (Bronfenbrenner 1943); and 4) an economy-wide minimum wage could shift income from entrepreneurs with lower propensities to consume to low-income workers who would spend their full incomes, leading to demand increases for consumer goods and services that promoted more output and employment (Brown 1940).

Among prominent economists who took the other side were Martin Bronfenbrenner (1943) and George Stigler (1946), who argued against all four points. For example, Stigler (1946, p. 358) sought to shift the focus of the debate, claiming that everybody agreed on the goal of eliminating extreme poverty. The important questions were: 1) do minimum wages diminish poverty? and 2) are there better alternatives? Stigler said "no" to the first question because the likely result of a minimum wage would be the discharge of "workers whose services are worth less than the minimum wage." After conceding that employment, wages, and output could be increased when employers have wage-setting power with "a skillfully set minimum wage," he was skeptical that policymakers possessed a "tolerably accurate method" of determining the optimum wages over time for each occupation, firm, and quality of worker. Hours worked and many other factors made the connection between the hourly wage and family income "remote and fuzzy." Like the earlier marginalists of the 1910s, Stigler believed that alternative policies would be more effective at helping the poor and added the idea of what we would now call a negative income tax (similar in form to

the modern Earned Income Tax Credit) to the list of policy options for helping low-wage workers offered in the 1910s by Fetter, Taussig, Bates Clark, and others.

In response, Richard Lester (1946, 1947) castigated Stigler and other marginalist thinkers for an inadequate understanding of the operation of labor markets and how employers made decisions. His surveys of southern entrepreneurs showed that they focused primarily on demand changes when choosing employment, did not think in terms of marginal analysis, and did not adjust their capital-labor ratios in response to North-South wage differentials. They responded to the minimum wage shock by improving management practices and increasing sales efforts. Fewer than 10 percent of those responding to his surveys mentioned reducing output. His own studies of wage determination had not found the single wage predicted by marginalists. Instead, he found diversity in pay for equally productive workers.

At the start of the 1960s, professional opinion about economists about the lessons to be drawn from empirical work on the minimum wage was strongly divided. At various times between 1910 and 1951, state labor departments and the US Department of Labor collected data on earnings and employment just prior to and after the adoption of a minimum wage. Most studies by government departments reported weak or no effects of the minimum wage on employment. Peterson (1957) argued that many of these reports had failed to dig very deeply into the data. In reexamining the data in a more disaggregated fashion, he found support for the "hypothesis that employment changes will be inversely related to wage increase imposed by a minimum" (p. 430). In a similar fashion, the initial Department of Labor reports on the increase in the minimum wage to $1 showed small employment effects, while later academic studies found larger effects (Macesich and Stewart Jr. 1960; Douty 1960).

The differences came to a head in a debate between Lester and Peterson in a 1960 issue of the *Industrial and Labor Relations Review*. As they criticized each other nearly line by line, the debate centered on empirical issues familiar to modern empirical economists. Both discussed "causal inference" and focused heavily on issues related to ceteris paribus conditions. As in the modern debates, they used difference-in-difference comparisons and sometimes difference-in-difference-in-difference comparisons to compare outcomes pre- and post-minimum wage in treatment and control areas. Their debate over ceteris paribus conditions examined simultaneous events and prior trends. In a study of the 70-cent minimum wage, Peterson mimicked a regression discontinuity design by showing cross-tabulations before and after the minimum wage for firms in several wage categories ascending from well below the minimum to well above, prior to the increase. Peterson also compared covered and uncovered firms within the same town and the same industry. These studies faced the problem that they had to rely on cross-tabulated aggregates and there were not enough observations for them to control effectively for multiple confounds with regression analysis. Peterson, citing the rough nature of the data, declared victory when the results showed negative effects on hours or employment in more than half of the comparisons. Lester remained unconvinced.

In the aftermath of this debate, the neoclassical view of negative employment consequences gained the upper hand, when most of the empirical papers in the late

1960s, and especially into the 1970s, found negative employment effects, particularly for the less-skilled, teens, African Americans, and the disadvantaged. In two prominent examples from this more extensive literature, Benewitz and Weintraub (1964) found employment effects from the 1962 New York City increase in the minimum wage to $1.50, and Campbell and Campbell (1969) found higher unemployment in cities with state minimum wage rates. For summaries of this empirical minimum wage literature up to about 1980, see Brown, Gilroy, and Kohen (1982) and Neumark and Wascher (2008).

In the 1960s, Milton Friedman became the most prominent public face of neoclassical economic opposition to minimum wages. In an oft-quoted 1966 *Newsweek* opinion column, Friedman (1966) noted that Congress had just raised the minimum wage, which would "add to the ranks of the unemployed." But Friedman was far from alone in this view. Prominent Keynesians like James Tobin wrote: "People who lack the capacity to earn a decent living need to be helped, but they will not be helped by minimum-wage laws. . . . The more likely outcome of such regulations is that the intended beneficiaries are not employed at all" (*Congressional Record* 1966, p. 11270). By the late 1960s, the view that high minimum wages reduced employment of low skilled workers featured widely in the *Congressional Record* (1966, p. 11301; 1974, p. 5719; and 1977, p. 29186).

As the real minimum increased and gaps in coverage were closed, the minimum wage became binding for more workers. Friedman argued that this would promote discrimination, arguing, "I am convinced that the minimum-wage law is the most anti-Negro law on our statute books—in its effect not its intent." Studies such as Adie (1973), Mincer 1976, and Ragan (1977) found negative employment effects for African American workers. The prospect of minimum wages resulting in labor market discrimination was frequently mentioned in the *Congressional Record* (1974, pp. 5720–30 and 1977, pp. 29186–87, 29455, 29463, 297303, 32707). In contrast, most Black leaders favored minimum wage increases as a way of increasing wages and reducing inequality. They also strongly resisted separate and lower minimum rates for African American workers, which they regard as implying inferiority (Schulman 1991, p. 56).

Epilogue

The real value of federal minimum wages, as shown in Figure 1, never again reached the same real level as in 1968. Through 2020, there have been seven additional amendments to the original Fair Labor Standards Act that increased the minimum wage in 14 different years. Inflation has eroded the real value of these increases and since the 1980s the minimum wage has bounced around an average real value similar to that of the 1950s.

Between 1968 and 1990, the minimum wage received relatively little attention in the political arena. The term "minimum wage" appeared only once in a presidential State of the Union address (by Jimmy Carter in 1981). In 1973, President

Nixon vetoed a proposed increase in the minimum wage that did not include a youth subminimum rate, because he believed the new minimum would harm teenage employment (*Congressional Record* 1974, p. 4706). While supporting the 1974 amendment, Nixon cautioned against raising it too high (*Congressional Record* 1974, p. 4706). The stagflationary years of the 1970s, with slow productivity and high unemployment spikes, may have seemed like an unwise time to keep raising the value of the federal minimum wage. The first federal version of the Earned Income Tax Credit became law in 1975, offering an alternative policy tool for increasing the take-home pay of low-wage workers with children—and arguably with less concern over potential negative effects on employment.

In 1981, Ronald Reagan became the first president to actively oppose minimum wage increases. During his time in office, no minimum wage increases were enacted, but a sub-minimum wage was introduced. A broad swath of expert opinion agreed with him. In January 1987, the *New York Times* published an editorial titled "The Right Minimum Wage: $0.00," which argued for the Earned Income Tax Credit and other mechanisms to be used in its place.

It seems likely that a substantial part of the shift against the minimum wage was a result of the shifting consensus in economic research that negative employment effects of a minimum wage were a real concern. By the mid-1970s, minimum wage opponents were entering the negative findings from academic studies by Adie (1973), Mincer (1976), Gramlich, Flanagan, and Wachter (1976), Welch (1974), Ragan (1977), and Cotterill and Wadycki (1976) into the *Congressional Record*. The newer studies had access to more data (in particular, a longer time series since the implementation of minimum wages) and increasingly sophisticated econometric methodology.[9] A survey by Brown, Gilroy, and Kohen (1982) reported a widely cited consensus that "time series studies typically find that a 10 percent increase in the minimum wage reduces teenage employment by one to three percent," while cross-sectional studies produced smaller and less precise estimates of 0 to 0.75 percentage points. A 1992 survey of economists found that 57 percent agreed and 21 percent disagreed that "a minimum wage increases unemployment among young and unskilled workers" (reported in Whaples 1996). The survey also found (p. 729) that 87 percent of labor economists agreed that minimum wages increased unemployment for teens and the unskilled. Their median estimate of the impact of a 10-percent rise in the minimum was 2 percent, similar to the earlier consensus.

In the 1990s, the minimum wage debate was reignited when Card and Krueger (1994, 1995, 2000) published a series of studies using firm-level panel data techniques and found weak to zero employment effects of higher minimum wages. Neumark and Wascher (2000, 2008) challenged their findings with alternative

---

[9] Researchers increasingly adopted time series techniques that used the real minimum wage or the Kaitz Index (the ratio of the minimum wage to average hourly earnings, multiplied by the rate of coverage) as the measure of the minimum. Looking back at this literature, Kennan (1995) noted that most of the primary variation in the Kaitz index did not come from the wage minimum or coverage; therefore, using it to estimate the effect of minimum wages was akin to "looking for a needle in a haystack."

methods and data sources. Waves of research (discussed in the other papers in this symposium) have followed. More recent polls of economists show much less certainty about negative employment effects of a minimum wage. One 2015 poll asked leading academic economists whether increasing the minimum wage to $15 (from the current level of $7.25) by 2020 would substantially reduce employment of low-skilled workers (IGM Forum 2020): 26 percent agreed or strongly agreed, 24 percent disagreed, 38 percent were uncertain, and the remainder did not answer.

Meanwhile, higher minimum wages continue to have popular support. A 2013 Gallup poll indicated that about three-quarters of Americans supported a minimum wage increase from the prevailing rate of $7.25 per hour. In a 2019 NPR/PBS News-Hour/Marist Poll, 56 percent responded that they believed a national minimum wage of $15 per hour would be a good idea (Polling Report 2020). The states have responded. In 1989, 15 states had minimums above the national level. The number fell back to four states after the national minimum wage increases in 1990 and 1991 but then had risen to 32 in 2007 just before the last national amendments (Neumark 2019). At present, the Congressional Budget Office (2019) reported that 60 percent of US workers live in a state where the minimum exceeds the federal minimum of $7.25 per hour. By 2025, about 30 percent of workers will live in states with a minimum wage of $15 or higher. The United States has returned to an era of substantial minimum wage differences across states, and the future course of the federal minimum wage may be determined by the influence of the state and local minimums on labor market outcomes in these areas.

References

Adie, Douglas K. 1973. "Teen-Age Unemployment and Real Federal Minimum Wages." *Journal of Political Economy* 81 (2): 435–41.

American Institute of Public Opinion. 1939. "American Institute of Public Opinion-Surveys, 1938–1939." *The Public Opinion Quarterly* 3 (4): 581–607.

Barrington, Linda, and Gordon M. Fisher. 2006. "Poverty." In *Historical Statistics of the United States: Earliest Times to the Present,* Vol. 2., edited by Susan B. Carter, Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright. Cambridge, UK: Cambridge University Press.

Benewitz, Maurice, and Robert E. Weintraub. 1964. "Employment Effects of a Local Minimum Wage."

Industrial and Labor Relations Review 17 (2): 276–88.

Bowden, Witt. 1937. "Hours and Earnings Before and After the N.R.A." *Monthly Labor Review* 44 (1): 13–36.

Brackett, Jean C. 1969. "New BLS Budgets Provide Yardsticks for Measuring Family Living Costs." *Monthly Labor Review* 92 (4): 3–16.

Bronfenbrenner, Martin. 1943. "Minimum Wages, Unemployability, and Relief: A Theoretical Note." *Southern Economic Journal* 10 (1): 52–59.

Brown, Charles, Curtis Gilroy, and Andrew Kohen. 1982. "The Effect of the Minimum Wage on Employment and Unemployment." *Journal of Economic Literature* 20 (2): 487–528.

Brown, Weir M. 1940. "Some Effects of a Minimum Wage upon the Economy as a Whole." *American Economic Review* 30 (1): 98–107.

Campbell, Colin D., and Rosemary G. Campbell. 1969. "State Minimum Wage Laws as a Cause of Unemployment." *Southern Economic Journal* 35 (4): 323–32.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.

Card, David, and Alan B. Krueger. 1995. *Myth and Measurement: The New Economics of the Minimum Wage.* Princeton: Princeton University Press.

Card, David and Alan B. Krueger. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply." *American Economic Review* 90 (5): 1397–1420.

Caro, Robert. 2002. *Master of the Senate: The Years of Lyndon Johnson.* New York: Alfred A. Knopf.

Carter Susan B., Scott Sigmund Gartner, Michael R. Haines, Alan L. Olmstead, Richard Sutch, and Gavin Wright, Eds. 2006. *Historical Statistics of the United States: Earliest Times to the Present,* Vol. 2. Cambridge, UK: Cambridge University Press.

Clark, John Bates. 1913. "The Minimum Wage." *Atlantic Monthly* 112 (9): 289–97.

Clark, Lindley D. 1921. *Minimum-Wage Laws of the United States: Construction and Operation.* Washington, D.C.: US Government Printing Office.

Columbia Law Review. 1948. "Portal to Portal Act. Good Faith Provisions," 48 (3): 443–50.

Commons, John R. 1935. *History of Labour in the United States.* New York: The Macmillan Company.

Complete State of the Union Addresses, from 1790 to 2006. 1790–2006. http://www.gutenberg.org/files/5050/5050-h/5050-h.htm (accessed October 1, 2020).

Congressional Budget Office. 2019. *The Effects on Employment and Family Income of Increasing the Federal Minimum Wage.* Washington, D.C.: Congressional Budget Office.

Congressional Record. Various years. Washington, D. C.: United States Government Publishing Office.

Cotterill, Philip G., and Walter J. Wadycki. 1976. "Teenagers and the Minimum Wage in Retail Trade." *Journal of Human Resources* 11 (1): 69–85.

Cushman, Barry. 1998. *Rethinking the New Deal Court: The Structure of a Constitutional Revolution.* Oxford: Oxford University Press.

Detroit Free Press. 1933. "Johnson Gives New Warning," August 12.

Douglas, Paul H. 1972. *In the Fullness of Time: The Memoirs of Paul H. Douglas.* New York: Harcourt Brace Jovanovich.

Douty, H. M. 1960. "Some Effects of the $1.00 Minimum Wage in the United States." *Economica* 27 (106): 137–47.

Eisenhower, Dwight D. 1955. "State of the Union Address.*" Speech, January 6, 1955. https://www.ontheissues.org/Archive/SOTU_Ike_Dwight_Eisenhower.htm (accessed November 4, 2020).

Federal Reserve Bank of St. Louis. 2019. FRED website. https://fred.stlouisfed.org/ (accessed October 1, 2020).

Fetter, Daniel K. 2016. "The Home Front: Rent Control and the Rapid Wartime Increase in Home Ownership." *Journal of Economic History* 76 (4): 1001–43.

Fetter, Frank. 1917. "A Doubtful Remedy: Statement of Frank Fetter Published in the Fourth Report of the New York Factor Commission in 1915." In *Selected Articles on Minimum Wage*, edited by Mary Katherine Reely, 149–50. New York: H.W. Wilson Company.

Fishback, Price V. 1998. "Operations of 'Unfettered' Labor Markets: Exit and Voice in American Labor Markets at the Turn of the Century." *Journal of Economic Literature* 36 (2): 722–65.

Fishback, Price V. 2018. "Rule of Law in Labor Relations, 1898–1940." Unpublished.

Frankfurter, Felix. 1955. "Mr. Justice John Roberts." *University of Pennsylvania Law Review* 104 (3): 311–17.

Friedman, Milton. 1966. "Minimum Wage Rates and Unemployment." *Newsweek*, September 11. https://miltonfriedman.hoover.org/objects/57909/minimumwage-rates.

Gramlich, Edward M., Robert J. Flanagan and Michael L. Wachter. 1976. "Impact of Minimum Wages on Other Wages, Employment, and Family Incomes." *Brookings Papers on Economic Activity* 7 (2): 409–61.

Grossman, Jonathan. 1978. "Fair Labor Standards Act of 1938: Maximum Struggle for a Minimum Wage." *Monthly Labor Review* 101 (6): 22–30.

Hammond, Matthew B. 1915. "Wages Boards in Australia: IV. Social and Economic Results of Wages Boards." *Quarterly Journal of Economics* 29 (3): 563–630.

IGM Forum. 2020. "$15 Minimum Wage." University of Chicago. http://www.igmchicago.org/surveys/15-minimum-wage [accessed 1/2/2020].

Kennan, John. 1995. "The Elusive Effects of Minimum Wages." *Journal of Economic Literature* 33 (4): 1950–65.

Kennedy, John F. 1960. "Remarks of Senator John F. Kennedy in the Senate, Washington, D. C. August 10, 1960." Speech, John F. Kennedy Presidential Library and Museum, August 10, 1960. https://www.jfklibrary.org/archives/other-resources/john-f-kennedy-speeches/united-states-senate-labor-standards-amendments-19600810 (accessed October 1, 2020).

Lamale, Helen H., and Margaret S. Stotz. 1960. "The Interim City Worker's Family Budget." *Monthly Labor Review* 83 (8): 785–808.

Leonard, Thomas C. 2005. "Retrospectives: Eugenics and Economics in the Progressive Era." *Journal of Economic Perspectives* 19 (4): 207–24.

Lester, Richard A. 1946. "Shortcomings of Marginal Analysis for Wage-Employment Problems." *American Economic Review* 36 (1): 63–82.

Lester, Richard A. 1947. "Marginalism, Minimum Wages, and Labor Markets." *American Economic Review* 37 (1): 135–48.

Lester, Richard A. 1960. "Employment Effects of Minimum Wages." *Industrial and Labor Relations Review* 13 (2): 254–64.

Macesich, George, and Charles T. Stewart, Jr. 1960. "Recent Department of Labor Studies of Minimum Wage Effects." *Southern Economic Journal* 26 (4): 281–90.

Marchingiglio, Riccardo, and Michael Poyker. 2020. "The Economics of Gender-Specific Minimum Wage Legislation." Unpublished.

Meeker, Royal. 1933. "The Great American Adventure." *The American Scholar* 2 (4): 463–70.

Mincer, Jacob. 1976. "Unemployment Effects of Minimum Wages." *Journal of Political Economy* 84 (4): 87–104.

Monthly Labor Review. 1940. "Costs and Standards of Living." *Monthly Labor Review,* 51 (4): 1008–23.

Neumark, David. 2019, "The Econometrics and Economics of the Employment Effects of Minimum Wages: Getting from Known Unknowns to Known Knowns." *German Economic Review* 20 (3): 293–329.

Neumark, David, and William Wascher. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Comment." *American Economic Review* 90 (5): 1362–96.

Neumark, David, and William L. Wascher. 2008. *Minimum Wages.* Cambridge, MA: MIT Press.

Neumann, Todd C., Jason E. Taylor, and Fishback, Price V. 2013. "Comparisons of Weekly Hours Over the Past Century and the Importance of Work Sharing Policies in the 1930s." *American Economic Review* 103 (3): 105–10.

New York Times. 1955. "Eisenhower Signs $1 Minimum Pay," August 13.

New York Times. 1987. "The Right Minimum Wage: $0.00." January 14. https://www.nytimes.com/1987/01/14/opinion/the-right-minimum-wage-0.00.html.

Peterson, John M. 1957. "Employment Effects of Minimum Wages, 1938–50." *Journal of Political Economy* 65 (5): 412–30.

Peterson, John M. 1959. "Employment Effects of State Minimum Wages for Women: Three Historical Cases Re-Examined." *Industrial and Labor Relations Review* 12 (3): 406–22.

Peterson, John M. 1960. "Employment Effects of Minimum Wages: Reply." *Industrial and Labor Relations Review* 13 (2): 264-73.

Phelps, Orme W. 1939. *The Legislative Background of the Fair Labor Standards Act.* Chicago: University of Chicago Press.

Pigou, Arthur Cecil. 1913. "A Minimum Wage for Agriculture." *The Nineteenth Century,* (74): 1167–84.

Polling Report. 2020. Polling Report Inc. https://www.pollingreport.com/work.htm (accessed October 1, 2020).

Prasch, Robert E. 2000. "John Bates Clark's Defense of Mandatory Arbitration and Minimum Wage Legislation." *Journal of the History of Economic Thought* 22 (2): 251–63.

Prasch, Robert E. 2007. "Professor Lester and the Neoclassicals: The 'Marginalist Controversy' and the Postwar Academic Debate Over Minimum Wage Legislation: 1945–1950." *Journal of Economic Issues* 41 (3): 809–25.

Ragan, James F., Jr. 1977. "Minimum Wages and the Youth Labor Market." *Review of Economics and Statistics* 59 (2): 129–36.

Robinson, Joan. 1933. *The Economics of Imperfect Competition*. London: Macmillan. EQ: year

Roper Center. 2020. Results of Search for Minimum Wage Polling Between 1935 and 1970. https://ropercenter.cornell.edu/CFIDE/cf/action/ipoll/ipollResult.cfm?keyword=minimum+wage&exclude=&topic=Any&organization=Any&fromDate=1935&toDate=1970&questionViewId=&label=&studyId=&sortBy=BEG_DATE_DESC&search=submit (accessed July, 1 2020).

Rose, Jonathan D. 2010. "Hoover's Truce: Wage Rigidity in the Onset of the Great Depression." *Journal of Economic History* 70 (4): 843–70.

Schulman, Bruce J. 1991. *From Cotton Belt to Sunbelt: Federal Policy, Economic, Development, and the Transformation of the South, 1938–1980*. Oxford: Oxford University Press.

Seltzer, Andrew. 1994. "Causes and Consequences of American Minimum Wage Legislation, 1911–1947." PhD diss. University of Illinois.

Seltzer, Andrew. 1995. "The Political Economy of the Fair Labor Standards Act of 1938." *Journal of Political Economy* 103 (6): 1302–44.

Seltzer, Andrew J. 1997. "The Effects of the Fair Labor Standards Act of 1938 on the Southern Seamless Hosiery and Lumber Industries." *Journal of Economic History* 57 (2): 396–415.

Seltzer, Andrew. 2002. "Monopsony and Minimum Wages: Evidence from the Tobacco Leaf-Processing Industry." *Essays in Economic and Business History* 20 (1): 167–82.

Stecker, Margaret Loomis. 1937. *Intercity Differences in Costs of Living in March, 1935, 59 Cities*. Washington, D.C.: US Government Printing Office.

Stigler, George J. 1946. "The Economics of Minimum Wage Legislation." *American Economic Review* 36 (3): 358–65.

Stitt, Louise, and Florence P. Smith. 1938. "Minimum Wages and Maximum Hours." *Monthly Labor Review* 46 (1): 191–204.

Taussig, Frank. 1916. "Minimum Wages for Women." *Quarterly Journal of Economics* 30 (3): 411–42.

Taylor, Jason E. 2011. "Work-Sharing During the Great Depression: Did the 'President's Reemployment Agreement' Promote Reemployment?" *Economica* 78 (309): 133–58.

Taylor, Jason E. 2019. *Deconstructing the Monolith: The Microeconomics of the National Industrial Recovery Act*. Chicago: University of Chicago Press.

US Bureau of Labor Statistics. 1933a. "Minimum Wage." *Monthly Labor Review* 36 (6): 1259–75.

US Bureau of Labor Statistics. 1933b. "Minimum Wage Legislation in the United States." *Monthly Labor Review* 37 (6): 1344–54.

US Bureau of Labor Statistics. 1951. "Family Budget of City Worker, October 1950." *US Bureau of Labor Statistics Bulletin No. 1021*. Washington, D.C.: Government Printing Office.

US Department of Labor. 2020. *History of Federal Minimum Wage Rates Under the Fair Labor Standards Act, 1938–2009*. Washington, D.C.: US Department of Labor.

US Supreme Court 1754–2020. *United States Reports*. Washington, D. C.: United States Government Publishing Office.

Webb, Sidney. 1912. "The Economic Theory of a Legal Minimum Wage." *Journal of Political Economy* 20 (10): 973–98.

Welch, Finis. 1974. "Minimum Wage Legislation in the United States." *Economic Inquiry* 12 (3): 285–318.

Whaples, Robert. 1996. "Is There Consensus among American Labor Economists? Survey Results on Forty Propositions." *Journal of Labor Research* 17 (4): 725–34.

Willis, Robert N. 1972. "The Evolution of The Fair Labor Standards Act." *University of Miami Law Review* 26 (3): 607–34.

Women's Bureau. US Department of Labor. 1928. *The Development of Minimum-Wage Laws in the United States, 1912 to 1927*. Washington, D.C.: US Government Printing Office.

Women's Bureau. US Department of Labor. 1939. *State Minimum-Wage Laws and Orders: An Analysis*: *Women's Bureau Bulletin No. 167*. Washington, D.C.: US Government Printing Office.

Women's Bureau. US Department of Labor. 1942. *State Minimum-Wage Laws and Orders: 1942: Bulletin of the Women's Bureau, 191*. Washington, D.C.: US Government Printing Office.

Women's Bureau. US Department of Labor. 1951. *State Minimum-Wage Laws and Orders: July 1, 1942 to July 1, 1950, Revised Supplement to Bulletin 191. Bulletin of the Women's Bureau, 227*, Revised. Washington, D.C.: US Government Printing Office.

Women's Bureau. US Department of Labor. 1953. *State Minimum Wage Laws and Orders: July 1, 1942– March 1, 1953: Women's Bureau Bulletin, No. 247*. Washington, D.C.: US Government Printing Office.

Women's Bureau. U. S. Department of Labor. 1958. *State Minimum Wage Laws and Orders: July 1, 1942 – July 1, 1958. Part II. Analysis of Rates and Coverage. Bulletin of the Women's Bureau, 267*, Washington, D.C.: US Government Printing Office.

Wright, Gavin. 1987. "The Economic Revolution in the American South." *Journal of Economic Perspectives* 1 (1): 161–78.

Estimating Judicial Ideology

Adam Bonica and Maya Sen

Justices on the US Supreme Court and other judges throughout the American judiciary inescapably make policy with their decisions, whether they want to or not. However, the traditionally dominant view of judging—historically within the legal academy but also among members of the public—is that judges merely apply the law, together with precedent from earlier court decisions, to a set of facts. As an example of this sentiment, US Supreme Court Chief Justice John Roberts famously commented that "[w]e do not have Obama judges or Trump judges, Bush judges or Clinton judges. What we have is an extraordinary group of dedicated judges doing their level best to do equal right to those appearing before them" (Sherman 2018). Judges, in his telling, are simply calling "balls and strikes."

As far back as the 1940s, scholarship began to challenge the assumptions behind this view. Today, the dominant view among social scientists is that ideology is indeed a key component predicting judicial rulings and judicial behavior. A judge's ideology shapes the law and, by extension, has significant social and economic consequences for individual litigants and society. Judicial ideology is, therefore, a topic of study and an important factor for understanding the economic and societal impact of the US legal system.

It was not until the last few decades that researchers tackled the question of how best to measure judicial ideology beyond simply using the party of the appointing

■ *Adam Bonica is Associate Professor of Political Science, Stanford University, Stanford, California. Maya Sen is Professor of Public Policy, John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts. Their email addresses are bonica@stanford. edu and maya_sen@hks.harvard.edu.*

president as a proxy. By far the largest area of this scholarship concerns the US Supreme Court. This is no surprise: the US Supreme Court is the most important court in the country and the final stopping point for many politically sensitive issues. Also, from a research standpoint, the Supreme Court lends itself relatively well to ideological measurement. First, unlike most other courts in the United States, all nine members of the Court hear and vote on cases together. Second, a small and tractable docket makes it possible to subjectively hand-code cases in order to estimate judicial ideology. However, researchers must also be cautious about extrapolating broadly about judicial behavior from the behavior of Supreme Court justices; not only does the very small sample of justices increase uncertainty, but it places considerable importance on nine idiosyncratic individuals who are relatively unconstrained in their position atop the American judicial hierarchy (Bailey 2017).

From the perspective of social scientists, the lower federal courts are perhaps a more important subject of study. Whereas the Supreme Court might hear 70 or 80 cases per year, the lower courts hear hundreds of thousands. In 2018, for example, 49,363 cases were filed in the US Courts of Appeals, one level below the US Supreme Court, and 358,563 were filed in the US District Courts, one level below that (according to the Federal Judicial Caseload Statistics website at https://www. uscourts.gov/statistics-reports/federal-judicial-caseload-statistics-2018). In addition, the lower federal courts rule not only on topics of constitutional importance but also on "bread and butter" economic issues and criminal justice questions—topics of significant interests to social scientists and policymakers. Despite their importance, the larger number of judges and the manner in which they hear cases make it more challenging to estimate ideology for judges at these levels. While the US Supreme Court has nine justices hearing cases as a single voting body, the lower courts include 663 federal district court judgeships and 179 courts of appeals judgeships spread across twelve geographic circuits and a federal circuit (according to the US Courts website at https://www.uscourts.gov/sites/default/files/allauth.pdf).

Despite these challenges, most contemporary researchers agree that ideology— usually measured via partisanship—is among the most important factors shaping judicial decision making. In this paper, we provide an overview of how scholars think about and measure judicial ideology. We begin by discussing the various measures of judicial ideology that have been estimated from a range of sources: classifying court rulings and judicial votes as conservative or liberal; newspaper editorials about US Supreme Court justices before they are confirmed; agreement and disagreement across the voting records of US justices; the party and ideological scores of the elected politicians who play a central role in appointing federal and state judges; campaign contributions made to and by judicial candidates; evidence from the random assignment of cases to lower federal court judges; and estimates generated by automated text analysis of judicial opinions. We also offer some brief comments on research pertaining to judicial ideology in other countries and on international courts. We also discuss some limitations of this literature: for example, the challenges inherent in comparing measures of judicial ideology across time and across the judicial hierarchy.

We then illustrate an important implication of judicial ideology: ideological polarization among judges. As we show, ideological polarization within the federal courts has risen in the past few decades. Along with this polarization have come increasingly fractious opinions, indicating growing discord and conflict within the courts. We conclude by discussing how these changes require a more complex and realistic perception of what is involved in the exercise of judging.

Measurement Methods

Classifying Decisions as Conservative or Liberal

A starting point for estimating judicial ideology has been to classify judicial rulings themselves as "conservative" or "liberal," looking closely at the votes cast and the reasoning contained therein. The best-known effort to classify judicial rulings—and votes—is available for the US Supreme Court via the Supreme Court Database at Washington University in St. Louis. It is publicly available at http://scdb.wustl.edu/index.php. As the website notes, "The Database contains over two hundred pieces of information about each case decided by the Court between the 1791 and 2018 terms. Examples include the identity of the court whose decision the Supreme Court reviewed, the parties to the suit, the legal provisions considered in the case, and the votes of the Justices." The Supreme Court Database has become a standard research tool in the study of the US Supreme Court and in the ideological valance of Supreme Court rulings.

The database relies on expert coding of all cases ruled on by the Supreme Court since 1946. Cases (and accordingly the votes of individual justices) as coded as being conservative, liberal, or unspecifiable. As an example, the liberal position on criminal cases would be the one generally favoring the criminal defendant; in civil rights cases, the liberal position would be the one favoring the rights of minorities or women, while in due process cases, it would be the anti-government side. For economic activity cases—which make up a perhaps surprisingly large share of the Supreme Court's docket—the liberal position will be the pro-union, anti-business, or pro-consumer stance. For cases involving the exercise of judicial power or issues of federalism, the liberal position would be the one aligned with the exercise of federal power, although this may depend on the specific issues involved. Finally, some decisions are categorized as "indeterminate," such as a boundary dispute between states.

Some papers have raised concerns about the nature of the Database's subjective coding and its reliability, including the possibility of miscoded ideological directionality (Harvey and Woodruff 2013). However, any classification system—human or machine coded—will have errors. As Bailey (2017) discusses, the percent of decisions each justice votes in a conservative direction using the classifications of the Supreme Court Database actually correlates quite well with other, more sophisticated ideological measures. Indeed, other key measures of the justices' ideology—including ones we discuss below—rely on the conservative/liberal distinctions in the Supreme Court database.

The kind of coding found in the Supreme Court Database is only currently available with this degree of transparency and with attendant covariates for the US Supreme Court. Categorizing the decisions of federal appeals and federal district courts in this way, or categorizing the decisions of state-level judges in this way, is a significant task. One extant and comprehensive data set is the Federal Judicial Center's Integrated Database (IDB) (https://www.fjc.gov/research/idb), which collects data from the Administrative Office of the Courts. The IDB "contains data on civil case and criminal defendant filings and terminations in the [federal] district courts, along with bankruptcy court and appellate court case information," but these data are purposefully de-identified in terms of the identity of the federal judges or judge ruling on the case. The IDB has, however, recently been merged with the Courtlistener database of judicial opinions (https://www.courtlistener.com), which facilitates extraction of judge-identifying information (as described by Lissner 2019).

Another data set for federal appeals court cases is the United States Courts of Appeals Databases, maintained by a team of researchers once led by the late Donald Songer (http://www.songerproject.org). These data also include the coding of various court of appeals case attributes as well as subjective coding on the ideological valance of case outcomes. However, this dataset extends from 1925 to only 2002 and contains only a small subset (approximately 2.6 percent) of published (versus unpublished) cases from the federal courts of appeals; in addition, other scholars have reported finding errors in this database (for example, Epstein, Landes and Posner 2013).

Pre-Confirmation Newspaper Editorials

We now turn to existing measures of ideology at the judge-level—as opposed to case-level. One approach relies on information available in the pre-confirmation process to predict the post-confirmation voting patterns for judges. For example, Segal and Cover (1989) code editorials from four leading conservative and liberal newspapers about US Supreme Court nominees and whether they describe the future justices as liberal, moderate, conservative, or "not applicable." For example, liberal statements would include those describing the candidate's "support for the rights of defendants in criminal cases, women and racial minorities in equality cases, and the individual against the government in privacy and First Amendment cases" (p. 559). The codings were then combined across different coders and newspapers in a systematic fashion. The Segal-Cover scores are periodically updated and accessible through the Washington University Supreme Court Database.

A strength of this approach is that it uses external (non-case specific) information to estimate judicial ideology. This makes it straightforward to use these scores to predict or better understand voting later on by the justices. Segal and Cover (1989) demonstrate the validity of these scores by comparing them to justices' votes on civil rights and civil rights cases, using the classifications from the Supreme Court Database. For example, they find a high correlation (0.80) between the coding scheme and the justices' voting on civil liberty cases (p. 561).

A limitation of this approach is that the scores are static: once a candidate is confirmed, such estimates can no longer be updated and therefore fail to capture

key Supreme Court dynamics. For example, these scores will not include "ideo-logical drift" among the Justices, which refers to the empirical observation that Justices occasionally move away from their original ideology, usually gravitating in a liberal direction during their tenures (Owens and Wedeking 2012). A well-known example of ideological drift is Justice Harry Blackmun who was appointed by Republican Richard Nixon but who shifted in an ideologically leftward direc-tion in the years following his seminal opinion in *Roe v. Wade* (410 US 113 [1973]).

Another limitation is that Segal-Cover scores are only available for judges for whom a wealth of pre-confirmation information is available. This tends to hold only for Supreme Court nominees, who garner significant confirmation coverage and are the subject of extensive journalistic editorializing. Such information is rarely available for lower federal court judges (appeals court or district judges) and state court judges. Lastly, Segal-Cover scores require a subjective assessment of subjective information (in the form of newspaper editorials), as opposed to being based on observed behavior of judges.

Voting Patterns

While Segal-Cover scores leverage pre-nomination information to measure ideology for Supreme Court Justices, other methodologies use the observed behavior of judges (or justices of the US Supreme Court) while on the bench. The most commonly used measures are those of Martin and Quinn (2002), which leverage overlapping tenures of US Supreme Court justices to estimate dynamic ideal points based on their voting behavior. These measures draw from item response theory (IRT), commonly used to estimate some latent variable from observed responses: for example, the technique is frequently used in the standardized testing litera-ture to estimate an unobserved quantity of interest, such as the test-taker's latent knowledge or ability. Underpinning the Martin and Quinn methodology is a unidi-mensional spatial model of voting, which assumes that justices maximize their utility by voting for the outcome nearest their ideal point, allowing for a degree of error. The model is similar to Poole and Rosenthal's NOMINATE scores, which estimate ideal points for members of Congress based on roll call votes (Poole and Rosenthal 1997; Poole 2005). At a high level, vote-based models of ideology essentially posi-tion individuals along a liberal-conservative dimension such that those who often vote together are placed near one another, while those who are less likely to vote together are further apart.

The Martin and Quinn (2002) approach relies on a dynamic Bayesian item response model. To model the dynamic component, they assume that ideal points follow a random walk process. Scores are estimated using Markov Chain Monte Carlo simulations. The eventual result, updated in each Supreme Court term, is a trendline of scores for each Supreme Court justice over time, as we will present later in the paper when discussing polarization that has emerged over time.

Martin-Quinn scores are widely used in the Supreme Court literature and the approach has some notable advantages. First, within a single term, it can estimate relative judicial ideology and generate useful estimates of uncertainty—similar to

what has been done in the past with Congressional ideology (Clinton, Jackman, and Rivers 2004). Second, Martin-Quinn scores do not require subjective coding of cases as conservative or liberal (Fischman and Law 2009). Third and related, Martin-Quinn scores rely on actual observed behavior (votes), rather than inferring ideology from third-person writings (such as editorials) or other kinds of subjective evaluations (such as case coding).

Martin-Quinn scores have some drawbacks as well. First and most important, estimation is only feasible for courts where judges decide cases together as a voting body. This largely restricts their usefulness beyond the US Supreme Court and state supreme courts.

Second, Martin-Quinn scores are estimated on the basis of justices' votes, meaning that using the scores to predict or analyze voting patterns will inevitably raise theoretical and empirical concerns about endogeneity. That is, trying to use Martin-Quinn scores as an independent variable to predict voting in the same term would mean that the same underlying vote data are being used in both the independent and dependent variables. The usual approach in dealing with this is to lag the Martin-Quinn scores by at least one term and then use them to predict or understand justices' votes in the following term.

Third, because Martin-Quinn scores are estimated every term (with a random walk prior), any estimates within a particular year incorporate the idiosyncrasies of that term's docket. Indeed, comparing the absolute Martin Quinn scores across different years is unwise, because of changing dockets. Indeed, comparing them across years implicitly relies on the assumption that "the distribution of case characteristics is constant over time" (Ho and Quinn 2010), which, given the discretionary nature of the Court's docket, is usually an unreasonably strong assumption. (This has led to the critical observation that Martin-Quinn scores are basically cardinal, not ordinal estimates.) A related point is that the scores themselves reflect the idiosyncrasies of that year's docket. To see this, suppose that an otherwise liberal leaning justice feels strongly that burning the US flag is not protected speech (as the otherwise reliably liberal Justice John Paul Stevens, an army veteran, did). If the Court in any one year decides to take on several cases involving flag-burning, then this justice's ideology would be estimated as being more conservative. But this would be reflective of a largely idiosyncratic docket as opposed to a meaningful shift in ideology.

Fourth, the Martin-Quinn scores are one-dimensional. Lauderdale and Clark (2014) develop a method to recover issue-specific preferences for Supreme Court justices across substantive legal issues. They find that allowing for multiple dimensions better predicts judicial voting behavior and that the identity of the median justice can vary across issues.

A final possible concern with the use of Martin-Quinn scores is that, because they are estimated on the basis of voting by US Supreme Court Justices only, they are not comparable with ideological estimates of other political actors—like those for members of Congress. To address this, Bailey (2007) uses instances of Supreme Court review of Congressional statutes and other cases on which there is presidential and Congressional input (for example, cases in which the executive has filed

an amicus brief with the Supreme Court) to bridge ideology across institutions and time. The results are ideological measures across key American political entities that are estimated on the same scale and, thus, easily comparable.

Appointment-based Measures

The seminal research of Segal and Spaeth (2002) showed that the party of a judge's appointing president is a powerful predictor of Supreme Court decision-making across a variety of subject matters. Perhaps the simplest way to "estimate" judicial ideology for all federal judges—Supreme Court Justices, federal appeals judges, and federal district judges—is to assign them either the partisan affiliation or the ideology of the US president who appointed them. The simplest operationalization is to compare Republican-appointed judges to Democratic-appointed judges. Several studies employ this strategy to make compelling arguments that Republican-appointed judges tend to vote in a more conservative direction than Democratic-appointed judges or that the partisan composition of three-judge panels predicts the ideological direction of rulings (for example, Sunstein et al. 2006).

Of course, the party of the appointing president is, at best, an inexact proxy of judicial ideology. Judges appointed by presidents of the same party can differ significantly in terms of their jurisprudence and policy preferences; for example, judges appointed by Donald Trump are, according to most reports and related measures, far more conservative than those appointed by Gerald Ford, a more moderate Republican. Even looking at judges appointed by the same president—for example, through the use of an indicator for the appointing president's identity— masks substantial variation. To take one example, George H.W. Bush made two US Supreme Court appointments: the first, Clarence Thomas, has been among the Court's most conservative members, while the second, David Souter, finished his career mostly voting with the Court's liberal members. In addition, focusing exclusively on the party of appointing president ignores the US Senate's role, which, per Article III of the U.S. Constitution, must confirm all federal judicial nominees. Suppose the Senate is controlled by the opposing party (or a more moderate or more extreme subset of the president's own party). In that case, the president is effectively constrained and often ends up nominating a candidate whose ideological preferences may not closely match his own. For example, in 2004, George W. Bush named Harriet Meiers, his own White House counsel, to replace Sandra Day O'Connor on the US Supreme Court. Senate Republicans, however, took an unusual position in opposing the nomination largely because Meiers was viewed as someone who might be intolerably moderate on the important conservative issue of abortion. Bush withdrew her nomination and instead nominated Samuel A. Alito, a reliable conservative more palatable to Senate Republicans.

More pragmatically, lower federal court appointments have been guided by the longstanding custom by which US Senators are closely consulted on nominees to federal courts located in their geographic area. Relying on this practice—known as "Senatorial Courtesy"—Giles, Hettinger, and Peppers (2001) assign to federal court of appeals judges either the estimated ideology of the appointing president (not

just party) or, in instances where the president or one of the senators from the state where the vacancy is located are of the same party, the ideology of the senator. In cases where both senators are of the same party as the appointing president, then the methodology assigns the average of the two senators' ideological scores.

The ideology of senators and presidents is estimated using the well-known DW-NOMINATE scores, a dynamic implementation of NOMINATE that permits legislator scores to change over their career (Poole and Rosenthal 1997; Poole 2005). The one-dimensional implementation of DW-NOMINATE assumes legislators decide between yea and nay outcomes on roll call votes as a function of their "spatial utility," which is determined by the distance between a legislator's ideal point and the location of the outcomes, allowing for a random, normally distributed error component. The legislators' "ideal points" and the yea and nay outcome components for bills are estimated simultaneously from roll call votes. This measure is conceptually similar to the Martin-Quinn scores in that it is based on vote patterns.

In further work, Epstein et al. (2007) rely on the intuition from Giles, Hettinger, and Peppers (2001) for their Judicial Common Space (JCS) scores, which map Martin-Quinn scores for the Supreme Court and scores for appeals court judges onto the same scale. They validate the approach by showing JCS scores predict Martin-Quinn scores for appeals judges who are later named to the Supreme Court. A separate dataset—Boyd (2015)—uses the same approach to generate scores for federal district court judges. Both datasets are used extensively by scholars of judicial politics.

Judicial Common Space scores and related appointment-based measures are not without drawbacks of their own. First, judges appointed within the same jurisdiction within the same rough two-year time frame by the same president will be assigned the identical score. This means that the scores are, by construction, measured with some error. Second, some courts—including the politically powerful US Court of Appeals for the District of Columbia—have no "home-state" senator. In these instances, the JCS scores simply assign judges the ideological score of the appointing President. For example, then-Judge Brett Kavanaugh, appointed to the DC Circuit Court of Appeals by George W. Bush in 2006, was assigned Bush's ideological score. It is also unclear whether the presumption of "Senatorial Courtesy" is applicable with the same force across judicial appointments (Nixon 2004), and norms have moved away from the practice in recent years. Lastly, JCS scores are available only for federal judges. We discuss one approach to estimating the ideology of state court judges that is similar—that of Brace, Langer, and Hall (2000)—below.

Estimating Judicial Ideology Using Campaign Contributions

Yet another strategy for measuring judicial ideology is to use political contributions made by judges. For federal judges, the inputs are contributions they made to political candidates or other political entities (such as political action committees or PACs) before being confirmed to a federal court. Federal judges may not make political contributions once on the bench; thus for them, this estimation method is based entirely on pre-confirmation observed behavior. For state judges, restrictions on their political activity may vary.

Bonica (2014) uses a campaign finance-based methodology to estimate the ideology of politicians, PACs, and individual donors. The ideology scores for donors and politicians (and the underlying contribution data) are publicly available as part of the Database on Ideology, Money in Politics, and Elections (DIME) (Bonica 2016). The logic behind contribution-based measures is that campaign contributions provide a costly and therefore informative signal about a donor's ideology. Similar to vote-based measures of ideology, the DIME model assumes a spatial utility model (allowing for an error component) and jointly estimates scores for donors and candidates from a contingency matrix of donation amounts. On an intuitive level, the DIME methodology assumes donors tend to prefer candidates with whom they are ideologically aligned. Thus, someone who is more conservative will be more likely to make political donations to conservative candidates, while the opposite holds for someone who is more liberal.[1] When merged with survey data, the scores are reliable predictors of individual-level policy preferences on a wide range of policy issues (like taxes, abortion, gun control, health care, and others) (Bonica 2019).

The DIME scores provide a means to measure judicial ideology based on judges' revealed preferences. Bonica and Sen (2017a) apply this measurement approach to federal judges and validate the scores by comparing across ideological measurements. They also show how the scores can be used to compare judges and lawyers arguing cases.

This approach is appealing for several reasons. First, contribution-based scores are available for anyone who has made political contributions or has run for office. This includes not just federal judges but also state judges (discussed below), political actors in federal, state, and local government (legislators, presidents, governors, attorneys general, and so on), interest groups, and tens of millions of individual donors (including lawyers). In this way, the DIME measure enables a broad range of inter-institutional studies, because the scores are estimated in a consistent manner for all actors and thus are directly comparable. An example can be found in Bonica and Sen (2017a), which examines the correspondence between US Supreme Court justice voting patterns and the ideology of the attorneys representing clients before the Court. Second, judges appointed by the same president in the same jurisdiction can and frequently are assigned different ideological estimates, making these scores more fine-grained than Judicial Common Space scores. This enables more in-depth inquiry based on cross-judge differences in ideology—useful for scholars who are exploiting random case assignment within jurisdictions as a causal identification strategy (which we discuss below).

We note some drawbacks to this approach, as well. First, not all judges have engaged in preconfirmation political activity, and this lack of activity could correlate

---

[1] To illustrate in simplified terms, suppose a person donates $500 to Hillary Clinton in 2016 and then $500 to 2020 presidential candidate Elizabeth Warren. DIME would assign this donor a score halfway between Warren (left) and Clinton (center-left). However, a donor giving $250 to Warren and $500 to Clinton would be assigned a score closer to Clinton than Warren, reflecting that they are likely to be closer to Clinton ideologically.

with ideology—in turn, suggesting non-random missingness. According to Bonica and Sen (2017b), some 81 percent of federal appeals judges appointed since 2001 have made campaign contributions—a large share compared to the general population but still short of perfect coverage. In Bonica and Sen (2017a), we impute scores for these missing judges from observed characteristics. Second, the contribution-based scores are estimated (for federal judges) using pre-confirmation information. This means that contribution-based scores can be of limited use in studying dynamic patterns, such as intellectual drift or responsiveness to current events after judges have taken the bench.

In this vein, Bonica et al. (2017) use the contribution-based scores from Bonica (2014) and impute scores for federal judges based on the ideology of their law clerks, as revealed by the clerks' political contributions. Each year federal judges hire several clerks to assist them in drafting opinions, doing legal research, and evaluating the arguments presented at oral arguments. (Each federal appeals judge is entitled to hire four clerks yearly, and each federal district judge is entitled to hire two.) The substantive literature on hiring—and statements made by law professors and other anecdotal evidence—strongly suggests that clerks are hired partly on the basis of shared ideology with the hiring judge. (Some judges are known to hire contrarian-leaning clerks—a phenomenon known as a hiring a "counter clerk.") Thus, averaging the scores for clerks provides an informative signal about a judge's ideology.

Because judges hire new cohorts of law clerks every year, the clerk-based ideology scores are dynamic. If a judge hires mostly conservative clerks early on in her career but later on begins to hire mostly liberal clerks, this is indicative of ideological drift. The clerk-based ideology scores reasonably predict other measures of judicial ideology. One advantage of the clerk-based ideology scores is that they provide estimates for both Supreme Court justices and federal lower court judges.

Random Assignment of Judges

A long-standing norm within the federal courts is that litigants should not be able to choose their judge. Of course, this does not prevent litigants from "forum-shopping"—that is, trying to choose a jurisdiction where they believe that the judges will give their case a more favorable hearing. But while litigants may have some discretion over jurisdiction, they have little control over which of the judges in that jurisdiction will hear their case. In the federal courts of appeals, where most cases are heard by three-judge panels, panels are formed on the basis of judges' schedules or other factors, and then (in many jurisdictions) cases are assigned "randomly" via a computer program or another unpredictable mechanism to a panel. In federal district courts, where judges individually hear cases, cases are typically randomly assigned to judges within a district, conditional on availability, vacation days, and workload balance.

Randomization in the selection of federal judges is extremely useful for scholars, because it can be used to estimate the causal effect of judge characteristics (including partisanship or ideology) on case outcomes. One of the most comprehensive studies in this vein is Sunstein et al. (2006), which relies on the random

assignment of cases to three-judge panels to estimate the causal impact of panel composition—specifically, how a case will fare with more or fewer Republican versus Democratic appointed judges. "Dampening" refers to the common pattern whereby a judge on a panel with two judges appointed by the other party is more likely to go along with a ruling at odds with their ideology. Conversely, "amplification" means that three judges appointed by the same party are more likely to reach a decision consistent with their ideology. In the Sunstein et al. evidence, judges appointed by Democrats are more susceptible to these effects than those appointed by Republicans. Another example is Epstein, Landes, and Posner (2013), which examines similar issues but looks more broadly at the entirety of the federal courts.

The strategy of using the random assignment of cases to judges can be leveraged to explore the role of ideology and partisanship across issue areas. For example, Cox and Miles (2008) examine federal court cases involving Section 2 of the Voting Rights Act. They find a substantial and growing gap in how judges appointed by Democrats and Republicans rule on voting rights cases. As another example, Cohen and Yang (2019) find that 65 percent of the unadjusted gap in sentence length between blacks and whites in federal district courts can be attributed to Republican-appointed judges giving black defendants longer sentences (of three months, on average) than similar nonblacks, as compared to Democratic-appointed judges. Thus, people being sentenced for a crime will randomly get a higher or lower sentence, depending on which judge they draw. Several other studies have used the random assignment of cases to judges as an instrument, investigating the effect of judging tendencies on a wide variety of downstream outcomes (Kling 2006; Dobbie, Goldin, and Yang 2018).

These findings that rely on the random assignment of cases to judges do not measure judicial ideology per se. Still, they do help identify the effect of judicial characteristics, of which partisanship is one. On a larger scale, leveraging random assignment allows scholars to put the impact of partisanship front and center in their research designs. However, an important caveat is that the randomization of cases in federal courts is very far from perfect, and scholars should be careful in claiming randomization for purposes of causal identification. Judges might recuse themselves or have a scheduling conflict; court clerks might try to balance workload, try to rotate cases so that judges sit with more colleagues, or have other considerations that lead them to break from true random assignment. Several studies have shown that case assignments for the federal appeals courts deviate significantly from what would be expected under true randomization (Chilton and Levy 2015), while others have documented instances where courts claim randomization but actually use non-random rubrics in case assignment (for example, Hildabrand 2019). Presumably, similar breaks from randomization occur for other federal courts and state courts that claim to employ random assignment. In addition, because lower court federal judges are grouped in panels of three for a particular sitting (where they might hear dozens of cases together), and then cases are randomly assigned to panels, thinking carefully about randomization and the appropriate unit of analysis (appeals panels versus individual judges) is essential.

Text-based Analysis

Automated text analysis is frequently used outside of the legal context to measure ideology. Scholars have, for example, analyzed floor speeches made on the US Congressional floor (for example, Diermeier et al. 2012) to estimate legislative ideology. The general idea is that the greater use of certain words is likely to be associated with certain ideologies—for example, "death tax" or "Obamacare" would be more likely to be associated with a conservative ideology. Most of these methods use machine learning methods for text classification, with some portion of text documents classified by hand as "conservative" or "liberal" as a training set. Examples of approaches that focus specifically on ideologically oriented text include Wordscores (Laver, Benoit, and Garry 2003); Wordfish (Slapin and Proksch 2008); and Wordshoal (Lauderdale and Herzog 2016).

The courts present a special challenge, however. Unlike elected officials, whose public statements are made with non-specialist members of the public in mind, lawyers communicate in a specific legal language, which creates challenges for mapping legal concepts and language directly onto ideology. Nonetheless, we note several recent attempts to scale case-level ideology, which could be extended to estimate judicial ideology. A notable example is Lauderdale and Clark (2016), which uses a conditional autoregressive preference measurement model to examine case-level voting at the US Supreme Court. Instead of trying to estimate a general measure of judicial ideology, this approach estimates the justices' latent preferences on every vote, looking closely at cut-points in each case. The authors use this model to generate case-specific preference estimates for US Supreme Court justices from 1946 to 2005, revealing substantively meaningful variation in the relative ideological ordering across cases.

Another recent attempt has been Hausladen, Schubert, and Ash (2020), which attempts to estimate the ideology of written opinions from the federal courts of appeals using supervised machine learning methods. The authors hand-coded around 5 percent of cases in their sample and then used this as a training set; the algorithm then accurately predicts federal appeals decisions hand-coded by Landes and Posner (2009) from the Songer et al. Court of Appeals dataset.

Judicial Ideology in State Courts

Most of the discussion so far has focused on the US federal courts. However, whereas the federal courts combined hear approximately several hundred thousand cases per year, state courts hear vastly more—an estimated 84 million per year (http://www.courtstatistics.org). In addition, while federal courts are limited in their jurisdiction to only those cases having a federal component or to those cases involving cross-state litigants, state courts hear the vast majority of within-state legal matters—including most matters pertaining to criminal punishment, family matters (including child custody and divorces), and trusts and estates processing. Thus, decisions by state courts are of vital importance to bread-and-butter economic and social issues and therefore studying them is of high value in economics, sociology, and political science.

However, the challenges for studying the behavior and backgrounds of state judges are formidable. The federal courts are a contained group of some 1,000

individuals whose identities and prior partisan and employment histories are easily tracked via extant resources. For state judges—spread across all 50 states and the District of Columbia—the task is vastly more difficult. A good starting point for researching the identities and backgrounds of state judges are new, open-source initiatives such as Courtlistener (https:// www.courtlistener.com/person/) or Ballotpedia (https://ballotpedia.org/Courts_ and_judges_by_state).

In terms of estimating judicial ideology, the difficulty extends beyond numbers. Federal judges are selected via the same procedure—nomination by the president with the "advice and consent" of the US Senate. This makes it straightforward to impute executive (or senatorial) ideology for the judges' ideologies (although as we note above, doing so by necessity introduces measurement error). In contrast, state judges are chosen via an amalgam of selection methods including, but not limited to, merit commissions, partisan elections, nonpartisan elections, legislative appointments, and executive (gubernatorial) appointments. The larger number of actors makes it intractable to use the same methodology for state judges as, for example, is followed by Judicial Common Space scores for federal judges.

Nonetheless, scholars have made strong inroads in the estimation of judicial ideology at the state level. The most prominent of these are Party-Adjusted surrogate Judge Ideology (PAJID) scores developed by Brace, Langer, and Hall (2000), which are focused on ideology for justices on state supreme courts. These scores build off of the intuition of Judicial Common Space scores and others by looking to the relevant political actors to be "surrogates" for judicial ideology; that is, these scores borrow the political ideology of pertinent political actors to impute the likely ideology scores of the judges in question. In states with appointments systems, this would be the ideologies of political elites; in states with elections, this would be the electorate. Thus, the authors use "elite ideologies for appointed judges and citizen ideologies for elected judges" (Brace, Langer, and Hall 2000, p. 397), further adjusting by the expressed partisanship of the judge. The result is scores that capture the variegated nature of judicial selection at the state level; however, as with other approaches that use ideological surrogates (like Judicial Common Space scores), the PAJID scores are not based on the judges' own revealed behavior..

Another approach to estimating the ideology of state judges is provided by Bonica and Woodruff (2015) and Bonica and Sen (2017a), which rely on the campaign contribution scores methodology presented in Bonica (2014). Similarly to the DIME scores for federal judges, these scores take judges' political activity as expressed through campaign contributions and assume that this political activity is reflective of their true ideological preferences. Again like federal judges, the coverage is reasonably strong, reflecting the politicized nature of judicial offices: of state high court judges, 71 percent are included in the contributions database. Moreover, because the methodology scales judges from across federal and state courts in the same fashion, the ideology estimates are comparable across different areas of the judiciary, legislative bodies, and other political actors. This facilitates comparisons across presidents, members of Congress, state legislators, state executives, legal elites, and even litigating parties.

Judicial Ideology on Other National Courts or International Courts

The literature on estimating the ideology of judges on national or subnational courts in other countries is too widespread for a brief synthesis. However, those interested in learning more might begin with estimates of the ideology of judges in Argentina (Bertomeu, Dalla Pellegrina, and Garoupa 2017), Canada (Songer et al. 2012), the United Kingdom (Hanretty 2013; Arvind and Stirton 2016), Spain and Portugal (Hanretty 2012); Taiwan (Dalla Pellegrina, Garoupa, and Lin 2012), and Philippines (Dalla Pellegrina, Escresa, and Garoupa 2014).

The estimation of judicial ideology has so far made little progress in the increasingly substantively important arena of international courts. The challenges here are numerous—these courts tend to have less regular terms and more idiosyncratic caseloads, making bridging across panels difficult. One notable effort here is the work of Frankenreiter (2018) on the EU Court of Justice.

## How Judicial Ideology Has Polarized over Time

As the salience of judicial ideology has grown in recent years, so has judicial polarization. Judges appointed to the federal courts, from both parties, are increasingly being selected based on their partisan bona fides and being drawn from the ideological extremes. While polarization is on full public display in recent Supreme Court nomination battles, this trend is also playing out in the federal courts more generally. In this section, we consider some of the evidence, using Martin-Quinn scores for Supreme Court justices and campaign contribution-based measures of judicial ideology for lower court judges.

One sign of polarization is that Supreme Court justices have sorted into distinct ideological voting blocks along party lines. The notion that a president would nominate a justice who would align with the opposing ideological camp, something that had been relatively common in the past, is now unthinkable. The lower courts have polarized alongside the Supreme Court, to the extent that federal district and circuit court judges are now nearly as polarized as the parties in Congress. These trends look poised to continue for the foreseeable future, which has important implications for judicial decision-making on politicized areas of law. It also threatens to diminish the legitimacy of the courts in the eyes of the public.

Polarization on the Supreme Court

Tracking polarization in the US Supreme Court is an important contribution of the literature on measuring judicial ideology. As the nation's highest court, many issues of economic, societal, and political importance are decided each term. The ideological composition of the court has become a major battleground for partisan conflict. Supreme Court nomination battles were front and center in the 2016, 2018, and 2020 elections, with both parties emphasizing to their respective bases the vital importance of Supreme Court appointments to their policy agendas.

Figure 1
Martin-Quinn Estimates of Supreme Court Justice Ideology

Source: https://mqscores.1sa.umich.edu
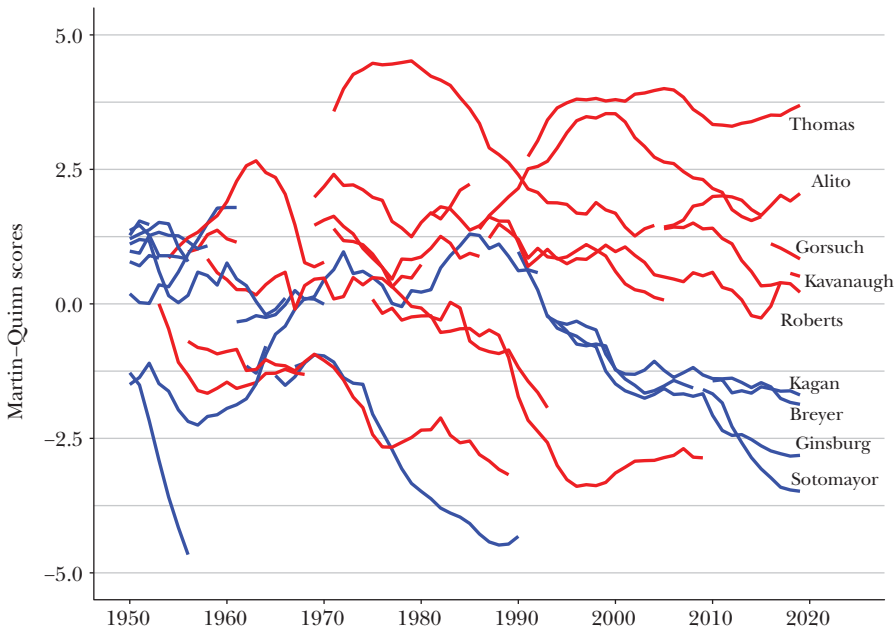
The Martin-Quinn dynamic ideal point trends show a growing partisan divide on the Supreme Court in recent decades, as seen in Figure 1. Prior to the 1990s, voting patterns reveal substantial ideological overlap of justices appointed by Republicans and Democrats. Following the retirement of John Paul Stevens in 2009, justices separated into two distinct ideological voting blocks along party lines.

While nomination battles and the ideological preferences of justices have become polarized, the business of the Court has not necessarily done so to the same extent. Even as justices appointed by Republicans and Democrats have sorted into distinct voting blocks, the percentage of unanimous decisions issued by the Court has increased in recent decades. Straight party-line voting, where justices vote as opposing partisan blocks, remains relatively infrequent, accounting for about 10 percent of cases decided during the 2018 and 2019 terms. This is a stark contrast with voting patterns in Congress, where party line votes account for more than 70 percent of floor votes. This pattern likely reflects that the Supreme Court continues to perform its institutional role of resolving circuit splits and providing uniformity on legal issues (Bartels 2015).

At the same time, some of the most controversial and politically consequential Supreme Court rulings over the past decade have been decided by close votes

along ideological lines. When the Supreme Court considers cases that have major implications for setting public policy, it tends to behave more like their partisan counterparts in Congress. Some examples of the landmark cases during the 2010s that were generally decided along ideological lines include *Citizens United v. FEC* (558 U.S. 310 [2010]) concerning campaign finance, *Shelby County v. Holder* (570 U.S. 529 [2013]) concerning voting rights, *Obergefell v. Hodges* (576 U.S. 644 [2015]) concerning same-sex marriage, *Arizona v. United States* (567 U.S. 387 [2012]) addressing conflicts between federal and state-level immigration law, *Janus v. AFSCME* (585 U.S. [2018]) on whether labor unions could collect fees from non-union members, *Burwell v. Hobby Lobby* (573 U.S. 682 [2014]) on the circumstances in which a closely held corporation could be exempt from federal regulation on religious grounds, and *Schuette v. Coalition to Defend Affirmative Action* (572 U.S. 291 [2014]) on whether states could ban affirmative action in education.

Republican and Democratic Federal Appointments over Time

While selecting judges to serve on the federal bench has never been entirely free from partisan considerations, the choices have historically been mediated to varying extents by concerns about judicial qualifications, norms of impartiality, and institutional arrangements designed to promote bipartisan compromise in the Senate. Figure 2 provides a sense of the extent to which the countervailing considerations have given way to a more purely partisan selection process. It plots the ideological distribution of federal judges appointed during recent presidential administrations based on the contribution-based DIME scores of judicial ideology discussed earlier. Box-and-whisker plots are included to show the median and relative dispersion of ideology scores for judges appointed by each president. Comparing the distributions of more recent administrations and earlier administrations reveals a general trend towards ideologically driven selection of judges.

This trend towards appointing ideologically aligned judges has exacerbated judicial polarization. Figure 3 plots the polarization trends for the US District and Circuit courts. To measure polarization, we draw from the standard approach in the literature on congressional polarization. We first group judges appointed by Democratic and Republican presidents and calculate each group's average ideology by year. We then calculate the distance between the party means as the measure of polarization in each year. A unit value on this scale is equivalent to a standard deviation in the scores for congressional candidates, or in practical terms, about the ideological distance between Senators Joe Biden (D-DE) and Susan Collins (R-ME). The polarization of circuit court judges has outpaced that of district court judges, increasing from 1.04 to 1.50 compared with an increase from 0.98 to 1.26, reflecting the appellate courts' higher political value. For reference, the federal appellate courts in 2018 were about as polarized as Congress was in the mid-2000s. This suggests that supposing federal court judges were instead serving in Congress, few would be considered moderates, with most behaving as partisans.

Figure 2

Ideological Distributions of Judges Appointed by Presidential Administrations
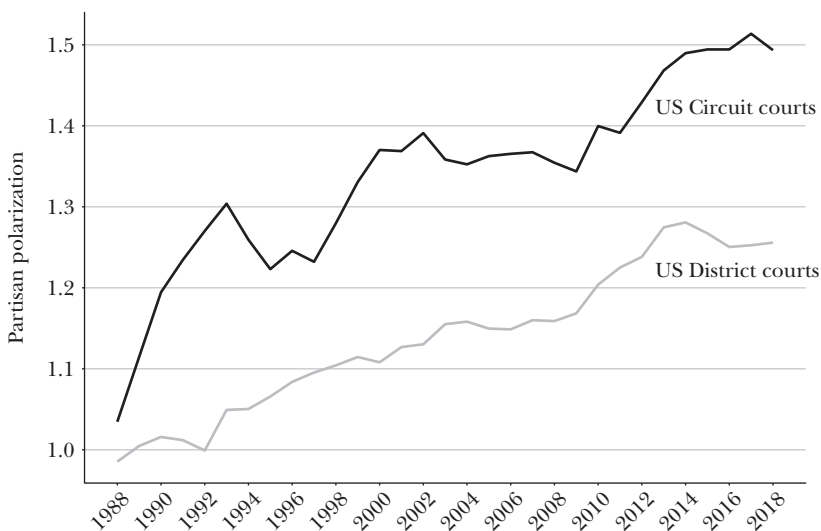
Figure 3
Ideological Distance between Federal Judges Appointed by Democratic and Republican Presidents
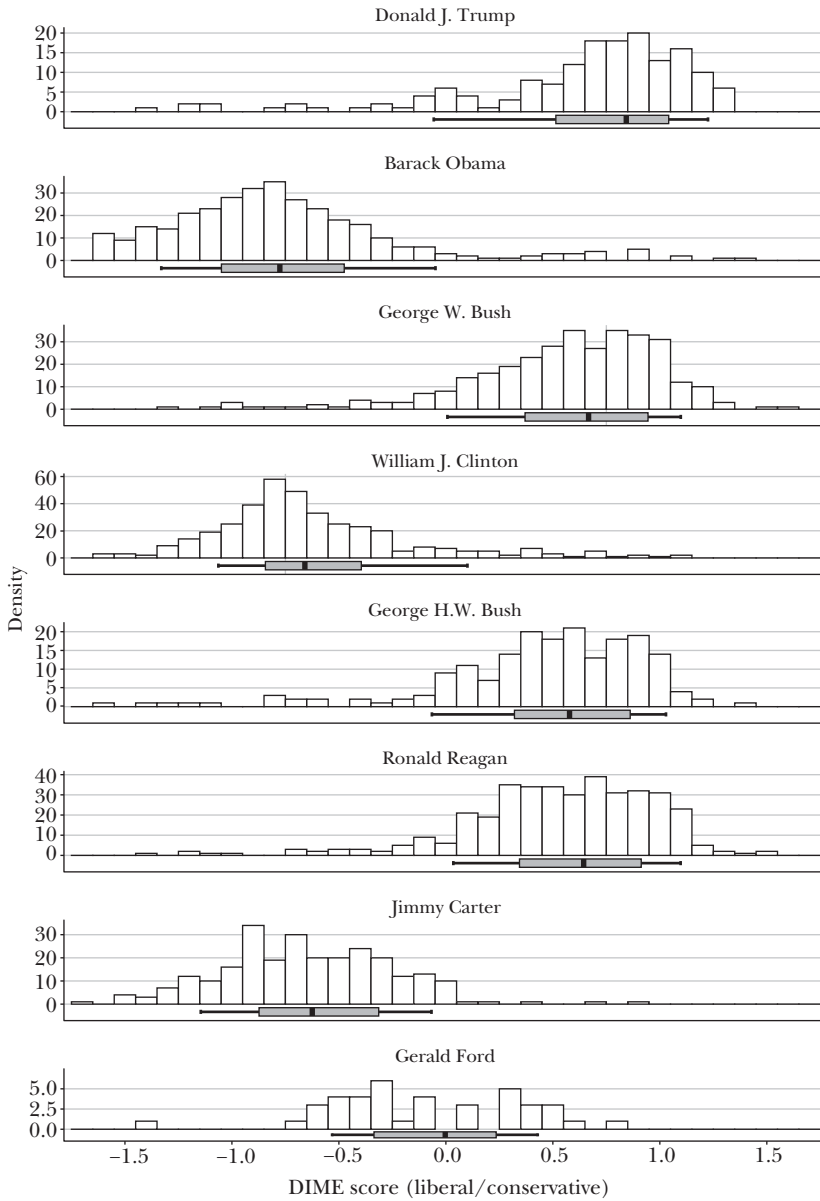
Source: Federal Judicial Center Biographical Directory, DIME

Consequences of Increasing Judicial Polarization

There is evidence that judicial polarization has influenced the judicial system in ways large and small. As ideology becomes more predictive of judicial disposition, it creates opportunities for "venue shopping," a process whereby litigants filing in the federal courts can strategically exercise discretion about which federal circuit to bring their case.

The effects of judicial polarization are perhaps most clearly observed in the judicial selection process (Devins and Baum 2017). State and federal courts routinely issue decisions on questions of paramount political importance, from health care policy and abortion rights to voting rights and redistricting. As a result, support from the courts is a highly sought-after political prize in American politics. As partisans compete to secure seats for ideologically friendly judges, they have sought to manipulate the institutional rules and mechanisms used to select judges (Bonica and Sen 2017b).

Judicial polarization also has downstream effects on the labor market for legal elites. The courts have polarized alongside a generational shift within the legal profession, with upwards of 90 percent of recent graduates from elite law schools identifying as Democrats (Bonica et al. 2018; Fisman et al. 2015). As demand for conservative judges has outpaced the supply of potential jurists from elite law schools, conservative graduates of these programs have become far more likely to be chosen for coveted federal clerkships and are much more likely to be appointed to positions on the federal bench than their more numerous liberal counterparts.

One especially fraught set of issues are cases concerning voting rights, campaign finance, and elections. The judicialization of politics, as Ferejohn (2002) has phrased it, refers to a long-term trend whereby courts have increasingly involved themselves in regulating political activity and adjudicating election disputes. The earlier discussion listed some high-profile Supreme Court cases which were decided largely along partisan lines, but for many people, perhaps the most notable example is *Bush v. Gore* (531 US 98 [2000]), which ended a set of disputes over the conduct of voting recounts in the state of Florida and had the proximate effect of Bush winning the presidential election. There is evidence that these patterns apply to the lower courts, as well. Cox and Miles (2008) find that partisanship predicts how federal judges decide cases related to Section 2 of the Voting Rights Act. Kang and Shepherd (2016) report a similar partisan pattern relating to election disputes in state courts.

Another symptom of judicial polarization that has presented itself in recent years relates to how the courts have ruled on lawsuits brought against President Donald Trump and his administration. Dozens of lawsuits have been filed against Trump regarding his personal and official actions. Many of these cases raise fundamental constitutional issues and extend beyond policy matters to questions of executive authority, congressional oversight, and President Trump's personal conduct and finances. Figure 4 plots the disposition of rulings (coded by rulings for or against the president) against the median ideology of three-judge US circuit court panels hearing the cases. Ideology emerges as a strong predictor of case outcomes. Right-of-center panels tend to rule in favor of the president, while left-of-center overwhelmingly rule against him. A similar analysis performed by the Institute for Policy Integrity (2019) tracked the outcomes of litigation regarding the Trump administration's use of federal agencies. It finds that 8 percent of lawsuits heard by Democratic judges were ruled favorably for the Trump administration compared with 44 percent of cases heard by Republican judges.

Conclusion

Substantial scholarship has documented that the ideology of judges plays an important role in shaping their judicial behavior. The accumulated research goes back decades, to the 1940s and 1950s. Although most of the work has focused on the US Supreme Court, evidence of the important role of judicial ideology in predicting judicial rulings has been found at the other levels of state and federal courts. The effects of judicial ideology do not just involve topics of federal interest, such as civil rights or voting rights, but also cases that are less ostensibly political, such as the sentencing decisions of criminal defendants. This evidence has been built on a number of measures of judicial ideology: the party of the politician who appointed the judge, scores based on voting patterns, informed opinions like newspaper editorials before a Supreme Court justice is appointed, campaign contributions made by judges before they were appointed, analysis of text for clues about political ideology, and others. We also show that judges have become more polarized.

Figure 4

Rulings on Cases Involving President Trump by Ideology of Circuit Court Panels

Courts of Appeals panel median ideology
(liberal/conservative)
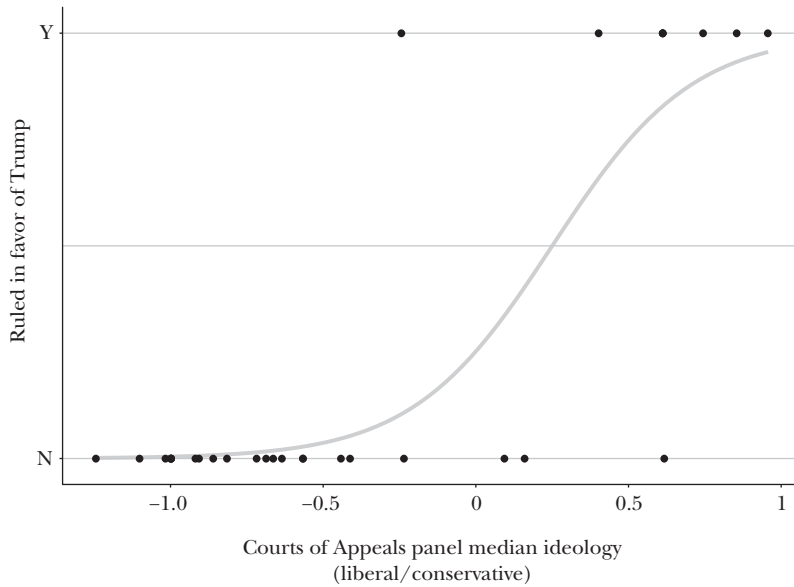
Source: Author's calculations, DIME.

Note: Each observation is a case decided by the federal circuit courts in which the President Trump or his administration was a party. The trend line is estimated from logistic regression.

However, this literature does not purport to suggest that ideology is the only or the primary predictor of judicial decision making. Judging operates in tandem with legal constraints and the institutional constraints that come along with the courts. Judges are also constrained by the need to justify their decisions, together with their desire to have not just the approval of their peers for reaching a certain result but also the respect of their peers for how that result was reached.

References

Arvind, T.T., and Lindsay Stirton. 2016. "Legal Ideology, Legal Doctrine and the UK's Top Judges." *Public Law* 2016 (July): 418–36.

Bailey, Michael A. 2007. "Comparable Preference Estimates across Time and Institutions for the Court, Congress, and Presidency." *American Journal of Political Science* 51 (3):433–48.

Bailey, Michael A. 2017. "Measuring Ideology on the Courts." In *Routledge Handbook of Judicial Behavior*, edited by Robert M. Howard and Kirk A. Randazzo, 62–83. New York: Routledge.

Bartels, Brandon. 2015. "The Sources and Consequences of Polarization in the US Supreme Court." In *American Gridlock: The Sources, Character, and Impact of Political Polarization*, edited by James A. Thurber and Antoine Yoshinaka, 171–200. Cambridge, UK: Cambridge University Press.

Bertomeu, Juan González, Lucia Dalla Pellegrina, and Nuno Garoupa. 2017. "Estimating Judicial Ideal Points in Latin America: The Case of Argentina." *Review of Law & Economics* 13 (1).

Bonica, Adam. 2014. "Mapping the Ideological Marketplace." *American Journal of Political Science* 58 (2): 367–86.

Bonica, Adam. 2016. "Database on Ideology, Money in Politics, and Elections: Public Version 2.0." http://data.stanford.edu/dime (accessed October 4, 2020).

Bonica, Adam. 2019. "Are Donation-Based Measures of Ideology Valid Predictors of Individual-Level Policy Preferences?" *The Journal of Politics* 81 (1): 327–33.

Bonica, Adam, Adam S. Chilton, Jacob Goldin, Kyle Rozema and Maya Sen. 2017. "Measuring Judicial Ideology Using Law Clerk Hiring." *American Law and Economics Review* 19 (1): 129–61.

Bonica, Adam, Adam Chilton, Kyle Rozema and Maya Sen. 2018. "The Legal Academy's Ideological Uniformity." *Journal of Legal Studies* 47 (1):1–43.

Bonica, Adam, and Maya Sen. 2017a. "A Common-Space Scaling of the American Judiciary and Legal Profession." *Political Analysis* 25 (1): 114–21.

Bonica, Adam, and Maya Sen. 2017b. "The Politics of Selecting the Bench from the Bar: The Legal Profession and Partisan Incentives to Introduce Ideology into Judicial Selection." *The Journal of Law and Economics* 60 (4): 559–95.

Bonica, Adam, and Michael J. Woodruff. 2015. "A Common-Space Measure of State Supreme Court Ideology." *Journal of Law, Economics, & Organization* 31 (3): 472–98.

Boyd, Christina L. 2015. "Federal District Court Judge Ideology Data." http://cLboyd.net/ideology.html. (accessed November 4, 2020).

Brace, Paul, Laura Langer, and Melinda Gann Hall. 2000. "Measuring the Preferences of State Supreme Court Judges." *The Journal of Politics* 62 (2): 387–413.

Chilton, Adam S., and Marin K. Levy. 2015. "Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals." *Cornell L. Rev.* 101 (1).

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355–70.

Cohen, Alma, and Crystal S. Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy* 11 (1): 160–91.

Cox, Adam B., and Thomas J. Miles. 2008. "Judging the Voting Rights Act." *Columbia Law Review* 108 (1): 1–54.

Dalla Pellegrina, Lucia, Laarni Escresa, and Nuno Garoupa. 2014. "Measuring Judicial Ideal Points in New Democracies: The Case of the Philippines." *Asian Journal of Law and Society* 1 (1): 125–64.

Dalla Pellegrina, Lucia , Nuno Garoupa, and Shirley Ching-ping Lin. 2012. "Judicial Ideal Points in New Democracies: The Case of Taiwan." *NTU L. Rev.* 7 (123).

Devins, Neal, and Lawrence Baum. 2017. "Split Definitive: How Party Polarization Turned the Supreme Court into a Partisan Court." *The Supreme Court Review* 2016: 301–65.

Diermeier, Daniel, Jean-François Godbout, Bei Yu, and Stefan Kaufmann. 2012. "Language and Ideology in Congress." *British Journal of Political Science* 42 (1): 31–55.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review* 108 (2): 201–40.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2013. *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice.* Cambridge, MA: Harvard University Press.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland. 2007. "The Judicial Common Space." The Journal of Law, Economics, and Organization 23 (2): 303–25.

Ferejohn, John. 2002. "Judicializing Politics, Politicizing Law." Law and Contemporary Problems 65 (3): 41–68.

Fischman, Joshua B., and David S. Law. 2009. "What Is Judicial Ideology, and How Should We Measure It?" *Washington University Journal of Law & Policy* 29 (133).

Fisman, Raymond, Pamela Jakiela, Shachar Kariv, and Daniel Markovits. 2015. "The Distributional Preferences of an Elite." *Science* 349 (6254).

Frankenreiter, Jens. 2018. "Are Advocates General Political? Policy Preferences of EU Member State Governments and the Voting Behavior of Members of the European Court of Justice." *Review of*

Law & Economics 14 (1).

Giles, Micheal W., Virginia A. Hettinger, and Todd Peppers. 2001. "Picking Federal Judges: A Note on Policy and Partisan Selection Agendas." *Political Research Quarterly* 54 (3): 623–41.

Hanretty, Chris. 2012. "Dissent in Iberia: The Ideal Points of Justices on the Spanish and Portuguese Constitutional Tribunals." *European Journal of Political Research* 51 (5): 671–92.

Hanretty, Chris. 2013. "The Decisions and Ideal Points of British Law Lords." *British Journal of Political Science* 43 (3):703–16.

Harvey, Anna, and Michael J. Woodruff. 2013. "Confirmation Bias in the United States Supreme Court Judicial Database." *The Journal of Law, Economics, & Organization* 29 (2): 414–60.

Hausladen, Carina I., Marcel H. Schubert, and Elliott Ash. 2020. "Text Classification of Ideological Direction in Judicial Opinions." *International Review of Law and Economics* 62.

Hildabrand, Clark L. 2019. "The Curiously Nonrandom Assignment of Sixth Circuit Senior Judges." *Kentucky Law Journal* 108.

Ho, Daniel E., and Kevin M. Quinn. 2010. "How Not to Lie with Judicial Votes: Misconceptions, Measurement, and Models." *California Law Review* 98 (3): 813–76.

Institute for Policy Integrity. 2019. "Roundup: Trump-Era Agency Policy in the Courts." https://policy-integrity.org/trump-court-roundup.

Kang, Michael S., and Joanna M. Shepherd. 2016. "The Long Shadow of Bush v. Gore: Judicial Partisanship in Election Cases." *Stanford Law Review* 68 (1411).

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–76.

Landes, William M., and Richard A. Posner. 2009. *The Economic Structure of Intellectual Property Law.* Cambridge, MA: Harvard University Press.

Lauderdale, Benjamin, and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58 (3): 754–71.

Lauderdale, Benjamin E., and Tom S. Clark. 2016. "Estimating Vote-Specific Preferences from Roll-Call Data Using Conditional Autoregressive Priors." *The Journal of Politics* 78 (4): 1153–69.

Lauderdale, Benjamin E., and Alexander Herzog. 2016. "Measuring Political Positions from Legislative Speech." *Political Analysis* 24 (3): 374–94.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97 (2): 311–31.

Lissner, Michael. 2019. "We've Integrated the FJC Integrated Database into CourtListener." The Integrated Database. https://free.law/2019/02/14/weve-integrated-the-fjc-integrated-database-into-courtlistener/ (accessed November 4, 2020).

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10 (2): 134–53.

Nixon, David. 2004. "APPENDIX A: Ideology Scores for Judicial Appointees." http://www2.hawaii.edu/~dnixon/PIMP/judicial.pdf.

Owens, Ryan J., and Justin Wedeking. 2012. "Predicting Drift on Politically Insulated Institutions: A Study of Ideological Drift on the United States Supreme Court." *The Journal of Politics* 74 (2): 487–500.

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting: Analytical Methods for Social Research.* Cambridge, UK: Cambridge University Press.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting.* Oxford: Oxford University Press.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83 (2): 557–65.

Segal, Jeffrey A., and Harold J. Spaeth. 2002. *The Supreme Court and the Attitudinal Model Revisited.* Cambridge, UK: Cambridge University Press.

Sherman, Mark. 2018. "Roberts, Trump Spar in Extraordinary Scrap over Judges." *Associated Press*, November, 21. https://apnews.com/article/c4b34f9639e141069c08cf1e3deb6b84.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. "A Scaling Model for Estimating Time-Series Party Positions from Texts." *American Journal of Political Science* 52 (3): 705–22.

Songer, Donald R., Susan W. Johnson, C.L. Ostberg, and Matthew E. Wetstein. 2012. *Law, Ideology, and Collegiality: Judicial Behaviour in the Supreme Court of Canada.* Montreal: McGill-Queen's Press.

Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki. 2006. *Are Judges Political?: An Empirical Analysis of the Federal Judiciary.* Washington, DC: Brookings Institution Press.

Can Structural Changes Fix the Supreme Court?

Daniel Hemel

The US Supreme Court, by many accounts, is facing a "legitimacy crisis" (Tomasky 2018; Waldman 2018; Epps and Sitaraman 2019; Johnsen 2020). Prominent scholars of the court characterize the confirmation process for new justices as "dysfunctional" (Barnett and Blackman 2016; Walsh 2016). They describe the behavior of the court's members as "nakedly partisan" (*New York Times* Editorial Board 2017; Epstein and Posner 2018). And they warn that the court is turning into a "supreme gerontocracy" (Calabresi and Lindgren 2005), whose members serve "ever longer terms" (Carrington and Crampton 2005, 468) and often linger on the bench even after entering a state of "mental decrepitude" (Garrow 2000; Garrow 2005, 273).

Reports of declining public confidence in the court are particularly threatening to an institution whose influence depends in large part upon both popular and elite opinion (Dahl 1957). US Supreme Court justices are appointed by the president and—if confirmed by a majority of the Senate—serve for life, subject only to the largely theoretical threat of impeachment by a majority of the House followed by conviction on a two-thirds vote of the Senate. Only one justice has ever been impeached (Samuel Chase in 1805), and the Senate did not vote to convict. But notwithstanding their almost impenetrable life-tenure protection, the justices are entirely reliant on other actors for their authority. The court has no funding source of its own and so must ask the House and Senate for money each year. The court has no law-enforcement authority (beyond a small force of approximately 160

■ *Daniel Hemel is Professor of Law and Ronald H. Coase Research Scholar at University of Chicago Law School, Chicago, Illinois. His email address is dhemel@uchicago.edu.*

police officers who guard the justices and the court grounds) and thus must depend upon the executive branch, helmed by the president, to implement its rulings. The court's practical power rests entirely upon the other branches and the citizenry accepting its decisions as binding. Legitimacy is thus its stock-in-trade.

It was for this reason that Alexander Hamilton (1788), during the debates over the ratification of the Constitution, described the judiciary as the "least dangerous" branch. But as the legal scholar Alexander Bickel (1962, 1) would later observe, "The least dangerous branch of the American government is the most extraordinarily powerful court of law the world has ever known." The Supreme Court has long asserted the power of "judicial review"—in other words, the power to strike down federal and state laws as unconstitutional—and although the text of the Constitution says nothing specifically about the subject, the other federal branches and the states have largely acquiesced to the court on this point. The court has wielded this power to invalidate laws that had segregated schools, banned abortions and contraceptives, and prohibited interracial and same-sex marriages. Through its exercise of judicial review, it has influenced virtually all aspects of American life.

With great power comes inevitable controversy. Since the nation's founding, public attitudes toward the Supreme Court have swung "from reverence to condemnation and back to reverence, often in one generation" (Steamer 1971, 4). Many of the current complaints about the court—that the confirmation process has broken down, that the justices have become "politicians in robes," and that the court's members remain on the bench for too long—are the same charges that were leveled against the court decades earlier (Roosevelt 1937; Rehnquist 1959; Sheldon 1970). Then and now, these complaints have motivated calls for structural reform.

This paper begins with a critical evaluation of claims that this institutional structure has produced a court in "crisis": a decline in public confidence, contentious confirmation hearings, polarized voting patterns on the court, uneven allocation of appointments between the political parties, and increasing tenure and age of justices. The discussion then turns to some prominent proposals for structural reform of the court and examines how these proposals can be expected to shape the selection and behavior of justices.

The proposal that has drawn the most attention and the broadest support is the idea of 18-year term limits for the justices. Proponents argue that term limits would lower the stakes of confirmation fights, making a brutally bitter process somewhat less so (for example, Ornstein 2014; Ackerman 2018). They also argue that term limits would make the court "less partisan" (American Academy of Arts and Sciences 2020, 31) and less likely to engage in "policymaking" (Graglia 1994). Further, they argue that term limits would prevent the emergence of a "supreme gerontocracy" whose members are superannuated, sometimes senile, and overly insulated from the mood of the electorate (Calabresi and Lindgren 2005). Upon closer examination, many of these claims seem overstated or implausible.

The paper then considers a number of alternatives to the term-limit proposal: imposing age limits on justices, selecting justices from the pool of lower-court judges by lottery, and imposing explicit political party balance requirements on the

court's membership. A less-discussed direction for court reform—and the one that this paper considers to be most promising—is to retain life tenure but decouple appointment opportunities from vacancies. The decoupling approach would produce an even allocation of appointment opportunities across presidential terms while bolstering the court's capacity to check the executive, without sacrificing advantages of the life-tenure status quo.[1]

To be sure, none of these reforms seem likely to pass in the near future, and even if they did, none seem likely to quell the controversy that has long surrounded the Supreme Court.[2] Nonetheless, considering such reforms serves to highlight the merits and flaws of the court's current structure. Analysis of reform proposals can thus enrich our understanding of the existing institution—an institution that, in all likelihood, will persist in its present form for some time.

## A Court in Crisis?

Public Confidence in the Court

Opinion polls provide an important source of information regarding public confidence in the court. In 2014, the polling organization Gallup announced that Americans' confidence in the Supreme Court had reached a "record low" (McCarthy 2014). Indeed, if one focuses on the percentage of Gallup respondents saying they have a "great deal" or "quite a lot" of confidence in the court, then the data do reflect a three-decade decline, from a peak of 56 percent in 1988 to a low of 30 percent in 2014. Notwithstanding a rebound since 2014, confidence by this metric still remained lower in 2019 than at any time before 2007. This is the measure most often emphasized by authors asserting that the court is in crisis (Seib 2018; Thomson-DeVeaux and Roeder 2018; Epps and Sitaraman 2019, 160).

A closer look at survey data paints a more complicated picture. Figure 1 illustrates the change over time in public confidence in the court as measured by Gallup and the General Social Survey (GSS), which is carried out by the National Opinion Research Center. The percentage of Gallup respondents reporting a "great deal" of confidence in the court in 2019 is equal to the mean for the entire 1974–2019

---

[1] The paper does not address proposals explicitly aimed at "packing" the court with justices of a particular party (for example, Millhiser 2019). The theoretical possibility of court packing may play a useful role as a democratic check on the court—a "nuclear option" that, like real nuclear options, is better left unexercised. As an actual strategy, it is likely to trigger tit-for-tat responses from the opposing party once power changes hands. The paper also does not address much more serious proposals to "disempower" the court, such as by stripping it of jurisdiction over cases involving the constitutionality of certain federal statutes (Doerfler and Moyn, forthcoming). To be sure, the normative premise that we should continue to empower the Supreme Court with broad judicial-review authority deserves thorough examination. I have sought to defend the premise elsewhere (Hemel 2018b). However, this paper takes that premise as a given.

[2] The website of the US Supreme Court, https://www.supremecourt.gov/about, provides a historical list of all justices and when they served, along with answers to basic questions about how the court works and links to decisions dating back to 1991.

Figure 1
Public Confidence in the US Supreme Court, 1973 to 2019
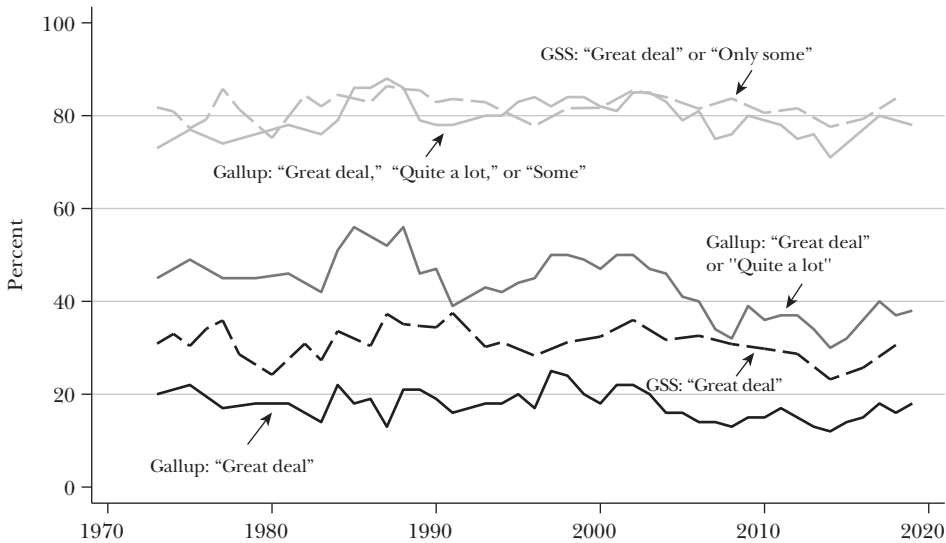
Source: Gallup, 1973–2019; General Social Survey (GSS), 1973–2018.

period (18 percent). The percentage of Gallup respondents saying they have at least "some" confidence in the court is not significantly different from the whole-period mean (78 percent in 2019 versus 79.6 percent mean). Likewise, the percentage of GSS respondents saying they have a "great deal" of confidence in the court in the most recent survey is equal to the mean over that survey's 1973–2018 period (31 percent). And the percentage of GSS respondents reporting at least "some" confidence is likewise roughly equal to (indeed, a hair above) the historical mean (83.7 percent in 2018 versus 82.1 percent mean).

Even if confidence in the court was declining (a claim not clearly borne out by the data), the normative implications would be ambiguous. When public confidence in the court is high, the political cost to the executive and legislature of not implementing the court's decisions increases, and the justices tend to be more willing to confront the other branches (Stephenson 2004; Clark 2011, 176). Those who want the justices to exert a strong check on the other branches might therefore be concerned about a lack of public confidence in the court. Conversely, those concerned about an "activist" court overriding the decisions of the people's elected representatives therefore may be worried about the court commanding too *much* public confidence. Without a fleshed-out theory of the appropriate judicial role, it is difficult to say whether declining public confidence in the court is a trend to be lamented.

Changes in the Confirmation Process

Conventional wisdom holds that the confirmation process for Supreme Court justices changed dramatically in July 1987, when the Senate rejected President Reagan's nominee Robert Bork by a 58–42 margin. Bork's credentials were unquestioned: a former Yale law professor and noted antitrust scholar, Bork had argued more than 40 cases at the high court as solicitor general under Presidents Nixon and Ford and had served as a judge on the prestigious US Court of Appeals for the DC Circuit. While previous nominees had been rejected because of thin qualifications, the Senate voted to reject Bork because of his judicial ideology. One commentator goes so far as to call the Bork confirmation "the beginning of the end of civil discourse in politics" (Nocera 2011, A17).

But what, exactly, changed with the Bork confirmation—and why? Figure 2, which describes the outcome of every Supreme Court nomination from 1789 through 2019, provides one way of contextualizing the Bork confirmation. The x's at the bottom show that a nominee was withdrawn before a Senate vote. Hollow circles show that a nominee was rejected on a roll call vote (with the share supporting the nominee measured on the vertical axis). The light-shaded circles show nominees confirmed on a roll-call vote, and the dark circles treat confirmation on a voice vote as receiving 100 percent support.

From the founding until the early 1890s, 19 nominations failed, either because the Senate declined to put them to a vote or because a majority of senators voted against confirmation. In some cases, these failures occurred because of concerns about the nominee's qualifications, but in other cases, opposition was strictly ideological or partisan (Abraham 2008, 102; Gerhardt 2002, 404–05).

Then, from the confirmation of Edward Douglass White in February 1894 through the confirmation of Thurgood Marshall in 1967, only one nomination failed. (In 1930, the Senate rejected President Herbert Hoover's nomination of the unapologetically racist Fourth Circuit judge John Parker.) One explanation for this remarkable run is that the president's party controlled the Senate for nearly this entire span (Stone 2010, 384). The only presidents to submit a nomination to a Senate controlled by the opposite party during this period were Grover Cleveland, a Democrat whose nomination of the conservative Rufus Peckham was approved via voice vote by a Republican-majority Senate in 1895, and Dwight Eisenhower, a Republican who submitted a series of relatively moderate and liberal jurists to a Democratic-majority Senate during his second term (Abraham 2008, 116–17, 205–16).

The era of overwhelmingly successful Supreme Court confirmations came to an end in 1968, when President Lyndon Johnson nominated then-Associate Justice Abe Fortas to succeed the retiring Earl Warren as chief justice. Opposition to Fortas was partly based on ethics concerns and partly on ideology: as Associate Justice, Fortas had been one of the court's more liberal members, and his Senate opponents were largely Republicans and conservative southern Democrats (Kalman 1990, 332). In the five decades since then, five more nominations have failed: two of President Nixon's nominees plus President Reagan's nomination of Bork were rejected by the Senate, President George W. Bush withdrew his nomination of Harriet Miers

Figure 2
Senate Votes on US Supreme Court Nominees, 1789 to 2019
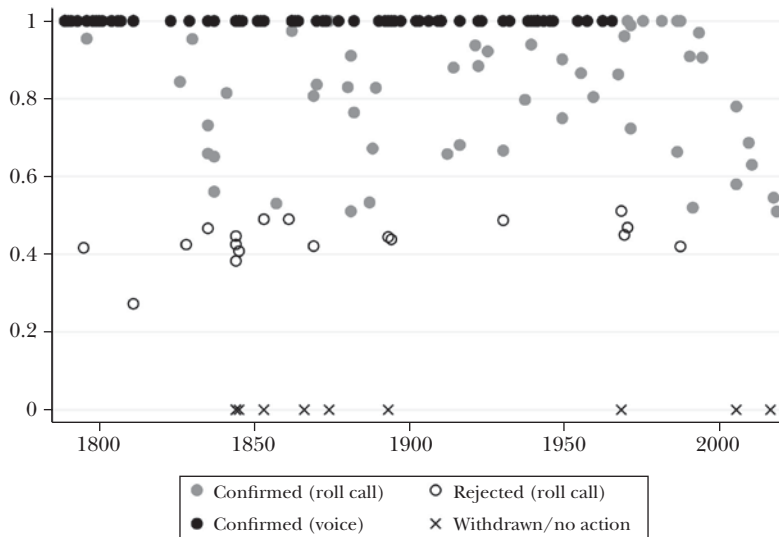
Source: US Supreme Court Justices Database (Spaeth et al. 2019).

under intense questioning of her qualifications, and the Senate never acted on President Obama's nomination of Judge Merrick Garland in 2016.[3] Except in the case of Miers, these failed nominations occurred at a time when the president's party was a minority in the Senate.

The data do reveal some changes in confirmation dynamics. Over the court's history, 73 justices have been confirmed by voice vote, which means that individual senators' votes were not recorded. None of those voice-vote confirmations, though, have occurred in the last half-century. Of those put to a roll call during the court's history, the mean percentage of support in the Senate has been 70 percent. None of the last six nominees who were put to a roll-call vote (including Amy Coney Barrett, whose 2020 confirmation is not reflected in Figure 2) reached that threshold. In that respect, confirmations *are* more contentious (and, because of the decline of voice votes, contentiousness is more transparent).

The normative implications of these changes, though, are ambiguous. The shift from voice votes to roll-call votes may allow the electorate to identify and evaluate the positions of their own senators on important issues. The increasing contentiousness

---

[3]President Reagan announced the nomination of Judge Douglas Ginsburg after Bork's failure in 1987, but Reagan never formally submitted Ginsburg's name to the Senate. Ginsburg withdrew his own name from consideration after he admitted that he had smoked marijuana several times while a professor at Harvard Law School.

of the confirmation process arguably leads to a court whose members' views have been interrogated—and in this way legitimated—by the people's elected representatives (Kagan 1995). Thus the period from 1894 to 1967, in which the Senate routinely rubber-stamped the president's appointees, should not necessarily be considered a golden age.

Increasing Polarization and Party Sorting

Proponents of Supreme Court reform often point to the court's increasing "polarization" (for example, Ornstein 2014; Leonhardt 2018; Epps and Sitaraman 2019; Schwarz 2019). The legal scholar Richard Hasen (2019, 267) writes that "[t]oday no one doubts that the Supreme Court is growing more polarized in its decisionmaking." But political scientists draw a distinction between polarization and party sorting (Fiorina and Levendusky 2006, 53)—and the second seems to be a more accurate description of what has happened at the court.

Polarization occurs when attitudes migrate from the center toward liberal and conservative extremes. Party sorting occurs when Republicans are consistently more conservative than Democrats: that is, when the fit between political ideology and party affiliation becomes tighter. Polarization is a longstanding feature of Supreme Court voting patterns: the justices have coalesced into recognizably liberal and conservative "blocs" as far back as the 1930s (Kalman 1999). What is different about today's court is that Republican appointees are now reliably more conservative and Democratic appointees are reliably more liberal.

Quantitative measures of polarization reflect this change. Martin-Quinn scores are the most commonly used measure of justices' preferences (Spruk and Kovac 2019, 2). Details of Martin-Quinn scores are discussed in more detail in the companion paper in this symposium by Sen and Bonica. In brief, Martin and Quinn (2002) use Markov chain Monte Carlo methods to fit a Bayesian model of ideal points based on votes to affirm or reverse in decided cases. Such scores are now available for all justices from the 1937 term to the 2018 term.

According to the Martin-Quinn measure, the current spread between the most liberal and most conservative justices is not especially wide by historical standards. For example, the 6.4-point gap between the most conservative justice (Clarence Thomas) and the most liberal justice (Sonia Sotomayor) as of the 2018 term is much smaller than the 12.3-point gap between the most conservative justice (William Rehnquist) and the most liberal justice (William Douglas) as of the 1975 term. What is distinct about the last decade is that—for the first time in the period covered by the Martin-Quinn estimates—all of the justices appointed by Democratic presidents have more liberal scores than all of the justices appointed by Republican presidents. Since the 2010 term, no Democratic-appointed justice has stood to the right of any Republican-appointed justice. The increase in party sorting has removed any overlap between these groups.

The ever-stronger linkage between the party of the president who appointed a justice and the way that justice votes on the court has at least three potentially significant consequences. First, some scholars have expressed concern that if Americans

come to view the court as a partisan or ideological institution, public confidence in the court will decline (Bybee 2012, 75–76; Kahan 2011). Empirical support for this prediction, though, is ambiguous. Analyzing survey data, Gibson and Caldeira (2011) find that support for the court is higher among individuals who think that justices' political views *are* relevant to their decisions than among individuals who think those views are irrelevant. In apparent tension with that result, the authors find that support for the court is lower among individuals who agree with the proposition that justices are "politicians in robes." To the extent that these observations allow us to draw any conclusions, they suggest that voters do not demand an entirely apolitical court, but also that a purely political court would risk losing public respect.

Second, party sorting makes it easier for members of the public to identify presidential candidates who will appoint—and Senate candidates who will confirm—justices whose views and values align with the voters' own. This phenomenon strengthens the link between the electorate's preferences and the justices' decisions and thus arguably bolsters the democratic foundation for the court's authority. In an earlier era, party affiliation was a noisy indicator of a politician's confirmation-related views. Even as recently as the Bork nomination, six Republicans crossed party lines to oppose the Republican Reagan's pick. Such defections are now much less frequent. Over the last four presidencies, only three senators have voted against confirmation of a justice appointed by a president of their own party (Republican Senator Lincoln Chafee of Rhode Island, who voted against the confirmation of Samuel Alito in 2006; Republican Senator Lisa Murkowski of Alaska, who voted against the Kavanaugh confirmation in 2018; and Republican Senator Susan Collins of Maine, who voted against the Barrett confirmation in 2020). Increasingly, a voter who wants to see more liberal (or more conservative) justices on the court can advance that interest straightforwardly at the ballot box by voting for Democratic (or Republican) presidential and Senate candidates.

Third, party sorting provides a potential check on ideological voting. Party sorting makes it easier for members of the public to evaluate whether justices are voting in line with their ideological preferences: simply knowing the party of the president who appointed the justice is an increasingly reliable indicator. If justices believe that public confidence in the court hinges on the view that they are not "politicians in robes," and party-line voting contributes to the belief that they are, then anticipated reputational costs may discourage justices from voting their partisan preferences. In other words, the same phenomenon that makes ideological voting by the justices *more transparent* also may make it *less likely*.

Uneven Allocation

A further complaint about the court is that appointments are allocated unevenly across presidential terms (*New York Times* Editorial Board 2018; Greenberg and Levin 2019, 273). Of the last 19 justices added to the court, 15 have been appointed by Republican presidents. The jurisprudential effect of the party disparity in appointments has been somewhat offset by the fact that three of the Republican-appointed justices in the last half century—Blackmun, Stevens, and Souter—ultimately drifted

into line with their Democratic-appointed counterparts. But as conservative activists exert greater influence over judicial appointments under Republican presidents (Zengerle 2018), the probability of a dramatic leftward drift by any of the court's current Republican appointees seems remote.

Insofar as Republicans have enjoyed an advantage in Supreme Court appointments, the advantage does not appear to be structural. President Johnson's strategic blunder in nominating then-Associate Justice Fortas to replace the retiring Earl Warren as chief justice cost Democrats two appointments. The failure of the Fortas nomination meant that the opportunity to appoint the next chief justice fell to President Nixon, and the disgraced Fortas's departure gave Nixon an opportunity to fill Fortas's associate-justice seat. In two other cases (William Douglas and Thurgood Marshall), failing health led a liberal justice to retire shortly before a Democrat would retake the White House. Ruth Bader Ginsburg's decision not to retire when President Obama enjoyed a Democratic majority in the Senate—notwithstanding a chorus of commentary urging her to leave (Kennedy 2011; Tracy 2013; Cohen 2014)—ultimately resulted in one more Republican appointment. Going forward, the uneven allocation of appointments across presidencies could favor either party.

Some scholars view the uneven allocation of appointments across presidencies as ipso facto problematic. For example, Peretti (2005, 440) writes that Jimmy Carter—the only full-term president who never had a Supreme Court vacancy to fill—was "denied [his] right to influence the Court." But no law or widely accepted norm dictates that every president will have such an opportunity. Carter was not denied his right to influence the court because no such right exists.

One still might worry that a disciplined ideological faction could, conceivably, maintain control of the court for decades through strategic retirements and appointments. But the concern here is largely hypothetical. Much to the chagrin of many liberals, Ginsburg and fellow liberal Stephen Breyer did not choose to retire during the Obama presidency, though they were both in their 70s when Obama won a second term with a Senate majority. Conservative septuagenarians Clarence Thomas and Samuel Alito likewise declined to depart when President Trump and a Republican-controlled Senate could have replaced them with much younger like-minded jurists. Although there is some evidence that justices time their retirements so that ideologically sympathetic presidents can choose their successors (Stolzenberg and Lindgren 2010), neither liberals nor conservatives have engaged in this practice persistently enough in order to entrench their faction for decades on end.

Increasing Tenure

A further trend cited by proponents of court reform is the "dramatic increase" in the average tenure of justices over the last several decades (Calabresi and Lindgren 2006, 778). Scholars of the court attribute this trend to three factors. First, rising life expectancy naturally pushes back the date of involuntary departures from the court (Carrington and Crampton 2005, 4; Meador 2005, 115). Second, party sorting may incentivize aging and ailing justices to hang on longer until a president of their preferred party wins the White House (Amar and Calabresi 2002; Calabresi

and Lindgren 2006, 802–03), though by the same token, party sorting may incentivize justices to retire *sooner* so they can be assured that a president who shares their views can name a replacement (for example, see Cray 1997, 497). Third, the job of Supreme Court justice is, by most accounts, becoming cushier (Powe 2005, 101–02). Each justice now has a team of four or five law clerks who review certiorari petitions, prepare the justice for oral arguments, and write first drafts of opinions. The number of cases decided by the court has also diminished dramatically in recent decades. Plus, justices no longer "ride circuit" to hear mostly perfunctory appeals across the country. Modern justices can occupy positions of great power and prestige while leading comfortable day-to-day lives.

In looking more closely at tenure length, though, the starkest contrast between the last half-century and early periods is not a rise in very long tenures. Instead, it is what Crowe and Karpowitz (2007, 425) describe as "the fall of the 'short-term' justice." Over the court's history, 40 justices have served for ten years or less. None of these quick departures occurred in the last half-century.

Several factors have contributed to the fall of the short-term justice. Fewer are dying young, and no justice since Fortas in 1969 has been forced to depart in disgrace. The justices also are now less likely to leave the court to pursue political careers in other branches. Contrast this with Charles Evans Hughes, who left the court in 1916 to accept the Republican nomination for president, and James Byrnes, who would serve as US Secretary of State and Governor of South Carolina in his post-judicial life. Also, with more nominees having previous judicial experience, fewer are arriving at the court and finding that they dislike the work.

The fall of the short-term justice is not obviously an unfortunate trend. The justices may now be more partisan, but the fact that they are less likely to harbor ambitions for elected office bolsters the court's independence from partisan politics. Today's justices have a less diverse range of experiences than their predecessors, but they also arrive at the court with better preparation for the specific tasks of their new job.

One implication of the fall of the short-term justice is that a divergence has arisen between the tenure of justices *at departure* and the mean tenure of *sitting* justices. Figure 3 shows tenure at time of departure for the justices. The absence of short-term justices in recent decades is reflected by the blank bottom-right corner of the figure. Figure 4 shows the mean tenure of *sitting* justices, with the dashed lines showing the overall tenure for each justice. In 2019, the mean tenure of sitting justices (14.2 years) remained well below the historic high of 19.5 years in 1828. Similarly, the median tenure (13 years) was well below the high of 21 (reached in 1833 and again in 2015). The shift in focus from *tenure of departing justices* to *tenure of sitting justices* significantly clouds the perception that tenure is steadily increasing.

Increasing Age

Concerns about the lengthening tenure of justices are closely linked to worries about the justices' increasing age. Calabresi and Lindgren (2006, 782) write that

Figure 3

Tenure at Time of Departure for US Supreme Court Justices, 1789 to Present
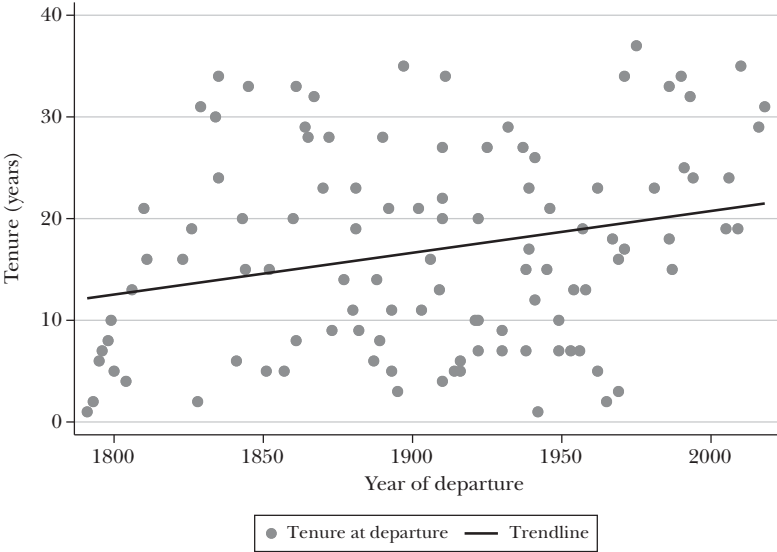
Source: US Supreme Court Justices Database (Epstein et al. 2019).

Figure 4

Tenure of Sitting US Supreme Court Justices, 1789 to 2019
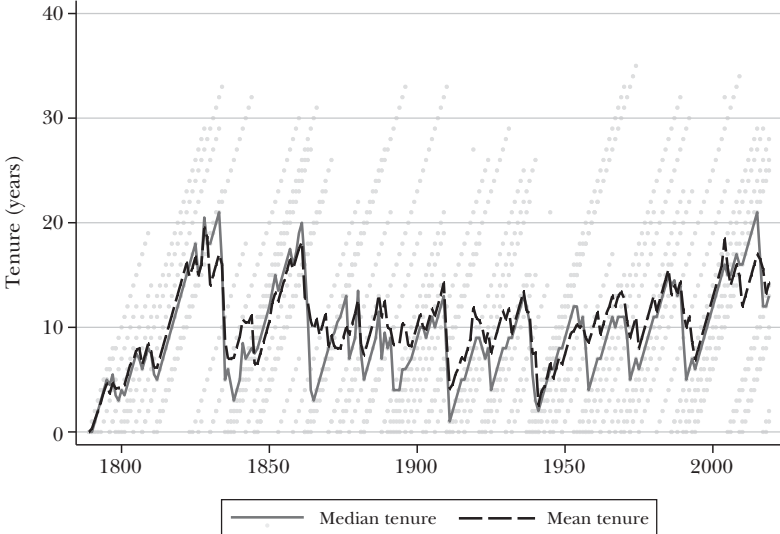
Source: US Supreme Court Justices Database (Epstein et al. 2019).
Note: The lighter dotted diagonal lines show the tenure for each individual Supreme Court justice.

"the average age at which Justices have left office . . . has dramatically increased for those retiring in the past thirty-five years," a pattern illustrated in Figure 6. Of the 16 justices age 80 or older in the court's history, nine have served in the last half-century. Even more dramatically, departures before age 65—once routine—have virtually stopped. Over the court's 231-year history, 34 justices have departed before age 65. None of those departures occurred in the last half-century.

But as with tenure, there is a notable divergence between the age of *departing* justices (Figure 5) and the age of *sitting justices* (Figure 6). The highest mean age of sitting justices was 71.1 years in 1936; the highest median age was 77 in 1985. As of 2019, the mean was 67 and the median was 65. One may think this is too old, but it is not unusually old by historical standards. The divergence between *age at departure* and *age of sitting justices* is a product of the same statistical phenomenon noted above with regard to tenure of the justices. A larger share of younger justices departing shortly after they join the court brings down the mean age at departure considerably, though the replacement of these short-term justices with similarly aged appointees has little effect on the average age of sitting justices. And again, the fall of the short-term justice—the phenomenon fueling the divergence between age at departure and age of sitting justices—does not have obviously negative normative implications.

Term Limits for the Supreme Court?

Proponents of court reform make predictions about the likely effects of institutional change on the selection and behavior of justices. This section focuses on the proposal with the broadest support: 18-year term limits for Supreme Court justices. In particular, I discuss how term limits are likely to affect the judicial confirmation process, polarization and party sorting, and the allocation of appointments across the two parties. The following sections then consider other proposals.

Term Limits and the Confirmation Process

A number of authors argue that 18-year term limits would lower the stakes of the confirmation process and thereby defuse some of its contentiousness (Calabresi and Lindgren 2005, 39–41; Carrington and Crampton 2005, 468). From the perspective of confirmation-process actors, a sympathetic justice has the features of an annuity with a term lasting the justice's tenure. Tenure limitations reduce the annuity's term and thus its value. The prediction is that confirmation-process actors will invest fewer resources in capturing the annuity if its value is lower.

This metaphor suggests that the effect of term limits on the contentiousness of the confirmation process will depend critically on the discount rate that confirmation-process actors assign. These discount rates may be high. For example, politicians are often more focused on the time until the next election rather than 18 years in the future. In addition, confirmation-process actors may recognize that the hot-button issues before the court two decades from now may be very different

Figure 5
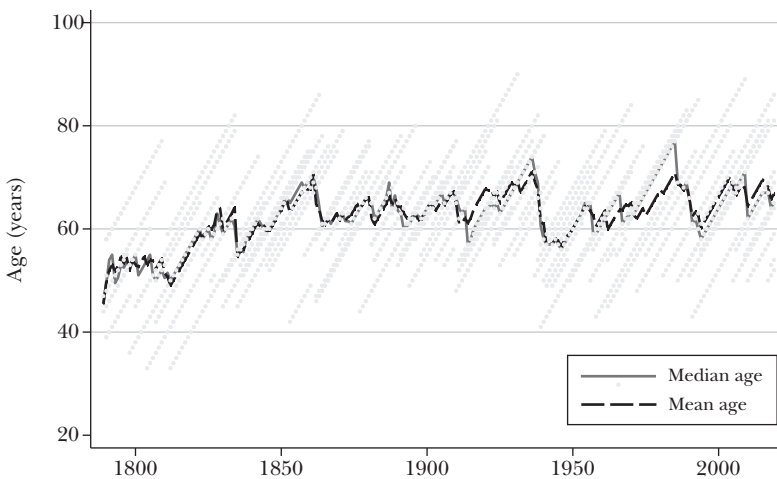**Age at Departure for US Supreme Court Justices, 1789 to 2019**
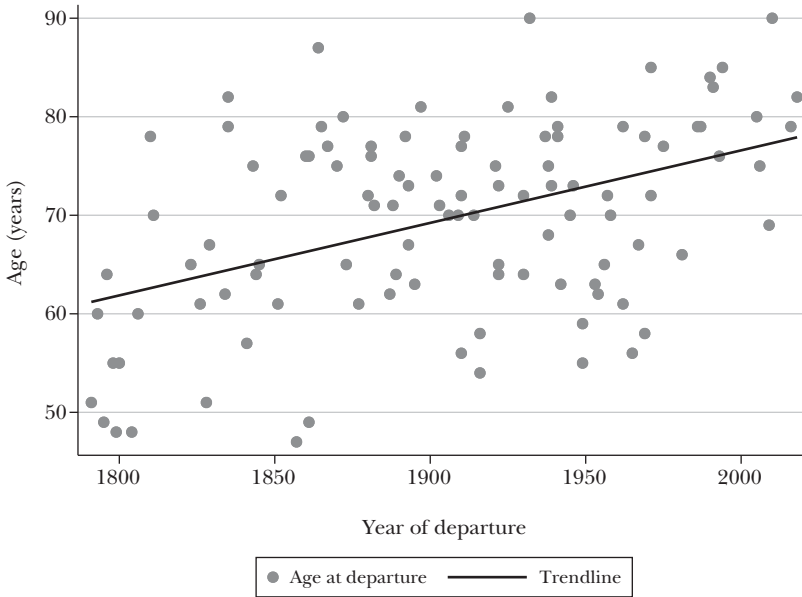
Source: US Supreme Court Justices Database (Epstein et al. 2019).

Figure 6
Age of Sitting US Supreme Court Justices, 1789 to Present

Source: US Supreme Court Justices Database (Epstein et al. 2019).
Note: The lighter dotted diagonal lines show the age for each individual Supreme Court justice during their time in office.

from the issues confronting the court today. Furthermore, the phenomenon of ideological drift (Epstein et al. 2007) suggests that confirmation-process actors should place some additional discount on the future because of the probability that a sympathetic justice today may no longer be a sympathetic justice many years down the road.

As a concrete if hypothetical example, consider the case of Brett Kavanaugh. At the time of his nomination, he was 53 years old, with an additional life expectancy of approximately 29 years (Social Security Administration 2020). How much would it matter to confirmation-process actors if Kavanaugh had only 18 years to serve on the court rather than approximately 29? With a 10 percent discount rate, an 18-year level-payment annuity has approximately 88 percent of the value of a 29-year annuity with the same level payments over a longer time. The key point is that 18-year term limits may have a relatively small effect on confirmation "stakes" if confirmation-process actors discount effects beyond that 18-year horizon at high rates.

Another reason to question the claim that term limits will reduce the contentiousness of the confirmation process is that some justices already appear to time their retirements so as to make the process of replacing them smoother. From the beginning of the Clinton presidency through the end of Trump's term, the same party controlled the White House and the Senate for approximately 16 of 28 years. During that time, six justices retired from the bench. (Three more died in office.) All six retirements occurred when the president's party was in charge of the Senate. Assuming that divided government continues to be a relatively common feature of the American political landscape, 18-year terms would likely mean that more vacancies and nominations would arise at times when the president and Senate are at loggerheads.

Finally, if a reduction in expected judicial tenure has a significant effect on the contentiousness of the confirmation process, older judicial nominees who have fewer years of service remaining should have smoother confirmations. There is no evidence to support such a claim. An age gap between confirmed and rejected/withdrawn nominees has emerged since 1937, but failed nominees tend to be *older* than the confirmed justices (mean for confirmed nominees post-1937: 53.5 years; mean for rejected/withdrawn nominees post-1937: 58.0). This is the exact opposite of what we might expect if confirmations are less contentious when the nominee has fewer years to serve.

To be sure, nominee age may be endogenous to other confirmation-process characteristics. For example, Merrick Garland was 63 at the time of his nomination in 2016, the oldest nominee in 45 years. One reason why President Obama may have chosen Garland is that Obama knew confirmation by a Republican-controlled Senate would be difficult and therefore wanted to lower the stakes by choosing an older nominee. Still, there is little evidence that the number of years a justice is expected to serve will have a significant effect on the contentiousness of the confirmation process.

Term Limits and Polarization

There are at least three reasons to doubt that term limits would have a depolarizing effect on the court and one reason to think that they might.

First, term limits would truncate the process of ideological drift. For the most part, Democratic appointees start out as relatively liberal (both according to everyday judgement and by quantitative measures as judged by the Martin-Quinn scores mentioned earlier), and Republican appointees begin their careers on the court as relatively conservative. But over time, party sorting becomes considerably messier, with some Democratic appointees (like Felix Frankfurter and Byron White) moving in a conservative direction, and some Republican appointees (like Harry Blackmun and John Paul Stevens) growing more liberal. A term-limited court would likely be a better-sorted court, because there would be less time for such drift to occur. If party sorting is a concern, term limits are likely to make it worse.

Second, term limits would potentially undermine incentives for liberals and conservatives on the Supreme Court to strike compromises. When judicial careers are longer, justices know that they may cycle in and out of the majority over the course of their careers. They may expect that cooperative behavior now will be rewarded later and uncooperative behavior will be punished (Axelrod 1984; Wahlbeck, Spriggs, and Maltzman 1999). In addition, a justice who expects to spend many more years on the court will likely care more about public legitimacy concerns, which in turn may encourage justices to build cross-ideological coalitions—or, at least, to adopt voting patterns that do not conform to partisan stereotypes. Term limits, by shortening the shadow of the future for justices, potentially undermine these reciprocity and legitimacy-based incentives.

Third, term limits—without additional limits on justices' post-court careers— would raise the probability of justices pursuing electoral politics after their tenures are over. Insofar as success in partisan politics depends upon clear identification as a liberal or conservative, this dynamic would potentially lead to a court even more clearly sorted along party lines.

To be sure, term limits also could cut in the opposite direction. By limiting the ability of justices to time their departures for periods in which their own party controls both the White House and the Senate, term limits could increase the frequency with which a president would have reason to nominate more moderate justices who would stand a fighting chance of confirmation in a hostile Senate. However, this scenario runs counter to the ostensible objective of reducing the contentiousness of the confirmation process: after all, the *most* contentious conformation fights tend to occur when the president's party holds a minority or very slim majority of Senate seats.

Term Limits and Allocation of Appointments

Supreme Court term-limit proposals generally envision a staggered-term system: a new 18-year term would begin every other year, and any replacement appointed for a justice who leaves before 18 years are up would serve only for the remainder of the term. These interim replacement justices presumably would not be eligible for reappointment, or else they would face potentially perverse incentives to cater to the interests of the sitting president and Senate. Each president

would be guaranteed two appointments per term, and perhaps more if an unfinished term needed to be filled out.[4]

One worry raised by such a proposal, aside from the issues of more frequent confirmation fights, is that by the end of a president's second term, at least four members of the court would be appointees of the sitting president. Justices are significantly more deferential to the administration of the president who appointed them than partisan or ideological variables would predict—a phenomenon that Epstein and Posner (2016) describe as the "loyalty effect." If one of the court's functions is to serve as a check against executive power, then having a large share of the court's members indebted to the sitting president for their seats would potentially weaken that check.

Term Limits and Tenure/Age

Term limits would, mechanically, reduce average tenure on the court. The effect on age is more ambiguous. Life tenure allows older justices to hang onto their seats, but also incentivizes presidents to appoint younger justices who will potentially serve well beyond 18 years. Conversely, term limits might lead to appointing more judges in their late 50s or into their 60s—because there is less to be gained from appointing someone younger.

If the overriding concern is the risk of mental decrepitude on the court, then compressing the age distribution of justices—even if it does not lower the mean or median age—may be salutary. Dementia incidence appears to increase exponentially with age after 65 (for example, Van der Flier and Scheltens 2005). But if the animating concern is age *diversity* on the court, then term limits probably will be counterproductive. And if the goal is a younger court, then age limits (discussed in the next section) almost certainly dominate term limits as the preferred policy tool.

Beyond Term Limits

The loose fit between term limits and the court's perceived problems has led some scholars and commentators to explore alternative reforms. These include age limits as well as judicial lotteries and partisan balance requirements.

Age Limits

In 1954, the Senate approved—by a 58–19 margin—a constitutional amendment that would have required justices to retire at 75, but the House never brought the amendment to a vote (Mazza 2003, 152–153). More recently, some scholars have suggested age limits as an alternative to term limits (for example, Farnsworth 2005, 443–47). On the other side, Calabresi and Lindgren (2006, 817–18) argue that age

---

[4]The staggered-term system would be similar to that used for the Federal Reserve Board of Governors, except with 18-year terms rather than the Fed's 14 years. And unlike Federal Reserve governors, justices who filled out an incomplete term would not be eligible for reappointment to a full term.

limits would be "unfair" because they would "blindly discriminate" against older justices. Roosevelt and Vassilas (2019) say that age limits would be "discriminatory and ineffective."

Such objections to age limits have not stopped 32 states and the District of Columbia from imposing age limits on state judges (Jones 2013). Assuming a limited number of seats on the court, one could make an "unfairness" objection to life tenure as well: the status quo allows an even smaller number of individuals to hog the privilege of serving as a Supreme Court justice rather than sharing the honor and responsibility among a (slightly) larger group. And, of course, age affects the selection of Supreme Court justices already. No one older than age 65 has been nominated to the court in nine decades.

Age limits likely *would* reduce the risk of mental decrepitude on the court, but age is an imperfect proxy for cognitive-impairment risk. Of the 14 twentieth-century justices whom Garrow (2000) identifies as having suffered mental decline on the bench, at least six experienced cognitive impairment before age 75 (Moody, Taft, Murphy, Minton, Whittaker, and Rehnquist), and four of those cases occurred before the justice reached 65.

On other dimensions—contentiousness of the confirmation process and polarizing of justices' voting—age limits are likely to have effects similar to those of term limits. For example, age limits would modestly reduce the stakes of confirmation for each individual judge. Age limits also would cut short the process of ideological drift and likely reduce incentives for coalition-building, thus generating a court that is more clearly sorted along party lines rather than less.

Lotteries

McGinnis (1999) and Epps and Sitaraman (2019) propose that Supreme Court justices be selected by lottery from among sitting federal judges. In the McGinnis version, justices selected by lottery would serve for six months to a year; in the Epps and Sitaraman version, they would serve for two weeks. The idea of rotating groups of Supreme Court justices chosen by lottery may lie so far outside of the Overton window of political possibility that its utility is primarily as a thought experiment. Even so, the thought experiment can shed light on some advantages and disadvantages of the status quo.

One advantage of the status quo (and a disadvantage of a lottery system) is doctrinal stability. In a lottery system, precedents decided on 5–4 votes would likely cycle in and out of case law as the composition of the court swings. Sundby and Sherry (2019) note that term limits suffer from a similar problem, although doctrinal instability when the composition of the court changes every two years is likely less severe than instability when composition changes every two weeks or six months.

Lotteries would likely make lower-court confirmation processes more contentious, though super-high-stakes battles over Supreme Court appointments would cease. Lottery justices also would have weaker incentives to build cross-ideological coalitions because they would not be repeat players. The risk of mental decrepitude

on a lottery court would probably be higher because lower-court judges operate outside the limelight, and mental decrepitude on the lower federal courts is harder to detect. While there is already relatively little diversity among the justices in terms of their professional backgrounds (Kagan is the only member of the current court who was not previously a federal appellate judge), a lottery system would eliminate the possibility of anyone other than a federal appellate judge ascending to the high-court bench.

Partisan Balance

An alternative proposal—which gained national attention when it was suggested by two 2020 Democratic presidential candidates (Lederman 2019)—is the "balanced bench" plan (Epps and Sitaraman 2019). The roster of permanent Supreme Court justices would expand from nine to ten, with five seats explicitly reserved for Democrats and five for Republicans. Every year, the ten permanent justices would appoint (either unanimously or by supermajority vote) a slate of five lower-court judges to serve as temporary justices for one-year nonrenewable terms. If the Republican and Democratic permanent justices could not agree on a slate, the court would cease to operate for that year.

It is difficult to see how the balanced-bench plan would address any of the court's perceived flaws. Rather than making the confirmation process for permanent justices less contentious, it could well have the opposite effect: Senators would debate not only whether a nominee is qualified to serve on the court, but also whether the nominee counts as a "real" Democrat or Republican. Explicitly identifying permanent justices by party would potentially strengthen the public perception of the court as an institution enmeshed in party politics (Sachs 2019, 98). Moreover, if one of the two parties receded (as did the Federalists in the 1820s and the Whigs in the 1850s), the balanced-bench plan could lead to a court quite distant from the political mainstream.

A New Direction: Decoupling Appointments and Departures

Another alternative to structural-reform proposals outlined above—and in my view, a more attractive one—would be to "decouple" the appointments and departures of justices.[5] Decoupling could be implemented as follows. Each president would have the opportunity to appoint two justices at the beginning of each term, regardless of how many vacancies have occurred or will occur. Those justices would join the bench at the beginning of the next presidential term. For example, President Trump, upon taking office in January 2017, would have had the opportunity

---

[5] Peretti (2005, 449–451) proposes that every president be guaranteed one appointment on the court, regardless of whether a vacancy occurs. The plan here goes several steps further—entirely decoupling appointments and departures and creating a new waiting period before a confirmed justice receives her commission. For earlier versions of the decoupling proposal, see Hemel (2018a) and Hemel (2018c).

to make two appointments. Those appointees—if confirmed—would receive their commissions in January 2021. The retirement or death of a justice would have no effect on the number of appointments the sitting president could make. Justices would continue to serve for life. Decoupling thus shares some similarities with the norm among university faculties, where senior members enjoy life tenure but the departure of one does not automatically and immediately trigger the addition of a new member.

The decoupling proposal would result in an equal allocation of appointments across presidential terms, though that is not its principal advantage. It would create new opportunities for compromise when the White House and Senate are at daggers drawn: Because appointments would come in pairs, a Democratic president could resolve an impasse with a Republican Senate (or vice versa) by appointing one liberal and one conservative. It would significantly reduce the risk that a substantial number of justices would be subject to the loyalty effect, since no more than two justices would ever be appointees of the sitting president (and only in that president's second term). The loyalty effect could be eliminated entirely by modifying the plan so that justices receive their commission only after the president who appointed them leaves office (that is, if Trump had been reelected in 2020, none of his appointees would join the court until January 2025).

The plan would likely have a modest effect on the size of the court. The mean tenure of justices who have left the court in the last half-century (since 1970) is 26.4 years, though one might expect tenure to be shorter if appointees had to wait four (or eight) years between confirmation and commission. If justices join the court at a slightly faster rate than they depart, the gradual growth in the court's size would be tolerable. Germany's constitutional court functions with 16 members; the Israeli and Japanese supreme courts have 15 (Turley 2012); the US Tax Court—located just blocks away from the Supreme Court—has 19; and the US Court of Appeals for the Ninth Circuit has 29. A larger court would serve the objective sometimes cited by term-limit proponents of reducing the influence of any individual jurist's idiosyncrasies over the shape of American law. It would also likely lessen the macabre obsession with the health of individual older justices.

The effect of decoupling on retirement incentives is ambiguous. A justice who is ready to leave the bench would no longer have an incentive to delay departure so that a like-minded president can appoint the successor. But as noted above, strategic retirement incentives can lead justices to accelerate as well as delay their departures. Under the decoupling proposal, the departure of a liberal justice—even under a liberal president—would still lead to the loss of one liberal vote. Decoupling would thus reduce the incentive for strategic delay of retirement but also reduce the incentive for strategic acceleration of retirement.

The decoupling approach is at this stage more of a thought experiment than a live agenda item. This cursory sketch certainly falls short of a definitive proof that decoupling would be welfare-enhancing. Further analysis—including simulations to evaluate the likely effect of decoupling on court size over time—remains for future

work. But unlike the other approaches, which would likely require constitutional amendments, the decoupling approach could be implemented via ordinary legislation. Congress can change the size of the court without a constitutional amendment, and Congress has power to prescribe the procedure for new justices to receive their commissions.

Conclusion

The analysis here does not lend itself easily to an urgent call for legislative action—if anything, it suggests that the court's "crisis" is more imagined than real. To be sure, confidence in the court is down from historic highs, the confirmation process is becoming more contentious, justices' votes are more easily predictable based on the party of their appointing president, and average tenures are growing longer. But the normative implications of all those trends are ambiguous. For a moment in the fall of 2020, it looked like the nation might be gearing up for a partisan battle over court reform, with Democrats potentially seeking to add additional justices so as to reverse the current court's conservative lean. But in the aftermath of the November 2020 election, that possibility now seems distant. The court's conservative majority appears to have raised its standing among liberals and moderates by quickly dispatching a Republican-backed lawsuit seeking to undo Democrat Joe Biden's 2020 electoral victory (Greenhouse 2020). While it was not yet certain as of this writing which party will control the Senate for the next two years, it is highly unlikely that Democrats will have the necessary votes to push a court-packing plan through the upper chamber (Stracqualursi 2020).

None of this, though, is intended to cut short conversations about court reform. To invert Paul Romer's famous quip (reported by Friedman 2004), the absence of a crisis is a terrible thing to waste. Periods of non-emergency—instead of inviting complacency—can offer opportunities for unrushed evaluation of institutional reforms that reduce the risk of a true crisis. In that spirit, creative court-reform proposals allow us to evaluate aspects of the status quo that we often take for granted. Considering such proposals can strengthen our understanding of the court as it is and to see new possibilities for what the court could become.

References

Abraham, Henry J. 2008. *Justices, Presidents, and Senators: A History of the U.S. Supreme Court Appointments from Washington to Bush II*. 5th ed. Plymouth, UK: Rowman & Littlefield Publishers.

Ackerman, Bruce. 2018. "Op-Ed: Trust in the Justices of the Supreme Court Is Waning. Here Are Three Ways to Fortify the Court." *Los Angeles Times*, December 20. https://www.latimes.com/opinion/op-ed/la-oe-ackerman-supreme-court-reconstruction-20181220-story.html.

Amar, Akhil Reed, and Steven G. Calabresi. 2002. "Term Limits for the High Court." *Washington Post*, August 9. https://www.washingtonpost.com/archive/opinions/2002/08/09/term-limits-for-the-high-court/646134cd-8e13-4166-9474-5f53be633d7c.

American Academy of Arts and Sciences. 2020. *Reinventing American Democracy for the 21st Century*. Cambridge, MA: Commission on the Practice of Democratic Citizenship.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Barnett, Randy E., and Josh Blackman. 2016. "Restoring the Lost Confirmation." *National Affairs,* Fall 2016. https://www.nationalaffairs.com/publications/detail/restoring-the-lost-confirmation.

Bickel, Alexander M. 1962. *The Least Dangerous Branch: The Supreme Court at the Bar of Politics*. Indianapolis: Bobbs-Merrill Company.

Bybee, Keith J. 2012. "Paying Attention to What Judges Say: New Directions in the Study of Judicial Decisionmaking." *Annual Review of Law and Social Science* 8: 69–84.

Calabresi, Steven G., and James Lindgren. 2005. "Supreme Gerontocracy." *Wall Street Journal*, April 8. https://www.wsj.com/articles/SB111292087188301557.

Calabresi, Steven G., and James Lindgren. 2006. "Term Limits for the Supreme Court: Life Tenure Reconsidered." *Harvard Journal of Law and Public Policy* 29 (3): 770–877.

Carrington, Paul D., and Roger C. Crampton. 2005. "The Supreme Court Renewal Act: A Return to Basic Principles." In *Reforming the Court: Term Limits for Supreme Court Justices,* edited by Paul D. Carrington and Roger C. Cramton, 467–72. Durham, NC: Carolina Academic Press.

Clark, Tom S. 2011. *The Limits of Judicial Independence*. Cambridge: Cambridge University Press.

Cohen, Michael. 2014. "Ruth Bader Ginsburg Should Do All Liberals a Favor and Retire Now." *The Guardian*, February 14. https://www.theguardian.com/commentisfree/2014/feb/14/ruth-bader-ginsburg-retire-liberal-judge.

Cray, Ed. 1997. *Chief Justice: A Biography of Earl Warren*. New York: Simon and Schuster.

Crowe, Justin, and Christopher F. Karpowitz. 2007. "Where Have You Gone, Sherman Minton? The Decline of the Short-Term Supreme Court Justice." *Perspectives on Politics* 5 (3): 425–45.

Dahl, Robert A. 1957. "Decision-Making in a Democracy: The Supreme Court as a National Policy-Maker." *Journal of Public Law* 6 (2): 279–95.

Doerfler, Ryan, and Samuel Moyn. Forthcoming. "Democratizing the Supreme Court." *California Law Review.*

Epps, Daniel, and Ganesh Sitaraman. 2019. "How to Save the Supreme Court." *Yale Law Journal* 129: 148–206.

Epstein, Lee, Andrew D. Martin, Kevin M. Quinn, and Jeffrey A. Segal. 2007. "Ideological Drift Among Supreme Court Justices: Who, When, and How Important?" *Northwestern University Law Review* 101 (4): 1483–1542.

Epstein, Lee, and Eric A. Posner. 2016. "Supreme Court Justices' Loyalty to the President." *Journal of Legal Studies* 45 (2): 401–36.

Epstein, Lee, and Eric Posner. 2018. "If the Supreme Court Is Nakedly Political, How Can It Be Just?" New York Times, July 9. https://www.nytimes.com/2018/07/09/opinion/supreme-court-nominee-trump.html.

Farnsworth, Ward. 2005. "The Regulation of Turnover on the Supreme Court." *University of Illinois Law Review* 2: 407–53.

Fiorina, Morris P., and Matthew S. Levendusky. 2006. "Disconnected: The Political Class versus the People - Rejoinder." In *Red and Blue Nation?: Characteristics and Causes of America's Polarized Politics*, edited by Pietro S. Nivola and David W. Brady, 95–110. Washington, DC: Brookings Institution Press and Stanford: Hoover Institution Press.

Friedman, Thomas L. 2004. "Kicking Over the Chessboard." *New York Times*, April 18. https://www.nytimes.com/2004/04/18/opinion/kicking-over-the-chessboard.html.

Gallup. 1973–2019. "Supreme Court." Gallup. https://news.gallup.com/poll/4732/supreme-court.aspx.

Garrow, David J. 2000. "Mental Decrepitude on the U.S. Supreme Court: The Historical Case for a 28th Amendment." *University of Chicago Law Review* 67 (4): 995–1087.

Garrow, David J. 2005. "Protecting and Enhancing the U.S. Supreme Court." In *Reforming the Court: Term Limits for Supreme Court Justices,* edited by Paul D. Carrington and Roger C. Cramton, 271–89. Durham, NC: Carolina Academic Press.

Gerhardt, Michael J. 2002. "Supreme Court Selection as War." *Drake Law Review* 50 (3): 393–410.

Gibson, James L., and Gregory A. Caldeira. 2011. "Has Legal Realism Damaged the Legitimacy of the U.S. Supreme Court?" *Law & Society Review* 45 (1): 195–219.

Graglia, Lino A. 1994. "Roundtable: The Supreme Court as a Political Institution: Do Judges Have a Policy-Making Role in the American System of Government?" *Harvard Journal of Law and Public Policy* 17 (1): 119–30.

Greenberg, Leah, and Ezra Levin. 2019. *We Are Indivisible: A Blueprint for Democracy After Trump.* New York: Atria.

Greenhouse, Linda. 2020. "Thank the Supreme Court, for Now." *New York Times*, December 17. https://www.nytimes.com/2020/12/17/opinion/supreme-court-trump-election.html.

GSS Data Explorer. 1973–2018. "Confidence in Supreme Court." GSS Data Explorer. https://gssdataexplorer.norc.org/trends/Politics?measure=conjudge.

Hamilton, Alexander. 1788. "The Federalist Papers: No. 78." https://avalon.law.yale.edu/18th_century/fed78.asp.

Hasen, Richard L. 2019. "Polarization and the Judiciary." *Annual Review of Political Science* 22: 261–76.

Hemel, Daniel. 2018a. "Instead of Term Limits for Justices, How About Term Delays?" *Medium: Whatever Source Derived*, October 5. https://medium.com/whatever-source-derived/instead-of-term-limits-for-justices-how-about-term-delays-66c120910257.

Hemel, Daniel. 2018b. "The Supreme Court Is a Bulwark of Democracy." *Slate*, October 11. https://slate.com/news-and-politics/2018/10/supreme-court-good-history.html.

Hemel, Daniel. 2018c. "Why Supreme Court Term Limits Won't Work and What to Do Instead." *Slate*, October 11. https://slate.com/news-and-politics/2018/10/supreme-court-court-packing-bad.html.

Johnsen, Dawn E. 2020. "The Supreme Court's Legitimacy Crisis and Constitutional Democracy's Future." Unpublished.

Jones, Ashby. 2013. "A New Lease for Old Judges." *Wall Street Journal*, March 5. https://www.wsj.com/articles/SB10001424127887323699704578328214137916682.

Kagan, Elena. 1995. "Confirmation Messes, Old and New." *University of Chicago Law Review* 62 (2): 919–42.

Kahan, Dan M. 2011. "Neutral Principles, Motivated Cognition, and Some Problems for Constitutional Law." *Harvard Law Review* 125 (1): 1–77.

Kalman, Laura. 1990. *Abe Fortas: A Biography.* New Haven, CT: Yale University Press.

Kalman, Laura. 1999. "Law, Politics, and the New Deal(s)." *Yale Law Journal* 108 (8): 2165–2213.

Kennedy, Randall. 2011. "The Case for Early Retirement." *The New Republic*, April 28. https://newrepublic.com/article/87543/ginsburg-breyer-resign-supereme-court.

Lederman, Josh. 2019. "Inside Pete Buttigieg's Plan to Overhaul the Supreme Court." *NBC News*, June 3. https://www.nbcnews.com/politics/2020-election/inside-pete-buttigieg-s-plan-overhaul-supreme-court-n1012491.

Leonhardt, David. 2018. "The Supreme Court Needs Term Limits." *New York Times*, September 18. https://www.nytimes.com/2018/09/18/opinion/columnists/brett-kavanaugh-supreme-court-term-limits.html.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10 (2): 134–53.

Mazza, Michael J. 2003. "A New Look at an Old Debate: Life Tenure and the Article III Judge." *Gonzaga Law Review* 39 (1): 131–62.

McCarthy, Justin. 2014. "Americans Losing Confidence in All Branches of U.S. Gov't." *Gallup*, June 30. https://news.gallup.com/poll/171992/americans-losing-confidence-branches-gov.aspx.

McGinnis, John O. 1999. "Justice Without Justices." *Constitutional Commentary* 16 (3): 541–48

Meador, Daniel J. 2005. "Thinking About Age and Supreme Court Tenure." In *Reforming the Court: Term Limits for Supreme Court Justices,* edited by Paul D. Carrington and Roger C. Cramton, 115–26. Durham, NC: Carolina Academic Press.

Millhiser, Ian. 2019. "Let's Think About Court-Packing." *Democracy: A Journal of Ideas* Winter 2019 (51).

New York Times Editorial Board. 2017. "The Supreme Court as Partisan Tool." New York Times, April 4.

https://www.nytimes.com/2017/04/04/opinion/the-supreme-court-as-partisan-tool.html.

New York Times Editorial Board. 2018. "How Brett Kavanaugh Failed." *New York Times*, October 5. https://www.nytimes.com/2018/10/04/opinion/brett-kavanaugh-supreme-court-vote.html.

Nocera, Joe. 2011. "The Ugliness Started with Bork." *New York Times*, October 21. https://www.nytimes.com/2011/10/22/opinion/nocera-the-ugliness-all-started-with-bork.html.

Ornstein, Norm. 2014. "Why the Supreme Court Needs Term Limits." *The Atlantic*, May 22. https://www.theatlantic.com/politics/archive/2014/05/its-time-for-term-limits-for-the-supreme-court/371415.

Peretti, Terri L. 2005. "Promoting Equity in the Distribution of Supreme Court Appointments." In *Reforming the Court: Term Limits for Supreme Court Justices*, edited by Paul D. Carrington and Roger C. Cramton, 435–54. Durham, NC: Carolina Academic Press.

Powe, L.A., Jr. 2005. "'Marble Palace, We've Got a Problem—with You." In *Reforming the Court: Term Limits for Supreme Court Justices,* edited by Paul D. Carrington and Roger C. Cramton, 99–114. Durham, NC: Carolina Academic Press.

Rehnquist, William H. 1959. "The Making of a Supreme Court Justice." *Harvard Law Record*. October 8, 7–10.

Roosevelt, Franklin D. 1937. "Fireside Chat on the Reorganization of the Judiciary." Speech, White House, March 9, 1937. http://docs.fdrlibrary.marist.edu/030937.html.

Roosevelt, Kermit, and Ruth-Helen Vassilas. 2019. "Supreme Court Justices Should Have Term Limits." *CNN*, September 30. https://www.cnn.com/2019/09/30/opinions/supreme-court-term-limits-law-roosevelt-vassilas/index.html.

Sachs, Stephen E. 2019. "Supreme Court as Superweapon: A Response to Epps & Sitaraman." *Yale Law Journal Forum* 129: 93–107.

Seib, Gerald F. 2018. "Kavanaugh Controversy Adds to Public's Crisis of Confidence." *Wall Street Journal*, September 24. https://www.wsj.com/articles/kavanaugh-controversy-adds-to-publics-crisis-of-confidence-1537799881.

Sheldon, Charles H. 1970. *The Supreme Court: Politicians in Robes*. Glencoe, IL: Glencoe Press.

Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2019. "LEGACY Database: SCDB Legacy 06." The Supreme Court Database. http://scdb.wustl.edu/data.php?s=6.

Spruk, Rok, and Mitja Kovac. 2019. "Replicating and Extending Martin-Quinn Scores." *International Review of Law and Economics* 60.

Stracqualursi, Veronica. 2020. "Centrist Democrat Says He Won't Back Expanding Supreme Court." *CNN*, November 10. https://www.cnn.com/2020/11/10/politics/joe-manchin-supreme-court-packing-cnntv/index.html.

Steamer, Robert J. 1971. *The Supreme Court in Crisis: A History of Conflict*. Amherst, MA: University of Massachusetts Press.

Stephenson, Matthew C. 2004. "Court of Public Opinion: Government Accountability and Judicial Independence." *Journal of Law, Economics, and Organization* 20 (2): 379–99.

Stolzenberg, Ross M., and James Lindgren. 2010. "Retirement and Death in Office of U.S. Supreme Court Justices." *Demography* 47 (2): 269–98.

Stone, Geoffrey R. 2010. "Understanding Supreme Court Confirmations." *Supreme Court Review* 2010 (2010): 381–467.

Thomson-DeVeaux, Amelia, and Oliver Roeder. 2018. "Is the Supreme Court Facing a Legitimacy Crisis?" *FiveThirtyEight*, October 1. https://fivethirtyeight.com/features/is-the-supreme-court-facing-a-legitimacy-crisis.

Tomasky, Michael. 2018. "The Supreme Court's Legitimacy Crisis." *New York Times*, October 5. https://www.nytimes.com/2018/10/05/opinion/supreme-courts-legitimacy-crisis.html.

Tracy, Mark. 2013. "Justice Ginsburg Is Wrong: She Should Step Down." *The New Republic*, December 18. https://newrepublic.com/article/115973/ruth-bader-ginsburg-should-retire-supreme-court.

Turley, Jonathan. 2012. "The Fate of Health Care Shouldn't Come Down to 9 Justices. Try 19." *Washington Post*, June 22. https://www.washingtonpost.com/opinions/the-fate-of-health-care-shouldnt-come-down-to-9-justices-try-19/2012/06/22/gJQAv0gpvV_story.html.

Van der Flier, W.M., and P. Scheltens. 2005. "Epidemiology and Risk Factors of Dementia." *Journal of Neurology, Neurosurgery, and Psychiatry*, 76 (5): 2–7.

Wahlbeck, Paul J., James F. Spriggs II, and Forrest Maltzman. 1999. "The Politics of Dissents and Concurrences on the U.S. Supreme Court." *American Politics Research* 27 (4): 488–514.

Waldman, Paul. 2018. "Yes, the Supreme Court is Facing a Legitimacy Crisis. And We Know Exactly Whose

Fault It Is." *Washington Post*, September 24. https://www.washingtonpost.com/blogs/plum-line/wp/2018/09/24/yes-the-supreme-court-is-facing-a-legitimacy-crisis-and-we-know-exactly-whose-fault-it-is.

Walsh, Kevin C. 2016. "The Court After Scalia." *First Things*, October, 2016. https://www.firstthings.com/article/2016/10/the-court-after-scalia.

Zengerle, Jason. 2018. "How the Trump Administration Is Remaking the Courts." *New York Times Magazine*, August 22. https://www.nytimes.com/2018/08/22/magazine/trump-remaking-courts-judiciary.html.

Staffing the Higher Education Classroom

David Figlio and Morton Schapiro

There are roughly 4,000 colleges and universities in the United States, including 1,600 publics, 1,700 private non-profits, and 700 private for-profits. Together, they enroll around 17 million undergraduates and another three million graduate and professional school students. Teaching is a primary mission at all of these institutions; for the vast majority of these schools, it is their singular mission. Therefore, a centrally important decision that colleges and universities face is making informed choices about how to staff their classrooms.

This decision has a number of important dimensions. Investing in a tenure-track faculty member could mean a commitment that spans several decades, and in the United States—since the 1994 change in federal law that abolished manda-tory retirement at a certain age—it is a commitment with an uncertain end date. A lack of ability to forecast future demand for certain subject areas may encourage institutions to hire more contingent faculty and fewer tenure-line faculty, a deci-sion that might affect instructional quality. Institutions also differ substantially in the degree to which the demographic characteristics of their faculty match those of their student bodies, and this might also affect students' outcomes. Moreover, while we know that teaching quality is massively heterogenous across the cate-gories of potential instructors, we also know that it is extremely challenging to

■ *David Figlio is Orrington Lunt Professor and Dean of the School of Education and Social Policy and Morton Schapiro is President and Professor of Economics, both at Northwestern University, Evanston, Illinois. Their email addresses are figlio@northwestern.edu and nu-president@northwestern.edu.*

measure teaching effectiveness. As we discuss below, student evaluations of classroom teaching are subject to considerable bias and room for manipulation.

One objective of this paper is to explore different ways to evaluate what students experience in the undergraduate classroom. A central question is whether there is a more objective way to evaluate a teacher than through the ubiquitous course evaluation survey. We develop an alternative approach that seeks to measure what students actually learn. Once we do that, we address a secondary question: whether especially charismatic teachers—defined as those who are particularly adept at inspiring students to major in their disciplines—are better (or worse) at having their students learn than their less captivating peers.

The relatively small number of institutions that pride themselves not just on their teaching output but also on the scholarly contributions of their faculty face a multi-tasking problem of the first order. Too much emphasis on research may leave their students at all levels—undergraduate, graduate, and professional—to fend for themselves. While there appears to be important peer effects among students attending selective institutions so that a good deal of learning occurs outside the classroom, one would hope that transformative learning is also taking place within the classroom. On the other hand, too much focus on teaching might mean that the self-described "research" university is that in name only.[1]

While we don't pretend to offer a magic formula for determining the ideal distribution of resources and focus between teaching and research, we present some clues about the relationship between the two. One question we explore, at least tentatively, is whether the most accomplished teachers sacrifice in terms of their scholarship, and, alternatively, whether relatively poor teachers make up for it in research excellence. Throughout this paper, we review the literature and then offer empirical results based on students and faculty at our own university, Northwestern. Whether the findings are generalizable to other schools is an open question that we hope others will be encouraged to explore.

[1] In addition to teaching and research, many prominent universities are also in the entertainment business. Their plays and concerts might occasionally attract attention beyond their local communities, but if they participate in "big time" sports, national and even global attention are virtually guaranteed. For example, when Northwestern played Ohio State for the Big Ten conference football championship in December 2018, it was viewed in 8.7 million households. While it is commonplace for major public research universities to participate at the highest level of athletics, it is unusual for their private counterparts to do the same. Of the 27 private universities that are members of the prestigious Association of American Universities, only five are among the 65 members of the athletic "power conferences": Stanford and USC in the Pac 12, Vanderbilt in the SEC, Duke in the ACC, and Northwestern in the Big Ten. Most of the rest are either in the Ivy League or compete in the NCAA Division III (which prohibits athletic scholarships).

Measuring Teaching Effectiveness

While it is commonplace to have students fill out evaluation forms at the conclusion of the semester or quarter, a large literature shows that the results appear to be biased by gender, race, and nationality. White American men are often given higher ratings than others and, without objective measures of student learning, it is impossible to evaluate whether those ratings are actually "earned." Mengel, Sauermann, and Zoelitz (2019) study a context at Maastricht University in which students are randomly assigned to male or female section instructors, and show that students systematically rate female instructors worse—a pattern driven by evaluation of male students. Despite these biases, there may be some value in student evaluations: Hoffmann and Oreopoulos (2009a), for example, provide evidence using administrative data from a large Canadian university that higher student perception of professor quality is associated not just with a reduction in course-dropping, but also in increased performance in that class. But this signal may be a noisy one: Deslauriers et al. (2019) examine Harvard undergraduates enrolled in large introductory physics classes and find that "attempts to evaluate instruction based on students' perceptions of learning" could be very misleading, misstating the actual learning (as determined by multiple choice tests) that takes place. They warn that "a superstar lecturer could create such a positive feeling of learning that students would choose those lectures over active learning" despite the fact that, according to their analysis, students in "active" classrooms actually learned more.[2]

Concerns about bias have led the American Sociological Association (2019) to caution against over-reliance on student evaluations of teaching, pointing out that "a growing body of evidence suggests that their use in personnel decisions is problematic" given that they "are weakly related to other measures of teaching effectiveness and student learning" and that they "have been found to be biased against women and people of color." The ASA suggests that "student feedback should not be used alone as a measure of teaching quality. If it is used in faculty evaluation processes, it should be considered as part of a holistic assessment of teaching effectiveness." Seventeen other scholarly associations, including the American Anthropological Association, the American Historical Association, and the American Political Science Association, have endorsed the ASA report (for discussion in the education press, see Flaherty 2019b; Supiano 2019). For all of these reasons, it's clear that student evaluations are far from excellent summative assessments of instructor effectiveness.

An alternative method to evaluate teaching is to examine whether faculty induce student interest in a subject, as measured by the likelihood that students take additional courses in that subject (Bettinger and Long 2010; Hoffmann

---

[2]Some other studies of student evaluations of instructional quality include Bavishi, Madera, and Hebl (2010), Beleche, Fairris, and Marks (2012), Boring (2017), Boring, Ottoboni, and Stark (2016), Braga, Paccagnella, and Pellizzari (2014), Rivera and Tilcsik (2019), Stark and Freishtat (2014), and Stroebe (2016). Flaherty (2016) provides a useful summary of some of the literature.

and Oreopoulos 2009a). While it is certainly reasonable to assume that teachers who inspire students to continue studying their discipline are especially good at conveying knowledge, that need not be the case. Charisma might not translate into teaching effectiveness. An instructor who increases the likelihood that students take additional classes in a subject may or may not produce "deep learning," in the words of Carrell and West (2010). One could imagine that charismatic instructors could even inhibit deep learning if they induce students into thinking that their *enjoyment* of a subject implies *mastery* of that subject.

Can we actually measure learning? One way to get a handle on this question, at least in part, would be to examine (with appropriate statistical controls) how students perform in subsequent classes. If faculty members are especially effective at teaching, say, introductory microeconomics, we would expect their students to perform unexpectedly well in intermediate micro and beyond. Perhaps those faculty might not be particularly good at generating majors, and they might not have the best teaching evaluations, but being a "popular" teacher and being a "successful" one are not necessarily the same thing. Carrell and West (2010) pioneered a measure of professor quality based on academic performance in subsequent courses within the same subject, a method continued by Figlio, Schapiro, and Soter (2015) and Ran and Xu (2019).

We are curious about whether a relationship exists between charismatic or inspirational teaching and the "deep learning" reflected by value added to subsequent classes. As an exploratory first step, in this paper we make use of registrar data on all Northwestern University freshmen who entered between fall 2001 and fall 2008, a total of 15,662 students, and on the faculty who taught them during their first quarter at Northwestern. Specifically, 170 tenured faculty members taught at least 20 first-quarter students across the eight annual cohorts. (We focus on tenured faculty because we also want to compare teaching effectiveness measures with research quality measures.) While Northwestern is an admittedly unrepresentative institution, our analysis should at least offer some clues regarding where and how to look at other colleges and universities.

We construct two measures of teaching quality. First, we look at the ability of professors to convert students into majors. A talented chemistry teacher in an introductory course may lead a declared chemistry major to keep that major, or convince an undeclared student to major in chemistry, or lead a student majoring in physics or economics to become a chemistry major. The ability to attract students to the major presumably reflects one dimension of teaching excellence. Second, we consider a "deep learning" measure described in considerable detail in Figlio, Schapiro, and Soter (2015). We look first at the likelihood of taking additional courses in a subject area, and then measure the deviation in the grade received by a student in follow-up courses in that subject. A successful undergraduate teacher in, say, introductory psychology, not only induces students to take additional psychology courses but leads those students to do unexpectedly well in those additional classes (based on what we would have predicted given their standardized test scores, other grades, grading standards in that field, and so on). In our 2015 paper, we lay out the

statistical techniques employed in controlling for course and student impacts other than those linked directly to the teaching effectiveness of the original professor.[3]

In essence, therefore, our two measures of teaching quality reflect, in the first case, inspiration, as indicated by the ability to convert students to a subject that they had not previously planned on studying in-depth and in the second case, value added (or "deep learning"), that is transferrable to subsequent classes in the subject.

There is considerable range in these variables: The 25th percentile tenured professor ranked according to the "inspiration" measure converted no students to the major in our data, while the 75th percentile tenured professor converted 25 percent of students to the major, and the 90th percentile tenured professor converted 46 percent of students to the major. When ranked according to "deep learning," the difference between the 90th and 10th percentile tenured professor is over one-tenth of a letter grade, and the difference between the 75th and 25th percentile tenured professor is over one-twentieth of a letter grade.

Are More Charismatic Teachers Also Better Teachers?

How are these two measures of teaching success related? Are charismatic professors who attract students to the major also unusually proficient at imparting knowledge that improves performance in future courses? The study by Deslauriers et al. (2019) of Harvard undergraduates in a physics class suggests they might actually be worse. This concern was expressed in an account of their work that appeared in the education press, with the provocative title: "Study: How Smooth-talking Professors Can Lull Students Into Thinking They've Learned More Than They Have" (Flaherty 2019a).

Based on our analysis of the 170 tenured Northwestern faculty, we find that those who are most successful in inspiring students to become majors in their subject are not any more distinguished in facilitating "deep learning" than their less charismatic counterparts: the correlation between the two measures of teaching quality is virtually zero (the correlation coefficient is trivial, at −0.025), suggesting that these two dimensions of teaching quality are essentially unrelated. That is, teachers who leave scores of majors in their wake appear to be no better or worse at teaching the material needed for future courses than their less inspiring counterparts; teachers who are exceptional at conveying course material are no more likely than others to

---

[3]While this measure of teaching excellence goes far beyond course evaluations, it is of course an imperfect measure of teaching success. An especially poor teacher may lead students to switch from that subject entirely—hence we would have no information on how their students would have done in subsequent courses. In addition, our dataset centers on success in teaching courses open to first-quarter freshmen, most of which are introductory courses. Perhaps some faculty are poor teachers for an intro course aimed at first-year undergraduates but do a better job teaching senior seminars for undergraduates or graduate courses. That said, this measure at least provides a (selected) analog to the "value added" measurements of K–12 instruction.

inspire students to take more courses in the subject area. We would love to see if this result would be replicated at other institutions.

Of course, attracting majors may be a reward in itself, at least at the departmental level, should more majors translate into additional faculty slots and more funding at the departmental level. At the institutional level, however, choice of major is basically a zero-sum game, as adding one major typically means losing a different one (unless a student decides to double or triple major). There may also be a zero-sum aspect of our measure of "deep learning" as well: If a faculty member inspires extra studying in one subject, does that reduce student attention to other subjects? Of course, the reverse might be true as well, if a particularly effective instructor helps foster deeper learning in other subjects. Unpacking this question could be a fruitful avenue for future research.

Is There a Tradeoff between Research and Teaching Excellence?

Our two measures of teaching excellence described above allow us to address empirically whether those faculty who do particularly well in the classroom pay a price in terms of their scholarship. But first, a word about our measures of scholarship.

While measuring scholarly excellence is somewhat less contentious than evaluating teaching effectiveness, it is nonetheless fraught. In some fields, well-received books indicate success, in others it is in-person performances, and still others it is highly-cited articles or the awarding of grants. How might one recognize stellar scholarship across chemistry and theater, engineering and music, economics and English, mathematics and anthropology?

As described in Figlio and Schapiro (2017), we employ two very different scholarship measures. For each year since 1988, Northwestern has had a faculty committee comprised of distinguished professors from a wide range of disciplines who review the scholarly accomplishments of the faculty in the previous academic year, and select a subset of faculty to be honored for their research excellence at an annual dinner. Reasons for being honored include: recognition by the leading scholarly organizations in their fields such as being elected into the National Academy of Sciences, Engineering, or Medicine, or into the American Academy of Arts and Sciences; receipt of prestigious fellowships such as those given by the MacArthur and Guggenheim Foundations; winning major research awards from top scholarly associations, and comparable achievements. The committee does also make certain to identify appropriate awards at all career stages, so more junior scholars winning a "rising star" award from a top professional society or a National Science Foundation CAREER grant would also make the cut for inclusion. Using this measure, 57 percent of the 170 tenured faculty in our data set have been recognized at least once as an extraordinary scholar.

As an alternative measure, we followed a more traditional approach and constructed for each faculty member a within-department indicator of how influential that person's scholarly work has been. Specifically, we compute a scholar's

h-index—an indicator that simultaneously measures frequency of publications and the scholarly influence of those publications, thereby capturing aspects of both a researcher's productivity and the significance of that person's work. The h-index is defined as the number of articles or papers with at least *h* citations: For example, a professor with 20 papers that had been cited at least 20 times would have an h-index of 20. We adjust this h-index so that we are comparing scholars only to colleagues within their own department at Northwestern; specifically, we standardize h-indices within each academic department of Northwestern. We carry out this within-Northwestern-department adjustment to take into account the fact that publication and citation norms vary dramatically across disciplines, and because some Northwestern departments are more eminent than others. Nonetheless, there exists very substantial within-department variation in tenured faculty h-indices—among tenured Northwestern faculty, the typical 75th percentile department-adjusted h-index is 1.5 standard deviations higher than the typical 25th percentile department-adjusted h-index, and the typical 90th percentile department-adjusted h-index is 2.4 standard deviations higher than the typical 25th percentile department-adjusted h-index—as one would expect given that this measure of scholarly work has some bias toward older and more established faculty members. (The first measure may also be modestly biased toward more established faculty members but not nearly as much as the h-index is.)
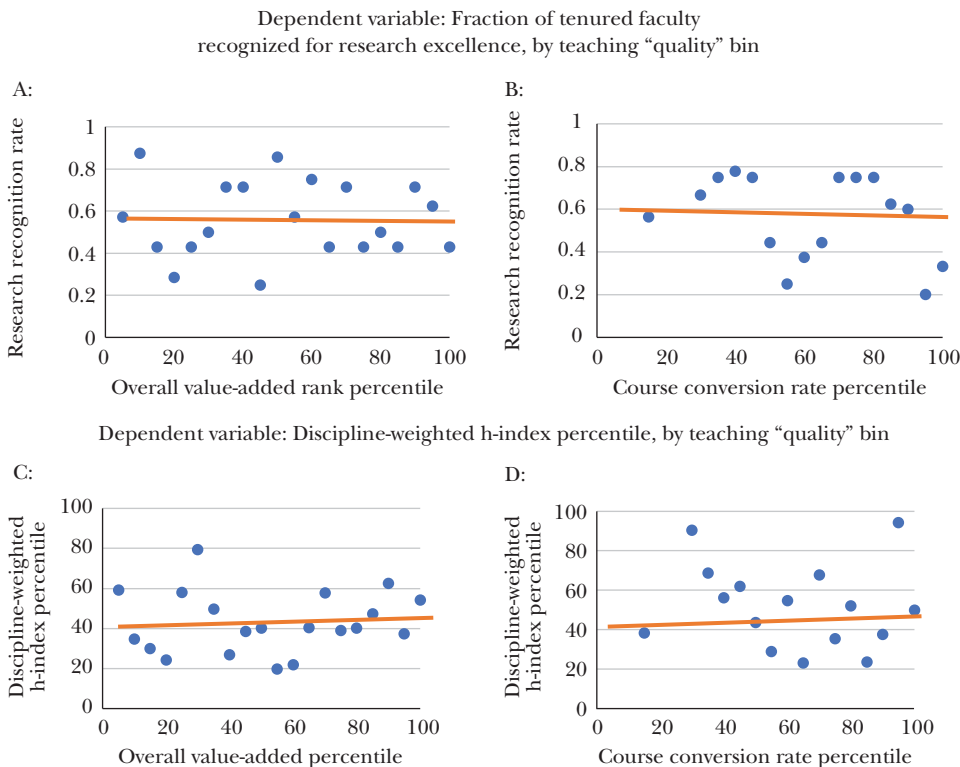
These two measures of research quality are much more highly correlated than our two measures of teaching success: Those Northwestern faculty members teaching introductory first-year courses whose research has been recognized by the university average in the 49th percentile of tenured faculty field-adjusted h-indices, while those tenured introductory course teachers who have not been recognized for their research average in the 36th percentile of tenured faculty field-adjusted h-indices. The fact both groups of faculty are at or below the university median suggests that introductory classes are disproportionately taught by faculty with somewhat less research success, at least by this measure. This last point, however, could be due in part to the previously mentioned bias of h-indices toward more senior faculty members, given that very senior faculty members are somewhat less likely to teach introductory courses.

With these two measures of teaching quality and two measures of research quality, we make four comparisons in Figure 1 of teaching quality and research quality among the *tenured* Northwestern faculty in our sample. Our bottom line is, regardless of our measure of teaching and research quality, there is no apparent relationship between teaching quality and research quality.

In Figure 1, panel A compares our value-added measure of teaching quality for future courses to the probability of being recognized for one's research. For ease of illustration, we group the faculty members into 20 equal-sized instructor quality bins, but we use the disaggregated data to estimate relationships. The relationship is essentially flat. With each percentile improvement in measured teaching value-added, a faculty member is 0.025 percentage points less likely to be recognized for research quality. This is a very precisely estimated zero: the standard error

Figure 1

Relationship between Measures of Teaching Excellence and Measures of Research Excellence

Dependent variable: Fraction of tenured faculty
recognized for research excellence, by teaching "quality" bin

Dependent variable: Discipline-weighted h-index percentile, by teaching "quality" bin

Source: This figure was previously published in Figlio and Schapiro (2017).
Note: Data include tenured faculty at Northwestern with 20 or more student observations. Instructor "value-added" is measured as the contribution to future course letter grades in the subject, calculated using the method developed by Figlio, Schapiro, and Soter (2015). The "course conversion rate" is the percentage of non-majors who ultimately take six or more subsequent courses in the subject.

of this estimate is just 0.14 percentage points. Put differently, an instructor at the 75th percentile of the instructor value-added distribution is only one percentage point less likely to be recognized for research evidence than would an instructor at the 25th percentile of the value-added distribution.

In Panel B, we see the same lack of a relationship when we instead measure instructor quality using a "conversion rate"—the fraction of initially non-majors who ultimately take six or more courses in the subject. As with the previous measure of instructor quality, we express this as a within-sample ranking. (The leftmost point on the graph is unevenly spaced because 32 percent of faculty studied convert zero undecided students to majors.) Again, we observe a precisely estimated

zero relationship between this alternative measure of instructor quality and the probability of research recognition: with each percentile increase in the instructor's conversion rate rank, a faculty member is 0.08 percentage points less likely to be recognized for research quality. The standard error of this estimate is just 0.13 percentile points. This means that an instructor at the 75th percentile of the "conversion rate" distribution is just four percentage points less likely to be recognized for research evidence than would an instructor at the 25th percentile of the distribution.

We repeat the same two analyses using the field-adjusted h-index as a measure of research excellence. Panel C compares our value-added measure of teaching quality to a faculty member's percentile rank in the field-adjusted h-index. Again, the relationship is virtually flat: with each percentile point improvement in measured teaching value-added, a faculty member is 0.067 percentile points higher in the h-index ranking. The standard error of this estimate is 0.114. Therefore, the difference between the 25th and 75th percentile of the teacher quality distribution, measured in terms of value-added, is just three percentile points in the h-index distribution (and the opposite signed relationship as seen with the other measure of research quality).

Finally, Panel D presents the same comparison, with the "conversion-rate" measure of instructor quality. With each percentage point improvement in measured teacher quality, a faculty member is 0.037 percentile points higher in the h-index ranking (standard error of 0.108), implying a difference in the h-index distribution of only two percentile points between the 25th and 75th percentile teachers.

In sum, regardless of our measure of effective teaching or exemplary scholarship, we find that top teachers are no more or less likely to be especially productive scholars than their less accomplished teaching peers. This is encouraging for those who fear that great teachers specialize in pedagogy at the expense of research. On the other hand, it is disappointing to observe that weak undergraduate teachers do not make up for their limitations in the classroom with disproportionate research excellence.

So what does this analysis imply in terms of staffing the undergraduate classroom? Our findings suggest that superb teaching does not come at the cost of diminished scholarship. Are great teachers poor scholars? Not according to our measures of teaching and research prominence. Of course, it's possible that this result arises because teaching or scholarship are imperfectly measured, but we are finding very "precise zeros" between variables with a reasonably high level of variation—that is, we don't find statistically significant relationships even though we have the statistical power in our data to detect even very modest relationships. At least in the scope of teaching by tenure-line Northwestern faculty, the factors that drive teaching excellence and those that determine research excellence are unrelated.

These findings have implications for university administrators and for policymakers. Some individuals in these groups prioritize research excellence over teaching quality, while others prioritize teaching excellence over research quality. Our analysis implies that policymakers worried about whether research efforts

will come at the expense of teaching, or vice versa, should have their fears at least partially allayed. But what if state legislators take seriously our finding that while top teachers don't sacrifice research output, it is also the case that top researchers don't teach exceptionally well? Why have those high-priced scholars in the undergraduate classroom in the first place? Surely it would be more cost-efficient to replace them in the classroom either with untenured, lower-paid professors, or with faculty not on the tenure-line in the first place. That, of course, is what has been happening throughout American higher education for the past several decades, as we discuss in detail in the section that follows. And, of course, there's the other potentially uncomfortable question that our analysis implies: Should we be concerned about the possibility that the weakest scholars amongst the tenured faculty are no more distinguished in the classroom than are the strongest scholars? Should expectations for teaching excellence be higher for faculty members who are on the margin of tenurability on the basis of their research excellence?

Should the Move toward Contingent Faculty in US Universities Concern Us?

The role of tenure in American higher education has declined dramatically in recent decades. In 1975, 57 percent of all faculty (full-time and part-time, excluding graduate students) were in the tenure system; by 2011, the proportion was about half that high at 29 percent, and it has continued to fall since. The American Association of University Professors website presents its Contingent Faculty Index summarizing data from the IPEDS Fall Staff Survey, and *The Chronicle of Higher Education* regularly summarizes the latest numbers. For example, Simonton (2019) reports that non-tenure-track appointments grew from 10 percent of all full-time faculty positions in 2008–09 to 27 percent in 2018–19 at the 870 institutions that participated in the American Association of University Professors' Faculty Compensation Survey in both years, and that 56 percent of full-time and part-time faculty members at four-year public institutions and 66 percent of those at four-year private nonprofit institutions were not on the tenure track in 2017. At two-year public colleges, fewer than 20 percent of faculty are in the tenure system.

There has been ongoing speculation in the education press about the level at which this decline in tenure might bottom out. A common prediction is that the share of tenure track/tenured faculty will eventually stabilize at between 15 to 20 percent, with tenure being largely limited to the flagship public and private research universities and the wealthiest of the liberal arts colleges (Wilson 2010). Morson and Schapiro (2015) predict that by 2040, only around 10 percent of faculty positions will be held by tenure-track/tenured professors.

This reduction in tenure trend seems to have accelerated since January 1, 1994, when the mandatory retirement age for college and university faculty was abolished by federal law. As a result, colleges that wanted an older faculty member to retire

have often sought to negotiate multi-year plans with incentives for a "phased retire-ment"—but a faculty member who does not wish to sign an agreement for such a plan cannot be required to do so. As a result, granting tenure involves additional end-of-career uncertainties, and shifting more instruction to contingent faculty mitigates those risks. Especially notable is the rise of the full-time, contingent faculty members at PhD-granting universities. Their representation within the entire group of full-time faculty went from 24 percent to 35 percent at public doctoral institu-tions from 1995 to 2011 and from 18 percent to 46 percent at private nonprofit doctoral institutions (Ehrenberg 2012).

The erosion of tenure has raised a number of concerns. Some observers have lamented the potential blow to academic freedom dealt by the decline of tenure, while others have focused on the often-challenging employment conditions under which many contingent faculty work (for example, June 2012; Wilson 2010). Further, McPherson and Schapiro (1999) point to efficiency gains from tenure; they outline its positive role in influencing the distribution of authority within colleges and universities. Here, we focus on educational outcomes: Do undergraduates taught by contingent faculty members learn as much as those taught by faculty who are tenured or on a tenure-track appointment?

On a national level, Ehrenberg and Zhang (2005) present evidence that hiring more part-time and contingent faculty lowers institutional graduation rates. This result is bolstered by Bettinger and Long (2010), who find a similarly negative effect on aggregate levels of persistence when they focus specifically on part-time adjuncts. These types of results indicate that even if contingent faculty are more popular with students—perhaps because of classroom behaviors that maximize student evaluations but not student learning—they nonetheless might not be successful in improving students' longer-term prospects.[4]

Several recent papers have made substantial steps toward understanding whether tenure-line faculty members outperform (or perform worse than) contingent faculty in the classroom. Bettinger and Long (2010) study this ques-tion using data largely centered on data from twelve public four-year colleges in the state of Ohio whose principal purpose is teaching. Their creative identifica-tion strategy involves treating short-term vacancies in departments as random events, and thus essentially analyzing the effects on learning from transitory adjunct faculty. While their analysis therefore may not speak to the effects of part-time faculty with longer-term contracts, and certainly does not address the effects of full-time contingent faculty, it does reflect the types of "one-off" contin-gent faculty hired with some regularity at many institutions. Bettinger and Long find some evidence that contingent faculty induce student interest in a subject, as measured by the likelihood that students take additional courses in that subject, though their data do not provide the opportunity to study how students perform in subsequent classes. Ran and Xu (2019), studying both two-year and four-year

---

[4] Carrell and West (2010) show that instructors who have better student evaluations tend to produce lower levels of "deep learning."

teaching-oriented colleges in an anonymous state university system, measure both "deep learning" in the sense of performing better in future courses à la Carrell and West (2010) and the likelihood of subsequent course-taking à la Bettinger and Long (2010). They find that contingent faculty perform worse on both measures than do tenure-line faculty. On the other hand, Feld, Salamanca, and Zoelitz (2020) study students randomly assigned to tutorials in a Dutch business school and find negligible differences across instructor type in effects on students' current and future academic outcomes, job satisfaction, and earnings. In sum, while the evidence varies from setting to setting, it is certainly not obvious that tenure-line faculty members are more effective instructors in teaching-intensive settings.

What about results specific to research universities, where the aforementioned multi-tasking problem of maximizing an objective function that includes both the production of cutting-edge research and the provision of outstanding undergraduate teaching may be more pronounced? Hoffmann and Oreopoulos (2009a) were the first to evaluate the instructional quality of tenure-line versus contingent faculty in a research university setting, though the Canadian research university context they study only permits analyzing the likelihood that students take additional classes in the same subject rather than observe their academic performance in future classes. They find no evidence that contingent faculty are either better or worse at inspiring students to take more classes in their subjects.

In Figlio, Schapiro, and Soter (2015), we carry out an analysis at Northwestern, where we also can measure both performance in future related courses and the likelihood of subsequent course-taking. While Northwestern contingent faculty do not have the employment protections that many at places like the University of California (which has a status of "Lecturer with Security of Employment") have, Northwestern is still unusual relative to most institutions in that contingent faculty members tend to have stable, long-term relationships with the university and a substantial majority are full time. (Generally, these are renewable three-year contracts, and most of the contingent faculty in our study are in their second or later contract.) Thus, our results should be viewed in the context of where non-tenure faculty at a major research university function as designated teachers (both full-time and part-time) with long-term relationships to the university. We find that, on average, tenure-line faculty members do not teach introductory undergraduate courses as well as do their (largely full-time, long-term) contingent faculty counterparts. In other words, our results suggest that on average, first-term freshmen learn more from contingent faculty members than they do from tenure track/tenured faculty.

Are our results driven by a handful of outliers or a by larger swath of the distribution? That is, do these differences arise because most tenure track/tenured faculty members perform similarly to most contingent faculty members? Or are the differences due to events at the tails of the distribution, with either the best contingent faculty teachers substantially outperforming the best tenure track/tenured teachers or the worst tenure track/tenured teachers performing considerably worse than the worst contingent faculty teachers?"

Distribution of Instructor "Value Added," Contingent versus Tenure-Line Faculty

Distribution of instructor "value added,"
contingent versus tenure-track faculty

Quantile of the contingent (or tenure-track) "value added" distribution
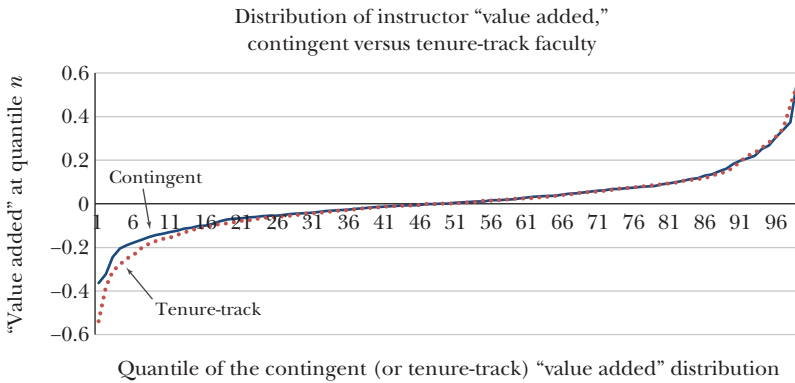
Source: This figure was first published in Figlio, Schapiro, and Soter (2015).

     To explore this question, Figure 2 compares the distributions of value-added in future related courses of individual contingent faculty teachers and tenure track/tenured teachers. In the figure, we plot a variant of the cumulative density function where the percentile in each distribution is on the horizontal axis and the corresponding value-added measure is on the vertical axis, which makes clear what part of the distribution is generating the results. An individual instructor's value-added is an instructor-specific fixed effect retained from our preferred specification (in which we estimate instructors' effect on grade points earned in the next course, controlling for both student fixed effects and next-course fixed effects). As can be seen in the figure, the top three-quarters of the contingent faculty and tenure track/tenured faculty distributions are virtually perfectly overlapping, so the most outstanding contingent faculty members and most outstanding tenure track/tenured faculty members perform essentially identically, and the same is true at other points in the distribution such as the median. But the bottom quarter of the tenure track/tenured faculty have lower value-added than the bottom quarter of the contingent faculty, and this difference is substantial for the bottom 13 percent of the distribution (around the weakest 150 instructors, by our definition, amongst those who taught introductory courses over the decade of time covered in our study). It is clear that our results are not being driven by a handful of outliers, but it is also clear that the difference in average outcomes is due to the differences at the bottom of the value-added distribution.

     In some ways, this pattern is exactly what we might have expected: Contingent faculty members who are hired to teach and who perform relatively poorly are less likely to be renewed than are those who perform well, while tenure-track faculty who are relatively poor teachers may be promoted and retained for reasons other than their teaching ability. But are there specific differences between contingent faculty members and tenure track/tenured faculty members that can explain our findings? To shed some light on this issue,

we collected curriculum vitae available through extensive web-searching. Two directly observable variables are years of experience (calculated based on time since PhD and employment history) and native language (calculated based on the country in which a faculty member earned his or her bachelor's degree or its equivalent). Tenure track/tenured faculty are modestly more likely to have attended undergraduate institutions in English-speaking countries (86.3 percent versus 79.2 percent for contingent faculty) and average dramatically more experience (21.9 years versus 11.6 years for contingent faculty). But these variables explain no more than a modest fraction of our results.

Rather, it is apparently the case that the bottom tail of the contingent faculty distribution are considerably more effective classroom instructors than are those at the bottom tail of the tenure-line faculty distribution, at least in the first-term freshman classroom. There are extraordinarily good reasons for top research universities facing multi-tasking problems to recruit, retain, and reward faculty members on the basis of research, but apparently this comes at the cost of having a fraction of the distribution of tenure-line faculty be disproportionately poor performers in the classroom—both in terms of inspiration and in terms of preparation for future courses.

It is, of course, also noteworthy how closely the top three-quarters of the tenure-line and contingent faculty distributions track one another in Figure 2. Most tenure-line faculty teaching freshmen at Northwestern perform just as well, according to our measures of teaching effectiveness, as do those who are recruited, retained, and rewarded on the basis of their teaching. This must be reassuring to those who worry that there's a stark tradeoff between teaching and research quality. (The results we present above, in which we look just at tenured Northwestern faculty, should also provide reassurance on this front.) Rather, the instructional quality gap between tenure-line and contingent faculty members is concentrated in a relatively small fraction of the teaching force, at least at Northwestern.

What about our other measure of teaching quality—the ability of charismatic faculty to inspire students to major in their subject? Contingent faculty members are modestly more likely to convert students into majors; the typical contingent faculty member converts 1.6 percent more students to majors than does the typical tenure-line faculty member.

An obvious question is the degree to which these findings can be generalized. Because a key part of our identification strategy is to limit our analysis to first-term freshman undergraduates, the evidence that contingent faculty produce better outcomes may not apply to more advanced courses. Further, Northwestern is among the most selective and highly ranked research universities in the world, and its ability to attract first-class contingent faculty may be different from that of most other institutions.

It is also important to note that a substantial majority of contingent faculty at Northwestern are full-time faculty members with long-term contracts and benefits, and therefore may have a stronger commitment to the institution than some of their contingent counterparts at other institutions. Northwestern's tenure track/tenured faculty members may also have different classroom skills from those at other schools.

Finally, Northwestern students come from a rarefied portion of the preparation distri-bution and are far from reflective of the general student population in the United States. That said, our results are strongest for the students and subjects that are most likely to generalize to a considerably wider range of institutions: The benefits of taking courses with contingent faculty appear to be stronger for the relatively marginal students at Northwestern (although they are still very well-prepared students), and our results are similar for top-ranked departments as for lower-ranked departments.

There are many aspects relating to changes in the tenure status of faculty, from the impact on research productivity to the protection of academic freedom. But certainly, learning outcomes are an important consideration in evaluating whether the observed trend away from tenure track/tenured toward contingent faculty is good or bad. Our results, coupled with that of others, provide evidence that the rise of full-time designated teachers at US colleges and universities may be less of a cause for alarm for the quality of teaching than many assume. Of course, it is important to note that our analysis is necessarily a partial-equilib-rium one: While our analysis tells us on the margin how faculty types compare, it does not tell us whether the current allocation of faculty slots to contingent versus tenure-track faculty is efficient. For one thing, because contingent faculty members teach more classes per year than tenure-track faculty members, the implication that contingent faculty manage to maintain comparable (or supe-rior) teaching with heavier teaching loads speaks to the potential tradeoffs between staffing the undergraduate classroom with more research-intensive versus more teaching-intensive faculty members.

In What Ways Does Instructor Gender, Race, and Ethnicity Matter?

In fall 2017, 11.6 percent of US college and university faculty (and 10.8 percent of tenure-line faculty) were Black, Hispanic, Native American, or mixed-race, despite 35 percent of the student body being one of these demographic groups (US Depart-ment of Education 2018; Tables 306.10 and 315.20). Also in fall 2017, 46.3 percent of faculty (and 43.2 percent of tenure-line faculty) were women, at a time when 56.7 percent of the student body were women. Large mismatches between student and faculty demographics like this have a number of consequences and risks: a lack of diversity in substantive representation, the potential for limiting the range of what is taught and how it is taught, and how it might affect major and career choices of students due to implicit stereotypes (Reuben, Sapienza, and Zingales 2014) or lack of role models (Zafar 2013).

A large literature in K–12 education suggests that outcomes such as test scores, attendance, and suspension rates are affected by the demographic match between teachers and students (Dee 2004; Dee 2005; Egalite, Kisida, and Winters 2015; Gershenson, Holt, and Papageorge 2016; Lim and Meer 2017; Lindsay and Hart 2017). This relationship appears to have a long reach: Gershenson et al. (2018) find that if a Black male student has at least one Black teacher in the third, fourth, or

fifth grade, he is significantly less likely to drop out of high school and more likely to aspire to attend a four-year college, as proxied by taking a college entrance exam, and that these effects are particularly pronounced for economically disadvantaged Black male students. Looking at data from schools in Israel, Lavy and Sand (2018) demonstrate the long-run effects on later courses, career choice, and earnings when primary teachers show gender bias: there is reason to believe that the same is true regarding race and ethnicity.

Does demographic match matter in college as well? Some recent evidence suggests that the answer is "yes." Using data from Canadian research universities, Hoffmann and Oreopoulos (2009b) show that college students are more likely to complete a course and to perform better in that course when they have a same-gender professor. Carrell, Page, and West (2010) show that having a female professor in a course in science, technology, engineering, or mathematics will substantially increase the likelihood that females will take more courses and eventually graduate with a degree in these fields. Fairlie, Hoffmann, and Oreopoulos (2014) demonstrate, using data from a large community college in California, that racial and ethnic minority faculty members reduce the minority achievement gap in class performance and dropout rate, while Kofoed and McGovney (2019) find that same-gender and same-race role models at the US Military Academy have a strong effect on occupational choice. Mansour et al. (2020) examine students at the US Air Force Academy and find that high-ability female students who were assigned a female professor had a substantial increase in the probability of working in a science, technology, engineering, or mathematics occupation and in the probability of receiving a master's degree in those fields. It seems likely, therefore, that hiring a racially, ethnically, and gender diverse faculty will benefit students in an environment, such as the current state of affairs in US higher education, where the student body is much more diverse than is the faculty responsible for teaching them.

Women and minorities are underrepresented not just among faculty members in general, but in particular among tenure-line faculty members. While Black, Hispanic, Native American, and mixed race faculty members represented 12.7 percent of assistant professors in fall 2017 and 11.6 percent of associate professors, they are only 8.2 percent of full professors (US Department of Education 2018). The gradient is even more striking for women, who comprised 51.8 percent of assistant professors—almost, if not quite, the female proportion of the student body—but 45.4 percent of associate professors and 32.8 percent of full professors in fall 2017. Women represented 49.4 percent of contingent faculty members in fall 2017, while Black, Hispanic, Native American, and mixed race individuals comprised 11.4 percent of contingent faculty. Looking at the intersection of gender and minority groups, 7.3 percent of assistant professors are women who are Black, Hispanic, Native American, or mixed race, but only 5.8 percent of associate professors and 3.2 percent of full professors are.

The role-model issue is relevant not just for students but for junior faculty as well: While in fall 2017 there were 1.79 white male full professors for every white male assistant professor, there was just 0.39 Black female full professors for every Black female assistant professor, 0.54 Hispanic female full professors for every

Hispanic female assistant professor, 0.75 Native American female full professors for every Native American female assistant professor, and 0.36 mixed race female full professors for every mixed race female assistant professor. Lundberg and Stearns (2019) identify a number of challenges that women in economics face and point to institutional policies and promotion and tenure processes that are biased against women. The disproportionate impacts of the current COVID-19 pandemic on female academics, and especially female academics of color, seem likely to further exacerbate gender inequities in the academy, and universal remedies like blanket tenure clock extensions might have the unintended consequence of disadvantaging women in the academy.[5] Of course, the issue starts before women become faculty members; for example, there exists considerable heterogeneity in the success of female economics graduate students across PhD programs (Boustan and Langan 2019).

The causes and specific consequences of these racial and gender differences in faculty rank are beyond the scope of this paper, but they signify a structural challenge that colleges and universities face. Two recent papers published in this journal, Buckles (2019) and Bayer, Hoover, and Washington (2020), provide advice for ways to make the economics profession more inclusive of and hospitable to female scholars and scholars of color at all career stages. However, the shift away from tenure-line faculty toward contingent faculty means that there are fewer opportunities to diversify the tenure ranks, putting several institutional objectives in direct tension with one another.

Conclusion

With non-tenure line faculty providing unusually good undergraduate teaching, and within the tenured group, top scholars neither better nor worse in teaching than their colleagues, why would anybody ever allocate the best senior scholars to undergraduate classes? Clearly, it would be more efficient to replace them with less expensive non-tenure eligible teachers or with younger non-tenured tenure-line faculty. But before leaping to this conclusion, we offer a few words of caution. Illustrious research faculty provide a draw for students and faculty alike. Even if their undergraduate teaching isn't exceptional, their presence often is. Having outstanding scholars teaching first-year students sends a signal to the community that the school takes undergraduate education seriously—that it isn't just research and the production of PhD students that matters. Meanwhile, while it may be tempting to hire female and minority faculty members into teaching-track positions in order to quickly provide more students with "a professor like me," the segregation of "professors like me" into lower-prestige teaching positions may reinforce stereotype threats and have deleterious effects on minority students.

---

[5] As one important example that is highly salient at the present moment, Antecol, Bedard, and Stearns (2018) show that gender-neutral tenure clock-stopping policies associated with parenting reduced female economists' tenure probabilities but increased male economists' tenure probabilities.

What about the recent move at the University of California and elsewhere towards effectively tenuring some of their full-time teaching faculty? Our analysis suggests that if the motivation for moving undergraduate teaching from faculty with responsibility for both teaching and research to faculty whose sole responsibility is teaching is to protect the time of the former group for scholarship, this approach needs to be questioned. We have shown that, at least at Northwestern, the gap in teaching performance between tenure-line and contingent faculty depends entirely on the weakest teachers among the tenured professors. Presumably, weak contingent faculty are not renewed. While we certainly see the strong benefit of offering job security for teaching-track faculty (and recognize that higher levels of job protections likely attract more excellent teachers to the university), giving de facto tenure to this group might reduce the power of this important lever for department chairs, deans, and provosts. On the other hand, if doing so permits universities to invest more in research faculty members' research time, this could help them to partially solve the university's multi-tasking problem but at the potential cost of reinforcing a two-tiered faculty system, albeit with more job protections for the less-prestigious tier of faculty.

Finally, a word about the role of research at those universities that take special pride in their scholarly output. The reason why most of the top-rated universities in the world are located in the United States is not what goes on in their classrooms; it is the research power of their faculties. Read a college guidebook or go on a college tour at a top research university. Over and over, you will see pictures of and hear stories about superstar research faculty including Pulitzer Prize winners, Nobel Laureates, National Academy members—all in the undergraduate classroom and often teaching first-year students. Whether those pictures properly represent reality is one question; what we address here is whether it should represent reality. Given that we do not find a tradeoff between great teaching and great research, we believe that having top research faculty do a share of the undergraduate and first-year teaching is an advantageous allocation of faculty talent.

The multi-tasking challenge faced by our nation's most prestigious research universities isn't easy to solve. But measuring both teaching and research outcomes seems to be a useful first step and could be helpful in cracking the nut of how to staff the higher education classroom. One key lesson from our analysis is that teaching-intensive institutions and research universities alike could be well-served to pay less attention to numerical student evaluations and instead work to develop alternative ways to measure and reward good teaching. Our analysis provides some ideas about ways institutions might try to go about doing this.

References

American Sociological Association. 2019. "Statement on Student Evaluations of Teaching." *American Sociological Association*, September. https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_feb132020.pdf.

Antecol, Heather, Kelly Bedard, and Jenna Stearns. 2018. "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?" *American Economic Review* 108 (9): 2420–41.

Bavishi, Anish, Juan Madera, and Michelle Hebl. 2010. "The Effect of Professor Ethnicity and Gender on Student Evaluations: Judged before Met." *Journal of Diversity in Higher Education* 3 (4): 245–56.

Bayer, Amanda, Gary A. Hoover, and Ebonya Washington. 2020. "How You Can Work to Increase the Presence and Improve the Experience of Black, Latinx, and Native American People in the Economics Profession." *Journal of Economic Perspective* 34 (3): 193–219.

Beleche, Trinidad, David Fairris, and Mindy Marks. 2012. "Do Course Evaluations Truly Reflect Student Learning? Evidence from an Objectively Graded Post-Test." *Economics of Education Review* 31 (5): 709–19.

Bettinger, Eric P., and Bridget Terry Long. 2010. "Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes." *The Review of Economics and Statistics* 92 (3): 598–613.

Boring, Anne. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145: 27–41.

Boring, Anne, Kellie Ottoboni, and Philip B. Stark. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research.*

Boustan, Leah, and Andrew Langan. 2019. "Variation in Women's Success across PhD Programs in Economics." *Journal of Economic Perspectives* 33 (1): 23–42.

Braga, Michela, Marco Paccagnella, and Michele Pellizzari. 2014. "Evaluating Students' Evaluations of Professors." *Economics of Education Review* 41: 71–88.

Buckles, Kasey. 2019. "Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage." *Journal of Economic Perspectives* 33 (1): 43–60.

Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125 (3): 1101–44.

Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy* 118: 409–32.

Dee, Thomas S. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *The Review of Economics and Statistics* 86 (1): 195–210.

Dee, Thomas S. 2005. "A Teacher Like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95 (2): 158–65.

Deslauriers, Louis, Logan S. McCarty, Kelly Miller, Kristina Callaghan, and Greg Kestin. 2019. "Measuring Actual Learning versus Feeling of Learning in Response to Being Actively Engaged in the Classroom." *Proceedings of the National Academy of Sciences of the United States of America* 116 (39): 19251–57.

Egalite, Anna J., Brian Kisida, and Marcus A. Winters. 2015. "Representation in the Classroom: The Effect of Own-Race Teachers on Student Achievement." *Economics of Education Review* 45: 44–52.

Ehrenberg, Ronald G. 2012. "American Higher Education in Transition." *Journal of Economic Perspectives* 26 (1): 193–216.

Ehrenberg, Ronald G., and Liang Zhang. 2005. "Do Tenured and Tenure Track-Faculty Matter?" *Journal of Human Resources* 40 (3): 647–59.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review* 104 (8): 2567–91.

Feld, Jan, Nicolás Salamanca, and Ulf Zoelitz. 2020. "Are Professors Worth It? The Value-Added and Costs of Tutorial Instructors." *Journal of Human Resources* 55 (3): 836–63.

Figlio, David N., and Morton Schapiro. 2017. "Are Great Teachers Poor Scholars?" *Brookings Institution Evidence Speaks Reports* 2 (6): 1–7.

Figlio, David N., Morton O. Schapiro, and Kevin B. Soter. 2015. "Are Tenure Track Professors Better Teachers?" *The Review of Economics and Statistics* 97 (4): 715–24.

Flaherty, Colleen. 2016. "Bias against Female Instructors." *Inside Higher Education*, January 11. https://www.insidehighered.com/news/2016/01/11/new-analysis-offers-more-evidence-against-student-evaluations-teaching.

Flaherty, Colleen. 2019a. "The Dangers of Fluent Lectures." *Inside Higher Education*, September 9. https://www.insidehighered.com/news/2019/09/09/study-how-smooth-talking-professors-can-lull-students-thinking-theyve-learned-more.

Flaherty, Colleen. 2019b. "Speaking out against Student Evals." *Inside Higher Education*, September 10 https://www.insidehighered.com/news/2019/09/10/sociologists-and-more-dozen-other-professional-groups-

speak-out-against-student.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas Papageorge. 2018. "The Long-Run Impacts of Same-Race Teachers." NBER Working Paper 25254.

Gershenson, Seth, Stephen B. Holt, and Nicholas W. Papageorge. 2016. "Who Believes in Me? The Effect of Student-Teacher Demographic Match on Teachers Beliefs." *Economics of Education Review* 52: 209–24.

Hoffmann, Florian, and Philip Oreopoulos. 2009a. "Professor Qualities and Student Achievement." *Review of Economics and Statistics* 91 (1): 83–92.

Hoffmann, Florian, and Philip Oreopoulos. 2009b. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources* 44 (2): 479–94.

June, Audrey Williams. 2012. "Adjuncts Build Strength in Numbers: The New Majority Generates a Shift in Academic Culture." *Chronicle of Higher Education*, November 5. https://www.chronicle.com/article/adjuncts-build-strength-in-numbers/.

Kofoed, Michael S., and Elizabeth McGovney. 2019. "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point." *Journal of Human Resources* 54 (2): 430–67.

Lavy, Victor, and Edith Sand. 2018. "On the Origins of Gender Gaps in Human Capital: Short- and Long-Term Consequences of Teachers' Biases." *Journal of Public Economics* 167: 263–79.

Lim, Jaegeum, and Jonathan Meer. 2017. "The Impact of Teacher–Student Gender Matches." *Journal of Human Resources* 52 (4): 979–97.

Lindsay, Constance A., and Cassandra M.D. Hart. 2017. "Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina." *Educational Evaluation and Policy Analysis* 39 (3): 485–510.

Lundberg, Shelly, and Jenna Stearns. 2019. "Women in Economics: Stalled Progress." *Journal of Economic Perspectives* 33 (1): 3–22.

Mansour, Hani, Daniel I. Rees, Bryson M. Rintala, and Nathan N. Wozny. 2020. "The Effects of Professor Gender on the Post-Graduation Outcomes of Female Students." NBER Working Paper 26822.

McPherson, Michael, and Morton Owen Schapiro. 1999. "Tenure Issues in Higher Education." *Journal of Economic Perspectives* 13 (1): 85–98.

Mengel, Friederike, Jan Sauermann, and Ulf Zoelitz. 2019. "Gender Bias in Teaching Evaluations." *Journal of the European Economic Association* 17 (2): 535–66.

Morson, Gary Saul, and Morton Schapiro. 2015. *The Fabulous Future? America and the World in 2040.* Evanston, IL: Northwestern University Press.

Ran, Florence Xiaotao, and Di Xu. 2019. "Does Contractual Form Matter? The Impact of Different Types of Non-Tenure-Track Faculty on College Students' Academic Outcomes." *Journal of Human Resources* 54 (4): 1081–1120.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales. 2014. "How Stereotypes Impair Women's Careers in Science." *Proceedings of the National Academy of Sciences* 111 (12): 4403–08.

Rivera, Lauren A., and András Tilcsik. 2019. "Scaling Down Inequality: Rating Scales, Gender Bias, and the Architecture of Evaluation." *American Sociological Review* 84 (2): 248–74.

Simonton, Teghan. 2019. "Tenure and Other Variations." *The Chronicle of Higher Education*, August 18. https://www.chronicle.com/article/tenure-and-other-variations/.

Stark, Philip, and Richard Freishtat. 2014. "An Evaluation of Course Evaluations." *ScienceOpen*.

Stroebe, Wolfgang. 2016. "Why Good Teaching Evaluations May Reward Bad Teaching: On Grade Inflation and Other Unintended Consequences of Student Evaluations." *Perspectives on Psychological Science* 11 (6): 800–816.

Supiano, Beckie. 2019. "Sociologists Caution Colleges Not to Over-Rely on Student Evaluations of Teaching." *The Chronicle of Higher Education*, September 9. https://www.chronicle.com/article/sociologists-caution-colleges-not-to-over-rely-on-student-evaluations-of-teaching/

U.S. Department of Education, National Center for Education Statistics. 2018. "Digest of Education Statistics." https://nces.ed.gov/programs/digest/d18/ (accessed December 10, 2020).

Wilson, Robin. 2010. "Tenure, RIP: What the Vanishing Status Means for the Future of Education." *Chronicle of Higher Education*, July 4. https://www.chronicle.com/article/tenure-rip-what-the-vanishing-status-means-for-the-future-of-education/.

Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources* 48 (3): 545–95.

The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System

John Bound, Breno Braga, Gaurav Khanna, and Sarah Turner

University learning has facilitated the flow of individuals and knowledge across national borders for centuries, but the recent scale of student flows and the magnitude of tuition revenues from foreign students across the globe is unprecedented. The number of students pursuing higher education degrees outside their home countries more than doubled between 2000 and 2017 to reach 5.3 million (UNESCO 2018).

For the United States, which has a large number of colleges and universities and a disproportionate share of the most highly ranked colleges and universities in the world, total enrollment of foreign students more than tripled between 1980 and 2017, from 305,000 to over one million students in 2017 (National Center for Enrollment Statistics 2018). This rising population of students from abroad has made higher education a major export sector of the US economy, generating $44 billion in export revenue in 2019, with educational exports being about as big as the total exports of soybeans, corn, and textile supplies combined (Bureau of Economic Analysis 2020).

Traditionally, talented undergraduate and graduate students from abroad have engaged with educational opportunities that exist in the United States at a

■ *John Bound is George E. Johnson Collegiate Professor of Economics, University of Michigan, Ann Arbor, Michigan. Breno Braga is Senior Research Associate, Urban Institute, Washington, DC. Gaurav Khanna is Assistant Professor of Economics, School of Global Policy and Strategy, University of California at San Diego, La Jolla, California. Sarah Turner is University Professor of Economics and Education, Souder Family Professor, University of Virginia, Charlottesville, Virginia. Bound and Turner are also Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are jbound@umich. edu, bbraga@urban.org, gakhanna@ucsd.edu, and set5h@virginia.edu.*

time when their home countries often had more limited high-quality university options. In addition, for students, especially those in fields related to science, technology, engineering, and mathematics, time spent studying in the United States has facilitated access to job opportunities, with the US visa system structured to encourage this behavior. Unlike work visas, student visas are not subject to a cap and constitute an important pathway for the foreign-born to enter the US labor market (Rosenzweig 2006; Bound et al. 2014). The participation of students from abroad in US higher education affects the global production of skills and ultimately alters the allocation of university-educated workers to labor markets in the United States and abroad. On the supply side of higher education, US colleges and universities saw the opportunity to recruit talented students and, in some cases, to generate revenue.

We begin with an overview of the basic evidence of student flows to US colleges and universities by degree level and type of institution and the visa policies which mediate these flows. We examine how factors driving the demand for higher education—reflecting socioeconomic and demographic change abroad, and supply-side factors, reflecting the behavior of US colleges and universities—impact these flows.

Finally, we explore the potential consequences of reductions in foreign student flows for talent development and labor markets in the United States and abroad. Even before the COVID-19 pandemic and accompanying recession, there was evidence that enrollment of foreign students in US higher education was slowing dramatically, driven by some combination of improved educational and employment opportunities in home countries and other non-US destinations as well as perceptions of rising US hostility to immigrants. Given the formidable levels of tuition revenue generated by foreign students, especially at the undergraduate and master's levels, any reduction in the flow of foreign students would have a direct and negative impact on university resources that would not be easily offset by other sources of support. While reductions in the flow of foreign students at the doctorate level would not lead to declines in tuition revenues—given that PhD students usually receive financial support from universities—disruptions in academic research are likely to follow, which would likely not be offset in full by growth in doctorate study among domestic students.

Trends in Higher Education Flows

The number of foreign students enrolled in US universities at both the undergraduate and graduate levels has grown considerably over the last four decades: as noted earlier, total foreign enrollment rose from 305,000 in 1980 to over one million students in 2017 (National Center for Enrollment Statistics 2018). Foreign students as a share of total enrollment increased from 2.5 to 5.1 percent over this interval. Turning to degrees awarded by US institutions, about 5 percent of all bachelor's degrees (BAs) were awarded to international students in 2017–18, 18 percent of master's degrees (MAs), and 13 percent of doctorate degrees, as illustrated in Figure 1.

While undergraduate and graduate enrollment maintain broadly similar trajectories over the past 40 years (as indicated in Figure 1), the underlying causes of

Figure 1
Total and Share of Non-Resident Degrees by Academic Level
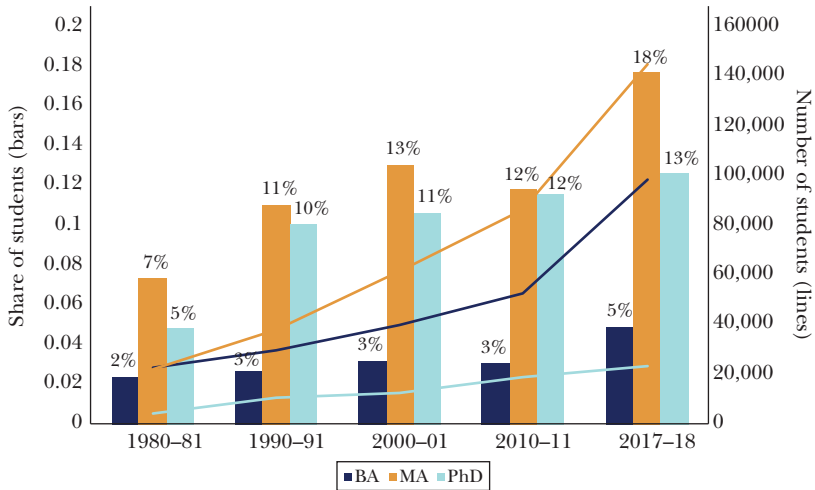
Source: IPEDS Degrees Awarded (1980–2017).
Note: BA is all undergraduate degrees, MA is all master's degrees, and PhD is all doctoral degrees granted to non-resident students. Left axis shows the share of total (resident plus non-resident) degrees, and the right axis shows the number of degrees granted to non-residents.

enrollment growth are quite distinct, as we discuss below. In addition, over the past two decades, the rise in master's-level enrollment has generated most of the increase in graduate enrollment. While the number of doctorate degrees awarded to international students increased by 22 percent between 2010 and 2017 (from 18,965 to 23,199), the number of master's degrees increased by 68 percent (from 163,827 to 184,074) over the same period.

At the level of countries most likely to send students to the United States, economies in transition and those with newly opened markets often lead with growth in US enrollment at the doctorate level because these programs offer financial support in the form of fellowships, research assistantships, and teaching assistantships (Bound, Turner, and Walsh 2009). Undergraduate enrollment at US institutions then follows only after there is expansion in the pool of students able to afford the tuition of US undergraduate programs; China exemplifies this pattern with doctorate-level enrollment expanding in the 1980s and 1990s, followed by undergraduate enrollment in the 2000s. Today, China is the largest source country for enrollment at both the undergraduate and graduate levels. Other countries with substantial student flows include India, South Korea, and Saudi Arabia.[1]

---

[1] There was a substantial growth in the number of students in the United States from Saudi Arabia over the decade from 2003 to 2013, but this is something of a special case. Much of this growth was concentrated at the undergraduate level, increasing from 2,022 students in 2003 to 26,865 in 2013. A clear

Table 1

International Students Enrolled in Post-Secondary Institutions by Destination Country

Year	2000	2010	2017
Panel A. Students from China			
Australia	5,008	87,588	128,498
Canada	4,701	26,298	66,161
United Kingdom	6,158	55,496	96,543
United States	50,281	126,498	321,625
Panel B. Students from India			
Australia	4,578	20,429	51,976
Canada	969	5,868	32,616
United Kingdom	3,962	38,205	16,421
United States	39,084	103,968	142,618
Panel C. Students from South Korea			
Australia	2,361	7,311	8,316
Canada	1,116	4,320	5,277
United Kingdom	2,165	4,347	5,157
United States	38,026	71,514	56,186

Source: UNESCO Institute for Statistics 2018.
Note: Numbers depict total number of international students enrolled across all degree statuses (undergraduate and graduate) in 2000, 2010, and 2017.

The increase in international students is not a uniquely US-centric phenomenon. Colleges and universities in Australia, Canada, and the United Kingdom also experienced a rapid increase in the enrollment of students from China and India since 2000, as shown in Table 1. Although the United States remains the largest destination country for students from these countries, the US higher education system is no longer as dominant as it was 20 years ago. As an illustration, student flows from China to the United States were more than 10 times larger than the flows to Australia and Canada in 2000; by 2017, those ratios fell to 2.5 to 1 and 3.3 to 1, respectively. Yet even as competition for international students has increased, the world market for higher education remains highly concentrated with just eight countries accounting for 58 percent of net student inflows in 2017 (UNESCO Institute for Statistics 2018)—the United States, United Kingdom, Australia, Germany, France, Russia, Canada, and Japan—and the first three of those countries accounting for 34 percent of all student imports.

---

impetus behind this increase was the introduction of the King Abdullah Scholarship Program, which stemmed from efforts to improve Saudi-US relations post-9/11, but has since grown into a substantial program aimed at boosting Saudi human capital. However, decreased budgets and new restrictions on approved universities have limited its growth since 2016. For discussion, see Saudi Arabian Cultural Bureau (2014).

Distribution of Students across Institutions and Fields of Study

The enrollment of international students varies considerably across postsecondary institutions. In the 21st century, foreign enrollment of undergraduate students is largely concentrated at public research universities, including large institutions like the University of Illinois Urbana-Champaign and Purdue University, which are somewhat less selective than top private research universities. The concentration of international students at these public universities reflects their scale, but also the fact that these universities have relied on tuition revenue from foreign students to cushion the effects of falling appropriations in the last decade (Bound et al. 2020). Still, the reliance of US colleges and universities on tuition revenue from abroad is not a 21st century phenomenon. In the late 1970s, the exposure of many private colleges to risk of a foreign enrollment shock became evident when relations (and financial flows) with Iran soured, and some colleges and universities found themselves at financial risk when payments from Iran ceased (Hechinger 1979).

As shown in Table 2, foreign students studying at the undergraduate level are most numerous at research-intensive public universities (about 32 percent of all bachelor's degrees), though they also enroll in substantial numbers at non-doctorate and less selective private and public institutions. Declining state support for public colleges and universities is one factor propelling the enrollment of foreign undergraduates at public universities. Since the mid-1980s, state appropriations per student for these institutions have fallen from about $12,000 per full-time equivalent to less than $7,000 in 2015. For public universities, the balance between state appropriations and tuition revenues has shifted markedly over time toward greater reliance on tuition revenues, which induced publicly funded universities to seek tuition revenue from full-fee paying international undergraduates (Bound et al. 2020).

The enrollment of international master's students is more difficult to characterize. For international students, incentives for pursuing an MA degree are diverse. One is the desire to live in a major US city like New York or Los Angeles; indeed, Columbia, New York University, and the University of Southern California stand out as institutions that awarded more than 3,000 MA degrees to foreign students in 2016. For some international master's students, these programs provide skills and credentials to strengthen applications to US doctorate programs, while for others, MA-level study yields direct access to employment options in the United States, particularly in areas where information technology-related industries are expanding.[2] In particular, the enrollment of students from India is typically concentrated in master's programs, with more than 90,000 Indian students enrolled in master's programs in 2015. It is also noticeable that some less selective public universities, such as the

---

[2] Focusing on MA degrees in computer science and information technology-related fields, Bound et al. (2014) note that there is substantial heterogeneity in the programs awarding degrees to temporary residents. Institutions awarding large numbers of master's degrees in computer science to temporary residents in 2013 include Carnegie Mellon University (464), Illinois Institute of Technology (397), University of Southern California (377), Columbia University in the City of New York (292), and University of Texas at Dallas (214). Ghose and Turner (2020) demonstrate the sensitivity of MA enrollment from foreign students to labor demand variation, with much of the MA enrollment changes concentrated among less-selective and for-profit institutions.

Table 2

Distribution of Foreign Students at Public and Non-Profit Universities and Tuition Status, 2015

Degree type	Institution type	Number of degrees	Tuition and fee	Expected student payment	Fully funded
Panel A. Private, non-profit universities					
BA	Non-Doctorate	16,518	$31,138	$20,355	13%
BA	Other Doctorate	5,930	$37,526	$30,486	5%
BA	Very High Research Doctorate	7,332	$45,266	$38,466	11%
MA	Non-Doctorate	20,452	$20,358	$17,748	3%
MA	Other Doctorate	14,102	$26,564	$24,313	3%
MA	Very High Research Doctorate	30,096	$45,512	$39,858	5%
PhD	Non-Doctorate	348	$27,353	$20,589	20%
PhD	Other Doctorate	908	$25,667	$10,645	51%
PhD	Very High Research Doctorate	4,116	$38,451	$4,394	87%
Panel B. Public universities					
BA	Non-Doctorate	15,435	$15,324	$13,099	6%
BA	Other Doctorate	10,549	$20,313	$18,321	6%
BA	Very High Research Doctorate	26,187	$29,245	$28,249	2%
MA	Non-Doctorate	15,128	$13,899	$12,918	5%
MA	Other Doctorate	21,649	$18,020	$15,517	10%
MA	Very High Research Doctorate	32,423	$27,410	$23,709	11%
PhD	Non-Doctorate	414	$14,613	$4,857	73%
PhD	Other Doctorate	2,936	$16,217	$4,498	72%
PhD	Very High Research Doctorate	9,771	$22,238	$4,504	79%

Source: IPEDS Degrees for the number of degrees and F-1 visa administrative data, from United States Citizenship and Immigration Services (USCIS) for the other statistics.
Note: "Very High Research Doctorate" are universities classified as having very high research activity by the Carnegie 2010 classifications. "Other Doctorate" are universities classified as having high research activity by the Carnegie 2010 classifications and other Doctoral-awarding universities. "Non-Doctorate" are all other post-secondary institutions. "Tuition & Fees" are the average tuition and fee charged to the foreign student. "Expected Student Payment" is the average tuition and fee not funded by the post-secondary institution. "Fully funded" is the share of students who are fully funded by the post-secondary institution.

University of Central Missouri and the University of Texas at Dallas, have a high number of foreign master's-level students. While lower prices than more highly ranked institutions may be part of the attraction of these institutions for foreign students (Redden 2017), it is also the case that many of these colleges actively pay foreign recruiters; for example, Chen and Korn (2015) report that Wichita State pays foreign agents between $1,000 and $1,600 per student for recruiting.

Naturally, doctorate-level students are concentrated at research universities (Table 2), with public research universities having a somewhat higher level of foreign PhD degrees. Public research universities provide both greater scale and large programs in engineering, science, and technology fields, which tend to attract foreign students.

International students represent a higher share of students in science, technology, engineering, and mathematics programs than in any other fields at the bachelor's,

master's, and PhD levels (Figure A1 in the online Appendix available with this article at the *JEP* website). In fact, about 17 percent of all BA degrees in mathematics were awarded to temporary residents in 2017. The concentration of international students in master's programs in the fields of science, technology, engineering, and mathematics is even more remarkable: for example, in 2017 foreign students received about 62 percent of all master's degrees in computer science and 55 percent in engineering. Nonetheless, the representation of foreign students was higher in 2017 than 2002 in virtually all fields in both bachelor's and master's programs.

Visa Policy for Foreign Study in the United States

International students enter the United States on F, J, or M student non-immigrant visas: the F-1 student visa is the primary mode for full-time foreign students, J-1 visas are for exchange students and researchers, and the less frequently used M-1 visa is for those attending vocational or technical education. F-1 students must first be accepted by a US higher education institution certified by the Student and Exchange Visitor Program (SEVP), which provides the student with a certificate of eligibility for non-immigrant student status (the I-20 form). The student pays a SEVIS (Student and Exchange Visitor Information System) fee to the US Citizenship and Immigration Services ($200 in 2020) and applies for a visa at a US Embassy before entering the United States. To remain in legal status, the student must maintain a full course load but can engage in part-time work at the college or university.

The term of an F visa can be extended beyond formal academic enrollment through participation in Optional Practical Training (OPT), which allows for temporary employment related to a student's major area of study. This option provides an extended period in the United States for a foreign student to search for employment outside the constraints of a numerical visa quota. In 2008, the duration of the OPT was extended from 12 to 29 months for those in science, technology, engineering, or mathematics fields. An administrative change extended the number of designated programs from about 90 to nearly 400 in June 2012, and another change extended the term to 36 months in 2016.

Student visas differ from work visas in that they are largely unconstrained in quantity. The primary work visa for those with a college degree is the H-1B, which requires that the employee be in a specialty occupation, defined as one that requires "theoretical and practical application of a body of highly specialized knowledge and attainment of a Bachelor's or higher, or its equivalent." H-1B visas require an employer application and sponsorship, and their use in the private sector is subject to a cap (currently binding at 65,000 per year with some additional allowances) for all foreign workers except those from five exempt countries (Canada, Mexico, Chile, Singapore, and Australia).[3] While binding in the private sector, higher education institutions,

---

[3] Although the original H-1 visa did not have a numerical cap, the Immigration Act of 1990 imposed an annual cap of 65,000 visas. This total was not reached during the early 1990s, but the cap became binding in the mid-1990s. In 1999 and 2000, the cap was raised to 115,000, and then to 195,000 in 2001. This limit held until 2004, when the H-1B cap reverted to 65,000 once again. In 2004, Congress authorized,

nonprofit research organizations, and government agencies are exempt from the H-1B visa cap, providing an additional pathway to the labor market for individuals studying in the United States on F visas. As a result, students from H-1B-dependent countries became more likely to work in academic institutions when the H-1B cap became binding in 2004 (Amuedo-Dorantes and Furtado 2019).

In Figure 2, panel A presents the different types of high-skill visas and the transitions between them. Panel B shows the number of visas for each category issued between 1997 and 2018. It demonstrates the lack of numerical constraints on student visas—since the mid-2000s, student visas have increased sharply, even as the numbers dwarf the frequently debated H-1B visa program. Student visas are an important pathway into the US labor market. Yet, as the figure shows, the transition rates from student visas to work visas have steadily declined over time because even as student visas have increased, the number of new H-1B visas has stayed roughly constant.[4]

After graduating from US higher education institutions, foreign students have three primary options. First, they may enroll in a different degree program with a new F-visa, such as when continuing from a bachelor's to a master's program or from a master's to a PhD program. Second, they may start working for a US employer either through an OPT extension on the same F-visa or through a work visa, such as an H-1B. Their third option is to leave the country. Alternative options include a direct path to permanent residency as a spouse/relation of a US citizen.

Because student visas are an important stepping-stone for participating in the US labor market, changes to visa policy and the availability of H-1B opportunities will affect decisions to study in the United States (Kato and Sparber 2013). Indeed, policy adjustments such as the extension of the OPT period for F-1 degree recipients in science, technology, engineering, and mathematics fields and a rule-making change favoring US advanced degree recipients in the allocation of H-1B visas potentially affect both foreign educational investments and persistence in the US labor market (Radnofsky 2019; Amuedo-Dorantes, Furtado, and Xu 2019). In effect, growth in the student visa reservoir and the pool of students persisting with OPTs increases the supply of foreign-born college-educated workers in the United States and effectively lengthens the queue for employment visas such as the H-1B and employment-based permanent residency.

In recent years, other countries have begun competing with the United States for high-skilled immigrants by offering policies that provide somewhat flexible options for transitions to employment. Both Canada and Australia have programs which allow foreign-born graduates to stay in the country after they finish their studies. They also use these programs to feed their permanent residency point-based selection program (Moltaji 2017; Crown, Faggian, and Corcoran 2020). For example, obtaining a Canadian post-secondary educational credential generates extra points in the Express Entry

---

through the Visa Reform Act, that an extra 20,000 H1-B visas could be issued to foreign workers holding advanced degrees from US universities.

[4] Since 2016, there has been a drop in new student visas, perhaps reflecting a change in the visa renewal requirements of Chinese students as well as other global trends in the demand for higher education from abroad. In 2014, Chinese students were given an extension for their F-1 student visas, making them valid for five years instead of one.

Figure 2
Transitions from Student to Work Visas

A: Description of transition paths out of the Student Visa
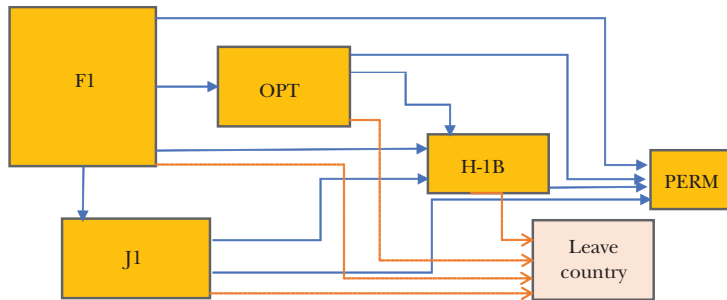
B: New Visas awarded and transition rates, 1997–2018
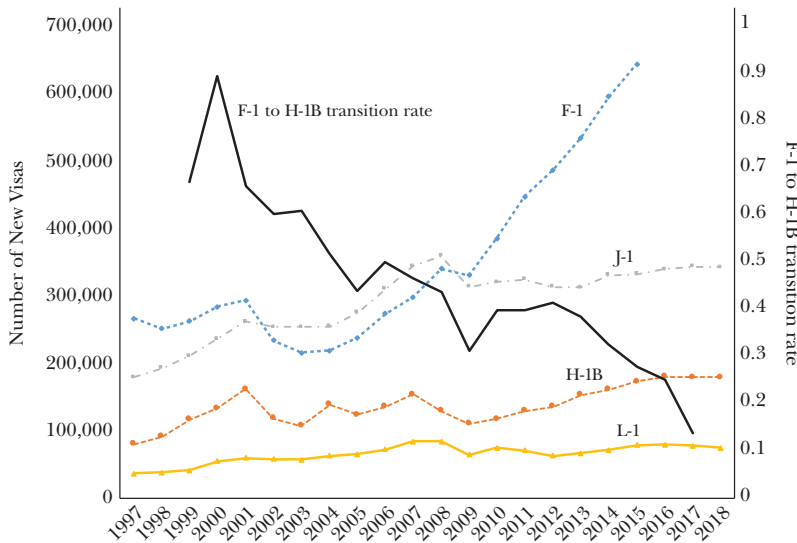
Source: United States Department of State (1997–2015)
Note: Panel A shows the transition paths for F-1 visa recipients. Panel B shows the trends over time in visas granted and transition rates from F-1 to H-1B visas. The F-1 visa is the student visa applicable to most students at certified US universities. OPT is Optional Practical Training which allows those on an F-1 visa to work for a US-based employer post graduation. The J-1 visa is for exchange students, researchers, and physicians undergoing training (including international medical graduates for medical residencies). The L-1 visa is for intra-company transfers of executives and managers. The H-1B visa is for high-skill workers in specialty occupations. PERM is applications for immigration status (green cards). The sizes of the boxes in Panel A crudely, but not accurately, depict the number of visas granted. In Panel B, the right-axis uses USCIS data to estimate the fraction of F-1 visas that converted to an (new, initial employment) H-1B visa each year. For the transition rates we use the 2000–2018 Characteristics of H-1B Specialty Workers Reports of the USCIS, and the 1999–2018 Completion Surveys by Race from the Integrated Post-Secondary Data System. The ratio of initial H-1B petitions processed to aliens in the United States to the number of foreign graduates of US universities in that class of graduation is an approximation of the transition rate from F-1 visas to H-1B for each year of graduation. We omit the F-1 visa data after 2015 because of the change of visa regime in visa renewals for Chinese students.

system used by Canada to select economic migrants (Canada Express Entry 2020). Applicants are also awarded extra points in the Australian Skilled Immigration Points Requirements if they completed their education from an Australian educational institution (Australian Department of Home Affairs 2020). Such competing immigration policies may, in turn, diminish the US advantage in attracting global talent.

The Demand for US Higher Education from Abroad

Demand for admission to US degree programs—like programs in other foreign countries—depends on the number of individuals academically prepared for post-secondary study, the availability of home country university options, individuals' capacity to pay for education abroad, and the extent to which enrollment provides a pathway to the US labor market. Because the home country supply of university opportunities tends to be fairly inelastic in the short term (particularly in the research-intensive sectors), enrollment in US degree programs is one way to satisfy demand for university education that cannot be immediately accommodated by expansion in home countries. These forces generate predictable patterns in which economic growth in a developing country fuels increases in US enrollment, with doctorate enrollment often leading undergraduate enrollment. Eventually, the country's enrollment flow to the United States (or other countries) will stabilize or decline as home-country capacity increases.

For many developing countries, the opening of education markets to the United States is a first step in a transition that includes improved diplomatic relations and the broader opening of markets to international trade. For instance, the establishment of diplomatic relations between China and the United States in 1979 dramatically increased the level of educational exchange between these two countries, particularly at the doctorate level, with similar patterns evident in the post-Cold War era for students from Eastern Europe and the former Soviet Union in the late 1980s and early 1990s tied to political changes in those countries (Blanchard, Bound, and Turner 2009). Conversely, political developments have also sometimes worked to close down foreign student enrollment (and trade more generally), as happened for students from China in the early 1950s, Hungary in the mid-1950s, and Iran after the 1979 Iranian Revolution.

Changes in educational attainment and personal incomes in developing countries have been a major driver in the overall growth in demand from abroad for US post-secondary education. For countries like South Korea, Singapore, Taiwan, China and India, the upward trend in secondary and post-secondary attainment in recent decades has been remarkable. China experienced an increase of 15 million in the number of students enrolled in secondary education between 1997 and 2017 (from about 68 million students) and an increase of 38 million in students enrolled in post-secondary education (from about 6 million) during the same period (UNESCO Institute for Statistics 2018). In 2017, India had 61 million more students enrolled in secondary education and 27 million more students enrolled in a post-secondary education than in 1997.

These dramatic growth trajectories were not matched by immediate expansion in home country university capacity, at a time in the 1980s and 1990s when opportunities for study at home country research universities comparable to highly ranked North American or European research universities were very limited. In 2003, no universities from China were among the 50 most highly ranked universities in the world, while universities from the United States occupied 39 of the top 50 spots. Today, two universities from China have entered this elite group, while US universities represent 31 of the 50 most highly ranked universities in the world (Shanghai Ranking 2020). Massive Chinese government investments in research and university education in the last two decades have produced expansion of home-country capacity for both undergraduate and graduate education: specifically, Project 985 promotes the 39 top universities in China, while Project 211 targets the top 112 universities.

At the undergraduate level, enrollment demand for US institutions from foreign students reflects the presence of types of programs rarely available in other countries, including liberal arts colleges and other broad-based programs of study, along with a greater supply of selective and resource-intensive options. Even as China and India have developed highly competitive elite universities, and their capacity has grown in the last two decades, seats are so scarce at these institutions that admission to top-ranked US colleges may be no more difficult; indeed, Najar (2011) provides evidence that some of the most qualified students in India are being crowded out of top Indian colleges.

For countries like China, enrollment in US and other foreign doctorate programs increased before the growth in enrollment in undergraduate and master's-level programs charging tuition. The upward trajectory in doctorate enrollment started in the 1980s even as the growth in undergraduate enrollment did not escalate until the 21st century (Bound, Turner, and Walsh 2009). In the 1980s and 1990s, US universities awarded more PhDs to students from China than did Chinese universities. Because foreign doctoral students commonly receive full support in the form of fellowships and teaching assistantships, their enrollment is often less sensitive to home country income.

Over the past 20 years, a substantial number of households in developing countries have experienced increasing income levels, which provide them with the capacity to pay for US higher education (Bird and Turner 2014). For instance, Bound et al. (2020) estimate that the fraction of Chinese families with incomes greater than the average amount charged by US public universities for out-of-state tuition and room and board grew from 0.005 percent in the year 2000 to more than 2 percent by 2013. This growth in the ability of Chinese families to pay for a US education in the first part of the 21st century allowed US universities to enroll increasing numbers of qualified full-fare paying students from abroad, particularly at the undergraduate and master's levels.

Chinese cities experiencing the largest income growth induced by increased goods exports were among those with the greatest outflow of international students to US universities (Khanna et al. 2020). In effect, Chinese families in locally booming economies used some of their new trade-liberalization driven wealth to send students abroad. This response of flows to income growth was strong for students

at the undergraduate and master's level and not detectable at the doctorate level. Accordingly, the response was also strongest among self-funded students. This is not only a demonstration of the effects of income growth on US enrollment but also demonstrates how the US trade deficit in goods with China partially cycled back as an export surplus in higher education services.

Beyond (potential) access to post-secondary options unavailable in their home countries, obtaining a US degree provides the advantage of potentially easier access to US employment options (Rosenzweig 2006; Bound et al. 2014).[5] Because most students graduating from a US university are eligible for an extension of their visas with the pursuit of Optional Practical Training, they are able to gain employment in the US labor market without needing an H-1B visa in the supply-constrained lottery. In addition, obtaining a degree from a US college or university may provide advantages for foreign students searching for jobs over those educated abroad, to the extent that US employers have more information on skills acquired at familiar educational institutions, and employers might find it more straightforward to interview candidates on-site. Moreover, the likelihood that foreign students stay in the United States after finishing their studies is also a function of economic conditions in their home countries. Generally, students from higher-income nations are less likely to convert their student status to OPT than students from lower-income countries, as shown in Figure 3. For example, about 65 percent of all US bachelor's graduates from India switched to OPT in 2015, while only 28 percent of Canadian graduates switched to OPT over the same period.

In turn, as the option value of pursuing employment in the US changes, we would expect enrollment demand from abroad to adjust. Using the number of takers of the Graduate Record Examination (GRE) as a measure of students' willingness to come to the United States for graduate education, for most of the past two decades, the demand from Indian students for a US education has been higher when US unemployment rates are low. Nonetheless, there has been a significant drop in the number of Indians taking the GRE since 2016, a period with several years of low unemployment rates. One potential explanation for this is a perception of less willingness of the United States to welcome immigrants after the 2016 election, which makes the United States less attractive to international students (Anderson and Svrluga 2018).

The Supply Side: How US Colleges and Universities Benefit from Foreign Students

US colleges and universities seek talent and resources from international students. The relative importance of academic skills and capacity to pay varies markedly by degree level and type of university.

---

[5] Amuedo-Dorantes, Furtado, and Xu (2019) find evidence that when the OPT policy was changed in 2008 to extend the time for science, technology, engineering, and mathematics graduates to stay in the United States after graduation, it induced an increasing number of foreign students to major in these fields.
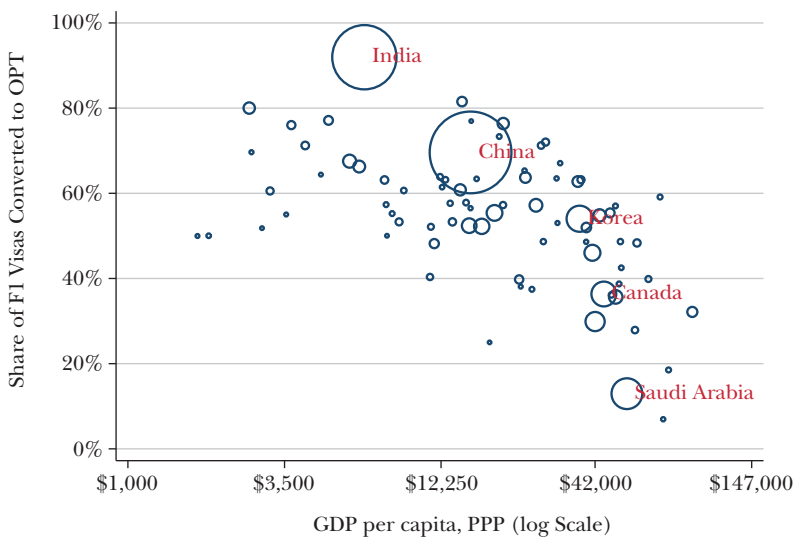
Figure 3

Share of F1 Visas Converted to Optional Practical Training by Country's Per Capita GDP

A: Bachelor's Degree

B: Master's Degree

Note: Size of each bubble is proportional to the number of foreign graduates in 2015. Sample is restricted to countries with more than 50 graduates in 2015.

For most doctoral students, capacity to pay is secondary (and often irrelevant), as admission to many PhD programs is accompanied by full tuition waivers and guaranteed living expenses for multiple years. At the other extreme, many master's degree and undergraduate programs have quite modest academic requirements and can attract foreign students who are able to pay tuition levels that relatively few US students are willing to pay in full. Occupying a middle ground are selective colleges and universities that face excess demand for undergraduate programs along with professional programs like the MBA. These institutions, often competing on quality, see both academic talent and capacity to pay undiscounted or out-of-state prices among foreign students as inputs in their objective functions.

University admission offices typically employ a variety of recruitment strategies to attract international students, ranging from utilizing social media to traveling abroad to meet with students, parents, counselors, and alumni at schools and education fairs. In addition, several institutions started experimenting with commission-based recruitment agents in the past years. If contracted by colleges, these commissioned agents are paid on a per capita basis. As of 2017, 38.5 percent of US colleges and universities reported using commission-based recruitment agents—up from 30 percent in 2010 (National Association for College Admission Counseling 2018).

Revenue Implications

It would be naïve to understate the revenue implications of foreign students for US colleges and universities. In 2019, foreign students poured nearly $44 billion into the US education system (Bureau of Economic Analysis 2020). The revenue implications from tuition differ markedly by degree level, as shown in Table 2. BA and MA students rarely have a "free ride," while it is quite common for doctorate-level students to have university resources cover tuition and fees. International undergraduate students pay nearly 96 percent of tuition costs from personal or home country sources at public research universities.

Among undergraduate students at private universities, the share of tuition paid by individuals is somewhat smaller but the levels are higher, reflecting the higher tuition prices at private institutions. Two different factors yield a modest wedge between the sticker price and what students pay for foreign students at the undergraduate level: first, a modest number of very wealthy private institutions like Princeton University provide some financial aid for foreign undergraduates, and second, somewhat less selective private universities regularly engage in "discounting," which refers to offering need-based financial aid to increase enrollment (Bowen and Breneman 1993).

Foreign students studying at the MA level represent a significant source of revenue in both the public and private sectors of higher education, representing $3 billion and $4.3 billion in revenues respectively. Although top research universities have the largest numbers of master's students and are able to extract the highest prices (net tuition revenue of $39,858 on a posted tuition and fee level averaging $45,512), there are also many full-pay foreign students outside this tier in the public and private sectors..

The importance of MA-level training for foreign students as a revenue source for universities has increased markedly in recent years, with a number of universities

adding revenue-generating programs precisely to cater to foreign students. For example, the number of master's programs in the United States where 80 percent or more of the graduates were foreign rose slowly, from 354 in 2000 to 512 by 2012.[6] But with declining state appropriations and stagnant federal funding for science over the last decade (National Science Board 2020, Figure 5B-4), the number of such programs leaped, reaching nearly 1,000 by 2016 (Education Data Portal 2020). The growth has primarily been in science, technology, engineering, and mathematics programs, where demand from full-pay students from abroad may cross-subsidize doctorate programs.

In an accounting sense, doctoral programs are cost drivers, not sources of revenue generation, and this pattern evidently holds true for foreign students as well as domestic students. Using data on total cost of attendance (which includes living expenses), universities in the private and public sectors make a substantial investment in foreign doctoral students. For 2015, average total expenses (tuition and living expenses) for doctorate students at top private research universities were nearly $62,000, with funds from universities averaging $55,572 (about 90 percent of the total); at public research universities, the comparable numbers are $39,803 in total expenses, with $34,396 funded from universities (although some of what appears as university funding may reflect grant funding from federal or private sources allocated at the university level). Foreign doctorate students contribute not just to the research and teaching enterprise but also to university prestige, particularly at universities looking to establish their competitive research credentials. Moreover, the presence of doctorate programs may help recruit and retain research-active faculty who are likely to gain from the capacity to teach small courses tied to research specialties (Courant and Turner 2019).

Consequences of International Student Flows

Dramatic increases (or decreases) in foreign student flows may have implications not only for the university sector, but also for labor markets and the broader economy in both the sending and receiving countries. A body of work examines such consequences, most notably focused on the consequences on the native-born, innovation, and higher-education institutions in the United States.

One obvious question is whether the spillovers of international flows of students on native students is beneficial. On the negative side, there is some indication that foreign PhD students "crowd out" domestic students (Borjas 2007), presumably in contexts where there are capacity constraints on enrollment along with an excess demand for slots among domestic students. In addition, there is some indication that at the undergraduate level, the concentration of international students in certain majors like business, economics, or science and engineering may dilute per-student resources or lead local students to concentrate in other fields (Anelli, Shih, and Williams 2017). Some suggest that growth in foreign students may have

---

[6]These calculations are restricted to programs with at least five students.

generated institution-level administrative challenges, while others have questioned how well foreign students are integrated at US universities (Jordan 2015; Redden 2014; Gareis 2012).

Yet much evidence also points to the potential cross-subsidization of native students. International students are an important source of revenue for public research universities facing declining state appropriations (Bound et al. 2020). Without a ready supply of foreign students, such universities would have had to navigate reductions in instructional resources per student or substantially raise in-state tuition. Such cross-subsidization may also be present in graduate programs, specifically in terms of revenue from master's programs (Shih 2017). Revenue-generating master's programs are not only more likely to charge full sticker price than subsidized PhDs, but they are also relatively more elastic in their supply. Many large research institutions now draw as much as 20 percent of their tuition revenue from foreign students (Larmer 2019). Universities may invest in programs and centers better aligned with the demands of foreign, rather than local, students. Yet, the revenue from international students may also help institutions better cater to the preferences of local students. Of course, this pattern also makes these institutions more dependent on foreign flows, which will fluctuate in response to global political crises, home-country economies, growth in home-country institutional quality, and competition from other developed economies like Canada, the United Kingdom, and Australia.

International student flows also help generate a ready supply of high-skill foreign workers for the US labor market. The OPT program mentioned earlier allows students between one and two years of additional labor market experience in the United States post-graduation, and the stringently capped high-skill H-1B program has a separate category of 20,000 visas that makes it easier for foreign citizens who have a graduate degree from US universities. Such features help facilitate the transition to the US workforce. In turn, the pool of foreign students considering US employment facilitates matches and reduces the monopsony power employers have over their foreign workers.[7]

Foreign students looking for work in the United States are likely to have spillover effects on US students for a combination of reasons. For example, the presence of foreign students who may be willing to accept a lower wage may disadvantage US-born students. There is reason to believe that had firms not been able to hire H-1B workers, the wages of US computer scientists would be even higher than they are (Bound et al. 2015; Bound et al. 2013). As a result of constrained wages, fewer Americans may decide to pursue fields likely dominated by foreign graduates.

Additionally, some limited and anecdotal evidence exists that the expansion of Optional Practical Training combined with the potential for limited employment while enrolled has contributed to the rise of fraudulent post-secondary institutions. One such example is the case of Tri-Valley University in California, which appeared

---

[7]Much of the criticism leveled at the H-1B program involves the limited mobility H-1B workers have. Similarly, to be eligible for the OPT extensions, foreign students need to find jobs within two months of finishing their degree.

to have nearly 1,000 students enrolled on F-1 visas in May 2010, with 185 listed as pursuing a doctorate degree in Computer Science. The institution was shut down in 2011 by Immigration and Customs Enforcement with the founder charged with fraud and money laundering.[8] While cases of outright fraud are likely to represent a very small share of the utilization of OPT and the extended provisions associated with degrees in science, technology, engineering, and mathematics, there is a legitimate question of how the extension of the OPT from 12 to 29 months in 2008 for recipients in these fields affected outcomes in the United States. Demirci (2015) finds increases in the incidence and duration of persistence in the United States for F-1 visa recipients at least in the immediate period after degree completion, with these effects particularly marked for master's degree recipients.

While the rates at which foreign students stay in the United States after receiving their degrees are difficult to measure at the undergraduate and master's level, the five-year stay rate for doctorate recipients exceeds 70 percent and is higher for those from China (84 percent) and India (86 percent) than from other countries (authors' calculations using the Survey of Earned Doctorates). Similarly, at the PhD level, Finn and Pennington (2018) find that 10-year stay rates (2002–03 to 2013) were highest among students from China and India (85 percent), with students from South Korea, Europe, and the Americas less likely to stay. Yet stay-rates for doctoral students are unlikely to translate to other levels of degree receipt; as Figure 2 showed earlier, given the capped nature of work visas and the rising number of international students, the transition rates from F-1 student to H-1B work visas have been steadily declining.

Transitions to the US workforce are often concentrated in high-skill sectors, such as information technology and health care (Bound et al. 2014). Foreign workers may help facilitate innovation and production by allowing firms to draw from a large pool of global talent abroad (Kerr 2018). Indeed, a number of studies have identified the outsized role played by immigrants in science and engineering innovation, including elite settings like membership in the National Academy of Sciences, Nobel prize receipt, and authorship of very highly cited papers (for example, Chellaraj, Maskus, and Mattoo 2008; Black and Stephan 2010; Stuen, Mobarak, and Maskus 2012; Gaulé and Piacentini 2013). Immigrants have also played prominent roles in tech entrepreneurship (Anderson and Platzer 2006; Saxenian 2000; Wadhwa et al. 2007). But not all immigration in the tech field is concentrated in the tail end of the distribution of innovation and productivity; for example, using patent data, Hunt and Gauthier-Loiselle (2010) find that, conditional on occupation, immigrants are equally likely to innovate as US-born workers.

---

[8] A report in the *Chronicle of Higher Education* suggests, "Tri-Valley is only the beginning. Other colleges—most of them unaccredited—exploit byzantine federal regulations, enrolling almost exclusively foreign students and charging them upward of $3,000 for a chance to work legally in the United States. They flourish in California and Virginia, where regulations are lax, and many of their practices—for instance, holding some classes on only three weekends per semester—are unconventional, to say the least. These colleges usher in thousands of foreign students and generate millions of dollars in profits because they have the power, bestowed by the US government, to help students get visas" (Bartlett, Fischer, and Keller 2011).

Although it is relatively straightforward to enumerate the contributions of skilled immigrants educated in the United States, assessing their overall effects on the US economy involves evaluating counterfactuals. If there is some crowd-out of US-born workers, then enumerating the contributions of skilled immigrants will exaggerate their net contributions. On the other hand, if the crowd out is less than total, as would seem likely in most cases, then the net contribution will be positive. Outside specific contexts, accurately gauging magnitudes is probably not possible.

Labor market opportunities may also have substantial impacts on home economies. Sending countries may experience "brain drain" as bright minds move abroad. On the other hand, the potential to migrate abroad may encourage the foreign-born to acquire skills (such as undergraduate engineering degrees) that are valued abroad. Such a "brain gain," combined with return migration at a later time, may facilitate the shifting of knowledge and production to home countries (Khanna and Morales 2019). Indeed, PhDs trained in the United States and other western countries may have fostered the growth of tertiary education and scientific research in a range of counties (Kahn and MacGarvie 2016).

Evaluating the effect that the availability of foreign students interested in and capable of attending US universities has on these institutions and the US economy more generally will often involve important feedback effects between the educational sector and the rest of the economy, as we have emphasized. For instance, changes in the US H-1B program are likely to have significant effects on the demand for education by both foreign and domestic students. Further, US immigration policy interacts with other features of the US economy including, for example, state higher education funding decisions. Evaluating the relevant counterfactuals essentially involves working implicitly or explicitly with general equilibrium models.

Moving Forward: The Future of the Higher Education Sector

The flow of foreign student revenues and talent from abroad has had a substantial impact on US higher education in recent decades. As such, market forces, political crises and the COVID-19 pandemic that can destabilize these flows are likely to result in reduced resources for US higher education institutions, with such shocks reverberating to the economy more broadly. To that end, universities have started taking precautionary measures like consciously diversifying their portfolio of origin countries, and even taking out insurance policies to cover themselves against losses to foreign-student revenue (Bothwell 2018).

Political concerns following the escalation of US-China trade relations in 2018, along with the handling of the pandemic that erupted in 2020, may curb the flow of foreign students from abroad. Khanna et al. (2020) estimate that if the US-China trade war continues, it could cost US universities about 30,000 Chinese students or $1.15 billion in revenue over the next 10 years. This loss, which would be about 8 percent of educational service exports to China, is likely an underestimate of overall economic losses for the US economy as it does not account for broader effects on local economies surrounding universities. More generally, changes to the

likelihood of obtaining a US work visa may discourage many students who were looking at US education as a stepping-stone to the labor market. For the first time in many decades, new foreign undergraduate enrollment has declined.

At the same time, universities in other parts of the world have become global players in this market and threaten the dominance of the US position in attracting foreign students. In particular, the expansion of home-country higher education capacity may keep students back in China or India. In India, the expansion of numerous Institutes of National Importance may stem the outflow of bachelor's students. These Institutes are primarily teaching-based, but do produce a stream of high-quality students ready for graduate programs.[9]

China has recently increased investments in both the instructional and research capacity of their higher education institutions. One of the most prominent global rankings for universities is the Academic Ranking of World Universities originated by China's Shanghai Jiao Tong University and thus commonly known as the Shanghai rankings. According to these rankings, the US share of the world's top 500 research universities fell about 7 percentage points from 2004 to 2018, from 35 to 28 percent, while the share of Chinese research universities in the top 500 accounted for most of this change by rising 8 percentage points from 2 to 10 percent (Appendix Figure A2). On the margin of the top 500, Chinese universities are "overtaking" some lower mid-tier US institutions. Such changes may affect the future flow of students from abroad. This, in turn, would affect the size of the science and technology workforce produced by and working in the United States, and the corresponding location of innovation and economic activity.

US universities are experiencing drastic revenue shortfalls in the second half of 2020 and beyond as a result of the COVID-19 crisis. For instance, these near-term losses are projected to be $250 million at the University of Delaware (as reported in Flaherty 2020), $300 million at the University of North Carolina Chapel Hill (Murphy 2020), and $500 million for university system in Maryland (Condon 2020). Possible in-person enrollment reductions in the summer and fall, and tightening visa and mobility restrictions, tend to exacerbate these shortfalls. As such, universities most reliant on foreign enrollment may be most adversely affected. These include schools in the University of California system, and large Midwestern universities, such as Purdue, University of Illinois at Urbana-Champaign, and Michigan State, all of which enroll a relatively large number of their incoming first-year undergraduates from China (Bound et al. 2020).

Visas for the academic year are usually granted between March (when admissions decisions are made) and September (when semesters begin). Between 2017 and 2019, about 290,000 visas were granted each year over these seven months (United States Department of State 2020). Between March and September 2020, only 37,680 visas were granted—an extraordinary drop of 87 percent. Visas for students from China dropped from about 90,000 down to only 943 visas between March and September 2020. A fall 2020 survey of 700 higher education institutions found that

---

[9] Institutes of National Importance specialize in both undergraduate and post-graduate education in technical fields like medicine, information technology, sciences, engineering, architecture or business.

one in five international students were studying online from abroad in response to the COVID-19 pandemic. Overall, new international enrollment (including those online) decreased by 43 percent, with at least 40,000 students deferring enrollment (Baer and Martel 2020). Not only does the absence of international students from campuses in the 2020–21 academic year impact college-town economies and university dorm revenues, but the disruption in the flow of new international enrollment may have lasting impacts on university finances and academic outcomes.

References

Amuedo-Dorantes, Catalina, and Delia Furtado. 2019. "Settling for Academia? H-1B Visas and the Career Choices of International Students in the United States." *Journal of Human Resources* 54 (2): 401–29.

Amuedo-Dorantes, Catalina, Delia Furtado, and Huanan Xu. 2019. "OPT Policy Changes and Foreign Born STEM Talent in the US." *Labour Economics* 61.

Anderson, Nick, and Susan Svrluga. 2018. "What's the Trump Effect on International Enrollment? Report Finds New Foreign Students Are Dwindling." *The Washington Post,* November 13.

Anderson, Stuart, and Michaela Platzer. 2006. "American Made: The Impact of Immigrant Entrepreneurs and Professionals on US Competitiveness." Washington, DC: National Venture Capital Association.

Anelli, Massimo, Kevin Shih, and Kevin Williams. 2017. "Foreign Peer Effects and STEM Major Choice." IZA Discussion Paper 10743.

Australian Department of Home Affairs. 2020. *Points Table for Skilled Independent Visa (Subclass 189).* Canberra: Government of Australia.

Baer, Julie, and Mirka Martel. 2020. *Fall 2020 International Student Enrollment Snapshot.* New York: Institute of International Education.

Bartlett, Tom, Karin Fischer, and Josh Keller. 2011. "Little-Known Colleges Exploit Visa Loopholes to Make Millions Off Foreign Students." *The Chronicle of Higher Education*, March 20. https://www.chronicle.com/article/Little-Known-Colleges-Make/126822.

Bird, Kelli, and Sarah Turner. 2014. "College in the States: Foreign Student Demand and Higher Education Supply in the U.S." https://curry.virginia.edu/sites/default/files/uploads/resourceLibrary/23_Bird_Foreign_Student_Demand.pdf.

Black, Grant C., and Paula E. Stephan. 2010. "The Economics of University Science and the Role of Foreign Graduate Students and Postdoctoral Scholars." In *American Universities in a Global Market*, edited by Charles T. Clotfelter, 129–61. Chicago: University of Chicago Press.

Blanchard, Emily, John Bound, and Sarah Turner. 2009. "Opening (and Closing) Doors: Country-Specific Shocks in U.S. Doctorate Education." Population Studies Center Report 09-674.

Borjas, George J. 2007. "Do Foreign Students Crowd Out Native Students from Graduate Programs?" In *Science and the University*, edited by Paula E. Stephan, and Ronald G. Ehrenberg. Madison: University of Wisconsin Press.

Bothwell, Ellie. 2018. "Insuring Against Drop in Chinese Students" *Inside Higher Ed,* November 29. https://www.insidehighered.com/news/2018/11/29/university-illinois-insures-itself-against-possible-drop-chinese-enrollments.

Bound, John, Breno Braga, Joseph M. Golden, and Gaurav Khanna. 2015. "Recruitment of Foreigners in the Market for Computer Scientists in the United States." *Journal of Labor Economics* 33 (S): S197–223.

Bound, John, Breno Braga, Joseph M. Golden, and Sarah Turner. 2013. "Pathways to Adjustment: The Case of Information Technology Workers." *American Economic Review* 103 (3): 203–07.

Bound, John, Breno Braga, Gaurav Khanna, and Sarah Turner. 2020. "A Passage to America: University Funding and International Students." *American Economic Journal: Economic Policy* 12 (1): 97–126.

Bound, John, Murat Demirci, Gaurav Khanna, and Sarah Turner. 2014. "Finishing Degrees and Finding Jobs: U.S. Higher Education and the Flow of Foreign IT Workers" *Innovation Policy and The Economy* 15 (2015).

Bound, John, Sarah Turner, and Patrick Walsh. 2009. "Internationalization of U.S. Doctorate Education." In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, edited by R. B. Freeman and D. L. Goroff, 59–98. Chicago: University of Chicago Press.

Bowen, William G., and Breneman, David W. 1993. "Student Aid: Price Discount or Educational Investment?" *The Brookings Review* 11 (1).

Bureau of Economic Analysis. 2020. "International Transactions, International Services, and International Investment Position Tables." International Data. https://apps.bea.gov/iTable/iTable.cfm?reqid=62&step=6&isuri=1&tablelist=245&product=4. (accessed November 1, 2020).

Canada Express Entry. 2020. Canada Express Entry Eligibility. https://www.canadaexpressentry.org/eligibility.php. (accessed November 1, 2020).

Chellaraj, Gnanaraj, Keith E. Maskus, and Aaditya Mattoo. 2008. "The Contribution of International Graduate Students to US Innovation." *Review of International Economics* 16 (3): 444–62.

Chen, Te-Ping, and Melissa Korn. 2015. "American Colleges Pay Agents to Woo Foreigners, Despite Fraud Risk" *Wall Street Journal*, September 30. https://www.wsj.com/articles/american-colleges-pay-agents-to-woo-foreigners-despite-fraud-risk-1443665884.

Condon, Christine. 2020. "University System of Maryland chancellor takes pay cut, warns employees could 'share in the pain' of pandemic" *The Baltimore Sun*, September 9. https://www.baltimoresun.com/education/bs-md-university-system-losses-chancellor-pay-cut-20200909-fmqwssllkncvjer22m37e3niiq-story.html.

Courant, Paul N., and Sarah Turner. 2019. "Faculty Deployment in Research Universities." In *Productivity in Higher Education*, edited by Caroline M. Hoxby and Kevin Stange, 177–208. Chicago: University of Chicago Press.

Crown, Daniel, Alessandra Faggian, and Jonathan Corcoran. 2020. "Foreign-Born Graduates and Innovation: Evidence from an Australian Skilled Visa Program." *Research Policy* 49 (9).

Demirci, Murat. 2015. "OPT and The Impact of High Skill Immigration on Career Decisions of College-Educated Natives." Unpublished.

Education Data Portal. 2020. "Version 0.9.0, Urban Institute, Center on Education Data and Policy." https://educationdata.urban.org/documentation/. (accessed April 1, 2020).

Finn, Michael, and Leigh Ann Pennington. 2018. *Stay Rates of Foreign Doctorate Recipients from U.S. Universities, 2013.* Oak Ridge, TN: Oak Ridge Institute for Science and Education.

Flaherty, Darryl. 2020. "The Cost of COVID-19 for the University of Delaware" *Delaware Online*, October 3. https://www.delawareonline.com/story/opinion/2020/10/03/cost-covid-19-university-delaware/5893911002/.

Gareis, Elisabeth. 2012. "Intercultural Friendship: Effects of Home Region and Sojourn Location." *Journal of International and Intercultural Communication* 5: 309–28.

Gaulé, Patrick, and Mario Piacentini. 2013. "Chinese Graduate Students and US Scientific Productivity." *Review of Economics and Statistics* 95 (2): 698–701.

Ghose, Devaki, and Sarah Turner. 2020. "Enrollment as Access to Employment: Evidence from Foreign Student Enrollment in CS." Unpublished.

Hechinger, Fred M.. 1979. "Iranians' Plight Puts a Spotlight on U.S. Colleges." *New York Times*, February 20. Section C, Page 4.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2 (2): 31–56.

Jordan, Miriam. 2015. "International Students Stream into U.S. Colleges." *Wall Street Journal*, March 24. http://www.wsj.com/articles/international-students-stream-into-u-s-colleges-1427248801.

Kahn, Shulamit, and Megan MacGarvie. 2016. "Do Return Requirements Increase International Knowledge Diffusion? Evidence from the Fulbright Program." *Research Policy* 45 (6): 1304–22.

Kato, Takao, and Chad Sparber. 2013. "Quotas and Quality: The effect of H-1B Visa Restrictions on the Pool of Prospective Undergraduate Students from Abroad." *The Review of Economics and Statistics* 95 (1): 109–26.

Kerr, William. 2018. *The Gift of Global Talent: How Migration Shaped Business, Economy, & Society.* Stanford,

CA: Stanford University Press.

Khanna, Gaurav, and Nicolas Morales. 2019. "The IT Boom and Other Unintended Consequences of Chasing the American Dream." http://www-personal.umich.edu/~moralesn/Khanna%20Morales.pdf.

Khanna, Gaurav, Kevin Shih, Ariel Weinberg, Mingzhi Xu, and Miaojie Yu. 2020. "Trade Liberalization and Chinese Students in US Higher Education." Unpublished.

Larmer, Brook. 2019. "One of America's Most Vital Exports, Education, Never Goes Abroad, but It Still Faces Threats." *The New York Times*, January 3. https://www.nytimes.com/2019/01/03/magazine/one-of-americas-most-vital-exports-education-never-goes-abroad-but-it-still-faces-threats.html.

Moltaji, Golbon. 2017 "Exploring the Sensory Terrain of Mobility among Graduates Who Immigrate to Canada via Post Graduate Work Permit Program" PhD diss. proposal. University of Ottawa.

Murphy, Kate. 2020. "More Employees See Cuts as NC Universities Face Sharp Financial Losses Due to COVID-19." *The News & Observer*, September 24. https://www.newsobserver.com/news/local/education/article245916490.html.

Najar, Nida. 2011. "Squeezed Out in India, Students Turn to U.S." *New York Times*, October 13. https://www.nytimes.com/2011/10/14/world/asia/squeezed-out-in-india-students-turn-to-united-states.html.

National Association for College Admission Counseling. 2018. "Trends in Commission-Based International Student Recruitment." https://www.nacacnet.org/globalassets/documents/knowledge-center/international/agent-series-_part-1-cr.pdf.

National Center for Education Statistics. 2018. Integrated Postsecondary Education Data System Enrollment. "Integrated Postsecondary Education Data System Enrollment." https://nces.ed.gov/ipeds/. (accessed April 1, 2020).

National Science Board. 2020. National Science Foundation, Science & Engineering Indicators. https://ncses.nsf.gov/pubs/nsb20202/. (accessed April 1, 2020).

Radnofsky, Louise. 2019. "U.S. Changes Visa Process for High-Skilled Workers." *Wall Street Journal*, January 31. https://www.wsj.com/articles/u-s-changes-visa-process-for-high-skilled-workers-11548879868.

Redden, Elizabeth. 2014. "Integrating International Students, So You Have X Number of International Students on Your Campus. So What?" *Inside Higher Ed*, February 20. https://www.insidehighered.com/news/2014/02/20/gathering-senior-international-educators-integration-international-students-was.

Redden, Elizabeth. 2017. "Boom in Indian Enrollments, Followed by Bust." *Inside Higher Ed*, October 30. https://www.insidehighered.com/admissions/article/2017/10/30/after-huge-surge-enrollment-india-central-missouri-sees-substantial.

Rosenzweig, Mark. 2006. "Global Wage Differences and International Student Flows." *Brookings Trade Forum* 57–86.

Saudi Arabian Cultural Bureau. 2014. "KASP Scholarship Program." https://www.saudibureau.org/en/inside.php?ID=16. (accessed April 1, 2020).

Saxenian, AnnaLee. 2000. *Silicon Valley's New Immigrant Entrepreneurs*. Santa Clara, CA: Public Policy Institute of California.

Shanghai Ranking, Consultancy. 2020. "Academic Ranking of World Universities 2018." http://www.shanghairanking.com/ARWU2018.html. (accessed April 1, 2020).

Shih, Kevin. 2017. "Do International Students Crowd-out or Cross-subsidize Americans in Higher Education?" *Journal of Public Economics* 156: 170–84.

Stuen, Eric T., Ahmed Mushfiq Mobarak, and Keith E. Maskus. 2012. "Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in US doctoral Programmes." *The Economic Journal* 122(565): 1143–76.

UNESCO Institute for Statistics. 2018. "Education Data." UIS database. http://data.uis.unesco.org/. (accessed April 1, 2020).

United States Department of State. 2020. *Monthly Non-Immigrant Visa Issuances*. Washington, DC: United States Department of State.

United States Citizenship and Immigration Services. 2015. "F-1 visa recipients (2004–2015)." obtained through a Freedom of Information request.

Wadhwa, Vivek, AnnaLee Saxenian, Ben Rissing, and Gary Gereff. 2007. "America's New Immigrant Entrepreneurs." https://www.immigrationresearch.org/report/duke-university/americas-new-immigrant-entrepreneurs-part-i.

World Bank. 2020. "GDP per capita, PPP (current international $)." *World Development Indicators database. The World Bank Group*. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD (accessed April 1, 2020.

Why Does the United States Have the Best Research Universities? Incentives, Resources, and Virtuous Circles

W. Bentley MacLeod and Miguel Urquiola

The educational performance of nations is not uniform across different dimensions, nor is it static. Around 1875, the United States had substantially higher levels of schooling than any European country, but that gap has now largely disappeared (Goldin and Katz 2008). By contrast, circa 1875 the United States had none of the world's leading research universities; today, it accounts for the majority of the top-ranked (Cole 2009; Clotfelter 2010).

One way to illustrate the rise of US research universities is by counting the times universities are mentioned in the biographies of Nobel prize winners—the idea being that universities that train or host Nobel winners are likely places that are active in research. These mentions are associated with, for example, the year a winner graduated from one institution or became a professor at another. One can use this to create a time series reaching back into the 1800s. Figure 1 shows that by this measure, around 1875, American universities badly lagged behind their European counterparts; today they are distinctly ahead.

Many observers cite events surrounding World War II as producing the turning point in American performance—for example, they highlight the migration of Jewish academics from Germany to the United States, and the rise of federal research funding (Graham and Diamond 1997; Cole 2009; Gruber and Johnson 2019). The emphasis on World War II is analytically attractive; it highlights factors that surely

■ *W. Bentley MacLeod and Miguel Urquiola are both Professors of Economics, Columbia University, New York, New York. They are also both Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts.*

Figure 1
University Nobel Prize Mentions, by Country
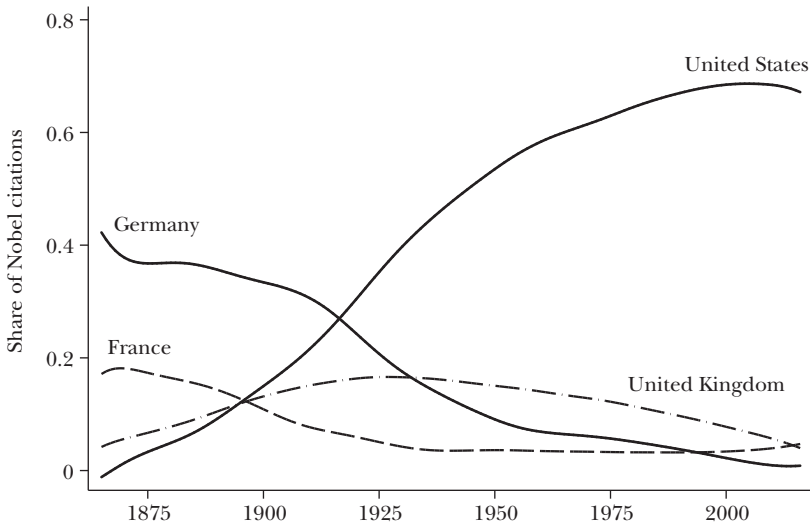
Source: Urquiola (2020).
Note: This figure describes the frequency with which Nobel winners' biographies mention universities in different countries. Specifically, the dates in the figure refer to the years in which laureates graduated from a given school if they were students there (this accounts for why there are dates for the late 1800s, before the prize was granted). If they were professors, the dates refer to the year of award for the university they were affiliated with at that point, and the year of joining for other schools. The figure plots fitted values of locally weighted regressions of each country's share of total mentions on the year. The data cover the four countries with universities accounting for the most mentions. They do not adjust for income or population because that does not change the two central facts the paper addresses: American universities' initial weakness and later leadership. For such adjustments, see Urquiola (2020).

strengthened American universities, and it addresses the relative German decline apparent in Figure 1.[1]

However, the raw data behind Figure 1 indicate that US universities had matched or surpassed most countries' well before World War II. By the measure used, they were ahead of all but Germany's by 1910, and of all countries' by 1920. To cite one anecdote illustrating this timing, in 1901, Theodore Richard—who went on to win a Nobel Prize in Chemistry—became the first American-trained professor to be offered a position at a German university. In short, a successful explanation of American universities' research dominance must begin in the late 1800s and must involve factors other than mid-20th century events in Europe.

---

[1] Analyses of recent university performance often rely on bibliographic databases. Such data have advantages but are of limited use here because databases like the Web of Science only have author affiliation information going a few decades back (for discussion, see Urquiola 2020).

The explanation we offer highlights reforms that began after the US Civil War and enhanced the incentives and resources the American university system directs at research.[2] Understanding these reforms benefits from two observations. First, the United States takes a relatively free market approach to higher education; for instance, it allows easy entry by schools seeking to satisfy customer demand. Second, universities do not supply their customers with a single "product"; they provide a complex set of services, among which is sorting. When a school produces sorting, it implicitly "sells" peer groups—a chance to be exposed to and associated with certain types of people. For example, some schools cater to students who are smart, wealthy, or artistic.

We begin our review of the history of US university research by noting that from colonial times and into the 1800s, American households demanded denominational sorting. For example, Presbyterians wished to attend college with Presbyterians, Episcopalians with Episcopalians, and so on. Further, most students preferred colleges close to home. Together, these factors produced massive entry; while the United States had nine colleges in 1776, about 900 more opened before the Civil War.

Mass entry meant that early US colleges tended to be small and underfunded. They offered a basic, narrow, and rigid curriculum delivered by unspecialized and often poorly paid professors. Colleges routinely hired faculty using criteria other than expertise and provided them with few incentives or resources to do research. It is not surprising that, on average, colleges did not excel in this dimension.

We then discuss how reforms allowed the system to begin providing incentives and resources. First, we sketch an agency model that asks how universities can incentivize professors to do research. It highlights that even if an institution is committed to research, producing this output benefits from reasonably precise research performance measures. Precision is enhanced by academic specialization—in this case, by organizing professors into disciplines.

We then describe how American colleges and universities gained an interest in research and increased specialization. To preview, during the last few decades of the 1800s, US colleges let a gap emerge between the skills they taught and those their customers demanded. With industrialization, interest grew around areas that the college curricula essentially ignored, like engineering and business. In the 1870s and 1880s, new universities like Cornell and Johns Hopkins showed that one could attract students by offering specialized and advanced instruction in a range of areas, rather than by supplying denominational sorting. Incumbents like Harvard and Columbia responded forcefully, forming specialized arts and sciences departments and creating professional schools.

Free entry allowed other schools to join the fray. This process was helped by private donations, which propelled universities like Chicago and Stanford. It was also aided by public resources, which were crucial for entrants like the University of California at Berkeley and the Massachusetts Institute of Technology.

---

[2] In emphasizing earlier events, our approach is closer to Veysey (1965) and Goldin and Katz (1999). We draw on Urquiola (2020).

All these schools began to seek professors who were specialists and therefore, researchers at the frontier of a field. The academic system responded by providing tools to measure research output, like specialized journals. Schools thus gained the ability to identify, bid for, and begin to recruit successful researchers.

These years also saw some American universities acquire tremendous amounts of resources. In the early 1900s, the schools implementing reforms sought and achieved rapid enrollment growth. However, they found that this threatened the match between the sorting they provided and that which key students demanded. In particular, enrollment growth brought in lower-income students and Jewish students, which—in a period of rising anti-Semitism—alienated the Protestant elites the colleges had traditionally served. By the early 1920s, Columbia implemented selective admissions, with several schools following soon after. Selectivity set in motion another sorting process, concentrating high ability/income students—and eventually high tuitions and donations—at such schools. This bolstered their ability to recruit desirable professors.

In addition, the early decades of the 1900s saw the emergence of "lumpy" rewards for research, the most salient being tenure. Agency theory implies that an up-or-out incentive system can promote performance, particularly if professors compete against individuals of similar ability. In concert with increasingly precise research performance measures, tenure furthered a sorting process that led to professors clustering in departments with colleagues of similar ability.

In short, top US schools came to enjoy a virtuous circle that gave them resources to invest in research, which they could effectively incentivize; this helped attract strong students and funding, which could go into further reforms and enhancements. At the same time, other schools struggled and lost ground, creating the between-school inequality evident in US higher education.

Throughout this period, such dynamics operated much more weakly in Europe, particularly on the continent. European states controlled their higher education systems and did not allow easy entry. This prevented the emergence of hundreds of struggling schools but also slowed the rise of extremely well-heeled magnets for global talent, like MIT or Stanford. Overall, sorting of all types was not as extensive in Europe. Concerning incentives, European professorial arrangements—whose origins lie far back in the high middle ages—used up-or-out schemes like tenure to a much lesser extent.

The Initial American Weakness in Higher Education

A demand for denominational sorting drove the creation of the nine American colonial colleges. For example, Massachusetts Puritans created Harvard to produce what they saw as a theologically sound education. They took this task seriously; for instance, the colony forced out Harvard's first President, Henry Dunster—even as it praised his management—because of his stance on infant baptism. Analogous objectives led to the creation of the College of New Jersey (Princeton) by Presbyterians,

King's (Columbia) by Anglicans, Queen's (Rutgers) by the Dutch Reformed, Brown by Baptists, and so on (Herbst 1982).

In addition, like school users to this day, the early founders desired proximity—to have schools close to home (for modern evidence, see Card 1995; Neilson 2017). For example, Connecticut-based Puritans created Yale because they perceived that the Harvard of Massachusetts-based Puritans was too physically distant (in addition to too religiously liberal). After independence, the desire for sorting and proximity led to massive entry: historians estimate 900 colleges opened before the Civil War (Tewksbury 1932; Hofstadter 1955; Burke 1982). Religious denominations competed to open colleges, including on the changing western frontier (Xiong and Zhao 2020).

Well into the 1800s, therefore, American colleges mainly catered to small local markets. It was common for a college to have fewer than 100 students and, say, five faculty members (Rudolph 1962; Geiger 1986). The latter were often hired with attention to their religious affiliation. For example, the desire for clarity on this front led colleges to prefer their own graduates and avoid sending them away for training. In 1879, Princeton President James McCosh wrote to an alumnus studying in Europe: "You are aware that the Trustees . . . are resolute in keeping the college a religious one. You have passed through varied scenes since you left us. . . . If a man has the root in him he will only be strengthened in the faith by such an experience. It will be profitable to me to find how you have stood all this" (Leslie 1992).

Wealth-related criteria mattered too. Rudolph (1962) writes: "From 1835 to 1852 chemistry at Williams was taught by a man of independent wealth who spent his token salary on laboratory equipment. The appointment of Henry D. Rogers . . . at . . . Pennsylvania . . . was a response to his offer to serve without salary. Amasa Walker joined Oberlin . . . on the same terms." This reflects that professorships often paid relatively little. Historians state that even in the 1860s, salaries at Harvard and Yale were "below the cost of living" (Flexner 1946).

While religion clearly influenced hiring, expertise could matter less. For example, in 1853 James Renwick, Columbia's sole professor of "natural philosophy," announced his retirement. Among the mainly Anglican trustees who searched for his replacement was Samuel Ruggles, a real estate investor. Ruggles favored hiring Wolcott Gibbs, a chemist with sterling credentials who would later teach at City College and Harvard. Some trustees, however, noticed a blemish on his record: he was a Unitarian. Ruggles tried to preempt the opposition, proposing a motion:

> Whereas the original charter incorporating King's . . . College . . . provides that [its] laws . . . shall not 'extend to exclude any person of any religious denomination whatever, from equal Liberty and advantage of Education' . . . RESOLVED that in filling [this] Professorship, the Trustees are legally and morally bound to select such Professor, with reference solely to his fitness for the place, without regard to his religious opinions (Thompson 1946).

The motion was rejected.

Other types of identity could matter too. In 1802, Yale President Timothy Dwight recruited Benjamin Silliman to teach chemistry. That Silliman would become an illustrious teacher would not have been obvious at the time. As Kelley (1974) notes, "Silliman had never studied chemistry and knew almost nothing about it. . . . Dwight chose Silliman because he knew it would be difficult to find anyone in America with knowledge of chemistry and natural history, and he was afraid to select a foreigner."

In short, the early US colleges did not prioritize specialized or even trained personnel. It is not surprising that they made an associated choice: they offered a basic curriculum featuring few, if any, electives. The course of study emphasized Latin, Greek, logic, rhetoric, mathematics, physical sciences ("natural philosophy"), and ethics and politics ("moral philosophy"). It excluded "practical" fields such as business and engineering. When the curriculum involved science, the emphasis was deductive rather than experimental/inductive (Butts 1939; Storr 1953). The teaching methods involved rote learning in the form of "recitations" in which students declaimed memorized passages.

These curricular and pedagogical choices meant that classes could be staffed by few instructors. One faculty member at Williams taught rhetoric, English literature, aesthetics, and political economy. Another at Dartmouth was hired to teach "English, Latin, Greek, Chaldee and such other languages as he shall have time for" (Rudolph 1962). Even more extreme, in the 1700s, individual tutors delivered all instruction to a given cohort (like the class of 1776) into the junior year. Given such a task, modern professors might warm to rote learning, too!

These practices kept costs down, which was necessary due to limited enrollments and revenues. Serious budgetary pressures were common; Boyer (2015) states that most colleges hovered between "genial penury and unmitigated fiscal disaster."

The whole setup also reflected the inability and/or unwillingness of schools to branch out. For example, to the extent that instruction in applied science developed, it was mainly outside colleges. A notable player was the Rensselaer Polytechnic Institute (RPI), which offered engineering, surveying, applied science, and some of the first laboratories in the country. However, schools like RPI also offered narrow curricula and little advanced instruction. Rudolph (1977) states that around 1830, a motivated college graduate could earn an RPI degree in 24 weeks. Further, the existence of such "institutes" facilitated the claim by colleges that science belonged elsewhere. When Williams hired Ira Remsen to teach chemistry, he asked for a small room to set up a self-funded laboratory. The response: "You will please keep in mind that this is a college and not a technical school" (Rudolph 1962).

All of the above is not to say that no research went on in the early American colleges. In the 1830s, Amherst's Edward Hitchcock completed the first geological survey of Massachusetts, and Harvard's Asa Gray corresponded with Charles Darwin on the *Origin of the Species*. But such efforts were exceptions rather than the rule.

Meanwhile, European universities were in a different situation. Since the Protestant Reformation, European states—particularly on the continent—had tended to control and eventually fund their universities in a relatively generous manner. Faculty were often well-paid civil servants, assigned by ministries of education to specialized "chaired" professorships (Paulsen 1906). They gradually began to absorb the research activity previously located in academies founded mainly in the 1600s (for example, the Académie Royale des Sciences, the Royal Society of London, and the Berlin Academy of Sciences).

This created a contrast evident in two schools—one American, one German—both chartered in the mid-1700s by George II (King of England and ruler of German lands). By the 1850s, the American school, Columbia, had six professors and 150 students; the German, Göttingen, about 90 professors and 1,600 students (Ruggles 1854). Table 1 returns to the data to list the 20 universities most mentioned in Nobel winners' biographies in roughly 40-year periods. For 1855–1900, Göttingen was among the top 10; Columbia—like all other American schools—did not make the top 20.

What happened since the 1850s? To start, many more schools entered the market. Today, the United States has about 4,700 post-secondary institutions. By the logic above, its universities might be even weaker. And yet, the United States has about 100 "Research I" universities.[3] Thanks to these schools, the United States accounts for 17 of the top 20 universities in the last column of Table 1 (for 1981–2016). The next sections review how this transformation happened.

Agency Theory and Research Performance

We begin by sketching an agency theory framework that helps make sense of what occurred (a mathematical version of this section appears as an online Appendix at the *JEP* website). Our starting point is the fundamental idea that individuals respond to rewards and that performance measures shape those rewards. As Merton (1968) argued, this applies to academic activities as well.

Suppose professors work at a university and exert effort on research and "outside activities." The outside activities may include consulting work and staying active in alumni networks or, in the 1800s, religious groups. We suppose the professors are intrinsically motivated to do research, and so they will engage in this activity to some extent even if they are not compensated for it. We also assume that all else equal, individuals and universities prefer a stable rather than risky income stream.[4]

---

[3] "Research University-I" is a category in the taxonomy of higher education institutions developed by the Carnegie Commission on Higher Education.

[4] In the case of individuals, this follows from risk aversion; in the case of universities, stable and egalitarian pay reduces organizational conflict (Lazear 1989; MacLeod 2003).

Table 1

University-level Nobel Prize Mentions in Given Time Periods

Total number of mentions for:			
1855–1900		*1901–1940*	
Humboldt University of Berlin	19	University of Cambridge	63
University of Cambridge	11	Harvard University	35
University of Munich	10	Humboldt University of Berlin	31
University of Strasbourg	10	Columbia University	29
University of Paris	9	University of Gottingen	27
Leiden University	8	University of Munich	25
University of Wurzburg	7	University of Paris	21
Heidelberg University	6	University of Chicago	20
University of Copenhagen	6	University of Manchester	18
University of Gottingen	6	University of Oxford	17
ETH Zurich	5	ETH Zurich	16
University of Bonn	5	California Institute of Technology	15
University of Vienna	5	University of Copenhagen	15
Ecole Normale Superieure	4	University of Wisconsin	15
University of Manchester	4	Johns Hopkins University	14
Leipzig University	3	University of California, Berkeley	12
University of Giessen	3	University of Marburg	12
University of Graz	3	Uppsala University	12
University of Lyons	3	Cornell University	11
University of Stockholm	3	Princeton University	11

1941–1980		*1981–2016*	
Harvard University	101	Harvard University	32
University of Cambridge	83	Stanford University	28
Columbia University	64	Massachusetts Institute of Technology	27
University of Chicago	64	University of California, Berkeley	20
University of California, Berkeley	62	University of Chicago	18
Massachusetts Institute of Technology	48	Yale University	17
Stanford University	46	Princeton University	16
University of Oxford	33	University of Cambridge	13
Princeton University	32	Columbia University	12
California Institute of Technology	29	University of Texas	12
Yale University	26	Johns Hopkins University	10
Cornell University	23	California Institute of Technology	9
University of Illinois	20	Rockefeller University	9
University of Paris	20	Cornell University	8
Carnegie Mellon University	19	Nagoya University	8
University of Pennsylvania	19	Northwestern University	8
Johns Hopkins University	15	University of Pennsylvania	8
New York University	15	University of California San Diego	7
Rockefeller University	15	University of California, Santa Barbara	7
Technical University of Munich	15	University of Oxford	7

A key question is whether compensation can be used to reward research performance, which depends on whether the university observes meaningful measures of research output. Suppose first that the university does not, and so faculty salary cannot vary with research effort.

This basic setup captures elements of the American college system well into the 1800s. There were few rewards for research, and yet some professors engaged in it. Research was likely easier for wealthier individuals with less need for outside activity. In addition, in the 1800s, the return to outside effort could have been substantial, even if one were focused on an academic career. The cases discussed above illustrate that one might more effectively land a professorship by signaling one's denomination than by doing research.

Now suppose that a measure of professors' research performance exists, and that the university places value on producing research. Principal-agent theory (as in Holmstrom and Milgrom 1991) implies that the university should make faculty compensation depend on this measure. Indeed, at modern universities, rewards like salaries and raises vary with research performance. Such compensation provides incentives, but it entails a cost because it exposes professors to risk. Thus, a more precise research performance measure is a "win-win": it reduces risk and allows for better incentives. A key question, therefore, is how one might enhance the precision of such a measure.

This question matters because it is usually difficult for non-experts to assess professors' research performance. This is particularly true in making comparisons across fields. For example, in the humanities, faculty tend to write books, while articles are the norm in the sciences. Similarly, a chemistry professor may be listed as a co-author on all papers produced by her laboratory, while an economic theorist writes alone—and it is not obvious who is being more productive.

To obtain more precise measures one useful step is to organize professors into fields and to compare their performance *within fields* (Lazear and Rosen 1981). With access to a relative performance measure, it makes more sense for the university to increase the rewards for research.

In short, our framework highlights factors that improve research performance: i) increasing specialization that organizes faculty into fields; ii) reducing the return to outside activities that compete with research for faculty effort; and iii) increasing the extent to which universities value research. We now discuss how reforms beginning in the 1800s promoted these changes.

US Reforms of Higher Education

As industrialization advanced in the second half of the 19th century, American colleges increasingly failed to satisfy some areas of educational demand. Some students were interested in subjects the colleges did not cover, such as business, engineering, and mining. Others wanted specialized modern

training in areas like chemistry and economics. Over the first part of the 1800s, reformers at several schools (like Brown, Harvard, and Union) attempted to address these shortcomings but failed to sustain reform (Butts 1939; Rudolph 1962).

Lasting change arrived with two entrants: Cornell (1865) and Johns Hopkins (1876). Each received substantial private support, the salient donations being those of Ezra Cornell and Johns Hopkins. But each also received major public support: Cornell got federal funds via the Morrill Act, and Hopkins eventually received support from Maryland (Curti and Nash 1965).[5]

These schools expanded the curriculum and began to offer specialized instruction. Cornell saw rapid enrollment growth, showing that a school could thrive by responding to curricular demand rather than by providing denominational sorting. For its part, Hopkins emulated the focus of German universities on advanced/graduate instruction. Its first president, Andrew Gilman, promoted hiring based on expertise and research ability, rather than aspects like denominational affiliation. He stated: "The institution we are about to organize would not be worthy the name of a University, if it were to be devoted to any other purpose than the discovery and promulgation of the truth; and it would be ignoble . . . if the resources . . . given by the Founder . . . should be limited to the maintenance of ecclesiastical differences or perverted to the promotion of political strife . . . sectarian and partisan preferences should have no control in the selection of teachers" (as quoted in Hawkins 1960). In our framework, Gilman was increasing the weight on research and decreasing that on outside activities.

Incumbent schools had to decide how to respond. Harvard and Columbia invested aggressively to compete with Cornell and Hopkins in offering electives, funding research, and creating graduate schools of arts and sciences and professional schools (like business, engineering, law, and medicine). Other well-heeled incumbents—notably Princeton and Yale—generally followed but retained a greater focus on their undergraduate colleges. Other incumbents, like Amherst and Williams, remained fully committed to undergraduate curricula. But even these last schools began to offer majors and electives and to split academic areas into specialized fields. For example, natural philosophy into astronomy, biology, chemistry, physics; political economy into economics, political science, sociology, and so on. Academic specialization was on the rise.

Reformers soon faced additional competition from other entrants aiming to be research universities. These included Stanford (1885), and the University of Chicago (1890), which relied on private donations, and schools that benefited from Morrill Act funding and/or state support—such as the universities of California, Illinois, Michigan, Minnesota, and Wisconsin, along with Michigan State, Penn State, and Purdue. Yet others, like MIT, enjoyed both private and public funding.

---

[5] The Morrill Act (1862) transferred shares of mostly Western federal land to states. The proceeds of their sale were to benefit higher educational institutions whose objective was "without excluding other scientific or classical studies, to teach such branches of learning as are related to agriculture and the mechanic arts." States retained wide latitude in the use of funds; for example, they could be awarded to private universities.

Many of these schools innovated in ways of their own. For example, Chicago was unusually aggressive in raiding other schools' faculty; Berkeley emphasized hiring junior professors and avoided "in-breeding" (another way of reducing the return to outside activities); and MIT and Stanford moved to collaborate with the private sector in ways that generated funding, setting a pattern for years to come.

It is worth noting that this scale of entry in US higher education stood in stark contrast to the situation in Europe. There, states controlled and mostly restricted the creation of universities. For example, around the time of the US Civil War, the United States had hundreds of separate colleges but England still had four universities. Further, many continental universities required state approval to venture into new fields, particularly where states funded faculty salaries.

As American universities competed to deliver advanced and specialized instruction, they began to seek professors at the cutting edge of their fields. Initially, such academics were in short supply and/or were hard to identify. For example, early in his long presidency of Harvard (1869–1909), Charles Eliot observed: "There is in this country a very considerable body of teachers who know how to teach Latin and Greek . . . but if you are in search of teachers to teach botany, chemistry, physics and so on, you cannot find them. They do not exist" (as quoted in Flexner 1946).

As the competition for research talent heightened, the supply of PhD-trained specialists grew. The academic system also began to emphasize specialization in ways that provided measures of the research performance of professors.

Faculty members founded professional associations, like the American Chemical Society (1877) and the American Historical Association (1884). These associations began to publish journals: for example, the American Economic Association (1885) introduced the *Publications of the American Economic Association*, which later evolved into the *American Economic Review*. University department journals supplemented such periodicals: for example, Harvard's *Quarterly Journal of Economics* (1886) and Chicago's *American Journal of Sociology* (1895).

Here the United States was catching up to Europe, where academic societies and specialized journals had existed long before. For example, consider mathematics, where European periodicals like the *Journal de l'Ecole Polytechnique* appeared in the 1700s (Bartle 1995). At Johns Hopkins, President Daniel Gilman hired an English mathematician, James Sylvester, who led the founding of the *American Journal of Mathematics* in 1878. Years later, in his retirement address, Sylvester disputed this role (as quoted in Flexner 1946):

> You have spoken of our Mathematical Journal. . . . Mr. Gilman is continually telling people that I founded it . . . I assert that he is the founder. Almost the first day I landed in Baltimore, . . . he began to plague me to found a Mathematical Journal on this side of the water . . . I said it was useless, there were no materials for it. Again and again he returned to the charge and again and again I threw the cold water I could on the scheme; nothing but obstinate persistence . . . brought his views to prevail.

As American periodicals formalized their procedures and made use of peer review, they differentiated in quality. Those journals that became better-regarded received more submissions, and their editors made them more selective, thus attracting more papers—a hierarchy evolved in each field.[6]

Charles Eliot reflected on the impact this had over his Harvard presidency: "The chief difficulty that I encountered was the procuring of teachers competent to give advanced instruction. There were really no guides to the discovery and invitation of the persons needed. Then none of the societies organized for the . . . mutual support of learned . . . men existed. By 1885 I could get some assistance . . . from the proceedings of the . . . scientific societies. At the beginning there was no such aid" (James 1930).

The demand for research talent improved the negotiating position for professors. To illustrate, Chicago's first president, William Harper, raided campuses for faculty. One of his targets was Clark University, which had made an underfunded attempt to compete with Johns Hopkins at graduate teaching; Harper hired away 15 of its professors. To compete, schools were soon offering higher salaries, reduced teaching loads, and sabbaticals. At Swarthmore, President Joseph Swain informed an astronomer he was recruiting that he would satisfy his demand for a 24-inch telescope. Still, he added a caveat: "[R]emember, this is a Friend's College and thee should give up thy smoking" (Leslie 1992). In the event, Swain had to accept the professor's tobacco habit too.

Resources and Admissions-Related Reform

How did universities manage to finance these changes? How were top departments able to pay talented researchers well and offer them sabbaticals and PhD students? How did the American system transition from one in which essentially all colleges experienced "genial penury" to one in which a (small) minority of universities spend lavishly? To elaborate, at present, US university spending is high—the country is an outlier in regressions of per-student expenditure on GDP per capita.[7] Beyond this, the US higher education sector displays distinct inequality. Hoxby (2016) estimates that the most selective universities spend about $150,000 per student—about six times the US average and about 15 times as much as their less well-heeled counterparts. How did this situation develop?

A key part of the story involves selective admissions. Before the 1920s, essentially all colleges had "open enrollment" policies, reflecting their desire to grow. Harvard's Charles Eliot (1908) stated: "The American universities have always and everywhere been desirous of increasing the number of their students; and this is

---

[6]It is possible that a more thorough hierarchy appeared in the United States than in Europe. Geiger (1986) suggests that Europe displays a greater tendency for academics to submit work to periodicals in their own institutes, universities, or countries; language differences may factor into the latter too.

[7]For cross country data, see Organization for Economic Cooperation and Development (2013).

a true instinct of university governors in a democratic country."[8] Such policies, combined with the reforms that rendered schools attractive, produced higher enrollments. The historically cash-strapped schools welcomed this development. In the 1910s, Columbia and Harvard—perhaps the most aggressive reformers— were the largest American universities by enrollment, larger than any public school (Rudolph 1962). At this time, size was seen as a signal of quality because it provided resources to pursue reforms.

However, this growth began to generate its own challenges. Because universities continued to serve relatively local markets, growth entailed the arrival of less prepared and/or lower-income students. The resulting heterogeneity threatened universities' ability to deliver the peer groups that their traditional "elite" customers (white and usually Protestant men) expected. In other words, at some schools, traditional customers might have felt that their educational spaces were being invaded, or at least losing some of their signaling value.

Economic theory suggests that if universities had low fixed costs, such concerns would lead to the existence of a large number of schools, each catering to individuals of specific types (Epple and Romano 1998; MacLeod and Urquiola 2015). But in reality, fixed costs and coordination challenges prevented elites from easily seceding from the schools they used. Instead, they began to form exclusive clubs within their schools. At Harvard, Princeton, and Yale, wealthier students sorted into private dormitories, "eating clubs," and secret societies; at Cornell, Michigan, and Wisconsin, into select fraternities and sororities (Synnott 1979; Karabel 2006; Urquiola 2020). Beyond providing exclusivity, these clubs performed tasks that growth had led schools to neglect—like supplying living spaces and dining rooms.

Many students coveted membership in these clubs. Franklin Roosevelt's experience is illustrative. Arriving at Harvard in 1900, he aspired to join Porcellian, the most exclusive club. He later described his failure to do so as the greatest disappointment of his life. Eleanor Roosevelt added that this experience led him to identify "with life's outcasts" (Karabel 2006). Similarly, a student at Princeton stated, on being rejected by a club: "The news came like a thunderbolt. With a cold, sick feeling the bottom dropped out of my college life" (Karabel 2006). Such reactions echo the disappointment some students today feel about not getting into this or that elite college. This illustrates that in this period the clubs—rather than the colleges themselves—began to provide the sorting students coveted.

Initially, many university leaders did not view this outcome as problematic. In 1892, for example, Harvard's Charles Eliot justified it: "Rich people cannot be made to associate comfortably with poor people, or poor with rich. They live, necessarily, in different ways, and each set will be uncomfortable in the habitual presence of

---

[8]Some eastern schools had admission exams but routinely enrolled students "with conditions," that is, despite having failed (Karabel 2006).

the other. Their common interests are unlike, and their pleasures are as different as their more serious occupations" (Veysey 1965).

But in time, university leaders realized that the clubs posed challenges. First, alumni might donate to their clubs rather than to their school, especially if the clubs were the source of valuable networks. Second, a school that loses control of the student experience loses control over its admissions and image; for instance, in the early 1900s some students avoided Yale because they perceived getting into its clubs as too difficult (Karabel 2006). Third, the clubs had their own agenda, one that often glorified unruly/unscholarly behavior; in the words of Princeton's President Woodrow Wilson (1902–1910): "The sideshows are so numerous, so diverting, so important, if you will, that they have swallowed up the circus" (Brubacher and Rudy 1958).

To complicate matters, this happened in a period of rising anti-Semitism when many of the new students—particularly at urban schools such as Columbia, Harvard, and Pennsylvania—were Jewish. For example, by 1920, the share of Jewish students at Columbia had reached 30 percent (Karabel 2006). These are only estimates, because colleges initially had little data on religious affiliation. But they began to work toward such data at this time. At Harvard, President Abbott Lowell asked a committee to use extensive personal data to assign students to three groups: J1—"conclusively Jewish," J2—"indicatively Jewish," and J3—"possibly, but not probably Jewish" (Synnott 1979).

In 1919, Columbia responded by implementing selective admissions. It capped its entering class size, started requiring more personal data from applicants, and began to reject some without stating a reason. Soon other schools (including Cornell, Harvard, MIT, Stanford, and Yale) implemented similar policies.

Despite its convoluted origins, selectivity gradually made for an academically stronger student body at these schools. Making enrollments predictable also allowed universities to better supply services like dormitories and dining rooms, aiding efforts to rein in student behavior.

In other words, selectivity set the stage for a massive sorting process—in this case involving students. The academically more talented began to congregate in a few schools. Just as happened with journals, a pecking order developed. Further, just as the appearance of journals facilitated the sorting of professors, standardized tests like the SAT (starting in 1926) facilitated student sorting—in both cases, observability was central. Among schools, student sorting created winners and losers. Many schools whose students had been quite similar to Harvard's in ability found that to no longer be the case (as documented in this journal by Hoxby 2009). Further, an "open enrollment" sector continued to exist, as theory predicts.

Selectivity produced a financially virtuous circle for the schools that moved early. These took in bright and/or well-connected students and provided them with good experiences, including teaching by the best professors (at least in terms of research ability). They gave their students networks with potentially valuable labor market and marriage market contacts. Their satisfied alumni donated to them, allowing these schools to make even more investments, and so forth.

Events around World War II added a powerful ingredient that helped raise spending and reinforced inequality: federal public research funding. Gruber and Johnson (2019) note that in 1938 federal and state governments spent a combined 0.08 percent of national income on research. By 1944, the US government was spending nearly 0.5 percent of national income on research, a figure which would reach 2 percent by the 1960s (thereafter declining to its pre-war level). Much of this funding is allocated to research projects as a function of quality assessed by panels of experts. To the extent that the top universities account for more than their fair share of research talent, it is not surprising that they receive a substantial share of these funds (Graham and Diamond 1997).[9]

Here again, a contrast emerges with Europe, where many states suppressed the type of inequality and sorting that characterizes the American university sector. For example, German states and countries like Spain—which largely control their universities—allocate resources to promote equality across schools. In a similar spirit, German states often set faculty salaries largely as a function of seniority and rank. Such restrictions limit universities' ability to differentiate pay and compete on a global hiring market.

Tenure

Higher salaries, lower teaching, and enhanced laboratory space illustrate some ways in which the compensation of professors began to reflect research performance. Furthermore, this period saw the emergence of tenure, a salient reward for research performance. Unlike salary, tenure is "lumpy"—it does not allow for small enhancements. Rather, tenure rewards reaching a threshold level of achievement and cannot be taken away. This section describes the emergence of tenure and discusses how agency theory suggests it can incentivize research performance.

The Emergence of Tenure

One can see the rise of tenure as a response to two related sets of demands. The first was a desire on the part of professors for academic freedom and protection from arbitrary dismissal. Several events illustrating this involve economists, in part because during this period many saw themselves as reformers in addition to researchers—this created tension with university presidents and donors. One famous case involved Richard Ely, who was hired (in 1881) by Johns Hopkins president Andrew Gilman. Ely was active in the progressive movement, supported labor unions, and helped found the American Economic Association (Barber 1987). He later moved to the University of Wisconsin at a time of labor tensions

---

[9] For more detailed data on government research efforts, see reports at the National Science Foundation website like "Higher Education Research and Development: Fiscal Year 2011" at https://www.nsf.gov/statistics/nsf13325/content.cfm?pub_id=4240&id=2.

in that state. Soon Wisconsin president Charles Adams heard complaints that Ely taught socialism and had "entertained a union organizer in his home" (Schrecker 1986). The university's board appointed a committee to investigate. The board not only supported Ely but made a broader statement (as quoted in Scott 2018):

> As Regents of a university with over a hundred instructors . . . who hold a vast diversity of views . . . we could not for a moment think of recommending the dismissal . . . of a teacher even if some of his opinions should, in some quarters, be regarded as visionary. Such a course would be equivalent to saying that no professor should teach anything which is not accepted by everybody as true. We cannot for a moment believe that knowledge has reached its final goal, or that the present condition of society is perfect. We must therefore welcome from our teachers such discussions as shall suggest the means and prepare the way by which knowledge may be extended, present evils be removed and others prevented.

However, this outcome was not preordained, as is evident in the case of Edward Ross, one of Ely's students. Ross obtained a PhD in political economy at Johns Hopkins and taught economics and sociology at Stanford, where he publicly defended the socialist Eugene Debs and opposed Asian immigration. This irritated Jane Stanford, whose fortune partially derived from her late husband's investment into railroads built using Chinese labor. She accused Ross of aiding the "vilest elements of socialism," and pressured President David Jordan to terminate Ross (Dorn 2017). Jordan obtained Ross's resignation (1900), but at the cost of bad press and eight additional faculty departures. The case contributed to creating the American Association of University Professors (AAUP), which in 1915 stated a desire to "safeguard freedom of inquiry . . . against both covert and overt attacks."

The second set of demands reflected universities' desire to incentivize and occasionally terminate professors. Harvard's early hiring practices illustrate some associated challenges. In the 1700s, Harvard hired tutors using renewable contracts. Its officials explained this stating that permanent contracts raised the risk that bad instructors might be "fixed on the college for life" (Metzger 1973). Further, renewable contracts could provide incentives; as put by Harvard's statutes, they could "excite tutors . . . to greater care and fidelity in their work" (Metzger 1973). A further benefit of finite terms is that a school does not have to cite a cause to dismiss a professor who performs poorly—it can simply let his contract end. Yet as any manager knows, terminating a worker is difficult even if it is feasible; when renewals come up, it is always tempting to "kick the can down the road." Metzger (1973) states that a 1760 Harvard rule limiting appointments within a rank to eight years "was to prevent incumbencies from being lengthened by reappointments given out of neglect or sympathy; it was intended to defeat the importance of kindness in the serious business of evaluating personnel."

These two sets of demands—for permanence/freedom and performance/ incentives—could become intertwined. For example, in the 1930s, Harvard

president James Conant wished to strengthen his school's social sciences, widely perceived to have fallen behind those at Chicago and Columbia. Two faculty members with somewhat limited publication records and labor union ties came up for reappointment. When they were terminated, some complained that politics had been the cause. A faculty committee found no evidence of this but questioned the timing of their review.

Cases like this illustrate that it is difficult to disentangle the desire for permanence/freedom and the desire to provide incentives. It is possible to see tenure—by creating a pre-defined trial period leading to an "up-or-out" date—addressing both sets of demands.[10] It is not entirely clear when tenure became formalized because, as happened with many features of the US higher education system, it appeared in a decentralized fashion. Salient parts of the package had emerged at Princeton by the 1920s and at several schools by the 1930s; by 1950, the full package was commonplace. However, in recent decades, tenure has become less common in the United States, and it is increasingly available only to research-focused professors at wealthier universities (Figlio, Schapiro, and Soter 2015).

Tenure and Agency Theory

Agency theory also provides ways to understand tenure. Even with improvements in the precision of research measures, there always remains an element of subjectivity, which can allow the university to shirk on its compensation commitments. A solution to this problem is for the school to commit in advance to providing rewards for relative performance (Lazear and Rosen 1981; Carmichael 1983; Malcomson 1984). For example, the school might create a teaching award, leaving in question only who will win it. [11]

However, if there is substantial variation in ability, then some individuals will have a very low (or high) probability of winning the prize, and for them, the scheme will not enhance performance. This implies that all else equal, tenure will more effectively promote research if combined with a process whereby professors sort into schools/departments according to their research ability.

Indeed, once measures of professors' research output were available, and once universities began to bid on faculty, stratification by ability/performance resulted. In other words, academic fields came to display a pecking order of departments/schools. This sorting process was facilitated by professors, given that they themselves have incentives to segregate by ability. For example, some believe there are externalities from proximity to talented colleagues or that belonging to an elite department enhances their reputation (MacLeod et al. 2017).

---

[10]As a matter of law, tenure does not guarantee permanence of employment, but rather, sets the bar for dismissal high. Faculty who violate behavioral policies or fail to teach, for example, can be dismissed.

[11]In practice, perfect performance measures do not exist. MacLeod (2003) shows that this necessarily leaves space for conflict, which decreases as the quality of information improves. In the online Appendix, we show that lower variation in ability increases the effectiveness of prizes; see Brown (2011) for evidence.

Tenure likely facilitated such sorting. The individuals most qualified to judge faculty in a specialized area are other academics in that area. This can create a problem when hiring—if incumbents' employment depends upon relative evaluations, they will naturally want to hire individuals less skilled than themselves. Carmichael (1988) observes that tenure helps solve this problem. If faculty with tenure have protected employment, they have no incentive to misrepresent the quality of new hires.

Agency theory provides additional insights into tenure. First, the length of the tenure clock may allow for a more precise measure of performance. Second, labor economists often treat labor as a standard commodity; for instance, workers produce a certain output in an hour and are paid for that hour. Work on research can be quite different; for instance, a researcher might work for years to make one important discovery—MacLeod (2007) calls this an *innovative* commodity. In such cases, optimal compensation can feature delayed rewards and bonuses (Levin 2003; Fuchs 2007; Manso 2011), which tenure helps introduce. Third, universities wish to hire professors with an intrinsic preference for research (Stern 2004). A long tenure clock and relatively low pay can result in the self-selection of faculty who will keep exerting research effort even after receiving tenure. Fourth, with tenure, faculty come to have a vested interest in their institution and will be more willing to contribute towards its advancement (Carmichael 1983, 1988).

Note that tenure mostly did not emerge in Europe. To be clear, European universities certainly have a set of senior professors who enjoy job security and other privileges. For example, chaired German professors enjoy arguably greater power in their schools than American full professors—they are the heirs of the "masters" that controlled ancient universities like Paris and Heidelberg.

However, younger academics in Europe are much less likely to face an internal "up-or-out" research-based evaluation. At some schools (for instance, in France and the United Kingdom) faculty essentially receive immediate tenure. Their progression through the ranks may indeed depend upon their performance, but their permanence is assured. In other settings (for example, in Germany), universities hold open contests for all senior professorships, in some cases even precluding applications by junior academics at the same school. This differs from the American arrangement where the junior "tenure-track" professor takes part in an internal evaluation and the only question is whether she has attained an absolute level of attainment.

To the extent that tenure has positive impacts, these are likely reinforced by the fact that it interacts with faculty sorting by research talent. Sorting implies that assistant professors are matched to a department at which each of them is not far, in expected output, from their colleagues; combined with a tenure goal, this will raise effort through a wide range of the ability distribution.

Conclusion

We have discussed why the United States accounts for a high proportion of the world's leading research universities. Our story is not one of success by design, but

rather of competition helping to create a confluence of incentive mechanisms that help explain current performance.

The rise of American higher education began in the late 1800s, when industrialization increased universities' interest in advanced/specialized instruction and hence in professors' ability to do research. The emergence of specialized fields (with accessories like journals) produced improved measures of performance. Armed with these, schools began to bid for and reward good researchers. This began to concentrate talent at a few schools, a process reinforced by the emergence of tenure that selected dedicated individuals and created incentives for them. Schools that aggressively pursued these reforms began to experience growth and responded to associated challenges by implementing selective admissions. This unleashed a further sorting process, allowing them to attract the most talented students along with high tuition payments and donations. In short, complementary and self-reinforcing dynamics concentrated research-talented professors and strong students—increasingly from all over the world—at a few schools.

In thinking about the emergence and persistence of US leadership in university research, it is worth contrasting this outcome with the situation in K–12 education. A country's performance in K–12—say, as measured by international tests—depends on the state of hundreds if not thousands of schools. In addition, the existence of the "teacher value-added" literature illustrates the difficulty of measuring performance in K–12. By contrast, when it comes to university research, it is easier to measure performance, sorting may be less of a concern, and top quality output is disproportionately important. Thus, national performance in higher education can depend on what happens at a few dozen universities.

The American system is well suited to producing top schools, although at the cost of inequality. To illustrate, del Corral (2020) compares the performance of Spain and the United States in a recent Academic Ranking of World Universities list often known as the "Shanghai ranking" (http://www.shanghairanking.com/). The United States accounts for 40 of the top 100 universities; Spain for 0. On the other hand, 83 percent of public Spanish universities (delivering in-person instruction) appear somewhere in the ranking that only 23 percent of their American counterparts do.

In closing, we note that agency theory highlights that performance depends crucially on high-quality performance metrics. Today, the academy is rife with complaints that individuals focus excessively, and with calls for "inter-disciplinarity." Our results suggest that a move in that direction is likely to entail other trade-offs, because it may bring on noisier evaluation systems.

References

American Association of University Professors. 1915. "1915 Declaration of principles on Academic Freedom and Academic Tenure." https://www.aaup.org/NR/rdonlyres/A6520A9D-0A9A-47B3-B550-C006B5B224E7/0/1915Declaration.pdf.

Barber, W.J. 1987. "Should the American Economic Association Have Toasted Simon Newcomb at its 100th birthday party?" *Journal of Economic Perspectives* 1 (1): 179–83.

Bartle, Robert G. 1995. "A Brief History of the Mathematical Literature." *Publishing Research Quarterly* 11: 3–13.

Boyer, John W. 2015. *The University of Chicago: A History.* Chicago: The University of Chicago Press.

Brown, Jennifer. 2011. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Super-stars." *Journal of Political Economy* 119 (5): 982–1013.

Brubacher, John S., and Willis Rudy. 1958. *Higher Education in Transition.* New York: Harper.

Burke, Colin B. 1982. *American Collegiate Populations: A Test of the Traditional View.* New York: New York University Press.

Butts, Robert Freeman. 1939. *The College Charts its Course: Historical Conceptions and Current Proposals.* New York: Arno Press.

Card, D. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." In *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp,* edited by N. Christofides, E. Kenneth Grant, and R. Swidinsky. Toronto: University of Toronto Press.

Carmichael, H. Lorne. 1983. "Firm-Specific Human Capital and Promotion Ladders." *The Bell Journal of Economics* 14 (1): 251–58

Carmichael, H. Lorne. 1988. "Incentives in Academia: Why is There Tenure?" *Journal of Political Economy* 96 (3): 453–72.

Clotfelter, Charles T. 2010. "Introduction." In *American Universities in a Global Market,* edited by Charles T. Clotfelter. Chicago: The University of Chicago Press.

Cole, Jonathan R. 2009. *The Great American University: Its Rise to Preeminence, Its Indispensable National Role, Why it Must be Protected.* New York: PublicAffairs.

Curti, Merle, and Roderik Nash. 1965. *Philanthropy in the Shaping of American Higher Education.* New Brunswick, NJ: Rutgers University Press.

Del Corral, J. 2020. "Ranking de Shanghai: ¿comparando España con Estados Unidos?" https://www.universidadsi.es/ranking-universidades-shanghai-espana-estados-unidos/.

Dorn, Charles. 2017. *For the Common Good: A New History of Higher Education in America.* Ithaca, NY: Cornell University Press.

Eliot, Charles W. 1908. *University Administration.* Boston: Houghton Mifflin Company.

Epple, Dennis, and Richard E. Romano. 1998. "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review* 88 (1): 33–62.

Figlio, David N., Morton O. Schapiro, and Kevin B. Soter. 2015. "Are Tenure Track Professors Better Teachers?" *The Review of Economics and Statistics* 97 (4): 715–24.

Flexner, Abraham. 1946. *Daniel Coit Gilman: Creator of the American Type of University.* New York: Harcourt, Brace and Company.

Fuchs, William. 2007. "Contracting with Repeated Moral Hazard and Private Evaluations." *American Economic Review* 97 (4): 1432–48.

Geiger, Roger L. 1986. *To Advance Knowledge: The Growth of American Research Universities, 1900–1940.* Oxford: Oxford University Press.

Goldin, Claudia, and Lawrence F. Katz. 1999. "The Shaping of Higher Education: The Formative Years in the United States, 1890 to 1940." *Journal of Economic Perspectives* 13 (1): 37–62.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology.* Cambridge, MA: The Belknap Press of Harvard University Press.

Graham, Hugh David, and Nancy Diamond. 1997. *The Rise of American Research Universities: Elites and Challenges in the Postwar Era.* Baltimore: Johns Hopkins University Press.

Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream.* New York: PublicAffairs.

Hawkins, Hugh. 1960. *Pioneer: A History of the Johns Hopkins University, 1874–1889.* Baltimore: The Johns Hopkins University Press.

Herbst, Jurgen. 1982. *From Crisis to Crisis: American College Government, 1636–1819.* Cambridge, MA: Harvard University Press.

Hofstadter, Richard. 1955. *Academic Freedom in the Age of the College.* Piscataway, NJ: Transaction Publishers.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7: 24–52.

Hoxby, Caroline M. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23 (4): 95–118.

Hoxby, Caroline. M. 2016. "The Productivity of U.S. Postsecondary Institutions." In *Productivity in Higher Education*, edited by Caroline M. Hoxby and Kevin Strange. Chicago: University of Chicago Press.

James, Henry. 1930. *Charles W. Eliot: President of Harvard University, 1869–1909,* Vol. 1. Boston: Houghton Mifflin Company.

Karabel, Jerome. 2006. *The Chosen: The Hidden History of Admission and Exclusion at Harvard, Yale, and Princeton.* Boston: Houghton Mifflin Company.

Kelley, Brooks Mather. 1974. *Yale: A History.* New Haven, CT: Yale University Press.

Lazear, Edward P. 1989. "Pay Equality and Industrial Politics." *Journal of Political Economy* 97 (3): 561–80.

Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89 (5): 841–64.

Leslie, W. B. (1992). *Gentlemen and Scholars: College and Community in the "Age of the University," 1865-1917.* University Park: Pennsylvania State University Press.

Levin, Jonathan. 2003. "Relational Incentive Contracts." *American Economic Review* 93 (3): 835–57.

MacLeod, W. Bentley. 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review* 93 (1): 216–40.

MacLeod, W. Bentley. 2007. "Reputations, Relationships, and Contract Enforcement." *Journal of Economic Literature* 45 (3): 595–628.

MacLeod, W. Bentley, Evan Riehl, Juan E. Saavedra, and Miguel Urquiola. 2017. "The Big Sort: College Reputation and Labor Market Outcomes." *American Economic Journal: Applied Economics* 9 (3): 223–61.

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review* 105 (11): 3471–88.

Malcomson, James M. 1984. "Work Incentives, Hierarchy, and Internal Labor Markets." *Journal of Political Economy* 92 (3): 486–507.

Manso, Gustavo. 2011. "Motivating Innovation." *The Journal of Finance* 66 (5): 1823–60.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159 (3810): 56–63.

Metzger, W. P. 1973. "Academic tenure in America: a Historical Essay." In *Faculty Tenure: A Report and Recommendations* by the Commission on Academic Tenure in Higher Education. San Francisco: Jossey-Bass, Inc.

Neilson, C. 2017. "Targeted Vouchers, Competition among Schools, and the Academic Achievement of Poor Students." Unpublished.

Organization for Economic Cooperation and Development. 2013. "Education Expenditures by Country. https://nces.ed.gov/programs/coe/pdf/Indicator_CMD/coe_cmd_2013_03.pdf.

Paulsen, Friedrich. 1906. *The German Universities and University Study.* New York: Charles Scribner's Sons.

Rudolph, Frederick. 1962. *The American College and University: A History.* Athens, GA: The University of Georgia Press.

Rudolph, Frederick. 1977. *Curriculum: A History of the American Undergraduate Course of Study since 1636.* San Francisco: Jossey-Bass Publishers.

Ruggles, Samuel Bulkley. 1854. *The Duty of Columbia College to the Community, and its Right to Exclude Unitarians from its Professorships of Physical Science.* New York: John F. Trow.

Schrecker, Ellen. 1986. *No Ivory Tower: McCarthyism and the Universities.* Oxford: Oxford University Press.

Scott, Joan Wallach. 2018. "Academic Freedom: The Tension between the University and the State." In *Academic Freedom: The Global Challenge*, edited by Michael Ignatieff and Stefan Roch, 11–26. Budapest: Central European University Press.

Stern, Scott. 2004. "Do Scientists Pay to be Scientists?" *Management Science* 50 (6).

Storr, Richard J. 1953. *The Beginnings of Graduate Education in America.* Chicago: The University of Chicago Press.

Synnott, Marcia Graham. 1979. *The Half-Opened Door: Discrimination and Admissions at Harvard, Yale, and Princeton, 1900–1970.* Greenwood Press.

Tewksbury, Donald G. 1932. *The Founding of American Colleges and Universities before the Civil War: With*

Particular Reference to the Religious Influences Bearing upon the College Movement. New York: Columbia University.

Urquiola, Miguel. 2020. *Markets, Minds, and Money: Why America Leads the World in University Research.* Cambridge, MA: Harvard University Press.

Veysey, Laurence R. 1965. *The Emergence of the American University.* Chicago: University of Chicago Press.

Xiong, Heyu, and Yiling Zhao. 2020. "Sectarian Competition and the Market Provision of Human Capital." Unpublished.

Taxing Our Wealth

Florian Scheuer and Joel Slemrod

A primary reason for imposing a wealth tax is to shift the tax burden toward the most affluent households. This goal is evident in the choice of wealth as a base for the tax, given that wealth is considerably more unequally distributed than the most common tax bases—income and consumption. In 2016, for example, the top 1 percent of households ranked by net worth held 40 percent of US wealth, while the top 1 percent ranked by income earned 24 percent of income (Wolff 2017). In addition, wealth taxes typically have high wealth thresholds before any tax is due and may also include a graduated structure of tax rates that apply above the thresholds. Heightened consideration of a wealth tax is almost certainly tied to the increase in inequality over the past four decades: although the details of this increase are disputable, in our view the conclusion that it has increased nontrivially is not (for discussion, see the "Symposium on Rising Inequality of Income and Wealth" in the Fall 2020 issue of this journal).

The wealth tax has had some prominent academic adherents: for example, Kaldor (1956) called for a wealth tax for developing countries, Allais (1977) proposed to replace most direct taxes with a 2 percent wealth tax in France, and Piketty (2014) called for a global progressive wealth tax. But even some of those who are concerned about rising economic inequality raise questions about whether a wealth tax is the appropriate policy response. Measuring some forms of wealth

■ *Florian Scheuer is the UBS Professor of Economics of Institutions, University of Zurich, Switzerland. Joel Slemrod is the Paul W. McCracken Collegiate Professor of Business Economics and Public Policy, Stephen M. Ross School of Business, and Professor of Economics, University of Michigan, Ann Arbor, Michigan. Their email addresses are florian.scheuer@uzh.ch and jslemrod@umich.edu.*

on an annual basis is very difficult: for example, how does one value pension accounts, life insurance, trusts, or closely held family corporations? Heterogeneous measurability of assets could cause horizontal inequity, a flight to more easily undervalued assets, and consequent understatement of net wealth. Recent experience is not encouraging: while a dozen high-income European countries levied wealth taxes in the recent past, now only three retain them, which suggests that the other nine countries determined that any benefits associated with such taxes do not justify their costs. In addition, there are a range of other policy options to tackle inequality.

In this essay, we begin with a basic conceptual discussion of the base and tax rates for a wealth tax. We then provide an overview of the recent wealth taxes in European countries. The United States has never levied an annual wealth tax, but during the 2019–20 contest for the Democratic presidential nomination, wealth tax proposals were put forward by Senators Bernard Sanders (D-VT) and Elizabeth Warren (D-MA). As we will see, none of the European wealth taxes either applied rates anywhere near the 6 or 8 percent top rates in these proposals nor established such a broad base, and only Switzerland raises a level of government revenue comparable to these US proposals. However, the United States does have experience with some taxes that have aspects similar to a wealth tax. As we discuss, the property tax is an annual tax on ownership of immovable property, the estate tax is a wealth tax imposed at time of death, and the capital gains tax is imposed on some increases in the value of some assets—albeit in a haphazard way.

We then turn to what we know about the behavioral effects of a wealth tax, including effects on real behavior, financial choices aimed at reducing the burden of a wealth tax, outright tax evasion, and administrative and compliance costs. Studies of the European wealth taxes often, but not always, find a substantial behavioral response, although the nature of the response varies. We emphasize that any lessons drawn from the European experience must be applied to the recent US proposals with substantial caution, because the design features of recent proposals—rate schedule, broadness of the base, and enforcement provisions—are very different from any previous wealth tax.[1]

Finally, we look to optimal tax theory—*should* we as a society decide to our tax wealth and, if yes, how so? We point out that the former conventional wisdom—that an optimal tax system would feature no taxes on capital—has been overturned. Instead, we review a series of arguments that justify some form of progressive taxation of wealth accumulation both in the short and long run. We also discuss under which conditions such taxes should take the form of a wealth tax versus alternative policies that have similar objectives. We conclude with an overview of some political economy arguments for taxes on wealth accumulation that go beyond the usual redistributional objectives.

---

[1] We do not address whether a wealth tax would be constitutional in the United States, a subject of some controversy. For the two views, compare Johnson and Dellinger (2018) and Jensen (2019).

What Is a Wealth Tax?

The Base of a Wealth Tax

In principle, the base of a wealth tax is net worth—the value of assets minus debts. Like all taxes, in practice, the base to which tax rates are applied could be narrowed by exemptions, deductions, or preferential treatment (for example, discounted valuation) of certain components of net wealth.

Determining the base for a wealth tax raises some thorny practical issues. For example, the market value of a closely held, family-run business, or of personal property perhaps received via inheritance, is difficult to estimate at high frequency. The value of many assets, including firms, consists of a projected flow of future income, which makes valuation highly sensitive to the applied discount rate. Another issue is that wealth and wealth tax liability is not always matched by disposable income in a given year, and, as a result, requiring the tax to be remitted annually may raise liquidity problems.

Past and current wealth taxes contain many base-narrowing features. For example, all wealth taxes exempt wealth below a certain threshold, which varies considerably across countries. Some wealth taxes do not apply to wealth held in a pension or life insurance account. Some have exemptions or reduced tax rates for the wealth in one's primary residence; more generally, wealth tax rules often differ across real estate and financial assets. There are reduced or deferred wealth taxes for certain business assets—for example, to prevent a situation where a family-owned firm would need to be liquidated to satisfy a wealth tax liability. Wealth tax bases often leave out trusts established to pass wealth to later generations. Finally, wealth taxes have not been applied to implicit wealth in the form of an individual's human capital, although this is sometimes hard to disentangle from the value of business partnerships (such as law firms or doctors' practices).

The Rate of a Wealth Tax

It is a useful starting point to think of a wealth tax as a tax on the "normal" rate of return to capital. A wealth tax at a rate of t_w is equivalent to a tax rate of t_w/r on capital income where r is the interest rate. For an asset whose rate of return is 8 percent, a 4 percent wealth tax corresponds to an annual 50 percent tax rate on capital income and an 8 percent wealth tax (the top rate in Sanders's proposal) would be equivalent to a 100 percent tax rate on capital income. Thus, the income-tax-equivalent of a given wealth tax rate is smaller, the higher is the rate of return.

However, a wealth tax differs from a capital income tax in an important way. For a given amount of wealth, the tax liability of a wealth tax does not depend on the amount of capital income the wealth actually generates: in contrast, a capital income tax liability is related to that flow. For example, if all of one's wealth were held in a zero-interest demand deposit, a capital income tax would generate no tax liability, while a wealth tax would. If wealth declines in a given year—that is, the return for that year is negative—the wealth tax still applies.

Because a wealth tax affects the rate of return to saving, it changes the relative price of consumption across time. For instance, in a world where the rate of

return $r = 0.07$, an 8 percent wealth tax turns a 7 percent rate of return before tax into a negative 1 percent return after tax.[2] To put it another way, an initial investment of \$1 with a 7 percent return will compound after 30 years to a total of \$7.61 (= $(1 + 0.07)^{30}$). However, with a negative 1 percent return arising from a 7 percent return and an 8 percent wealth tax, \$1 after 30 years will fall to \$0.74 (= $(1 - 0.01)^{30}$), or just one-tenth as much. Tax rates that might sound low in the income or sales tax context are actually much higher when they become part of the annual rate of return that is compounded over time.

Wealth Taxes: Existing and Proposed

European Wealth Taxes

In 1990, twelve European countries levied an annual tax on net wealth. By 2018, only four—France, Norway, Spain, and Switzerland—levied such a tax, with Switzerland raising more than three times as much revenue as a fraction of total revenues (3.9 percent) as any of the other three countries (OECD 2018). In 2018, France replaced its annual wealth tax with a tax only on immovable property. Italy levies an annual tax on financial assets in the form of a stamp duty on bank and securities accounts, and the Netherlands has a hybrid system with similarities to an annual wealth tax, imputing an asset-type-specific rate of return to assets and assessing a 30 percent tax on those imputed returns.[3]

Table 1 provides some summary statistics about wealth taxes in high-income European countries, and then in the bottom two rows compares them to the proposals made by Senators Sanders and Warren in 2019.

For the European wealth taxes, the average top rate was about 1 percent. Finland, Sweden, and Switzerland had top rates reaching nearly 4 percent in the past, and the highest current rate is in the Spanish region Extremadura at 3.75 percent. But the Spanish system, certain Swiss cantons, and, in the past, some other countries feature a cap on the sum of wealth and income taxes as a fraction of taxable income. Such a cap limits the liquidity problem of a high ratio of income plus wealth tax liability to disposable income—and has the effect of imposing a zero marginal tax on wealth for those at the cap. The cap also provides an additional incentive to reduce reported taxable income.

Many of the wealth taxes described in Table 1 feature exemptions or preferential treatment of some forms of assets, notably one's main residence, life insurance proceeds, pension wealth, and business assets. The exemption thresholds of these countries' wealth taxes average about €500,000 for married couples.

---

[2] The real average growth rate of the total wealth held by those on the Forbes 400 list of the wealthiest Americans has been 7.3 percent per year between 1982 and 2018.

[3] Several non-European countries have had wealth taxes, including Argentina, Bangladesh (more recently a net-worth-triggered income tax surcharge or net wealth tax, whichever is higher), Colombia, India (repealed in 2015), Indonesia (abolished in 1985), Pakistan (removed in 2003 and reinstated in 2013), and Sri Lanka (1959–1993).

At first blush, it does not bode well for wealth taxes that of the dozen European countries that have had them in the last three decades, only one-quarter of them still do.[4] Why did the other three-quarters abandon them? A 2018 OECD report lists a number of concerns: efficiency costs, risk of capital flight, failure to meet redistributive goals, and high administrative costs. In Germany, the Federal Constitutional Court deemed the wealth tax unconstitutional in 1995 on the grounds that the tax's distinction between property and financial assets was an infringement against the fiscal principle of tax equality (Drometer et al. 2018). In Sweden, it was argued that the special treatment of business equity made the wealth tax regressive—taxing middle-class wealth (housing, financial assets) and exempting the wealthiest individuals' assets (large, closely held firms)—and it was blamed for spurring tax avoidance and evasion, including capital flight to tax havens (Waldenström 2018).

The Swiss Example

Of the three European countries that still levy a wealth tax, Switzerland raises by far the most revenue as a share of overall tax revenue, amounting to 1.1 percent of GDP in 2018, which is comparable to the revenue projected for the recent US proposals. Hence, the Swiss example is of particular interest for the wealth tax debate in the United States.

The wealth tax in Switzerland has a long history, and in fact, predates the modern income tax. The Swiss tax system is generally structured in three layers: the federal, cantonal, and municipal level. There is no federal wealth tax, but all cantons must levy a comprehensive wealth tax, which they have significant freedom in designing. Eight cantons impose flat rates (above some exemption level), and the other 18 feature graduated rate schedules. Each municipality then chooses a multiplier that is applied proportionally to the cantonal tax rate schedule. Hence, an individual's overall tax liability depends on both the canton and municipality of residence. This highly decentralized system induces local tax competition and migration.

In 2018, the combined cantonal and municipal marginal wealth tax rates in the top bracket ranged between 0.1 percent (canton of Nidwalden) and 1.1 percent (canton of Geneva). In 16 of the 26 canton capitals, the annual top wealth tax rate was below 0.5 percent. There is also some variation in the tax-exempted amounts, ranging in 2018 from about $55,000 in the canton of Jura to $250,000 in the canton of Schwyz (for married couples). Hence, even though it raises similar overall revenue as the estimates for the US proposals, the Swiss wealth tax is targeted at a larger share of the population and is substantially less progressive.

The base of the Swiss wealth tax is broad: in principle, all assets, including those held abroad, are taxable. Only foreign real estate, common household assets, and pension wealth are exempt. The tax liability is based on net wealth, so taxpayers can deduct mortgages and other debt. The annual reporting requirements for assets

---

[4]The Florida intangibles tax offers a cautionary tale closer to home. It could be avoided by putting intangibles in trust in December and distributing them out of trust in January, to the point that it became commonly known as a voluntary tax; it was repealed in 2007.

Table 1
European Wealth Taxes and the Sanders and Warren Proposals

Country/plan	Years of enforcement [1]	Revenue as % of GDP [2]	Top marginal rate		Exemption level	
			Most recent [3a]	Ever [3b]	Single [4a]	Married [4b]
Austria	1954–1994	0.14%	1.00%	1.00%	None	None
Denmark	1903–1997	0.06%	0.70%	2.20%	€ 320,657	€ 641,314
Finland	1919–2006	0.08%	0.80%	4.00%	€ 250,000	€ 500,000
France	1982–1986, 1989–2017	0.22%	1.50%	1.80%	€ 1,300,000	€ 1,300,000
Germany	1952–1997	0.11%	1.00%	2.50%	€ 61,355	€ 122,710
Iceland	1096–2006, 2010–2015	0.48%	2.00%	2.00%	€ 473,248	€ 630,997
Ireland	1975–1978	0.10%	1.00%	1.00%	€ 88,882	€ 126,974
Luxembourg	1934–2006	0.55%	0.50%	0.50%	€ 2,500	€ 5,000
Netherlands	1965–2001	0.18%	0.70%	0.80%	€ 90,756	€ 113,445
Norway	1892–present	0.45%	0.85%	1.10%	€ 157,833	€ 315,666
Spain	1977–2008, 2011–present	0.18%	0%–3.75%	3.75%	€ 400,000–700,000	€ 800,000–1,400,000
Sweden	1947–1991; 1991–2007	0.19%	1.50%	4.00%	€ 166,214	€ 221,619
Switzerland	(1840–1970)–present	1.08%	0.1%–1.1%	3.72%*	€ 25,380–116,250	€ 51,150–232,500
Sanders	n/a	1.56%	8.00%	n/a	$16,000,000	$32,000,000
Warren	n/a	1.34%	6.00%	n/a	$50,000,000	$50,000,000

(continued)

and liabilities allow the cantonal tax authorities to track the year-to-year evolution of wealth and cross-check it against reported income (the so-called "wealth development test"), so the wealth tax serves a supporting role for income tax enforcement.

Several aspects of the broader Swiss tax system provide context for the greater role of the wealth tax there. First, there is no capital gains tax on movable assets (for example, shares of stock in a company) unless the owner professionally trades securities.[5] Second, almost all cantons have abolished taxes on gifts and bequests from parents to children. The Swiss wealth tax therefore serves as a backstop to at least partly substitute for a capital gains tax and an estate tax, which are common in other countries. Third, due to the institution of bank secrecy within Switzerland, third-party reporting of financial assets is precluded, which constrains enforcement. Fourth, while there are some guidelines for the valuation of privately held business assets based on a weighted average of capitalized earnings in recent years and net asset holdings (Hongler and Mauchle 2020), it remains subject to considerable

[5] As for real estate, there is a special capital gains tax at the cantonal level as well as a property tax at the municipal level.

Table 1

European Wealth Taxes and the Sanders and Warren Proposals *(continued)*

					Treatment of:		Estate/
Country/plan	No. of marginal rates [5]	Cap on liability? [6]	Main residence [7a]	Life insurance [7b]	Pension [7c]	Business assets [7d]	inheritance/ gift tax? [8]
Austria	1	N	T	T	T	T	N
Denmark	1	Y	T	E	E	T	Y
Finland	1	Y	TP	E	E	T	Y
France	5	Y	TP	T	E	E	Y
Germany	1	N	T	T	E	TP	Y
Iceland	2	N	T	E	E	TP	Y
Ireland	1	Y	E	T	E	TP	Y
Luxembourg	1	N	T	T	E	TP	Y
Netherlands	1	Y	TP	E	E	TP	Y
Norway	1	N	TP	E	E	TP	N
Spain	8	Y	TP	T	E	E	Y
Sweden	1	Y	T	T	E	E	N
Switzerland	1–1,000	Y*	TP	T	E	TP	Y
Sanders	8	N	T	T	T	T	Y
Warren	2	N	T	T	T	T	Y

Notes and sources: Countries with currently active wealth taxes are shaded in gray. For inactive countries, the table reports information as of the most recent active year. The Sanders and Warren plans are as shown on campaign websites as of April 20, 2020. Data for Iceland come from OECD 2018, Herd and Thorgeirsson 2001, Krenek and Schratzenstaller 2018, and conversations with Thorolfur Matthiasson. [1]: OECD 2018, p. 76, Table 4.1. France abolished its wealth tax in 2017 and replaced it with a tax based on real estate; in 1096 (not a typo) Icelanders began paying a 1% tax on wealth (in fact, a tithe based on a 10% tax applied to an assumed 10% return on assets); Sweden made a major change in 1991; Switzerland's cantons introduced the tax gradually, with full adoption by 1970. [2]: OECD Global Revenue Statistics Database, line 4210 (individual recurrent taxes on net wealth). For active countries the value is for 2018. For inactive countries, the value is for the most recent full active year. For Ireland, McDonnell 2013, p. 24; for Iceland and Luxembourg, Krenek and Schratzenstaller 2018, pp. 23–24. [3]: OECD 2018, p. 88, Figure 4.2 and various historical sources (Du Rietz and Henrekson 2014; Kessler and Pestieau 1991; Lehner 2000, Sandford and Morrisey 1985). The Spanish central government top rate is 2.5%; some regions levy higher rates (e.g., Extremadura's top rate is 3.75%) while others levy lower rates (e.g., Madrid's 100% credit results in an effective rate of 0%). Rates differ across Swiss cantons. *The average top marginal rate in 1942 across all cantonal capitals was 3.72% (Eidgenössische Steuerverwaltung 1969), which is higher than the cantonal plus municipal rates that we are able to confirm in any other year. [4]: OECD 2018, p. 81, Table 4.2. Austria had no specific threshold but implicitly wealth below € 11,000 was untaxed. Denmark's exemption levels were provided in Krone by Katrine Jakobsen and converted to Euro at the 12/31/1996 exchange rate. Krenek and Schratzenstaller 2018, p. 23 notes Luxembourg's € 2,500 exemption per person. Taxpayers are taxed individually in Finland and Spain. The Spanish central government statutory individual exemption is € 700,000; some regions have lower exemptions, including Aragon at € 400,000 and Catalonia at € 500,000. Exemptions differ across Swiss cantons. [5]: OECD 2018, pp. 87–88. The Swiss canton of Basel-Country has a schedule with rates increasing for each CHF1,000 of reported wealth, up to CHF 1 million. [6]: OECD 2018, pp. 88–89, and Silfverberg (2009) (Sweden). *In some cantons, but not all. [7]: OECD 2018, p. 84, Table 4.3. For Denmark, Jakobsen et al. 2019. T = fully taxed, E = full exemption, TP = tax preference. [8]: OECD 2018, p. 24, Table 1.1. Austria abolished its inheritance and gift taxes in 2008, though there is a 2.0–3.5% tax on the transfer of assets with the lower rates for transfers between close relatives.

discretion on the part of cantonal tax authorities, which may contribute to an equilibrium where the wealthy are treated rather leniently.[6]

---

[6]Some cantons offer foreigners who live but do not work in Switzerland an exemption from regular taxation, subjecting them instead to a flat-rate tax based on their living expenses, which has allowed some very rich households to enjoy low tax burdens. The minimum tax rules under this alternative tax regime

The Swiss wealth tax enjoys broad political support, as evidenced by the fact that it keeps being reaffirmed by citizens in Switzerland's system of direct democracy, where most tax policy decisions must be put directly to voters. But its design and the role it plays in the overall tax system are quite different from what is currently discussed in the United States. In particular, it is not geared towards a major redistribution of wealth, and indeed, wealth concentration in Switzerland remains high in international comparisons (Föllmi and Martínez 2017).

Comparisons with Recent US Wealth Tax Proposals

During the 2019–20 Democratic presidential nomination campaign, two prominent candidates, Bernie Sanders and Elizabeth Warren, proposed that the United States enact an annual wealth tax. The Sanders proposal featured graduated rates starting at 1 percent on net worth above $32 million for a married couple, rising to a marginal tax rate of 8 percent on net worth above $10 billion, while Warren proposed a 2 percent rate on net worth in excess of $50 million and a 6 percent rate above $1 billion. The candidates claimed these levies would raise $3.75 trillion and $4.35 trillion over 10 years, or approximately, 1.34 percent and 1.56 percent of GDP and 7.9 percent and 9.1 percent of federal revenues.[7]

As Table 1 details, these US proposals differ quite substantially from the wealth taxes across Europe—past and present—in a number of ways. First, the top marginal rates of both the Sanders and Warren proposals are far higher than any top rate of the European wealth taxes. Second, neither the Sanders nor Warren proposal has a cap on annual tax payments as a share of income. Third, both the Sanders and Warren wealth tax proposals have what is by European standards an exceptionally high exemption level. Under Warren's plan, 75,000 households would be subject to the tax, while Sanders's plan would apply to 180,000 households. The very top rates would affect an even smaller group: Warren's top rate kicks in at $1 billion, applying to about 600 people as of October 2019 (according to *Forbes*), while Sanders's top rate begins at $10 billion, applying to only about 50 people.[8] Fourth, both Sanders and Warren proposed unprecedently broad bases, including, for example, assets held in trust, and (in Warren's case) including retirement assets and assets held by minor children. Fifth, the revenue to be raised from these wealth taxes as a share of GDP is one-third to one-half above the Swiss level, which in turn, is substantially higher than the revenue raised in the other countries. Finally, the US wealth tax proposals would be federal, rather than decentralized as in Spain or Switzerland.

Supporters of these proposals for a US wealth tax often discount the relevance of European experience with wealth taxes—and the wide abandonment of those taxes—on the grounds that the details of the US plans, especially the high exemption level, rate schedule, broadness of the base, and enforcement provisions, are

---

have been tightened recently, and it currently affects fewer than 5,000 individuals (Federal Department of Finance 2019).

[7] Emmanuel Saez and Gabriel Zucman advised Senator Warren regarding her proposal, produced the revenue estimates, and have written a detailed explanation and defense of it (Saez and Zucman 2019a).

[8] The extreme targeting of top wealth levels has led some readers of an earlier draft to suggest that a more appropriate title is "Taxing Their Wealth."

very different than these other systems. This distinction is true, but it is a double-edged sword as these unprecedented design features also make it difficult to learn from experience and to predict its consequences with much confidence.

What Other Taxes Does a Wealth Tax Resemble?

Although the United States has never had an annual wealth tax, it has long experience with several other related taxes.

Property Taxes

Local governments in the United States rely heavily on an annual tax on one form of wealth, often called "immovable" property, in the form of property taxes. Property taxes account for nearly half of own-source local government revenues. The rate of tax levied on immovable property varies widely. Harris and Moore (2013) report that, for the period 2007–2011, the mean property tax burden as a share of house prices was 1.15 percent.

For a number of reasons, a property tax is not identical to an equivalent-rate wealth tax. First, US property taxes do not allow for a deduction for debt. Second, the Tiebout (1956) theory of property taxes emphasizes that households can choose among many different communities with varying levels of local public goods. In this setting, the property tax, unlike a wealth tax, becomes a non-distorting price for local services, and has much different implications than a broader wealth tax not tied to public services. Third, immovable property is only one component of wealth, and because the relative size of this component varies dramatically across levels of household net worth, a property tax is not well-targeted at the very wealthy. Using the 2016 Survey of Consumer Finances, Wolff (2017) estimates that, while principal residences (the major but not only component of the property tax base) account for 62 percent of the gross assets of individuals in the middle three quintiles of net worth, they comprise only 8 percent of gross assets for the top 1 percent. Fourth, most local property taxes in the United States feature a fixed rate, occasionally with an exempt level of property value. Of course, other rate structures are possible. For example, the "council tax" in the United Kingdom features a graduated rate structure based on the property value. Many US states do levy a surcharge on the highest-value homes or have a progressive bracket structure through their real estate transfer tax system, sometimes referred to as "mansion taxes."

Estate and Inheritance Taxes

The US federal government has levied a tax on a base close to net worth since it enacted an estate tax in 1916 and added a gift tax in 1924. This tax requires a valuation of taxable wealth at death, coinciding with the probate process, which attempts to locate and determine the net worth of the deceased. The US estate tax has a sizeable exemption level, which as of 2020 is $11.58 million for singles and $23.16 million for married couples and features a flat rate of 40 percent over the exemption. The revenue it generates has eroded over time recently because of legislated

increases in the exemption level; in fiscal year 2019 it raised $16.7 billion, or slightly less than 0.5 percent of federal revenues—less than one-tenth of what the Sanders and Warren proposals projected to collect—from about 0.06 percent of decedents.

In principle, the estate tax is designed to target the superrich, but in practice many features of the law allow the wealthy to reduce their exposure significantly. Notably, the effective estate tax rate is reduced by extensive undervaluation of wealth transfers via, for example, family limited partnerships, which are holding companies owned by two or more family members created to retain a family's business interests, real estate, publicly traded and privately held securities. Due to the lack of control and lack of marketability that limited partners possess, these interests can be transferred to future generations at a discount to market value.

What is known about the consequences of an estate tax? Kopczuk (2013) surveys the evidence and concludes that the literature suggests an elasticity of reported estates—which includes changes in real wealth accumulation, avoidance, and evasion—with respect to the net-of-tax rate between 0.1 and 0.2. Eller, Erard, and Ho (2001) analyze estate tax evasion based on data from a stratified random sample of federal estate tax returns as filed and audit assessments, and they estimate the estate tax underreporting gap due to noncompliance to be 13 percent, but this figure may substantially understate the true magnitude of the gap, in part because it does not account for any noncompliance not detected during the IRS examination process.

Capital Gains Taxes

Any analysis of how the wealthy are taxed must confront how capital gains are taxed. For the superrich, realized capital gains represent a very high fraction of reported income. For example, IRS data shows that in tax year 2014 realized capital gains represented 60 percent of total adjusted gross income (AGI) for the 400 highest AGI Americans. In tax year 2016, those with adjusted gross income over $10 million reported net capital gains corresponding to 46 percent of their adjusted gross income, whereas it is a negligible fraction for those earning less than $200,000 (Scheuer and Slemrod 2020).

A capital gains tax is of course not a wealth tax, because it applies to gains rather than to the total value of the financial asset. Under current US law, capital gains are taxed at a lower rate than other income and are taxed upon realization (usually sale) rather than accrual, which generates a so-called deferral (or interest) advantage. Most importantly, capital gains unrealized at death completely escape income taxation. Instead, there is a "step-up" provision under which the value of an asset at the time of death becomes the tax basis for the inheritor, so that if sold the taxable capital gain is calculated as if value at time of inheritance was the purchase price. Even though some of the income that gives rise to the appreciation of assets (such as corporate stock) is subject to taxation at the corporate level, corporate tax rates have come down over time. Taken together with the extreme concentration of capital gains at the top, these provisions have led to concerns that the overall progressivity of the income tax is eroding.

There have been a number of proposals to alter the capital gains tax in ways that would restore tax progressivity without resorting to an annual wealth tax. As

one example, President-elect Joe Biden released a plan during the campaign to tax capital gains and dividends at the same rate as ordinary income for taxpayers with incomes exceeding $1 million and also to tax unrealized capital gains at death. Combined with other proposed changes in the income tax code, Biden's tax plan would raise the top marginal tax rate for capital gains from 20 percent to 39.6 percent. While Biden's plan would eliminate two of the preferential tax provisions for capital gains, it would retain the current system of taxation based on realization rather than accrual other than at death, thereby preserving the advantage of being able to defer taxes within a lifetime.

In view of this, calls for the taxation of accrued capital gains have been made. For example, Batchelder and Kamin (2019) offer a menu of "incremental" revenue options, including an accrual-based capital gains tax consisting of an annual mark-to-market tax on publicly traded assets plus a retrospective accrual tax for illiquid assets. Under a retrospective scheme, the capital gains tax is assessed upon realization, but the statutory tax rate rises as the holding period lengthens, effectively charging interest on past gains when realization occurs. This eliminates the need to value assets that are not actually being sold while minimizing liquidity problems and the incentive to defer such realization (Auerbach 1991).

Consequences of a Wealth Tax

Imposing a wealth tax will tend to reduce the amount of taxed wealth, due to some combination of changes in wealth accumulation, shifts in financial choices, and outright evasion. Here, we consider these various consequences, along with some discussion of the administrative and compliance costs of levying such a tax. We focus mostly on evidence from the European wealth taxes, but also consider some evidence from related taxes like the estate tax.

Evidence on the Overall Response to Wealth Taxes

Empirical studies of the behavioral response to wealth taxes are much sparser than for income taxes, largely because wealth taxes themselves are much rarer than income taxes. Moreover, because tax bases and relevant enforcement details vary widely, applying the evidence on the effect of one country's tax to another is problematic. Indeed, some potentially critical enforcement instruments, such as cross-country information exchange agreements that are designed to constrain tax evasion using foreign accounts, postdate essentially all of the studies.

We begin with a number of recent studies that find taxable wealth to be highly responsive to its tax rate. Brülhart et al. (2019) take advantage of variations in the Swiss wealth tax rate across cantons and over time and find that a 1 percentage-point decrease in wealth taxes increases reported taxable wealth after six years by at least 43 percent (and by 96 percent for the subset of large reforms). Comparing administrative tax records from two cantons suggests that about one-fourth of the effect comes from taxpayer mobility and another one-fifth from house price capitalization. They argue that savings responses cannot explain more than a small

fraction of the remainder, suggesting sizable evasion responses in this setting with no third-party reporting of financial wealth.

Jakobsen et al. (2020) examine changes in the Danish wealth tax that was cut back beginning in 1989 and abolished in 1997, taking advantage of two design aspects of this tax: a doubling of the exemption threshold for married couples and a cap on the ratio of income, payroll, and wealth taxes as a fraction of income that renders the marginal wealth tax equal to zero for those at the cap. For the very wealthy, they conclude that reducing the wealth tax rate by 1 percentage point would raise taxable wealth by 21 percent after eight years. Because the estimated effect grows over time, they argue that it could not be all a one-time avoidance effect. Instead, half of the long-run effect is mechanical since a higher wealth tax reduces wealth even when behavior is unchanged.

Also sizeable are the estimated elasticities of Durán-Cabré, Esteller-Moré, and Mas-Montserrat (2019) based on an analysis of the surprise reintroduction of a wealth tax in Catalonia in 2011. They find no evidence of it reducing wealth accumulation, but find that the tax triggered substantial tax avoidance via taxpayers changing their asset composition toward exempt assets (mainly company shares) and also that the return of the tax induced taxpayers to reduce taxable income to take advantage of an income-related cap on the sum of income and wealth tax liability. They find that a 1 percentage point reduction in the average wealth tax rate would lead to an increase in taxable wealth of 32 percent over four years. Agrawal, Foremny, and Martínez-Toledano (2020) look more closely at the migration response, focusing on the fact that all Spanish regions levied positive wealth tax rates except for Madrid. They conclude that, by five years after the reform, the number of wealthy individuals residing in Madrid for tax purposes increased by 10 percent relative to other regions, but conclude that misreporting rather than physical location change is likely the main factor. As in Brülhart et al. (2019), this applies to sub-national variation in wealth tax rates, where migration (or reported migration) is likely to be large relative to cross-national migration.

Zoutman (2018) studies a major reform to wealth and capital income taxation in the Netherlands that occurred in 2001. Comparing households that were similar in wealth and income, but treated differently by the reform, he concludes that a 1 percentage point decrease in the wealth tax rate leads to a long-run increase in accumulated wealth of 14 percent.

On the other side, Seim (2017) finds considerably smaller effects. Exploiting bunching around a kink in the Swedish wealth tax rate schedule where the rate changes from 0 percent to 1.5 percent, he estimates that a reduction in the wealth tax rate by 1 percentage point increases reported wealth by 0.10 to 0.27 percent. He concludes that the elasticity mainly represents reporting responses, and finds no evidence of households changing their saving or portfolio composition.[9]

----

[9] In a recent study of Colombia, Londoño-Velez and Ávila-Mahecha (2019) find evidence of bunching responses of reported wealth below notches in the tax rate structure and estimate that, in the short run, a 1 percentage point wealth tax cut increases reported wealth by 2 percent. They conclude that these

In sum, recent studies of the European experience suggest that the behavioral response to wealth taxation can be substantial, but that the anatomy of the response—real versus avoidance versus evasion—varies a lot, in large part because of differences in the broadness of the tax base. Advani and Tarrant (2020) offer a comprehensive review of these empirical studies and attempt to explain the varying results based on design features, contextual factors, and methodological differences.

There is an important interaction among these behavioral reactions. As Slemrod (2001) details in a more general context, the tax disincentive to real behavior depends on how the marginal cost of avoidance and evasion interacts with the real decision. In the extreme, a tax that can be costlessly evaded will provide no disincentives for real behavior. This insight suggests that when estimating the effects of a wealth tax, it is essential to understand how design differences might affect the costs of evading the tax. Indeed, supporters of the prominent US wealth tax proposals have suggested several reasons why it might be harder to evade than the European experience would indicate: for example, the United States is a much larger country, its tax system is citizenship-based rather than residence-based, the proposals involve much higher exemption thresholds, they are accompanied by plans to enhance tax enforcement, and their implementation would post date the adoption of the Foreign Account Tax Compliance Act (FATCA) in 2010. We take up some of these issues below.

Real Behavioral Responses

A wealth tax reduces the after-tax return to saving. The most important potential real behavioral response is in terms of reduced saving and capital accumulation. This effect is qualitatively the same as under other taxes on capital accumulation, such as a capital income tax (for an overview, see Bernheim 2002). As seen above, though, one difference is that a wealth tax can translate into higher capital income tax rates than are commonly imposed (potentially exceeding 100 percent), which presumably leads to larger effects.

Taxes that appear to be levied on the wealthy may instead be borne by others via tax-induced changes in pre-tax prices. For example, if a wealth tax reduces capital accumulation, in the long run it may reduce average wage rates. Such an argument figured prominently in the debate preceding the Tax Cuts and Jobs Act of 2017, when supporters argued that the proposed cut in the rate of corporate income tax would, via increased business investment and eventually a larger capital stock, increase average annual wages by as much as $9,000; this suggests an avenue through which taxing "their" wealth ends up affecting "our" wealth. This conclusion is highly controversial, however (for an overview of the arguments made at this time, see Slemrod 2018 in this journal).

responses reflect predominantly avoidance and evasion, such as misreporting wealth items subject to less third-party reporting. They also find that wealthy taxpayers increased compliance in response to incentives for the disclosure of previously hidden wealth as well as in response to an exogenous increase in the risk of detection and punishment due to the publication of the "Panama Papers."

A wealth tax could also affect work effort, but there is no consensus on the relevant labor supply elasticity. Notably, a substantial fraction of the very wealthy are either themselves or descendants of principals in a rather successful business venture: for example, of the wealthiest Americans on the 2018 Forbes 400 list, 69 percent were "self-made" founders of their business (Scheuer and Slemrod 2020). As a result, the relevant margin is probably not hours of work in the narrow sense. Instead, the key effects may be on the incentives for entry into entrepreneurship (Cullen and Gordon 2007; Scheuer 2014; Shourideh 2014)) and on the ownership and control structure of business enterprises.

Due to the highly progressive nature of the wealth tax, it could, for example, discourage entrepreneurial risk taking. Hall and Woodward (2020) document that entrepreneurial risk is highly skewed, with most venture-capital backed start-up companies faring poorly and a few performing exceptionally well. Due to incentive problems, this risk cannot be diversified, which limits the attractiveness of entrepreneurship under reasonable risk aversion, so further reducing entry might seem like a bad idea. However, because a risk-averse individual will have relatively low marginal utility in case of very good outcomes, the effect on decisions to participate in entrepreneurship of a wealth tax that applies only in those low-probability states of the world could be modest.

Another concern is that a wealth tax might force entrepreneurs to reduce their ownership in a company whose valuation increases over time in order to pay the tax liability. Even if such founders are not primarily motivated by monetary incentives, but instead are mostly interested in being able to realize their ideas, such an anticipated dilution of control rights could have discouraging effects on entrepreneurial activity.

Might a US wealth tax induce some people to move out of the country? Because the US taxes on the basis of citizenship rather than residence, moving does not relieve an American citizen of any tax obligations—instead, citizenship renunciation is required. There are some prominent examples: Facebook co-founder Eduardo Saverin dropped his US citizenship in favor of Singapore just prior to the Facebook initial public offering in 2012. But overall, US citizenship renunciation by the wealthy has been very small. Between 2005 and 2017, more than 30,000 individuals dropped their US citizenship, of whom fewer than 100 reported net worth greater than $100 million (Organ 2020). Overall, however, about one-third of those dropping citizenship were millionaires in terms of wealth, compared with only about 5–6 percent in the US population. An increase in renunciations in the 2010s was probably due to increased enforcement of tax evasion using offshore accounts, prompting renunciation by dual citizens already residents abroad. However, there is no historical precedent to help gauge the renunciation response to a wealth tax at rates far above existing levels.[10]

---

[10] Senator Warren proposed a 40 percent exit tax on the net worth above $50 million of any US citizen who renounces their citizenship, while Senator Sanders proposed a 40 percent exit tax on the net value of all assets under $1 billion and a 60 percent exit tax for those with wealth exceeding $1 billion. If enforced, these measures might greatly limit any potential exit responses.

Avoidance

One way to reduce wealth tax liability is to substitute assets that face lower tax rates, or to hold assets for which the value is harder to monitor and thus easier to understate successfully. Spain offers a stark example: when it exempted some forms of closely held businesses from its wealth tax base, the share of the exempted stock as a share of all closely held businesses increased from 15 to 77 percent (Alvaredo and Saez 2009).

In a US context, a wealth tax might lead some high net-worth individuals to shift into assets that are harder to value, such as keeping businesses private rather than going public. Start-up firms might forego equity infusions to avoid new valuation rounds, which could constrain their expansion, or they could start issuing non-standard, less transparent types of stocks. Hemel (2019) offers the example of companies deciding not to offer their shares on public equity markets, even if a public offering would be the most efficient means of raising capital, because a more transparent valuation will lead to a larger wealth tax liability for its shareholders. Much wealth of the Forbes 400, for example, is currently held in publicly traded stock, but this feature cannot be taken as unresponsive to a potential wealth tax. This is an example of a potentially substantial and distorting behavioral response of which there is no trace in existing data; how likely it is to occur and what enforcement responses might constrain it are very hard to know. Such shifting into less visible assets would also have repercussions for our measures of wealth inequality: it might look like a wealth tax reduces concentration when in reality it partly shifts top wealth into forms that are less susceptible to accurate measurement.[11]

Evasion

Government auditors typically lack the resources to trace sophisticated means of wealth tax evasion—say, methods that work through layers of financial intermediaries. High-profile leaks from these intermediaries, such as the 2007 leak from HSBC Bank in Switzerland and the 2015 "Panama Papers" from the firm Mossack Fonseca, have allowed researchers to gain insights into these forms of tax evasion. Alstadsæter, Johannesen, and Zucman (2019) link the account names from the HSBC leak with individual tax data for Norway, Sweden, and Denmark and find that 95 percent of these foreign account-holders did not report the existence of the account to the tax agency. They show that evasion rates rise sharply across the income distribution and conclude that the top 0.01 percent in the income distribution evade about 25 percent of the income and wealth taxes they owe. Guyton et al. (2020) combine random audit data with data on offshore bank accounts and show that tax evasion for US taxpayers through offshore financial institutions is highly concentrated at the very top of the income distribution, and that random audits virtually never detect this form of evasion.

Despite this new evidence, we do not yet know the extent to which a wealth tax at much higher rates would be susceptible to evasion, although some of the studies

---

[11] Another avenue of wealth tax avoidance is inter vivos gifts. Research suggests that these gifts are tax-sensitive (for example, see Bernheim, Lemke, and Scholz 2004; Joulfaian and McGarry 2004).

of European wealth taxes suggest substantial evasion. Its extent will certainly depend on the enforcement environment, which is evolving. The Foreign Account Tax Compliance Act (FATCA) of 2010 set up third-party reporting requirements based on existing tax information exchange agreements. Through threat of a punitive withholding tax for non-complying foreign financial institutions, FATCA provides US tax subjects with strong incentives to report to the IRS the value and income generated by their foreign accounts.[12] Both the Sanders and Warren wealth tax plans would expand enforcement further, proposing significant increases in the IRS enforcement budget and a minimum audit rate for taxpayers subject to the wealth tax.

How effective such expanded enforcement would be in restraining evasion has been controversial. Saez and Zucman (2019a) claim that evasion would shrink the wealth tax base by just 15 percent. Kopczuk (2019) expresses skepticism, noting that the most effective tax enforcement relies on market transactions reported by third parties, which would be absent for much wealth. This is not purely an enforcement problem because, as mentioned, the valuation of many assets is objectively hard. Clever ideas have been put forward to address this problem; for example, Allais (1977) proposes that wealth owners self-report the value of their assets but then the government (or any other private bidder) could acquire these assets at a surcharge of 40 percent (respectively, 50 percent). Such schemes come with their own difficulties, though, especially with opaque assets, not to mention the political concerns about the government owning a large share of businesses in the economy.

One difference between the wealth tax and the estate tax is that the former requires reporting at a much higher frequency. While this potentially raises compliance costs, the upside is that any evasion strategy must engineer an entire path of reports that is plausible on a yearly basis, notably relative to yearly income, rather than just one end-of-life snapshot. This may make it harder to conceal wealth systematically than in the case of the estate tax, which allows for decades of planning without generating much data for tax authorities.

Administrative and Compliance Costs

A wealth tax imposes costs of collection, including the compliance costs borne by taxpayers and third parties and the administrative costs borne by the government. Leiserson (2019) extrapolates from experience with the US estate tax to estimate the ratio of private compliance costs to revenues from a 2 percent wealth tax at 19 percent, which is approximately double the conventional wisdom about the US income tax; he estimates government administrative costs to be just 0.6 percent of revenue. Troup, Barnett, and Bullock (2020) estimate the compliance cost to be 1 to 1.5 percent of total wealth in the first year based on the legal costs of the probate process in the United Kingdom, which (depending on the tax rate) would

---

[12] The international version of FATCA, known as the Automatic Exchange of Financial Account Information in Tax Matters (the AEOI Standard), began in September 2017 and, by 2019, 94 countries had exchanged information. Johannesen et al. (2020) provide evidence on the impact of pre-FATCA enforcement policies aimed at foreign accounts held by US taxpayers.

mean costs roughly equivalent to tax revenue. However, the compliance costs would fall in subsequent years due to repetition effects. Moreover, due to the fixed-cost nature of valuation and reporting efforts, the compliance cost relative to revenue declines if a higher wealth tax rate is applied.

Are Wealth Taxes Part of an Optimal Tax System?

We now turn to the normative question whether wealth accumulation should be taxed and, if so, the extent to which a wealth tax should be a preferred mechanism for doing so. A growing literature in public economics has started to incorporate more realistic labor markets into models of optimal tax policy, accounting for phenomena such as rent-seeking, skill-biased technological change, and superstar effects, to name just a few advances (for an overview, see Scheuer and Slemrod 2020). This line of work has focused on the optimal design of labor income taxes in static models that capture recent trends in occupational sorting and wage inequality. Because a growing concentration of earnings can affect, through savings, the degree of wealth inequality down the road, this raises the question whether these trends also affect the optimal taxation of capital or wealth.

The Atkinson-Stiglitz Benchmark

Suppose first that all wealth inequality is driven by inequality in labor incomes. In this case, the Atkinson-Stiglitz (1976) theorem provides a classic benchmark. It states that, if a nonlinear labor income tax is available, any distortion of savings is Pareto-inefficient whenever preferences satisfy two conditions: (i) they are separable between consumption and labor; and (ii) all individuals have the same utility over consumption across time. In other words, if individuals only differ in their labor productivities, and the recent rise in wealth inequality is the result of changes in labor markets, then the policy response should be to adjust the progressivity of labor income taxes. The taxation of capital income or wealth on top of that is not justified.

This theorem is a conceptually useful baseline, but the underlying assumptions are not realistic. First, Saez (2002) has suggested a positive correlation between labor productivities and savings propensities. This violates the Atkinson-Stiglitz condition (ii), because individuals will differ in their discount rates in a way that is related to earnings abilities. Similarly, when individuals differ in their rates of return on their wealth, this lends support to the additional taxation of capital (Gerritsen et al. 2020). Second, disentangling labor and capital income can be challenging in practice, so capital income taxes may be needed to minimize revenue-losing tax-base shifting from labor to capital income (Christiansen and Tuomala 2008). Third, when agents face uncertainty and are risk averse, taxing capital income can improve incentives for labor supply (Golosov, Tsyvinski, and Werning 2006). Fourth, current policymakers do not face a blank slate, but instead face a situation with preexisting wealth inequality. Individuals already differ in the wealth they own, either because they have inherited it from previous generations or because they themselves have saved in the past.

A One-Time Tax on Existing Wealth

In principle, preexisting wealth inequality could be alleviated in a lump-sum fashion through a one-time, unanticipated wealth tax. Indeed, historically, various countries have used one-time wealth taxes to deal with revenue shortfalls, such as wartime spending shocks. In 1999, Donald J. Trump, then a candidate for the Reform Party presidential nomination, proposed a 14.25 percent one-time "net-worth tax" on individuals and trusts worth more than $10 million in order to eliminate the US national debt. More recently, calls have been made for a time-limited, progressive wealth levy to stem the fiscal burden arising from the coronavirus pandemic (for example, Landais, Saez, and Zucman 2020).

From an optimal tax perspective, such policies are attractive because they avoid behavioral distortions by only touching wealth that has already been accumulated.[13] But this appealing feature critically hinges on the ability of policymakers to implement such policies on short notice and on their commitment not to make such taxes permanent or to reintroduce them periodically when similar times come about in the future, which would lead to reputational damage. In the past, originally one-off war taxes have often turned into long-lasting tax policies.

Taxing Future Wealth Accumulation

If an unexpected, distortion-free redistribution of existing wealth is not feasible, one policy option is to adjust the labor income tax going forward. If initial wealth and earnings abilities are positively correlated, a more progressive labor income tax could be used to target both determinants of inequality, at least indirectly. The alternative is to introduce a tax on future wealth accumulation, which will of course distort the savings incentives of individuals.

In the online Appendix available with this article at the *JEP* website, we provide a formal demonstration that a tax on wealth accumulation will be part of the optimal tax mix, even if preferences satisfy conditions (i) and (ii), and that the optimal schedules of the wealth and labor income tax are closely linked. For example, if the importance of initial wealth relative to labor income inequality increases toward the top of the distribution, then the wealth tax should be more progressive than the labor income tax (and vice versa). Moreover, we show that the optimal marginal wealth tax is decreasing in the intertemporal elasticity of substitution (because the savings distortions increase) and increasing in the Frisch elasticity of labor supply (because this increases the distortions from the labor tax, making the wealth tax relatively more attractive). We also provide a sufficient-statistics formula for the top marginal wealth tax that can be calibrated using the shape of the income and wealth distribution.

In sum, unless there is already a fully equalized wealth distribution, it is generally optimal to introduce progressive taxes on future wealth accumulation on top of labor income taxes, despite their distortive effects, at least for some amount of time. In the long run, of course, the effect of initial wealth on overall inequality

---

[13] Taxes on existing wealth can be replicated, in principle, by consumption taxes coupled with subsidies on labor income. Moreover, even though taxing preexisting wealth has no incentive effects, it has redistributive effects across age cohorts because those who are older tend to have more wealth.

will diminish. Indeed, the influential Chamley-Judd result argued that zero capital taxation is optimal in the long run; Judd (1985) argues that this result holds even in the face of extreme wealth inequality, while Chamley (1986) instead considers long-run capital taxation in a representative agent framework. However, Straub and Werning (2020) have recently demonstrated that Judd's result is invalid whenever the intertemporal elasticity of substitution is at most one—which seems the empirically relevant level—and the long-run tax on capital should in fact be positive and significant. For higher elasticities, it converges to zero but possibly at a very slow rate—for decades, or even centuries.[14]

Wealth versus Capital Income Taxes

We conclude that the modern theory of optimal taxation lends support to taxing wealth accumulation. However, the existing literature does not pin down the appropriate tax instruments to use for this purpose. As discussed earlier, in standard models, the wealth tax is equivalent to a tax on the returns to capital income. Given that most countries already have progressive capital income taxes, for instance, what might justify levying a wealth tax instead, or in addition?

When individuals differ in the rates of return r on their wealth, there is a tradeoff between wealth taxes and capital income taxes that depends on the source of these differences. Because the capital-income-tax equivalent of a given wealth tax rate t_w is given by t_w/r, Allais (1977) points out that a wealth tax favors wealth-holders with high rates of return relative to a capital income tax. Hence, relative to a capital income tax, a wealth tax encourages the reallocation of capital from "idle" wealthholders to productive entrepreneurs. In a quantitative model, Guvenen et al. (2019) find significant efficiency gains from this effect compared to a uniform capital income tax.

There is, however, an opposing effect. If heterogeneous returns reflect heterogeneous windfall gains, rents, or excess profits (perhaps due to market power or inside information), rather than actual productivity differences, then taxing away such gains has well-known efficiency benefits (Rothschild and Scheuer 2016). But a wealth tax gets this exactly reversed—it taxes the normal rate of return and leaves the excess returns untouched. For example, if all investors have a real rate of return of 3 percent, but some earn additional excess profits on their investments, then a 3 percent wealth tax would not target any of those rents, whereas a capital income tax would.

A related issue is that much of what shows up as return to capital on the tax reports of the superrich (for example, in the form of realized capital gains) is arguably compensation to labor from the work that went into building a successful company or picking high-performing assets. In Scheuer and Slemrod (2020), we argue that the ability to convert this kind of labor income into preferentially taxed capital gains is a key margin of behavioral response to taxes at the top, which we

[14]Chamley's (1986) model imposes an upper bound on capital taxes, and Straub and Werning (2020) provide conditions under which this bound is binding forever at the optimum, also implying positive long-run capital taxes. Saez and Stantcheva (2018) show that when wealth enters utility directly, in addition to the consumption it finances, the optimal long-run capital tax is also positive.

refer to as the "plasticity" of the tax base. A wealth tax only taxes some normal return, whereas a capital income tax hits the full extent of such shifted labor compensation.

Some progress has been made in measuring the extent and nature of return heterogeneity (for example, Fagereng et al. 2020). However, a comprehensive decomposition into actual productivity differences versus differential rents or shifted labor compensation has not yet been accomplished.

Political Economy

The case for a wealth tax often reaches beyond specifically economic questions of tax incidence and redistribution and is based on a concern that rising inequality of income and wealth may lead to adverse political outcomes. As one example, excessive inequality might allow the rich to capture the political system and tilt it in their favor. Concerns from a somewhat different angle go back to Karl Marx, who predicted that an increase in the concentration of wealth would lead to a revolution and to radical redistribution. We briefly consider both perspectives and the extent to which they can justify a wealth tax.

Wealth and Political Power

Even in a "one-person, one-vote" democracy, the superrich can affect politics more than others through campaign contributions, ownership of media outlets, or lobbying activities. Gilens (2014) and Bartels (2016) collect evidence that political decisions often are more sensitive to the preferences of the rich than those of the median voter. Accordingly, proponents of a progressive wealth tax have argued that reducing the wealth of the superrich is a desirable objective in itself, beyond the revenue it could raise to effect redistribution. Indeed, Saez and Zucman (2019b) propose setting wealth tax rates above the revenue-maximizing rate, expressing a willingness to reduce the wealth of the superrich in the interest of preventing an "oligarchic drift" that would otherwise undermine democracy. Of course, there is some tension between enacting a wealth tax to fund redistribution initiatives (such as "Medicare for All" in some recent US proposals) and enacting a wealth tax with the goal to reduce top wealth rather than collect revenue.

Even if concerns about an extreme concentration of wealth and political power are warranted, the jury is still out on the extent to which a wealth tax is the appropriate tool to address the problem. Other instruments may be better targeted at ensuring a more equal political representation, such as regulating campaign contributions and public financing of political campaigns. Some European countries offer examples of democracies where money plays a smaller role in politics than in the United States, and Brechenmacher (2018) concludes that wealthy elites exerting disproportionate political influence is a distinctly US phenomenon. One particular concern with the wealth tax is that it might encourage the wealthy to become more politically active, in an attempt to reduce their wealth tax liability through, for example, political donations.

Politically Sustainable Tax Policy

One can argue that the primary role of a wealth tax is to make tax policy and the resulting inequality more stable, so that it can resist the threat of political upheaval (Farhi et al. 2012; Piketty 2014). Such threats were important drivers of tax and welfare state policies in 19th- and 20th-century Europe, when the socialist movement gained momentum (Esping-Andersen 1990), and they are palpable today in many South American countries. If political instability is an urgent problem, annual wealth taxes are able to compress the wealth distribution relatively quickly compared to, say, taxes on bequests or capital income.

One approach to modelling this question is to focus on tax policies that will maintain the support of a majority of citizens over time. Scheuer and Wolitzky (2016) show that the optimal sustainable tax policy involves a positive marginal tax on the wealth accumulation of the rich, while subsidizing that of the middle class. At any given time, there is always a temptation to impose wealth taxes, because at this point, wealth accumulation is sunk. However, if the future is likely to bring near-confiscatory wealth taxes, then individuals anticipating this outcome would save very little in the first place, leading eventually to a poor outcome for everyone. Hence, it is better to tax the savings of the rich at least to some degree and create a middle class that accumulates enough wealth to successfully oppose more extreme redistribution in the future.

Of course, if the issue is reducing the impetus for political disruption based on tensions related to economic inequality, the wealth tax needs to be compared to a range of other political alternatives: for example, a combination of a progressive increase in income tax rates, more tax audits, expanding the estate tax, reforms to capital gains taxation, refocusing government spending on those with lower income levels, or an expansion of social insurance programs.

Conclusion

In recent years, many European countries decided that a wealth tax did not belong in their armory of tax instruments. Although the United States has never had such a tax, perceptions of unacceptably high income and wealth inequality have recently galvanized support for one, and two prominent US senators have produced detailed proposals. These proposals differ quite substantially from the experience of their European counterparts. Thus, the evidence about the consequences of wealth taxation in Europe is in any event of limited usefulness. On one hand, the broader base along with promised expanded enforcement will limit the revenue leakage and distortion from avoidance and evasion, while exacerbating real behavioral responses. On the other hand, the higher top rates and targeting of the superrich will concentrate the revenue pressure on those taxpayers with the best means and strongest incentives to avoid the tax. Hence, when evaluating these US wealth tax proposals, one can at best hold one's breath and extrapolate broadly from the European wealth tax experience and the US experience with similar taxes, and gain insight from optimal tax reasoning about whether to tax capital via an

annual wealth tax. Given rising economic inequality in the United States, proposals for taxes on wealth accumulation in some form are likely to remain an ongoing subject of debate.

References

Advani, Arun, and Hannah Tarrant. 2020. "Behavioural Responses to a Wealth Tax." Wealth Tax Commission Evidence Paper 5.

Agrawal, David, Dirk Foremny, and Clara Martínez-Toledano. 2020. "Paraísos Fiscales, Wealth Taxation, and Mobility." Unpublished.

Allais, Maurice. 1977. "L'Impôt sur le Capital et la Réforme Monétaire." Paris: Hermann.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman. 2019. "Tax Evasion and Inequality." *American Economic Review* 109 (6): 2073–2103.

Alvaredo, Facundo, and Emmanuel Saez. 2009. "Income and Wealth Concentration in Spain from a Historical and Fiscal Perspective." *Journal of the European Economic Association* 7 (5): 1140–67.

Atkinson, A., and J. Stiglitz. 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6: 55–75.

Auerbach, Alan. 1991. "Retrospective Capital Gains Taxation." *American Economic Review* 81 (1): 167–78.

Bartels, Larry. 2016. *Unequal Democracy: Political Economy of the New Gilded Age*. Princeton: Princeton University Press.

Batchelder, Lily, and David Kamin. 2019. "Taxing the Rich: Issues and Options." Unpublished.

Bernheim, B. Douglas. 2002. "Taxation and Saving." *Handbook of Public Economics* 3: 1173–1249.

Bernheim, B. Douglas, Robert Lemke, and John Karl Scholz. 2004. "Do Estate and Gift Taxes Affect the Timing of Private Transfers?" *Journal of Public Economics* 88 (12): 2617–34.

Brechenmacher, Saskia. 2018. *Comparing Democratic Distress in the United States and Europe*. Washington, DC: Carnegie Endowment for International Peace.

Brülhart, Marius, Jonathan Gruber, Matthias Krapf, and Kurt Schmidheiny. 2019. "Behavioral Responses to Wealth Taxes: Evidence from Switzerland." CEPR Discussion Paper 14054.

Chamley, Christophe. 1986. "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives." *Econometrica* 54 (3): 607–22.

Christiansen, Vidar, and Tuomala, Matti. 2008. "On Taxing Capital Income with Income Shifting." *International Tax and Public Finance* 15: 527–45.

Cullen, J., and R. Gordon. 2007. "Taxes and Entrepreneurial Risk-Taking: Theory and Evidence for the U.S." *Journal of Public Economics* 91: 1479–1505.

Drometer, M., M. Frank, M. Hofbauer Pérez, C. Rhode, S. Schworm, and T. Stitteneder. 2018. *Wealth and Inheritance Taxation: An Overview and Country Comparison*. Munich: Ifo Institute.

Duran-Cabré, José, Alejandro Esteller-Moré, and Mariona Mas-Montserrat. 2019. "Behavioural Responses to the (Re)Introduction of Wealth Taxes. Evidence from Spain." IEB Working Paper 2019/04.

Du Rietz, Gunnar, and Magnus Henrekson. 2014. "Swedish Wealth Taxation, 1911–2007." IFN Working Paper 1000.

Eidgenössische Steuerverwaltung. 1969. *Steuerbelastung in der Schweiz.* Bern: Eidgenössische Steuerverwaltung.

Eller, Martha Britton, Brian Erard, and Chih-Chin Ho. 2001. "Noncompliance with the Federal Estate Tax." In *Rethinking Estate and Gift Taxation*, edited by William G. Gale, James R. Hines Jr., and Joel Slemrod, 375–410. Washington, DC: Brookings Institution Press.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism.* Princeton: Princeton University Press.

Fagereng, Andreas, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri. 2020. "Heterogeneity and Persistence in Returns to Wealth." *Econometrica* 88 (1): 115–70.

Farhi, Emmanuel, Christopher Sleet, Iván Werning, and Sevin Yeltekin. 2012. "Non-linear Capital Taxation without Commitment." *The Review of Economic Studies* 79 (4): 1469–93.

Federal Department of Finance. 2019. "Besteuerung nach dem Aufwand." https://www.efd. admin. ch/efd/de/home/themen/steuern/steuern-national/besteuerung-nach-dem-aufwand/ fb-besteuerung_nach_aufwand.html.

Föllmi, Reto, and Isabel Martínez. 2017. "Volatile Top Income Shares in Switzerland? Reassessing the Evolution between 1981 and 2010." *The Review of Economics and Statistics* 99 (5): 793–809.

Gerritsen, Aart, Bas Jacobs, Alexandra Victoria Rusu, and Kevin Spiritus. 2020. "Optimal Taxation of Capital Income with Heterogeneous Rates of Return." CESifo Working Paper 8395.

Gilens, Martin. 2014. *Affluence and Influence: Economic Inequality and Political Power in America.* Princeton: Princeton University Press.

Golosov, Mikhail, Aleh Tsyvinski, and Ivan Werning. 2006. "New Dynamic Public Finance: A User's Guide." *NBER Macroeconomics Annual* 21 (2006).

Guvenen, Fatih, Gueorgui Kambourov, Burhanettin Kuruscu, Sergio Ocampo-Diaz, and Daphne Chen. 2019. "Use It or Lose It: Efficiency Gains from Wealth Taxation." Federal Reserve Bank of Minneapolis Working Paper 764.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman. 2020. "Tax Evasion by the Wealthy: Measurement and Implications." In *Measuring and Understanding the Distribution and Intra/Inter-Generational Mobility of Income and Wealth*, edited by Raj Chetty, John Friedman, Janet Gornick, Barry Johnson, and Arthur Kennickell.

Hall, Robert, and Susan Woodward. 2010. "The Burden of the Nondiversifiable Risk of Entrepreneurship." *American Economic Review* 100 (3): 1163–94.

Harris, Benjamin H., and Brian David Moore. 2013. *Residential Property Taxes in the United States.* Washington, DC: Urban-Brookings Tax Policy Center.

Hemel, Daniel. 2019. "Taxing Wealth in an Uncertain World." *National Tax Journal* 72 (4): 755–76.

Herd, Richard, and Thorsteinn Thorgeirsson. 2001. "Increasing Simplicity, Neutrality and Sustainability: A Basis for Tax Reform in Iceland." OECD Economics Department Working Paper 292.

Hongler, Peter, and Fabian Mauchle. 2020. "Is Switzerland a Role Model for Wealth Taxes?" *Tax Notes International* 97 (6): 645–53.

Jakobsen, Katrine, Kristian Jakobsen, Henrik Kleven, and Gabriel Zucman. 2020. "Wealth Taxation and Wealth Accumulation: Theory and Evidence from Denmark." *Quarterly Journal of Economics* 135 (1): 329–88.

Jensen, Erik. 2019. "Is a Tax on Wealth Constitutional?" *Journal of Taxation of Investments* 36 (3): 79-86.

Johannesen, Niels, Patrick Langetieg, Daniel Reck, Max Risch, and Joel Slemrod. 2020. "Taxing Hidden Wealth: The Consequences of US Enforcement Initiatives on Evasive Foreign Accounts." *American Economic Journal: Economic Policy* 12 (3): 312–46.

Johnsen, Dawn, and Walter Dellinger. 2018. "The Constitutionality of a National Wealth Tax." Indiana Law Journal 93.

Joulfaian, David, and Kathleen McGarry. 2004. "Estate and Gift Tax Incentives and Inter Vivos Giving." *National Tax Journal* 57 (2): 429–44.

Judd, Kenneth. 1985. "Redistributive Taxation in a Simple Perfect Foresight Model." *Journal of Public Economics* 28: 59–83.

Kaldor, Nicholas. 1956. *Indian Tax Reform.* Delhi: Ministry of Finance, Government of India.

Kessler, Denis, and Pierre Pestieau. 1991. "The Taxation of Wealth in the EEC: Facts and Trends." *Canadian Public Policy* 17 (3): 309–21.

Kopczuk, Wojciech. 2013. "Taxation of Intergenerational Transfers and Wealth." In *Handbook of Public Economics*, Vol. 5, edited by Alan Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, 329–90. Amsterdam: Elsevier.

Kopczuk, Wojciech. 2019. "Comment on 'Progressive Wealth Taxation' by Saez and Zucman." *Brookings Papers on Economic Activity* 2019: 512–26.

Krenek, Alexander, and Margit Schratzenstaller. 2018. "A European Net Wealth Tax." WIFO Working Paper 561.

Landais, Camille, Emmanuel Saez, and Gabriel Zucman. 2020. "A Progressive European Wealth Tax to Fund the European COVID Response." *Vox EU*, April 3. https://voxeu.org/article/progressive-european-wealth-tax-fund-european-covid-response.

Lehner, Moris. 2000. "The European Experience with Wealth a Tax: Comparative Discussion." *Tax Law Review* 53: 615–92.

Leiserson, Greg. 2019. "Taxing Wealth." Unpublished.

Londoño-Velez, Juliana, and Javier Ávila-Mahecha. 2019. "Can Wealth Taxation Work in Developing Countries? Quasi-Experimental Evidence from Colombia." Unpublished.

McDonnell, Thomas A. 2013. "Wealth Tax: Options for its Implementation in the Republic of Ireland." NERI Working Paper 6.

OECD. 2018. *The Role and Design of Net Wealth Taxes in the OECD*. OECD.

Organ, Paul. 2020. "Expatriation and the U.S. Tax System." Unpublished.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press.

Rothschild, Casey, and Florian Scheuer. 2016. "Optimal Taxation with Rent-Seeking." *The Review of Economic Studies* 83 (3): 1225–62.

Saez, Emmanuel. 2002. "The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes." *Journal of Public Economics* 83 (2): 217–30.

Saez, Emmanuel, and Stefanie Stantcheva. 2018. "A Simpler Theory of Optimal Capital Taxation." *Journal of Public Economics* 162: 120–42.

Saez, Emmanuel, and Gabriel Zucman. 2019a. "Progressive Wealth Taxation." *Brookings Papers on Economic Activity* 2019: 437–511.

Saez, Emmanuel, and Gabriel Zucman. 2019b. *The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay*. New York: W. W. Norton & Company.

Sandford, Cedric, and Oliver Morrissey. 1985. "The Irish Wealth Tax: A Case Study in Economics and Politics." Economic and Social Research Institute Paper 123.

Scheuer, Florian. 2014. "Entrepreneurial Taxation with Endogenous Entry." *American Economic Journal: Economic Policy* 6 (2): 126–63.

Scheuer, Florian, and Joel Slemrod. 2020. "Taxation and the Superrich." *Annual Review of Economics* 12: 189–211.

Scheuer, Florian, and Alexander Wolitzky. 2016. "Capital Taxation under Political Constraints." *American Economic Review* 106 (8): 2304–28.

Seim, David. 2017. "Behavioral Responses to Wealth Taxes: Evidence from Sweden." *American Economic Journal: Economic Policy* 9 (4): 395–421.

Shourideh, Ali. 2014. "Optimal Taxation of Wealthy Individuals." Unpublished.

Silfverberg, Christer. 2009. "The Swedish Net Wealth Tax—Main Features and Problems." https://scandinavianlaw.se/pdf/44-22.pdf.

Slemrod, Joel. 2001. "A General Model of the Behavioral Response to Taxation." *International Tax and Public Finance* 8: 119–28.

Slemrod, Joel. 2018. "Is This Tax Reform, or Just Confusion?" *Journal of Economic Perspectives* 32 (4): 73–96.

Straub, Ludwig, and Iván Werning. 2020. "Positive Long-Run Capital Taxation: Chamley-Judd Revisited." *American Economic Review* 110 (1): 86–119.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (5): 416–24.

Troup, Edward, John Barnett, and Katherine Bullock. 2020. "The Administration of a Wealth Tax." Wealth Tax Commission Evidence Paper 11.

Waldenström, Daniel. 2018. "Inheritance and Wealth Taxation in Sweden." Munich: Ifo Institute.

Wolff, Edward. 2017. *A Century of Wealth in America*. Cambridge, MA: Harvard University Press.

Zoutman, Floris. 2018. "The Elasticity of Taxable Wealth: Evidence from the Netherlands." Unpublished.

Melissa Dell: Winner of the 2020 Clark Medal

Daron Acemoglu

The 2020 John Bates Clark Medal of the American Economic Association was awarded to Melissa Dell, professor of economics at Harvard University, for her path-breaking contributions in political economy, economic history, and economic development. Her work is distinguished and admired for its combination of painstaking data collection, careful empirical implementation, and audacious ambition.

Melissa is a native of Enid, Oklahoma, a relatively low-income rural community with a population of about 50,000. She was the first member of her family to go to college, and also the first person from her high school, the Oklahoma Bible Academy, to study at Harvard, where she began her undergraduate studies in 2001. She started in the Social Studies program at Harvard but soon realized that the central questions that interested her necessitated a more systematic empirical analysis and switched to the Department of Economics. Her performance as an undergraduate was beyond impressive: from the Harvard economics department, she was awarded the John Williams prize for the best student and the Seymour Harris prize for the best thesis. From outside Harvard, she was awarded a Truman Fellowship and a Rhodes Scholarship, which she used to complete a master's degree in economics at Oxford. She continued her studies with a PhD in economics at MIT (where I was fortunate to be her advisor).

In this essay, I will attempt to put Melissa's achievements and research in the broader context of what I am going to call *the fundamental questions of long-run economic*

■ *Daron Acemoglu is Institute Professor, Massachusetts Institute of Technology, Cambridge, Massachusetts. His email address is daron@mit.edu.*

Melissa Dell

development: Why are some countries or regions so much richer than others? Why is there so much inequality? Why has growth since the middle of the 18th century been so unequal, and why did it start taking off in the second half of the 18th and early 19th centuries in some parts of the world but not others? Furthermore, what do these outcomes have to do with democracy, political participation, and constitutions or a lack thereof?

These fundamental questions have fascinated and energized not just economists but a wide array of social scientists, historians, and philosophers. Many aspiring economists (and I count my younger, 18 year-old self among their ranks) are attracted to economics by these questions. Yet many young economists are soon discouraged from delving into these issues; sometimes they are told that these questions are too broad or too big for careful empirical work, and sometimes they are even told that the big questions fall outside the boundaries of economics.

However, Melissa's work has broadened and illuminated the way we think about the fundamental questions. She has used state-of-the-art microeconometric techniques, like regression discontinuity approaches, to produce more credible evidence on the role of institutions and historical factors in the fundamental questions of long-run economic development. Instead of comparing countries that differ in many dimensions, she has shown how it is possible to focus on more granular variation, such as between villages or municipalities. It turns out that zeroing in on specific settings not only makes causal inference more straightforward, but it also allows a much deeper understanding of the economic impacts of institutions, how institutions persist or atrophy, and how various dimensions of institutions may work very differently depending on the historical and social context. For example, what were the effects of the labor coercion as practiced by Spanish colonialists in

the Andes region for mining? What were the long-lasting consequences of Mexican revolution, which seem to have arisen from the policies used to de-escalate that conflict? Why did the sugar cultivation system imposed by Dutch colonists in Java have unexpected sectoral effects? How did the way in which local states took shape in Vietnam influence the long-run development of different parts of the country?

I begin with a few words of background on the fundamental questions of long-run economic development and then take a closer look at some of Melissa's most prominent work concerning these topics and others in three broad areas: 1) colonial history, institutions, and modern development; 2) conflict, law enforcement, and politics; 3) other research including climate change and economic growth and new methods for data extraction. Throughout, I refer to her key papers by number as listed in Table 1.

Background: Fundamental Questions of Long-Run Economic Development

Most of the modern work in the fields of economic growth and development economics does not seek to address the fundamental questions of long-run economic development. Researchers in economic growth have typically worked in a framework focused on changes in physical capital, human capital, and technology, while looking at topics like "how does capital or human capital accumulate" or "how can we model endogenous and sustained technological change." Development economists have focused on health conditions, schooling, imperfect credit markets, discrimination, and other issues critical for the lives of the poor in the developing world. But neither line of research looks at how historical and institutional factors shape political and economic trajectories and divergences.

Of course, there has always been a tradition in economics that was motivated by and tried to grapple with these fundamental questions. After all, the full title of Adam Smith's (1776) magnum opus is *An Inquiry into the Nature and Causes of the Wealth of Nations*, and Smith situated his analysis firmly in its institutional context. Marx (1867), too, had an "institutional" theory of divergence, first linking much of the potential of an economy as well as the nature of the inequalities that it created in particular to the ownership of the means of production. This "mode of production" then shaped other aspects of society, including its political and social system, and determined economic growth via its impact on the rate of capital accumulation and technological change.[1] Many 20th-century economists who focused on long-run economic development and institutions were influenced—or in some

---

[1] Marx did at times sound like a technological determinist, linking the mode of production and all else to technology: for example, he asserted in *The Poverty of Philosophy* that "the hand-mill gives you society with the feudal lord, the steam-mill society with the industrial capitalist" (1847 [1920], p. 49). But Marx's thinking was often more nuanced, recognizing the autonomous role of institutions, perhaps most distinctively in his *18th Brumaire of Louis Napoleon* (Marx 1852).

Table 1

Selected Papers by Melissa Dell

1. "The Persistent Effects of Peru's Mining Mita." 2010. *Econometrica* 78 (6): 1863–1903.

2. "Productivity Differences between and within Countries" (with Daron Acemoglu). 2010. *American Economic Journal: Macroeconomics* 2 (1): 169–88.

3. "Path Dependence in Development: Evidence from the Mexican Revolution." MIT Working Paper.

4. "The Historical State, Local Collective Action, and Economic Development in Vietnam" (with Nathan Lane and Pablo Querubín). 2018. *Econometrica* 86 (6): 2083–2121.

5. "The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java" (with Benjamin Olken). 2020. *Review of Economic Studies* 87 (1): 164–203.

6. "Trafficking Networks and the Mexican Drug War." 2015. *American Economic Review* 105 (6): 1738–79.

7. "The Violent Consequences of Trade-Induced Worker Displacement in Mexico" (with Benjamin Feigenberg and Kensuke Teshima). 2019. *American Economic Review: Insights* 1 (1): 43–58.

8. "Nation Building Through Foreign Intervention: Evidence from Discontinuities in Military Strategies" (with Pablo Querubín). 2018. *Quarterly Journal of Economics* 133 (2): 701–64.

9. "Temperature and Income: Reconciling New Cross-Sectional and Panel Estimates" (with Benjamin F. Jones and Benjamin A. Olken). 2009. *American Economic Review* 99 (2): 198–204.

10. "Temperature Shocks and Economic Growth: Evidence from the Last Half Century" (with Benjamin F. Jones and Benjamin A. Olken). 2012. *American Economic Journal: Macroeconomics* 4 (3): 66–95.

11. "What Do We Learn from the Weather? The New Climate-Economy Literature." (with Benjamin F. Jones and Benjamin A. Olken). 2014. *Journal of Economic Literature* 52 (3): 740–98.

12. "Information Extraction from Text Regions with Complex Tabular Structure" (with Kaixuan Zhang, Zejiang Shen, and Jie Zhou). 2019. In *33rd Conference on Neural Information Processing Systems*. Vancouver, Canada.

13. "OLALA: Object-Level Active Learning Based Layout Annotation." (with Zejiang Shen, Jian Zhao, Yaoliang Yu, and Weining Li). 2020. Working Paper.

cases triggered—by Marx's writings. This not only included Marxist economists,[2] but historians such as Christopher Hill, Perry Anderson, and Robert Brenner, and non-Marxist early institutionalists including Thorstein Veblen, Joseph Schumpeter, Karl Polanyi, and Barrington Moore, who revolutionized our understanding of the evolution of early modern institutions.[3]

This early work did not offer a systematic study of how differences in institutions determine the economic and political trajectories of nations. This was mostly because of an absence of theory of institutional dynamics—that is, how different historical, international, geographic and political characteristics of countries created

---

[2] For example, Baran and Sweezy (1968) and Dobb (1975).
[3] See Hill (1961), Anderson (1974), Brenner (1976), Veblen (1899), Schumpeter (1942), Polanyi (1944), and Moore (1966).

tendencies for different types of institutions to evolve.[4] There was no conceptual framework for studying whether and how institutions are "efficient" —meaning whether they enable the organization of the productive capacity of the economy in a way that helps generate a high level of output, or whether they generate systematic "inefficiencies" and suboptimal organizations of the economy. These works also lacked a clear description of the set of channels through which history matters and systematic empirical work for evaluating various institutional channels.

A group of pioneering scholars more familiar with economic mechanisms and economic data made the next major contributions in this area. Mancur Olson, Douglass North, Oliver Williamson, and Barry Weingast, for example, begin clarifying questions related to the efficiency of institutions, emphasizing the inseparability of the distribution of resources in society from the efficiency question—put simply, we cannot presume that institutions maximize the size of a pie, which will then be divided in whatever way we feel appropriate; the size of the pie is linked to its slicing. Elinor Ostrom and Williamson initiated the study of how important market and non-market institutions work. North, Joel Mokyr, Stan Engerman, and Ken Sokoloff started building comprehensive accounts of how certain aspects of institutions persist, for example in Engerman and Sokoloff's famous work on New World colonies, because of how inequality breeds inequality and via this channel, shapes the economic and political development trajectory of a country.[5]

The next important milestone was the realization by a number of economists, led by Allan Meltzer and Scott Richard (1981), that political economy ideas did not just belong to the normative or theoretical realm but could also be useful for understanding major social and economic changes—in Meltzer and Richard's case, the increase in the size of government that arose from the expanding voting franchise. Many modern economists took off from this insight and applied political-economic reasoning to shed light on questions related to investment, growth, and inequality, for example, by studying how and why inequality may retard economic growth and how property rights are determined that impact investment and growth.[6]

My own work with Simon Johnson and James Robinson built on this literature, attempting to develop a more systematic framework for the study of the fundamental questions of long-run economic development. James and I attempted to develop a theory of institutional change, with special emphasis on understanding the comparative statics of institutions: why does democracy emerge in some places

---

[4]I certainly do not mean to claim that early scholars in this tradition did not understand the importance of comparative statics. For instance, Barrington Moore's (1966) ambition in his seminal *Social Origins of Dictatorship and Democracy* was a type of comparative static—to uncover the historical and social origins of communism, fascism, and market capitalism. But these attempts did not crystallize into a coherent framework. In particular, Moore's explanation could not easily be generalized to apply to Africa, the Americas, or Asia, or to different eras.

[5]For relevant examples by these authors, see North (1981), Olson (1982), Williamson (1985), North and Weingast (1989), Mokyr (1990), Ostrom (1990), and Engerman and Sokoloff (2011).

[6]See Alesina and Rodrik (1994), Grossman (1994), Perotti (1994), Persson and Tabellini (1994), Besley and Coate (1997), and Benabou (2000).

and not others, and why are certain institutions more growth-enhancing in countries with certain histories (for example, see Acemoglu and Robinson 2000, 2006, 2012)? On the empirical front, Simon, James, and I used some of these ideas in order to probe the causes of economic disparities across countries. We tried to build our empirical work on core ideas from theory—for example, to describe the conditions in which dominant groups, such as European colonizers in Asia or the New World, accept or allow property rights for a broad cross-section of society versus when they will try to monopolize resources and take control of labor. In this way, we attempted to move towards establishing a causal role of institutions in cross-country disparities and inequalities. We also strove to document the role of the colonial period in the economic and political development of many societies around the world (Acemoglu, Johnson, and Robinson 2001, 2002).

Melissa was attracted to economics by these fundamental questions of long-run economic development and was aware of this intellectual background from early on. James Robinson recounts how he met Melissa while teaching at Harvard. At the time, Melissa was taking his undergraduate course on the political economy of Africa and Jeffrey Williamson's graduate economic history class. She had come across our then-recent papers, "Colonial Origins" (2001) and "Reversal of Fortune" (2002). She walked into James's office asking whether he was the same James Robinson who had co-authored these papers and proceeded to discuss not just the economic issues, but the comparative historical questions that these papers were raising.

I, too, remember my first conversation with Melissa, when she was a prospective graduate student trying to decide between several institutions vying to attract her. It was clear to me that I was in the presence of somebody with great drive, ability, and insight, not least because we started talking about how parts of various parts of Bolivia, despite their initial similarities, developed very differently. She was already aware that this might have something to do with the long-run effects of coercive labor market institutions—a perspective that would become the basis of one of her most important works.

Colonial History, Institutions, and Modern Development

When Melissa came on the research scene in the mid-2000s, she already understood that the economic literature on the fundamental questions was too "macro," both theoretically and empirically. In particular, it focused on country-level variation. Not only was much of the relevant inequality and variation within countries, but the ways in which different institutions worked—the specific pathways of persistence and their political and social foundations—varied greatly at the subnational and micro level. No area illustrates the nature of Melissa's contributions better than her work on the impact of colonial institutions and history on modern economic and political outcomes.

As I mentioned above, work by several scholars in the 1990s and early 2000s had started providing evidence and new ways of thinking about how European

colonial policies had shaped the divergent fortunes of former colonies. But there are obvious problems in comparing the development trajectories of the United States and Mexico from the 17th century onwards, because the two countries differ in so many ways.

Melissa, who is a fluent Spanish speaker and erudite concerning Latin American history, had started thinking about divergences within Latin America from the moment she walked into James Robinson's office as an undergraduate. Her deep dive into Latin American colonial history ultimately led to what is arguably her most well-known and important paper, "The Persistent Effects of Peru's Mining *Mita*" [1].

The Spanish Viceroyalty of Peru established in 1573 relied on a system of coerced labor, the mining *mita*. The *mita* was a compulsory labor draft on villages located near mines, in particular the Potosí silver mines in what is today Bolivia, from which one-seventh of the adult male population would be forced to go and work in the mines for extended periods. Although the mining *mita* was not as extensive and brutal as the chattel slavery that came to dominate parts of Brazil, Colombia, the Caribbean, and southern United States, it was nonetheless an onerous burden, bolstered by a suite of repressive institutions. The labor draft was abolished at the beginning of the 19th century. The key question is whether this particular colonial institution has had persistent effects during the 200 years since then.

Poring over archival and historical material on the colonial mining system, Melissa learned that there was a specific catchment area for the mines, and villages outside of the catchment area were not subject to the labor draft. She was able to locate the exact boundaries of the catchment area. This allowed a geographic regression discontinuity design, whereby one could compare adjacent villages on either side of the boundaries of the catchment area.

There are at least four challenges confronting this particular strategy for identifying causality, and how Melissa dealt with each one of them shows the creativity and maturity of her thinking. First, there is a danger that one could throw the baby out with the bathwater in undertaking such a micro approach. The mining *mita* was coupled with other institutions in the Viceroyalty of Peru that enabled the control of labor by force and repression, and these institutions persist in some form throughout Bolivia today. If so, by comparing two neighboring places within this area, one might filter out the true effects of coercive labor institutions. In response to this challenge, Melissa lays out a plausible case for why many of the effects of labor coercion will be at a more local level. As she documents, many of the affected municipalities were isolated then and now, and the Bolivian state is weak and has limited reach. All the same, as Melissa recognizes, one might wish to interpret estimates of the effects of the *mita* system from such a strategy as lower bounds, because some of the effects of the institutions undergirding labor coercion will also affect the control municipalities.

Second, the area where mines are located may not be random, and thus the catchment area for mining labor may not be random, either. Indeed, part of the catchment area boundary follows the edge of the Altiplano, a high plateau area located in central South America where altitude and several other geographic characteristics change sharply. From an econometric point of view, this raises a concern

that villages inside and outside the catchment area may vary in other systematic ways. Melissa recognized this problem and determined that the northern part of the Altiplano area did not suffer from this geographic discontinuity. The northern boundary was also attractive for her purposes because its location was fixed in the 16th century according to estimates of how long it would take laborers to walk from their village to Potosí (a capital city for the area during the colonial period). Her empirical work then focused on this northern boundary.

Third, one would need to understand the organization of the local economy to determine how significant differences could persist for over 200 years between neighboring villages. What mechanisms might separate the trajectories of these nearby villages? This, of course, is both a challenge and an opportunity, because it may be easier to make progress on understanding the mechanisms of persistence within a specific country, rather than in cross-country work.

Finally, in contrast to the most common applications of regression discontinuity at the time, where there is a single running variable, spatial regression discontinuity design has to grapple with the fact that distances from the relevant boundaries are two-dimensional. Melissa carefully and transparently dealt with this issue in a way that has enabled many other researchers since then to build on her work.

Melissa's regression discontinuity estimates using the northern border of the mining *mita* reveal sharp differences between neighboring municipalities on either side of the catchment boundary. For example, places inside of the catchment area have household-equivalent consumption levels that are somewhere between 20 and 30 percent lower today than neighboring municipalities outside of the catchment area. They also have significantly higher rates of child stunting (low height-for-age among children). This is an eye-popping finding. Imagine that a block of streets in your town has 30 percent lower income per capita than a neighboring block because they had a different history 200 years ago! Labor coercion must have been hugely scarring to generate such durable consequences.

But the issue is even more complex and more interesting than the headline finding might suggest because these municipalities are not like two city blocks in a modern American metropolis. Understanding the difference in their social, political, and economic terrain is the second major contribution of Melissa's paper.

Melissa documents that the economies of neighboring villages remain very distinct, in part because these places are isolated and poorly connected. In fact, one of the proximate causes of poverty in the catchment area is likely to be the low prevalence and quality of basic public services, such as roads, water, sewage, and education. But why might labor coercion 200 years ago cause lower public good provision today? To answer this question, Melissa delves into the history of Spanish colonial politics and economics.

Paradoxically, it turns out that the Spanish colonial state wanted to "protect" the people in the catchment area so that they could be a reliable source of labor for the mines. This meant, in particular, shielding them from the exploitation of local Spanish settlers, who would set up their own haciendas and recruit and sometimes coerce the indigenous population to work for them for low wages. The most

straightforward way for the Spanish colonial state to do this was to prevent settlers from building haciendas inside the catchment area.

Melissa documents that haciendas are common outside of the *mita* and much less common inside. The hacienda owners became increasingly powerful in the 19th century. They became the ones who controlled state investments in roads, local amenities, and schools. They directed such investments to the areas where their haciendas were, because they would need to transport their produce and also would benefit from these other services directly or indirectly via their workers. Consequently, municipalities inside the catchment area were often ignored and deprived of critical public services. Because they could not reach the market, subsistence farming and economic self-sufficiency became more prevalent in these areas.[7]

Since its publication in 2010, this paper [1] has become very well-known and features in many graduate and some undergraduate reading lists. It combines painstaking data collection and high-quality econometric analysis to create a nuanced picture of colonial institutions and their persistent effects. This combination emerges again and again in Melissa's work.

Many of the qualities that make [1] so important for the literature are seen in a second major paper Melissa wrote on Latin American colonial history, "Path Dependence in Development: Evidence from the Mexican Revolution" [3]. This paper investigates the long-term consequences of the armed conflict in the context of the Mexican Revolution, which raged across Mexico in the 1910s following the overthrow of the longtime Mexican dictator Porfirio Diaz. The paper is interesting for its discussion of the unique and rich context of early 20th century Mexico. But it also raises the big question of institutional persistence. All countries in South America became independent in the first half of the 19th century and many of them have experienced various social revolutions. Why didn't these upheavals erase the effects of the colonial institutions of the past?[8]

The Mexican Revolution is an ideal case for studying these questions because it destroyed many parts of the existing system, ushered in a new era of politics in Mexico and a new dominant party, and fundamentally changed the contours of Mexican politics. While the causes of rebellion in many parts of the country were variegated, much of it was caused by the desire of local groups to challenge the authority of the state. Eventually, the Mexican Revolution led to the victory of the "constitutionalists," whose plan was to centralize political power in Mexico.

Melissa's idea is that the centralizing powers would treat local areas that rose up against central power differently, and this would have long-run influences through politics in these areas. Melissa shows that severe droughts between 1906 and 1910 were a major predictor of greater revolutionary activity and armed conflict, and goes

---

[7]I should also mention a more minor paper that Melissa and I co-authored, "Productivity Differences Between and within Countries" [2], where we document the extent of within-country between-municipality inequality, and link it, as in [1], to public good provision and especially road infrastructure.

[8]In Acemoglu and Robinson (2008, 2012), we suggested that initial institutions influenced subsequent political and economic development through a process of "path-dependent change," but we could not nail the exact mechanisms and did not have systematic evidence on how this process worked.

through her usual careful work to document that this is a causal effect. Interestingly, these places experiencing greater revolutionary activity are significantly poorer today and also have access to fewer public goods and lower education. Why this persistence?

Melissa's investigation of Mexican history provides a convincing explanation. The constitutionalists' Institutional Revolutionary Party (the PRI) that dominated Mexican politics until the 2000s controlled these areas by setting up patron-client relations—for example, in the form of agricultural communes, the *ejidos*. Political reform and public goods were provided but always embedded within this system of distorted property rights and highly politicized resource allocation. This political channel then created a new type of persistence: long-run underdevelopment resulting from inefficient ways of appeasing the local community in order to ensure order within the context of a non-representative, largely extractive regime.

Put differently, Melissa proposed a new channel for persistence. The political disruption created by temporary shocks—droughts between 1906 and 1910—during this critical juncture generated persistent political and economic effects. Given the Mexican context and the nature of PRI's tenuous control over this vast country, this response took the form of concessions and resources for these areas but in a particularly inefficient form which then generated persistent economic costs.

Given Melissa's deep knowledge of Latin American history, one might have expected her to continue to focus on this area. Instead, she turned to colonial history in Asia. One fundamental difference here is that several Asian nations managed to embark upon a process of rapid economic growth over the last 70 years, unlike most of Latin America. The causes of this "second divergence" are still poorly understood. Motivated by these issues, Melissa started voraciously reading and learning about this history. Here, I will discuss two important papers in this area and omit her very promising work-in-progress on Japan and Taiwan.

In "State Capacity, Local Governance, and Economic Development in Vietnam" [4], jointly written with Nathan Lane and Pablo Querubín, Melissa not only shifts focus from Latin America to Asia but also turns to a new set of political economy mechanisms: the role of bureaucratic state capacity at the local level.

There is a growing literature on state capacity, and much of this work focuses on within-country variation. But various problems afflict this literature. It is often difficult to pinpoint either the mechanisms that led to variation in state capacity or even how state capacity works. Moreover, it is difficult to dispel concerns that state capacity co-varies with other institutional or cultural characteristics. The paper seeks to overcome these concerns by exploiting the different types of institutions that became deep-rooted within Vietnam because of the influence of two competing empires: the Dai Viet in the north and the Khmer Empire in the south. While the Dai Viet in the north had a centralized political structure, empowering village institutions as a continuation of the centralized state, the Khmer Empire in the south collected tribute from the areas it controlled via patron-client type relationships (a type of "indirect rule") that did not invest in or develop bureaucratic, institutionalized states at the local level.

Melissa, Nathan, and Pablo use this historical and geographic context to shed light on the role of state capacity and its historical development in Vietnam. They exploit the fact that there was a clear boundary between the two empires and obtain regression discontinuity estimates of the effects of these different types of political institutions on the long-term development of state capacity and contemporary outcomes.

As is typical of Melissa's careful historical work, the paper uses very rich data, mostly hand-collected over several years, to document persistent differences in living standards, favoring the north of the boundary with its history of better-developed state capacity. Melissa, Nathan, and Pablo document that the effects largely work through better provision of public goods. Once again, the ability of Melissa (and her co-authors) to develop a nuanced understanding of how institutions work on the ground is deeply impressive.

One important finding is that the strong centralized institutions of Dai Viet appear to have "crowded in" (rather than crowded out) stronger civil society and cooperative relations within villages, which are ultimately responsible for the provision and distribution of these public goods. This thesis is of particular interest because several scholars have observed or conjectured that what distinguishes Asian political economy (especially in places such as China, Vietnam, and Korea) from Africa and Latin America is the greater capacity of bureaucracies and state institutions that have evolved over millennia. But how does this capacity emerge in reality? Is it a hallmark of top-down imperial control as in the Chinese case, where efforts to build a capable bureaucracy under Imperial command go back more than two millennia? This paper is thought-provoking in this context because it suggests that there are multiple paths to high state capacity, and in fact state capacity can be strengthened by local political organization and cooperation (which is rather different than the most common interpretation of the Chinese path).

The other paper on persistent effects of Asian institutions is Melissa's recent work with Ben Olken, "The Developmental Effects of Extractive Colonial Economy: The Dutch Cultivation System in Java" [5]. This paper is also based on meticulous data collection from Dutch archives that Melissa located. It uses a regression discontinuity design as well as a creative placebo strategy, each based on historical differences in the reach of the Dutch colonial apparatus.

There are some parallels between this paper and Melissa's work on the mining *mita* [1]. There is once more a catchment area, subject to colonial coercive activities, this time related to sugar cultivation. Again, production is organized under coercion with the chief purpose of enriching European colonists who kept control via their military superiority. But institutions work quite differently depending on historical context. In this instance, Melissa and Ben find that villages in the catchment area of the colonial sugar production system in Dutch-colonized Java are no poorer than the control groups, and in fact, have higher levels of education and more manufacturing activity. Why would this be?

The mining *mita* and sugar cultivation in Java have some important differences. For one, sugar workers in Java were not moved to the mine, but cultivation instead

took place in their local community—which meant investments in roads and amenities. Even more importantly, sugar cultivation, though taking place under harshly extractive conditions, also came with sugar mills to process the sugar. Melissa and Ben show that these mills, as well as the infrastructure, stayed after the forced cultivation regime disappeared. They further document via a placebo strategy—by comparing locations where the mills were placed to plausible counterfactual locations along the same rivers—that the mills led to more economic activity, and this effect persisted for more than 100 years after the closure of the mills. It therefore appears that the economic structure implemented by the Dutch, centered on sugarcane production and processing, created a platform for continued economic activity after the colonial period, and was supplemented by schools and other public good investments. The economic opportunities generated by these investments had persistent effects.

This is another very well-crafted, careful paper, based on a massive amount of historical data collection. Yet again, what is most impressive is the detailed institutional context and the deep dive into how these institutions work on the ground, depending on historical conditions.

Conflict, Law Enforcement, and Politics

Melissa has also made defining contributions concerning the interplay of conflict, law enforcement, and politics. Of course, there is a large literature on crime and law enforcement both in developed and developing economies. But previous to Melissa's work, their relationship to politics had not been central. For example, might political factors determine when law enforcement is effective and when it backfires?

In "Trafficking Networks and the Mexican Drug War" [6], Melissa asks how violent conflict affects economic opportunities and outcomes. She recognizes that this question cannot be answered without understanding the institutional context in which violent conflict is taking place and the policies that central authorities are using to contain it.

The context on which Melissa focuses is Mexico's drug wars and the policies that the National Action Party (PAN) took against drug cartels in the second half of the 2000s. Felipe Calderon from PAN ran for president in 2006 on an agenda to fight the drug cartels and emphasized how these cartels had started controlling local politics. Calderon's strategy involved using federal police against drug cartels in municipalities where they had taken root. Implementing these policies needed at least some support of local politicians, however, so whether local mayors were aligned with PAN was critical in the war against drug cartels.

Melissa first shows that local PAN mayors have a causal effect on the fight against drug cartels. To do this, she has to resolve the usual omitted variable bias and selection problems associated with certain areas having different political preferences and ideologies (or desperately turning to a party promising a cleanup in their worst

hour). Melissa addresses this problem using a different type of regression discontinuity design; in this case, comparing municipalities that narrowly elected mayors from PAN to those that narrowly elected mayors from a different party (typically PRI). She finds that places that narrowly elected PAN mayors, relative to those that narrowly did not, experienced greater anti-drug cartel action, which resulted in a significant increase in violence. While one might expect this to have been due to police versus drug cartel confrontations, it turns out that the rise in violence is mostly driven by fighting between drug gangs.

The federal action appears to have been successful in weakening the dominant cartels and sometimes removing their kingpins. This and other evidence suggest that the harder line against drug cartels destabilized the market and triggered a turf war between rival gangs for the control of the municipality and the local routes. Thus, in the presence of weak political institutions, fighting drug cartels can create unintended consequences—especially so long as the drug trade remains lucrative.

Melissa also wanted to understand how these anti-drug actions have reshaped the drug network in Mexico and the implications of these changes for the lives of ordinary Mexicans. Originally, certain municipalities were well-located to serve the large US market for drugs, but once the turf wars erupted, these municipalities ceased to be as attractive for the established drug cartels. Based on this observation, Melissa maps the road network taking drugs from their origins to US entry points. She finds a shift of drug cartel activity from the municipalities affected by turf wars towards those that form alternative feasible paths to destinations. This analysis is, to the best of my knowledge, the first carefully established instance of spillover of criminal activity through spatial networks.

The paper then uses this variation to tackle the initial economic question: the effects of drug cartel activity on economic opportunities and outcomes for residents. Melissa uses the source of variation generated by Calderon's drug war to answer this question. Rather than improving the lot of ordinary Mexicans, Melissa's evidence shows that the war against drug cartels had significant negative economic consequences for a municipality's residents, most likely because of the collateral damage and the instability that they suffered during the turf wars.

In many ways, this is a very different research agenda than investigating the historical persistence of colonial institutions. But it shares the keen-eyed focus on how the details of an institutional setting matter for the effects of policies. In a different historical context and institutional environment, the federal action against drug cartels could have been very effective and it may have even improved the economic opportunities for residents. Yet the situation in Mexico, with weak local institutions, limited federal control, and continuing lucrative opportunities from the US demand for drugs, made the outcome very different.

A more recent paper, "The Violent Consequences of Trade-Induced Worker Displacement in Mexico" [7], cowritten with Benjamin Feinberg and Kensuke Teshima, is also related. This paper turns to the converse problem: what are the impacts of economic shocks (in this instance trade shocks coming from manufacturing job losses due to Chinese imports) on drug trafficking and violence. When

negative trade shocks reduce local labor market opportunities, more low-skill residents (those that are especially hard-hit by manufacturing job loss) appear to turn to illicit jobs in the drug trade or gangs. This increases criminal activity and also leads to greater violence, once again likely related to turf wars—this time because economic (trade) shocks have disrupted the balance between different cartels.

An even more ambitious and original paper within this broad agenda is "Nation Building through Foreign Intervention: Evidence from Discontinuity in Military Strategies" [8], jointly written with Pablo Querubín. The question is whether foreign intervention, in this instance US reactions during the Vietnam War, is effective in holding insurgencies at bay. The narrower question is whether bombing and military tactics focused on the use of robust force against insurgents "work." This issue is often complicated by the fact that such strategies affect civilians among whom insurgents hide and live, and the damage that the civilians suffer in the process might motivate them to support the insurgents. This question is relevant not just to the broad area of law enforcement and politics but also for military strategy and the study of anti-insurgent activities in areas with weak state presence.

Melissa and Pablo tackle this question with a clever regression discontinuity design. The US Department of Defense collected detailed information in every Vietnamese hamlet using about 170 monthly and quarterly questions on security, politics, and economics. These questions were aggregated into an overall security score, which then became the basis of the Air Force's decisions on which parts of Vietnam to target with airstrikes. Because of limited computing resources at the time, these security scores were rounded to the nearest integer (in particular, printing continuous security scores from memory was deemed too costly and time-consuming). The source data were not discarded, and Melissa and Pablo were able to locate and recover these continuous security scores.

The paper then uses the rounding thresholds as a discontinuity cutpoint in the likelihood of being bombed. Villages that ended up just below the rounding threshold were deemed to have higher insurgency risk and experienced a higher likelihood of bombing than otherwise comparable villages just above the rounding threshold. The striking result is that US Air Force bombing campaigns, rather than discouraging them, helped the Viet Cong insurgents. The residents of bombed villages, compared to similar villages with essentially the same risk and sympathies for insurgents, became more likely to cooperate with the Viet Cong after being bombed.

The paper also compares the different strategies of the US Army and Marines, where the latter pursued an early form of the "winning hearts and minds" campaign. The identification here is based on a very different approach, this time rooted in the effects of different military traditions. The Marines were much less heavy-handed than the Army given their traditions and prior experiences, so they approached the problem of fighting the Viet Cong very differently. Using this different strategy, Melissa and Pablo estimate similar effects, confirming that the heavy-handed bombing tactics backfired in containing the insurgency.

The potential lessons from this paper are obvious beyond the context of the Vietnam War. It is another thought-provoking paper, which has already affected the literature and will continue to trigger more innovative work in this new area that Melissa is now carving out.

Other Work

Melissa has other important lines of work, which I briefly discuss now. Two papers Melissa co-authored with Ben Jones and Ben Olken, [9, 10], study the effects of climate on long-run economic development. The question is whether and how climate (and by implication, climate change) matters for economic outcomes.

In "Temperature Shocks and Economic Growth: Evidence from the Last Half Century" [9], the authors exploit within-country climate and temperature fluctuations to look at the effects of temperature on the level of income and economic growth. The paper finds statistically significant and substantial negative effects. Given the nature of the aggregate cross-country data, the authors are limited in their ability to understand the root causes of this surprising headline effect. Nevertheless, they show that temperature affects a variety of industries and may even impact political instability. The second paper, [10], is an offshoot that further explores the same questions.

Melissa has also contributed to the practice of high-quality empirical work. In "What Do We Learn from the Weather? The New Economy-Climate Literature" [11], Melissa, Ben, and Ben provide a lucid exposition of best practices in studies using microdata on weather variation across space. The paper shows how such variation can be useful for answering a variety of questions but also points out various pitfalls. For example, temperature and rainfall data are often interpolated or extrapolated from satellite data or a few weather stations, which can create spatial correlation and the possibility of spurious inference.

In [12, 13], Melissa and her co-authors develop new computational image processing methods for extracting information from non-standard (and foreign-language) text data, which have become increasingly common with expanding digital access to archival records.

Conclusion

Melissa has set a new standard in applied research because she has shown by example that one does not need to shy away from the fundamental questions of long-run economic development and political economy—which are also some of the fundamental questions of the social sciences as a whole—to do credible and sometimes literature-changing empirical research. The demands of the academic research and publication environment for credible identification strategies need

not limit researchers to narrow contexts and questions. Melissa's path-breaking approach has shown how to use painstaking and meticulous data collection and sophisticated, well-executed empirical approaches so that economists and other social scientists can understand the way institutions work. Her work is rooted in the history, context, and details of specific institutions. In the course of her broad research agenda, she has forged several distinctive lines of research.

As this essay has expounded, Melissa's insight, energy, and creativity are unrivaled. But equally impressive for all of us who have known her is her stupendous work ethic and commitment to academic excellence. I remember vividly how whenever we were on the same flight on the way to a conference, I would see her engrossed in the pages of a history book even as she was lining up on the jet bridge. This is the work ethic that has made her not only an amazing scholar but also a top long-distance runner (appropriate training for somebody who would doggedly work to change her field, a task which is not unlike a marathon).

We will all continue to admire Melissa's work, but it is the next generation of young scholars in economic development and political economy who will be the real beneficiaries of her track record. There is much for them to learn from Melissa's work ethic, sophisticated approach to history and empirical work, and ambition and courage to tackle big questions. Indeed, Melissa has helped make the fundamental questions of long-run economic growth cool for ambitious applied economists.

References

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369–1401.

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117 (4): 1231–94.

Acemoglu, Daron, and James A. Robinson. 2000. "Why Did the West Extend the Franchise? Democracy, Inequality, and Growth in Historical Perspective." *Quarterly Journal of Economics* 115 (4): 1167–99.

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy.* Cambridge, UK: Cambridge University Press.

Acemoglu, Daron, and James A. Robinson. 2008. "Persistence of Power, Elites, and Institutions." *American Economic Review* 98 (1): 267–93.

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty.* New York: Crown.

Alesina, Alberto, and Dani Rodrik. 1994. "Distributive Politics and Economic Growth." *Quarterly Journal of Economics* 109 (2): 465–90.

Anderson, Perry. 1974. *Lineages of the Absolutist State.* London: Verso, 2013.

Baran, Paul A., and Paul M. Sweezy. 1968. *Monopoly Capital: An Essay on the American Econom. and Social Order.* London: Penguin Books.

Benabou, Roland. 2000. "Unequal Societies: Income Distribution and the Social Contract." *American Economic Review* 90 (1): 96–129.

Besley, Timothy, and Steven Coate. 1997. "An Economic Model of Representative Democracy." *Quarterly Journal of Economics* 112 (1): 85–114.

Brenner, Robert. 1976. "Agrarian Class Structure and Economic Development in Preindustrial Europe." *Past and Present* 70: 30–75.

Dobb, Maurice. 1975. *Theories of Value and Distribution since Adam Smith: Ideology and Economic Theory.* Cambridge, UK: Cambridge University Press.

Engerman, Stanley L., and Kenneth L. Sokoloff. 2011. *Economic Development in the Americas since 1500: Endowments and Institutions.* Cambridge, UK: Cambridge University Press.

Grossman, Herschel I. 1994. "Production, Appropriation and Land Reform." *American Economic Review* 84 (3): 705–12.

Hill, Christopher. 1961. *The Century of Revolution 1603–1714.* New York: W.W. Norton & Co.

Marx, Karl. 1847. *The Poverty of Philosophy.* Chicago: Charles H. Kerr & Company, 1920.

Marx, Karl. 1852. *The 18th Brumaire of Louis Bonaparte.* Rockville, MD: Wildside Press LLC, 2008.

Marx, Karl. 1867. *Capital*, Vol. 1. London: Penguin, 2004.

Meltzer, Allan H., and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy* 89 (5): 914–27.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress.* Oxford: Oxford University Press.

Moore, Barrington, Jr. 1966. *The Social Origins of Dictatorship and Democracy* Boston: Beacon Press.

North, Douglass C. 1981. *Structure and Change in Economic History.* New York: W.W. Norton & Co. New York.

North, Douglass C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History* 49 (4): 803–32.

Olson, Mancur. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities.* New Haven, CT: Yale University Press.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK: Cambridge University Press.

Perotti, Roberto. 1994. "Income Distribution and Investments." *European Economic Review* 38 (3–4): 827–35.

Persson, Torsten, and Guido Tabellini. 1994. "Is Inequality Harmful for Growth?" *American Economic Review* 84 (3): 600–21.

Polanyi, Karl. 1944. *The Great Transformation.* Boston: Beacon Press.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy.* London: Routledge. London.

Veblen, Thorstein. 1899. *The Theory of the Leisure Class.* Boston: Houghton Mifflin, 1973.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting.* New York: Free Press.

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2020—commonly known as the Nobel Prize in economics—was awarded to Paul R. Milgrom and Robert B. Wilson "for improvements to auction theory and inventions of new auction formats." The Nobel prize committee also published "Popular science background: The quest for the perfect auction" (October 2020, https://www.nobelprize.org/prizes/economic-sciences/2020/popular-information/) and "Scientific Background: Improvements to auction theory and inventions of new auction formats" (October 2020, https://www.nobelprize.org/prizes/economic-sciences/2020/advanced-information/). "Every day, auctions distribute astronomical values between buyers and sellers. This year's Laureates, Paul Milgrom

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

and Robert Wilson, have improved auction theory and invented new auction formats, benefitting sellers, buyers and taxpayers around the world. . . . Robert Wilson was the first to create a framework for the analysis of auctions with common values, and to describe how bidders behave in such circumstances. In three classic papers from the 1960s and 1970s, he described the optimal bidding strategy for a first-price auction when the true value is uncertain. Participants will bid lower than their best estimate of the value, to avoid making a bad deal and thus be afflicted by the winner's curse. His analysis also shows that with greater uncertainty, bidders will be more cautious and the final price will be lower. Finally, Wilson shows that the problems caused by the winner's curse are even greater when some bidders have better information than others. . . . In most auctions, the bidders have both private and common values. . . . An energy company that bids on the right to extract natural gas is concerned with both the size of the gas reservoir (a common value) and the cost of extracting the gas (a private value, as the cost depends on the technology available to the company). . . . Analysing bids in auctions with private and common values turned out to be an even trickier problem . . . The person who finally cracked this nut was Paul Milgrom, in a handful of papers published around 1980."

Claudia Goldin delivered the 2020 Martin Feldstein Lecture on the topic "Journey across a Century of Women" (NBER Reporter, October 2020, https://www.nber.org/reporter/2020number3/journey-across-century-women). "Five distinct groups of women can be discerned across the past 120 years, according to their changing aspirations and achievements. Group One graduated from college between 1900 and 1919 and achieved 'Career or Family.' Group Two was a transition generation between Group One, which had few children, and Group Three, which had many. It achieved 'Job then Family.' Group Three, the subject of Betty Friedan's *The Feminine Mystique*, graduated from college between 1946 and 1965 and achieved 'Family then Job.' Group Four, my generation, graduated between 1966 and 1979 and attempted 'Career then Family.' Group Five continues to today and desires 'Career and Family.' College-graduate women in Group One aspired to 'Family or Career.' Few managed both. In fact, they split into two groups: 50 percent never bore a child, 32 percent never married. . . . More Group Two college women aspired to careers, but the Great Depression intervened, and this transitional generation got a job then family instead. As America was swept away in a tide of early marriages and a subsequent baby boom, Group Three college women shifted to planning for a family then a job. Just 9 percent of the group never married, and 18 percent never bore a child. Even though their labor force participation rates were low when they were young, they rose greatly—to 73 percent—when they and their children were older. But by the time these women entered the workplace, it was too late for them to develop their jobs into full-fledged careers. . . . Group Four . . . aided by the Pill, delayed marriage and children to obtain more education and a promising professional trajectory. Consequently, the group had high employment rates when young. But the delay in having children led 27 percent to never have children. Now, for Group Five the goal is career and family, and although they are delaying marriage and childbirth even more than Group Four, just 21 percent don't have children."

The Credit Suisse Research Institute has published "Collectibles: An integral part of wealth" (October 2020, https://www.credit-suisse.com/media/assets/corporate/docs/about-us/research/publications/csri-collectibles-2020.pdf). From the opening chapter by Nannette Hechler-Fayd'herbe and Adriano Picinati di Torcello: "[W]e estimate that people with net worth exceeding USD 30 million accounted for USD 26.3 trillion of global wealth prior to the outbreak of the COVID-19 pandemic. . . . Conservatively estimated, an approximate share of 3–6 percent in collectibles would bring the value of collectibles owned by private individuals to around USD 1.2 trillion. . . . [F]or illustrative purposes, we compare the historical evolution of the Sotheby's Mei Moses index for fine arts with the Liv-ex Fine Wine 100 Index, the HAGI Top 100 Index for classic cars, the AMR indices for watches and jewelry, and a luxury handbag index . . . Over the last ten years, most collectible categories have gained in value, but with substantial differences from one to the other. On aggregate, wines and fine art have returned the least. Watches and jewelry have been effective stores of value, with cumulative 10-year returns between 27 and 61 percent. Classic cars were by far the best-performing collectibles category. . . . Naturally, this trend is time- and index-dependent, and other periods will show different developments."

Andrea Ciani, Marie Caitriona Hyland, Nona Karalashvili, Jennifer L. Keller, Alexandros Ragoussis, and Trang Thu Tran have written *Making It Big: Why Developing Countries Need More Large Firms* (September 2020, World Bank, https://openknowledge.worldbank.org/handle/10986/34430). "This report shows that large firms are different than other firms in low- and middle-income countries. They are significantly more likely to innovate, export, and offer training and are more likely to adopt international standards of quality. Their particularities are closely associated with productivity advantages—that is, their ability to lower the costs of production through economies of scale and scope but also to invest in quality and reach demand. Across low- and middle-income countries with available business census data, nearly 6 out of 10 large enterprises are also the most productive in their country and sector. These distinct features of large firms translate into improved outcomes not only for their owners but also for their workers and for smaller enterprises in their value chains. Workers in large firms report, on average, 22 percent higher hourly wages in household and labor surveys from 32 low- and middle-income countries—a premium that rises considerably in lower-income contexts. . . . Besides higher wages—which are strongly associated with higher productivity—large firms more frequently offer formal jobs, secure jobs, and nonpecuniary benefits such as health insurance that are fundamental for welfare in low- and middle-income countries. . . . Smaller and lower-income markets tend to host smaller firms. But even in relative terms, there are too few larger firms in these countries relative to the size of the economy and the number of smaller firms—there is a 'missing top.'"

Florian Scheuer investigates "Taxing the superrich: Challenges of a fair tax system" (UBS Center Public Paper #9, November 2020, https://www.ubscenter.uzh.ch/en/publications/public_papers/taxing-the-superrich.html#challenges-of-a-fair-tax-system). "Most countries' tax systems treat capital gains favorably relative to

ordinary labor income (Switzerland being an extreme case where most capital gains are untaxed). Realized capital gains represent a very high fraction of the reported income of the superrich. For example, realized capital gains represented 60% of total gross income for the 400 highest-income Americans in tax year 2014. . . . [F]or tax year 2016, those earning more than $10 million report net capital gains corresponding to 46% of their total income, whereas capital gains are a negligible fraction of income for those earning less than $200k. . . . Five OECD countries levy no tax on shareholders based on capital gains (Switzerland being a prominent example). Of those that do, all tax is on realization rather than on accrual. Five more countries apply no tax after the end of a holding period test, while four others apply a more favorable rate afterwards. The tax rate varies widely, with the highest as of 2016 being Finland, at 34%. With a few exceptions, the accrued gains on assets in a decedent's estate escape income taxation entirely, because the heir can treat the basis for tax purposes as the value upon inheritance." The paper offers a complement to the essay on wealth taxation by Scheuer with Joel Slemrod in this issue.

Siro Aramonte suggests "Mind the buybacks, beware of the leverage" (*BIS Quarterly Review*, September 2020, pp. 49–59, https://www.bis.org/publ/qtrpdf/r_qt2009d.htm). "Corporate stock buybacks have roughly tripled in the last decade, often to attain desired leverage, or debt as a share of assets. . . . In 2019, US firms repurchased own shares worth $800 billion . . . Net of equity issuance, the 2019 tally reached $600 billion. . . . In a number of cases, repurchases improve a firm's market value. For instance, if managers perceive equity as undervalued, they can credibly signal their assessment to investors through buybacks. In addition, using repurchases to disburse funds when capital gains are taxed less than dividends increases net distributions, all else equal. Furthermore, by substituting equity with debt, firms can lower funding costs when debt risk premia are relatively low, especially in the presence of search for yield. And, by reducing funds that managers can invest at their discretion, repurchases lessen the risk of wasteful expenditures. . . . [B]uybacks were not the main cause of the post-GFC [global financial crisis] rise in corporate debt. After 2000, internally generated funds became more important in financing buybacks. . . . There is, however, clear evidence that companies make extensive use of share repurchases to meet leverage targets. The initial phase of the pandemic fallout in March 2020 put the spotlight on leverage: irrespective of past buyback activity, firms with high leverage saw considerably lower returns than their low-leverage peers. Thus, investors and policymakers should be mindful of buybacks as a leverage management tool, but they should particularly beware of leverage, as it ultimately matters for economic activity and financial stability."

The IMF *World Economic Outlook* report includes a chapter on "Dissecting the Economic Effects" (October 2020, https://www.imf.org/en/Publications/WEO/Issues/2020/09/30/world-economic-outlook-october-2020). "This chapter's first goal is to shed light on the extent to which the economic contraction was driven by the adoption of government lockdowns instead of by people voluntarily reducing social interactions for fear of contracting or spreading the virus. . . . If lockdowns were largely responsible for the economic contraction, it would be reasonable

to expect a quick economic rebound when they are lifted. But if voluntary social distancing played a predominant role, then economic activity would likely remain subdued until health risks recede. . . . In fact, the analysis suggests that lockdowns and voluntary social distancing played a near comparable role in driving the economic recession. The contribution of voluntary distancing in reducing mobility was stronger in advanced economies, where people can work from home more easily and sustain periods of temporary unemployment because of personal savings and government benefits."

Symposia

Leslie Willcocks has written "Robo-Apocalypse cancelled? Reframing the automation and future of work debate" (*Journal of Information Technology*, 35:4, pp. 286–302, https://journals.sagepub.com/doi/full/10.1177/0268396220925830), which is followed by four comments. From the abstract: "Robotics and the automation of knowledge work, often referred to as AI (artificial intelligence), are presented in the media as likely to have massive impacts, for better or worse, on jobs skills, organizations and society. The article deconstructs the dominant hype-and-fear narrative. . . . The term AI has been hijacked, in order to suggest much more going on technologically than can be the case. The article reviews critically the research evidence so far, including the author's own, pointing to eight major qualifiers to the dominant discourse of major net job loss from a seamless, overwhelming AI wave sweeping fast through the major economies. The article questions many assumptions: that automation creates few jobs short or long term; that whole jobs can be automated; that the technology is perfectible; that organizations can seamlessly and quickly deploy AI; that humans are machines that can be replicated; and that it is politically, socially and economically feasible to apply these technologies. A major omission in all studies is factoring in dramatic increases in the amount of work to be done. Adding in ageing populations, productivity gaps and skills shortages predicted across many G20 countries, the danger might be too little, rather than too much labour. The article concludes that, if there is going to be a Robo-Apocalypse, this will be from a collective failure to adjust to skills change over the next 12 years."

The *Brookings Papers on Economic Activity* devotes its Fall 2020 issue to papers about "COVID-19 and the Economy." Papers, short readable overviews, presentation slides, and video are all available at https://www.brookings.edu/events/bpea-fall-2020-covid-19-and-the-economy/. Here's the agenda: "Business Credit Programs in the Pandemic Era," by Samuel G. Hanson, Jeremy C. Stein, Adi Sunderam, and Eric Zwick; "Fiscal Effects of COVID-19," by Alan J. Auerbach, William Gale, Byron Lutz, and Louise Sheiner; "Has the Paycheck Protection Program Succeeded?" by Glenn Hubbard and Michael R. Strain; "Epidemiological and Economic Effects of Lockdown," by Alexander D. Arnon, John A. Ricco, and Kent A. Smetters; "Macroeconomic Outcomes and COVID-19: A Progress Report," by Jesús Fernández-Villaverde and Charles I. Jones; "Will the Secular Decline in Exchange Rate and

Inflation Volatility Survive COVID-19?" by Ethan Ilzetzki, Carmen Reinhart, and Kenneth Rogoff; "Temporary Unemployment and Labor Market Dynamics during the COVID-19 Recession," by Jessica Gallant, Kory Kroft, Fabian Lange, and Matthew J. Notowidigdo; and "Sizing Up Corporate Restructuring in the COVID Crisis," by Robin Greenwood, Benjamin Iverson, and David Thesmar.

Interviews with Economists

Gordon Rausser and David Zilberman have "A Conversation with Angus Deaton" in the *Annual Review of Resource Economics* (2020, 12: pp. 1–22, https://www.annualreviews.org/doi/pdf/10.1146/annurev-resource-111219-042601). Here's Deaton on economic development: "In the back of Arthur's book on economic growth (Lewis 1955), he raises this question, which doesn't get asked enough: 'Why? Why do we care about this at all?' Because a lot of people don't. The Pope doesn't really seem to care about economic growth very much. I don't know whether Arthur actually talks about Mozart, but he talks about kids growing up in absolute poverty and how they never have the opportunity to develop what may be extraordinary innate skills. There are these buried talents—lost Mozarts, or lost Einsteins—a great term someone's been talking about recently. What development does is give people what Amartya Sen would call capabilities, which you just don't have if you're living in grinding poverty. The expressions of human genius and human creativity are going to be stifled and stamped out if you don't have economic development. That's why you should have development."

David A. Price interviews Joshua Gans "On managing pandemics, allocating vaccines, and low-cost prediction with AI" (*Econ Focus*: Federal Reserve Bank of Richmond, Second/Third Quarter 2020, pp. 18–22, https://www.richmondfed.org/publications/research/econ_focus/2020/q2-3/interview). "I now see these pandemics as manageable things. Policymakers have to react right away and stay the course, but pandemics can be managed. If I had to guess how history is going to judge this period, the judgment is going to be that this shouldn't have been a two- to three-year calamity, it should have been a three-month calamity. The need for testing aggressively at the beginning had to be appreciated. You aggressively isolate people you find who are infected, you trace who they had contact with, and you aim for quick, complete suppression. The countries that had experience with pandemics—Hong Kong, South Korea, Taiwan, most of Africa—got it right away. They knew what the problems would be if they didn't do anything about it. So experience with viruses was definitely a factor. . . . But once the virus breaks out, then you've got a problem. Then you've got to do the complete lockdown. And we're seeing places that did a complete lockdown—like they did in Italy, France, and Spain—squash it all the way down. Locking down is terribly painful; that's why you don't want to go through it in the first place. But you may have to. . . . Early in the crisis, people in the United States and Canada were not talking about the virus as something we needed to suppress completely. The discussion was mainly, 'We're

going to push down the curve, and then we'll wait for a vaccine.' But the evidence both historically and now with this virus is that, as I said, you can achieve suppression in months if you act quickly. You have to keep working at it because if you don't have a vaccine, the disease can crop up again, but it's manageable."

Douglas Clement offers "Seeing the margins: An interview with Columbia University economist Sandra Black," which addresses, as the subtitle says, "education, family wealth, her time at the White House, COVID-19, and the cost of bad policy" (*For All*, Opportunity & Inclusive Growth Institute at the Minneapolis Federal Reserve, Fall 2020, https://www.minneapolisfed.org/article/2020/seeing-the-margins-an-interview-with-columbia-university-economist-sandra-black).   "We decided to look at what's driving the correlations we see across generations in wealth using the Swedish data. . . . What's unique to the Swedish data is that we can observe adopted children and, importantly, we can observe both their biological and their adopted parents. So when you observe a child's wealth as an adult, you can see how correlated it is to their adopted parents' wealth and to their biological parents' wealth. . . . [W]e find that environment, or the adoptive parent, matters a lot, and more than the biological parent, unlike outcomes such as education or even income, which had more of a biological component. This is really important because it says people aren't wealthy because there's something inherently different about them that makes them better able to accumulate wealth; they're wealthy because they have these opportunities."

Discussion Starters

Matthew Higgins considers "China's Growth Outlook: Is High-Income Status in Reach?" (*Economic Policy Review*, Federal Reserve Bank of New York, October 2020, 26:4, pp. 69–98, https://www.newyorkfed.org/research/epr/2020/epr_2020_china-growth-outlook_higgins.html). From the abstract: "Our key finding is that China would need to sustain total factor productivity growth at the top end of the range achieved by its high-income Pacific Rim neighbors in order to match their success in raising living standards. While fast-growing working-age populations boosted per capita income growth elsewhere in the Pacific Rim, a rapidly aging population will act as a powerful drag on income growth in China's case. Moreover, China's already capital-intensive production structure will make it difficult to match those countries' gains from capital deepening. These restraints mean that a sustained and exceptionally high pace of productivity growth will be needed for Chinese per capita incomes to reach even 50 percent of the US level by 2040. We argue that lagging institutional development represents the chief obstacle to the needed productivity gains."

J. Mark Iwry, Claire Haldeman, William G. Gale, and David C. John discuss "Retirement Tontines: Using a Classical Finance Mechanism as an Alternative Source of Retirement Income" (Brookings Institution, October 2020, https://www.brookings.edu/research/retirement-tontines-using-a-classical-finance-mechanism-as-an-

alternative-source-of-retirement-income/). "Tontines are investment pools where members commit funds irrevocably and where the resources and income claims of members who die are given to members who survive. Tontines can be adapted to a wide variety of financial structures. They have financed everything from European wars to colonial-era capital projects to Americans' retirement. They were quite popular in the United States in the late 1800s and early 1900s until they were effectively (though not literally) outlawed in response to corrupt insurance company management. The 'tontine principle'—that surviving group members benefit financially from the death of other members—can evoke strong reactions, and has inspired murder plots in novels, movies, and even a *Simpsons* episode. But the mechanisms involved are not very different from how group annuities operate, and members of modern tontines would be mutually anonymous in any case. In recent years, analysts have revisited tontines as a theoretical tool, and several countries have created pension plans that incorporate tontine principles."

Alexander Monge-Naranjo and Qiuhan Sun ask "Will Tech Improvements for Trading Services Switch the U.S. into a Net Exporter?" (*Regional Economist*, Federal Reserve Bank of St. Louis, Fourth Quarter 2020, https://www.stlouisfed.org/publications/regional-economist/fourth-quarter-2020/tech). "The U.S. is a world leader in most high-skilled professional service sectors, such as health, finance and many sectors of research and development. Moreover, leading American producers have been ahead of others in the adoption of ICT in their production networks. The global diffusion of ICT—including possibly the expansion of 5G networks—is prone to make many of these services tradeable for servicing households and businesses. . . . Similarly, the day-to-day activities of many businesses all involve tasks that can be automated and/or performed remotely and, of course, across national boundaries. Thus, a natural prediction would be that the U.S. should become a net exporter of high-skilled, knowledge-intensive professional services because of its comparative advantage."

The American Economic Association

MIX
Paper from
responsible sources
FSC
www.fsc.org
FSC™ C132124

The Journal of
Economic Perspectives

Symposia

Minimum Wage

Alan Manning, "The Elusive Employment Effect of the Minimum Wage"
Arindrajit Dube and Attila Lindner, "City Limits:
What Do Local-Area Minimum Wages Do?"
Jeffrey Clemens, "How Do Firms Respond to Minimum Wage Increases?
Understanding the Relevance of Non-Employment Margins"
Price V. Fishback and Andrew J. Seltzer, "The Rise of American Minimum
Wages, 1912–1968"

Polarization in Courts

Adam Bonica and Maya Sen, "Estimating Judicial Ideology"
Daniel Hemel, "Can Structural Changes Fix the Supreme Court?"

Economics of Higher Education

David Figlio and Morton Schapiro, "Staffing the Higher Education Classroom"
John Bound, Breno Braga, Gaurav Khanna, and Sarah Turner, "The Globalization
of Postsecondary Education: The Role of International Students in the US Higher
Education System"
W. Bentley MacLeod and Miguel Urquiola, "Why Does the United States Have the
Best Research Universities? Incentives, Resources, and Virtuous Circles"

Articles

Florian Scheuer and Joel Slemrod, "Taxing Our Wealth"
Daron Acemoglu, "Melissa Dell: Winner of the 2020 Clark Medal"

Recommendations for Further Reading

AMERICAN
ECONOMIC
ASSOCIATION