The Journal of

Economic Perspectives

A journal of the
American Economic Association

Summer 2013

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@jepjournal.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA
Printed by R. R. Donnelley Company, Jefferson City, Missouri, 65109, USA

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Summer 2013, Vol. 27, No. 3. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $20.00, $30.00, or $40.00 depending on income; for an additional $15.00, you can receive this journal in print. E-reader versions are free. For details and further information on the AEA go to http://www.vanderbilt.edu/AEA/.
Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 27 • Number 3 • Summer 2013

Symposia

Features

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.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 3–20

The Top 1 Percent in International and Historical Perspective[†]

Facundo Alvaredo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez

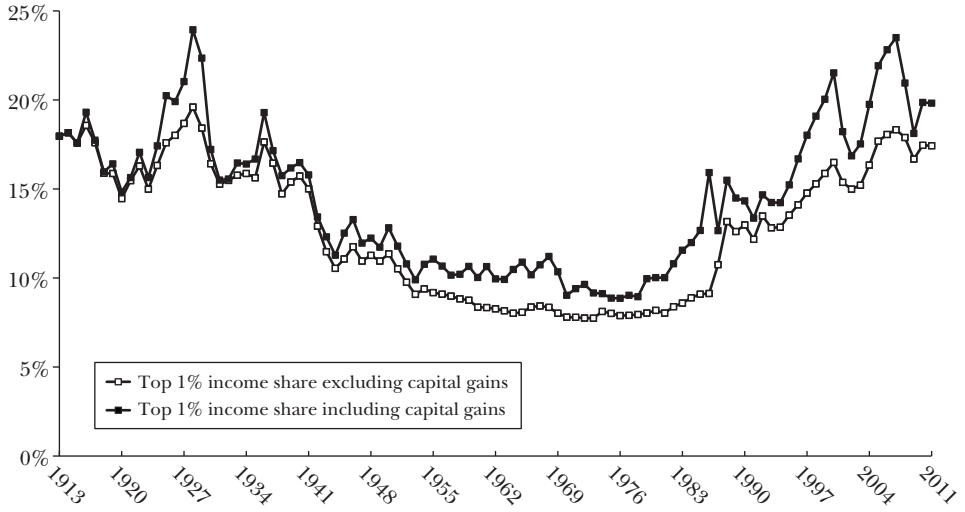
For three decades, the debate about rising income inequality in the United States has centered on the dispersion of wages and the increased premium for skilled/educated workers, attributed in varying proportions to skill-biased technological change and to globalization (for example, see Katz and Autor 1999 for a survey). In recent years, however, there has been a growing realization that most of the action has been at the very top. This has attracted a great deal of public attention (as witnessed by the number of visits to and press citations of our World Top Incomes Database at http://topincomes.parisschoolofeconomics.eu/) and has represented a challenge to the economics profession. Stories based on the supply and demand for skills are not enough to explain the extreme top tail of the earnings distribution; nor is it enough to look only at earned incomes. Different approaches are necessary to explain what has happened in the United States over the past century and also to explain the differing experience in other high-income countries over recent decades. We begin with the international comparison in the first section and then turn to the causes and implications of the evolution of top income shares.

■ *Facundo Alvaredo is Research Fellow at Nuffield College and Department of Economics, Oxford, United Kingdom, and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Buenos Aires, Argentina, and Affiliate Member, Paris School of Economics, Paris, France. Anthony B. Atkinson is Fellow of Nuffield College, Oxford, and Centennial Professor at the London School of Economics, London, United Kingdom. Thomas Piketty is Professor of Economics at the Paris School of Economics, Paris, France. Emmanuel Saez is Professor of Economics, University of California at Berkeley, United States. Their email addresses are alvaredo@gmail.com, tony.atkinson@nuffield.ox.ac.uk, piketty@ens.fr, and saez@econ.berkeley.edu, respectively.*

[†] To access the disclosure statements, visit
http://dx.doi.org/10.1257/jep.27.3.3

doi=10.1257/jep.27.3.3

Figure 1
Top 1 Percent Income Share in the United States



Source: Source is Piketty and Saez (2003) and the World Top Incomes Database.
Notes: The figure reports the share of total income earned by top 1 percent families in the United States from 1913 to 2011. Income is defined as pre-tax market income; it excludes government transfers and nontaxable fringe benefits. The figure reports series including realized capital gains (solid squares) and series excluding realized capital gains (hollow squares).

We should start by emphasizing the factual importance of the top 1 percent. It is tempting to dismiss the study of this group as a passing political fad due to the slogans of the Occupy movement or as the academic equivalent of reality TV. But the magnitudes are truly substantial. Based on pre-tax and pre-transfer market income (excluding nontaxable fringe benefits such as health insurance but including realized capital gains) per family reported on tax returns, the share of total annual income received by the top 1 percent has more than doubled from 9 percent in 1976 to 20 percent in 2011 (Piketty and Saez, 2003, and the World Top Incomes Database). There have been rises for other top shares, but these have been much smaller: during the same period, the share of the group from 95th to 99th percentile rose only by 3 percentage points. The rise in the share of the top 1 percent has had a noticeable effect on overall income inequality in the United States (Atkinson, Piketty, and Saez 2011, Section 2.2).

The United States Top 1 Percent in International Perspective

Figure 1 depicts the US top 1 percent income share since 1913. Simon Kuznets (1955) famously hypothesized that economic growth would first be accompanied by a rise in inequality and then by a decline in inequality. At first glance, it is tempting

to conclude from Figure 1 that the Kuznets curve has been turned upside-down. But this suggestion is too facile. After all, the interwar period did not exhibit a secular downward trend in shares of top incomes. Apart from the bubble of the late 1920s, the US top 1 percent share was between 15 and 20 percent throughout this time. At the time of Pearl Harbor in 1941, the share of the top 1 percent was essentially the same as in 1918. The downward trend in top shares started at the time of World War II and continued until the end of the 1960s. There was then a sharp reversal such that the top share is today back in the same range as in the 1920s. Interestingly, the Great Recession of 2008–2009 does not seem to have reversed the upward trend. There was a fall in the top 1 percent share in 2008–2009 but a rebound in 2010. This would be consistent with the experience of the previous economic downturn: top income shares fell in 2001–2002 but quickly recovered and returned to the previous trend in 2003–2007. Another piece of evidence that is consistent with this interpretation is the smaller cyclical variation in the series excluding capital gains (shown by the hollow squares in Figure 1).

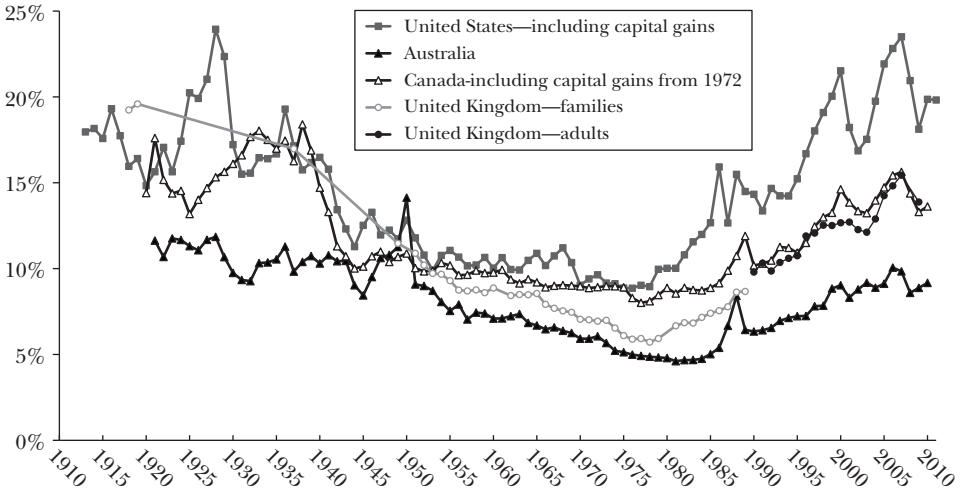
Has the US experience been reproduced in other high-income countries? The evolution of the shares of the top 1 percent is shown for four Anglo-Saxon countries in Figure 2A and for France, Germany, Sweden, and Japan in Figure 2B (it should be noted that the estimates for France and the United Kingdom do not include capital gains, the estimates for Canada, Germany, Japan, and Sweden include realized capital gains after the year therein shown, and the estimates for Australia include them only partially and at varying degrees over time). The other Anglo-Saxon countries—Australia, Canada, and the United Kingdom—all show a strong asymmetric U-shape. However, the rises were less marked in two of these countries. Over the period 1980 to 2007, when the top 1 percent share rose by some 135 percent in the United States and the United Kingdom, it rose by some 105 percent in Australia and 76 percent in Canada (and by 39 percent in New Zealand, not shown). The experience is markedly different in continental Europe and Japan, where the long pattern of income inequality is much closer to an L-shaped than a U-shaped curve. (Sweden and other Scandinavian countries such as Norway (not shown) are intermediate cases.)[1] There has been some rise in recent years in the top shares in these countries, but the top 1 percent shares are not far today from their levels in the late 1940s, whereas in the United States the share of the top 1 percent is higher by more than a half.

To us, the fact that high-income countries with similar technological and productivity developments have gone through different patterns of income inequality at the very top supports the view that institutional and policy differences play a key role in these transformations. Purely technological stories based solely upon supply and demand of skills can hardly explain such diverging patterns. What is more, within countries, we have to explain not only why top shares rose (in the U-shaped countries) but also why they fell for a sustained period of time earlier in

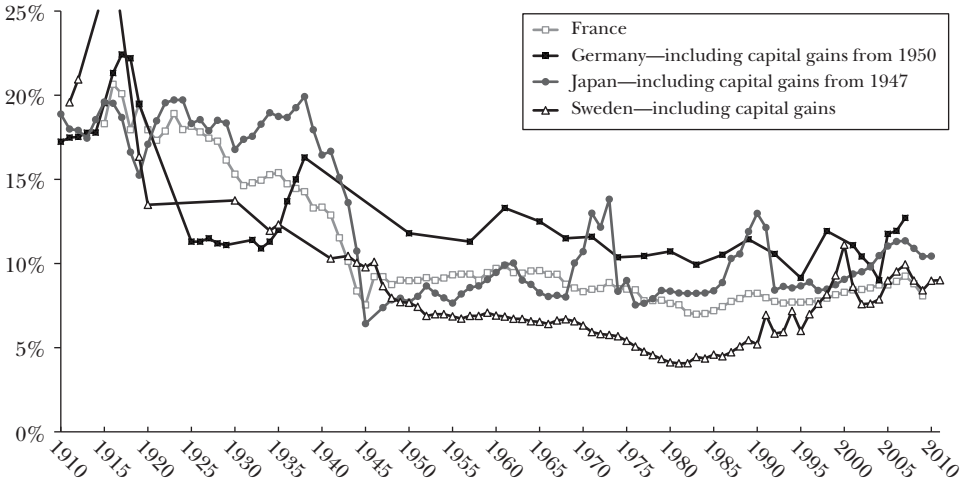
[1] The Swedish top 1 percent share was very high during World War I. The same is observed in Denmark—see the discussion in Atkinson and Søgaard (2013).

Figure 2
The Evolution of the Shares of the Top 1 Percent in Different Countries

A: Top 1 Percent Income Shares in English-speaking Countries (U-Shape)

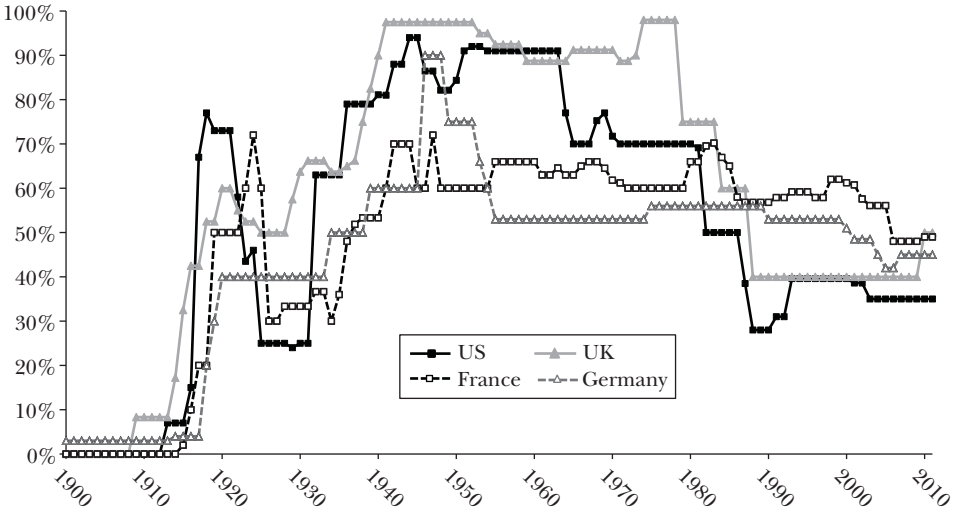


B: Top 1 Percent Income Shares in Continental Europe and Japan (L-Shape)



Source: The World Top Incomes Database.
Notes: The figure reports the share of total income earned by the top 1 percent in four English-speaking countries in panel A, and in four other OECD countries (Japan and three continental European countries) in panel B. Income is defined as pre-tax market income. The estimates for Australia include realized capital gains partially and at varying degrees over time.

Figure 3
Top Marginal Income Tax Rates, 1900–2011



Source: Piketty and Saez (2013, figure 1).
Notes: The figure depicts the top marginal individual income tax rate in the United States, United Kingdom, France, and Germany since 1900. The tax rate includes only the top statutory individual income tax rate applying to ordinary income with no tax preference. State income taxes are not included in the case of the United States. For France, we include both the progressive individual income tax and the flat rate tax "Contribution Sociale Generalisée."

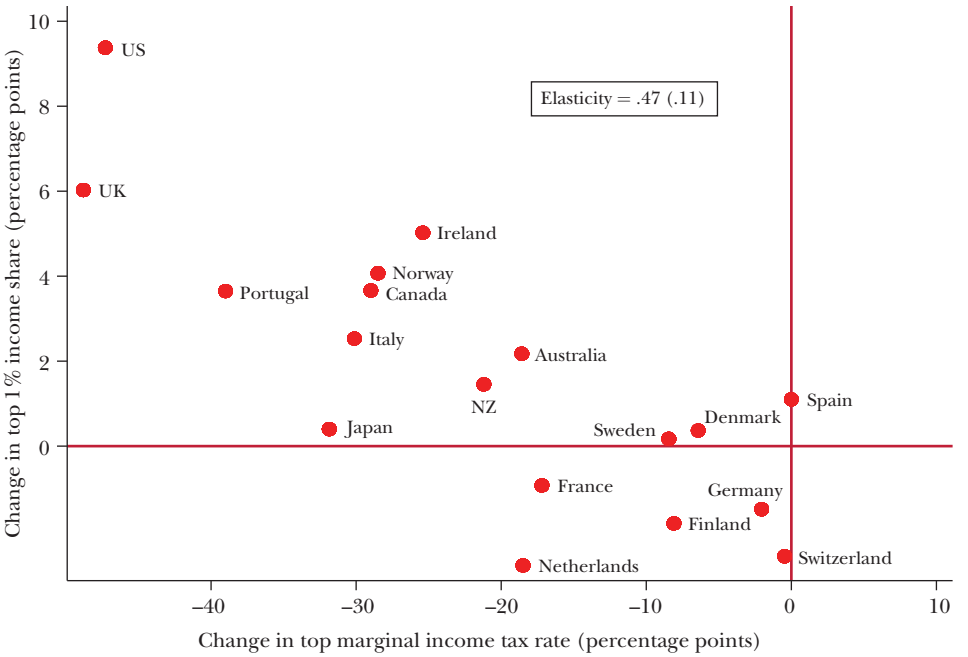
the twentieth century. The most obvious policy difference—between countries and over time—regards taxation, and it is here that we begin.

Taxes and Top Shares

During the twentieth century, top income tax rates have followed an inverse U-shaped time-path in many countries, as illustrated in Figure 3. In the United States, top income tax rates were consistently above 60 percent from 1932 to 1981, and at the start of the 1920s, they were above 70 percent (of course, varying proportions of taxpayers were subject to the top rate). High income tax rates are not just a feature of the post-World War II period, and their cumulative effect contributed to the earlier decline in top income shares. While many countries have cut top tax rates in recent decades, the depth of these cuts has varied considerably. For example, the top tax rate in France in 2010 was only 10 percentage points lower than in 1950, whereas the top tax rate in the US was less than half its 1950 value.

Figure 4 plots the changes in top marginal income tax rates (combining both central and local government income taxes) since the early 1960s against the changes over that period in top 1 percent income shares for 18 high-income countries in the World Top Incomes Database. It shows that there is a strong correlation between the reductions in top tax rates and the increases in top 1 percent

Figure 4
Changes in Top Income Shares and Top Marginal Income Tax Rates since 1960
(combining both central and local government income taxes)



Source: Piketty, Saez, and Stantcheva (2011, revised October 2012, figure 3). Source for top income shares is the World Top Incomes Database. Source for top income tax rates is OECD and country-specific sources.
Notes: The figure depicts the change in the top 1 percent income share against the change in the top income tax rate from 1960–64 to 2005–2009 for 18 OECD countries. If the country does not have top income share data for those years, we select the first available five years after 1960 and the most recent 5 years. For the following five countries, the data start after 1960: Denmark (1980), Ireland (1975), Italy (1974), Portugal (1976), Spain (1981). For Switzerland, the data end in 1995 (they end in 2005 or after for all the other countries). Top tax rates include both the central and local government top tax rates. The correlation between those changes is very strong. The elasticity estimates of the ordinary least squares regression of $\Delta\log(\text{top 1\% share})$ on $\Delta\log(1 - \text{MTR})$ based on the depicted dots is 0.47 (0.11).

pre-tax income shares. For example, the United States experienced a reduction of 47 percentage points in its top income tax rate and a 10 percentage point increase in its top 1 percent pre-tax income share. By contrast, countries such as Germany, Spain, or Switzerland, which did not experience any significant top rate tax cut, did not show increases in top 1 percent income shares. Hence, the evolution of top tax rates is strongly negatively correlated with changes in pre-tax income concentration.

This negative correlation can be explained in a variety of ways. As pointed out originally by Slemrod (1996), it is possible that the rise in top US income shares occurred because, when top tax rates declined, those with high incomes had less

reason to seek out tax avoidance strategies. This argument has more recently been used to deny that any real increase in income concentration actually took place—that it is a pure statistical artifact. Under this scenario, the real US top income shares were as high in the 1960s as they are today, but a smaller fraction of top incomes was reported on tax returns. While this factor may have affected the pattern of the data at certain times—for example, the jump in top US income shares following the 1986 Tax Reform Act—closer examination of the US case suggests that the tax avoidance response cannot account for a significant fraction of the long-run surge in top incomes. Top income shares based on a broader definition of income that includes realized capital gains, and hence a major portion of avoidance channels, have increased virtually as much as top income shares based on a narrower definition of income subject to the progressive tax schedule (see Figure 1 and Piketty, Saez, and Stantcheva 2011 for a detailed analysis).

The explanation that changes in tax rates in the top tax brackets do lead to substantive behavioral change has indeed received some support. After noting that top US incomes surged following the large top marginal tax rate cuts of the 1980s, Lindsey (1987) and Feldstein (1995) proposed a standard supply-side story whereby lower tax rates stimulate economic activity among top earners involving more work, greater entrepreneurship, and the like. In this scenario, lower top tax rates would lead to more economic activity by the rich and hence more economic growth.

Behavioral change is at the heart of the optimal income tax analysis pioneered by Mirrlees (1971) and publicly evoked in the debate about top tax rates in the UK, where the Chancellor of the Exchequer has argued that reducing the top tax rate below 50 percent (for broadly the top 1 percent) will not reduce revenue. The standard optimal tax formula (Diamond and Saez 2011) implies, with an elasticity of taxable income of 0.5, that the revenue-maximizing top tax rate would be 57 percent.[2] When allowance is made for other taxes levied in the United Kingdom, such as the payroll tax, this implies a top income tax rate in the United Kingdom of some 40 percent (Atkinson 2012).

Richer Models of Pay Determination

The optimal tax literature has, however, remained rooted in an oversimplified model of pay determination that takes no account of developments in labor economics, and the same applies to the explanations of changing top income shares. Changes in the pay of a worker are assumed to have no impact on either the other side of the labor market or on other workers. The worker generates more output and pay adjusts by the same amount. Each person is an island. However, in the now-standard models of job-matching, a job emerges as the result of the costly creation of a vacancy by the employer and of job search by the employee. A match

[2] The revenue-maximizing top tax rate formula takes the form $\tau = 1/(1 + a \cdot e)$ where a is the Pareto parameter of the top tail of the income distribution, and e is the elasticity of pre-tax income with respect to the net-of-tax rate $1 - \tau$. With $e = 0.5$ (as estimated from Figure 4) and $a = 1.5$ (the current Pareto parameter of the US income distribution), we get $\tau = 1/(1 + 0.5 \cdot 1.5) = 57$ percent.

creates a positive surplus, and there is Nash bargaining over the division of the surplus, leading to a proportion β going to the worker and $(1 - \beta)$ to the employer. Typically, β is assumed fixed, but it is possible that what we have observed, at least at the top, is an increase in β, which can lead to changes in the distribution of income.[3]

Why should β have increased? The extent to which top earners exercised bargaining power may have interacted with the changes in the tax system. When top marginal tax rates were very high, the net reward to a highly paid executive for bargaining for more compensation was modest. When top marginal tax rates fell, high earners started bargaining more aggressively to increase their compensation. In this scenario, cuts in top tax rates can increase top income shares—consistent with the observed trend in Figure 1—but the increases in top 1 percent incomes now come at the expense of the remaining 99 percent.

One can also weave this notion of greater incentives for bargaining into a broader scenario, in which the improved information and communications technology and globalization were increasing the demand for high-skilled labor, and the deregulation of finance and of other industries was both raising the demand for skill at the top and changing the rules under which compensation had been calculated in the past. In this perspective, high marginal tax rates had served as a brake on the level of surplus extraction in the past, but then this brake was released at the same time that economic and institutional conditions allowed for higher compensation at the top of the income distribution (Piketty, Saez, and Stantcheva 2011).

In this scenario, the higher share of income going to the top 1 percent does not reflect higher economic growth—which is a key difference with the supply-side scenario. It is even possible that reductions in top marginal tax rates may have adverse effects on growth, as may be seen if we go back to the theories of managerial firms and the separation of ownership and control developed by Oliver E. Williamson, William Baumol, and Robin Marris in the 1960s and 1970s (for discussion, see Solow 1971). In these models, managers are concerned with their remuneration (both monetary and nonmonetary) but also with other dimensions such as the scale or rate of growth of their firms, and allocate their effort accordingly. Where top tax rates were high, there was a low return to effort spent on negotiating higher pay. Top corporate executives may have concentrated on securing alternative sources of utility, such as unproductive corporate expenses, but they may also have ploughed back profits into securing faster expansion than in the traditional stock market valuation-maximizing firm. Cuts in top tax rates, however, meant that top executives switched efforts back to securing a larger share of the profits, in which case increases in remuneration, or bonuses, may have come at the expense of employment and growth.

The correlation shown in Figure 4 between top marginal tax rates and changes in top income shares may of course reflect in part coincidence rather than causality. The political factors that led to top tax rate cuts—such as those by Reagan and Thatcher

[3] Kleven, Landais, Saez, and Schultz (2013) find evidence of such bargaining effects in the pay determination of high earners, using the Danish preferential tax scheme for highly paid immigrants.

in the 1980s in the United States and the United Kingdom—were accompanied by other legislative changes, such as deregulation, which may have caused top incomes to rise, not least on account of the impetus they gave to the growth of the financial services (Philippon and Reshef 2012) and legal services sectors. More generally, the effects of taxation may interact with other changes, such as those in remuneration practices. Where there is a surplus to be shared, the division may reflect relative bargaining strength, as above, but it may also be influenced by social norms. Notions of fairness, or a "pay code," may come into play to remove the indeterminacy where "individual incentives are not by themselves . . . sufficient to determine a unique equilibrium" (MacLeod and Malcomson 1998, p. 400). A "pay code" limits the extent to which earnings are individually determined, a situation that both workers and employers accept on reputational grounds. As argued in Atkinson (2008), there may be a tipping-point where there is a switch from a high level of adherence to such a code to a situation where pay becomes largely individually determined. This has been documented in the case of the United States by Lemieux, MacLeod, and Parent (2009), who find an increase in the proportion of performance-pay jobs over the period 1976 to 1998. As they note, the increased extent of performance-pay may be a channel by which other factors are expressed in greater wage dispersion, and they stress the effect at the top end of the wage distribution.

Top Tax Rates and Growth

If we look at the aggregate outcomes, we find no apparent correlation between cuts in top tax rates and growth rates in real per capita GDP (Piketty, Saez, and Stantcheva 2011). Countries that made large cuts in top tax rates such as the United Kingdom or the United States have not grown significantly faster than countries that did not, such as Germany or Switzerland. This lack of correlation is more consistent with a story that the response of pre-tax top incomes to top tax rates documented in Figure 4 is due to increased bargaining power or more individualized pay at the top, rather than increased productive effort. Naturally, cross-country comparisons are bound to be fragile; exact results vary with the specification, years, and countries. However, the regression analysis by Piketty, Saez, and Stantcheva (2011), using the complete time-series data since 1960, shows that the absence of correlation between economic growth and top tax rates is quite robust. By and large, the bottom line is that rich countries have all grown at roughly the same rate over the past 40 years—in spite of huge variations in tax policies.

More specifically, international evidence shows that current pay levels for chief executive officers across countries are strongly negatively correlated with top tax rates even controlling for firm's characteristics and performance, and that this correlation is stronger in firms with poor governance (Piketty, Saez, and Stantcheva 2011).[4] This finding also suggests that the link between top tax rates and pay of chief

[4] Governance is measured with an index that combines various governance measures: insider ownership, institutional ownership, the ratio of independent board directors, whether the CEO is also chairman of the board, and the average number of board positions held by board members.

executive officers does not run through firm performance but is likely to be due to bargaining effects.

Such findings have strong implications for top tax rate policies. The optimal top tax rate rises dramatically if a substantial fraction of the effect of top tax rates on pre-tax top incomes documented in Figure 4 above is due to wage-bargaining effects instead of supply-side effects. Using mid-range parameter values where the response of top earners to top tax rate cuts is three-fifths due to increased bargaining behavior and two-fifths due to increased productive work, Piketty, Saez, and Stantcheva (2011) find that the top tax rate could potentially be set as high as 83 percent—as opposed to 57 percent in the pure supply-side model.[5]

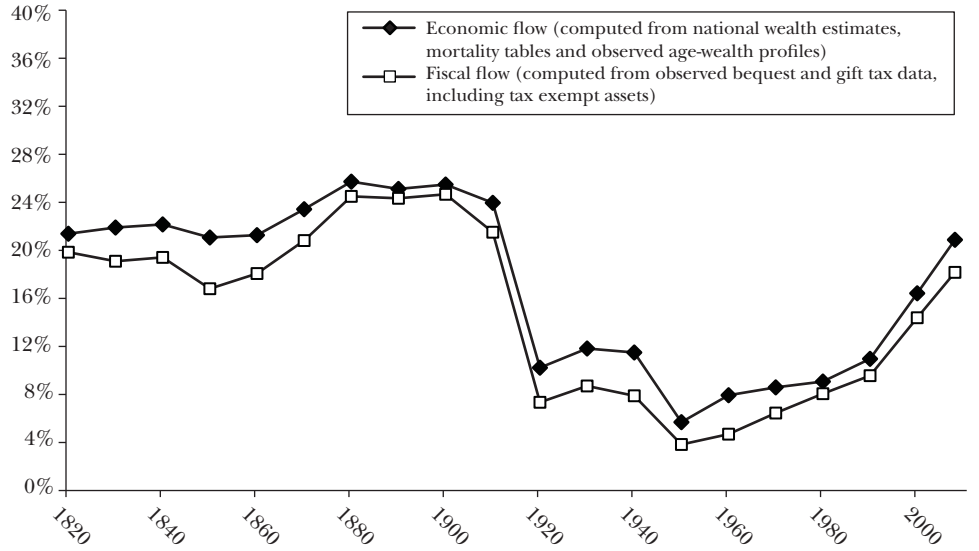
Capital Income and Inheritance

The analysis just cited focused—like much of the literature—on what is commonly called "earned incomes," referring to income received in return for work. But capital income is also an important part of the story. Of course, the distinction between the two types of income can become blurry in some cases—notably, entrepreneurial income can have elements of both compensation for work and a return to capital investment. Here, we define "capital income" as rents, dividends, interest, and realized capital gains. The decline of top capital incomes is the main driver of the falls in top income shares that occurred in many countries early in the twentieth century. For example, from 1916 to 1939, capital income represented 50 percent of US top 1 percent incomes, whereas by the end of the century from 1987 to 2010, the share had fallen to one-third (Piketty and Saez 2003, tables A7 and A8). In the United Kingdom, the corresponding share fell from 60 percent in 1937 to under 20 percent by the end of the century (Atkinson 2007, figure 4.11). At the same time, it should be borne in mind that these calculations depend on the definition of taxable incomes. In times past, a number of income tax systems like those in France and the United Kingdom included imputed rents of homeowners in the income tax base, but today imputed rents are typically excluded. Where the tax base has been extended, this has in some cases taken the form of separate taxation (as with realized capital gains in the United Kingdom), so that this element of capital income is not covered in the income tax data. As a result of these developments, the share of capital income that is reportable on income tax returns has often significantly decreased over time.

Earlier we referred to the cumulative effect of progressive taxation. A long period of high top rates of income taxation, coupled with high top rates of taxation on the transmission of wealth by inheritance and gift, reduced the capacity of large

[5] With wage-bargaining effects, the optimal top tax rate formula becomes $\tau = (1 + s \cdot a \cdot e)/(1 + a \cdot e)$ where s is the fraction of the total behavioral elasticity due to bargaining effects. With $a = 1.5$, $e = 0.5$ (as above), and $s = 3/5$, we obtain $\tau = 83$ percent. In the standard model with no wage-bargaining effects, we had $s = 0$ and $\tau = 57$ percent.

Figure 5
Annual Inheritance Flow as a Fraction of Disposable Income, France 1820–2008



Source: Piketty (2011).
Notes: The annual inheritance flow is defined as the total market value of all assets (tangible and financial assets, net of financial liabilities) transmitted at death or through *inter vivos* gifts. Disposable income was as high as 90–95 percent of national income during the 19th century and early 20th century (when taxes and transfers were almost nonexistent), while it is now about 70 percent of national income.

wealth-holders to sustain their preeminence. The key factor in determining the capacity to transmit wealth is the difference between the "internal rate of accumulation" (the savings rate times the rate of return net of taxes) and the rate of growth of the economy. This means that the taxation of income and wealth transfers can cause the share of top wealth-holders to fall, as in the United Kingdom over the first three-quarters of the twentieth century (Atkinson and Harrison 1978), contributing to the downward trajectory of top income shares. Alongside this was the growth of "popular wealth" owned by the bottom 99 percent. Back in 1908 in the United Kingdom, the 17th Earl of Derby had a rent roll of some £100,000, which was more than 1,000 times the average income at the time. Many of these houses are now owned by their occupiers.

In recent decades, however, the relation between the internal rate of accumulation of wealth holdings and the rate of growth of capital has now been reversed as a result of the cuts in capital taxation and the decline in the macroeconomic growth rate (Piketty 2011). As a result, a number of countries are witnessing a return of inheritance as a major factor. Figure 5 shows the estimates of Piketty (2011) for France for the period 1820 to 2008 of the annual inheritance flow (the amount passed on through bequests and gifts *inter vivos*), expressed as percentage of

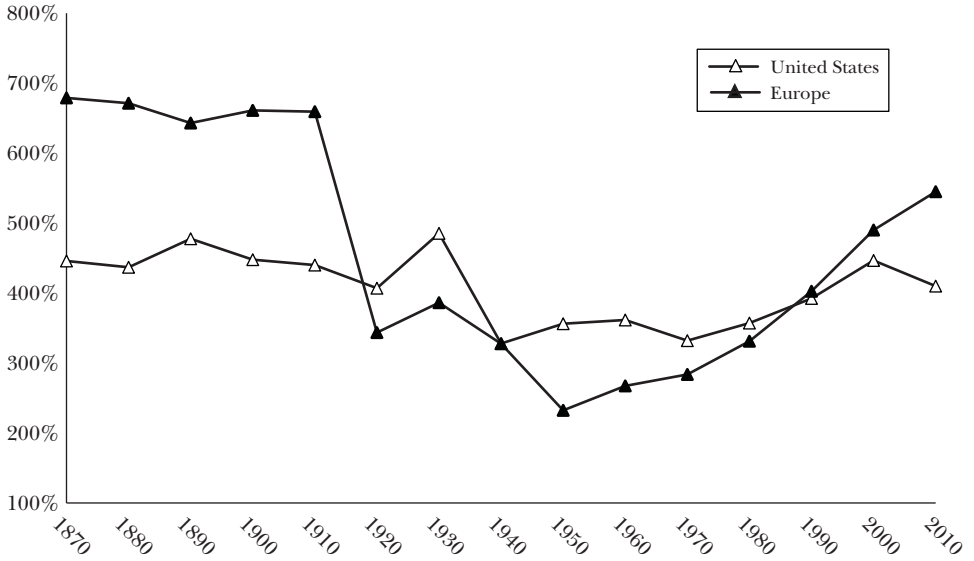
disposable income.[6] Two methods are employed: a constructive calculation from national wealth figures, mortality rates, and observed age-wealth profiles, and an estimate based on the estate and gift tax records. The two methods differ in levels (the fiscal flows are lower), but the time-paths are very similar.

The inheritance flow in France was relatively stable around 20–25 percent of disposable income throughout the 1820–1910 period (with a slight upward trend), before being divided by a factor of about 5 to 6 between 1910 and the 1950s. Since then, it has been rising regularly, with an acceleration of the trend during the past 30 years. These truly enormous historical variations bring France back to a situation similar to that of 100 years ago. An annual inheritance flow around 20 percent of disposable income is very large. It is typically much larger than the annual flow of new savings and almost as big as the annual flow of capital income. This implies that inheritance is again becoming a very important factor of lifetime economic inequality. As shown in Piketty and Saez (2012), in a world where inheritance is quantitatively significant, those receiving no bequests will leave smaller-than-average bequests themselves and hence should support shifting labor taxation toward bequest taxation. In this situation, inheritance taxation (and more generally capital taxation, given capital market imperfections) becomes a powerful and desirable tool for redistribution toward those receiving no inheritance.

The return of inherited wealth may well differ in magnitude across countries. The historical series available so far regarding the inheritance flows are too scarce to reach firm conclusions. Existing estimates suggest that the French U-shaped pattern also applies to Germany, and to a lesser extent to the United Kingdom and the United States (Atkinson 2013; Schinke 2012; see Piketty and Zucman, forthcoming, for a survey). Such variations could be due to differences in pension systems and the share of private wealth that is annuitized (and therefore nontransmissible). From a theoretical perspective, it is unclear however why there should be much crowding out between lifecycle wealth and transmissible wealth in an open economy (that is, the fact that individuals save more for their pension should not make them save less for their children; the extra pension wealth coming from the lifecycle motive should be invested abroad). It could be that there are differences in tastes for wealth transmission. Maybe wealthy individuals in the United Kingdom and in the United States have less taste for bequest than their French and German counterparts. However it should be kept in mind that there are important data problems (in particular, wealth surveys tend to vastly underestimate inheritance receipts), which could partly explain why the rise of inheritance flows in the recent period appears to be more limited in some

[6] It is critical to include both bequests (wealth transmitted at death) and gifts (wealth transmitted *inter vivos*) in our definition of inheritance, first because gifts have always represented a large fraction of total wealth transmission, and second because this fraction has changed a lot over time.

Figure 6
Private Wealth/National Income Ratios, 1870–2010



Source: Piketty and Zucman (2013).
Notes: Europe is the (unweighted) average of France, Germany, and the United Kingdom. Private wealth is defined as the sum of nonfinancial assets, financial assets, minus financial liabilities in the household and nonprofit sectors.

countries than in others.[7] Another source of difference between countries could come from variations in the total magnitude of wealth accumulation. There may in this respect be an important difference between the United States and Europe, as is indeed suggested when we look at total private wealth (expressed as a ratio to national income), shown in Figure 6 (see Piketty and Zucman, 2013, for a discussion on the differences between private and national wealth).

As may be seen from Figure 6, the twentieth century has seen a U-shaped time-path in the ratio of private wealth to national income that is more marked in Europe than in the United States. Private wealth in Europe was around six times

[7] In particular, the smaller rise of the UK inheritance flow (as compared to France and Germany) is entirely due to the much smaller rise of recorded *inter vivos* gifts, which according to fiscal data barely rose in the United Kingdom during recent decades, while they have become almost as large as bequests in France and Germany. This might simply be due to the fact that gifts are not properly recorded by the UK tax administration (Atkinson 2013). In the United States, due to the limitations of federal fiscal data on bequests and gifts, scholars often use retrospective wealth survey data. The problem is that in countries with exhaustive administrative data on bequests and gifts (such as France, and to some extent Germany), survey-based self-reported flows appear to be less than 50 percent of fiscal flows. This probably contributes to explaining the low level of inheritance receipts found in a number of US studies. An example of such a study is Wolff and Gittleman (2011); one additional bias in this study is that inherited assets are valued using asset prices at the time these assets were transmitted, and no capital gain or income is included.

national income in 1910, and then fell after the World Wars to less than two and a half times in 1950. In the past 60 years, it has risen sharply to reach more than five times national income. This pattern suggests that capital is "back" and that the low wealth–income ratios observed in Europe from the 1950s to the 1970s were an anomaly. This can be well accounted for by the long-run wealth accumulation formula $\beta = s/g$, where β is the Harrod–Domar–Solow wealth/income ratio, s is the saving rate, and g is the growth rate including both real per capita and population growth. For a given saving rate (say $s = 10$ percent), you accumulate a lot more wealth relative to income in the long run when the growth rate is 1.5 to 2 percent than if the growth rate is 2.5 to 3 percent. Given the large and continuing difference in population growth rates between Old Europe and the New World, this can explain not only the long-run changes but also the difference in levels between Europe and the United States (Piketty 2011; Piketty and Zucman 2013).[8]

On the other hand, it should be noted that *wealth concentration* (as opposed to wealth accumulation) is significantly greater in the United States, where the top 1 percent owns about 35 percent of aggregate wealth (for comparison, the share is about 20–25 percent in Europe). So far, existing studies have found that the increase in US wealth concentration since the 1970s and 1980s has been relatively moderate in contrast to the huge increase in US income concentration documented above (Kennickell 2009; Kopczuk and Saez 2004). However, we should be modest about our ability to measure the trends in top billionaire wealth. With low and diminishing growth rates and high global returns to capital, the potential for divergence of the wealth distribution is naturally quite large.

Joint Distribution of Earned and Capital Income

We have discussed earned income and capital income. The last piece of the puzzle concerns the *joint* distribution of earned and capital incomes—an aspect that is rarely given explicit consideration. Yet it is important to know whether the same people are at the top of both the distribution of capital income and the distribution of earned income. Suppose that we imagine asking the population first to line up along one side of a room in increasing order of their earned income and then to go to the other side of the room and line up in increasing order of their capital income. How much will they cross over? In the Ricardian class model, the crossing is complete: the capitalists come at the top in one case and at the bottom in the other. Has a negative correlation in the nineteenth century been replaced today by a zero correlation? Or is there a perfect correlation, so that people cross straight over? The pattern of crossing is given by the copula, which represents the joint distribution in terms of a function of the ranks in the two distributions of earnings and capital income. Because the copula compares ranks, it is not affected by whether the distributions themselves are widening or narrowing.

[8] In a way, this is equivalent to the explanation based upon lower bequest taste: with higher population growth and the same bequest taste (per children), the United States should save more. However a significant part of US population growth historically comes from migration, so this interpretation is not fully accurate.

Table 1
Relation between Top Labor Incomes and Top Capital Incomes in the United States

	Year	
	1980	*2000*
A: Percent of top 1% capital incomes in various top labor income groups		
Labor income groups:		
Top 1%	17%	27%
Top 5%	27%	45%
Top 10%	32%	52%
Top 20%	38%	61%
B: Percent of top 1% labor incomes in various top capital income groups		
Capital income groups:		
Top 1%	17%	27%
Top 5%	36%	50%
Top 10%	47%	63%
Top 20%	68%	80%

Source: Aaberge, Atkinson, Königs, and Lakner (forthcoming).
Notes: Panel A reports the percent of top 1 percent capital income earners in various top labor income groups in 1980 (column 1) and 2000 (column 2). In 2000, 27 percent of top 1 percent capital income earners were also in the top 1 percent of labor incomes, 45 percent were in the top 5 percent of labor incomes, etc. Panel B reports the percent of top 1 percent labor income earners in various top capital income groups in 2000 (column 1) and 1980 (column 2). The computations are based on the public use US tax return micro-datafiles (see Aaberge et al., forthcoming, for complete details).

What can be learned by considering the copula? Table 1 shows results for the United States in 2000 and in 1980 based on tax return data analysis from Aaberge, Atkinson, Königs, and Lakner (forthcoming). Three conclusions may be drawn. First, the joint distribution is asymmetric. In 2000, of those in the top 1 percent of capital income, 61 percent were in the top 20 percent of earned income. However, turning things round, of those in the top 1 percent of earned income, a larger proportion of 80 percent were in the top 20 percent of capital income. In fact, 63 percent of the top 1 percent of earners were in the top *10* percent of capital income. Such asymmetry could easily be missed by the use of a measure such as the correlation coefficient or a parametric form for the copula function. Second, the degree of association appears strong. Even for capital income, over half of the top 1 percent find themselves in the top tenth of earners. A quarter are in the top 1 percent for both. Third, the numbers for 1980 are all smaller than their counterparts for 2000. The degree of association increased between 1980 and 2000: in 1980 only 17 percent were in the top 1 percent for both. The proportion of the top 1 percent of earners who were in the top 5 percent of capital income rose from one-third to one-half, and the reverse proportion rose from 27 to 45 percent.

To understand the changing relationship between earned and capital incomes, we need to consider the mechanisms that link the two sources. In one direction, there is the accumulation of wealth out of earned income. Here the opportunities have changed in Anglo-Saxon countries. A third of a century ago, Kay and King (1980, p. 59) described the hypothetical position of a senior executive with a large corporation in the United Kingdom who had saved a quarter of his after-tax earnings: "[F]eeling . . . that he has been unusually fortunate in his career and unusually thrifty . . . he may be somewhat surprised to discover that there are in Britain at least 100,000 people richer than he is." Today, a chief executive officer may be both better paid and more able to accumulate. In the other direction, there is the effect of large family wealth on earnings. In the past, the link may have been negative, whereas today it may be socially unacceptable to live purely off unearned income. Wealth/family connections may provide access to high-paying employment (to assess this, it is necessary to investigate the cross-generation correlation of all income, not just earnings).

Conclusions

The rise in top income shares in the United States has been dramatic. In seeking explanations, however, it would be misleading to focus just on the doubling of the share of income going to the top 1 percent of the US distribution over the past 40 years. We also have to account for the fact that a number of high-income countries have seen more modest or little increase in top shares. Hence, the explanation cannot rely solely on forces common to advanced countries, like the impact of new technologies and globalization on the supply and demand for skills. Moreover, the explanations have to accommodate the falls in top income shares earlier in the twentieth century that characterize the countries discussed here.

In this paper, we have highlighted four main factors that have contributed to the growing income shares at the very top of the income distribution, noting that they may operate to differing extents in the United States and other countries, particularly in continental Europe. The first is tax policy: top tax rates have moved in the opposite direction from top pre-tax income shares. The second factor is a richer view of the labor market, where we have contrasted the standard supply-side model with the alternative possibility that there may have been changes to bargaining power and greater individualization of pay. Tax cuts may have led managerial energies to be diverted to increasing their remuneration at the expense of enterprise growth and employment. The third factor is capital income. In Europe—but less so in the United States—private wealth (relative to national income) has followed a spectacular U-shaped path over time, and inherited wealth may be making a return, implying that inheritance and capital income taxation will become again central policy tools for curbing inequality. The final, little-investigated, element is the correlation between earned income and capital income, which have become more closely associated in the United States.

■ *We are grateful to the journal's Editor, David Autor, the Managing Editor, Timothy Taylor, and Coeditors, Chang-Tai Hseih and Ulrike Malmendier, for most helpful comments. Financial support from the MacArthur Foundation, the Center for Equitable Growth at UC Berkeley, the Institute for New Economic Thinking at the Oxford Martin School, and the ESRC-DFID Joint Fund is thankfully acknowledged.*

References

Aaberge, Rolf, Anthony B. Atkinson, Sebastian Königs, and Christoph Lakner. Forthcoming. "Wages, Capital and Top Incomes." Unpublished paper, not completed.

Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez. 2011. The World Top Incomes Database, online at http://topincomes.g-mond.parisschoolofeconomics.eu/, May 15, 2013.

Atkinson, Anthony B. 2007. "The Distribution of Top Incomes in the United Kingdom 1908–2000." In *Top Incomes over the Twentieth Century—A Contrast between Continental European and English-Speaking Countries,* edited by A. B. Atkinson and T. Piketty, 82–140. Oxford University Press.

Atkinson, Anthony B. 2008. *The Changing Distribution of Earnings in OECD Countries.* Oxford University Press.

Atkinson, Anthony B. 2012. "The Mirrlees Review and the State of Public Economics." *Journal of Economic Literature* 50(3): 770–80.

Atkinson, Anthony B. 2013. "Wealth and Inheritance in Britain from 1896 to the Present." Unpublished paper.

Atkinson, Anthony B., and Allan J. Harrison. 1978. *Distribution of Personal Wealth in Britain.* Cambridge University Press.

Atkinson, Anthony B., and Thomas Piketty. 2007. *Top Incomes over the Twentieth Century—A Contrast between Continental European and English-Speaking Countries.* Oxford University Press.

Atkinson, Anthony B., and Thomas Piketty. 2010. *Top Incomes: A Global Perspective.* Oxford University Press.

Atkinson, Anthony B., Thomas Piketty, and Emmanuel Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49(1): 3–71.

Atkinson, Anthony B., and Jakob E. Søgaard. 2013. "The Long-Run History of Income Inequality in Denmark: Top Incomes from 1870 to 2010." EPRU working paper 2013-01, Economic Policy Research Unit, University of Copenhagen.

Diamond, Peter, and Emmanuel Saez. 2011 "The Case for a Progressive Tax: From Basic Research to Policy Recommendations." *Journal of Economic Perspectives* 25(4): 165–90.

Feldstein, Martin. 1995. "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy* 103(3): 551–72.

Katz, Lawrence F., and David H. Autor. 1999. "Changes in the Wage Structure and Earnings Inequality." Chap. 26 in *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, Vol. 3A. Amsterdam: North-Holland.

Kay, John A., and Mervyn A. King. 1980. *The British Tax System.* Oxford University Press.

Kennickell, Arthur B. 2009. "Ponds and Streams: Wealth and Income in the U.S., 1989 to 2007." Finance and Economics Discussion Series 2009-13, Federal Reserve Board, Washington, DC.

Kleven, Henrik Jacobsen, Camille Landais, Emmanuel Saez, and Esben Schultz. 2013. "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark." NBER Working Paper 18885.

Kopczuk, Wojciech, and Emmanuel Saez. 2004. "Top Wealth Shares in the United States, 1916–2000: Evidence from Estate Tax Returns." *National Tax Journal* 57(2, part 2): 445–88.

Kuznets, Simon. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45(1): 1–28.

Lemieux, Thomas, W. Bentley MacLeod, and Daniel Parent. 2009. "Performance Pay and Wage Inequality." *Quarterly Journal of Economics* 124(1): 1–49.

Lindsey, Lawrence B. 1987. "Individual Taxpayer Response to Tax Cuts: 1982–84: With Implications for the Revenue Maximizing Tax Rate." *Journal of Public Economics* 33(2): 173–206.

MacLeod, W. Bentley, and Malcomson, James M. 1998. "Motivation and Markets." *American Economic Review* 88(3): 388–411.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38(2): 175–208.

Philippon, Thomas, and Ariell Reshef. 2012. "Wages and Human Capital in the U.S. Finance Industry: 1909–2006." *Quarterly Journal of Economics* 127(4): 1551–1609.

Piketty, Thomas. 2011. "On the Long-run Evolution of Inheritance: France 1820–2050." *Quarterly Journal of Economics* 126(3): 1071–1131.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1): 1–39. (Series updated to 2011 in January 2013; available at Saez's website: http://elsa.berkeley.edu/~saez/.)

Piketty, Thomas, and Emmanuel Saez. 2012. "A Theory of Optimal Inheritance Taxation." CEPR Discussion Paper 9241. (Forthcoming in *Econometrica*.)

Piketty, Thomas, and Emmanuel Saez. 2013. "Optimal Labor Income Taxation." Chap. 9 in *Handbook of Public Economics*, Vol. 5, edited by A. Auerbach, R. Chetty, M. Feldstein, and E. Saez. Elsevier-North Holland.

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva. 2011. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities." NBER Working Paper 17616. (Forthcoming in *American Economic Journal: Economic Policy*.)

Piketty, Thomas, and Gabriel Zucman. 2013. "Capital is Back: Wealth-Income Ratios in Rich Countries, 1700–2010." http://piketty.pse.ens.fr/files/PikettyZucman2013WP.pdf.

Piketty, Thomas, and Gabriel Zucman. Forthcoming. "Wealth and Inheritance in the Long Run." In *Handbook of Income Distribution*, Vol. 2, edited by A. Atkinson and F. Bourguignon. Elsevier-North Holland.

Schinke, Christoph. 2012. "Inheritance in Germany 1911 to 2009: A Mortality Multiplier Approach." DIW-Berlin SOEP Paper 462.

Slemrod, Joel. 1996. "High Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response." In *Empirical Foundations of Household Taxation*, edited by Martin Feldstein and James M. Poterba, 169–92. University of Chicago.

Solow, Robert. 1971. "Some Implications of Alternative Criteria for the Firm." In *The Corporate Economy*, edited by R. Marris and A. Wood, 318–42. Macmillan.

Wolff, Edward N., and Maury Gittleman. 2011. "Inheritances and the Distribution of Wealth or Whatever Happened to the Great Inheritance Boom?" BLS Working Paper 445, US Bureau of Labor Statistics.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 21–34

Defending the One Percent

N. Gregory Mankiw

Imagine a society with perfect economic equality. Perhaps out of sheer coincidence, the supply and demand for different types of labor happen to produce an equilibrium in which everyone earns exactly the same income. As a result, no one worries about the gap between the rich and poor, and no one debates to what extent public policy should make income redistribution a priority. Because people earn the value of their marginal product, everyone has the appropriate incentive to provide the efficient amount of effort. The government is still needed to provide public goods, such as national defense, but those are financed with a lump-sum tax. There is no need for taxes that would distort incentives, such as an income tax, because they would be strictly worse for everyone. The society enjoys not only perfect equality but also perfect efficiency.

Then, one day, this egalitarian utopia is disturbed by an entrepreneur with an idea for a new product. Think of the entrepreneur as Steve Jobs as he develops the iPod, J. K. Rowling as she writes her Harry Potter books, or Steven Spielberg as he directs his blockbuster movies. When the entrepreneur's product is introduced, everyone in society wants to buy it. They each part with, say, $100. The transaction is a voluntary exchange, so it must make both the buyer and the seller better off. But because there are many buyers and only one seller, the distribution of economic well-being is now vastly unequal. The new product makes the entrepreneur much richer than everyone else.

The society now faces a new set of questions: How should the entrepreneurial disturbance in this formerly egalitarian outcome alter public policy? Should public policy remain the same, because the situation was initially acceptable and

■ *N. Gregory Mankiw is the Robert M. Beren Professor of Economics, Harvard University, Cambridge, Massachusetts. His email address is ngmankiw@harvard.edu.*

http://dx.doi.org/10.1257/jep.27.3.21 doi=10.1257/jep.27.3.21

the entrepreneur improved it for everyone? Or should government policymakers deplore the resulting inequality and use their powers to tax and transfer to spread the gains more equally?

In my view, this thought experiment captures, in an extreme and stylized way, what has happened to US society over the past several decades. Since the 1970s, average incomes have grown, but the growth has not been uniform across the income distribution. The incomes at the top, especially in the top 1 percent, have grown much faster than average. These high earners have made significant economic contributions, but they have also reaped large gains. The question for public policy is what, if anything, to do about it.

This development is one of the largest challenges facing the body politic. A few numbers illustrate the magnitude of the issue. The best data we have on the upper tail of the income distribution come from Piketty and Saez's (2003, with updates) tabulations of individual tax returns. (Even these numbers, though, are subject to some controversy: the tax code changes over time, altering the incentives to receive and report compensation in alternative forms.) According to their numbers, the share of income, excluding capital gains, earned by the top 1 percent rose from 7.7 percent in 1973 to 17.4 percent in 2010. Even more striking is the share earned by the top 0.01 percent—an elite group that, in 2010, had a membership requirement of annual income exceeding $5.9 million. This group's share of total income rose from 0.5 percent in 1973 to 3.3 percent in 2010. These numbers are not easily ignored. Indeed, they in no small part motivated the Occupy movement, and they have led to calls from policymakers on the left to make the tax code more progressive.

At the outset, it is worth noting that addressing the issue of rising inequality necessarily involves not just economics but also a healthy dose of political philosophy. We economists must recognize not only the limits of what we know about inequality's causes, but also the limits on the ability of our discipline to prescribe policy responses. Economists who discuss policy responses to increasing inequality are often playing the role of amateur political philosopher (and, admittedly, I will do so in this essay). Given the topic, that is perhaps inevitable. But it is useful to keep in mind when we are writing as economists and when we are venturing beyond the boundaries of our professional expertise.

Is Inequality Inefficient?

It is tempting for economists who abhor inequality to suggest that the issue involves not just inequality per se, but also economic inefficiency. Discussion of inequality necessarily involves our social and political values, but if inequality also entails inefficiency, those normative judgments are more easily agreed upon. The Pareto criterion is the clearest case: if we can make some people better off without making anyone worse off, who could possibly object? Yet for the question at hand, this criterion does not take us very far. As far as I know, no one has proposed any

credible policy intervention to deal with rising inequality that will make everyone, including those at the very top, better off.

More common is the claim that inequality is inefficient in the sense of shrinking the size of the economic pie. (That is, inefficiency is being viewed through the lens of the Kaldor–Hicks criterion.) If the top 1 percent is earning an extra $1 in some way that reduces the incomes of the middle class and the poor by $2, then many people will see that as a social problem worth addressing. For example, suppose the rising income share of the top 1 percent were largely attributable to successful rent-seeking. Imagine that the government were to favor its political allies by granting them monopoly power over certain products, favorable regulations, or restrictions on trade. Such a policy would likely lead to both inequality and inefficiency. Economists of all stripes would deplore it. I certainly would.

Joseph Stiglitz's (2012) book, *The Price of Inequality*, spends many pages trying to convince the reader that such rent-seeking is a primary driving force behind the growing incomes of the rich. This essay is not the place for a book review, but I can report that I was not convinced. Stiglitz's narrative relies more on exhortation and anecdote than on systematic evidence. There is no good reason to believe that rent-seeking by the rich is more pervasive today than it was in the 1970s, when the income share of the top 1 percent was much lower than it is today.

I am more persuaded by the thesis advanced by Claudia Goldin and Lawrence Katz (2008) in their book *The Race between Education and Technology*. Goldin and Katz argue that skill-biased technological change continually increases the demand for skilled labor. By itself, this force tends to increase the earnings gap between skilled and unskilled workers, thereby increasing inequality. Society can offset the effect of this demand shift by increasing the supply of skilled labor at an even faster pace, as it did in the 1950s and 1960s. In this case, the earnings gap need not rise and, indeed, can even decline, as in fact occurred. But when the pace of educational advance slows down, as it did in the 1970s, the increasing demand for skilled labor will naturally cause inequality to rise. The story of rising inequality, therefore, is not primarily about politics and rent-seeking, but rather about supply and demand.

To be sure, Goldin and Katz focus their work on the broad changes in inequality, not on the incomes of the top 1 percent in particular. But it is natural to suspect that similar forces are at work. The income share of the top 1 percent exhibits a U-shaped pattern: falling from the 1950s to the 1970s, and rising from the 1970s to the present. The earnings differentials between skilled and unskilled workers studied by Goldin and Katz follow a similar U-shaped pattern. If Goldin and Katz are right that the broad changes in inequality have been driven by the interaction between technology and education, rather than changes in rent-seeking through the political process, then it would seem an unlikely coincidence that the parallel changes at the top have been driven by something entirely different. Rather, it seems that changes in technology have allowed a small number of highly educated and exceptionally talented individuals to command superstar incomes in ways that were not possible a generation ago. Erik Brynjolfsson and Andrew McAfee (2011) advance this thesis forcefully in their book *Race Against the Machine*. They write,

"Aided by digital technologies, entrepreneurs, CEOs, entertainment stars, and financial executives have been able to leverage their talents across global markets and capture reward that would have been unimaginable in earlier times (p. 44)."

Nonetheless, to the extent that Stiglitz is right that inefficient rent-seeking is a driving force behind rising inequality, the appropriate policy response is to address the root cause. It is at best incomplete and at worst misleading to describe the situation as simply "rising inequality," because inequality here is a symptom of a deeper problem. A progressive system of taxes and transfers might make the outcome more equal, but it would not address the underlying inefficiency. For example, if domestic firms are enriching themselves at the expense of consumers through quotas on imports (as is the case with some agribusinesses), the solution to the problem entails not a revision of the tax code, but rather a change in trade policy. I am skeptical that such rent-seeking activities are the reason why inequality has risen in recent decades, but I would support attempts to reduce whatever rent-seeking does occur.

An especially important and particularly difficult case is the finance industry, where many hefty compensation packages can be found. On the one hand, there is no doubt that this sector plays a crucial role. Those who work in commercial banks, investment banks, hedge funds, and other financial firms are in charge of allocating capital and risk, as well as providing liquidity. They decide, in a decentralized and competitive way, which firms and industries need to shrink and which will be encouraged to grow. It makes sense that a nation would allocate many of its most talented and thus highly compensated individuals to this activity. On the other hand, some of what occurs in financial firms does smack of rent-seeking: when a high-frequency trader figures out a way to respond to news a fraction of a second faster than a competitor, the vast personal reward may well exceed the social value of what is produced. Devising a legal and regulatory framework to ensure that we get the right kind and amount of financial activity is a difficult task. While the solution may well affect the degree of equality and the incomes of the 1 percent, the issue is primarily one of efficiency. A well-functioning economy needs the correct allocation of talent. The last thing we need is for the next Steve Jobs to forgo Silicon Valley in order to join the high-frequency traders on Wall Street. That is, we shouldn't be concerned about the next Steve Jobs striking it rich, but we want to make sure he strikes it rich in a socially productive way.

Equality of Opportunity as a Desideratum

Closely related to the claim of inefficiency is concern about inequality of opportunity. Equality of opportunity is often viewed as a social goal in itself, but economists recognize that the failure to achieve such equality would normally lead to inefficiency as well. If some individuals are precluded from pursuing certain paths in life, then they might be unable to contribute fully to growing the economic pie. To be specific, if children from poor families are unable to continue their education because of financial constraints, they do not accumulate the optimal amount of

human capital. The outcome from underinvestment in education is both unequal and inefficient.

Measuring the degree of equality of opportunity is difficult. In his book, Stiglitz (2012) proposes a metric: the intergenerational transmission of income. He writes (p. 18), "If America were really a land of opportunity, the life chances of success—of, say, winding up in the top 10 percent—of someone born to a poor or less educated family would be the same as those of someone born to a rich, well-educated, and well-connected family." In other words, under this definition of equality of opportunity, people's earnings would be uncorrelated with those of their parents. Needless to say, in the data, that is not at all the case, which leads Stiglitz to conclude that we are falling short of providing equal opportunity.

Yet the issue cannot be settled so easily, because the intergenerational transmission of income has many causes beyond unequal opportunity. In particular, parents and children share genes, a fact that would lead to intergenerational persistence in income even in a world of equal opportunities. IQ, for example, has been widely studied, and it has a large degree of heritability. Smart parents are more likely to have smart children, and their greater intelligence will be reflected, on average, in higher incomes. Of course, IQ is only one dimension of talent, but it is easy to believe that other dimensions, such as self-control, ability to focus, and interpersonal skills, have a degree of genetic heritability as well.

This is not to say that we live in a world of genetic determinism, for surely we do not. But it would be a mistake to go to the other extreme and presume no genetic transmission of economic outcomes. A recent survey of the small but growing field of genoeconomics by Benjamin et al. (2012) reports, "Twin studies suggest that economic outcomes and preferences, once corrected for measurement error, appear to be about as heritable as many medical conditions and personality traits." Similarly, in his study of the life outcomes of adopted children, Sacerdote (2007) writes, "While educational attainment and income are frequently the focus of economic studies, these are among the outcomes least affected by differences in family environment." (He reports that family background exerts a stronger influence on social variables, such as drinking behavior.) This evidence suggests that it is implausible to interpret generational persistence in income as simply a failure of society to provide equal opportunities. Indeed, Sacerdote estimates (in his table 5) that while 33 percent of the variance of family income is explained by genetic heritability, only 11 percent is explained by the family environment. The remaining 56 percent includes environmental factors unrelated to family. If this 11 percent figure is approximately correct, it suggests that we are not far from a plausible definition of equality of opportunity—that is, being raised by the right family does give a person a leg up in life, but family environment accounts for only a small percentage of the variation in economic outcomes compared with genetic inheritance and environmental factors unrelated to family.

To the extent that our society deviates from the ideal of equality of opportunity, it is probably best to focus our attention on the left tail of the income distribution rather than on the right tail. Poverty entails a variety of socioeconomic maladies,

and it is easy to believe that children raised in such circumstances do not receive the right investments in human capital. By contrast, the educational and career opportunities available to children of the top 1 percent are, I believe, not very different from those available to the middle class. My view here is shaped by personal experience. I was raised in a middle-class family; neither of my parents were college graduates. My own children are being raised by parents with both more money and more education. Yet I do not see my children as having significantly better opportunities than I had at their age.

In the end, I am led to conclude that concern about income inequality, and especially growth in incomes of the top 1 percent, cannot be founded primarily on concern about inefficiency and inequality of opportunity. If the growing incomes of the rich are to be a focus of public policy, it must be because income inequality is a problem in and of itself.

The Big Tradeoff

In the title of his celebrated book, Arthur Okun (1975) told us that the "big tradeoff" that society faces is between equality and efficiency. We can use the government's system of taxes and transfers to move income from the rich to the poor, but that system is a "leaky bucket." Some of the money is lost as it is moved. This leak should not stop us from trying to redistribute, Okun argued, because we value equality. But because we are also concerned about efficiency, the leak will stop us before we fully equalize economic resources.

The formal framework that modern economists use to address this issue is that proposed by Mirrlees (1971). In the standard Mirrlees model, individuals get utility from consumption C and disutility from providing work effort L. They differ only according to their productivity W. In the absence of government redistribution, each person's consumption would be WL. Those with higher productivity would have higher consumption, higher utility, and lower marginal utility.

The government is then introduced as a benevolent social planner with the goal of maximizing total utility in society (or, sometimes, a more general social welfare function that could depend nonlinearly on individual utilities). The social planner wants to move economic resources from those with high productivity and low marginal utility to those with lower productivity and higher marginal utility. Yet this redistribution is hard to accomplish, because the government is assumed to be unable to observe productivity W; instead, it observes only income WL, the product of productivity and effort. If it redistributes income too much, high-productivity individuals will start to act as if they are low-productivity individuals. Public policymakers are thus forced to forgo the first-best egalitarian outcome for a second-best incentive-compatible solution. Like a government armed with Okun's leaky bucket, the Mirrleesian social planner redistributes to some degree but also allows some inequality to remain.

If this framework is adopted, then the debate over redistribution turns to questions about key parameters. In particular, optimal redistribution depends on the

degree to which work effort responds to incentives. If the supply of effort is completely inelastic, then the bucket has no leak, and the social planner can reach the egalitarian outcome. If the elasticity is small, the social planner can come close. But if work effort responds substantially to incentives, then the bucket is more like a sieve, and the social planner should attempt little or no redistribution. Thus, much debate among economists about optimal redistribution centers on the elasticity of labor supply.

Even if one is willing to accept the utilitarian premise of this framework, there is good reason to be suspicious of particular numerical results that follow from it. When researchers implement the Mirrlees model, they typically assume, as Mirrlees did, that all individuals have the same preferences. People are assumed to differ only in their productivity. For purposes of illustrative theory, that assumption is fine, but it is also false. Incomes differ in part because people have different tastes regarding consumption, leisure, and job attributes. Acknowledging variation in preferences weakens the case for redistribution (Lockwood and Weinzierl 2012). For example, many economics professors could have pursued higher-income career paths as business economists, software engineers, or corporate lawyers. That they chose to take some of their compensation in the form of personal and intellectual freedom rather than cold cash is a personal lifestyle choice, not a reflection of innate productivity. Those who made the opposite choice may have done so because they get greater utility from income. A utilitarian social planner will want to allocate greater income to these individuals, even apart from any incentive effects.

Another problem with the Mirrlees framework as typically implemented is that it takes a simplistic approach to tax incidence. Any good introductory student of economics knows that when a good or service is taxed, the buyer and seller share the burden. Yet in the Mirrlees framework, when an individual's labor income is taxed, only the seller of the services is worse off. In essence, the demand for labor services is assumed to be infinitely elastic. A more general set of assumptions would acknowledge that the burden of the tax is spread more broadly to buyers of those services (and perhaps to sellers of complementary inputs as well). In this more realistic setting, tax policy would be a less well-targeted tool for redistributing economic well-being.

The harder and perhaps deeper question is whether the government's policy toward redistribution is best viewed as being based on a benevolent social planner with utilitarian preferences. That is, did Okun and Mirrlees provide economists with the right starting point for thinking about this issue? I believe there are good reasons to doubt this model from the get-go.

The Uneasy Case for Utilitarianism

For economists, the utilitarian approach to income distribution comes naturally. After all, utilitarians and economists share an intellectual tradition: early utilitarians, such as John Stuart Mill, were also among the early economists. Also, utilitarianism seems to extend the economist's model of individual decision making

to the societal level. Indeed, once one adopts the political philosophy of utilitarianism, running a society becomes yet another problem of constrained optimization. Despite its natural appeal (to economists, at least), the utilitarian approach is fraught with problems.

One classic problem is the interpersonal comparability of utility. We can infer an individual's utility function from the choices that individual makes when facing varying prices and levels of income. But from this revealed-preference perspective, utility is not inherently measurable, and it is impossible to compare utilities across people. Perhaps advances in neuroscience will someday lead to an objective measure of happiness, but as of now, there is no scientific way to establish whether the marginal dollar consumed by one person produces more or less utility than the marginal dollar consumed by a neighbor.

Another more concrete problem is the geographic scope of the analysis. Usually, analyses of optimal income redistribution are conducted at the national level. But there is nothing inherent in utilitarianism that suggests such a limitation. Some of the largest income disparities are observed between nations. If a national system of taxes and transfers is designed to move resources from Palm Beach, Florida, to Detroit, Michigan, shouldn't a similar international system move resources from the United States and Western Europe to sub-Saharan Africa? Many economists do support increased foreign aid, but as far as I know, no one has proposed marginal tax rates on rich nations as high as the marginal tax rates imposed on rich individuals. Our reluctance to apply utilitarianism at the global level should give us pause when applying it at the national level.

In a 2010 paper, Matthew Weinzierl and I emphasized another reason to be wary of utilitarianism: it recommends a greater use of "tags" than most people feel comfortable with. As Akerlof (1978) pointed out, if the social planner can observe individual characteristics that are correlated with productivity, then an optimal tax system should use that information, in addition to income, in determining an individual's tax liability. The more the tax system is based on such fixed characteristics rather than income, the less it will distort incentives. Weinzierl and I showed that one such tag is height. Indeed, the correlation between height and wages is sufficiently strong that the optimal tax on height is quite large. Similarly, according to the utilitarian calculus, the tax system should also make a person's tax liability a function of race, gender, and perhaps many other exogenous characteristics. Of course, few people would embrace the idea of a height tax, and Weinzierl and I did not offer it as a serious policy proposal. Even fewer people would be comfortable with a race-based income tax (although Alesina, Ichin, and Karabarbounis, 2011, propose in earnest a gender-based tax). Yet these implications cannot just be ignored. If you take from a theory only the conclusions you like and discard the rest, you are using the theory as a drunkard uses a lamp post—for support rather than illumination. If utilitarianism takes policy in directions that most people don't like, then perhaps it is not a sound foundation for thinking about redistribution and public policy.

Finally, in thinking about whether the utilitarian model really captures our moral intuitions, it is worth thinking for a moment about the first-best outcome for

a utilitarian social planner. Suppose, in contrast to the Mirrlees model, the social planner could directly observe productivity. In this case, the planner would not need to worry about incentives, but could set taxes and transfers based directly on productivity. The optimal policy would equalize the marginal utility of consumption across individuals; if the utility function is assumed to be additively separable in consumption and leisure, this means everyone consumes the same amount. But because some people are more productive than others, equalizing leisure would not be optimal. Instead, the social planner would require more-productive individuals to work more. Thus, in the utilitarian first-best allocation, the more-productive members of society would work more and consume the same as everyone else. In other words, in the allocation that maximizes society's total utility, the less-productive individuals would enjoy a higher utility than the more productive.

Is this really the outcome we would want society to achieve if it could? A true utilitarian would follow the logic of the model and say "yes." Yet this outcome does not strike me as the ideal toward which we should aspire, and I suspect most people would agree. Even young children have an innate sense that merit should be rewarded (Kanngiesser and Warneken 2012)—and I doubt it is only because they are worried about the incentive effects of not doing so. If I am right, then we need a model of optimal government taxes and transfers that departs significantly from conventional utilitarian social planning.

Listening to the Left

In recent years, the left side of the political spectrum has focused much attention on the rising incomes of the top 1 percent. This includes President Obama's proposals to raises taxes on higher incomes, the Occupy Wall Street movement, and a rash of books about economic inequality. Even though I don't share the Left's policy conclusions, I find it is worthwhile to listen carefully to their arguments to discern what set of philosophical principles and empirical claims underlie their concerns.

It is, I believe, hard to square the rhetoric of the Left with the economist's standard framework. Someone favoring greater redistribution along the lines of Okun and Mirrlees would argue as follows: "The rich earn higher incomes because they contribute more to society than others do. However, because of diminishing marginal utility, they don't get much value from their last few dollars of consumption. So we should take some of their income away and give it to less-productive members of society. While this policy would cause the most productive members to work less, shrinking the size of the economic pie, that is a cost we should bear, to some degree, to increase utility for society's less-productive citizens."

Surely, that phrasing of the argument would not animate the Occupy crowd! So let's consider the case that the Left makes in favor of greater income redistribution. There are three broad classes of arguments.

The first is the suggestion that the tax system we now have is regressive. Most famously, during the presidential campaign of 2008, at a fund-raiser for Hillary

Clinton, the billionaire investor Warren E. Buffett said that the rich were not paying enough. Mr. Buffett used himself as an example. He asserted that his taxes in the previous year equaled only 17.7 percent of his taxable income, while his receptionist paid about 30 percent of her income in taxes (Tse 2007). In 2011, President Obama proposed the "Buffett rule," which would require taxpayers with income over a million dollars to pay at least 30 percent of their income in federal income taxes.

There are, however, good reasons to be skeptical of Buffett's calculations. If his receptionist was truly a middle-income taxpayer, then to get her tax rate to 30 percent, he most likely added the payroll tax to the income tax. Fair enough. But for Buffett's tax rate to be only 17.7 percent, most of his income was likely dividends and capital gains, and his calculation had to ignore the fact that this capital income was already taxed at the corporate level. A complete accounting requires aggregating not only all taxes on labor income but also all taxes on capital income.

The Congressional Budget Office (2012) does precisely that when it calculates the distribution of the federal tax burden—and it paints a very different picture than did Buffett's anecdote. In 2009, the most recent year available, the poorest fifth of the population, with average annual income of $23,500, paid only 1.0 percent of its income in federal taxes. The middle fifth, with income of $64,300, paid 11.1 percent. And the top fifth, with income of $223,500, paid 23.2 percent. The richest 1 percent, with an average income of $1,219,700, paid 28.9 percent of its income to the federal government. To be sure, some taxpayers aggressively plan to minimize taxes, and this may result in some individual cases where those with high incomes pay relatively little in federal taxes. But the CBO data make clear that these cases are the exceptions. As a general rule, the existing federal tax code is highly progressive.

A second type of argument from the Left is that the incomes of the rich do not reflect their contributions to society. In the standard competitive labor market, a person's earnings equal the value of his or her marginal productivity. But there are various reasons that real life might deviate from this classical benchmark. If, for example, a person's high income results from political rent-seeking rather than producing a valuable product, the outcome is likely to be both inefficient and widely viewed as inequitable. Steve Jobs getting rich from producing the iPod and Pixar movies does not produce much ire among the public. A Wall Street executive benefiting from a taxpayer-financed bailout does.

The key issue is the extent to which the high incomes of the top 1 percent reflect high productivity rather than some market imperfection. This question is one of positive economics, but unfortunately not one that is easily answered. My own reading of the evidence is that most of the very wealthy get that way by making substantial economic contributions, not by gaming the system or taking advantage of some market failure or the political process. Take the example of pay for chief executive officers. Without doubt, CEOs are paid handsomely, and their pay has grown over time relative to that of the average worker. Commentators on this phenomenon sometimes suggest that this high pay reflects the failure of corporate boards of directors to do their job. Rather than representing shareholders, the argument goes, boards are too cozy with the CEOs and pay them more than they are worth to their organizations.

Yet this argument fails to explain the behavior of closely-held corporations. A private equity group with a controlling interest in a firm does not face the alleged principal–agent problem between shareholders and boards, and yet these closely-held firms also pay their CEOs handsomely. Indeed, Kaplan (2012) reports that over the past three decades, executive pay in closely-held firms has outpaced that in public companies. Cronqvist and Fahlenbrach (2013) find that when public companies go private, the CEOs tend to get paid more rather than less in both base salaries and bonuses. In light of these facts, the most natural explanation of high CEO pay is that the value of a good CEO is extraordinarily high (a conclusion that, incidentally, is consistent with the model of CEO pay proposed by Gabaix and Landier 2008).

A third argument that the Left uses to advocate greater taxation of those with higher incomes is that the rich benefit from the physical, legal, and social infrastructure that government provides and, therefore, should contribute to supporting it. As one prominent example, President Obama (2012) said in a speech, "If you were successful, somebody along the line gave you some help. There was a great teacher somewhere in your life. Somebody helped to create this unbelievable American system that we have that allowed you to thrive. Somebody invested in roads and bridges. If you've got a business—you didn't build that. Somebody else made that happen. The Internet didn't get invented on its own. Government research created the Internet so that all the companies could make money off the Internet. The point is that when we succeed, we succeed because of our individual initiative, but also because we do things together."

In the language of traditional public finance, President Obama was relying less on the ability-to-pay principle and more on the benefits principle. That is, higher taxation of the rich is not being justified by the argument that their marginal utility of consumption is low, as it is in the frameworks of Okun and Mirrlees. Rather, higher taxation is being justified by the claim that the rich achieved their wealth in large measure because of the goods and services the government provides and therefore have a responsibility to finance those goods and services.

This line of argument raises the empirical question of how large the benefit of government infrastructure is. The average value is surely very high, as lawless anarchy would leave the rich (as well as most everyone else) much worse off. But like other inputs into the production process, government infrastructure should be valued at the margin, where the valuation is harder to discern. As I pointed out earlier, the average person in the top 1 percent pays more than one-quarter of income in federal taxes, and about one-third if state and local taxes are included. Why isn't that enough to compensate for the value of government infrastructure?

A relevant fact here is that, over time, an increasing share of government spending has been for transfer payments, rather than for purchases of goods and services. Government has grown as a percentage of the economy not because it is providing more and better roads, more and better legal institutions, and more and better educational systems. Rather, government has increasingly used its power to tax to take from Peter to pay Paul. Discussions of the benefits of government services should not distract from this fundamental truth.

In the end, the Left's arguments for increased redistribution are valid in principle but dubious in practice. If the current tax system were regressive, or if the incomes of the top 1 percent were much greater than their economic contributions, or if the rich enjoyed government services in excess of what they pay in taxes, then the case for increasing the top tax rate would indeed be strong. But there is no compelling reason to believe that any of these premises holds true.

The Need for an Alternative Philosophical Framework

A common thought experiment used to motivate income redistribution is to imagine a situation in which individuals are in an "original position" behind a "veil of ignorance" (as in Rawls 1971). This original position occurs in a hypothetical time before we are born, without the knowledge of whether we will be lucky or unlucky, talented or less talented, rich or poor. A risk-averse person in such a position would want to buy insurance against the possibility of being born into a less-fortunate station in life. In this view, governmental income redistribution is an enforcement of the social insurance contract to which people would have voluntarily agreed in this original position.

Yet take this logic a bit further. In this original position, people would be concerned about more than being born rich or poor. They would also be concerned about health outcomes. Consider kidneys, for example. Most people walk around with two healthy kidneys, one of which they do not need. A few people get kidney disease that leaves them without a functioning kidney, a condition that often cuts life short. A person in the original position would surely sign an insurance contract that guarantees him at least one working kidney. That is, he would be willing to risk being a kidney donor if he is lucky, in exchange for the assurance of being a transplant recipient if he is unlucky. Thus, the same logic of social insurance that justifies income redistribution similarly justifies government-mandated kidney donation.

No doubt, if such a policy were ever seriously considered, most people would oppose it. A person has a right to his own organs, they would argue, and a thought experiment about an original position behind a veil of ignorance does not vitiate that right. But if that is the case, and I believe it is, it undermines the thought experiment more generally. If imagining a hypothetical social insurance contract signed in an original position does not supersede the right of a person to his own organs, why should it supersede the right of a person to the fruits of his own labor?

An alternative to the social insurance view of the income distribution is what, in Mankiw (2010), I called a "just deserts" perspective. According to this view, people should receive compensation congruent with their contributions. If the economy were described by a classical competitive equilibrium without any externalities or public goods, then every individual would earn the value of his or her own marginal product, and there would be no need for government to alter the resulting income distribution. The role of government arises as the economy departs from this classical benchmark. Pigovian taxes and subsidies are necessary to correct externalities,

and progressive income taxes can be justified to finance public goods based on the benefits principle. Transfer payments to the poor have a role as well, because fighting poverty can be viewed as a public good (Thurow 1971).

This alternative perspective on the income distribution is a radical departure from the utilitarian perspective that has long influenced economists, including Okun and Mirrlees. But it is not entirely new. It harkens back about a century to the tradition of "just taxation" suggested by Knut Wicksell (1896, translated 1958) and Erik Lindahl (1919, translated 1958). More important, I believe it is more consistent with our innate moral intuitions. Indeed, many of the arguments of the Left discussed earlier are easier to reconcile with the just-deserts theory than they are with utilitarianism. My disagreement with the Left lies not in the nature of their arguments, but rather in the factual basis for their conclusions.

The political philosophy one adopts naturally influences the kind of economic questions that are relevant for determining optimal policy. The utilitarian perspective leads to questions such as: How rapidly does marginal utility of consumption decline? What is the distribution of productivity? How much do taxes influence work effort? The just-deserts perspective focuses instead on other questions: Do the high incomes of the top 1 percent reflect extraordinary productivity, or some type of market failure? How are the benefits of public goods distributed across the income distribution? I have my own conjectures about the answers to these latter questions, and I have suggested them throughout this essay, but I am the first to admit that they are tentative. Fortunately, these are positive questions to which future economic research may provide more definitive answers.

To highlight the difference between these approaches, consider how each would address the issue of the top tax rate. In particular, why shouldn't we raise the rate on high incomes to 75 percent, as France's President Hollande has recently proposed, or to 91 percent, where it was through much of the 1950s in the United States? A utilitarian social planner would say that perhaps we should and would refrain from doing so only if the adverse incentive effects were too great. From the just-deserts perspective, such confiscatory tax rates are wrong, even ignoring any incentive effects. By this view, using the force of government to seize such a large share of the fruits of someone else's labor is unjust, even if the taking is sanctioned by a majority of the citizenry.

In the final analysis, we should not be surprised when opinions about income redistribution vary. Economists can turn to empirical methods to estimate key parameters, but no amount of applied econometrics can bridge this philosophical divide. I hope my ruminations in this essay have convinced some readers to see the situation from a new angle. But at the very least, I trust that these thoughts offer a vivid reminder that fundamentally normative conclusions cannot rest on positive economics alone.

■ *I am grateful to David Autor, Nathaniel Hilger, Chang-Tai Hseih, Steven Kaplan, Ulrike Malmendier, Deborah Mankiw, Nicholas Mankiw, Lisa Mogilanski, Alexander Sareyan, Lawrence Summers, Timothy Taylor, Jane Tufts, and Matthew Weinzierl for helpful comments and discussion.*

References

Akerlof, George A. 1978. "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." *American Economic Review* 68(1): 8–19.

Alesina, Alberto, Andrea Ichin, and Loukas Karabarbounis, 2011. "Gender-Based Taxation and the Division of Family Chores." *American Economic Journal: Economic Policy* 3(2): 1–40.

Benjamin, Daniel J., David Cesarini, Christopher F. Chabris, Edward L. Glaeser, David I. Laibson, Vilmundur Guðnason, Tamara B. Harris, Lenore J. Launer, Shaun Purcell, Albert Vernon Smith, Magnus Johannesson, Patrik K. E. Magnusson, Jonathan P. Beauchamp, Nicholas A. Christakis, Craig S. Atwood, Benjamin Hebert, Jeremy Freese, Robert M. Hauser, Taissa S. Hauser, Alexander Grankvist, Christina M. Hultman, and Paul Lichtenstein. 2012. "The Promises and Pitfalls of Genoeconomics." *Annual Review of Economics* vol. 4, pp. 627–62.

Brynjolfsson, Erik, and Andrew McAfee. 2011. *Race Against the Machine*. Lexington, MA: Digital Frontier Press.

Congressional Budget Office. 2012. "The Distribution of Household Income and Federal Taxes, 2008 and 2009." July. Available at: http://www.cbo.gov/publication/43373.

Cronqvist, Henrik, and Rüdiger Frahlenbrach. 2013. "CEO Contract Design: How Do Strong Principals Do It?" *Journal of Financial Economics* 108(3): 659–74.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123(1): 49–100.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology*. Cambridge, MA: Harvard University Press.

Kanngiesser Patricia, and Felix Warneken. 2012. "Young Children Consider Merit when Sharing Resources with Others." *PLoS ONE* 7(8): e43979.

Kaplan, Steven N. 2012. "Executive Compensation and Corporate Governance in the U.S.: Perceptions, Facts, and Challenges." NBER Working Paper 18395.

Lindahl, Erik. 1958. "Just Taxation—A Positive Solution." In *Classics in the Theory of Public Finance*, edited by Richard Musgrave and Alan Peacock, 98–123. London: Macmillan.

Lockwood, Benjamin, and Matthew Weinzierl. 2012. "*De Gustibus non est Taxandum*: Theory and Evidence on Preference Heterogeneity and Redistribution." Harvard Business School Working Paper 12-063.

Mankiw, N. Gregory. 2010. "Spreading the Wealth Around: Reflections Inspired by Joe the Plumber." *Eastern Economic Journal* 36(3): 285–98.

Mankiw, N. Gregory, and Matthew Weinzierl. 2010. "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution." *American Economic Journal: Economic Policy* 2(1): 155–76.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimal Income Taxation." *Review of Economic Studies* 38(2): 175–208.

Obama, Barack. 2012. "Remarks by the President at a Campaign Event in Roanoke, Virginia." July 13. http://www.whitehouse.gov/the-press-office/2012/07/13/remarks-president-campaign-event-roanoke-virginia.

Okun, Arthur. 1975. *Equality and Efficiency: The Big Tradeoff*. Brookings Institution.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1): 1–39. (Tables and figures updated to 2011 at Saez's website: http://elsa.berkeley.edu/~saez/.)

Rawls, John. 1971. *A Theory of Justice*. Belknap Press.

Sacerdote, Bruce. 2007. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *Quarterly Journal of Economics* 122 (1): 119–157.

Stiglitz, Joseph E. 2012. *The Price of Inequality: How Today's Divided Society Endangers Our Future*. W. W. Norton and Company.

Thurow, Lester C. 1971. "The Income Distribution as a Pure Public Good." *Quarterly Journal of Economics* 85(2): 327–36.

Tse, Tomoeh Murakami. 2007. "Buffett Slams Tax System Disparities." *Washington Post*, June 27.

Wicksell, Knut. 1958. "A New Principle of Just Taxation." In *Classics in the Theory of Public Finance*, edited by Richard Musgrave and Alan Peacock, 72–118. London: Macmillan.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 35–56

It's the Market: The Broad-Based Rise in the Return to Top Talent[†]

Steven N. Kaplan and Joshua Rauh

Back in 1981, when the rise in inequality over recent decades was barely underway, Sherwin Rosen (1981) wrote a prescient article on "The Economics of Superstars." Rosen argued that technological change, particularly in information and communications, can increase the relative productivity of highly talented individuals, or "superstars." Essentially, such superstars become able to manage or to perform on a larger scale, applying their talent to greater pools of resources and reaching larger numbers of people. Those who are able to do so receive higher compensation.

Of course, other explanations of the rise in inequality have been offered, including arguments that managerial power has increased in a way that allows those at the top to receive higher pay (Bebchuk and Fried 2004), that social norms against higher pay levels have broken down (Piketty and Saez 2006), and that tax policy affects the distribution of surpluses between employers and employees (Piketty, Saez, and Stantcheva forthcoming). This paper offers some evidence bearing on these disputes. We first look at differences in occupations in the United States across those with the highest income levels. The increase in pay at the highest income levels is broad-based; for example, it is not primarily or solely a phenomenon of

■ *Steven N. Kaplan is Neubauer Family Distinguished Service Professor of Entrepreneurship and Finance, University of Chicago Booth School of Business, Chicago, Illinois. Joshua Rauh is Professor of Finance, Stanford Graduate School of Business, Stanford, California. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are skaplan@uchicago.edu and Rauh_Joshua@gsb.stanford.edu.*

[†] To access the disclosure statements, visit
http://dx.doi.org/10.1257/jep.27.3.35

doi=10.1257/jep.27.3.35

publicly traded companies. We also discuss some evidence on the income share of the top 1 percent over time.

We then turn to evidence on inequality of wealth at the top. In looking at the wealthiest Americans, those in the Forbes 400 are less likely to have inherited their wealth or to have grown up wealthy. The Forbes 400 of today also are those who were able to access education while young and apply their skills to the most scalable industries: technology, finance, and mass retail.

We conclude by analyzing which of the different theories are more consistent with the patterns in the data. We believe that the US evidence on income and wealth shares for the top 1 percent is most consistent with a "superstar"-style explanation rooted in the importance of scale and skill-biased technological change. In particular, we interpret the fact that the top 1 percent is spread broadly across a variety of occupations as most consistent with an important role for skill-biased technological change and increased scale. These facts are less consistent with an argument that the gains to the top 1 percent are rooted in greater managerial power or changes in social norms about what managers should earn.

Income Inequality at the Very Top

The increase in pay in the US economy at the very top is broad-based. It is not solely a phenomenon of one sector, but has occurred for public company executives, private company executives, financial executives, corporate lawyers, and professional athletes. In this section, we attempt to understand what professions and forces have driven those increases, using and building on earlier results from Kaplan and Rauh (2010), Bakija, Cole, and Heim (2012), and Kaplan (2012).

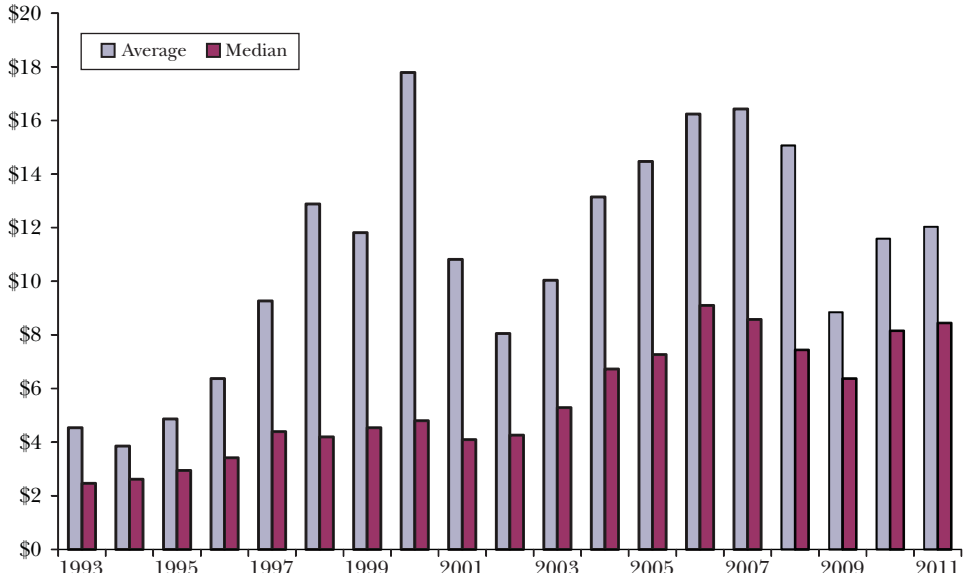
Evidence on Top Incomes across Professions

In this section, we report evidence on the increase in pay over time of public company executives, private company executives, hedge fund and private equity investors, Wall Street employees, lawyers, and professional athletes. All of these groups have experienced marked increases in pay over the last 20 to 30 years.

We update Kaplan (2012) by reporting time series information on the pay of US chief executive officers from 1993 to 2011. This analysis tracks the pay of CEOs of Standard and Poor's 500 companies from Standard and Poor's ExecuComp database. These are the largest publicly traded US companies. These data also include pay for CEOs of the other companies covered by ExecuComp—companies that at one time have been in Standard & Poor's 1500 index—but are not in the S&P 500.

Figure 1 shows "realized" pay—that is, in each year it includes the value of the stock options the executive exercised that year, all in millions of 2010 dollars to remove the effects of inflation. Average realized pay for chief executive officers rose sharply in the 1990s, peaked in 2000, dipped by more than 50 percent by 2002, rebounded close to 2000 levels by 2007, dipped markedly again in 2009 and rebounded somewhat in 2010–2011. Average pay in 2011, at $12 million, is

Figure 1
Average and Median Total Pay (Realized) of S&P 500 Chief Executive Officers from 1993 to 2011
(millions of 2010 dollars)

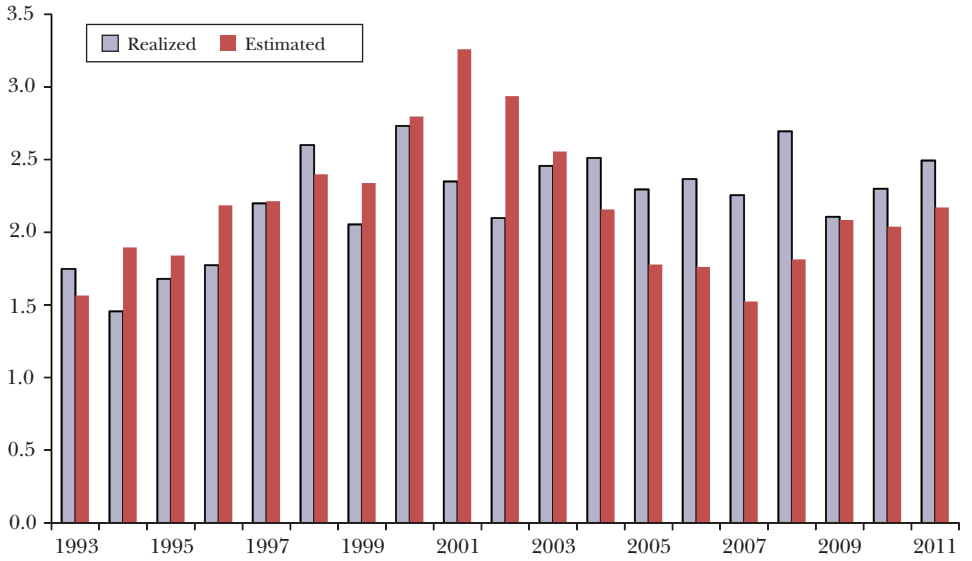


Source: We update Kaplan (2012) by reporting time series information from Standard and Poor's ExecuComp database.
Note: Figure 1 shows "realized" pay—that is, including the value of the stock options the executive exercised each year, in millions of 2010 dollars.

32 percent below its peak in 2000. Median pay fluctuates less, and has a generally upward trend over this period. It peaked in 2006 at a value of just over $9 million (again, expressed in 2010 dollars). The increase in the median is the result of an increased use of restricted stock, rather than stock options, as a compensation tool. The patterns for CEOs of firms outside the Standard and Poor's 500 are qualitatively similar.

In Kaplan and Rauh (2010), we compared how well-off chief executive officers and other top executives were relative to other top earners. Figure 2 updates this analysis by comparing the average pay of CEOs of the Standard and Poor's 500 to the average income of US households in the top 0.1 percent from 1993 to 2011. Here we show two different measures: realized pay (as discussed above), and estimated pay, which is based on the financial value of stock options when they are granted as opposed to when they are realized. In 2011, the adjusted gross income cutoff for the top 0.1 percent was $1.7 million; the average adjusted gross for taxpayers in the top 0.1 percent was $5.0 million. Because there were roughly 140,000 such taxpayers in this group in 2011, the 500 CEOs have only a minimal effect on the average gross income of this group. As the figure shows, actual pay for the CEOs of

Figure 2
Average Pay (Estimated and Realized) of S&P 500 Chief Executive Officers Relative to Average Adjusted Gross Income of Top 0.1% of Taxpayers from 1993 to 2011



Source: We update Kaplan and Rauh (2010), using data from ExecuComp and Piketty and Saez (2013).
Note: Realized pay includes the financial value of stock options when they are realized; estimated pay includes the financial value of stock options when they are granted.

Standard and Poor's 500 companies has been stable since 1997 at 2 to 2.75 times the average pay of the top 0.1 percent, with the latest figure at 2.5 times and with the peaks in 1998, 2000, and 2008. From 1993 to 1996, realized pay was somewhat lower, at roughly 1.75 times that of top 0.1 percent taxpayers. In other words, CEOs of Standard & Poor's 500 firms have seen little change in their realized pay relative to top 0.1 percent taxpayers since the late 1990s. For CEOs of companies outside the Standard and Poor's 500, realized pay relative to average income for top 0.1 percent taxpayers has remained roughly constant since the early 1990s.

One of the arguments sometimes heard about the pay of top executives of public companies is that these executives have too much power to shape a compliant board of directors and thus to gain power over their own pay. In evaluating the importance of this argument, it is natural to look at other groups of highly paid executives who operate under different corporate governance structures and have their pay determined in different ways.

One useful set of comparisons comes from Bakija, Cole, and Heim (2012), who study IRS tax return data for a number of years between 1979 and 2005. The advantage of the IRS data is that it by definition incorporates the pay of executives from all companies, both publicly traded and privately held. If the governance structure of

public companies has played a major role in rising income inequality, then we should see pay rising more for chief executive officers of public companies than for those of private companies. Public companies are potentially vulnerable to agency and managerial power problems, while private or closely-held companies are, by definition, controlled by large shareholders or the executives themselves, and therefore should face much more limited agency problems. The pay of executives of most privately-held companies is generally not publicly disclosed unlike the pay of public company executives. So if the loosening of social norms against higher pay levels have played a strong role, then we should see pay rising more for public company CEOs. While loosening social norms might lead to an increase in the pay of public company executives, it should have no effect on private company executives because their pay has not been and still is not visible to the general public.

Although the IRS data cannot definitively identify whether the executives in the sample work for private or public companies, nor whether executives are chief executive officers, the data allow for some informative approximations. Bakija, Cole, and Heim (2012) compare executives who receive the majority of their income in salary and wages to those executives who receive the majority of their income from self-employment, partnership, and S-corporation related income. They argue that the former are more likely to include public company executives, while the latter are most likely to include executives of closely-held business. Their data show that the pay of closely-held business executives has risen substantially as a share of the top 0.1 percent. Pay of closely-held business executives comprised 22 percent of the top 0.1 percent in 2005, up from 18 percent in 1993 and 9 percent in 1979. In contrast, the pay of salaried executives, whom they argue are more likely to be with public companies, has actually fallen relative to the top 0.1 percent. Pay of salaried executives comprised 20 percent of the top 0.1 percent in 2005, down from 28 percent in 1993 and 38 percent in 1979.

This evidence is not supportive of the arguments that the top incomes have been driven by managerial power or poor corporate governance in public companies. Public company executives, who should be more subject to problems of managerial power problems, saw their pay and relative standing increase less over this period than executives of closely-held company businesses that are, by definition, controlled by large shareholders or the executives themselves and are subject to more limited agency problems. Furthermore, the Bakija, Cole, and Heim (2012) findings are not consistent with loosening social norms being an important factor in the increase in incomes at the top, as it is the pay of closely-held businesses—where executive pay is private and undisclosed—that increased the most.

Finance professionals are another potentially useful comparison group. In Kaplan and Rauh (2010), we find that when looking at those with top incomes, finance industry executives are at least as important as public company executives. In the Bakija, Cole, and Heim (2012) data, taxpayers in finance increased from 0.4 percent of the top 0.1 percent in 1979 to 2.0 percent in 2005, or a factor of five times. As discussed in a symposium in the Spring 2013 issue of this journal, the financial sector markedly increased its share of US GDP in the last years.

Table 1
Average Pay of Top Hedge Fund Managers and Law Partners

Year	***Average pay of top 25 hedge fund managers*** *(Millions of $2010 [and relative to average adjusted gross income of top 0.1%])*	***Average profit per partner at top 50 law firms*** *(Millions of $2010 [and relative to average adjusted gross income of top 0.1%])*
1994		$0.704 [0.268]
1996		$0.784 [0.219]
1998		$0.997 [0.200]
2000		$1.084 [0.167]
2002	$133.7 [34.6x]	$1.099 [0.285]
2004	$289.5 [55.7x]	$1.286 [0.247]
2006	$616.2 [90.3x]	$1.491 [0.218]
2008	$469.8 [82.1x]	$1.449 [0.253]
2010	$882.8 [177.6x]	$1.557 [0.313]
2012	$537.2 [115.7x]	

Source: Authors using data from *Absolute Return + Alpha* magazine, *American Lawyer*, and Piketty and Saez (2013).

We can also consider data for financial professionals working in the specific areas of hedge funds, private equity funds, and venture capital funds. Since 2001, *Absolute Return + Alpha* magazine has published an annual "Rich List" of the 25 highest paid hedge fund managers. These estimates include the annual income of these managers from fees and their capital invested in their funds. As a result, these data overstate the income of these managers attributable to their employment per se, as separate from their investment income. Nevertheless, the results are informative.

As shown in the first column of Table 1, the average pay of these top 25 hedge fund managers was $134 million in 2002, and it was $537 million in 2012 (all in 2010 dollars), standing at almost 116 times the pay of the average adjusted gross income of the top 0.1 percent. In fact, in each year since 2004, the 25 highest paid hedge fund managers have earned more than all of the chief executive officers of the Standard and Poor's 500 companies combined. In 2010, the 25 top hedge fund managers combined earned roughly four times as much as the CEOs of Standard and Poor's 500 firms.

In Kaplan and Rauh (2010), we also document a large increase in fees to private equity and venture capital investors through 2005. The payments to all but a few of the individuals within these firms are not in the public record. However, the fees received by these firms can be inferred by looking at their assets under management and their profits. We start with the assumption that private equity and venture capital investors earn fees on capital committed to their firms over the previous seven years. Capital raised or committed is obtained from Private Equity Analyst. We assume that the private equity firms earn an annual management fee of 1.5 percent on that capital, while venture capital firms earn a 2.0 percent management fee. In addition, we assume that private equity and venture capital firms receive 20 percent of the profits earned by funds in a given year. Profits are estimated using the average

return earned by private equity and venture capital funds in a given year reported by Cambridge Associates.

Using this approach, we estimate that the fees for private equity firms have increased substantially over time. Back in the late 1980s, the private equity industry was relatively small, with $1 to $2 billion in annual fees. By 1996–98, it was consistently in the range of $10 billion per year in fees. Private equity fees then spiked in 1999 and 2000 at about $20 billion per year, before falling back in 2001 and 2002. Private equity funding and fees picked up substantially after that. From 2005 to 2011, private equity fees averaged roughly $34 billion per year (in 2010 dollars). This represents an increase of almost three times the average over the previous ten years.[1]

Consistent with the large increase in fees, the founders of the large private equity firms earned very large amounts in 2012. The three founders of Carlyle received a total of over $300 million on distributions. The founder of Blackstone received over $200 million; and the two founders of KKR received more than $130 million each.

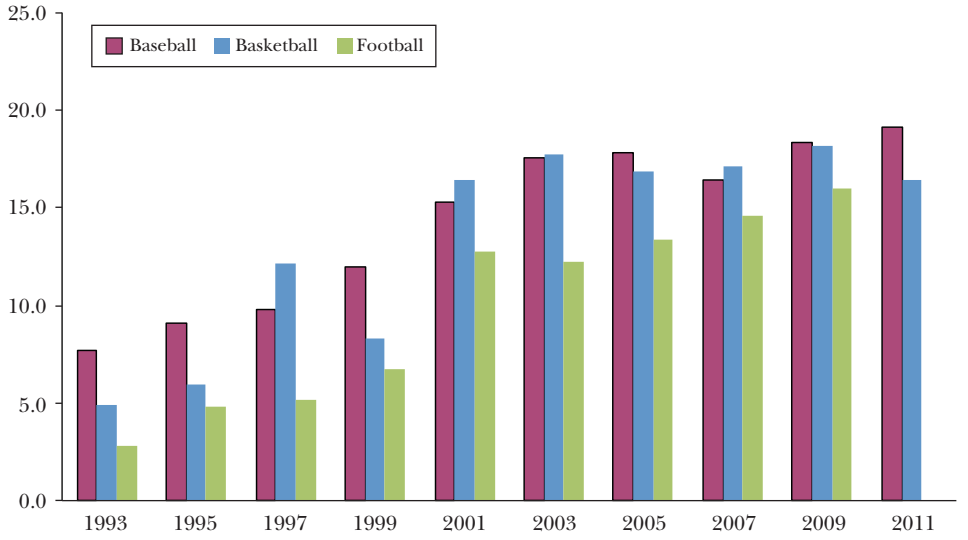
Venture capital firms have had a more volatile record. These firms typically earned fees of $1–$4 billion through the 1990s, until the Internet boom at the turn of the century, when their fees spiked to almost $30 billion in 1999 and more than $70 billion in 2000. Venture capital fees then dropped back to $4–$5 billion per year from 2001 to 2004. However, from 2004 to 2007 and, after the financial crisis years, again in 2011, venture capital fees exceeded $10 billion per year. Nevertheless, both analyses suggest that fees have increased markedly since the early 1990s.[2]

Lawyers at top law firms also have done very well. Because the law firms are partnerships and their fees are negotiated in an arms-length manner with clients, the pay of partners at those law firms is less subject to concerns that managerial power is leading to the increase in pay. The second column of Table 1 reports average profit per partner at the 50 top law firms from *American Lawyer* surveys from 1994 to 2010. This calculation measures the total partner profits at all of these firms divided by the total number of partners. The average profit per partner provides an estimate of the average partner's average gross income earned from employment at the law firm. The average profit per partner increased from $0.7 million in 1994 to almost $1.6 million in 2010 (in 2010 dollars). The numbers in brackets show that the average partner's income increased from 0.27 to 0.31 of the average income of the top 0.1 percent. These numbers may understate how well the very top partners

[1] In an alternative analysis, we assume that private equity firms earn overall fees of 4 percent while venture capital firms earn fees of 5 percent on capital raised over the previous seven years. This assumes that the profit share has a value of roughly 2.5 percent per year for private equity firms and 3 percent per year for venture capital firms. Those assumptions are consistent with the fee estimates in Metrick and Yasuda (2010). Under this methodology, private equity firm fees have averaged $26 billion per year since 2005.

[2] Investment banking firms offer another possible comparison group. In Kaplan and Rauh (2010), we use financial statements of publicly traded investment banking firms and assumptions of the pay distribution within these firms to estimate the pay of the most highly compensated people. We find that investment bankers constitute a meaningful fraction of the top 0.1 percent. Bell and Van Reenen (2010) use income data from the United Kingdom to study financial sector incomes there. They find that the pay of bankers' increased substantially since 1999 and contributed meaningfully to the increase in income inequality.

Figure 3
Average Top 25 Salaries in Professional Baseball, Basketball, and Football
(in millions of 2010 dollars)



Source: USA Today Sports Salary Database.

at these law firms have done. The number of partners increased over this period from 7,000 to 12,000. If it were possible to compare the pay of the top 7,000 partners in 2010 and 1994, the increase for top lawyers would be greater.

As a final comparison, we also consider pay for professional athletes in baseball, basketball, and football. We extend the analysis in Kaplan and Rauh (2010) by looking at the average pay of the top 25 most highly paid athletes in those sports (every other year) from 1993 to 2011. Figure 3 shows that the top 25 athletes in each sport average over $15 million per year in salary. Compared to Table 1, we see that these pay levels are higher than the average of top law partners (in the $1–2 million range) but lower than those of top hedge fund managers (in the hundreds of millions of dollars). This graph understates the total income of the athletes because it does not include nonsalary income such as endorsements. Pay at the top has increased markedly since 1993, with baseball, basketball, and football players earning, respectively 2.5, 3.3, and 5.8 times as much in 2009 as in 1993. The greater part of the increase, particularly for baseball and basketball, occurred by the early 2000s.

Again, the key reason for making these comparisons is that the mechanisms for determining compensation vary quite a bit across executives of public companies, executives of private companies, financial professionals, law firms, and professional athletes. If the reason for growth of incomes at the very top is, say, managerial power in publicly owned companies, then one would expect the increases in income at the

top levels to be much larger for that group. But the breadth of the occupations that have seen a rise in top income levels is much more consistent with the argument that the increase in "superstar" pay (or pay at the top) has been driven by the growth of information and communications technology, and the ways this technology allows individuals with particular skills that are in high demand to expand the scale of their performance.

Time Series Evidence on Income Inequality

One of the most widely cited measures of income inequality at the very top is the share of taxable income accruing to the top 1 percent of earners in the economy as tabulated by Piketty and Saez (2003, 2013). As of 2011, the top 1 percent of earners received 19.8 percent of all "market income" (labor income, business income, capital income, capital gains). This measure peaked at 23.9 percent in 1928, fell below 10 percent in the late 1960s and early 1970s, then increased markedly from 1980 to 2000. Since 2000, the share of the top 1 percent has fluctuated, peaking at 23.5 percent in 2007, dipping to 18.1 percent in 2009, and increasing to 19.8 percent in 2011, roughly the same as it was in 1998 and 1999. These measures of course do not include taxes or government transfers. The Congressional Budget Office (2012) has presented shares of the top 1 percent after transfers and before taxes, and after both transfers and taxes. In Figure 4, we present all three series since 1979, which is the first year in the Congressional Budget Office time series and also a time when the Piketty and Saez measure of the income of the top 1 percent is near its low point. Much can be said about this data, but here, we emphasize three points.

First, while the share of income going to the top 1 percent has clearly increased by any of these measures, the long-term extent of the increase is not best judged by looking at the peaks in 1999 or 2007.

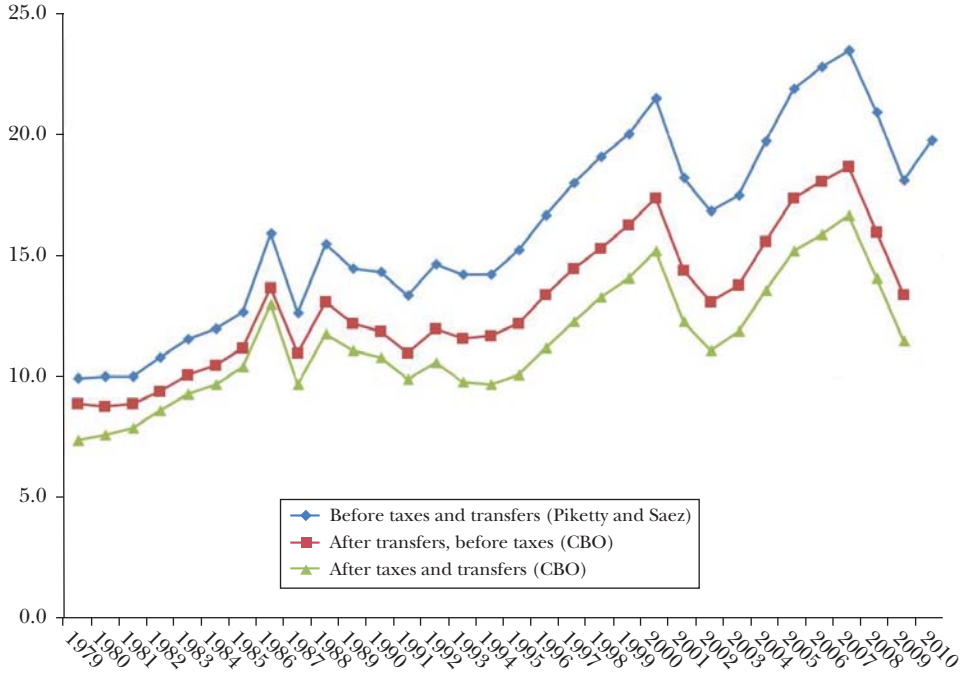
Second, the share of income going to the top 1 percent is quite volatile. In particular, it seems to rise during financial markets booms and then drop back.

Third, adding transfers and taxes to the picture and using the Congressional Budget Office data, transfer and tax policy appear to be having some effect in restraining the growth of inequality. In the most recent data from 2009, the after-tax, after-transfer income share of the top 1 percent was around the same level as in 1987–1988, 1996, and 2001.

Background of the Wealthiest Americans

In this section, we look at the Forbes 400 data to gather data on family background. The Forbes 400 is a list of the 400 wealthiest individuals in the United States by net worth. It has been published annually since 1982. We collected the lists in 1982, 1992, 2001, and 2011, approximately one year each decade, to get an overview. In constant 2011 dollars, the Forbes 400 wealth amounted to $214 billion in 1982, $483 billion in 1992, $1.197 trillion in 2001, and $1.525 trillion in 2011. As

Figure 4
Share of Income Accruing to the Top 1 Percent



Source: Picketty and Saez (2013) and Congressional Budget Office (2012).

a fraction of total US stock market value, this wealth amounted to, respectively, 7.0, 7.0, 7.0, and 8.5 percent.

In Kaplan and Rauh (2013), we provide a more detailed description of the list. This section draws on that research and deepens the analysis, particularly in comparing the Forbes 400 in the US to the list of international billionaires, also from *Forbes* magazine.

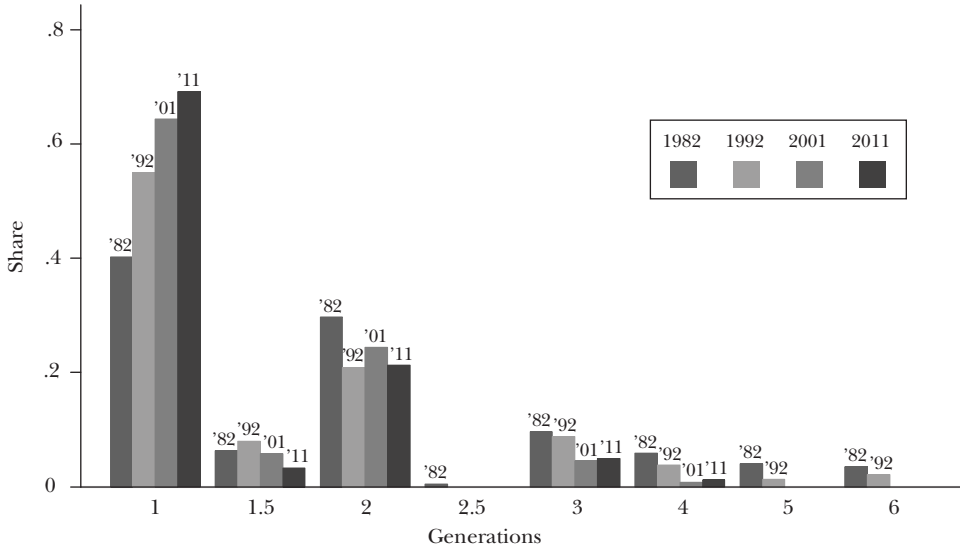
For each individual in the domestic and international datasets, we collect and code biographical details, using *Who's Who* and Internet searches as our primary sources. We identify the founding date of the business that generated the individual's wealth and then determine in what generation the individual is within the family of the founder of that business. The normal coding for the generation is an integer. However, if the individual inherited a modest business and built it into a much larger one, we input the generation as 1.5. For example, Edward C. Johnson III, of Fidelity Investments, receives a 1.5 in our dataset. Fidelity Investments was founded by the father of Johnson III but was built primarily by Johnson III himself.

We separately code the extent to which the individual grew up wealthy, defining three categories: little or no wealth in the family, some wealth in the family, or wealthy. For example, the Koch brothers grew up wealthy. Bill Gates, whose father

Figure 5

Generation of the Wealth-Creating Businesses of Forbes 400 Individuals in 1982, 1992, 2001, and 2011

(share among Forbes 400 individuals)

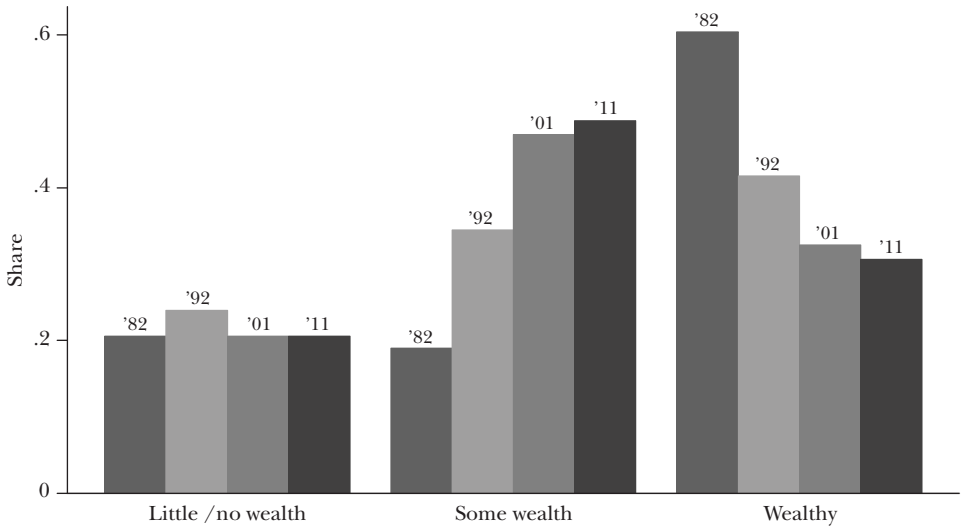


Source: Authors, using *Who's Who* and Internet searches as primary sources.
Notes: The numbers along the *x*-axis give the generation of a Forbes 400 individual within the family of the founder of his or her business. The normal coding for the generation is an integer. However, if the individual inherited a modest business and built it into a much larger one, we input the generation as 1.5.

cofounded a successful law firm, grew up with some wealth, as did, for example, sons and daughters of US Congressmen (Warren Buffett of Berkshire Hathaway), factory owners (James Simons of the hedge fund Renaissance Technologies), newspaper publishers (Philip Knight of Nike), retail owners (Stephen Schwarzman of the Blackstone Group), and psychiatrists (Dustin Moskovitz, a cofounder of Facebook). We view the "some wealth" category as the equivalent of an upper middle class upbringing.

Figure 5 shows that in the United States, the share of Forbes 400 individuals who are the first generation in their family to run their businesses has risen dramatically from 40 percent in 1982 to 69 percent in 2011. The inheritance of a small family business and the building of it into a larger business were never very common, but it represents a trivial fraction of the sample by 2011. Much of the increase in the first generation businesses has come at the expense of the second-generation and third-generation inherited businesses or wealth. The share of those belonging to fourth-, fifth-, and sixth-generation inherited businesses or wealth has virtually evaporated. These findings are very similar when the results are weighted by wealth.

Figure 6
Did the Forbes 400 Grow Up Wealthy?
(share of Forbes 400 individuals for 1982, 1992, 2001, and 2011 with each upbringing)



Source: Authors using *Who's Who* and Internet searches as primary sources.
Notes: Figure 6 shows the share of Forbes 400 individuals for 1982, 1992, 2001, and 2011 who grew up with little or no wealth, who grew up with some wealth, and who grew up wealthy. In coding the data, we view the "some wealth" category as the equivalent of an upper middle class upbringing.

Figure 6 illustrates that the percent that grew up wealthy fell from 60 to 32 percent while the percent that grew up with some money in the family rose by a similar amount. The percent who grew up with little or no wealth remained about flat. Again, these findings are very similar when the results are weighted by wealth.

Overall, Figures 5 and 6 show a trend in the Forbes 400 list away from people who grew up wealthy and inherited businesses towards those who grew up with more modest wealth in the family and started their own businesses. These changes largely occurred between 1982 and 2001. From 2001 to 2011, the percentage of Forbes 400 that started their businesses increased only slightly, while the percentage that grew up wealthy declined only slightly. The results therefore point to an increase in wealth mobility at the very top, although the added mobility comes from those who would be considered upper middle class.

Access to education also appears to be of increasing importance. The share of the Forbes 400 who graduated from college rose from 77 to 87 percent between 1982 and 2011. The share of college dropouts (like Bill Gates and Mark Zuckerberg) also rose from 6 to 8 percent. At the same time, the share of those without any college dropped markedly from 17 to 5 percent. These results are very similar when the observations are weighted by wealth.

Table 2

Categories of the Wealth-Creating Businesses behind the Forbes 400

(share of Forbes 400 businesses)

	1982	*1992*	*2001*	*2011*	*Change from 1982 to 2011*
Industrial					
Retail/Restaurant	0.053	0.118	0.132	0.150	+0.097
Technology–Computer	0.033	0.053	0.130	0.123	+0.090
Technology–Medical	0.005	0.018	0.021	0.023	+0.017
Consumer	0.131	0.174	0.125	0.108	–0.023
Media	0.136	0.132	0.164	0.100	–0.036
Diversified/Other	0.207	0.205	0.156	0.123	–0.084
Energy	0.214	0.089	0.062	0.098	–0.117
Finance and investments					
Hedge funds	0.005	0.011	0.018	0.075	+0.070
Private equity/leveraged buyout	0.018	0.034	0.039	0.068	+0.050
Money management	0.018	0.055	0.062	0.045	+0.027
Venture capital	0.003	0.005	0.008	0.015	+0.012
Real estate	0.179	0.105	0.081	0.075	–0.104

Source: Authors' calculations from the Forbes 400.

We next code industries of the wealth-generating firms into both broad and more narrowly defined categories. The broad categories are industrial, finance/investments, and real estate. Within the industrial category, we further subdivide into seven subcategories: computer technology, medical technology, retail/restaurant, energy, consumer, media, and diversified/other. Within the finance category, we divide into four subcategories: hedge funds, private equity/leveraged buyouts, venture capital, and money management. This leaves us with 12 separate categories.

Table 2 documents the industries of the wealth-generating businesses of the Forbes 400 members in each year of our sample. The "finance and investments" category grew in representation by around 16 percentage points, technology (both computer and medical) by 11 percentage points, and retail/restaurant by 10 percentage points. Energy shrank by 12 percentage points, real estate shrank by 10 percentage points, and the remaining groups that lost share were the nontechnology industrial businesses. The rise in computer technology and the decline in energy is even more pronounced in value-weighted results.

We also create an indicator variable for whether the business had an important technology component. We begin by assuming that all the technology businesses have a technology component. We then code other businesses as having a technology component if they are pharmaceutical firms, energy firms that develop new extraction technologies (such as fracking), financial firms that exploit new technologies (such as online brokerage), and venture capitalists who invest heavily in technology firms. Even in the businesses started by the Forbes 400 that are not computer

technology businesses per se, technology has become more important. The share of businesses with some technology component increased from 7.3 percent in 1982 to 17.8 percent in 2011. On a value-weighted basis, the percentage increased from 7.1 percent in 1982 to 25.5 percent in 2011, over one-quarter of the total wealth in the 2011 Forbes 400.

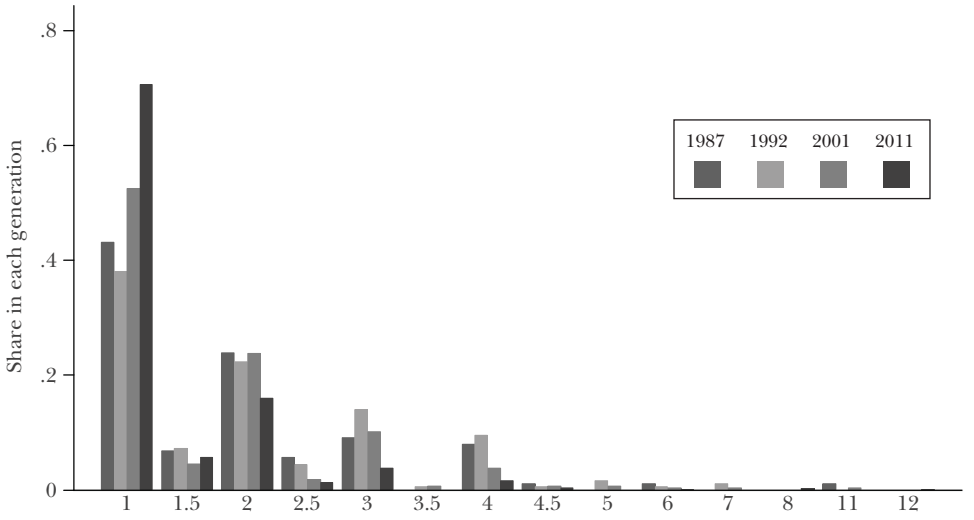
The growth in the technology share occurred largely in the 1990s. The share of computer and medical technologies in the Forbes 400 businesses peaked in 2001 at 15.1 percent, before declining slightly to 14.6 percent in 2011. The share of industries with a technology component rose from 9.1 to 17.5 percent between 1992 and 2001, but only very slightly thereafter. The growth in financial businesses—private equity, hedge funds, and venture capital—on the other hand, occurred primarily in the 2000s, while the share of media, consumer, and diversified businesses declined during that decade.

These findings appear to be consistent with theories of technological change that favor skill in scalable areas (Gabaix and Landier 2008; Kaplan and Rauh 2010). Entering the elite group of the wealthiest individuals is less likely to be linked to having already grown up rich, but having some wealth still confers advantages, particularly in access to education. The wealthiest individuals increasingly comprise individuals who accessed this education while young and then applied their skills to the industries that turned out to be highly scalable. In these industries, the combination of increasing technology and returns to skill allow for the greatest generation of wealth: finance, technology, and also retail. The findings are less consistent with the rise in inequality being the result of broken corporate governance or cultural changes.

Some of these patterns are reflected globally, but others are not. Figure 7 shows that the share of non-US billionaires who are first-generation in the business rose by a similar amount abroad as in the United States, and the composition of the declines in the other categories are very similar to the US experience. Strikingly, however, the share of non-US billionaires who grew up with little or no wealth has risen from under 30 percent in 1987 to over 50 percent in 2012, as shown in Figure 8. The share that grew up with some, but not large wealth has hovered around 20 percent, whereas the share that grew up wealthy plummeted. Recall that while the share of ultra wealthy that grew up wealthy also fell in the United States, the rise of those from the poorest group globally as opposed to the middle group in the US represents a major difference. We can only speculate about the sources of these differences, but one possibility is that in the United States there is better access to education when the family has some wealth, and such access is increasingly important to success in the US economy.

Table 3 shows the business in which the non-US billionaires earned their fortunes over time. Because of their greater prevalence outside the US, Table 3 adds two categories—construction and mining/metals—that were included in "diversified/other" in Table 2. The technology component has become more important globally, but not so important as in the United States. Computer technology and money management are increasingly represented among billionaires globally, but the category that gained

Figure 7
Generation of the Wealth-Creating Businesses in the Forbes Non-US Billionaires in 1987, 1992, 2001, and 2012



Source: Authors using *Who's Who* and Internet searches as primary sources.
Notes: The numbers along the *x*-axis give the generation of a Forbes non-US billionaire within the family of the founder of his or her business. The normal coding for the generation is an integer. However, if the individual inherited a modest business and built it into a much larger one, we input the generation as 1.5.

the most is mining/metals. There was a substantial increase in the business category of energy among the ultra rich globally, whereas this category of business fell among the ultra rich in the United States. Looking at energy together with metal/mining, it is clear that natural resources have played a much larger role outside the United States in creating wealth in the last 20 years.

With the large improvements in information technology and the substantial increase in value of the securities markets over the last 30 years, skilled individuals can now apply their talent to much larger pools of capital. The changes in the composition of the wealthiest individuals in the United States suggest that these trends are shaping wealth more than they used to, while being born to extensive wealth and inheriting family businesses have become less important. Having access to education also has become more important in wealth generation, but it remains an open question what proportion of those gains are due to gains in skill and what proportion are due to improved access to networks.

The rise in the overall college wage premium in the United States may have flattened somewhat in the past decade (James 2012), but our evidence from the identity of the super-rich suggest that the premium for technological skill has continued to rise in the right-tail of wealth outcomes. These findings are most consistent with

Figure 8
Did the Forbes Non-US Billionaires Grow Up Wealthy?
(share of Forbes non-US billionaires in 1987, 1992, 2001, and 2011 with each upbringing)

Source: Authors' using *Who's Who* and Internet searches as primary sources.
Notes: Figure 8 shows the share of Forbes non-US billionaires for 1987, 1992, 2001, and 2011 who grew up with little or no wealth, who grew up with some wealth, and who grew up wealthy. In coding the data, we view the "some wealth" category as the equivalent of an upper middle class upbringing.

the frameworks of Goldin and Katz (2008), in which technological progress widens inequality among skill groups. Over time, this effect might be countered by the continuing broad-based accumulation of human capital, particularly (as pointed out by Acemoglu and Autor 2012) when there are deep interactions between skills and technologies in accomplishing job tasks.

Wealth Inequality

Inequality in wealth appears to be more stable than inequality in income. Estimates of household wealth are available in the triennial Survey of Consumer Finances started in 1983, and from a couple of reasonably comparable earlier surveys, like the Survey of Financial Characteristics of Consumers from 1962. Wolff (2010, 2012) compiles and compares the evidence. He finds that the top 1 percent of US households held 33–35 percent of total net worth in surveys from 1962, 1969, and 1993. In the four surveys done from 1989 to 1998, the top 1 percent of US households held 37–39 percent of net worth. This share of wealth for the top 1 percent fell back below 35 percent in the surveys of 2001, 2004, and 2007, but reached 35.4 percent of total wealth in 2010. In short, the wealth share of the top 1 percent has been more stable than inequality in income, and the latest readings for wealth are below their historic highs.

Table 3

Categories of the Wealth-Creating Businesses behind the Forbes Non-US Billionaires

(shares of the businesses of Forbes non-US billionaires)

	1987	*1992*	*2001*	*2011*	*Change 1987–2011*
Industrial					
Retail/Restaurant	0.101	0.101	0.109	0.124	+0.023
Technology – Computer	0.011	0.039	0.091	0.068	+0.056
Technology – Medical	0.045	0.050	0.042	0.039	−0.006
Consumer	0.157	0.128	0.113	0.098	−0.060
Media	0.067	0.067	0.053	0.030	−0.037
Diversified/Other	0.247	0.346	0.362	0.304	+0.057
Energy	0.022	0.022	0.034	0.049	+0.026
Construction	0.090	0.061	0.042	0.034	−0.056
Mining and metals	0.011	0.011	0.026	0.070	+0.059
Finance and investments					
Hedge funds	0.000	0.000	0.004	0.008	+0.008
Private equity/leveraged buyout	0.022	0.000	0.011	0.010	−0.012
Money management	0.034	0.061	0.072	0.065	+0.031
Venture capital	0.000	0.000	0.004	0.010	+0.010
Real estate	0.101	0.101	0.109	0.124	+0.023

Source: Authors' calculations from the Forbes non-US billionaires.
Note: Table 3 shows the business in which Forbes non-US billionaires the earned their fortunes over time.

These calculations may not capture all of the issues. For example, if the top earners have been engaging in notably more tax avoidance or evasion by offshoring taxes (for some discussion, see Zucman 2013), then wealth inequality might be rising without it being apparent in US-based statistics.

If the proportion of total net worth held by those in the top 1 percent has indeed been fairly stable, then how can this pattern be reconciled with the rising share of income going to the top 1 percent? One possibility is that the top earners decreased their savings rates relative to the rest of the distribution. Definitive conclusions about consumption inequality at the very top are made difficult by data challenges. Looking across the income distribution as a whole, Krueger and Perri (2006) argue that consumption inequality has not tracked income inequality, whereas Aguiar and Bils (2011) argue that "income inequality has been mirrored by consumption inequality." However, no study has shown that the top incomes are actually consuming an even larger share of their income, which is what would be required for consumption patterns to be smoothing out wealth inequality.

Another possible reason that wealth inequality has been more stable than consumption inequality relates to the effects of taxation. As shown above, increases in after-tax, after-transfer inequality in the top 1 percent have been more modest than increases in the before-tax, before-transfer income inequality. If the consumed

share of pretax income were relatively constant, then the increase in the share of income going to taxes would tend to mute the effect of increases in income inequality on wealth inequality.

Conclusions

Inequality in pretax income as measured by the income share of the top 1 percent has increased substantially in the US since 1980. The rise in after-tax, after-transfer income of the top 1 percent has been more muted, and the current levels are closer to the late 1990s and not so high as in the mid-2000s. The sources of rising pre-tax inequality have been long debated. Of course, these theories are not mutually exclusive and will overlap in various ways, but it can be a useful shortcut to think about these theories in two broad categories.

In one category, rising inequality is being driven by top earners finding ways to extract additional income. This category includes explanations of greater inequality based on theories of managerial power (Bebchuk and Fried 2004), social norms (Piketty and Saez 2006; Levy and Temin 2007), and how changes in marginal tax rates alter incentives for managers to seek higher income for themselves (Piketty, Saez, and Stantcheva forthcoming). In many of these theories, top earners obtain rents in the sense that they distort the economic system to extract resources in excess of their marginal products.

In the other category, the increase in inequality has been driven by economic factors that have altered the marginal productivity of certain kinds of labor. This category would include explanations of greater inequality based in skill-biased technological change (Katz and Murphy 1992; Garicano and Rossi-Hansberg 2006; Autor, Katz, and Kearney 2006; Garicano and Hubbard 2007), greater scale (Tervio 2008; Gabaix and Landier 2008), superstars (Rosen 1981), and trade or globalization (Hecksher 1931; Ohlin 1933; Stolper and Samuelson 1941). In most of these theories, the idea is that top earners have scarce and unique talents that allow them to command a premium due to the increasing value of their talents in markets of increasing size.

Our results tend to support the second broad category of explanations more than the first. For example, one version of the managerial power or extraction hypothesis argues that corporate governance has deteriorated in a way that allows top executives to have increased their compensation substantially (Bebchuk and Fried 2004). But our evidence suggests that poor corporate governance cannot be more than a small part of the picture of increasing income inequality, even at the very upper end of the distribution. Other groups that do not report to a potentially compliant board of directors, particularly private company executives, have experienced equal or larger increases in their contribution to the top income brackets. Furthermore, the Forbes 400 lists include virtually no public company chief executives who are not founders or who became the chief executive officer after the company went public.

A related theory of extraction argues that high earners have a greater incentive to extract from lower earners when marginal tax rates are low, because they get to keep a larger share of the returns from this bargaining activity (Piketty, Saez, and Stantcheva forthcoming). Our evidence that the wealthiest individuals in the US are increasingly populated by technology entrepreneurs is not directly consistent with this, as the fortunes of the developers of new technologies are arguably quite insensitive to the strength of their bargaining against lower-wage workers. The fact that the top 1 percent share in the wealth distribution has in fact been much more stable than the income distribution, and that the rise in the share of after-tax, after-transfer income held by the top 1 percent has been much less, also suggests that the top earners may have actually not kept as much of these income gains.

Our evidence also is not obviously consistent with those who suggest that the increase in pay at the top is driven by a recent removal of social norms regarding pay inequality. While top executive pay has increased, so has the pay of other groups, who are and were less subject to disclosure and, arguably, less subject to social norms. This is particularly true of private company executives and hedge fund and private equity investors.

Overall, we believe that our evidence remains more favorable toward the theories that root inequality in economic factors, especially skill-biased technological change, greater scale, and their interaction. Skill-biased technological change predicts that inequality will increase if technological progress raises the productivity of skilled workers relative to unskilled workers and/or raises the price of goods made by skilled workers relative to those made by unskilled workers. For example, computers and advances in information technology may complement skilled labor and substitute for unskilled labor. This seems likely to provide part, or even much, of the explanation for the increase in pay of professional athletes (technology increases their marginal product by allowing them to reach more consumers), Wall Street investors (technology allows them to acquire information and trade large amounts more easily) and executives, as well as the surge in technology entrepreneurs in the Forbes 400. Globalization may have contributed to greater scale, but globalization cannot drive the increase in inequality at the top levels given the breadth of the phenomenon across the occupations we study.

Tervio (2008) and Gabaix and Landier (2008) analyze "assignment models" for chief executive officers in which more-able CEOs can add more value to larger firms. In a competitive equilibrium, their models predict that top executives will be paid more as their firms and other firms they can work for become larger, because larger size increases the returns to hiring the more productive people. Tervio (2008) concludes that his model can explain the high levels of CEO pay. Gabaix and Landier (2008) attribute the large increase in pay of chief executive officers in US public companies to the large growth of those companies over time—the typical large US firm increased in market value by four to seven times in real terms from 1980 to 2003. Public companies, financial services firms, venture capital funds, private equity funds, hedge funds, and law firms all have grown larger, in many instances by orders of magnitude. Although the two papers do not seek to explain

why firms (and funds) have been able to become so much larger over time, the same technological change that is biased towards skills may have helped firms and funds to become larger during our sample period.

These theories, interacted with the incentive effects of taxes, regulations, and institutions, also help explain why income inequality has not risen as much in other countries such as Sweden and France. While these countries have implemented productivity-enhancing technologies, there has been considerably less direct development of the most profitable new technologies there than in the United States. While the private equity, venture capital, and hedge fund industries have grown in these places, they have not been scaled up to the extent that they have in the United States. Disincentives related to tax and regulatory policy may have caused both financiers and innovators in such countries to locate elsewhere, in places like London and Silicon Valley.

■ *We thank JEP Editor David Autor along with Austan Goolsbee, Chang-Tai Hseih, Ulrike Malmendier, Timothy Taylor, and conference participants at the 2013 AEA meetings for helpful comments. We are grateful to Donny Chi, Kirill Demtchouk, Yifei Feng, and Kimberley Ho for research assistance. Kaplan acknowledges research support from the Fama-Miller Center and the Initiative for Global Markets at Chicago Booth. Kaplan serves on public company and mutual fund boards and has been a consultant to private equity and venture capital investors.*

References

Acemoglu, Daron, and David Autor. 2012. "What Does Human Capital Do? A Review of Goldin and Katz's *The Race between Education and Technology*." *Journal of Economic Literature* 50(2): 426–63.

Aguiar, Mark A., and Mark Bils. 2011. "Has Consumption Inequality Mirrored Income Inequality?" NBER Working Paper 16807.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2006. "The Polarization of the U.S. Labor Market." *American Economic Review* 96(2): 189–94.

Bakija, Jon, Adam Cole, and Bradley T. Heim. 2012. "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data." http://web.williams.edu/Economics/wp/BakijaColeHeimJobsIncomeGrowthTopEarners.pdf.

Bebchuk, Lucian, and Jesse Fried. 2004. *Pay without Performance: The Unfulfilled Promise of Executive Compensation*. Harvard University Press.

Bell, Brian, and John Van Reenen. 2010. "Bankers' Pay and Extreme Wage Inequality in the UK." Open Access publications from London School of Economics and Political Science, no. CESP21.

Congressional Budget Office. 2012. "The Distribution of Household Income and Federal Taxes, 2008 and 2009." http://www.cbo.gov/publication/43373.

Dolan, Kelly A. 2011. "The Forbes 400 Methodology." http://www.forbes.com/sites/kerryadolan/2011/09/21/the-forbes-400-methodology/.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123(1): 49–100.

Garicano, Luis, and Thomas N. Hubbard. 2007.

"The Returns to Knowledge Hierarchies." NBER Working Paper w12815.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121(4): 1383–1435.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology*. Harvard University Press.

Hecksher, Eli. 1931. *Merkantilismen*. Stockholm: Norstedt & Söner.

James, Jonathan. 2012. "The College Wage Premium." Economic Commentary 2012-10, Federal Reserve Bank of Cleveland. http://www.clevelandfed.org/research/commentary/2012/2012-10.cfm.

Kaplan, Steven N. 2012. "Executive Compensation and Corporate Governance in the United States: Perceptions, Facts and Challenges." *Cato Papers on Public Policy*, vol. 2, pp. 99–157.

Kaplan, Steven N., and Joshua Rauh. 2010. "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?" *Review of Financial Studies* 23(3): 1004–1050.

Kaplan, Steven N., and Joshua Rauh. 2013. "Family, Education, and Sources of Wealth among the Richest Americans, 1982–2012." *American Economic Review* 103(3): 158–62.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107(1): 35–78.

Krueger, Dirk, and Fabrizio Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *Review of Economic Studies* 73(1): 163–93.

Levy, Frank, and Peter Temin. 2007. "Inequality and Institutions in 20th Century America." NBER Working Paper 13106.

Metrick, Andrew, and Ayako Yasuda. 2010. "The Economics of Private Equity Funds." *Review of Financial Studies* 23(6): 2303–41.

Ohlin, Bertil. 1933. *Interregional and International Trade*. Cambridge: Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1): 1–39.

Piketty, Thomas, and Emmanuel Saez. 2006. "The Evolution of Top Incomes: A Historical and International Perspective." *American Economic Review* 96 (2): 200–206.

Piketty, Thomas, and Emmanuel Saez. 2013. Tables and figures from "Income Inequality in the United States" (2003), updated to 2011 at Saez's website: http://elsa.berkeley.edu/~saez/.

Piketty, Thomas, Emmanuel Saez., and Stefanie Stantcheva. Forthcoming. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities." *American Economic Journal: Economic Policy*.

Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review* 71(5): 845–58.

Stolper, Wolfgang F., and Paul A. Samuelson. 1941. "Protection and Real Wages." *Review of Economic Studies* 9(1): 58–73.

Tervio, Marko. 2008. "The Difference that CEOs Make: An Assignment Model Approach." *American Economic Review* 98(3): 642–68.

Wolff, Edward N. 2010. "Recent Trends in Household Wealth in the United States: Rising Debt and the Middle Class Squeeze—An Update to 2007." Working Paper 589, Levy Economics Institute of Bard College.

Wolff, Edward N. 2012. "The Asset Price Meltdown and the Wealth of the Middle Class." NBER Working Paper 18559.

Zucman, Gabriel. 2013. "The Missing Wealth of Nations: Are Europe and the US Net Debtors or Net Creditors?" *Quarterly Journal of Economics* 128(3): 1321–64.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 57–78

The Pay of Corporate Executives and Financial Professionals as Evidence of Rents in Top 1 Percent Incomes

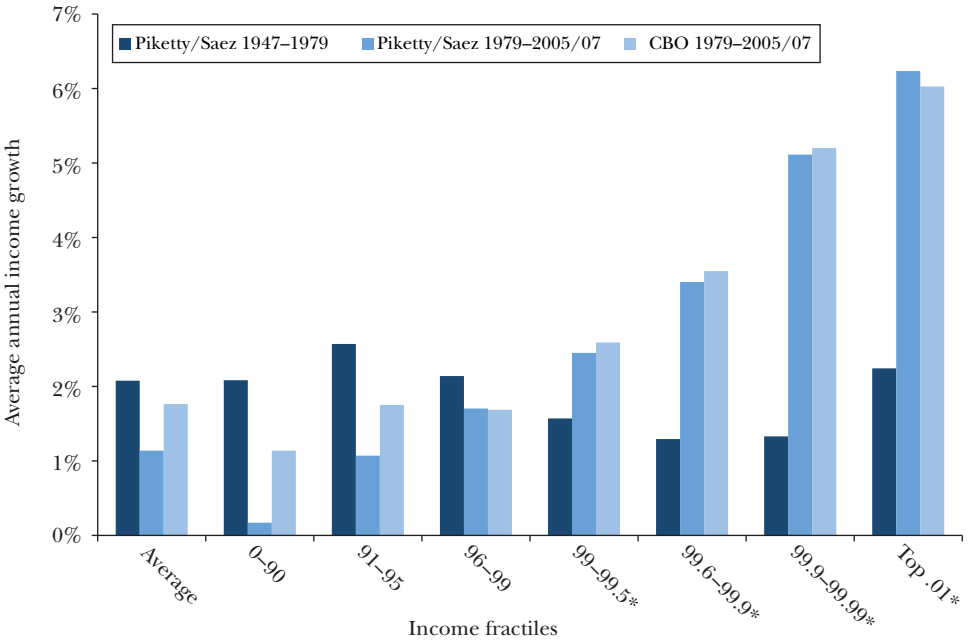
Josh Bivens and Lawrence Mishel

The debate over the extent and causes of rising inequality of American incomes and wages has now raged for at least two decades. In this paper, we will make four arguments. First, the increase in the incomes and wages of the top 1 percent over the last three decades should be interpreted as driven largely by the creation and/or redistribution of economic rents, and not simply as the outcome of well-functioning competitive markets rewarding skills or productivity based on marginal differences. This rise in rents accruing to the top 1 percent could be the result of increased *opportunities* for rent-shifting, increased *incentives* for rent-shifting, or a combination of both. Second, this rise in incomes at the very top has been the primary impediment to having growth in living standards for low- and moderate-income households approach the growth rate of economy-wide productivity. Third, because this rise in top incomes is largely driven by rents, there is the potential for checking (or even reversing) this rise through policy measures with little to no adverse impact on overall economic growth. Lastly, this analysis suggests two complementary approaches for policymakers wishing to reverse the rise in the top 1 percent's share of income: dismantling the institutional sources of their increased ability to channel rents their way and/or reducing the *return* to this rent-seeking by significantly increasing marginal rates of taxation on high incomes.

■ *Josh Bivens is Research and Policy Director and Lawrence Mishel is President, Economic Policy Institute, Washington, D.C. Their email addresses are jbivens@epi.org and lmishel @epi.org.*

http://dx.doi.org/10.1257/jep.27.3.57 doi=10.1257/jep.27.3.57

Figure 1
Average Annual Income Growth by Fractiles, by Time Period and Dataset



Source: Authors' analysis of updated data from Piketty and Saez (2003) and Congressional Budget Office (2012).
* Indicates data series end in 2005.

Evidence on the Rise in Income Shares at the Top

The facts about this rise in income shares at the very top are well-known, so our summary will be brief. Figure 1 summarizes the main points. It shows average annual income growth using updated data from Piketty and Saez (2003) on cash, market-based incomes in two time-periods, 1947–1979 and 1979–2007. For comparison, the figure also looks at Congressional Budget Office (2012) data for the latter period on comprehensive incomes, which include noncash benefits from employers and government transfers. The implications of these growth rates for the shares of overall growth claimed by various groups are startling. The Piketty and Saez data indicate, for instance, that between 1979 and 2007, the top 1 percent of American tax units accounted for 59.8 percent of average growth in cash, market-based incomes compared to just 9 percent of average growth accounted for by the bottom 90 percent over the period. While including transfers and noncash incomes reduces the share of growth received by the top 1 percent significantly, as shown in the Congressional Budget Office data, the top 1 percent *still* account for 38.3 percent of growth, more than the 31.0 percent share received by the bottom 80 percent.

The figure also illustrates several other points. First, the Piketty and Saez data indicate strongly that the rise in inequality is only a feature of the latter period: that is, on the figure, the darker bars on the left showing income growth for the 1947–79 period are fairly similar between the 0–90th percentile and the other measures shown. However, the measures for the more recent period from 1979 to 2007 show much faster income growth at the upper levels of income.

Second, while measuring income using the cash, market-based data from Piketty and Saez does have some shortcomings, in fact, movements in this type of income do overwhelmingly drive trends in inequality even in the more comprehensive income dataset tracked by the Congressional Budget Office. This finding should hardly be a shock, because cash, market-based incomes account for roughly 80 percent of all incomes even in the CBO data. Both of these points hold even more strongly when looking just at incomes in the top 1 percent and above: growth rates for these incomes are nearly identical in the latter period in both datasets, as the dominance of cash, market-based incomes is even greater for the highest income households.

Third, the divergence in income growth rates for the bottom 90 percent in these two datasets is striking. In the Piketty and Saez data, the bottom 90 percent saw average annual growth of just 0.2 percent as compared to 1.1 percent in the CBO data. We have noted elsewhere that more than half of the income growth for households in the middle of the income distribution between 1979 and 2007 was driven by government transfers (dominated by Social Security and Medicare) and pensions currently received for *past* labor market service (Mishel, Bivens, Gould, and Shierholz 2012, table 2.13). These influences accounted for more than 80 percent of the rise in average incomes for the middle fifth. Cash wages and salaries, conversely, accounted for just 6.1 percent of overall income growth for these families (while employer-sponsored health insurance contributions contributed 12 percent of overall growth in comprehensive incomes). We view the Piketty and Saez data as most reflective of how well the market economy has been providing income gains for households in the bottom 90 percent (that is, not very well), while the CBO data provides a better measure of the actual living standards attained by this group.

Lastly, in both the Piketty and Saez data as well as the Congressional Budget Office data, average income growth by fractile does not equal or exceed *overall average* growth below the 96th to 99th percentile.

Table 1 provides an overview of the sources of income growth for the top 1 percent in the three decades before the Great Recession, using the broader definition of income from the Congressional Budget Office (2012) data. The first two columns show the share of each income category claimed by the top 1 percent. They indicate that between 1979 and 2007 the top 1 percent's share of each major income category rose significantly, indicating that concentration *within* all forms of market income significantly lifted this share.

The next two columns show the share of total economy-wide income accounted for by each income category. The most striking finding in these columns is the large

Table 1
Sources of Top 1 Percent Incomes

	Top 1 percent share		*Income category share in whole economy*	
	1979	*2007*	*1979*	*2007*
Labor compensation	4.1%	8.8%	69.8%	60.3%
Capital and business income	31.8%	56.2%	18.3%	22.8%
Dividends, interest and rent*	26.9%	43.8%	10.2%	8.7%
Capital gains	58.5%	74.2%	3.6%	8.0%
S-corporation dividends and proprietors' income**	21.3%	50.6%	4.5%	6.1%
Other income	5.0%	7.1%	3.2%	6.3%
Transfers	1.0%	1.0%	8.7%	10.7%
Total	8.9%	18.7%	100.0%	100.0%

Source: Author's analysis of Congressional Budget Office (2012).
* Not including S-corporation dividends, this entire category is classified by the CBO as "capital income."
** Classified by CBO as "business income."

decline in labor compensation's share of overall income, falling from 69.8 percent in 1979 to 60.3 percent in 2007. Conversely, the combined share of capital income (including capital gains) and business income rose substantially, from 18.3 percent in 1979 to 22.8 percent in 2007. (We have noted elsewhere—Mishel, Bivens, Gould, and Shierholz 2012—that the rise in capital income's share is driven overwhelmingly by a higher profit rate, not a rise in capital-output ratios.) Finally, both "Other income" (mostly pension payments for past labor services) and transfer income's shares of total incomes rose, by a combined 5.1 percent between 1979 and 2007. Because the top 1 percent have much larger shares of capital income (56.2 percent in 2007) as compared to labor compensation (8.8 percent in 2007), the overall shift from labor- to capital-based incomes mechanically pushes up the top 1 percent income share. This shift from less- to more-concentrated sources of income accounts for over a quarter of the rise in the top 1 percent share, with rising concentration *within* income categories accounting for the rest.

Besides their sources of income, the occupations of the top 1 percent have also been investigated. In a study of tax returns from 1979 to 2005, Bakija, Cole, and Heim (2010) show the trend in the shares of total income of US households accruing to the top 1.0 and top 0.1 percent of households. They establish that the increases in income at the top were disproportionately driven by households headed by someone who was either an "executive" (including managers and supervisors and hereafter referred to as "executives") in nonfinancial sectors or someone who is in the financial sector as an executive or other worker. Households headed by a nonfinance executive were associated with 44 percent of the growth of the top 0.1 percent's income share and 36 percent in the growth among

Table 2
CEO Compensation and Stock Growth, 1965–2012
(2012 dollars)

	CEO annual compensation (thousands of dollars)	*Stock market indices (in 2011 dollars)*	
		S&P 500	*Dow Jones*
Year			
1965	807	562	5,805
1973	1,054	496	4,268
1978	1,442	310	2,652
1989	2,685	578	4,488
1995	5,684	810	6,731
2000	19,880	1,903	14,298
2007	18,274	1,636	14,593
2010	12,286	1,200	11,235
2011	12,484	1,294	12,206
2012	14,074	1,379	12,965
Percent change			
1965–78	79%	–45%	–54%
1978–2000	1,279%	513%	439%
2000–12	–29%	–28%	–9%
1978–2012	876%	344%	389%

Sources: Mishel and Sabadish (2013) analysis of Compustat ExecuComp database, and Federal Reserve Economic Data (FRED) from Federal Reserve Bank of St. Louis.
Notes: CEO compensation includes salary, bonus, restricted stock grants, options exercised, and long-term incentive payouts for CEOs at the top 350 firms ranked by sales. The results are similar if we use the value of options granted rather than options exercised. In Mishel and Sabadish (2013), we present a CEO compensation series where stock options are measured as the value of options granted.

the top 1.0 percent. Those in the financial sector were associated with nearly a fourth (23 percent) of the expansion of the income shares of both the top 1.0 and top 0.1 percent. Together, finance and executives accounted for 58 percent of the expansion of income for the top 1.0 percent of households and an even greater two-thirds share (67 percent) of the income growth of the top 0.1 percent of households. Relative to others in the top 1 percent, households headed by executives had roughly average income growth, those headed by someone in the financial sector had above average income growth, and the remaining households (nonexecutive, nonfinance) had slower than average income growth. In our view, this analysis of household income data understates the role of executives and the financial sector since they do not account for the increased spousal income from these sources.

Mishel and Sabadish (2013) examine chief executive officers of 350 firms with largest revenue in any given year and show that their compensation grew 79 percent between 1965 and 1978 during a period when the stock market (as measured by the Dow Jones and Standard & Poor's indices) fell by about half—as shown in Table 2.

Average CEO compensation grew strongly over the 1980s but then exploded in the 1990s and peaked in 2000 at nearly $20 million, growing by a multiple of 13 from 1978 to 2000.[1] This growth in compensation for chief executive officers far exceeded even the substantial rise in the stock market, which grew roughly five-fold in value over the 1980s and 1990s.

Since 2000, compensation for chief executive officers first dropped after the fall in the stock market in the early 2000s, rebounded by about 2007, fell again in the financial crisis of 2008–2009, and then has rebounded again. By 2012, average compensation had returned to $14.1 million. The compensation of chief executive officers in 2012 is high by any metric, except when compared with its own peak in 2000 after the 1990s stock bubble. From 1978 to 2012, even with the setbacks provided by the 2001 and 2008 stock market crashes, CEO realized compensation grew 876 percent, more than double the real growth in the stock market. In contrast, the hourly compensation of a private sector production/nonsupervisory worker grew a meager 5 percent.

This analysis of a narrow group of occupations' role in driving top 1 percent incomes allows us to narrow the argument about the role of rents in high incomes in a useful way. If one could establish that high compensation in just these two occupations (or sectors) is indeed heavily influenced by rents and are not just the efficient marginal return to differences in skill and ability, then one has effectively made this case for rents playing an important role in the compensation of the majority of the top 1 percent.

Rent-Shifting and the Rise in Top 1 Percent Incomes

In many discussions, "rent-seeking" refers loosely to ill-gotten gains. For example, it often refers to gains obtained by lobbying the government for some form of subsidy or preferential treatment, or for profits that stem from anticompetitive behavior. Such rents do exist, and powerful economic actors do indeed manage to influence policy to protect themselves from competition, and these probably do have nontrivial influences on some incomes in the top 1 percent. Legal and policy measures that enforce intellectual property claims, for example, surely play a key role in high incomes in Silicon Valley.

However, in this paper we are referring to rents in a broader sense: in this discussion, a "rent" means only that the income received was in excess of what was needed to induce the person to supply labor and capital to these respective markets. As an illustrative example, it seems likely that many top-level professional athletes

[1] The multiple is 14 if we consider options granted rather than options realized. In this paper, for simplicity, we will be discussing compensation with options realized, rather than options granted. Results using the value of options granted and the value of exercised options are very similar. Mishel and Sabadish (2013) present a CEO compensation series where stock options are measured as the value of options granted.

would continue to supply essentially the same amount of labor to their sport even if their salary was reduced by some substantial fraction, because even the reduced salary would be much higher than their next-best options. Thus, we are making a positive argument, not a normative one, that the rise in income for the top 1 percent income was not necessary to entice the people in that group to seek those jobs nor to provide effort in those jobs. One implication of this argument is that rents that do not necessarily interfere with efficient allocation of talent (for example, those described in Gabaix and Landier, 2008, and Terviö, 2008) can be redistributed via progressive marginal taxation or some other method without introducing economic distortions. And where rents do stem from institutional arrangements that block the efficient allocation of resources (as in many of the models of financial sector rents or CEO "pay-skimming" noted below or models with labor market rents like Terviö 2009), then dismantling the sources of these rents will self-evidently lead to increased efficiency.

Since rents are rarely observable, direct evidence on their role in driving income dynamics is scarce. However, we will focus on a wealth of suggestive evidence consistent with rents being important in income determination and then will point to direct evidence suggesting that the large increases in top 1 percent income shares has clearly not been associated with improved economic performance overall. We take this as further confirmation that well-designed policies can brake or reverse the rising top 1 percent shares without harming overall economic growth.

Executive Pay

The contribution made by concentration of labor compensation to the growth in the income share of the top 1 percent can largely be explained by a pay-setting institution that had existed for a long time, but took off with particular force in the last couple of decades: exercised stock options and bonuses. Kruse, Blasi, and Freeman (2011), for example, note that in 2006 roughly $65.1 billion in labor compensation was actually the result of exercised stock options, while Jaquette, Knittel, and Russo (2003) have estimated that total "spread income" (the exercise of nonqualified stock options) was $126 billion in 2000, and was even $78 billion in 2001, following the stock market decline. Stock options and bonuses are particularly relevant to the pay of high-level corporate executives.

Bebchuk and Fried (2004) compile ample evidence in favor of the claim that top executive pay is largely the result of rent-extraction. Perhaps their most persuasive argument is to point out that a well-designed contract for executive pay should offer rewards based on relative performance. For example, an executive for a company in an industry where stock prices are down across the board should be rewarded for performing less poorly than others, while an executive who runs a company in which the stock price is up, but up by less than every other firm in the industry, should not be rewarded. However, real-world compensation arrangements for chief executive officers are typically "camouflaged" to look like they are the result of contractual arrangements that reward relative performance, but generally they do not reward relative performance.

One example of this camouflage is the hiring of compensation consultants and the construction of "peer groups" to benchmark top executive salaries. While at first glance benchmarking to insure that shareholders are not overpaying for managers may seem like sound corporate practice, these consultants and peer-group constructions can often be used to justify inflated corporate pay. Bizjack, Lemmon, and Nguyen (2011), for example, find evidence that "peer groups are constructed in a manner that biases compensation upward."

Another example of such camouflage is construction of stock options—an instrument that *could* be consistent with aligning manager and shareholder interests—that largely reward the luck of whether the stock market rises or falls, rather than specific performance. Bertrand and Mullainathan (2001), for example, have found that the pay for luck is actually as large as pay for performance, and they interpret this finding as evidence in support of the rent-extraction hypothesis for pay of chief executive officers.

As yet another example, Bebchuk and Fried (2010) note that "standard pay arrangements have commonly failed to restrict the use of financial instruments that can weaken or eliminate entirely the incentive effects of equity-based instruments awarded as part of compensation arrangements." They note that in a study by Schwab and Thomas (2006) of 375 employment contracts governing pay of chief executive officers, not a single one restricted hedging away the risk of the option grants.

A last bit of evidence that flawed corporate governance has allowed US corporate executives to receive inefficiently high pay is the high ratio of the pay of US chief executive officers relative to their international peers. Fernandes, Ferreira, Matos, and Murphy (2012) show US compensation for chief executive officers in 2006 to be twice that of other advanced nations at both the median and mean. A survey by Towers Perrin, a consulting firm, shows US CEO compensation was triple that of other advanced nations in 2003, up from 2.1 times foreign CEO compensation in 1988 (Mishel, Bernstein, and Allegretto 2004, table 2.47). Towers Perrin also reports that US CEO compensation was 44 times that of the average worker whereas the non-US ratio was 19.9.

Fernandes et al. (2012) seek to challenge the claim that US CEOs are paid significantly more than their foreign counterparts. As noted above, they find US CEOs to be paid double that of their counterparts. However, even after controlling for firm-level characteristics—size, leverage, stock return, stock volatility, and Tobin's Q—they find a US pay premium of 88 percent. It is only when they control for "inside and institutional ownership" that they knock the pay premium down to a still substantial 31 percent. It is not clear to us that US institutional board arrangements are reflective of the skill of chief executive officers and should be included as a control when estimating the pay premium. One could, in fact argue simply that US-style governance features are associated with excessive pay for chief executive officers both here and abroad.

A Closer Look at Rent-Shifting in the Financial Services Sector

Whether rents are driving much of the increase in top salaries in the financial sector is a debate that is largely inseparable from the broader question of the social

value of marginal increases in the activities of the finance sector and whether the large expansion of the sector between the 1970s and 2007 was of benefit or detriment to the larger economy. This question has been interestingly addressed in a symposium in the Spring 2013 issue of this journal. Although we will not reiterate the arguments in detail here, we are convinced by arguments that the wider economy has not benefited from this expansion of finance and that it largely represents overpayment for financial intermediation services that more competitive markets could have delivered more efficiently. Moreover, this expansion of finance actually imposed large negative externalities on the wider economy through the increase in systemic risk that has accompanied the rise in large, complex financial institutions.

Again, the evidence on rent-shifting behavior should be viewed not as conclusive, but as highly suggestive. First, the rise in finance's share of overall economic activity and the steep rise in top incomes in this sector coincide with a range of legislative and regulatory changes that vastly expanded the range of activities in which financial firms could engage. Regulatory prohibitions from earlier eras were explicitly dismantled or made moot (for a good summary of many of these changes, see Sherman 2009). The result of these regulatory changes was a large rise in bank concentration, following a generation of economic history that saw concentration ratios roughly hold steady. There is very little evidence that the large rise in bank concentration reflected economies of scale or scope that were passed on to consumers in lower prices of intermediation (Haldane 2010a).

Second, it has become clear that, in the aftermath of the Great Recession, some potentially substantial share of the income for large financial institutions is based on implicit insurance against bankruptcy (if it can still be called "implicit" after all the financial sector loans and bailouts) that large financial firms receive from the government. These subsidies can be economically significant, both in fiscal costs from clean-ups after crises happen (Laeven and Valencia 2008) as well as reduced financing costs for firms perceived to be too big or too interconnected or too politically influential to fail (Baker and McArthur 2009). Further, the value of this implicit insurance rises with the risk of the underlying activities, and given that deregulation in this sector allowed a wider range of (often quite risky) activities, the value of this implicit insurance surely got larger as well.

Third, some financial firms seem to extract large rents largely by hiding financial risk, rather than managing it. Haldane (2010b) has highlighted many of the means through which financial firms have in recent decades assumed risk in a search for high returns while also managing to hide this risk from their sources of finance. Bebchuck, Cohen, and Spamann (2009) provide a stark example of the large gap between value produced by financial sector institutions and value claimed by their managers in their examination of the compensation provided to executives at Bear Stearns and Lehman Brothers—two of the most spectacular failures in American finance during the crisis. They show that even after including the losses suffered by top management from the loss of value of their holdings at the time of each banks' respective crash, that managers at these firms were able to obtain staggering payoffs over the entire 2000–2008 period: $650 million for Bear Stearns'

top executive team and $400 million for Lehman's team. Biais, Rochet, and Wooley (2010) note that financial managers often have an opportunity, with a combination of asymmetric information and the inability of outside investors to punish moral hazard fully (because of limited liability), to shift rents to themselves by failing to assess the true underlying risks of new financial innovations when they manage principals' money.

The overall pay of financial sector workers relative to others in the economy has risen substantially. Our own calculations, using National Income and Product Accounts Data, show the unadjusted ratio of financial sector pay (annual compensation per full-time employee) relative to the pay of workers in the rest of the economy since 1948. Between 1952 and 1982, this ratio never exceeded 1.1. By 2007, after decades of steady growth, it had reached 1.83. The rise in financial sector pay persists in the data even when standard wage-equation controls for experience and education are introduced. Philippon and Reshef (2009) also chart a rapid rise in the pay of financial sector workers, and they construct a time-series to chart the tight correlation between above-average pay in finance and the historical ebb and flow of financial regulation and deregulation (see also their paper in the Spring 2013 issue of this journal). They note that a significant pay premium to working in finance persists even in regressions with multiple controls, and conclude that roughly 30–50 percent of pay premium in finance seems due to rents.

In short, the financial sector illustrates that in one of the most important sectors driving top 1 percent incomes in recent years, there was an extraordinary divergence between what top managers took home and even what shareholders (surely a privileged group compared to the wider US economy) gained. This type of divergence seems like powerful evidence to us that a substantial part of the extraordinary rise of top 1 percent incomes is not a result of well-functioning markets allocating pay according to value generated, but instead resulted from shifting institutional arrangements leading to shifting of rents to those at the very top.

Objection: But It's Other Professions, Too

One response sometimes heard to our analysis of rent-shifting behavior by executives and financial professionals is that their pay has largely risen in proportion to the pay of others in the top 1 percent. Thus, the argument goes, it is implausible to argue that rent-shifting was happening to the same extent across different professions, so a supply-and-demand explanation about higher rewards to those with very rare skills is a more plausible answer. While we cannot discuss every occupation that has contributed to top 1 percent pay, we will note that rents seem extraordinarily important to many of the other occupations, and especially those represented outside of executives and finance professionals in the very top—say the top 0.1 or 0.01 percent.

Lawyers, for example, are often hired *explicitly* to redistribute returns to productive activity. In one research paper, Murphy, Shleifer, and Vishny (1991) explicitly use pre-law college majors as a proxy for the presence of rent-seeking behavior. As another example, the health care sector is one of the few that has actually outpaced

finance in terms of claiming an ever-larger share of overall economic activity, yet much analysis finds not only that a large fraction of provided (and billed-for) care does not measurably improve health outcomes, but that prices in the American system are vastly higher than in the health systems of our advanced-country peers (Anderson, Reinhardt, Hussey, and Petrosyan 2003; Cutler and Sheiner 1999). Further, the growth of spending on pharmaceuticals and medical devices constitute a significant share of overall health spending over the last generation, and these are sectors within health care given explicit government protection in the form of patents (Davis, Schoen, Guterman, Shih, Schoenbaum, and Weinbaum 2007).

More broadly, the existence and growth of rent-seeking sectors can pull up wages and incomes in other sectors. For example, Laugesen and Glied (2011) have demonstrated that physician salaries (orthopedists, in their study) in the United States are significantly higher even compared to that in our high-income industrial peers. The authors then make an astute point: "One explanation for the higher incomes of U.S. physicians may lie in the broader U.S. income structure. The share of income received by people in the top 1 percent of the U.S. income distribution far exceeds the corresponding share in the comparison countries." Empirical support for their point can be found in the work of Kedrosky and Stangler (2011) and Goldin and Katz (2008); both papers chart a large increase in the share of graduates from elite universities choosing to enter finance rather than other fields like medicine or hard sciences.

In short, just to keep a constant-quality workforce in the face of rent-driven increases in pay for chief executive officers and financial professionals, even competitive labor markets in other occupations near the top of the income distribution would have to see pay rise. In addition, it appears to us that the enormous pay increases received by chief executive officers of large firms have spillover effects, in the sense that the pay of other executives and managers rises in tandem with the pay of chief executive officers, although we do not know of any systematic studies that have established the scale of this impact.

One prominent example of the "it's other professions, too" argument comes from Kaplan (2012b), who argues that the pay of chief executive officers has risen in line with that of other highly-paid workers and that this is evidence against managerial power and rent-seeking driving the pay trends for chief executive officers. As noted a moment ago, even if the pay of chief executive officers was rising in line with others in the top 1 percent, we are not sure that this actually would be evidence against the managerial power theory of high pay for chief executive officers. But in addition, our reading of Kaplan's own data and more precise data yields an opposite conclusion. Table 3 presents the ratio of the average compensation of chief executive officers of large firms (the series developed by Kaplan) to two benchmarks: The first is the one Kaplan uses, the average *household income* of those in the top 0.1 percent developed by Piketty and Saez (2003, updated). The second is the average annual earnings of the top 0.1 of *wage earners* based on a series developed by Kopczuk, Saez, and Song (2010) and updated in Mishel et al. (2012). Each ratio is presented as a simple ratio and a log value (to convert to a "premium"). The

Table 3
Growth of Relative CEO and College Wages, 1979–2010

	Simple ratio			*Log ratio*		
	CEO compensation to income of top 1% households	*CEO compensation to wages of top 1% wage earners*	*College–High School wage premium*	*CEO compensation to income of top 1% households*	*CEO compensation to wages of top 1% wage earners*	*College–High School wage premium*
Ratio						
1979	1.18	3.16	1.40	0.164	1.151	0.338
1989	1.14	2.55	1.57	0.128	0.936	0.454
1993	1.56	2.95	1.63	0.443	1.083	0.488
2000	2.90	7.53	1.75	1.063	2.019	0.557
2007	1.49	4.23	1.76	0.396	1.442	0.568
2010	2.06	4.70	1.77	0.725	1.548	0.574
Change in ratio						
1979–2007	0.31	1.07	0.36	0.23	0.29	0.23
1979–2010	0.89	1.54	0.37	0.56	0.40	0.24
1989–2010	0.93	2.15	0.20	0.60	0.61	0.12

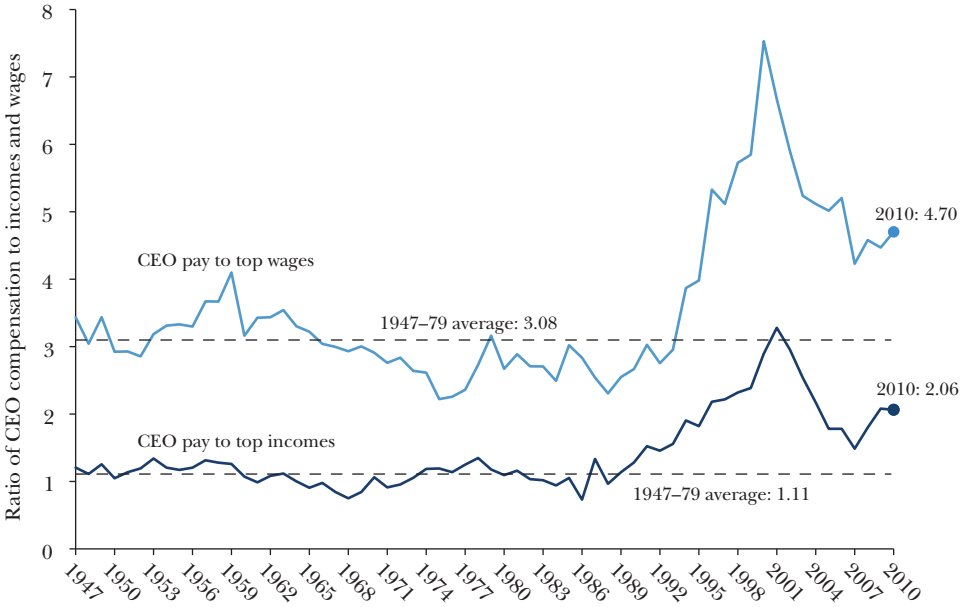
Sources: CEO compensation from Kaplan (2012b); top 0.1% wages from Mishel, Bivens, Gould, and Shierholz (2012, table 4.8); wages by education from Mishel et al. (2012, table 4.14); top 0.1 incomes from Piketty and Saez (2003, updated).

wage benchmark seems the most appropriate one since it avoids issues of household demographics—changes in two-earner couples, for instance—and limits the income to labor income and excludes capital income. Both of these ratios clearly understate the relative wage of chief executive officers as executive pay is a nontrivial share of the denominator, a bias that has probably grown over time simply because CEO relative pay has grown.[2] For comparison purposes, Table 3 also shows the changes in the gross (not regression-adjusted) college–high school wage premium.

Between 1989 and 2010, compensation of chief executive officers grew from 1.14 to 2.06 times that of the incomes of the top 0.1 percent households, the metric Kaplan (2012b) employs to measure CEO pay relative to that of other

[2] In 2007, according to the Capital IQ database (tabulations kindly provided by Temple University professor Steve Balsam), there were 38,824 executives in publicly held firms. There were 9,692 in the top 0.1 of wage earners, a group whose average W-2 earnings were $4,400,028. Using our Mishel, Bivens, Gould, and Shierholz (2012) estimates of top 0.1 wages, the executive wages comprised 13.3 of total top 0.1 percent wages. One can gauge the bias of including executives in the denominator by noting that the ratio of executive wages to all top 0.1 percent wages in 2007 was 2.14 but the ratio of executive wages to nonexecutive wages was 2.32. Unfortunately, we do not have data that permit an assessment of the bias in 1979 or 1989. We also do not have information on the number and wages of executives in privately held firms; their inclusion would clearly indicate an even larger bias. The IRS reports there were nearly 15,000 corporate tax returns in 2007 of firms with assets exceeding $250 million, indicating there are many more executives of large firms than just those in publicly held firms.

Figure 2
Comparison of CEO Compensation to Top Incomes and Wages, 1947–2010



Sources: Author's analysis of Kaplan (2012b) and Mishel, Bivens, Gould, and Shierholz (2012, table 4.8).
Notes: Figure 2 presents the ratios displayed in Table 3 back to 1947. See Figure 2 and text for details.

highly paid people. CEO pay relative to top 0.1 percent wage earners grew even more, from 2.55 to 4.70, in that same time frame, a rise (2.15) equal to the pay of more than two very high earners. The log ratio of CEO relative pay grew roughly 60 log points from 1989 to 2010 using either top household income or wage earners as the comparison.

Is this a large increase? Kaplan (2012b, p. 14) does not find this increase in relative pay (that is, CEO pay relative to top household incomes) to be large, simply noting that the ratio "remains above its historical average and the level in the mid-1980s." To put this in context, Figure 2 presents the ratios displayed in Table 3 back to 1947. In fact, Kaplan's ratio of CEO pay to top household incomes in 2010 (2.06) was nearly double the historical average, 1.11, a gain roughly equivalent to the total income of a top 0.1 percent household. CEO pay relative to top wage earners in 2010 was 4.70 in 2010, 1.62 higher than the historical average (a gain of 1.5 high wage earners). As the data in Table 3 show, the increase in the logged CEO pay premium since 1979, and particularly since 1989, was far in excess of the rise in the college–high school wage premium which is widely and appropriately considered to have substantially grown. Presumably, CEO relative pay has grown further since 2010: as Table 2 showed, CEO compensation rose 15 percent between 2010 and

2012.[3] If checking whether CEO pay grows faster than that of other high earners is a way to test for the presence of rents, then we would conclude that executives receive substantial rents.

Kaplan (2012a, 2012b) also suggests that pay of chief executive officers grew on par with firm size in the last few decades and that this is evidence of market-determined pay. However, Frydman and Saks (2010, footnote 25) comment that "the strong correlation in more recent decades may be due to an upward trend in both variables instead of a causal effect of firm size on pay." They also note "the strong correlation between compensation and aggregate firm size was limited to the 1980s and 1990s. For all other decades in our sample, average market value accounts for less than 1 percent of the variation in executive pay." Further, as Gordon and Dew-Becker (2007) point out, the available evidence does not support a unitary elasticity between firm size and pay of chief executive officers, either historically back to 1936 (Frydman and Saks 2010) or in annual cross-sections from 1990 to 2004. Fernandes, Ferreira, Matos, and Murphy (2012) report an elasticity of 0.4 from pay of chief executive officers to firm size, which they report is in line with prior studies.

The Rising Incentive for Rent-Shifting

So far, we have argued that the evidence supports the case that rents are an important component of top 1 percent incomes. We have also argued that in some cases (particularly in the financial sector), evidence suggests that there have been increased opportunities for shifting rents to boost incomes and wages in recent decades. We admit that the case for a rise in rent-shifting as the primary driver of rising inequality is not yet ironclad, although it is certainly consistent with lots of evidence.

However, the case that *incentives* for rent-shifting have changed is completely unambiguous. From the late 1960s into the mid-1980s, in particular, top federal marginal income tax rates fell substantially. For example, the top marginal income tax rate was above 70 percent in 1970, but had fallen to 28 percent in 1986. Those extremely high marginal income tax rates in previous decades mainly applied to the upper slice of the top 1 percent of the income distribution—but as we have argued, the rise in incomes for that group is a major factor in increasing the share of income going to the top 1 percent.

The incentive effect of lower marginal tax rates could well affect top incomes if they are significantly composed of rents. For example, in the model of Bebchuk and Fried (2004), well-placed individuals who have some ability to shift rents will balance the costs and benefits of exerting more influence to boost their own incomes, where

[3] This measure of compensation uses realized value of options rather than the estimated value of options granted. Estimated pay (using options granted rather than options realized) actually fell 4 percent between 2010 and 2012. Kaplan (2012b, p. 22) argues that realized pay is the preferred measure of performance: "Critics confuse estimated pay—what the boards give to the CEOs as estimated pay—and realized pay. The key question is whether CEOs who perform better earn more in realized pay." In this paper, we focus on Kaplan's preferred measure of "realized" pay, but the qualitative pattern is much the same if looking at "estimated" pay over the entire time period.

one of the costs is whether they encounter an "outrage constraint." Lower marginal tax rates at top income levels will provide a greater incentive for well-placed individuals at that income level to spend more energy on rent-shifting.

Piketty, Saez, and Stantcheva (2012) have shown that the link between falling marginal rates and higher pre-tax top 1 percent shares is significant both in time-series data for the US as well as across countries. In their paper in this symposium, Alvarado, Atkinson, Piketty, and Saez discuss these issues.

Policy and Institutional Changes beyond the Top 1 Percent

We have been arguing so far mostly within a framework amendable to microeconomists—at least microeconomists of the broad-minded and institutionalist variety—positing that developments within specific sectors and occupational labor markets have boosted the ability of well-placed groups to redistribute rents their way. However, the levers of rent-shifting can include both changes that shift bargaining power to those at the top of income distribution generally and changes that subvert the bargaining power of those at the bottom and middle.

In Mishel, Bivens, Gould, and Shierholz (2012), we document the ways in which a range of policy developments over the last generation have disproportionately damaged the wage prospects of low- and moderate-wage workers, including the declining real value of the minimum wage and the failure to update labor law to provide a level playing field in the face of growing employer hostility to union organizing efforts. Indeed, many of these policy changes were intentional and pursued with *much* greater vigor in the last generation than the previous one (Hacker and Pierson 2010). There is considerable evidence that these kinds of institutional changes can shift rents.[4] Too often the assumption is that policy variables like the real value of the minimum wage cannot be relevant to top 1 percent incomes as they are, by definition, nonbinding on high wages. Yet one person's income is another person's cost. If a declining value of the minimum wage, or increased effectiveness in blocking union organizing, keeps wages in check at, say, Walmart, then it is hardly a shock that this could well lead to higher pay for corporate managers and higher returns to Walmart shareholders (for example, Draca, Machin, and Van Reenen, 2011, offer evidence that in the UK, higher minimum wages reduce firm profitability—but with no significant impact on employment).

[4] Many studies, while not focusing on the top 1 percent, are strongly consistent with the interpretation that institutional changes can shift economic rents, which in turn can affect the level of inequality. For example, studies of the minimum wage line by DiNardo, Fortin, and Lemeiux (1996), Lee (1999), and Autor, Manning, and Smith (2010) have identified strong effects of the minimum wage in driving "lower-tail" inequality, while numerous other studies like Card and Krueger (1995), Allegretto, Dube, and Reich (2011), and Manning (2003) have not found employment losses following these increases in the minimum wage. The combination of these results strongly suggests that the primary effect of minimum wage increases is to redistribute economic rents, rather than to affect employment levels. Levy and Temin (2007) have identified the breakdown of a range of rent-shifting institutions, which they shorthand "The Treaty of Detroit," as driving inequality between the top (roughly 90th percentile) and middle of the wage and income distribution.

Further, it is likely that the role of globalization—a mixture of exogenous and policy-induced changes—also looms large. Textbook Stolper–Samuelson models explicitly show (at least in the older textbooks!) that trade openness can increase capital incomes and reduce labor compensation in rich countries like the United States. Rodrik (1999) and Jayadev (2007) have similarly noted that capital account openness, which is largely a policy choice, could well tilt bargaining power away from workers and towards capital-owners, resulting in higher capital shares not just in developed countries (the standard Stolper–Samuelson result) but in developing countries as well—a nonstandard result that has shown up strongly in the data.

Will Putting a Brake on Top 1 Percent Growth Harm *Overall* Growth Rates?

If the rise in top 1 percent incomes has accrued largely from shifting rents, then multiple possibilities exist for redistributing these rents without slowing overall economic growth or distorting economic efficiency. Successful redistributions would then translate directly into increased living standards for low- and moderate-income households. Besides the evidence assembled above indicating that the growth of these incomes are largely rents, a number of recent studies have looked directly at the issue of shifting top shares on overall economic growth. For the US economy, the broad historical pattern is a strong association between stable top income shares and faster overall growth in early post-World War II economic history, followed by rising top income shares and notably slower growth in the three decades before the Great Recession. This broad association between greater inequality and less growth is clearly not reversed in systematic attempts to establish a link between rising top shares and aggregate economic performance, nor by looking at international or state-level data.

For example, Piketty, Saez, and Stantcheva (2012) examine the relationship between top marginal tax rates, top income shares, and aggregate economic performance, both in US time-series as well as using an international panel of 18 OECD countries. They find strong evidence that falling top marginal tax rates are associated with higher pre-tax top income shares. However, they do not find a strong association either between falling top marginal rates and rising economic growth, or (for the US data) rising top income shares and faster economic growth. They also find significant evidence that falling top marginal tax rates are associated with slower income growth for the bottom 99 percent of households. They take this constellation of evidence as supporting a "bargaining model where gains at the top have come at the expense of the bottom."

Andrews, Jencks, and Leigh (2011) find slightly mixed evidence on the larger issue of top shares and subsequent growth, with increases in the share of income accruing to the top 10 percent positively related (and generally statistically significantly across regression specifications) to subsequent overall growth in their preferred regression models. They note the modest economic impact implied by

their results: "But at the very least, the 95 percent confidence intervals for our preferred estimates appear to rule out the claim that a rise in top income shares causes a large short-term increase or decrease in economic growth. The claim that inequality at the top of the distribution either benefits or harms everyone therefore depends on long-term effects that we cannot estimate very precisely even with these data." Most importantly for the question at hand, these results are driven by what is happening between the 90th and 99th percentiles. They note: "The top 1 percent's share is never both positively and significantly related to the growth rate."

Thompson and Leight (2012) have recently used a different sort of panel to examine the relationship between top income shares and growth—looking at the top 1 percent within individual US states. Their analysis finds that rising top 1 percent income shares are associated with *falling* subsequent growth in incomes and earnings for households in the middle of the distribution while having no significant effect on growth at the bottom of the distribution. Further, their finding on the statistical significance of the depressing effects of rising top shares on middle-incomes is fairly robust and survives the inclusion of a range of covariates (though the economic impact is still relatively modest).

Some advocates for reversing the rise of the income share of the top 1 percent occasionally make strong claims that the rise at the top has harmed overall economic growth. Our claim here is more modest: the empirical evidence that has directly examined the effect of rising top 1 percent shares on overall economic growth certainly does not suggest that they are strongly and robustly associated. But as long as the shift to the top 1 percent is not associated with *improved* growth, then the rest of the income distribution is harmed by this increase in top 1 percent shares.

What to Do? Attack Rents Directly and Raise Taxes

We think the suggestive evidence that the rise in top 1 percent shares stems from the creation and/or redistribution of rents, as well as the direct evidence that changing shares in the US economy do not seem to affect aggregate outcomes significantly, means that there is ample room for policymakers to act to stabilize or reverse top 1 percent shares. Taking much more ambitious steps (so long as they are intelligently directed) to halt or reverse the concentration of income at the very top will not kill any golden goose of economic growth. Instead, it will just lead to more income for those at the bottom and middle of the income distribution.

As we see it, there are two broad categories to describe what can be done about the rise of the top 1 percent: try to attack the source of their ability to shift rents directly and try to reduce the incentives for rent-shifting.

Attacking the source of the top 1 percent's ability to claim rents means acting on a long laundry list of policy changes. Baker (2011) provides a compelling argument behind many of these needed changes: Corporate governance reform that gives not just shareholders but other stakeholders as well real influence over executive pay decisions (for example, DiNardo, Hallock, and Pischke, 1997, find that

unionized firms more successfully restrict managerial pay); reform to ensure that financial firms are less likely to seek profits by hiding risks or exploiting information asymmetries; reconstituting labor standards that boost bargaining power at the low and middle-end of the wage-scale (higher minimum wages and labor law reform that allows willing workers to bargain collectively if they choose); the dedicated pursuit of genuinely full employment; and reform of intellectual property law that greatly reduces the legal monopoly granted to sectors like pharmaceuticals, software, medical devices, and entertainment.

Attacking the sources of rent-shifting at the top may also provide additional benefits. If corporate governance reforms better align the incentives of managers and important stakeholders, this could be efficiency enhancing. If reform of intellectual property laws reduces the price of pharmaceuticals and medical devices, money will be saved by ultimate consumers of health care.

Reducing the *incentives* for the top 1 percent to shift rents can be achieved with a much shorter list: significantly raising the marginal tax rates on high incomes, including a reduction in the current gap in taxation of labor earnings (which we should note includes the return to *human* capital) versus income derived from ownership of financial capital. Raising these marginal rates would also address other problems in American political economy—closing long-run fiscal gaps and providing revenue needed (during times of full-employment) to undertake productive public investment and maintain social insurance programs. Importantly, a rise in top tax rates—so long as they keep after-tax income rising with pre-tax income—would redistribute rents without harming economic efficiency even in cases where rents coexist with an efficient allocative equilibrium.

More specifically, the tax benefit received by corporations that allow them to deduct even executive compensation in excess of $1 million annually so long as it is "performance-based" has given a quasi-official blessing to executive pay practices like stock options and bonuses. But as noted before, such practices are not guarantees that executive pay and performance are tightly linked, and noncash compensation often serves to disguise the true extent of corporate executive pay. Balsam (2012) presents a full discussion of this provision and its effectiveness. In short, this tax preference clearly does not appear to spur better corporate governance, so it should be repealed.

In short, there is much to recommend both broad courses of action, and there's no reason for those concerned about the rise of income concentration at the very top to choose only one route. The economic stakes are large: in the last generation, the large rise in income inequality has been the primary barrier to low- and moderate-income households seeing growth in their living standards that comes close to matching *overall* income growth rates. In previous work (Mishel et al. 2012), we show that cumulative growth in middle-fifth comprehensive incomes in between 1979 and 2007 would have been more than twice as rapid if they had matched overall average rates (19.2 percent actual versus 51.5 percent overall growth). Further, we find that nearly 60 percent of the cumulative gap between growth of the middle-quintile and overall average growth (which we have labeled an implicit "inequality

tax" on these middle-quintile incomes) between 1979 and 2007 can be accounted for *solely* by growth of the top 1 percent. In short, the stakes are large in debates about why top 1 percent incomes have grown so fast and what can be done about it.

■ *Thanks to David Autor, Dean Baker, Jacob Hacker, Chang-Tai Hsieh, and Ulrike Malmendier for comments on a previous draft, and to Timothy Taylor for very useful suggestions on both substance and organization. The research on CEO compensation was made possible by financial support from the Stephen Silberstein Foundation.*

References

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): 205–240.

Anderson, Gerald F., Uwe E. Reinhardt, Peter S. Hussey, and Varduhi Petrosyan. 2003. "It's the Prices, Stupid: Why the United States Is So Different from Other Countries." *Health Affairs* 22(3): 89–105.

Andrews, Dan, Christopher Jencks, and Andrew Leigh. 2011. "Do Rising Top Incomes Lift All Boats?" *B.E. Journal of Economic Analysis & Policy* 11(1).

Autor, David H., Alan Manning, and Christopher L. Smith. 2010. "The Contribution of the Minimum Wage to U.S. Wage Inequality over Three Decades: A Reassessment." http://economics.mit.edu/files/3279.

Baker, Dean. 2011. *The End of Loser Liberalism: Making Markets Progressive.* Washington, DC: Center for Economic Policy Research.

Baker, Dean, and McArthur, Travis. 2009. "The Value of the 'Too Big to Fail' Big Bank Subsidy." Center for Economic and Policy Research Issue Brief. September. http://www.cepr.net/documents/publications/too-big-to-fail-2009-09.pdf.

Bakija, Jon, Adam Cole, and Bradley T. Heim. 2010. "Job and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence from U.S. Tax Return Data." http://www.williams.edu/Economics/wp/BakijaColeHeimJobsIncomeGrowthTopEarners.pdf

Balsam, Steven. 2012. "Taxes and Executive Compensation." August 14, Economic Policy Institute Briefing Paper 344.

Bebchuk, Lucian A., Alma Cohen, and Holger Spamann. 2009. "The Wages of Failure: Executive Compensation at Bear Stearns and Lehman 2000–2008." Harvard Law and Economics Discussion Paper 657.

Bebchuk, Lucian, and Jesse Fried. 2004. *Pay Without Performance: The Unfulfilled Promise of Executive Renumeration.* Cambridge, MA: Harvard University Press.

Bebchuk, Lucian A., and Jesse M. Fried. 2010. "How to Tie Equity Compensation to Long-term Results." *Journal of Applied Corporate Finance* 22(1): 906–106.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Are CEOs Rewarded for Luck? The Ones without Principals Are." *Quarterly Journal of Economics* 116(3): 901–932.

Biais, Bruno, Jean-Charles Rochet, and Paul Woolley. 2010. "Innovations, Rents and Risk." The Paul Woolley Centre Working Paper Series No. 13, Discussion Paper 659. http://www2.lse.ac.uk/fmg/researchProgrammes/paulWoolleyCentre/workingPapers/dp659PWC13.pdf.

Bizjak, John, Michael Lemmon, and Thanh Nguyen. 2011. "Are All CEOs above Average? An Empirical Analysis of Compensation Peer Groups and Pay Design." *Journal of Financial Economics* 100(3): 538–55.

Bronfenbrenner, Kate. 2000. "Uneasy Terrain: The Impact of Capital Mobility on Workers, Wages,

and Union Organizing." Submitted to the US Trade Deficit Review Commission. http://govinfo.library.unt.edu/tdrc/research/bronfenbrenner.pdf.

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

Congressional Budget Office. 2012. "The Distribution of Household Income and Federal Taxes, 2008 and 2009." Congressional Budget Office Report 4441. http://www.cbo.gov/sites/default/files/cbofiles/attachments/43373-06-11-HouseholdIncomeandFedTaxes.pdf.

Cutler, David M., and Louise Sheiner. 1999. "The Geography of Medicare." *American Economic Review* 89(2): 228–33.

Davis, Karen, Cathy Schoen, Stuart Guterman, Tony Shih, Stephen C. Schoenbaum, and Ilana Weinbaum. 2007. "Slowing the Growth of U.S. Health Care Expenditures: What are the Options?" Prepared for The Commonwealth Fund/Alliance for Health Reform 2007 Bipartisan Congressional Health Policy Conference. http://www.commonwealthfund.org/usr_doc/Davis_slowinggrowthUShltcareexpenditureswhatareoptions_989.pdf.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux. 1996. "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach." *Econometrica* 64(5): 1001–1044.

DiNardo, John, Kevin Hallock, and Jorn-Steffan Pischke. 1997. "Unions and Managerial Pay." NBER Working Paper 6318.

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

Fernandes, Nuno G., Miguel A. Ferreira, Pedro P. Matos, and Kevin J. Murphy. 2012. "Are US CEOs Paid More? New International Evidence." EFA 2009 Bergen Meetings Paper; AFA 2011 Denver Meetings Paper; ECGI - Finance Working Paper No. 255/2009. http://ssrn.com/abstract=1341639.

Frydman, Carola, and Raven E. Saks. 2010. "Executive Compensation: A New View from a Long-Term Perspective, 1936–2005." *Review of Financial Studies* 23(5): 2099–2138.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123(1): 49–100.

Goldin, Claudia, and Lawrence F. Katz. 2008. "The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005." NBER Working Paper 12984.

Gordon, Robert J., and Ian Dew-Becker. 2007. "Selected Issues in the Rise of Income Inequality." *Brookings Papers on Economic Activity,* no. 2, pp. 169–190.

Hacker, Jacob, and Paul Pierson. 2010. *Winner-Take-All Politics: How Washington Made the Rich Richer—and Turned Its Back on the Middle Class*. New York, NY: Simon and Schuster.

Haldane, Andrew G. 2010a. "The $100 Billion Question." Comments given at the Institute of Regulation & Risk, Hong Kong, March 30, 2010. http://www.bankofengland.co.uk/publications/Documents/speeches/2010/speech433.pdf.

Haldane, Andrew. 2010b. "The Contribution of the Financial Sector: Miracle or Mirage?" Speech prepared for the Future of Finance Conference, London, July 14, 2010. Text is taken from a coauthored chapter by Andrew Haldane, Simon Brennan, and Vasileios Madouros in "The Future of Finance: The LSE Report," published in 2010 by the London School of Economics. http://www.bankofengland.co.uk/publications/Documents/speeches/2010/speech442.pdf.

Jacquette, Scott, Matthew Knittel, and Karl Russo. 2003. "Recent Trends in Stock Options." OTA Working Paper 89, Office of Technical Assistance, US Department of the Treasury.

Jayadev, Arjun. 2007. "Capital Account Openness and the Labour Share of Income." *Cambridge Journal of Economics* 31(3): 424–43.

Kaplan, Steven N. 2012a. "Executive Compensation and Corporate Governance in the U.S.: Perceptions, Facts, and Challenges." Martin Feldstein Lecture given July 10.

Kaplan, Steven N. 2012b. "Executive Compensation and Corporate Governance in the U.S.: Perceptions, Facts and Challenges." NBER Working Paper 18395.

Kedrosky, Paul, and Dane Stangler. 2011. "Financialization and Its Entrepreneurial Consequences." Kauffman Foundation Research Series: Firm Formation and Economic Growth, March. http://www.kauffman.org/uploadedfiles/financialization_report_3-23-11.pdf.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *Quarterly Journal of Economics* 125(1): 91–128.

Kruse, Douglas, Joseph Blasi, and Richard Freeman. 2011. "Does Shared Capitalism Help the Best Firms Do Even Better?" Prepared for British Academy Seminar "The Economics of Share Ownership and Gainsharing: Findings and Policy Implications." Centre for Economic Performance, London School of Economics, May 26, 2011.

Laeven, Luc, and Fabian Valencia. 2008. "Systemic Banking Crises: A New Database." IMF Working Paper WP/08/224. http://www.imf.org/external/pubs/ft/wp/2008/wp08224.pdf.

Laugesen, Miriam J., and Sherry A. Glied. 2011.

"Higher Fees Paid To US Physicians Drive Higher Spending for Physician Services Compared to Other Countries." *Health Affairs* 30(9): 1647–56.

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.

Levy, Frank, and Peter Temin. 2007. "Inequality and Institutions in 20th Century America." Industrial Performanace Center, Massachusetts Institute of Technology Working Paper Series, MIT-IPC -07-0002. http://web.mit.edu/ipc/publications /pdf/07-002.pdf.

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

Mishel, Lawrence, Jared Bernstein, and Sylvia Allegretto. 2004. *The State of Working America: 2004–05*. An Economic Policy Institute book. Ithaca, NY: Cornell University Press.

Mishel, Lawrence, Josh Bivens, Elise Gould, and Heidi Shierholz. 2012. *The State of Working America, 12th Edition*. An Economic Policy Institute Book. Ithaca, NY: Cornell University Press.

Mishel, Lawrence, and Natalie Sabadish. 2012. "CEO Pay and the Top 1%." Economic Policy Institute Issue Brief 331.

Mishel, Lawrence, and Natalie Sabadish. 2013. "CEO Pay in 2012" Economic Policy Institute Issue Brief 367. Economic Policy Institute, Washington, DC.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1991. "The Allocation of Talent: Implications for Growth." *Quarterly Journal of Economics* 106(2): 503–30.

Philippon, Thomas, and Ariell Reshef. 2009. "Wages and Human Capital in the U.S. Financial Industry: 1909–2006." NBER Working Paper 14644.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1): 1–41. Tables and figures updated to 2010 in Excel format, March 2012).

Piketty, Thomas, Emmanuel Saez, and Stefanie Stantcheva. 2011, revised 2012. "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities." NBER Working Paper 17616.

Rodrik, Dani. 1999. "Globalization and Labor, Or: If Globalization is a Bowl of Cherries, Why Are There so Many Glum Faces Around the Table?" Chap. 5 in *Market Integration, Regionalism and the Global Economy*, edited by Richard E. Baldwin, Daniel Cohen, André Sapir, and Anthony Venables. Cambridge University Press for CEPR, New York.

Schwab, Stewart J., and Randall S. Thomas. 2006. "An Empirical Analysis of CEO Employment Contracts: What Do Top Executives Bargain For?" *Washington and Lee Law Review* 63(1): Article 6.

Sherman, Matthew. 2009. "A Short History of Financial Deregulation in the United States." Report, July 2009. Center for Economic and Policy Research, Washington, DC. http://www .cepr.net/documents/publications/dereg -timeline-2009-07.pdf.

Terviö, Marko. 2008. "The Difference That CEOs Make: An Assignment Model Approach." *American Economic Review* 98(3): 642–68.

Terviö, Marko. 2009. "Superstars and Mediocrities: Market Failure in the Discovery of Talent." *Review of Economic Studies* 76(2): 21–49.

Thompson, Jeffrey P., and Elias Leight. 2012. "Do Rising Top Income Shares Affect the Incomes or Earnings of Low and Middle-Income Families?" *B. E. Journal of Economic Analysis & Policy* 12(1): 1–38.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 79–102

Income Inequality, Equality of Opportunity, and Intergenerational Mobility

Miles Corak

The "American Dream" is a very broad concept with many meanings and certainly broader than any single statistic can measure. However, when the Economic Mobility Project (2009), supported by the Pew Charitable Trusts, conducted a nationally representative poll that asked Americans what they understood this phrase to mean, some typical answers included: "Being free to say or do what you want"; "Being free to accomplish almost anything you want with hard work"; and "Being able to succeed regardless of the economic circumstances in which you were born." These meanings have historically not only made the American Dream a defining metaphor of the country, they are also likely a reason why Americans have been willing to tolerate a good deal more inequality of outcomes than citizens of many other rich countries. Bénabou and Ok (2001) have called this the "prospect of upward mobility" hypothesis, the idea that those with lower incomes are not especially strong advocates of redistributive policies because of the belief that they, or in the least their children, are likely to climb the income ladder.

However, an emerging body of evidence suggests that more inequality of incomes in the present is likely to make family background play a stronger role in determining the adult outcomes of young people, with their own hard work playing a commensurately weaker role. The OECD (2011a, p. 40) has gone so far as to state that rising income inequality "can stifle upward social mobility, making it harder for talented and hard-working people to get the rewards they deserve.

■ *Miles Corak is Professor of Economics, University of Ottawa, Ottawa, Canada, and Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany. He can be reached at http://milescorak.com or through Twitter @MilesCorak.*

http://dx.doi.org/10.1257/jep.27.3.79 doi=10.1257/jep.27.3.79

Intergenerational earnings mobility is low in countries with high inequality such as Italy, the United Kingdom, and the United States, and much higher in the Nordic countries, where income is distributed more evenly."

This suggestion that higher inequality skews opportunity and lowers intergenerational mobility is the starting point of this paper. In particular, my focus is on the degree to which increasing inequality in the high-income countries, particularly in the United States, is likely to limit economic mobility for the next generation of young adults.

The paper offers a descriptive, yet structured, discussion of the underlying drivers of opportunity that generate the relationship between inequality and intergenerational mobility. The goal is to explain why America differs from other countries, how intergenerational mobility will change in an era of higher inequality, and how the process is different for the top 1 percent. To lay the foundation, I begin by presenting evidence that countries with more inequality at one point in time also experience less earnings mobility across the generations, a relationship that has been called "The Great Gatsby Curve." I also outline how to interpret the common statistic measuring intergenerational earnings mobility and its relationship to the broader concept of equality of opportunity. My overview of the causal factors determining intergenerational mobility is based upon a framework drawn from some influential economic models often used to examine the intergenerational transmission of inequality. This framework focuses attention on the investments made in the human capital of children influencing their adult earnings and socioeconomic status.

The interaction between families, labor markets, and public policies all structure a child's opportunities and determine the extent to which adult earnings are related to family background—but they do so in different ways across national contexts. Both cross-country comparisons and the underlying trends suggest that these drivers are all configured most likely to lower, or at least not raise, the degree of intergenerational earnings mobility for the next generation of Americans coming of age in a more polarized labor market. This trend will likely continue unless there are changes in public policy that promote the human capital of children in a way that offers relatively greater benefits to the relatively disadvantaged. At the same time, the substantial rise in the income shares of the top 1 percent, their access to sources of high-quality human capital investment for their children, and the intergenerational transmission of employers and wealth will imply a much higher rate of transmission of economic advantage at the very top, in a way that many will perceive as evidence of inequality in opportunity.

The Great Gatsby Curve

Countries with greater inequality of incomes also tend to be countries in which a greater fraction of economic advantage and disadvantage is passed on between parents and their children. It is now common to represent this relationship with

what Alan Krueger has referred to as "The Great Gatsby Curve."[1] Figure 1 depicts an example.

The figure ranks countries along two dimensions. The horizontal axis shows income inequality in a country as measured by the Gini coefficient from about a generation ago. During the early to mid 1980s, Finland, Sweden, Norway, and Denmark were the most equal; the United Kingdom and the United States, the least. The vertical axis is a measure of intergenerational economic mobility: specifically, the elasticity between paternal earnings and a son's adult earnings using data on a cohort of children born, roughly speaking, during the early to mid 1960s and measuring adult outcomes in the mid to late 1990s. In countries like Finland, Norway, and Denmark, the tie between parental economic status and the adult earnings of children is weakest: less than one-fifth of any economic advantage or disadvantage that a father may have had in his time is passed on to a son in adulthood. In Italy, the United Kingdom, and the United States, roughly 50 percent of any advantage or disadvantage is passed on.

The "intergenerational earnings elasticity," is derived from a regression-to-the-mean model, usually as the least-squares estimate of the coefficient β in the equation

$$\ln Y_{i,t} = \alpha + \beta \ln Y_{i,t-1} + \varepsilon_i,$$

with Y representing "permanent earnings" for individuals from a particular family indexed by i, across two generations, t and $t-1$. In much of the literature, Y refers to the earnings of fathers and sons to avoid the more complicated analyses needed to address the changing role of women in the labor force. It is not that studies of mothers, daughters, and the marriage market do not exist, only that father–son analyses are more common and permit a broader set of cross-country comparisons. In this equation, ε represents all other influences on the child's adult earnings not correlated with parental income. The constant term α captures the trend in average incomes across generations, due, for example, to changes in productivity, international trade, technology, or labor market institutions. The coefficient β indicates the degree to which earnings are "sticky" across generations within the same family, the percentage difference in child earnings for each percentage point difference in parental earnings. The higher the value of β, the more that knowing a parent's place in the earnings distribution will tell us about where we can expect the child's place to be; the lower the value, the less stickiness so that a parent's relative earnings are a weak predictor of the child's rung on the earnings ladder of the next generation. Mulligan (1997) offers a more detailed description of how this model should be interpreted.

[1] Krueger used this label for the first time in a speech, "The Rise and Consequences of Inequality," to the Center for American Progress on January 12, 2012, in his capacity as the Chairman of the Council of Economic Advisors. The curve has been drawn in different varieties by, to the best of my knowledge, Andrews and Leigh (2009), Björklund and Jäntti (2009), Blanden (2013), and myself (Corak 2006, 2013), Ermisch, Jäntti, Smeeding, and Wilson (2012) in addition to appearing in Krueger's speech and the 2012 *Economic Report of the President*, which stress the relevance of measuring inequality when the children are growing up.

Figure 1
The Great Gatsby Curve: More Inequality is Associated with Less Mobility across the Generations

Source: Corak (2013) and OECD.
Notes: Income inequality is measured as the Gini coefficient, using disposable household income for about 1985 as provided by the OECD. Intergenerational economic mobility is measured as the elasticity between paternal earnings and a son's adult earnings, using data on a cohort of children born, roughly speaking, during the early to mid 1960s and measuring their adult outcomes in the mid to late 1990s. The estimates of the intergenerational earnings elasticity are derived from published studies, adjusted for methodological comparability in a way that I describe in the appendix to Corak (2006), updated with a more recent literature review reported in Corak (2013), where I also offer estimates for a total of 22 countries. I only use estimates derived from data that are nationally representative of the population and which are rich enough to make comparisons across generations within the same family. In addition, I only use studies that correct for the type of measurement errors described by Atkinson, Maynard, and Trinder (1983), Solon (1992), and Zimmerman (1992), which means deriving permanent earnings by either averaging annual data over several years or by using instrumental variables.

Figure 1, showing the relationship between income inequality and intergenerational economic mobility, uses estimates of the intergenerational earnings elasticity derived from published studies that I adjust for differences in methodological approach (see notes to the figure for details). So these estimates are offered, not as the best available estimates for any particular country, but rather as the appropriate estimates for comparisons across countries. (Analyzing a broader group of countries, I find that many of the lower-income countries occupy an even higher place on the Great Gatsby Curve than depicted for the OECD countries in Figure 1, but this is likely due to structural factors not as relevant to a discussion of the high-income countries.)

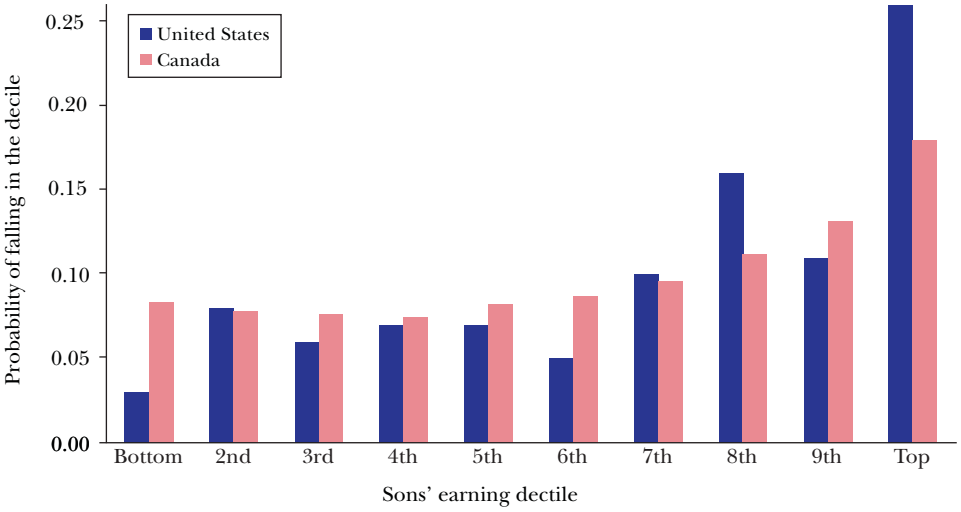
There are certainly many other ways of measuring intergenerational mobility that focus attention upon particular aspects of the process. The intergenerational elasticity, for example, offers an overall average measure of the degree of mobility without saying anything about the direction of change. The cross-country differences illustrated in the Great Gatsby Curve could reflect differences in the degree of upward mobility for those born to low-income fathers, or differences in the stickiness of intergenerational status for those born to top income parents. Figures 2 and 3 contrast, by way of illustration, mobility in the United States and Canada for sons raised by fathers with incomes in the top 10 percent, and for those raised by fathers in the bottom 10 percent. In Corak (2010), I argue that the comparison of these two countries is particularly apt in part because of similarities in the underlying data used, but also because they share many other things in common, an issue to which I return below.

It turns out that the American intergenerational earnings elasticity, at about 0.5, is twice as high as the Canadian, and this has little to do with the degree of mobility of children raised by families in broad swaths of the middle part of the distribution. Indeed, a comparison of the full decile transition matrices reveals a good deal of mobility in both countries, to the point that there is little relationship between family background and child outcomes (Corak and Heisz 1999; Mazumder 2005). It is at the extremes of the distribution that the two countries differ, as illustrated in Figures 2 and 3. In the United States, sons raised by top and bottom decile fathers are more likely to occupy the same position as their fathers than they are in Canada. More than half of sons raised by top decile American fathers fall no further than the 8th decile, and about half of those raised by bottom decile fathers rise no further than the third decile. In Canada, there is less stickiness at the top, and a much higher proportion of bottom decile sons also rise to the top half of the earnings distribution.

All this said, if one number is to summarize the degree to which inequality is transmitted across the generations, just as sometimes one number, like a Gini coefficient, is used to summarize the degree of inequality at a point in time, then the intergenerational elasticity is an appropriate statistic to use. But this does not mean that it measures "equality of opportunity" or the even-more-elusive "American Dream." Roemer (2004, 2012) and Jencks and Tach (2006), among others, are clear on this point, emphasizing that in no sense is an intergenerational elasticity of zero an optimum and noting that in order to make an inference about equality of opportunity from the degree of intergenerational earnings mobility we must draw a line between differences in *circumstances*—for which individuals should in some sense be compensated—and *personal choices*, for which they should be responsible.

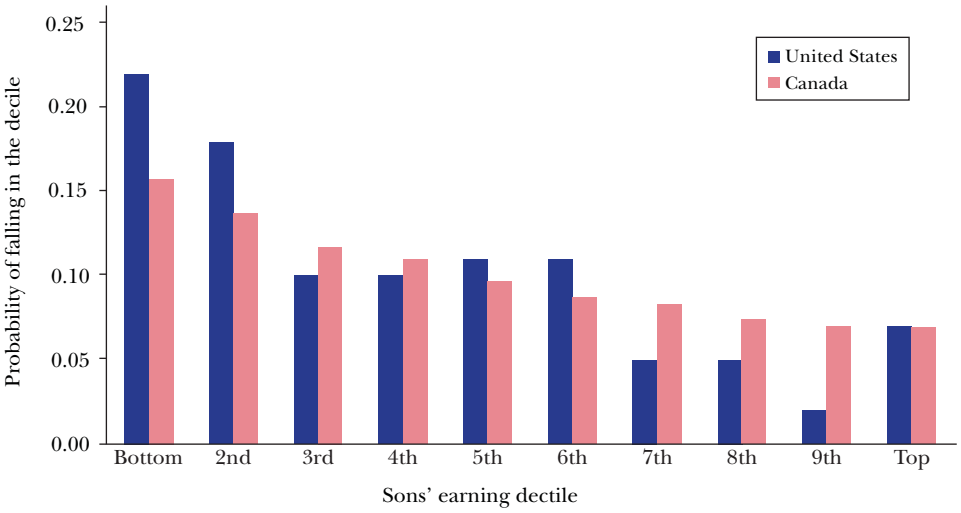
Another branch of the empirical literature seeks to draw this line by deriving indices of equality of opportunity that remove the influence of factors over which individuals have no control: for example, race, mother and father's schooling, region of birth, and father's occupation (Ferreira and Gignoux 2011; Lefranc, Pistolesi, and Trannoy 2008; Paes de Barros, Ferreira, Molinas Vega, and Saavedra Chanduvi 2009). Brunori, Ferreira, and Peragine (2013) offer a particularly clear

Figure 2
Earnings Deciles of Sons Born to Top Decile Fathers: United States and Canada



Source: Corak and Heisz (1999, table 6); Mazumder (2005, table 2.2).

Figure 3
Earnings Deciles of Sons Born to Bottom Decile Fathers: United States and Canada



Source: Corak and Heisz (1999, table 6); Mazumder (2005, table 2.2).

overview and find that indices of inequality of opportunity are in fact strongly correlated with indicators of intergenerational mobility, be it in earnings or in education. It is in this sense that the Great Gatsby Curve can be understood to be signaling—rather than directly measuring—a negative relationship between inequality and equality of opportunity. As these authors put it (p. 17): "[I]nequality of opportunity is the missing link between the concepts of income inequality and social mobility; if higher inequality makes intergenerational mobility more difficult, it is likely because opportunities for economic advancement are more unequally distributed among children."

A Framework for Comparisons across Space and Time

The Great Gatsby Curve is not a causal relationship, but it is too glib to dismiss it by saying "correlation does not imply causation." Theories of child development and economic mobility suggest it is reasonable to juxtapose measures of inequality and mobility as a starting point for understanding the causal process and its policy implications.

The adult outcomes of children reflect a series of gradients between their attainments at specific points in their lives and the prevailing socioeconomic inequalities to which they are exposed. Knudsen, Heckman, Cameron, and Shonkoff (2006), to cite only one example, summarize the literature on child development—and in particular some of the work by James Heckman—in a way that relates child development to adult social and labor market outcomes through a recursive process. Socioeconomic status influences a child's health and aptitudes in the early years—indeed even in utero—which in turn influences early cognitive and social development, and readiness to learn. These outcomes and the family circumstances of children, as well as the quality of neighborhoods and schools, influence success in primary school, which feeds into success in high school and college. Family resources and connections affect access to good schools and jobs, and the degree of inequality in labor markets determines both the resources parents have and ultimately the return to the education children receive. This entire process then shapes earnings in adulthood. The Great Gatsby Curve is a summary of all of these underlying gradients, reflecting the outcome of a host of ways that inequality of incomes affects children.

The usual starting points for discussions of causality are the models developed by Becker and Tomes (1986, 1979). Solon (2004) has adapted their research in a way appropriate for making comparisons across countries and over time. Very broadly speaking, the reasons for the differences in the intergenerational elasticity across countries have to do with the different balances struck between the influence of families, the labor market, and public policy in determining the life chances of children. These institutions determine the degree to which traits valuable in the labor market are passed on between parents and children, the efficacy of private and public investments in generating human capital, and the labor market returns

to human capital, all of which are important drivers of the degree of intergenerational income mobility.

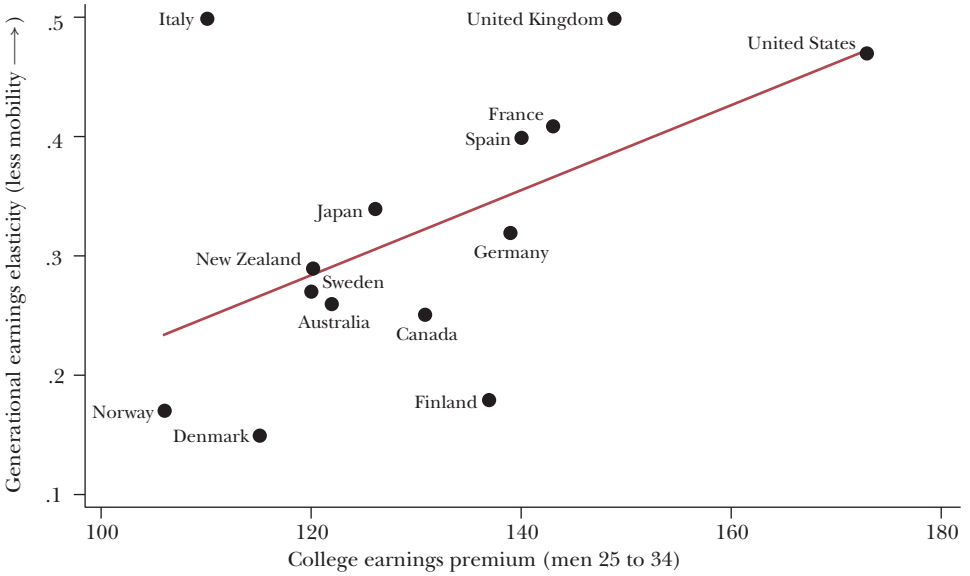
One perspective on the Great Gatsby Curve focuses on the heritability of traits between parents and their children. Becker and Tomes (1979, p. 1158) refer to these as "endowments of capital that are determined by the reputation and 'connections' of their families, the contribution to the ability, race, and other characteristics of children from the genetic constitutions of their families, and the learning, skills, goals, and other 'family commodities' acquired through belonging to a particular family culture." If these traits are strongly transmitted across generations, and if they are valued by labor markets over time, then there will also be an intergenerational association of incomes. At the same time, if there is significant demographic diversity across countries then we should not be surprised that there is an upward slope to the Great Gatsby Curve even if all societies are equally meritocratic. Becker (2013) and Roemer (2012) clearly articulate this interpretation. For this reason we should not think of the Great Gatsby Curve as a recipe for changing outcomes. Rather, it invites us to look at the differences between countries to appreciate the underlying drivers, assess the extent to which they are relevant for public policy, and in this sense recognize that some comparisons are more appropriate than others. Denmark, with a small relatively homogenous population, may not be a template for a large demographically diverse country like the United States; but a comparison of the United States with Canada, a diverse country sharing many fundamental values and institutions with America, may indeed be more appropriate.

Another perspective on the Great Gatsby Curve derives from the focus Solon (2004) places on the returns to education. He takes the rate of return to schooling as an indicator of the degree of inequality in the labor market and shows that societies with labor markets characterized by more cross-sectional inequality—reflecting in part a higher return to education—will be less generationally mobile. Parents with more human capital not only have a higher capacity to invest in the education of their children by virtue of their higher incomes, but also the incentives to do so are greater.

Furthermore, Solon (2004) also suggests that public policy can either accentuate or dampen the influence of labor market inequality, showing that intergenerational mobility is promoted by "progressive" public programs that are of relatively more benefit to the relatively less well-off. Two countries may spend the same fraction of their gross domestic product on education, but if this spending is directed to high-quality early childhood education and to primary and secondary schooling accessible for all, then it is likely to be of relatively more benefit to families lower in the socioeconomic scale than if it were directed to high-quality private tertiary education accessible to only a few. Indeed, this perspective should be applied to all aspects of public actions that influence the relationship between families and the labor market, which in addition to expenditures on schooling include other sources of human capital like health care, taxes and transfers, as well as regulations and policies helping parents to balance work and family life.

Figure 4

Higher Returns to Schooling are Associated with Lower Intergenerational Earnings Mobility



Source: Author using data from OECD (2011b, table A8.1), and Corak (2013).
Notes: The earnings premium refers to the ratio of average earnings of men 25 to 34 years of age with a college degree to the average earnings of those with a high school diploma. This is measured as the average employment income in 2009 of men 25 to 34 years of age with a college degree relative to the average income of their counterparts with a high school diploma (OECD 2011b, table A8.1). Intergenerational economic mobility is measured as the elasticity between paternal earnings and a son's adult earnings, using data on a cohort of children born, roughly speaking, during the early to mid 1960s and measuring adult outcomes in the mid to late 1990s (see notes to Figure 1).

Labor Market Inequalities and the Returns to Human Capital

Labor market outcomes have become more unequal in the United States and many other high-income countries since the late 1970s and early 1980s. This pattern is now very well-documented, as have been many of the underlying causes associated with skill biases in technical change, its interaction with globalization, and the capacity of the supply of skilled workers to keep up with demand. But institutional differences have also implied that changes in inequality and the returns to skills have varied across countries.

Figure 4 is inspired by the main hypothesis put forward by Solon (2004), and it relates the intergenerational earnings elasticity to the earnings premium a college graduate has over a high school graduate. The earnings premium is measured as the average employment income in 2009 of men 25 to 34 years of age with a college degree relative to the average income of their counterparts with a high school diploma (OECD 2011b, table A8.1). As the figure illustrates, in countries where the

return to college education is higher, intergenerational mobility tends to be lower. Clearly, this is a tendency, and there are notable outliers—particularly Italy and to a degree also Finland—suggesting that other forces dominate in those countries. But the premium is higher in the United States than any other country included in the figure: a college graduate earns about 70 percent more than a high school graduate, compared to about 30 percent in Canada.

This correlation between a higher skill premium and lower intergenerational mobility of earnings also holds over time. Aaronson and Mazumder (2008) derive estimates of the intergenerational elasticity of earnings for the United States from about 1940 to about 2000 and compare them with the evolution of the return to education documented by Goldin and Katz (1999). Figure 5, adapted from Mazumder (2012), shows that the father-son earnings elasticity moves over the decades in tandem with the return to education—the changes after 1980 being the most notable. The increase in the returns to college from 9 percent in 1980 to about 13 percent two decades later was matched by a significant increase in the intergenerational earnings elasticity, from 0.38 to about 0.55.[2]

Of course, the average premium for higher education isn't informative about the distribution of that premium. As Lemieux (2006) points out, relative wages among the highly educated have become much more dispersed since the mid 1970s. He suggests this is the result of a strong heterogeneity in the returns to higher education. Much of the increase in labor market inequality is the result of this heterogeneity and reflects higher wages at the very top of the distribution. Wage growth has been higher for those with a college education in the United States but higher still for those among the college educated with graduate and professional credentials.

This pattern is consistent with the significant increase in the share of total earnings and incomes accruing to the very top of the distribution. Rising top shares are an important component of rising inequality, and while they have been particularly notable in English-speaking countries, they have increased more in the United States than most anywhere else, though the United Kingdom and Canada also experienced large increases (OECD 2011a, p. 39).

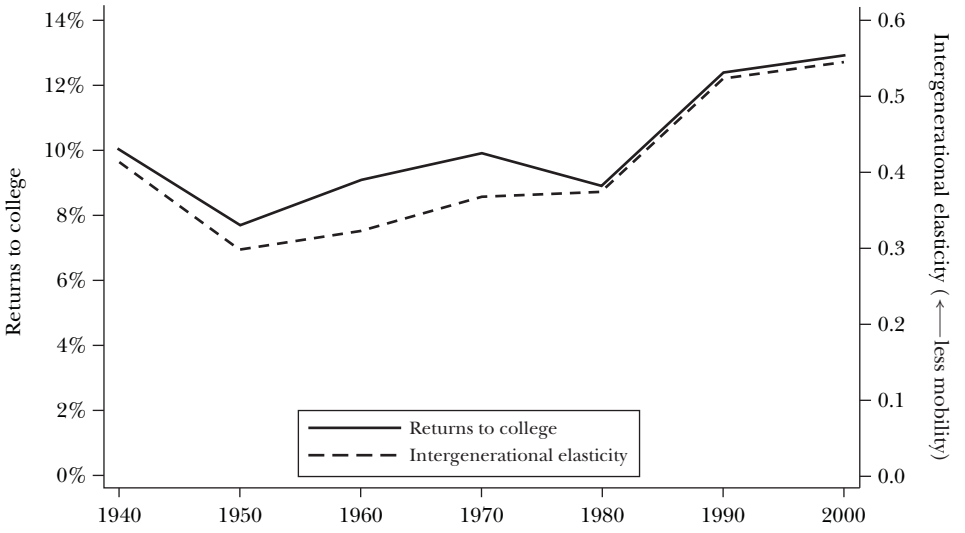
These patterns are likely to strengthen the tie between the economic outcomes of parents and children at the top. The intergenerational earnings elasticity is a measure of the average degree of relative mobility, but the underlying issue here is whether it is stronger (that is, nonlinear) for high levels of parental income. In some countries this was already evident for the generation that came of age just as top income shares started increasing.

Bratsberg et al. (2007) find that the intergenerational elasticity of income for Denmark, Finland, and Norway is flat across the lower parts of the parental distribution, and then rises at the higher end: that is, being raised by a low-income

[2] It should be noted that some other researchers have not found statistically significant changes in the intergenerational elasticity of earnings for the United States over the postwar period. For example, see Mayer and Lopoo (2004) and Lee and Solon (2009). However, the Aaronson and Mazumder paper is distinguished by its use of Census-based information that offers much larger samples sizes.

Figure 5

The Higher the Return to College, the Lower the Degree of Intergenerational Mobility: United States, 1940 to 2000



Source: Adapted by the author from Mazumder (2012, Figure 1).
Notes: Information on the returns to college and the intergenerational earnings elasticity were provided to the author by Bhashkar Mazumder. As reported in Mazumder (2012), these are respectively from Goldin and Katz (1999) and Aaronson and Mazumder (2008, table 1 column 2). The 1940 estimate of the elasticity is a projection using Aaronson and Mazumder (2008, table 2 column 2).

father confers no disadvantage, but being raised by a high-income father confers an advantage. Björklund, Roine, and Waldenström (2012) and my colleagues and I (Corak and Heisz 1999; Corak and Piraino 2010, 2011) document roughly similar patterns in Swedish and Canadian data with the intergenerational elasticity for top earners being two to three times greater than the overall average. However, Bratsberg et al. (2007) reject this convex pattern for the United Kingdom and United States, suggesting that a linear specification is a better fit. These differences may be substantive, or they may also reflect limitations in the size of the sample available from survey-based data used in the United Kingdom and United States. This is a major limitation in the American literature. In the other countries, the analyses are based upon administrative data with substantially larger sample sizes, and likely better representation at the extremes of the distribution.

Families and Investment in Human Capital

On the one hand, the impact of the returns to education on the degree of intergenerational mobility can be interpreted as reflecting an important role for the

transmission of innate ability between parents and children. If endowments of this sort are strongly inherited, then their impact on earnings is heightened if returns to schooling are higher: when returns are higher, mobility is as a result lower. But this interpretation also has to account for nonlinear patterns both in the returns to schooling, and the transmission of incomes across the generations. Nonlinearities would seem to imply that top earners are either particularly talented and have, in some sense, more of the characteristics valuable in the labor market to pass on to their children, or that these characteristics are more strongly transmitted between top-earning parents and their children than in middle- or lower-income families.

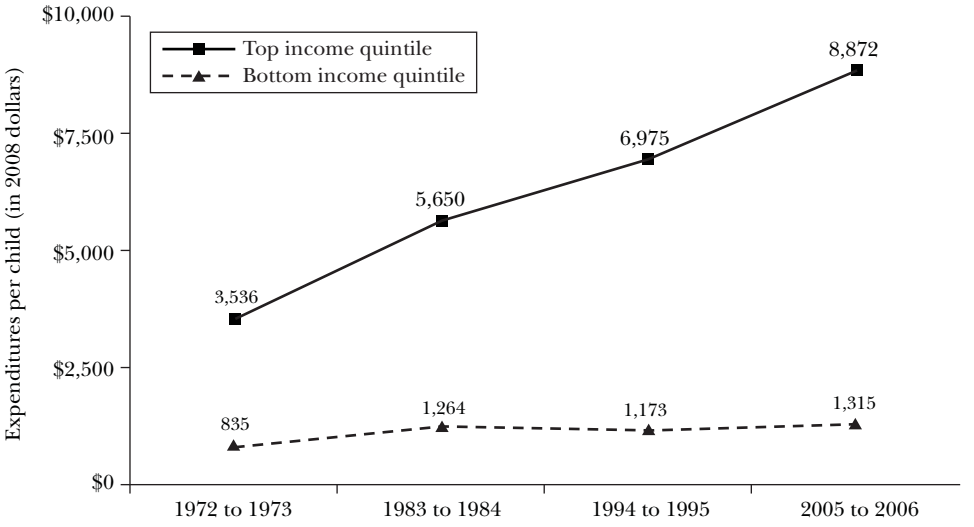
On the other hand, endowments should not be thought of as fixed traits transmitted mechanically across generations. Anything that boosts inequality reduces mobility because it differentially changes both opportunities and incentives for families to invest in their children. Families with more human capital invest more in their children. These investments are surely influenced by money: high-income parents having more scope to develop their children's skills and attitudes and to enrich their day-to-day experiences particularly during the early years. But the relevant investments are also nonmonetary, reflecting the development of behavior, motivation, and aspirations, as well as the possibility that high-income families can offer their children connections to selective schools and even to particular employers.

A college education is increasingly a gateway to higher incomes, but in the United States this effect is especially strong for a higher-level education from a selective college. The gap in college completion between children from low- and high-income families has increased significantly during the last two to three decades of increasing income inequality. Bailey and Dynarski (2011) show that the rate of college graduation increased by about 4 percentage points among a cohort of young people born in the early 1980s to low-income parents compared to their counterparts born in the early 1960s. However, among the cohorts born to relatively high-income parents, the rate of college graduation increased by almost 20 percentage points. Certainly the children of high-income families will find it easier to afford college. Belley and Lochner (2007) examine the relationship between family income and education outcomes in more detail and find that, even when controlling for cognitive skills, the strength of the relationship between family income and college attendance increased significantly over this period, about doubling in its impact. This pattern also holds when looking instead at the quality of the college attended. They suggest that the families of children coming of age during an era of increasing inequality, those born in the 1980s, are more likely to be borrowing-constrained than those raising children born during the 1960s and 1970s.

While family income matters, and while in the United States it increasingly matters, it is not everything. Belley, Frenette, and Lochner (2011) illustrate the importance of differences in financial aid in determining postsecondary attendance in the United States and Canada. While the strength of the tie between family income and postsecondary attendance is much weaker in Canada, even when controlling for cognitive skills, their analysis suggests that, at least in the case of public tuition

Figure 6

Money Matters: Higher-Income Families in the United States Have Higher Enrichment Expenditures on Their Children



Source: Duncan, Greg J. and Richard J. Murnane. Figure 1.6 "Enrichment Expenditures on Children, 1972–2006." In Whither Opportunity, edited by Greg J. Duncan and Richard J. Murnane, © 2011 Russell Sage Foundation, 112 East 64th Street, New York, NY 10065. Reprinted with permission.
Note: "Enrichment expenditures" refers to the amount of money families spend per child on books, computers, high-quality child care, summer camps, private schooling, and other things that promote the capabilities of their children.

fees and associated financial aid packages, the United States is more generous in its support to children from low-income families than Canada.

One way to explain all this is that the children of low-income families, especially in the United States, may not have the guidance and culture from their families that encourages college attendance, so that the offer of financial aid in and of itself is not enough. A field experiment conducted by Bettinger, Long, Oreopoulos, and Sanbonmatsu (2009) points out that a relatively small amount of help given to low-income families in completing a Free Application for Federal Student Aid, or FAFSA, form substantially raises the chances that high school seniors attend college. In other words, the patterns in the United States reflect—to a degree that they don't in Canada—more than the financial capacity of capable high school seniors.

The development of these capabilities during the years before high school graduation has also become more unequal in the way predicted by Solon (2004). Monetary investments outside of formal schooling help promote a child's human capital in the primary school years, and likely raise the odds of having both the skills and also the aptitudes, to successfully apply to a college when the time comes. These investments have been increasingly unequally distributed over time. Figure 6, adapted from Duncan and Murnane (2011), contrasts the evolution of "enrichment

expenditures" by families in the top 20 percent of the US income distribution with those in the bottom 20 percent. These expenditures refer to money spent on books, computers, high-quality child care, summer camps, and private schooling, among other things that promote the capabilities of children. Annual expenditures rose significantly for families in the bottom 20 percent, from about $835 dollars per child during the early 1970s to over $1,300 per child in the mid 2000s. But this pales in comparison to the increase among households in the top 20 percent: the significant gap between the two groups already present in the early 1970s ballooned over these decades as spending by those at the top went from $3,500 to almost $9,000 per child.

The nonmonetary capacities of families are also likely to be more limited for low-income families, and this may be especially apparent in America. In Corak, Curtis, and Phipps (2011), we find that along a host of dimensions—the age of the mother, the education of the mother, the incidence of living with both biological parents, as well as the incidence of living in a single parent household—Canadian children, particularly relatively disadvantaged children, live on average in a more enriching family environment than American children. McLanahan (2004) offers a clear presentation of the US trends in similarly defined family resources. She finds that there is a growing divergence in the parental resources available to children according to the education levels of their mothers. The evolution of mothers' age, mothers' employment, single motherhood, and fathers' involvement since the 1970s all indicate as much. Children born to the relatively more educated mothers are increasingly likely to be raised by an older, more mature mother who is working in a better paying job and more likely to be in a stable union. These children are also likely to be spending more time with their fathers. Children born to relatively less-educated mothers are increasingly likely over time to be making less-significant gains, indeed to be experiencing losses, in the parental resources available to them.

The upshot of all this is that increasing divergence in both monetary and nonmonetary investments in children during an era of increasing inequality may well lead to an increasing divergence in cognitive attainments and achievements that are the necessary prerequisites for college success. Reardon (2011) gathers information on math and reading test scores from a variety of sources for birth cohorts from about 1940 to the 2000s and charts the standardized gap between children raised in families with incomes at the 90th percentile and those raised in families at the 10th. Though some of the early trends are not conclusive, they seem to suggest that the 90/10 test score gap did not change that much from the 1950s to about the mid 1970s. For subsequent birth cohorts, however, there is a substantial increase amounting to about 30 to 40 percent between the mid to late 1970s and 2001. Even when the most able children of low-income families go to college, they tend not to apply to more selective colleges when compared to children of equal ability from high-income families. Hoxby and Avery (2012) document this tendency and suggest that it occurs in spite of the fact that when the more generous financial aid of selective colleges is considered, the eventual costs of these colleges would often be lower than for nonselective schools.

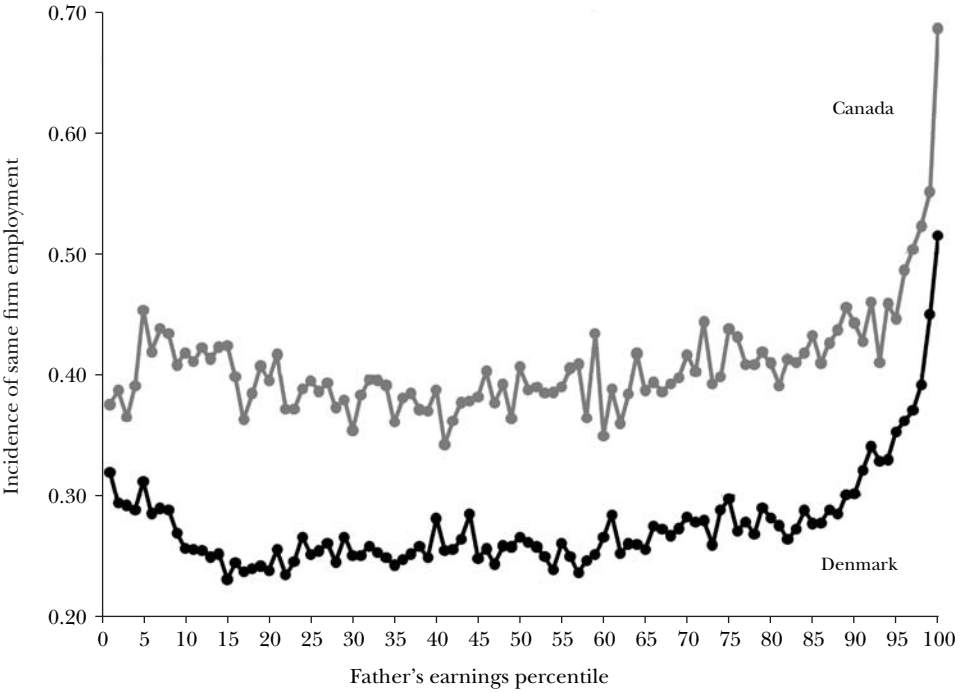
Finally, the traits relevant for success extend well beyond cognitive development before and during the school years. Families support their children through all the transitions they must make on the way to adulthood, including the transition to active and full-time engagement in the labor market. A more polarized and unequal labor market makes this more of a challenge for some than for others and also implies that family connections will matter all the more.

In Corak and Piraino (2010, 2011) and Bingley, Corak, and Westergård-Nielson (2012), the evidence suggests a strong tendency for labor market connections, in some sense, to matter for child outcomes. We document a very strong transmission of economic status at the top even in relatively mobile countries like Canada and Denmark. In particular, we show that the intergenerational transmission of earnings at the very top is associated with the intergenerational transmission of employers. Sons of top-earning fathers are more likely to fall from the top strata if they did not work for the very same employer for which their father had also worked. Figure 7 documents the intergenerational transmission of employers across the percentiles of the paternal earnings distribution for the two countries we analyze in Bingley, Corak, and Westergård-Nielson (2012). This is also based upon the broadest of the definitions we use: for sons in their early 30s, the incidence of ever having worked for an employer that had ever employed their fathers. The patterns in these two relatively mobile countries are remarkably alike: the overall levels differ, but there is a distinct tendency for the proportion to be much higher at the upper tail. Overall, about four out of every ten young Canadian men at some point held a job with an employer who in the past also employed their fathers. The intergenerational transmission of employers rises for those born to fathers in the top 10 percent of the income distribution, and sharply so for those born to fathers at the very top. Almost seven out of ten Canadian sons born to top 1 percent fathers had a job with an employer for which they had also worked, and in Denmark a little over half of sons of fathers at this level did so.

There is no direct evidence that these patterns also characterize the American labor market, though Datcher Loury (2006) suggests that in the United States up to half of jobs are found through family, friends, or acquaintances. She also shows that the highest wages are paid to those who find jobs through "prior generation male" relatives who actually knew the potential employer or served as a reference. While this information does not appear to be available across the US earnings distribution, the literature on the succession of chief executive officers in family firms hints at the possibility that the incidence could be higher at the very top. Pérez-González (2006) examines just over 300 CEO transitions and finds that in more than one-third, the new CEO had a family connection. In addition, these transitions were associated with a decline in firm performance, particularly so when the newly appointed family member did not attend a select college. Bennedsen, Nielson, Pérez-González, and Wolfenzon (2007) offer a similar, but more detailed analysis with Danish data, and using instrumental variables, more firmly document a causal impact of family succession on declining performance.

While these patterns may reflect simple nepotism, and the historical review by Bellow (2003) suggests that possibility, other interpretations are also possible. If

Figure 7
Proportion of Sons Currently Employed or Employed at Some Point with an Employer their Father had Worked for in the Past: Canada and Denmark
(by father's earnings percentile)



Source: Bingley, Paul, Miles Corak, and Niels Westergard-Nielson. Figure 18.2 "Sons Employed at Some Point with Employer Fathers Worked for, by Fathers' Earnings." In From Parents to Children: The Intergenerational Transmission of Advantage, edited by John Ermisch, Markus Jantti, and Timothy Smeeding, © 2012 Russell Sage Foundation, 112 East 64th Street, New York, NY 10065. Reprinted with permission.

there is intergenerational transmission of firm-specific skills, then children inherit human capital that has a higher return when they are employed by the family firm. In this sense, the intergenerational transmission of employers might be interpreted as another reflection of the transmission of skills and traits valuable for labor market outcomes. But the decline of firm performance upon the succession of a family member would seem to suggest that family members do not on average have a distinctly more valuable set of skills or managerial talent.

In Corak and Piraino (2010, 2011) and Bingley, Corak, and Westergård-Nielson (2012), my coauthors and I show that the intergenerational transmission of employers is higher when fathers report self-employment income, and presumably have control over a firm and its hiring decisions. But we also show that the patterns are much broader and not due simply to firm ownership. Other factors, like information about the labor market or "connections" (in the sense used by Becker and

Tomes) help to structure a child's job search and play a role in generating the intergenerational transmission of employers across the entire parental income distribution, but particularly at the top.

My own sense is that in the United States, and also the United Kingdom, this channel between parent and child economic status due to connections probably works more strongly for top earners through college choice, and particularly through the select colleges. Anecdotal evidence is often used to suggest that access to unpaid internships, which permit the development of on-the-job training and firm-specific human capital, is also tilted toward children of the relatively well-to-do, whose families have the resources to finance them.

Public Policy as Leveling or Tilting the Playing Field

Public policy can affect the investments made in children across the entire income distribution. It can also affect how families interact with labor markets. The United States stands out in the degree to which government programs are of relatively more benefit to the advantaged. As such, they are more likely to exacerbate rather than blunt the degree to which labor market inequalities are passed on across generations.

When the Pew Charitable Trusts asked Canadians what they understood the good and successful life to be—the dimensions of what might thought of as the "Canadian Dream"—the responses were uncannily similar to how Americans defined the "American Dream." In Corak (2010), I report that the citizens of both countries value the ideal of equality of opportunity and define it—almost exactly to the same degree—in terms of individual freedoms. They also recognize the importance of individual responsibilities and have an equal aversion to "equality of outcomes" as a desirable end. The biggest difference in this comparative analysis of similarly worded public opinion polls concerned the view of government and public policy. Americans were more inclined to view government as doing more harm than good in their pursuit of the American Dream; at the same time, they viewed a whole host of possible public policy interventions as effective in promoting economic mobility. From this, I surmise that they had less confidence that their federal, state, and local government could implement and manage effective policy changes.

As a result, there are significant differences in the broader social circumstances under which children in the United States and Canada are being raised. Carasso, Reynolds, and Steuerle (2008) attempt to estimate the global incidence of US federal government spending on programs, like education, that promote mobility, placing them into a broader context of total government spending. They find that the US government spends considerable amounts in this way, up to 1.6 percent of GDP in 2006, but that only about one-quarter of these expenditures are to the benefit of lower- to moderate-income individuals.

A notable example is the education system. At almost $15,000 per student, America spends more on the schooling of its children than almost any other

high-income country (OECD 2011b). But the American education system does not promote mobility to the extent that it could because its educational spending is more likely to benefit the relatively well-to-do. The OECD suggests that the higher levels of spending in the United States—both private and public—are driven by much higher spending on tertiary education. For every $1 spent on primary education, $3 are spent on tertiary education, the highest ratio of all high-income countries. Further, tertiary spending is dominated by private sources of financing, which makes up over 60 percent of all spending on this level of education. Education spending, in other words, is allocated to make higher education relatively more of a priority, and in a way that is of relatively more benefit to the relatively advantaged.

The demand for high-quality college education among the relatively well off expresses itself in a demand for high-quality primary and secondary schooling that offers a gateway to a good college education. While America also spends more on primary education per pupil than many other countries, significant inequalities in parental resources express themselves in the structure of the system, leading to variations in financing, quality, and access in a way that does little to level the playing field. The OECD (2012, p. 30) summarizes its research on this issue in this way: "Currently the United States is one of only three OECD countries that on average spend less on students from disadvantaged backgrounds than on other students. . . . Moreover, the most able teachers rarely work in disadvantaged schools in the United States, the opposite of what occurs in countries with high-performing education systems."

At the same time, socioeconomic differences in readiness to learn among children just starting school are larger in the United States than in other countries, making the challenge faced by the schooling system all the greater. In Bradbury, Corak, Waldfogel, and Washbrook (2012), we study vocabulary development and behavioral problems among children who were 4–5 year-olds in 2000 in Australia, Canada, the United Kingdom, and the United States. We find inequalities according to family income and mother's education in all four countries. But in general, these inequalities are notably greater in the United States and most muted in Canada.

In Corak, Curtis, and Phipps (2011), my coauthors and I look at a wider cohort of children from newborns to about 13 years of age during the late 1990s—that is, those who were among the first to be raised in an era of rising inequality and who will in the coming years be the subject of the next generation of intergenerational mobility studies—and find that they are much more affluent in the United States than in Canada, having on average almost one-third more income. Though children in both countries are distributed across their countrywide income distributions in the same way, the gap between bottom and top children differs in the two countries. In the United States it is much greater: a child in the top decile of the income distribution has 14 times as much as a bottom decile child. In Canada a top decile child has only 7.5 times as much economic resources as a bottom decile child. When we placed Canadian children in the American income distribution, adjusting their incomes using an index of Purchasing Power Parity, they tended to be lower-middle income in status. However, while Canadian children are much less likely to be in

the top half of the American income distribution, they are also less likely to be in the bottom 10 percent, so their low-income, in this absolute sense, is not as great.

Public regulations and provision of goods associated with human capital likely do more to level this playing field in Canada than in the United States. In Corak, Curtis, and Phipps (2011), we show that mental and physical health, school readiness, and some education outcomes are on average higher in Canada and less tied to family circumstance. It is not a simple task to attribute these outcomes to the public provision of goods in a causal sense. However, we suggest that universal provision of health care is associated with more preventative care for children that reduces the number and severity of health shocks that could have longer-term consequences. In addition, parents have more flexibility in making childcare and work arrangements in Canada. For the study period we consider, the late 1990s, there seemed to be more part-time employment in Canada, and a significant policy change in the mid 1990s extended paid parental leave for up to almost one year after a child's birth and gave parents the right to return to their job. Income support to families was also reformed at around the same time, delivered through the income tax system, and was more targeted and generous for lower-income families. The program is substantially more generous than its American counterpart and is more likely to reach all families with children because tax-filing rates are nearly universal. In fact, more recently some provincial governments have introduced full-time kindergarten for four year-olds.

In contrast, total hours of household labor supplied by household members were higher in the United States during this period, but also polarized across families. This pattern is associated with a more limited system of parental leave. While relatively well-off households are able to afford high-quality child-care or have one partner, usually the mother, withdraw from the labor market, lone parents have fewer child-care options and are likely to continue working. The methods in Corak, Curtis, and Phipps (2011) certainly fall short of establishing a causal impact on child attainments, whether in the long-run nor in the short-run, but our study does demonstrate that public policy is contributing to parents balancing the demands of work and family in different ways between the United States and Canada.

Conclusion

Relatively less upward mobility of the least advantaged is one reason why intergenerational mobility is lower in the United States than in other countries to which Americans are often compared. But it is not the only reason. Intergenerational mobility is also lower because children of top-earning parents are more likely to become top earners in their turn. An era of rising inequality will be more likely to heighten these differences than diminish them. The cohort of American children raised since the 1980s, who will reach their prime working years in the coming decade, is likely to experience an average degree of intergenerational income mobility as low—if not lower—than previous cohorts who were raised in an era of less inequality.

Inequality lowers mobility because it shapes opportunity. It heightens the income consequences of innate differences between individuals; it also changes opportunities, incentives, and institutions that form, develop, and transmit characteristics and skills valued in the labor market; and it shifts the balance of power so that some groups are in a position to structure policies or otherwise support their children's achievement independent of talent.

Thus, those who are concerned about equality of opportunity should also care about inequality of outcomes, but only to the extent that these differences in outcomes are due, in the words of John Roemer (2004), to "differential circumstances." Roemer considers three categories of circumstances through which parents may give their children an advantage. First, parents may transmit economic advantages through social connections facilitating access to jobs, admission to particular schools or colleges, or access to other sources of human capital. Second, parents may influence life chances through the genetic transmission of characteristics like innate ability, personality, and some aspects of health that are valued in the labor market. Third, parents may influence the lifetime earnings prospects of their children in subtle ways, like through a family culture and other monetary and nonmonetary investments that shape skills, aptitudes, beliefs, and behavior. When it comes to "equal opportunity," a common pattern is that people tend to support policies that would assure a level playing field in access to jobs and education, less willing to take steps to offset genetic advantages, and conflicted about what steps might be appropriate in counterbalancing within-family investments. But my main point here is that deciding which circumstances should be offset by policy steps of some kind, and as a result the fraction of parental income advantage passed on to children that is consistent with "equality of opportunity," is a value judgment that different societies may well make differently.

The demographic diversity between the high-income countries, and their underlying values, imply that it may be impossible, and indeed not even desirable, to change the degree of mobility in countries like the United Kingdom or the United States into the rates observed in Denmark. Rather, the cross-country comparison of intergenerational mobility of the sort offered by the Great Gatsby Curve invites us to reflect on what makes one country different than another so we may clarify the underlying drivers and determine whether these are forces that can change and whether we want them to change. This is one reason why parts of this overview have focused on the differences between the United States and Canada, and more importantly on changes within the United States over time.

The inequality literature has paid little attention to the intergenerational consequences of the increasing top income shares that it has so carefully documented. Freeland (2012) graphically documents the degree to which the top 1 percent, by virtue of the magnitude of their income, are divorced from the rest of the population in their work arrangements, consumption behavior, and beliefs. I have argued here that the top 1 percent are also different in the way advantages are passed on to the next generation, which certainly involves much higher-quality schooling and other investments of human capital from the early years onward, but may well also

involve nepotism in the allocation of jobs. Children of top earners are more likely to grow up to be top earners. Indeed, at some point the high levels of earnings accrued by the top 1 percent will be reflected in capital accumulation, and eventually lead to stronger intergenerational transmission of wealth, a topic not addressed at all in this paper. This dynamic at the top, and its underlying drivers, are likely very different from the configuration of forces determining intergenerational mobility for those in the lower half of the income distribution. Even so, some countries are likely to combine a good deal of intergenerational mobility with higher top shares because the balance in the lower parts of the income distribution between labor market inequalities, the health and vitality of the family as an institution, and broad, high-quality, and accessible public investments in human capital will not be (much) skewed by top earners. This pattern may well be the case in Sweden and Canada: Björklund, Roine, and Waldenström (2012) and my coauthor and I in Corak and Piraino (2010) suggest that, for these two countries, high mobility for most coexists with a "dynasty" for the top 1 percent.

A similar dynamic seems unlikely to unfold in the United States. While the imagined prospect of upward mobility for those in the lower part of the income distribution shares little in common with the generational dynamics of the top 1 percent, the latter may continue to be an important touchstone for those in, say, the top fifth of the US income distribution. After all, this group too has experienced significant growth in its relative standing, which partly reflects an increasing return to the graduate and other higher degrees for which they exerted considerable effort but is also linked to a background of nurturing families and select colleges. This group has both the resources and incentives to turn more intensely to promoting the capacities of their children. With effort and a bit of luck, it is not unreasonable for them to believe they may yet cross the threshold into the top 1 percent, and they can certainly imagine that their children stand just as good a chance, if not better. For them, the "American Dream" lives on, and as a result they are likely not predisposed, with their considerable political and cultural influence, to support the recasting of American public policy to meet its most pressing need, the upward mobility of those at the bottom.

■ *I gratefully acknowledge very helpful comments from Chang-Tai Hseih, Ulrike Malmendier, and especially David Autor and Timothy Taylor. I would also note that this paper is a revised version of a discussion paper originally circulated under the title "How to Slide Down the Great Gatsby Curve" by the Center for American Progress (available at http://www.americanprogress.org), which itself grew out of a presentation I made to the "Social Mobility Summit" in London, United Kingdom, in May 2012, which was organized by the Sutton Trust and the Carnegie Foundation. The support and comments of Heather Boushey, Lee Elliot Major, and Eric Wanner are also appreciated as is the opportunity to make presentations at a meeting of the Canadian Institute for Advanced Research, the Prentice Institute of the University of Lethbridge, and the conference in honor of Thomas J. Courchene held at Queen's University in October 2012.*

References

Aaronson, Daniel, and Bhashkar Mazumder. 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources* 43(1): 139–72.

Andrews, Dan, and Andrew Keith Leigh. 2009. "More Inequality, Less Social Mobility." *Applied Economics Letters* 16(15): 1489–92.

Atkinson, Anthony Barnes, Alan K. Maynard, and Chris G. Trinder. 1983. *Parents and Children: Incomes in Two Generations.* London: Heinemann Educational Books.

Bailey, Martha J., and Susan M. Dynarski. 2011. "Inequality in Postsecondary Education." In Chap. 6 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

Becker, Gary. 2013. "Meritocracies and Intergeneration Mobility." The Becker–Posner Blog, January 1. http://www.becker-posner-blog.com/2013/01/meritocracies-and-intergeneration-mobility-becker.html.

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy* 87(6): 1153–89.

Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4(3, Part 2): S1–S39.

Belley, Philippe, Marc Frenette, and Lance Lochner. 2011. "Post-Secondary Attendance by Parental Income in the U.S. and Canada: What Role for Financial Aid Policy?" NBER Working Paper 17218.

Belley, Philippe, and Lance Lochner. 2007. "The Changing Role of Family Income and Ability in Determining Educational Attainment." NBER Working Paper 13527.

Bellow, Adam. 2003. *In Praise of Nepotism: A Natural History.* New York: Random House.

Bénabou, Roland, and Efe A. Ok. 2001. "Social Mobility and the Demand for Redistribution: The Poum Hypothesis." *Quarterly Journal of Economics* 116(2): 447–87.

Bennedsen, Morten, Kasper Meisner Nielson, Francisco Pérez-González, and Daniel Wolfenzon. 2007. "Inside the Family Firm: The Role of Families in Succession Decisions and Performance." *Quarterly Journal of Economics* 122(2): 647–49.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2009. "The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment." NBER Working Paper 15361.

Bingley, Paul, Miles Corak, and Niels C. Westergård-Nielson. 2012. "Equality of Opportunity and Intergenerational Transmission of Employers." Chap. 18 in *From Parents to Children: The Intergenerational Transmission of Advantage,* edited by John Ermisch, Markus Jäntti, and Timothy Smeeding. New York: Russell Sage Foundation.

Björklund, Anders, and Markus Jäntti. 2009. "Intergenerational Income Mobility and the Role of Family Background." Chap. 20 in *The Oxford Handbook of Economic Inequality,* edited by Wiemer Salverda, Brian Nolan, and Timothy M. Smeeding. Oxford University Press.

Björklund, Anders, Jesper Roine, and Daniel Waldenström. 2012. "Intergenerational Top Income Mobility in Sweden: Capitalist Dynasties in the Land of Equal Opportunity?" *Journal of Public Economics* 96(5): 474–84.

Blanden, Jo. 2013. "Cross-Country Rankings in Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology." *Journal of Economic Surveys* 27(1): 38–73.

Bradbury, Bruce, Miles Corak, Jane Waldfogel, and Elizabeth Washbrook. 2012. "Inequality in Early Child Outcomes." Chap. 4 in *From Parents to Children: The Intergenerational Transmission of Advantage*, edited by John Ermisch, Markus Jäntti, and Timothy M. Smeeding. New York: Russell Sage Foundation.

Bratsberg, Bernt, Knut Røed, Oddbjørn Raaum, Robin Naylor, Markus Jäntti, Tor Eriksson, and Eva Österbacka. 2007. "Nonlinearities in Intergenerational Earnings Mobility: Consequences for Cross-Country Comparisons." *Economic Journal* 117(519): C72–C92.

Brunori, Paolo, Franciso H. G. Ferreira, and Vito Peragine. 2013. "Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons." Policy Research Working Paper 6304, World Bank, Development Research Group.

Carasso, Adam, Gillian Reynolds, and C. Eugene Steuerle. 2008. "How Much Does the Federal Government Spend to Promote Economic Mobility and for Whom?" Washington DC: Economic Mobility Project, Pew Charitable Trusts.

Corak, Miles. 2006. "Do Poor Children Become Poor Adults? Lessons for Public Policy from a Cross-Country Comparison of Generational Earnings Mobility." *Research on Economic Inequality,* Vol. 13: *Dynamics of Inequality and Poverty*, edited by John Creedy and Guyonne Kalb, 143–88. The Netherlands: Elsevier Press. (Available in an unabridged

form as IZA Discussion Paper 1993, http://ftp.iza.org/dp1993.pdf.)

Corak, Miles. 2010. "Chasing the Same Dream, Climbing Different Ladders: Economic Mobility in the United States and Canada." Washington DC: Economic Mobility Project, Pew Charitable Trusts.

Corak, Miles. 2013. "Inequality from Generation to Generation: The United States in Comparison." Chap. 6 in *The Economics of Inequality, Poverty, and Discrimination in the 21st Century,* edited by Robert S. Rycroft. Santa Barbara, CA: ABC-CLIO.

Corak, Miles, Lori Curtis, and Shelley Phipps. 2011. "Economic Mobility, Family Background, and the Well-being of Children in the United States and Canada." In *Persistence, Privilege and Parenting: The Comparative Study of Intergenerational Mobility,* edited by Timothy Smeeding, Markus Jäntti, and Robert Erickson. New York: Russell Sage Foundation.

Corak, Miles, and Andrew Heisz. 1999. "The Intergenerational Earnings and Income Mobility of Canadian Men: Evidence from Longitudinal Income Tax Data." *Journal of Human Resources* 34(3): 504–33.

Corak, Miles, and Patrizio Piraino. 2010. "Intergenerational Earnings Mobility and the Inheritance of Employers." IZA Discussion Paper 4876.

Corak, Miles, and Patrizio Piraino. 2011. "The Intergenerational Transmission of Employers." *Journal of Labor Economics* 29(1): 37–68.

Council of Economic Advisors. 2012. *Economic Report of the President.* Washington DC: United States Government Printing Office.

Datcher Loury, Linda. 2006. "Some Contacts are More Equal than Others: Informal Networks, Job Tenure, and Wages." *Journal of Labor Economics* 24(2): 299–318.

Duncan, Greg J., and Richard J. Murnane. 2011. "Introduction: The American Dream, Then and Now." Chap. 1 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances,* edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

Economic Mobility Project. 2009. "Opinion Poll on Economic Mobility and the American Dream." Washington DC: Pew Charitable Trusts. http://www.pewstates.org/research/analysis/opinion-poll-on-economic-mobility-and-the-american-dream-85899378651.

Ermisch, John, Markus Jäntti, Timothy M. Smeeding, and James A. Wilson. 2012. "Advantage in Comparative Perspective." Chap. 1 in *From Parents to Children: The Intergenerational Transmission of Advantage,* edited by John Ermisch, Markus Jäntti, and Timothy M. Smeeding. New York: Russell Sage Foundation.

Ferreira, Francisco H. G., and Jérémie Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *Review of Income and Wealth* 57(4): 622–57.

Freeland, Chrystia. 2012. *Plutocrats: The Rise of the New Global Super-Rich and the Fall of Everyone Else.* Toronto: Doubleday.

Goldin, Claudia, and Lawrence F. Katz. 1999. "The Returns to Skill in the United States across the Twentieth Century." NBER Working Paper 7126.

Hoxby, Caroline M., and Christopher Avery. 2012. "The Missing 'One-Offs': The Hidden Supply of High-Achieving, Low Income Students." NBER Working Paper 18586.

Jencks, Christopher, and Laura Tach. 2006. "Would Equal Opportunity Mean More Mobility?" In *Mobility and Inequality: Frontiers of Research from Sociology and Economics,* edited by Stephen B. Morgan, David L. Grusky, and Gary S. Fields. Stanford University Press.

Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff. 2006. "Economic, Neurobiological, and Behavioral Perspectives on Building America's Future Workforce." *PNAS,* July, 103(27): 10155–62.

Krueger, Alan. 2012. "The Rise and Consequences of Inequality." Presentation made to the Center for American Progress, January 12th. http://www.americanprogress.org/events/2012/01/12/17181/the-rise-and-consequences-of-inequality/.

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91(4): 766–772.

Lefranc, Arnaud, Nicolas Pistolesi, and Alain Trannoy. 2008. "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike." *Review of Income and Wealth* 54(4): 513–46.

Lemieux, Thomas. 2006. "Post-secondary Education and Increasing Wage Inequality." NBER Working Paper 12077.

Mayer, Susan E., and Leonard M. Lopoo. 2004. "What Do Trends in the Intergenerational Economic Mobility of Sons and Daughters in the United States Mean?" Chap. 5 in *Generational Income Mobility in North America and Europe,* edited by Miles Corak. Cambridge University Press.

Mazumder, Bhashkar. 2005. "The Apple Falls Even Closer to the Tree than We Thought: New and Revised Estimates of the Intergenerational Inheritance of Earnings." In *Unequal Chances: Family Background and Economic Success,* edited by Samuel Bowles, Herbert Gintis, and Melissa Osborne Groves. New York: Russell Sage Foundation.

Mazumder, Bhashkar. 2012. "Is Intergenerational Economic Mobility Lower Now than in the Past?" Chicago Fed Letter 297, April.

McLanahan, Sara. 2004. "Diverging Destinies: How Children Are Faring under the Second Demographic Transition." *Demography* 41(4): 607–627.

Mulligan, Casey B. 1997. *Parental Priorities and Economic Inequality.* University of Chicago Press.

OECD. 2011a. *Divided We Stand: Why Inequality Keeps Rising.* Organization for Economic Cooperation and Development.

OECD. 2011b. *Education at a Glance 2011: OECD Indicators.* Organization for Economic Cooperation and Development.

OECD. 2012. *OECD Economic Surveys: United States.* June 2012. Organization for Economic Cooperation and Development.

Paes de Barros, Ricardo, Francisco H. G. Ferreira, José R. Molinas Vega, and Jaime Saavedra Chanduvi. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean.* World Bank.

Pérez-González, Francisco. 2006. "Inherited Control and Firm Performance." *American Economic Review* 96(5): 1559–88.

Reardon, Sean F. 2011. "The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations." Chap. 5 in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage Foundation.

Roemer, John E. 2004. "Equal Opportunity and Intergenerational Mobility: Going Beyond Intergenerational Income Transition Matrices." Chap. 3 in *Generational Income Mobility in North America and Europe*, edited by Miles Corak. Cambridge University Press.

Roemer, John E. 2012. "What is the Justification of Studying Intergenerational Mobility of Socioeconomic Status?" Chap. 20 in *From Parents to Children: The Intergenerational Transmission of Advantage*, edited by John Ermisch, Markus Jäntti, and Timothy M. Smeeding. New York: Russell Sage Foundation.

Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review* 82(3): 393–408.

Solon, Gary. 2004. "A Model of Intergenerational Mobility Variation over Time and Place." Chap. 2 in *Generational Income Mobility in North America and Europe*, edited by Miles Corak. Cambridge University Press.

Zimmerman, David J. 1992. "Regression Toward Mediocrity in Economic Stature." *American Economic Review* 82(3): 409–29.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 103–124

Why Hasn't Democracy Slowed Rising Inequality?

Adam Bonica, Nolan McCarty, Keith T. Poole, and Howard Rosenthal

During the past two generations, democratic forms have coexisted with massive increases in economic inequality in the United States and many other advanced democracies. Moreover, these new inequalities have primarily benefited the top 1 percent and even the top .01 percent. These groups seem sufficiently small that economic inequality could be held in check by political equality in the form of "one person, one vote."

Indeed, the notion that inequality should be at least partially self-correcting in a democracy has a long pedigree in economic theory. In the canonical model of Meltzer and Richard (1981), increased inequality (in the form of median incomes falling relative to average incomes) leads the median voter to demand more redistribution, so that politics should limit after-tax and -transfer inequality. Redistribution is limited, however, by the consequences of how the higher rates of taxation reduce labor supply. A stripped-down version of this model, with similar implications, is the model developed by Bolton and Roland (1999), where redistribution is limited through deadweight loss in taxation. These early approaches (see also Romer 1975) assume that politics is majoritarian, equal (one person, one vote) and with full participation (all economic agents vote).

■ *Adam Bonica is Assistant Professor of Political Science, Stanford University, Stanford, California. Nolan McCarty is Susan Dod Brown Professor of Politics and Public Affairs, Woodrow Wilson School, Princeton University, Princeton, New Jersey. Keith T. Poole is Philip H. Alston Jr. Distinguished Professor of Political Science, University of Georgia, Athens, Georgia. Howard Rosenthal is Professor of Politics, New York University, New York City, New York. Their email addresses are bonica@stanford.edu, nmccarty@princeton.edu, ktpoole@uga.edu, and howardrosenthal@nyu.edu.*

http://dx.doi.org/10.1257/jep.27.3.103 doi=10.1257/jep.27.3.103

Other authors, notably Benabou (2000), look at more complex and realistic environments where theory does not provide clear predictions. In economies where there are efficiency gains to redistribution, there may be more support for redistribution at low levels of inequality than at high levels. If the 1 percent is not exceptionally rich relative to the mean, the efficiency gains may offset additional taxes. But as the 1 percent get relatively richer, they turn against redistribution. In any economy, unequal turnout or political power can create a situation where the pivotal voter under majority rule may have income greater than the median. When a society has large numbers of relatively poor residents without voting rights, as arises with substantial immigration, redistribution will be limited even when all of the enfranchised participate fully (McCarty, Poole, and Rosenthal 2006). When these and other concerns are brought into play, many relationships between democracy and inequality are theoretically possible.

In this paper, we explore five possible reasons why the US political system has during the last few decades failed to counterbalance rising inequality.

First, both Republicans and many Democrats have experienced an ideological shift toward acceptance of a form of free market capitalism which, among other characteristics, offers less support for government provision of transfers, lower marginal tax rates for those with high incomes, and deregulation of a number of industries. Financial deregulation, in particular, has been a source of income inequality (Philippon and Reshef 2008). The mass public may well embrace such an ideological shift if rising inequality nonetheless "trickles down" to rising incomes and home ownership for all. In recent years, there has been a serious financial crisis, declining median incomes, and declining home ownership rates. This raises important questions as to why these ideological trends persist and remain politically powerful.

Second, immigration and low turnout of the poor have combined to make the distribution of voters more weighted to high incomes than is the distribution of households (McCarty, Poole, and Rosenthal 2006). Turnout, of course, can also be influenced by legal and administrative measures that make it relatively costly for the poor to vote.[1]

Third, rising real income and wealth has made a larger fraction of the population less attracted to turning to government for social insurance. When the insured consumption can be self-provided at relatively low risk, there will be less support for government insurance than when risk is high and private insurance markets are of dubious value. There is a telling contrast between Social Security, which provides old age payments for general consumption, and Medicare and the Patient Protection and Affordable Care Act of 2010, which cover unforeseen, potentially catastrophic, expenses for health care. Social Security does not have a strongly progressive tax and benefit structure. Attempts to increase progressivity by removing the cap on the payroll tax or taxing unearned income have little political traction. By contrast, the

[1] See Alvarez, Bailey, and Katz (2008) for evidence on the disparate impact of recent tougher voter identification laws on the turnout of the poor.

Medicare tax on earned income was uncapped during the administration of George H. W. Bush and raised in the 2010 health care legislation. In addition, the Patient Protection and Affordable Care Act imposed a 3.8 percent tax on unearned income of high earners. Thus, public expenditure for high-risk health care consumption has expanded more than expenditure for other, less-risky, forms of old age consumption. Although the 1 percent might self-insure for health risks, the broader population has supported social insurance in this area.

Fourth, the rich have been able to use their resources to influence electoral, legislative, and regulatory processes through campaign contributions, lobbying, and revolving door employment of politicians and bureaucrats.

Fifth, the political process is distorted by institutions like gerrymandering that reduce the accountability of elected officials to the majority. Other political institutions, including a bicameral legislature with a filibuster, combine with political polarization to create policy gridlock, which in turn inhibits efforts to update social safety nets and regulatory frameworks in response to changing conditions.

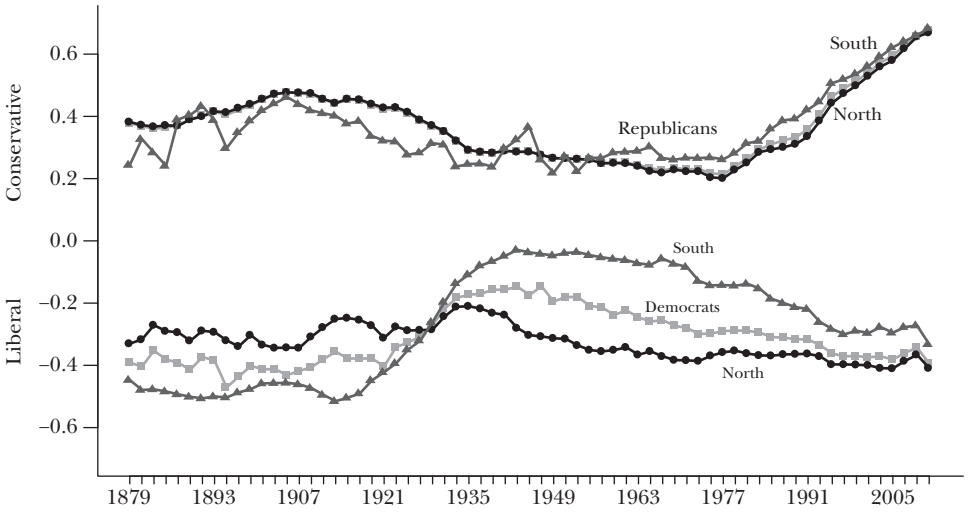
An Ideological Shift and Greater Polarization

Since roughly 1980, the political system has polarized (McCarty, Poole, and Rosenthal 2006) with the ideological gap between Democrats and Republicans widening while moderates vanish from Congress. At the same time, the Republicans enjoyed increased electoral success. The Democratic control of the federal government initiated with Franklin Delano Roosevelt was broken. During the 34 years from 1981–2014, the Republicans will have held the presidency for 20 years and at least one house of Congress for 22 years. Except for a brief period in 2009, the Republicans have had the 41 votes needed to sustain a Senate filibuster.[2]

Our data source for liberal–conservative positions of members of Congress are the widely used DW-NOMINATE scores, based on roll call votes. The one-dimensional implementation of DW-NOMINATE (Poole and Rosenthal 1997; McCarty, Poole, and Rosenthal 1997; Poole 2005) assumes that legislators make choices between yea and nay outcomes on a roll call vote as a function of the distance between the legislator's "ideal point" on the liberal–conservative dimension and the locations of the outcomes. The "spatial utility" a legislator assigns to an outcome is a Gaussian function of the distance between the outcome and that legislator's ideal point on the liberal–conservative dimension. The total utility of the outcome is the spatial utility plus a random, normally distributed component. The legislators' "ideal points" and the outcome components are estimated simultaneously from roll call votes. The

[2] In addition, a majority of the Supreme Court has been appointed by Republican presidents since 1970. Over time, some Republican Supreme Court appointees such as Earl Warren, John Paul Stephens, and David Souter often voted with Democratic appointees. However, there has been a conservative majority on the court since 1982 according to a scoring system developed by political scientists Andrew Martin and Kevin Quinn and available at http://www.targetpointconsulting.com/scotusscores-labels.html.

Figure 1
Party Means on Liberal–Conservative Dimension for the US House of Representatives, 1879–2012



Source: Author's calculations using DW-NOMINATE scores of the liberal–conservative positions of members of Congress, which are based on roll call votes. For methodological details, see Poole and Rosenthal 1997; McCarty, Poole, and Rosenthal 1997; and Poole 2005.

average ideal point of a legislator is constrained to lie in a continuum from −1 (far left) to +1 (far right).

Polarization is measured as the distance between the average ideal point of Republican legislators and that of Democrats. (Results are similar for other measures of dispersion, such as the variance of ideal points.) Overlapping cohorts of legislators permit measurements that are comparable across time.[3]

As Figure 1 shows, most of the polarization for the House of Representatives has been produced by a rightward movement of Republicans. The pattern for the Senate is similar. More precisely, entering Republicans like Ted Cruz have tended to vote to the right of continuing members, and exiting members like Olympia Snowe have been relatively moderate. Republicans have become sharply more conservative. The slight liberal drift of the Democrats is compositional in nature. Moderate Democrats from the South have been replaced by conservative Republicans. The few remaining southern Democrats are now heavily African-American, representing minority-majority districts. The change has made for a smaller, more liberal group of southern Democrats. However, Democrats as a whole have not moved much to the left. Overall, entering and exiting Democrats have looked much like those that have continued to serve.

[3] Legislator ideal points are restricted to linear trends throughout the legislator's career. For example, a legislator can trend from liberal to conservative but cannot go back and forth.

Our findings do not, however, suggest that the Democrats continue to support policies that would reduce inequality as much as they did in the New Deal. After all, nineteenth-century Democrats, centered on populist southern whites, supported railroad regulation and antitrust legislation, and in this sense were to the economic left of current Democrats. The Democratic party pushed through the financial regulation of the 1930s, while the Democratic party of the 1990s undid much of this legislation in its embrace of unregulated financial capitalism, including the Interstate Banking and Branching Efficiency Act of 1993, which eliminated previous restrictions on interstate banking and branching; the Gramm–Leach–Bliley Act of 1999, which repealed the 1933 Glass–Steagall Act that had separated commercial banking from other financial services; and the Commodity Futures Modernization Act of 2000, which prevented the Commodity Futures Trading Commission from regulating most over-the-counter derivative contracts, including credit default swaps.

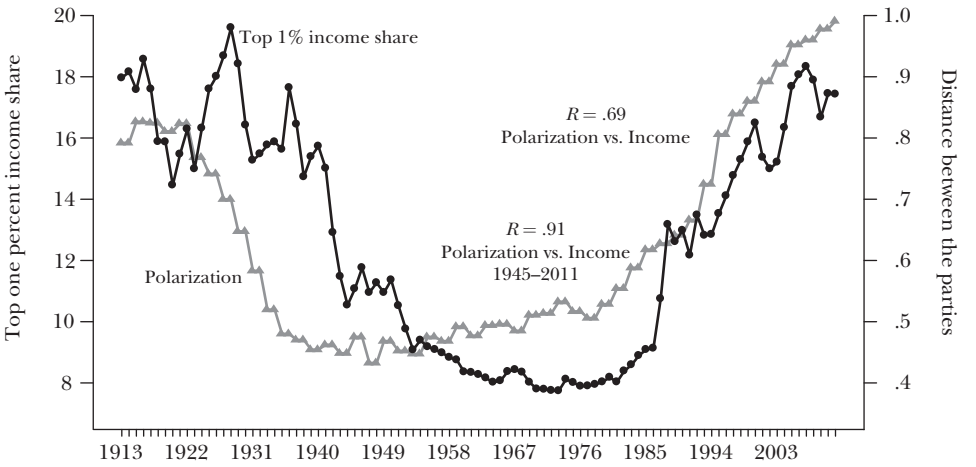
There appears to have been a substantial change in how issues of market regulation and fiscal policy have mapped onto the relatively stable ideological positions of Democratic members of Congress. What makes a legislator liberal or conservative evolves over time. Part of the changed mapping is found in the Democrats' embrace of environmental protection, a public and normal good where demand is likely to increase with income. Similarly, in a quantitative analysis of party platforms, Geering (1998) has documented that the Democratic agenda has shifted away from general social welfare to policies that target ascriptive identities of race, ethnicity, gender, and sexual orientation.

The extent of political polarization, as measured by the average difference between the parties, is positively correlated with the extent of inequality. Figure 2 shows the dramatic increases in the difference between the mean Republican and the mean Democrat in the House. Again, results for the Senate are very similar. The figure also shows that polarization and income inequality track each other closely over the past 100 years; indeed, when this measure of political polarization is lagged ten years, thus making it a leading indicator, its correlation with the income share of the top 1 percent exceeds 0.9.[4]

This correlation raises questions of cause and effect. What causes increased political polarization? How might greater political polarization reduce the efforts of government to push back against greater income inequality? These are important questions for research. In later sections of this paper, we will discuss the arguments suggesting that lower support for policies that would influence inequality may result from an increase in polarization coupled with a greater adherence to free-market ideology, which is in turn supported by campaign contributions from the rich.

[4] Polarization is not only a phenomenon at the federal level. Shor and McCarty (2011) show that many state legislatures are as polarized as Congress and have become more polarized over the last 15 years. Controlling for region, there is also a statistically significant cross-sectional correlation between state legislative polarization and the 80/20 wage ratio. Also see Garand (2010) for evidence that income inequality at the state level is correlated with the voting of US senators.

Figure 2
Top 1 Percent Income Share and Polarization in the US House of Representatives, 1913–2008



Source: Authors calculations using the polarization data described under Figure 1, and data on income from Piketty and Saez (2013).

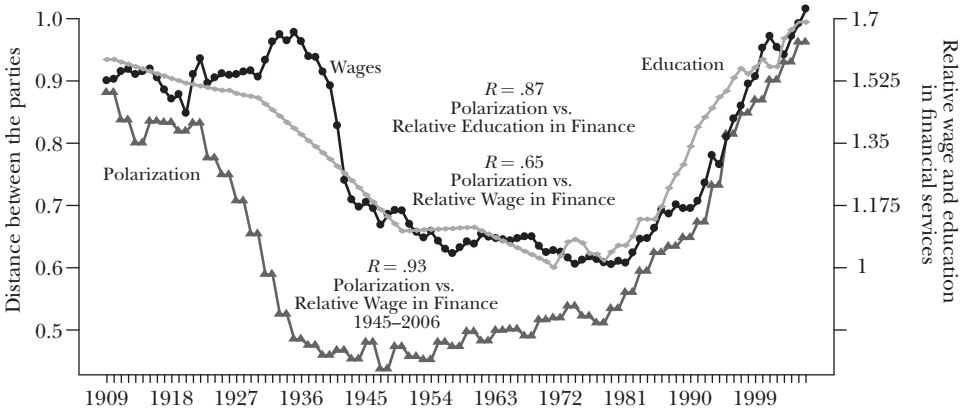
But clearly, efforts to reduce inequality are not especially popular. For example, Barack Obama campaigned in 2008 and again in 2012 to raise marginal tax rates only on the relatively small slice of households earning over $250,000 and eventually in early 2013 accepted increases for only the tiny sliver of households (filing joint returns) at $450,000 and above. There has been no groundswell of political support for sharp reforms of the financial sector, nor for actions to reduce mortgage foreclosures substantially, nor for expanding investment in the human capital of children from low-income households.

It's important to remember that the left-to-right political spectrum described here involves many issues, not just economic ones. When Democrats have won elections, their victory has been due in large part to voters' concerns about issues other than redistribution to the poor. Such issues might include the environment, foreign policy (Iraq), and identity issues for women and the lesbian, gay, and transgender communities. Arguably, the electoral success of the Democrats in 2006 and 2008 owed more to the Bush administration's Iraq policy than to redistributive policies espoused by Democrats.

Voters and citizens also tend to make clear distinctions among the holders of wealth based on how it was acquired. To the extent that public furor is directed against the 1 percent, it is largely directed against the financial sector. We had Occupy Wall Street, but no Occupy Silicon Valley or Occupy Walmart. Political animus toward the 1 percent is largely directed at those perceived to have caused direct harm to the general public, like executives from major financial institutions and BP. Rage is accentuated when the malefactors are

Figure 3

Polarization in the House of Representatives, and Relative Wage and Education in Financial Services Sector, 1909–2006



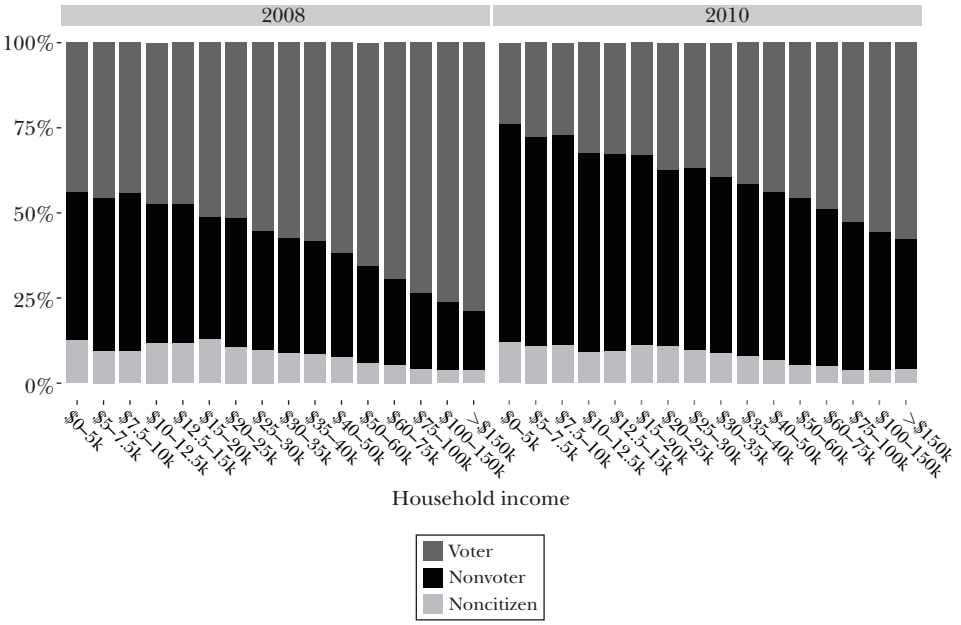
Source: Authors calculations using the political polarization data described under Figure 1, and financial data from Philippon and Reshef (2008).
Note: Figure 3 shows that political polarization in the US House of Representatives tracks wages and educational levels in the financial sector relative to the remainder of the economy.

perceived to benefit from taxpayer money, as in the TARP and AIG bailouts of 2008 (McCarty, Poole, and Rosenthal 2013). But there is no widespread sense that high incomes are illegitimate per se. Executive compensation went essentially unscathed in the Dodd–Frank financial regulation legislation passed in 2010. Entrepreneurial wealth derived from providing valued goods and services is admired, even revered.

Hard evidence on mass public attitudes toward the 1 percent is thin. But a 60 Minutes/ *Vanity Fair* poll in July 2010 asked, "Which one of the following professions do you think most deserves the large salaries its members receive?" The responses from 1,167 telephone interviews were: Bankers, 12 percent; Professional athletes, 15 percent; Film stars, 9 percent, Lobbyists, 5 percent, Internet entrepreneurs, 33 percent; None of the above (an unsolicited, volunteered response), 14 percent; and Don't know/ No answer, 15 percent. In short, Internet entrepreneurs, who populate far more of the Forbes 400 wealthiest Americans than athletes, film stars, and lobbyists, are viewed as deserving. Even bankers, less than two years after the collapse of financial markets, found some support. The framing of the question limits what we can draw from the responses—"Who most deserves . . .?" allows for the interpretation that none deserve it very much. But at least we can say that these responses are hardly a smoking gun of widespread public resentment over rising income inequality.

However, there are some tantalizing linkages from our measure of political polarization to the financial sector. Figure 3 shows that political polarization tracks

Figure 4
Voter Turnout by Household Income and Citizenship, 2008 and 2010



Source: Current Population Survey (2008:*N*=80,443; 2010:*N*=101,338).
Note: Figure 4 shows the relationship between voter turnout and voter household income and citizenship for the 2008 presidential and 2010 midterm elections.

wages and educational levels in the financial sector relative to the remainder of the economy. (Again, the correlation of polarization between both of these series is maximized at 0.9 by lagging polarization by 10 years.) Philippon and Resheff (2008) in turn show that these financial variables go hand in hand with financial deregulation; that is, a deregulated financial sector attracts those with more education, who in turn are paid a higher wage. The top 1 percent contains a sizable share of individuals from the financial sector. Thus, a set of interconnections seem to exist here between political polarization, policy choices affecting the financial sector, and greater inequality as one outcome of those policy choices.

Voting Participation and Inequality

If those with lower incomes are less likely to vote, then the political system will be less responsive to a rise in inequality. The relationship between voter turnout and income (and citizenship) is shown in Figure 4 for the 2008 presidential and 2010 midterm elections. The data is drawn from the large samples in the November Current Population Survey, which asks questions on citizenship, turnout of citizens,

and household income.[5] In the figure, the dark grey bars at the top represent the voters; the black bars in the middle, nonvoting citizens; and the light grey bars at the bottom, noncitizens.

Two effects are worth noting. First, higher fractions of the poor are noncitizens. Second, among the poor who are citizens, turnout is very low. Fewer than half the households with incomes under $15,000 reported voting in the presidential election of 2008, even though turnout of the poor increased over the 2000 level (shown in McCarty, Poole, and Rosenthal 2006, figure 4.1.) In contrast, over four-fifths of those with incomes over $150,000 reported voting. The contrast is greater for midterm elections. Turnout for the 2010 midterm election decreased only slightly from levels in 2008 for top incomes but decreased substantially for those with household incomes below $15,000 so that only about one-third of that group reported voting. In 2008, those reporting incomes above $150,000 represented 9.9 percent of the voters as against 8.0 percent of the voting age population. In 2010, the numbers were 9.6 and 7.4 percent.

Schlozman, Verba, and Brady (2012) show that there is a significant socioeconomic skew to all forms of political participation with the exception of participating in a protest—an activity in which very few Americans participate. Voting tilts the policy scales in favor of top incomes. Another important distortion comes from campaign contributions, which we discuss in the next section.

Campaign Contributions and Political Outcomes

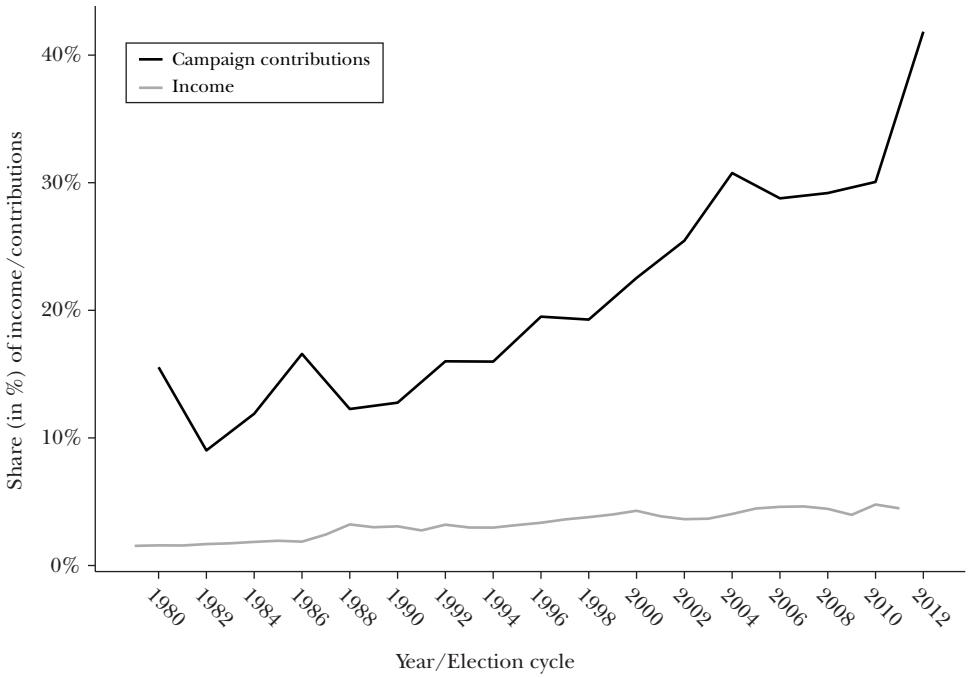
Campaign contributions by individuals have grown over time, with 3,138,564 individuals making itemized contributions in 2012 compared to 224,322 in 1980. But this increased participation has also been marked by increased inequality in contributions. Figure 5 shows that the share of total income received by the top 0.01 percent of households is about 5 percent but that the share of campaign contributions made by the top 0.01 percent of the voting age population is now over 40 percent.[6]

This change over several decades does not seem to be primarily driven by legal factors governing donations: broadly speaking, the legal framework for wealthy individuals to donate unlimited amounts was set in place decades earlier by

[5] There is overreporting of citizenship and turnout and probably underreporting of income. Ansolabehere and Hersh (2012) show that, in sample election surveys, turnout is heavily overreported and that the tendency to overreport is increasing in income. However, the same authors show that, for 2008, the Current Population Survey turnout rate among citizens is only 2 percent higher than the turnout of the Voting Eligible Population computed by McDonald (2011). Thus, overreport bias in the Current Population Survey appears unlikely to account for the strong differences shown in Figure 4.

[6] We have used the 0.01 percent rather than the 0.1 percent or the 1 percent because in 1982 less than 0.1 percent of the voting age population made itemized contributions. Although contributing has become more widespread, it has also become more concentrated, with the top 0.01 percent dominating total contributions.

Figure 5
Concentration of Income and Campaign Contributions in the Top 0.01 Percent of Households and Voting Age Population



Source: For income data, Piketty and Saez (2013).
Notes: The dark line tracks the share of campaign contributions in all federal elections donated by the top 0.01 percent of the voting age population. The number of donors included in the 0.01 percent share of voting age population grew from 16,444 in 1980 to 24,092 in 2012. During the same period, the minimum amount given to be included in the top 0.01 percent grew in real terms from $5,616 to $25,000 (in 2012 dollars). The shaded line tracks the share of total income (including capital gains) received by the top 0.01 percent of households. The figure includes individual contributions to Super PACs and 527 organizations but excludes contributions to nondisclosing 501c(4) organizations, which are recorded to have spent approximately $143 million in 2010 and $318 million in 2012, much of which was raised from wealthy individuals. Were it possible to include contributions to nondisclosing 501c(4)'s, the trend line would likely be 1–2 percentage points higher in 2010 and 2012.

Buckley v. Valeo (424 US 1 [1976]). Rather, it reflects the rising wealth of the superrich and an increased willingness to spend large sums on elections.

One stark indication of increased willingness to spend comes from a comparison of the largest individual contributors in federal elections over time. In 1980, the top contributor was Cecil R. Haden, owner of the tugboat operator Bay-Houston Towing, who gave $1.72 million (in 2012 dollars), nearly six times the amount given by the next largest contributor. In 2012, the two largest donors were Sheldon and Miriam Adelson, who gave $56.8 million and $46.6 million, respectively. Other members of the Forbes 400 accompany the Adelsons; 388 current members are on record as having made political contributions. They account for 40 of the

155 individuals who contributed $1 million or more to state and federal elections during the 2012 election cycle.

Another wealthy group is represented by a sample of 4,493 board members and chief executive officers from Fortune 500 firms. More than four in five in this group reported contributions in federal elections, a proportion that looks even higher when observing that many noncontributors are foreign nationals and are prohibited by law from contributing to US political campaigns. Overall, the corporate sample gave over $170 million during the 2012 election cycle. These contributors have given a lifetime average over the period 1979–2012 of $172,513 per individual and a lifetime median of $30,359. While these numbers are miniscule compared to philanthropic giving, the participation rate of the nation's wealthiest individuals is quite astounding when one considers that less than 4 percent of citizens have given in amounts large enough to be itemized. (Small contributions—less than $500 to a candidate before 1990, and $200 since then—are not reported on Federal Election Commission disclosure forms.)

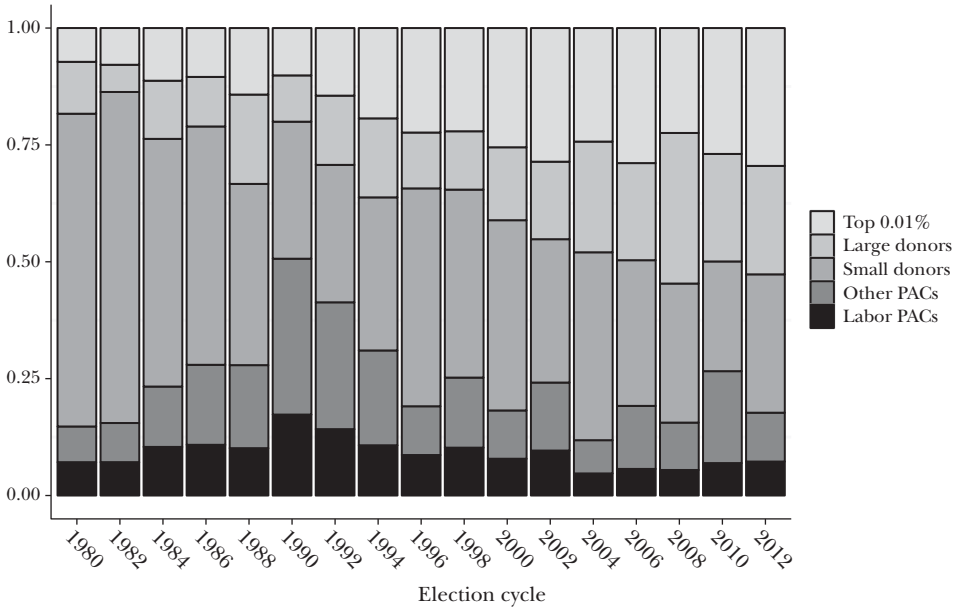
Consistent with Figure 5, there appears to be a growing reliance on the top 0.01 percent of donors. Democrats as well as Republicans rely on big donors. The sources of contributions to Democrats are shown in Figure 6. The Democrats have come to rely, increasingly, on contributions from individuals, particularly big money. Contributions from organized labor, never dominant, have fallen in importance.[7] The relative proportions of funds raised by Democrats from the top 0.01 percent and from organized labor provide a telling comparison. The top 0.01 percent, whose donations had been roughly on par with those of labor during 1980s and early 1990s, outspent labor by more than a 4:1 margin during the 2012 election cycle. While Republicans had a slight advantage in fundraising from the top 0.01 percent during the 1980s, this trend had reversed by the mid-1990s, with Democrats raising more than Republicans from the top 0.01 percent in six out of eight election cycles between 1994 and 2008. Only in the last two election cycles did Republicans regain the advantage in fundraising from the top 0.01 percent.[8] While it is difficult to gauge the effect of the Democrat's reliance on contributions from the wealthy, it does likely preclude a strong focus on redistributive policies.

For individual campaign donors, we have explored who gives how much to whom in terms of the ideology of the donors and candidates. We measure the ideology of donors and candidates jointly in an estimation procedure (for details, see Bonica 2013a, 2013b) reasonably analogous to DW-NOMINATE estimation (see Figures 1–3) for roll call data. Campaign finance scores reported here are based on individual contributions in federal races. Similar results for candidates are also obtained by using PAC ("political action committee") contributions with controls

[7] For a contrary view on the importance of the decline of contributions from labor, see Hacker and Pierson (2010).

[8] In 2012, 62 percent of contribution dollars raised from the top 0.01 percent went to Republicans, accounting for 40 percent of the party's total campaign dollars.

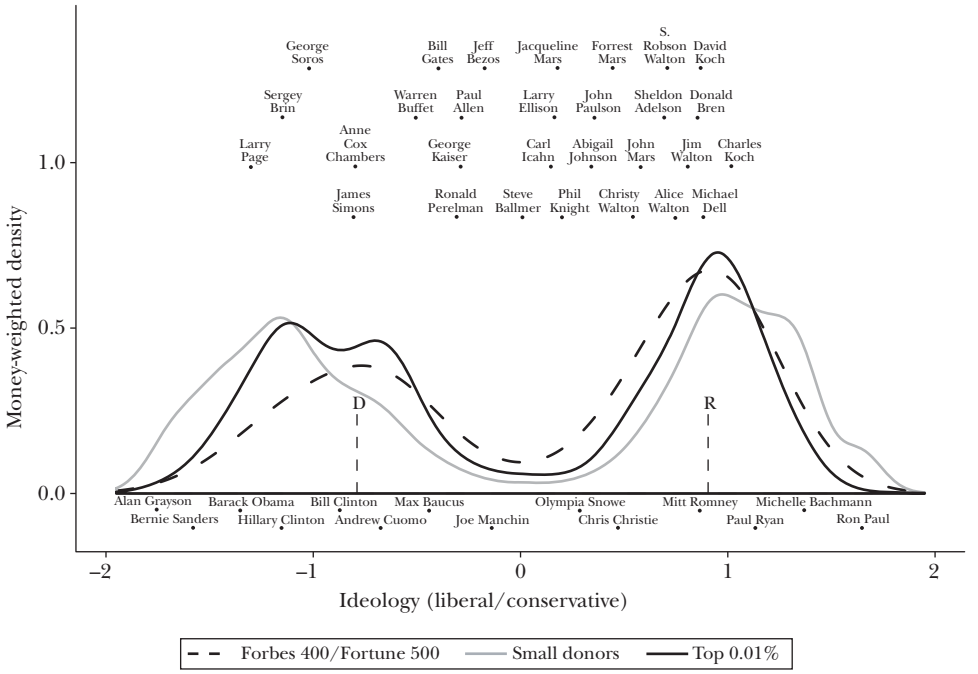
Figure 6
Democratic Fundraising Sources



Sources: The Federal Election Commission and the Internal Revenue Service.
Notes: Includes funds raised by federal candidates, national party committees, Super PACs, federal-oriented 527s, and other independent groups aligned with the Democratic Party (for example, Priorities USA Action and the MoveOn.org 527). The amounts from labor and other PACs include direct contributions to candidates and committees, soft-money contributions to national party committees (prior to 2004), and independent expenditures (including those made through 527s) to support Democrats or oppose Republicans in the general elections. The total amounts raised from individuals are divided into three categories. The first category includes all donations made by individuals included in the top 0.01 percent of donors in the voting age population, similar to Figure 5. The category labeled "Large Donors" includes all individual donors that gave more than $1,500 (in 2012 dollars) during the election cycle but not enough to be included in the top 0.01 percent. The category labeled "Small Donors" includes contributions from donors giving less than $1,500 during the election cycle plus the total funds raised in nonitemized amounts. The cutoff at $1,500 approximates the inflation-adjusted $500 minimum donation required for disclosure in 1980.

for standard covariates such as incumbency and seniority. Figure 7 shows how the top 0.01 percent of donors and smaller donors distribute their donations in terms of the ideology of candidates (including losing candidates). To further explore the ideological leanings of the wealthiest members of the business community, we include a third density of donations from a group composed of the Forbes 400 list of wealthiest Americans and of directors and executives of Fortune 500 companies. In addition to the ideological distribution of dollars allocated by each group of donors, the figure displays the estimated ideal points for several well-known politicians and the 30 richest Americans that have made contributions. The vertical lines

Figure 7
The Ideological Distribution of Dollars from Small Donors and the Top 0.01 Percent



Source: Authors calculation's using "CF" scores, as described in Bonica (2013b), as measures of candidate and contributor ideology. Data on political contributions are from the Federal Election Commission.
Notes: While the DW-NOMINATE scores range from –1 to 1, the Bonica CF scores range from –2 to 2. The kernel densities show the distribution of candidate ideal points weighted by the total amounts raised from small donors (defined as donors giving $500 or less during a two-year election cycle), from the top 0.01 percent of donors, and a group of the very wealthy donors composed of members of the Forbes 400 and directors and CEOs of Fortune 500 firms. The densities are based on contributions made to candidates for federal office from 2004 to 2012. The names above the density plots mark the estimated ideal points of the 30 richest Americans that have made contributions. In order to aid in interpreting the scale, the ideal points for several well-known politicians are included below the densities (read along the *x*-axis only; the varying heights of these points do not mean anything). The vertical bars show the average positions of Democratic and Republican candidates.

in the figure show the mean positions of Democratic and Republican members of Congress during the period.

Although contributions from the Forbes 400/Fortune 500 skew somewhat to the right, a sizable percentage of their contribution dollars go to support candidates who are left of center. In coding their positions, there are three channels for placing these wealthy individuals in the moderate category. First, the individual might donate nearly exclusively to moderates of one of the two major parties. Second, the individual might spread contributions across moderates in both parties. Third, the contributions could be dispersed across the ideological spectrum. The third channel is unimportant empirically. As a rule, individual donors

exhibit high levels of partisanship in their giving patterns, and the super-rich are no exception. The vast majority have given at least 90 percent of their contribution dollars to one or the other party. Of those who split their contributions between parties, most give predominantly to the moderate wings of both parties. As such, the bimodal density of contribution dollars reflects the ideological diversity in the contributors, rather than individuals purchasing access or otherwise acting strategically in ways that cause them to disperse their contributions widely. The level of ideological diversity is apparent in the estimated ideal points of the 30 richest Americans, which are displayed above the densities. Only three of the 30—George Soros, Larry Page, and Sergey Brin—would be placed in the "progressive" wing of the Democrats. Similarly, only one of the 30, Charles Koch, is to the right of the mean Republican member of Congress. Nonetheless, the densities in Figure 7 are bimodal, with the big contributions appearing as moderate only in the limited sense that they are less polarized than the small contributions.

The ideological diversity of corporate elites is not simply a function of firm-specific incentives that would cause a firm to stake out an ideological position. Bipartisan boardrooms are the norm. One way a firm maintains political access is to have both high-profile Democratic donors and high-profile Republican donors within the firm. A consequence of the ideological diversity of the corporate community is to help keep the political financing system competitive for both parties, while at the same time ensuring that firms remain well connected in both parties. Among the directors of JPMorgan Chase, to cite a typical example, are Crandall Bowles, former chair and chief executive officer of Springs Global and wife of Erskine Bowles, and James Bell, a former Boeing Co. executive. Crandall Bowles and James Bell have ideal points as far left as Bill and Hillary Clinton. The board also contains former ExxonMobil chief executive officer Lee Raymond and former Johnson & Johnson chief executive officer William Weldon whose positions are as far right as Paul Ryan. JPMorgan Chase chief executive officer Jamie Dimon is moderate left.

We have seen that the 30 richest Americans are diverse but relatively moderate. This observation is echoed in comparing the ideological densities of contribution dollars from Forbes 400/Fortune 500 individuals and small donors. While the contributions from the Forbes 400/Fortune 500 go disproportionately to candidates who are interior to the party means, contributions from small donors go disproportionately to candidates on the ideological extremes. The distribution of contributions from the top 0.01 percent of the voting age population is more extreme than the Forbes 400/Fortune 500 but less extreme than the small donors.

While Forbes 400/Fortune 500 individuals are more moderate than small donors, their contributions tilt toward Republicans. The top 0.01 percent of donors give pretty evenly to Democrats and Republicans. Why are big donors a bit less polarized than other donors? With regard to the left, the answer seems straightforward: those who are left and rich still have a limited tolerance for redistribution. For the right, the answer seems a bit more complicated. The wealthy are often more pro-business than pro-market (a distinction explored by Zingales 2012), and are often not anti-statist.

This section has emphasized political spending in the form of campaign contributions. We should also note that even more is spent on lobbying. While a good deal of lobbying reflects corporate competition—for example, will Citicorp or Wells Fargo be allowed to purchase Wachovia—much of it gets directed in ways that increase inequality. One example is the persistence of the carried interest provision for hedge funds, often championed by liberal Democrat Senator Charles Schumer of New York. Another is the subprime mortgage origination industry, which was defended by lobbyists and does not seem to have helped the poor. Black homeownership rates in 2012 have fallen back to their level in the mid-1990s (McCarty, Poole, and Rosenthal 2013). Originators do seem to have benefited. Igan, Mishra, and Tressel (2009) have shown that the bulk of the lobbying on subprime was carried out by financial institutions that offered the riskiest products and that turned out to be most likely to fail.

We would also be amiss not to emphasize the role of the revolving door between politics and the private sector as another nontrivial form of political expenditure. For example, Robert Rubin, Treasury Secretary under Clinton, went to Citicorp from the US Treasury. The recently confirmed Obama Treasury Secretary, Jack Lew, after serving during the Clinton administration and doing a highly compensated stint at New York University, was chief operating officer of Citicorp from 2006 to 2008. One view is that these people are dedicated public servants who sacrifice financial gain to serve the American public. The other is that the revolving door, even with the best of intentions, clouds their decision making. Revolving door jobs, lobbying, and campaign contributions by the rich, when coupled with free market ideological proclivities in the voting population, are likely to have abetted the increase in inequality.

We might wonder what these individuals receive for their contributions? A decade ago, Ansolabehere, de Figueiredo, and Snyder (2003) asked in this journal, "Why is There So Little Money in U.S. Politics?" They concluded that money was largely directed at ideological consumption rather than buying political benefits; for an opposite view, see Gordon, Hafer, and Landa (2007). The question we focus on is a little different: Why have total contributions and the share of total contributions from the top .01 percent risen so much in the last few decades? One possibility is that campaign contributions are just another fad for the rich in the twenty-first century. Standard economic arguments are that rational people won't make large contributions in the hope of different policy outcomes for the same reason that rational people will not expect their personal vote to influence the outcome of elections.[9] However, when contributions become very large, then the notion that your contributions will be completely overshadowed by others may

[9] There is, moreover, an ongoing academic debate about the effect of campaign spending on electoral outcomes. For the no-effect position, see Levitt (1994). For an opposing view, see Erikson and Palfrey (2000). We simply ask the reader to do the thought experiment as to what American politics would look like if one party were permitted no expenditure on polling, data analysis, grass roots organization, and television advertising.

not be as true. Large contributors may also be serving as "bundlers" (fundraisers who solicit checks from other individuals and then pass the checks along to candidates and committees), similar to the collective funding of industry lobbies such as the American Bankers Association. Contributions may help carry the day on very specific issues that relate to income inequality—like the provisions that allow the "carried interest" income received by private equity investors to be taxed at a lower capital gains tax rate rather than the higher marginal income tax rate.

More broadly, there does seem to be evidence that members of Congress represent the views of their high-income constituents much more than those of low-income ones (for example, Bartels 2008).[10] Gilens (2012) presents considerable evidence showing that the policy outputs of the US national government are far more responsive to preferences of high-income voters, especially in policy domains where the opinions of rich and poor diverge. When the rich and poor (90th and 10th percentile incomes) disagree more than 10 percentage points on a policy question, the odds of a policy change are completely unresponsive to views of the poorer voters. But if 80 percent of high-income voters support the change, it has a 50 percent chance of passing compared to only a 32 percent chance of passing with 80 percent support from the poor.[11]

Political Institutions and Reactions to Inequality

The United States is more majoritarian (winner takes all) in its political institutions than many nations, but also has a higher degree of inequality, which suggests the possibility that less-majoritarian systems may in fact have less inequality. Empirically, the use of proportional representation—in which at least some members of the legislature are chosen in proportion to the vote of an area, not in elections in single-person districts—correlates with lower levels of economic inequality. Austen-Smith (2000) and Iversen and Soskice (2006) match this stylized fact in theoretical models. Although the proposed mechanism varies, the central insight is that plurality rule (where the candidate with the most votes wins) is conducive to two-party political systems (Duverger 1954). In turn, political bargaining in two-party systems tends to privilege coalitions of the middle and upper classes against the poor. A related institutional distinction compares parliamentary regimes, which lack a strong separation of powers between the legislative and executive branch

[10] The Bartels (2008) claim is based on correlating DW-NOMINATE scores with public opinion data. The claim, however, has been criticized on methodological grounds by Bhatti and Erikson (2011). Using much larger sample sizes, Tausanovitch (2011) confirms the Bartels claim but then argues that the poor are nonetheless represented. His argument is that even if the DW-NOMINATE scores correlate more highly with the expressed preferences of those with high incomes, the scores also respond to the proportion of poor constituents. Resolving the representation controversy is undoubtedly complicated by the fact, documented above, that the rich are a very heterogeneous group.

[11] He does find however that the gap in responsiveness of the preferences of the rich and poor is largest on noneconomic issues such as abortion and homosexuality, areas in which high-income voters tend to take the liberal positions.

(because the executive is selected by the party with a majority in the legislative branch), to presidential-congressional ones, with separation of powers. Here the theoretical analysis of Persson, Roland, and Tabellini (2000) points to less redistribution in presidential-congressional regimes like the United States.

It is an open question as to the robustness of the conclusions of various theoretical models of the politics of inequality. In at least some cases, the theoretical predictions are not robust to changes in assumptions. For example, Olszewski and Rosenthal (2004) show that political coalitions and after-tax outcomes can shift dramatically in certain models when a quasi-linear utility function is replaced by a Cobb–Douglas utility function.

In the past, most political economy models that depart from majoritarian settings imposed some form of compromise bargaining in the political process (for some different approaches, see Baron and Ferejohn 1989; Alesina and Rosenthal 1995; Grossman and Helpman 1999). However, as compromise has been expunged from the Washington political dictionary, political science models of policy formation have largely turned to *pivot* models following the seminal work of Krehbiel (1998). Pivot models embed the institutional structure of the political system of the federal government. In order to change policy from the status quo, legislation must pass over hurdles that include passage by simple majority in both Houses of Congress, the 60 votes needed to shut down possible filibusters in the Senate, and the two-thirds majorities in both houses needed to override vetoes by the president, followed by challenges in the courts and possible foot-dragging or noncompliance in the states. Models of these institutional processes identify the actors whose support is necessary for policy change. These actors are known as *pivots*. Potential pivots include the median representative in the House (or perhaps, the median member of the majority party), the filibuster pivot in the Senate, and veto override pivots in both houses. At a given time, which of the potential pivots will be the pivotal legislator depends on the distribution of liberal–conservative preferences in the two chambers. The internal rules of the two Houses of Congress also constitute hurdles.

Of course, the membership of the US Congress is not simply determined by popular majorities, which leads to additional hurdles. Article V of the US Constitution has one unamendable provision that gives each state two votes in the Senate, without regard to population. The provision creates a situation in which some senators represent far fewer people than others. A pivot hurdle on the Affordable Care Act was Ben Nelson, the Democratic senator furthest to the right. Nelson is from Nebraska, which has its two of the 100 senators but only three of the 435 representatives. Majority representation in the House is distorted by geographic sorting of the population and, to a lesser extent, by gerrymandering that favors the Republicans. Gerrymandering explanations are rampant in the media but they should not be exaggerated. Democrats lost the House in 2010, but they won it with the same apportionment in 2006 and 2008. While gerrymandering has a Republican bias, it does not seem to contribute to polarization. Both geographic sorting and gerrymandering are discussed in more detail in McCarty, Poole, and Rosenthal (2009).

Multiple pivots imply that status quo policies may become gridlocked. The range of gridlocked policies expands theoretically, and has also expanded empirically, as polarization of legislative preferences increases. Expanded polarization has resulted in diminished legislative output (McCarty, Poole, and Rosenthal 2006, pp. 177–183).

Even if gridlock is sometimes overcome, policy change will need to be moderated to attract the support of critical pivots. In the first two years of the Obama presidency, for example, the stimulus package, the Patient Protection and Affordable Care Act, and the Dodd–Frank Act all passed. However, each piece of legislation was sharply limited by the need to obtain 60 votes to avoid a Senate filibuster. All three initially passed the Senate with exactly 60 votes. To obtain passage of the stimulus bill, the administration needed to cut $200 billion of assistance to state and local governments. Among the many concessions made in Dodd–Frank, a tax on banking transactions was removed to obtain the pivotal vote of Republican Senator Scott Brown of Massachusetts (McCarty, Poole, Romer and Rosenthal 2010). The Patient Protection and Affordable Care Act did raise certain taxes on those with high incomes, but in many ways it was constructed to appease insurance and pharmaceutical companies. Taxes were also increased on high incomes in the "fiscal cliff" legislation of 2012, but rates remain well below the levels in effect before the Reagan presidency. Moreover, this legislation ended payroll tax cuts that had provided low-income tax relief during the Great Recession. In the 2010 midterm elections, the Republicans regained control of the House of Representatives. The pivot was thus shifted from the relatively moderate 60th (from the left) position in the Senate to the quite conservative House median. The 2012 tax legislation reflected this shift.

Legislative gridlock also allows policies to drift. In principle, this effect could increase or decrease inequality. For example, an unindexed minimum wage falls in real terms if not updated by law. In contrast, unindexed income tax rates lead to "bracket creep" and higher taxes as inflation gradually pushes incomes into higher tax brackets. But in practice, indexing seems to have been applied more often where it prevents the rich from suffering the effects of a drifting policy: income tax brackets are indexed for inflation, but minimum wages are not.

Gridlock may also produce bureaucratic drift: those who are regulated can turn to a variety of different regulators at the federal level, as well as in the states and the courts, and when the legislature is gridlocked, these regulators become freer to pursue their own policy objectives without fear of legislative override (Ferejohn and Shipan 1990). This dynamic was very important in financial deregulation (McCarty, Poole, and Rosenthal 2013). For example, the barriers between commercial banking, investment banking, and insurance enshrined in Glass–Steagall were whittled to nearly nothing by regulators and court decisions well before Congress officially repealed the law. Regulatory conflict between the Federal Reserve, the Office of the Comptroller of the Currency, and the FDIC also facilitated lobbying as a gridlocked Congress first took two years from the financial crisis to enact Dodd–Frank and then produced a bill with substantial regulatory discretion (Bair 2012). Financial

industry lobbyists have to date been very successful in modifying and delaying new rules, many of which are still being written. But legislative polarization makes it exceedingly unlikely that Congress will be able to revisit Dodd–Frank and correct its shortcomings or overturn regulatory decisions.

Conclusion: Politics and Reacting to Inequality

Economists may be inclined to downplay the role of politics and public policy in generating and perpetuating inequality. Of course, economic shifts like globalization, technological shifts like information and communications technology, and social changes like the greater propensity of high-income earners to marry each other can lead to increases in inequality. But these changes are not orthogonal to political decisions that are taken—or decisions not taken.

For example, Piketty and Saez (2003) suggest that sharp changes in income inequality have been driven by destruction of assets in economic depressions and wars and by changes in fiscal policy. The economic policies that abet and respond to depressions are the subject of political processes, while "war is the continuation of politics by other means" (von Clausewitz 1832 [2009]). Open immigration policies, as argued by some economists like Borjas (1999), may increase inequality. The influence of globalization reflects political decisions about the free movement of goods and services. Globalization of financial services allows the wealthy to benefit from tax avoidance strategies by moving funds abroad or not bringing them home, depending on the policies embedded in the US tax code and the enforcement vigor of the IRS. The pace of technological development is intertwined with public policies related to innovation and intellectual property, and the applications of technology are intertwined with how, often slower-moving, regulatory agencies struggle to keep pace. The complexity in derivatives that emerged, with its disastrous consequences for the global economy, required both explicit acquiescence by the government in the form of legislation such as the Commodity Futures Modernization Act of 2000 and implicit acquiescence in the failure to exercise supervision. The deregulation and lack-of-regulation of the financial sector in the 1990s and early 2000s helped to shape the practices and compensation in that industry.

Thus, our general argument is that politics and public policy need to be considered explicitly in any discussion of the causes and consequence of inequality. The feedback from politics can also arise as a result of laws that were frozen in place, or that were never enacted. Top marginal tax rates in the United States declined sharply from the 1960s through the 1980s, but since then have fluctuated only mildly. US spending programs are increasingly focused on the elderly rather than on broader assistance to those of working age or those with low incomes. Overall, the kinds of government policies that could have ameliorated the sharp rise in inequality have been immobilized by a combination of greater polarization, lack of voter participation, feedback from high-income campaign contributors, and political institutions that must overcome a series of key pivots before making significant changes.

References

Alesina, Alberto, and Howard Rosenthal. 1995. *Partisan Politics, Divided Government, and the Economy*. Cambridge University Press.

Alvarez, R. Michael, Delia Bailey, and Jonathan N. Katz. 2008. "The Effect of Voter Identification Laws on Turnout." California Institute of Technology Social Science Working Paper 1267R. Available at SSRN: http://ssrn.com/abstract=1084598.

Ansolabehere, Stephen, John M. de Figueredo, and James M. Snyder. 2003. "Why is There So Little Money in U.S. Politics?" *Journal of Economic Perspectives* 17(1): 105–130.

Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20(4): 437–57.

Austen-Smith, David. 2000. "Redistributing Income under Proportional Representation." *Journal of Political Economy* 108(6): 1235–69.

Bair, Sheila. 2012. *Bull by the Horns: Fighting to Save Main Street from Wall Street and Wall Street from Itself*. The Free Press.

Baron, David, and John Ferejohn. 1989. "Bargaining in Legislatures." *American Political Science Review* 83(4): 1181–1206.

Bartels, Larry M. 2008. *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton University Press.

Benabou, Roland. 2000. "Unequal Societies: Income Distribution and the Social Contract." *American Economic Review* 90(1): 96–129.

Bhatti, Yosef, and Robert S. Erikson. 2011 "How Poorly are the Poor Represented in the U.S. Senate?" Chap. 8 in *Who Gets Represented?* edited by Peter K. Enns and Christopher Wlezein. Russell Sage Foundation Press.

Bolton, Patrick, and Gérard Roland. 1999. "The Breakup of Nations." *Quarterly Journal of Economics* 112(4): 1057–90.

Bonica, Adam. 2013a. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57(2): 294–311.

Bonica, Adam. 2013b. "Mapping the Ideological Marketplace." *American Journal of Political Science*, forthcoming.

Borjas, George J. 1999. *Heaven's Door: Immigration Policy and the American Economy*. Princeton University Press.

Duverger, Maurice. 1954. *Political Parties*. New York: Wiley.

Erikson, Robert S., and Thomas Palfrey. 2000. "Equilibria in Campaign Spending Games: Theory and Data." *American Political Science Review* 94(3): 595–609.

Ferejohn, John and Charles Shipan. 1990. "Congressional Influence on Bureaucracy." *Journal of Law, Economics, and Organization* 6(0):1–20.

Garand, James C. 2010. "Income Inequality, Party Polarization, and Roll-Call Voting in the U.S. Senate." *Journal of Politics* 72(4): 1109–28.

Geering, John. 1998. *Party Ideologies in America: 1828–1996*. New York: Cambridge University Press.

Gilens, Martin. 2012. *Affluence and Influence: Economic Inequality and Political Power in America*. Princeton University Press.

Gordon, Sanford, Catherine Hafer, and Dmitri Landa. 2007. "Consumption or Investment? On Motivations for Political Giving." *Journal of Politics* 69(4):1057–72.

Grossman, Gene M., and Elhanan Helpman. 1999. "Competing for Endorsements." *American Economic Review* 89(3): 501–524.

Hacker, Jacob S., and Paul Pierson. 2010. *Winner-Take-All Politics: How Washington Made the Rich Richer—And Turned Its Back on the Middle Class*. New York: Simon and Schuster.

Igan, Deniz, Prachi Mishra, and Thierry Tressel. 2009. "A Fistful of Dollars: Lobbying and the Financial Crisis." IMF Working Paper 09/287.

Iversen, Torbe, and David Soskice. 2006. "Electoral Institutions and the Politics of Coalitions: Why Some Democracies Redistribute More than Others." *American Political Science Review* 100(2): 165–81.

Jones, Jeffrey M. 2009. "Most Americans Favor Gov't. Action to Limit Executive Pay." Gallup, June 16. http://www.gallup.com/poll/120872/americans-favor-gov-action-limit-executive-pay.aspx.

Krehbiel, Keith. 1998. *Pivotal Politics: A Theory of U.S. Lawmaking*. Chicago University Press.

Levitt, Steven D. 1994. "Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House." *Journal of Political Economy* 102(4): 777–98.

McCarty, Nolan, Keith T. Poole, Thomas Romer, and Howard Rosenthal. 2010. "The Price of Principle." *Huffington Post,* The Blog, July 20.

McCarty, Nolan M., Keith T. Poole, and Howard Rosenthal. 1997. *Income Redistribution and the Realignment of American Politics*. Washington, DC: AEI Press.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge: MIT Press.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2009 "Does Gerrymandering Cause Polarization?" *American Journal of Political Science* 53(3): 666–80.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2013. *Political Bubbles: Financial Crises and the Failure of American Democracy*. Princeton University Press.

McDonald, Michael P. 2011. "Turnout, 1980–2010." United States Elections Project. http://elections.gmu.edu/voter_turnout.htm (accessed July 10, 2013).

Meltzer, Allan H., and Scott F. Richard. 1981. "A Rational Theory of the Size of Government." *Journal of Political Economy* 89(5): 914–27.

Olszewski, Wojciech, and Howard Rosenthal. 2004. "Politically Determined Income Inequality and the Provision of Public Goods." *Journal of Public Economic Theory* 6(5): 707–735.

Persson, Torsten, Gérard Roland, and Guido Tabellini. 2000. "Comparative Politics and Public Finance." *Journal of Political Economy* 108(6): 1121–61.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118(1): 1–39.

Piketty, Thomas, and Emmanuel Saez. 2013. Tables and figures from "Income Inequality in the United States" (2003), updated to 2011 at Saez's website: http://elsa.berkeley.edu/~saez/.

Philippon, Thomas, and Ariell Reshef. 2008. "Wages and Human Capital in the U.S. Financial Industry: 1909–2006." Working paper, December 2008, with the Excel file with the data series available at http://pages.stern.nyu.edu/~tphilipp/research.htm (accessed July 10, 2013).

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting*. New York: Cambridge University Press.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.

Romer, Thomas. 1975. "Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax." *Journal of Public Economics* 4(2): 163–85.

Schlozman, Kay Lehman, Sidney Verba, and Henry E. Brady. 2012. *The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy*. Princeton University Press.

Shor, Boris, and Nolan McCarty. 2011. "The Ideological Mapping of American Legislatures." *American Political Science Review* 105(3): 330–51.

Tausanovitch, Chris. 2011. "Income and Representation in the United States Congress."http://www.americanideologyproject.com/IncomeRepresentation.pdf.

von Clausewitz, Carl. 1832 (2009). *On War*. Translated by J. J. Graham. Brownstone Books.

Zingales, Luigi. 2012. *A Capitalism for the People: Recapturing the Lost Genius of American Prosperity*. New York: Basic Books.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 125–144

What Is European Integration Really About? A Political Guide for Economists

Enrico Spolaore

As an economic and financial crisis unfolds across the European Union, critics argue that European institutional integration has gone too far, blame misguided political motivations, and assert that the monetary union has failed (for example, Feldstein 2012). On the other side, supporters of European integration attribute the euro crisis to institutional incompleteness—what Bergsten (2012) called a "half-built house." They argue that the solution to Europe's woes should be sought in additional integration: a banking union, a fiscal union, or perhaps even a full political union, and the formation of a federation. In sum, the political design of European institutions is at the center of the current debate about the euro.

In fact, Tommaso Padoa-Schioppa (2004, p. 1), the economist and central banker who played a key role in the birth of the euro, wrote: "[T]he euro was the result of a long-term development that started in the aftermath of World War II. After experiencing political oppression and war in the first half of the twentieth century, Europe undertook to build a new order for peace, freedom, and prosperity. Despite its predominantly economic content, the European Union is an eminently political construct. Even readers primarily interested in economics would hardly understand the euro if they ignored its political dimension." This political guide for economists takes a step back and looks at the creation of the euro within the bigger picture of European integration. How and why were European institutions established? What is European integration really about?

The history of European integration is complicated, with a big cast of actors including governments, technocrats, interest groups, and voters, who in turn pursue a range of economic and political goals. This complexity is reflected in a variety of

■ *Enrico Spolaore is Professor of Economics, Tufts University, Medford, Massachusetts, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is enrico.spolaore@tufts.edu.*

http://dx.doi.org/10.1257/jep.27.3.125 doi=10.1257/jep.27.3.125

interpretations by political scientists and political economists (for overviews, see Gilpin 2001, chap. 13; Eichengreen 2006; Sadeh and Verdun 2009). This article discusses facts and theories about European integration from a political economy perspective, building on ideas and results from the economic literature on the formation of states and political unions (for overviews, see Spolaore 2006, 2012). Specifically, we look at the motivations, assumptions, and limitations of the European strategy of partially integrating policy functions in a few areas with the expectation that more integration will follow in other areas in a sort of chain reaction towards an "ever-closer union." The euro with its current problems is a child of that strategy and its limits.

A European Federation?

The idea of a new sovereign federation across Europe goes back a long time, but it received a big push from the first half of the twentieth century. At the end of World War II, the promoters of European integration looked back at the previous decades and saw a continent fragmented in independent and unconstrained nation states which had pursued costly beggar-thy-neighbor policies during the Great Depression and engaged in two major wars. The goal of European integration was to create a system where nation states would no longer follow such unilateral and destructive policies.

In 1943, a group led by Altiero Spinelli founded the European Federalist Movement. In 1946, Winston Churchill argued for the creation of "the United States of Europe" (which in his view did not include Britain). By definition, a federation would have eliminated national borders and international conflict (but not civil conflict) among Europeans. However, no European federation was created immediately after World War II.

Instead, the founding document of European integration is the Schuman Declaration of May 9, 1950, named after France's foreign minister Robert Schuman and inspired by Jean Monnet, a businessman and civil servant who played a crucial role in starting European institutions in the following years. The declaration proposed that "Franco-German production of coal and steel as a whole be placed under a common High Authority, within the framework of an organization open to the participation of the other countries of Europe." The plan was motivated by security as a way "to make it plain that any war between France and Germany becomes not merely unthinkable, but materially impossible." The pooling of coal and steel production was ambitiously defined as "a first step in the federation of Europe."

The Schumann Declaration led in 1951 to the European Coal and Steel Community (ECSC) among six countries: France, West Germany, Italy, the Netherlands, Belgium, and Luxembourg. The ECSC was then used as the institutional template for two proposed communities: the European Defense Community and the European Political Community, which included the formation of a common army, a common budget, and common institutions with significant legislative and executive powers. It would have basically amounted to a European federation (Moravcsik 1998; Rector 2009). A treaty was signed among the six countries in 1952 but failed to obtain ratification in the French parliament and never took effect. In 1955 several

politicians, including Jean Monnet, created an "Action Committee for the United States of Europe." But, again, no United States of Europe actually formed.[1]

The fundamental reasons behind these failures to form a federation have bedeviled the supporters of a United States of Europe, then and since. Two issues are key to understanding the beginning of the integration process, its setbacks, and the following path of European integration. One issue is a general problem in political economy: the trade-off between costs and benefits when heterogeneous groups are politically integrated under a common authority. The other issue involves the particular role of Germany, the country that played a central role in World Wars I and II.

The Political Economy of Heterogeneous Populations

The formation of a European federation across heterogeneous populations, sharing diverse social and economic structures, languages, cultures, and identities, would come with several benefits, but also with high costs. The trade-off between these costs and benefits is central to the political feasibility and stability of institutional integration among these populations (Alesina and Spolaore 1997, 2003).

Potential benefits from full political unification include economies of scale in the provision of federal public goods, such as defense and security, and the ability to internalize positive and negative externalities over a large area. A European federation with its own budget and redistribution policies could also provide insurance against asymmetric shocks that only affect some of its regions, whether natural like an earthquake or man-made like the bursting of a housing bubble. These benefits from fiscal federalism are often stressed when comparing Europe to the United States (for example, Sala-i-Martin and Sachs 1992; Krugman 2012) and are now at the forefront of the debate about the European sovereign debt crisis (Lane 2012).

However, political unification comes with significant costs when various groups speak different languages, share different cultural norms and identities, and have different preferences for public policies and institutions that cannot be decentralized at the sub-federal level (Alesina and Spolaore 1997, 2003). Among those institutions is the ultimate "public good": the federal government itself, with all its constitutional and legal traits, policies, official language(s), and so on, about which German or Dutch people may have very different views from those prevalent in France or Italy.

A growing literature has explored the links between measures of heterogeneity and political outcomes such as the provision of public goods, the extent of redistribution, the quality of government, and the likelihood of civil and international conflict. Microeconomic evidence links ethnic heterogeneity to underprovision of public goods at the local level (Alesina and La Ferrara 2005). There is also macroeconomic evidence of negative correlations between ethnolinguistic fractionalization and government performance, although causality and robustness are less clear-cut (Alesina, Devleeschauwer, Easterly, Kurlat, and Wacziarg 2003). In addition, Montalvo and Reynal-Querol (2005) and Esteban, Mayoral, and Ray (2012) find that

[1] Subsequent less-ambitious attempts to integrate European defense and foreign policy have not been very successful either. For instance, see Alesina and Perotti (2004) for a critical discussion of the more recent experience of the European Union in these and other areas.

ethnolinguistic polarization is associated with civil conflict. Desmet, Ortuño-Ortín, and Weber (2009) show that when distances between languages are accounted for, linguistic diversity has a significant negative effect on redistribution. Desmet, Ortuño-Ortín, and Wacziarg (2012) find that deep linguistic distances are good predictors of civil conflict and redistribution, while even finer distinctions between languages, such as those among different dialects, matter for economic growth and public goods provision. The bottom line of this literature is that measures of ethnic, linguistic, and cultural diversity have significant effects on policy outcomes, redistribution, and the provision of public goods. A European federation would be quite heterogeneous by most of these measures and likely to face significant political costs when choosing common public goods and policies at the federal level.[2]

The example of defense and security—which played a fundamental role in Europe's early attempts to integrate—can illustrate these issues. These public goods have high economies of scale, but also high heterogeneity costs stemming from diverse preferences across populations. Military power has historically been a central tool to ensure a government's monopoly of legitimate use of coercion over a territory. Integration of defense and security under one authority usually goes hand in hand with the centralization of this monopoly of coercion—that is, with the formation of a sovereign state or federation (Alesina and Spolaore 2003). However, different populations with different histories, cultures, and identities are likely to disagree over the type of government in charge of such a federation. Moreover, coercion can then be used to collect taxes, finance a larger set of other public goods, and redistribute resources across different groups. This redistribution is more likely to be resisted when groups are different not only economically but also along ethnic and linguistic lines. For instance, western Germans may be more willing (or less unwilling) to redistribute resources to eastern Germans than to Greeks or Italians. Consequently, centralized provision of defense and security across large and diverse populations usually takes place when dictatorial rulers are able to ignore the heterogeneity costs of the populations they conquer, and/or when there are overwhelming benefits of scale from defense that offset high heterogeneity costs (Alesina and Spolaore 2005, 2006). The two most successful federal republics, Switzerland and the United States, emerged in response to external security threats, and the unification of Germany in the nineteenth century resulted from conquest by Prussia (Riker 1964; Gilpin 2001).

Military and political union is not the only way to deal with security threats. Heterogeneous sovereign states can benefit from economies of scale in defense by forming military alliances while still maintaining their political and fiscal independence. But military alliances, where each state can autonomously decide its own level of military spending and pay for it, can lead to undersupply of defense from the perspective of the whole alliance because of free riding (for a discussion, Spolaore 2012). Western Europeans failed to form a federation even when faced with an

[2] Fractionalization is maximized when each individual belongs to a different group, while polarization is maximized when there are only two large groups of equal size. A larger European federation formed by many groups would be more fractionalized but less polarized than a smaller federation dominated by a couple of groups (for example, Germans and French).

existential threat from the Soviet Union and relied instead on an international alliance (NATO) where issues of undersupply and free riding were in part addressed by the dominant role of the United States.

If heterogeneity can explain failures to integrate in the past, does it need to be an obstacle to future political integration? Over time, couldn't a federal Europe change political and social interactions and affect cultures and identities among Europeans, leading to a shared identity within a "European nation"? After all, nineteenth century France famously turned "peasants into Frenchmen" through public policies and modernization (Weber 1976).

This question is part of the broader debate on the persistent political and economic effects of historical and cultural traits, and the extent to which culture itself can be changed by policies and institutions (for recent discussions, Bisin and Verdier 2010; Spolaore and Wacziarg 2013).[3] In the long run, people can learn new languages, modify their cultural traits and identities, and transmit different traits to their children in response to changing incentives, including public policies. However, it is at best a gamble to hope that political integration of modern democratic nations will lead to cultural integration. Historically, nation building and attempts to "homogenize" populations were implemented by rulers of undemocratic societies who had an interest in reducing heterogeneity costs in order to maximize their own rents (Alesina and Spolaore 2003, pp. 76–78) or pursue their own preferences (Alesina and Reich 2013). Realistic supporters of European integration understand that convergence of political preferences through reduction of linguistic and cultural barriers, if it is going to occur at all, will be a slow and gradual process, which should take place naturally and consensually.

For Europeans, heterogeneity has been a source of benefits as well as of costs. When people have different preferences and traits, societies can benefit economically and culturally through specialization, learning, and exchange of goods and services, as well as ideas and innovations. Benefits from heterogeneity, however, are mostly about interactions over *rival* goods, not public goods, which are nonrival. *Similar* preferences over the same rival goods can lead closely related groups to conflict and war (Spolaore and Wacziarg 2012), while *different* preferences over rival goods can facilitate peaceful exchanges and a better allocation of resources. In contrast, diverse preferences over public goods, like a federation's government, laws, and public policies, will be much harder to reconcile because one kind must apply to everyone within the federation, whether everyone likes it or not. As a result, heterogeneity of preferences is mostly beneficial when people interact about rival goods, but costly when sharing nonrival goods. This is an important reason why, as we will see, the European project has been much more successful when fostering economic exchanges and a common market, while it has stalled when attempting to pool "federal" public goods, such as defense and security.

[3] There is also an extensive political literature debating whether the social and political relevance of ethnic and linguistic divisions can be altered by politics and institutional change (for discussion, see Fearon 2006).

The Role of Germany

In hindsight, as we look back at the 1952 treaty that would have established a European Defense Community and a European Political Community, what's perhaps more surprising is not that France rejected it, but that other states ratified the treaty. A reason is that the other two largest states at that time, West Germany and Italy, had just emerged from a severe military defeat and faced significant constraints to their own defense and foreign policy. West Germany was the more extreme case: a divided country, technically under military occupation until 1955. In those circumstances, the costs of constraints on German sovereignty by pooling defense and security were low and could be traded against other political and economic benefits. As Germany's status as a sovereign state "normalized" over time, its incentives to join a security-based federal union decreased.

The agreement for a European Coal and Steel Community is often interpreted from a similar perspective. According to Milward (1984), France proposed the coal and steel community to constrain German control of its own industry in response to US plans in 1949 to allow a Germany relatively free of Allied supervision. Germany agreed to the Schuman plan because, by sharing management of its coal and steel, it could obtain important concessions, such as "the removal of ceilings on permissible levels of industrial production" (Eichengreen 2006, p. 802). According to Berger and Ritschl (1995), French access to German coal was "the most important element of the Monnet Plan for France's reconstruction."

These examples illustrate a continuing issue in the history and politics of European integration: the extent to which European supranational institutions can be interpreted as tools to constrain German power in the interest of its neighbors, especially France. This theme has come to the forefront again with the creation of the euro. A popular view is that giving up its currency was the price that Germany had to pay to overcome France's opposition to German reunification (Garrett 1993; Marsh 2011), a deal summarized by the witticism quoted by Garton Ash (2012, p. 6): "[T]he whole of Deutschland for Kohl, half the deutsche mark for Mitterrand." Literally taken, as a quid pro quo, this interpretation is not held by most scholars (Sadeh and Verdun 2009). It is questionable that a French threat to veto the reunification of Germany could be credible. Moreover, key decisions about the single currency had already been taken before the fall of the Berlin Wall in 1989, and German politicians and interest groups (like exporters) had other strong reasons to favor a monetary union (Moravcsik 1998; Frieden 2002).

However, it is not fully coincidental that the implementation of the euro took place during and right after German reunification and the opening of political and economic relations between Western and Eastern Europe. Germany's chancellor at that time, Helmut Kohl, viewed the euro as a big step in the broader process of European integration, which he considered essential to reassure Germany's neighbors about his enlarged country's commitments to peace, security, and economic cooperation (Garton Ash 2012). And even though the process leading to economic and monetary union had started before the fall of the Berlin Wall, a detailed analysis of the interactions among key participants in the negotiations

show that German reunification led to a reassessment of the relative payoffs from economic and monetary union, and was used "to reshape . . . negotiations" (Dyson and Featherstone 1999, p. 16).

The increase in Germany's potential power might also have affected the borders of the future euro area, making it much larger than predicted by efficiency criteria, such as the theory of optimal currency areas. For example, Eichengreen (2012, p. 125) mentions the view that France and others pushed for the inclusion of many countries at the "periphery," like Southern Mediterranean countries, to "balance" Germany's larger size and influence within the monetary union.

Whether these *Realpolitik* interpretations are fully persuasive, the French government saw a close link between German reunification and European integration. According to an adviser to the French President, "Mitterrand did not want [German] reunification without advances toward greater European integration, and the currency was the only topic that was open to debate" (*Spiegel* 2010).

How had a monetary issue become "the next step" in the process of European integration? What was (and is) such a process about? To answer these questions we need to go back to what happened after the rejection of the defense and political communities in the mid-1950s.

From Common Market to Economic and Monetary Union: Jean Monnet's Chain Reaction?

From the successful creation of the European Coal and Steel Community and the rejection of the European Defense Community, Jean Monnet and the other supporters of European integration learned a lesson in political realism (Duchêne 1994). Partial integration in narrowly defined areas, such as coal and steel, was feasible, while more ambitious integration in broader areas such as defense and policy coordination would meet too much political opposition. Their next step was the creation in 1957 of a community similar to ECSC for civilian atomic energy (EURATOM) and, more importantly, a European Economic Community (EEC) to set up a customs union: the "common market." The institutions of the three communities were later merged and became known as the European Community. The treaties of Maastricht (1992) and Lisbon (2009) reorganized and replaced the European Community with the European Union.

The Treaty of Rome of 1957 establishing the European common market no longer referred to steps "toward a federation" but included the vaguer objective of laying the "foundations of an ever-closer union among the peoples of Europe." The signatories' main stated goal was "to ensure the economic and social progress of their countries by common action to eliminate the barriers which divide Europe," while claiming that this would strengthen peace and security. To foster those goals, European states created two sets of institutions: supranational institutions such as the European Commission, Parliament, and Court of Justice, and intergovernmental institutions, such as the Council of Ministers and, later, the European Council, formed by the heads of state or government of the member states.

Over time, policy functions have been delegated to European institutions in an increasing range of areas. Nonetheless, national governments have kept control over fundamental decisions and must decide unanimously on all changes to the international treaties that set Europe's informal "constitution." An attempt to establish a formal "Constitution for Europe" failed when it was rejected by French and Dutch voters in 2005.

The history of European integration reflects this tension between the role of supranational institutions and the power of national governments. The conflict is also mirrored by the two most influential political theories about European integration: functionalism and intergovernmentalism. This terminology can be confusing for the uninitiated.[4] In a nutshell, the theories are distinguished by how they answer the question: who is in charge of European integration?

Intergovernmentalists believe that national governments are in charge, and that supranational institutions are tools of the national states, which use them to pursue their own goals. Moravcsik (1993, 1998), an influential proponent of this theory, believes that national governments have built European institutions to pursue the economic interests of their domestic constituencies. In this spirit, Moravcsik (2012) views the euro as an economic gamble, mostly reflecting the interests of powerful national producers. This interpretation fits within a broader literature emphasizing the link from domestic economic interests to national attitudes and policies towards European integration (for example, Frieden 1998, 2002). The political economy approach to regional integration based on domestic economic interests is familiar to the economics profession, and therefore I will not say more here. I will focus instead on the alternative theory of functionalism, which is much less known among economists, even though it has played a significant role in the ideology and practice of European integration and the creation of the euro.

Functionalists believe that European integration is not primarily driven by national governments and their voters, but mostly pushed by elites and interest groups that transcend national boundaries. They stress the role of supranational entrepreneurs and civil servants like Jean Monnet in the 1950s and Jacques Delors in the 1980s and 1990s. The theory is called "functionalism" because it is about the dynamic effects of transferring specific "functions" to supranational institutions: for example, regulation of coal and steel production to the European Coal and Steel Community or monetary policy to the European Central Bank. Although this integration starts in economic areas, integration in one area may well lead to further integration in many other areas, not only economic but also political (Haas 1958, 1964; Pierson 1996; Sandholtz and Stone Sweet 1998). Thus, while intergovernmentalists believe that European integration is rooted in the pursuit of

[4] The jargon is furthermore complicated by "neo" prefixes and other qualifications. Haas (1958, 1964), the father of the functionalist approach to European integration, called his theory "neo-functionalism," to distinguish it from a previous theory of international cooperation developed by David Mitrany. Moravcsik (1993, 1998) calls his approach "liberal intergovernmentalism" to distinguish it from "realist" theories that also place national states at the center of the analysis but emphasize power and interstate rivalry rather than domestic economic interests. In this article, I only use the simpler terms.

national economic interests, functionalists believe that it is about economic integration as a path towards political integration.

The theory of functionalism was directly inspired by Jean Monnet's strategy to delegate specific functions to supranational institutions in relatively narrow areas, mostly technical and economic, with the expectation that it would lead to more institutional integration in other areas over time. Functionalists believe that moving only some policy functions to the supranational level while leaving other functions at the national level creates pressure for more integration through positive and negative mechanisms. A positive mechanism would work through learning: as politicians and interest groups observe the benefits of integrating a few functions, they will want more. This idea is implicit in the Schuman Declaration, which stated that "Europe will not be made all at once, or according to a single plan. It will be built through concrete achievements." Another mechanism is assumed to work by changing people's preferences: as groups cooperate on specific functions, barriers to communication and interaction will decline, which will bring an "endogenous" convergence of values and norms and a demand for more integration. This rather optimistic outlook was inspired by Deutsch's (1964) influential research on communication theory and political integration.

A darker mechanism through which partial integration could lead to more integration is, paradoxically, by generating problems and crises. Because integration is only partial, important complementary functions are missing at each step. For the functionalists, such incompleteness is not a bug but a feature, because it creates pressure for further integration. Monnet's method was explained by his collaborator George Ball (1994, p. 10):

> There was a well-conceived method in this apparent madness. All of us working with Jean Monnet well understood how irrational it was to carve a limited economic sector out of the jurisdiction of national governments and subject that sector to the sovereign control of supranational institutions. Yet, with his usual perspicacity, Monnet recognized that the very irrationality of this scheme might provide the pressure to achieve exactly what he wanted—the triggering of a chain reaction. The awkwardness and complexity resulting from the singling out of coal and steel would drive member governments to accept the idea of pooling other production as well.

A challenge for this story is to explain why national politicians don't anticipate Monnet's chain reaction. The implicit assumptions here are that integration is irreversible and that national politicians or voters would prefer limited integration to either more integration or no integration. But then, if politicians see that limited integration will lead to more integration, they should either agree to the outcome of more integration right away or they should object to starting the process at all. What factors could allow elites and supranational technocrats to move ahead with initiatives leading to outcomes that national politicians or voters would not have approved in advance? A first possible explanation proposed by functionalists is that national politicians have short horizons: they approve the first step but do not care

about the next steps. A second explanation is asymmetric information. The initial steps of functional integration are taken in narrow and technical areas, such as coal and steel in the 1950s or, later, antitrust regulations and monetary issues. In those matters, national politicians and voters are much less informed than technocrats, political elites, and supranational entrepreneurs. Hence, it is difficult for them to monitor these agents and anticipate the consequences of their actions (Pierson 1996; see Eichengreen 2006 for a discussion). A third, even less-flattering reason why the mechanism may work is that European supranational institutions and bureaucracies have been set up (on purpose?) with little democratic accountability—the so-called "democratic deficit"—reducing the opportunities of national voters to monitor the technocrats, who can therefore move ahead with integration in areas that would not have been approved in advance by the voters (for a discussion, see Alesina and Spolaore 2003, chap. 12).

Functionalism was the dominant theory of European integration in the 1950s and 1960s, then came to seem less plausible (Haas 1975) following a series of political setbacks to integration. A major setback was the "Empty Chair Crisis," when French President Charles de Gaulle boycotted European institutions because he objected to their plans for more supranational integration. The crisis was resolved in a truly "intergovernmentalist" way with the Luxembourg compromise of 1966, in which *de facto* veto power was given to every member state on issues of "very important national interest." However, the functionalist view returned to fashion with the revival of European integration in the 1980s and 1990s when Jacques Delors was head of the European Commission. Functionalism continues to be very influential not only academically but also among European policymakers and supranational civil servants (perhaps not surprisingly, given that they play the main role according to the theory).

In 1992, the members of the European Community signed a Treaty on European Union at Maastricht, which reorganized European institutions and designed an Economic and Monetary Union (EMU)[5] establishing the institutional foundations for the euro. Jacques Delors and his Committee for the Study of Economic and Monetary Union, also known as the "Delors Committee," played a crucial role, as documented in a detailed analysis of the negotiations leading to the economic and monetary union (Dyson and Featherstone 1999). The design and rationale for the European economic and monetary union, as laid out in official documents and studies (Committee for the Study of Economic and Monetary Union, 1989; Commission of the European Communities, 1990), was deeply influenced by the functionalist view of European integration (Sadeh and Verdun 2009, p. 283).

[5] "EMU" is a confusing acronym. It does *not* stand for "European monetary union" (a widespread and understandable confusion). Instead, it means "economic and monetary union," including both the monetary union *and* the single market (for the official definition, http://ec.europa.eu/economy_finance/euro/emu/). In the 1990s, some even used EMU as shorthand for the whole Maastricht agreement, which included several other provisions besides those about economic and monetary union. In contrast, many now use EMU in a narrower sense, referring only to the monetary union. Given such ambiguities, I avoid the acronym EMU and spell out "economic and monetary union" whenever possible in this article.

An important functionalist argument was based on the "inconsistent quartet:" the mutual incompatibility of free trade, mobility of capital, fixed exchange rates, and independence of national monetary policies (Padoa-Schioppa 2004). Assuming that fixed exchange rates were essential for Europe's single market, then moving from commercial integration to liberalization of capital movements had to lead to the loss of national monetary autonomy.[6] In fact, Padoa-Schioppa (2004, p. 14), one of the architects of the economic and monetary union and key member of the Delors Committee, explained the path to the euro in terms that explicitly echoed the chain-reaction metaphor: "[T]he road toward the single currency looks like a chain reaction in which each step resolved a preexisting contradiction and generated a new one that in turn required a further step forward. The steps were the start of the EMS [European Monetary System] (1979), the re-launching of the single market (1985), the decision to accelerate the liberalization of capital movements (1986), the launching of the project of monetary union (1988), the agreement of Maastricht (1992), and the final adoption of the euro (1998)." Also, in the functionalist tradition, each step in this chain reaction was viewed as irreversible. A joke attributed to Padoa-Schioppa refers to how EMU, the economic and monetary union, was like the Australian bird with the same name, in that neither could walk backward.

Not only was the path to the euro explained in functionalist terms from a technical perspective, but it was also viewed, in Schumann and Monnet's tradition, as "a further step—and as a prerequisite for yet other steps—in the political unification of Europe" (Padoa-Schioppa 2004, p. 6). Wim Duisenberg, the first President of the European Central Bank, said (as quoted in Van Overtveldt 2011, p. 63): "EMU is, and was always meant to be a stepping stone on the way to a united Europe." German Chancellor Helmut Kohl famously said in 1991 (as quoted in Marsh 2011, p. 301): "It is absurd to expect in the long run that you can maintain economic and monetary union without political union."

From the perspective of Monnet's method, such an "absurd" economic and monetary union without political union should create pressures for still more integration. The euro area lacked many institutions historically associated with a successful monetary union: for example, a central bank that could really act as market maker and lender of last resort, a banking union, and a fiscal union. But this incompleteness could be rationalized as a natural and unavoidable feature of partial integration in the functionalist tradition. Even though present political constraints prevented the immediate implementation of a more comprehensive design, the launching of an "incomplete" monetary union would set the steps for further integration in due course, as predicted by functionalist theories. For example, people would learn with time about the large benefits from economic and monetary union and ask for more integration in other areas. Also, supporters of the euro embraced two arguments mirroring the long-standing functionalist view that preferences and

[6] Fixed exchange rates were viewed as essential for free trade within Europe, not only for economic reasons but also and perhaps chiefly for political reasons, because of the fear that competitive devaluations could trigger a protectionist policy reaction by other member states, leading to the eventual unraveling of the single market (Eichengreen and Frieden 2001, pp. 12–13).

behavior endogenously converge *following* integration. First, regions will become economically more homogeneous *after* they share a common currency (Frankel and Rose 1998).[7] Secondly, the economic and monetary union was supposed to provide discipline to governments, including those that used to pursue erratic policies. As a result, all member states would eventually converge to common values and policies emphasizing macroeconomic stability. Supranational institutions could provide the necessary sanctions if national governments deviated from agreed rules of stability. No-bailout rules would also be enforced. If, in spite of these positive effects and precautions, future crises were to occur, they could be resolved with more institutional integration.

Assuming that such logic could really work—and events of the last few years certainly sound a skeptical note—where would Jean Monnet's chain reaction lead in the long run? Monnet himself was ambiguous about his long-term vision of European integration (Hoffmann 1966; Duchêne 1994). He oscillated between two visions. One was the original federalist dream of the United States of Europe, in which the current Europe was an "incomplete federation" to be completed. The other vision was of a "post-modern" world where traditional sovereign states, including classic federations like the United States of America, would play a marginal role compared to supranational institutions and norms, which would represent a novel way to organize interdependence among individuals and groups—a vision of Europe as a "post-federation." This same ambiguity is present in the conflicting views about the euro among its supporters: is it a currency without a state *yet*, or is it a currency without a state *ever*?

The Limits to Monnet's Chain Reaction

Since the Schuman Declaration of 1950 which launched the Monnet strategy of partial integration, European institutions have grown from a coal and steel community of six countries to a European Union of 28 countries (as of summer 2013), building along the way a customs union, a single market, an economic and monetary union, and much more. This list of achievements has brought several benefits, to which we will return. Nonetheless, the "functionalist" view, deeply embodied in the ideology and practice of the European Union, that each step is part of a chain reaction leading to ever closer integration has serious limitations.

As a starting point, the functionalist emphasis on the rising power and autonomy of supranational institutions compared to national governments must be taken with a grain of salt. National governments do agree to delegate responsibilities to supranational institutions as commitment devices to achieve collective goals which are in each government's long-term interest. To be credible, those institutions must have some autonomous power and independence; the rules for the autonomy of the European Central Bank come to mind. In addition, as in all complex organizations,

[7] In contrast, the antifunctionalist argument that integration can lead to specialization and *more* heterogeneity (Krugman and Venables 1996) received much less attention in Brussels.

supranational agents cannot be perfectly monitored by their principals (in this case, national governments and voters), and some principal–agent slack always exists. However, none of this means that Europe's supranational institutions can go very far against the ultimate interests of national governments. While supranational institutions and procedures are important in the day-by-day working of Europe, they "could not work for a week in the absence of the will to cooperate of the member states, especially the largest ones—Germany and France above all" (Gilbert 2012, p. 3).

A well-known illustration of how centralized discipline does not work when ultimate power is in the hands of sovereign governments is the spectacular failure of the Stability and Growth Pact, which came into force in the late 1990s and included mechanisms to ensure that member states would hold their annual budget deficits below 3 percent of GDP and their accumulated government debt below 60 percent of GDP. The Stability and Growth Pact was never credible, and became moot after 2003, when France and Germany used their political power to prevent sanctions against their own violation of the pact's fiscal rules. In general, the success of supranational agents' ability to take autonomous decisions can only be sustained in matters where the extent of disagreement among national governments over policy outcomes is relatively low, like the enforcement of trade liberalization agreements. But success in those areas does not imply that supranational institutions and rules could also trump national institutions and rules in other areas with much higher heterogeneity of preferences and interests, like fiscal policies.

The role of the European Court of Justice is instructive. In a series of landmark cases, the Court enunciated the doctrine that European Community norms have direct effect in member states and trump domestic law and that individuals can directly invoke European law before national and European courts. For example, the 1963 case of *Van Gend en Loos v Nederlandse Administratie der Belastingen* (case 26/62) was decided in favor of a Dutch importer of German chemical products that had objected to a tariff charged by the Dutch authorities in violation of article 12 of the Treaty of Rome, which forbids member states from raising customs duties between themselves or introducing new ones. The aggressive interpretation of its role by the European Court of Justice in this and other cases went beyond the legal framework that had been formally agreed with the Treaty of Rome, and, according to some scholars, brought Europe close to a federal system from a legal perspective (Weiler 1991; Krasner 1999), expanding supranational powers beyond the control of national governments (Pierson 1996; Stone Sweet 2000). However, these new legal doctrines were established to enforce norms consistent with national governments' own collective objectives, such as trade liberalization. The acceptance of these decisions by national governments and courts did not imply that *any* European norms would be as easily accepted in the future. In more recent years, the German Constitutional Court has elaborated the legal theory of *conditional* acceptance of European Union norms, according to which Germany only accepts the supremacy of EU law insofar as it is consistent with fundamental German rights. In a famous decision on the constitutionality of the Maastricht Treaty (*BVerfGE 89,155* of October 12, 1993), the German Court ruled that there

is a legal limit to the powers of EU norms, defined by their effects on national democratic sovereignty. In a landmark ruling on the Lisbon Treaty (*BVerfG, 2 BvE 2/08* of June 6, 2009), the German Court explicitly stated that the national states are "the masters of the treaties" and "therefore must see to it that there are no uncontrolled, independent centralization dynamics" within the European Union (quoted in *Spiegel* 2009), a clear and explicit brake on functionalist dynamics.

In general, the central problem with the chain-reaction method is the unwarranted expectation that gradual integration, which has been successful in areas with low costs of heterogeneity, can continue unabated when moving to areas with much higher heterogeneity costs. The source of this problem is the lack of a realistic assessment. Successful integration is more likely to take off in areas such as commercial integration, where heterogeneity costs are relatively low, and partly offset by the benefits from diversity. As integration proceeds to other areas, after low-hanging fruits are picked, heterogeneity costs continue to increase along a convex curve. At some point, those high costs become politically prohibitive, and the pressure from spillovers, inefficiencies, and crises will no longer lead to further integration, but just to losses—and possibly even the collapse of the whole system. The chain-reaction approach does not anticipate that heterogeneity costs and constraints will eventually become binding and stop the process for good. Followers of this approach are therefore prone to setting up incomplete and inefficient arrangements, relying on the overoptimistic expectation that such inefficiencies can always be addressed at a later stage through additional integration.

For these reasons, there is no guarantee that regional integration in economic areas, such as a common market, should lead to political unification down the road. The example of the German customs union (*Zollverein*) in the nineteenth century, often mentioned in this respect, is misleading, because the main force behind commercial integration was political integration pushed by Prussia's military power (Gilpin 2001). In fact, international cooperation and political unification can be viewed as *substitute* ways to lower barriers to trade. If two regions can already agree to reduce their trade barriers with each other while remaining independent, they are going to obtain *smaller* additional gains from trade if they also form a political union with a unified domestic market. This is a direct negative effect of economic integration on the incentives to form a political union. In practice, this negative effect is likely to be larger than possible indirect effects of economic integration in lowering the costs of political integration through communication and coordination, which were stressed by Monnet and his followers.

The historical record up to now indeed suggests that international economic integration is more likely to go hand in hand with political *dis*integration (Alesina and Spolaore 1997, 2003; Alesina, Spolaore, and Wacziarg 2000).

The euro is a child of the functionalist method. The method of partial integration provided the institutional framework and rationale for monetary integration. Without Monnet's idea of delegating specific policy functions and prerogatives to supranational institutions, the euro would not have come into existence. Of course this does not mean that the euro was created exclusively for "functionalist" reasons and goals, nor that the decisionmakers were only supranational civil servants and

elites. There would be no euro without the actions of powerful national politicians pursuing their own geopolitical and domestic objectives or without the backing of powerful economic interests (like German exporters). However, statesmen with political goals and producers with economic interests exist elsewhere but do not end up with a "currency without a state." Such a currency was only possible—politically, technically, and intellectually—in the exceptional institutional framework provided by European integration.

By creating the euro, the chain reaction crossed the border between "pure" economic integration, which can be achieved through international cooperation in the form of liberalization of trade and capital flows, and the form of monetary integration that historically had only been obtained by a sovereign state using its power of coercion to establish one currency within its borders.

The exceptional nature of the euro does not mean that Europe's monetary union is unsustainable in the long run or only sustainable if Europe becomes a sovereign federation. The parameters of the questions have been well stated by Mario Draghi (2012), the head of the European Central Bank, who said that "those who claim only a full federation can be sustainable set the bar too high." Instead, Draghi focuses on the "minimum requirements to complete economic and monetary union." In such framework, the future of the euro depends on a key political variable: the heterogeneity costs associated with the minimum set of functions that must be pooled or delegated for a currency union to work.

If the costs associated with heterogeneity were small, the euro area crisis of the last few years could perhaps be addressed with deep fiscal and political integration. This outcome seems out of reach at present given the historical experience of European integration. In principle, monetary union could lead to a fiscal and political union, even if heterogeneity remains high, if the costs from leaving the euro are even higher and fiscal and political integration are perceived by national governments and voters as the only solution. That outcome could be seen as a vindication of the darker version of Monnet's chain reaction: heterogeneous Europeans would have been "trapped" in a fiscal and political union because they took an irreversible decision to enter a monetary union without anticipating the spillover to further integration. However, political union on such grounds would hardly be a solid start for a European federation; it would be very unlikely that such political union could trigger the positive cultural changes that would be the only sustainable foundation for a cohesive federation in the long run.

A more promising way to ensure the stability of monetary union in Europe is to focus on a narrower set of minimum requirements, as suggested by Draghi (2012). High priority is likely to be given to banking and financial integration. Those gains could in part be secured with tools and institutions similar to those already profitably employed in areas where the European Union has been most successful, like commercial integration, antitrust regulation, and the formation of a single market (for instance, Enderlein et al. 2012). At the same time, though, the close links between banks and sovereign states in Europe can create dangerous spillovers, crises, and clashes between supranational authorities and national governments.

Back Where We Started: The Benefits from European Integration

European integration in the aftermath of World War II was encouraged to counteract the threats of war and protectionism. We saw the timid beginnings with coal and steel, the defeat of the ambitious plan to pool defense, the establishment of a common market, and the building of ever more complex and ambitious supranational institutions—eventually including monetary union. We discussed the serious problems with the expectation that this would necessarily be a chain reaction towards ever closer political integration. Now, it is time to offer a few final comments on the sources of strengths of the European integration project.

While the chain-reaction method can be carried too far, key aspects of Monnet's strategy have in fact contributed to the concrete successes of the European project when the tools of integration have been applied to the appropriate areas: those with lower heterogeneity costs and higher economies of scale. These aspects include: partial integration with a focus on economic areas; deep "institutionalization," with the delegation of substantial prerogatives to supranational institutions going well beyond the institutional framework of more traditional international organization; and integration of several functions, creating useful "linkages."

Trade integration is a good example of the effectiveness of partial integration in economic areas. Trade is an area where costs of heterogeneity are offset by benefits from heterogeneity and large economies of scale. The removal of trade barriers was in the general interest of Europeans even though specific sectors and groups within each country benefited from protectionist policies. European supranational institutions provided a way to coordinate trade liberalization and to lock in the commitment not to raise barriers unilaterally when faced with domestic political pressure. In this respect, European integration was one of the earliest and most successful examples of regional arrangements set up to solve coordination problems and to provide credible commitments (Eichengreen 2006). Partial institutional integration in different areas also allowed "linkages" between issues and provision of credible side-payments to potential losers from commercial integration. For instance, the notoriously wasteful Common Agricultural Policy has been often explained as a political compromise between France and Germany: German manufacturers gained access to the French market, and German taxpayers helped subsidize French farmers.

Trade integration within Europe has also benefited peace and security. The view that international trade can reduce the risk of war goes back to Montesquieu and Kant and has spurred a large empirical literature. Multilateral openness (or "globalization") does not reduce the risk of war between pairs of countries. However, *bilateral* trade, by increasing the opportunity cost of conflicts between two partners, reduces the probability of conflict between that pair of countries (Martin, Mayer, and Thoenig 2008), even when controlling for historical, linguistic, and cultural similarities between populations (Spolaore and Wacziarg 2012). In fact, Martin, Mayer, and Thoenig (2010) find that country pairs with a high frequency of old wars are more likely to sign regional trade agreements, and this can be explained as a consequence of the complementarity between economic and political gains from trade. They also show that multilateral trade openness reduces the opportunity cost

of bilateral conflict, thereby increasing the risk of war between pairs of countries which can trade with third partners. Therefore, globalization also increases the political incentive to sign regional agreements for security reasons.

This interaction between economic and political factors can explain important aspects of European integration. For instance, it can shed light on why Konrad Adenauer, the Chancellor of West Germany from 1949 to 1963, pushed for a geographically narrower but institutionally deeper customs union with France, Germany's old enemy, therefore reducing the risk of war between the two countries. Adenauer overruled his economic minister Ludwig Erhard, who was primarily interested in economic benefits and would have preferred a broader free-trade area, which France would have been unlikely to join given its own commercial and political interests (Garton Ash 1993; Moravcsik 1998).

An open question is whether European integration has played a central or only a marginal role in securing peace in Europe. Skeptics of the "pacifying" effect of European institutions stress the crucial involvement of the United States and NATO in Europe during the Cold War and afterwards, and the failure of Europeans to deal with the breakup of Yugoslavia on their own. Moreover, peace has also held between Japan, the other loser of World War II, and its neighbors, and trade has prospered among them in the absence of Asian institutions analogous to the European Union. However, the Cold War ended in Europe with the fall of the Berlin Wall, Germany is now unified, and European institutions have played a very significant role in the process of democratization and integration of Eastern and Central Europe. In contrast, the relation between a still formally communist China and Taiwan remains tense and unresolved, and Korea is still divided and even at risk of a nuclear war, which could spread to Japan and other neighboring countries. On balance, whether because of European integration or other factors, in recent decades Europe has fared quite well in terms of peace and democracy relative to other areas of the world.

In general, pooling and delegating functions and policies to supranational institutions to take advantage of economies of scale and scope while maintaining other prerogatives at the national (or subnational) level has brought substantial benefits to Europeans when appropriately implemented in areas with relatively low heterogeneity costs. Those benefits have been obtained while keeping ultimate sovereign control and the monopoly of the legitimate use of coercion at the national state level. A centralized European authority could provide a broader range of public goods with large economies of scale and scope while using coercion to prevent free riding. But that would come with much higher heterogeneity costs. Europeans probably have been wise in not yet traveling all the way to a sovereign federation given their existing differences in preferences and cultures. If those preferences change, Europe may benefit from a reorganization in a federal direction, but of course that should be decided only through broad and democratically expressed consensus.

At the moment, Europeans are sticking to the current system of cooperation among sovereign states within a supranational organization. Within those boundaries, and in spite of the serious limitations that we discussed, European institutions have provided useful commitments to overcome some (but not all) problems from free riding and beggar-thy-neighbor policies. European cooperation has at least

turned out better than the alternative system of destructive unilateral national policies plaguing European history until 1945. In this respect, Moravcsik (2012), a leading "intergovernmentalist" scholar of European integration, is hopefully not far from the truth when he writes: "Whatever the outcome of the crisis, the EU will remain without rival the most ambitious and successful example of voluntary international cooperation in world history."

■ *I am grateful to Jeff Frieden, Yannis Ioannides, Deborah Menegotto, Stelios Michalopoulos, Romain Wacziarg, and the editors (David Autor, Chang-Tai Hseih, and Timothy Taylor) for their detailed comments. I also benefited from helpful feedback and conversations with many people, including Lorenzo Bini-Smaghi, Giancarlo Corsetti, Henrik Enderlein, Kai Konrad, Athanasios Orphanides, Lucas Papademos, and Daniela Schwarzer, and participants in the political economy discussion group at Harvard and a conference at the Condorcet Center for Political Economy in Rennes.*

References

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8(2): 155–94.

Alesina, Alberto, and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43(3): 762–800.

Alesina, Alberto, and Roberto Perotti. 2004. "The European Union: A Politically Incorrect View." *Journal of Economic Perspectives* 18(4): 27–48.

Alesina, Alberto, and Bryony Reich. 2013. "Nation Building." NBER Working Paper 18839.

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112(4): 1027–56.

Alesina, Alberto, and Enrico Spolaore. 2003. *The Size of Nations*. Cambridge, MA: MIT Press.

Alesina, Alberto, and Enrico Spolaore. 2005. "War, Peace, and the Size of Countries." *Journal of Public Economics* 89(7): 1333–54.

Alesina, Alberto, and Enrico Spolaore. 2006. "Conflict, Defense Spending, and the Number of Nations." *European Economic Review* 50(1): 91–120.

Alesina, Alberto, Enrico Spolaore, and Romain Wacziarg. 2000. "Economic Integration and Political Disintegration." *American Economic Review* 90(5): 1276–96.

Ball, George W. 1994. "Forward" to *Jean Monnet: The First Statesman of Interdependence,* by François Duchêne, 9–14. New York: Norton.

Berger, Helge, and Albrecht Ritschl. 1995. "Germany and the Political Economy of the Marshall Plan, 1947–52: A Re-revisionist View." In *Europe's Postwar Recovery*, edited by Barry Eichengreen, pp. 199–245. Cambridge University Press.

Bergsten, C. Fred. 2012. "Why the Euro Will Survive: Completing the Continent's Half-Built House." *Foreign Affairs*, September/October.

Bisin, Alberto, and Thierry Verdier. 2010. "The Economics of Cultural Transmission and Socialization." Chapter 9 in *Handbook of Social Economics,* Vol. 1, edited by Jess Benhabib, Alberto Bisin, and Matt Jackson. Elsevier.

Commission of the European Communities. 1990. *One Market, One Money: An Evaluation of the Potential Benefits and Costs of Forming an Economic and Monetary Union*. Study of the Directorate General for Economic and Financial Affairs, October. *European Economy* no. 44.

Committee for the Study of Economic and Monetary Union. 1989. *Report on Economic and Monetary Union in the European Community* [also known as *The Delors Report*]. Luxembourg: European Communities.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Romain Wacziarg. 2012. "The Political Economy

of Linguistic Cleavages." *Journal of Development Economics* 97(2): 322–38.

Desmet, Klaus, Ignacio Ortuño-Ortín, and Shlomo Weber. 2009. "Linguistic Diversity and Redistribution." *Journal of the European Economic Association* 7(6): 1291–1318.

Deutsch, Karl W. 1964. "Communication Theory and Political Integration." In *The Integration of Political Communities*, edited by Philip E. Jacob and James V. Toscano. Philadelphia: J. B. Lippincott.

Draghi, Mario. 2012. "The Future of the Euro: Stability through Change." Published in German in *Die Zeit,* August 29, 2012; English version: http://www.ecb.int/press/key/date/2012/html/sp120829.en.html.

Duchêne, François. 1994. *Jean Monnet: The First Statesman of Interdependence.* New York: Norton.

Dyson, Kenneth, and Kevin Featherstone. 1999. *The Road to Maastricht: Negotiating Economic and Monetary Union*. Oxford University Press.

Eichengreen, Barry. 2006. "European Integration." In *Oxford Handbook of Political Economy*, edited by Barry R. Weingast and Donald A. Wittman, 799–813. Oxford University Press.

Eichengreen, Barry. 2012. "European Monetary Integration with Benefit of Hindsight." *Journal of Common Market Studies* 50(S1): 123–36.

Eichengreen, Barry, and Jeffry A. Frieden. 2001. "The Political Economy of European Monetary Unification: An Analytical Introduction." In *The Political Economy of European Monetary Unification,* edited by Barry Eichengreen and Jeffry Frieden, 1–21. Boulder: Westview Press.

Enderlein, Henrik, Peter Bofinger, Laurence Boone, Paul de Grauwe, Jean-Claude Piris, Jean Pisani-Ferry, Maria Joao Rodrigues, Andrés Sapir, and Antonio Vitorino. 2012. *Completing the Euro—A Roadmap towards Fiscal Union in Europe.* Report of the "Tommaso Padoa-Schioppa Group." Paris: Notre Europe.

Esteban, Joan, Laura Mayoral, and Debraj Ray. 2012. "Ethnicity and Conflict: An Empirical Study." *American Economic Review* 102(4): 1310–42.

Fearon, James D. 2006. "Ethnic Mobilization and Ethnic Violence." In *Oxford Handbook of Political Economy*, edited by Barry R. Weingast and Donald A. Wittman, 852–68. Oxford University Press.

Feldstein, Martin. 2012. "The Failure of the Euro: The Little Currency That Couldn't." *Foreign Affairs*, January/February.

Frankel, Jeffrey A., and Andrew K. Rose. 1998. "The Endogeneity of the Optimum Currency Area Criteria." *Economic Journal* 108(449): 1009–1025.

Frieden, Jeffry A. 1998. "The Euro: Who Wins? Who Loses?" *Foreign Policy*, no 112, pp. 24–40.

Frieden, Jeffry A. 2002. "Real Sources of European Currency Policy: Sectoral Interests and European Monetary Integration." *International Organization* 56(4): 831–60.

Garrett, Geoffrey. 1993. "The Politics of Maastricht." *Economics and Politics* 5(2): 105–23.

Garton Ash, Timothy. 1993. *In Europe's Name: Germany and the Divided Continent.* New York: Random House.

Garton Ash, Timothy. 2012. "The Crisis of Europe." *Foreign Affairs*, September/October.

Gilbert, Mark. 2012. *European Integration: A Concise History.* Lanham: Rowman & Littlefield.

Gilpin, Robert. 2001. *Global Political Economy: Understanding the International Economic Order.* With the assistance of Jean M. Gilpin. Princeton University Press.

Haas, Ernst B. 1958. *The Uniting of Europe: Political, Social, and Economic Forces, 1950–1957.* London: Stevens.

Haas, Ernst B. 1964. *Beyond the Nation State: Functionalism and International Organization*. Stanford: Stanford University Press.

Haas, Ernst B. 1975. *The Obsolence of Regional Integration Theory.* Institute of International Studies, University of California.

Hoffmann, Stanley. 1966. "Obsolete or Obstinate? The Fate of the Nation-State and the Case of Western Europe." *Daedalus* 95(3): 862–915.

Krasner, Stephen D. 1999. *Sovereignty: Organized Hypocrisy.* Princeton University Press.

Krugman, Paul. 2012. "Revenge of the Optimum Currency Area." In *NBER Macroeconomics Annual*, Vol. 27, edited by Daron Acemoglu, Jonathan Parker, and Michael Woodford, 439–48. University of Chicago Press.

Krugman, Paul, and Anthony J. Venables. 1996. "Integration, Specialization, and Adjustment." *European Economic Review* 40(3–5): 959–67.

Lane, Philip R. 2012. "The European Sovereign Debt Crisis." *Journal of Economic Perspectives* 26(3): 49–68.

Marsh, David. 2011. *The Euro: The Battle for the New Global Currency.* New Haven: Yale University Press.

Martin, Philippe, Thierry Mayer, and Mathias Thoenig. 2008. "Make Trade Not War?" *Review of Economic Studies* 75(3): 865–900.

Martin, Philippe, Thierry Mayer, and Mathias Thoenig. 2010. "The Geography of Conflicts and Free Trade Agreements." CEPR Discussion Paper 7740.

Milward, Alan S. 1984. *The Reconstruction of Western Europe, 1945–51.* Methuen.

Montalvo, José G., and Marta Reynal-Querol. 2005. "Ethnic Polarization, Potential Conflict and Civil Wars." *American Economic Review* 95(3): 796–816.

Moravcsik, Andrew. 1993. "Preferences and

Power in the European Community: A Liberal Intergovernmentalist Approach." *Journal of Common Market Studies* 31(4): 473–524.

Moravcsik, Andrew. 1998. *The Choice for Europe: Social Purpose and State Power from Messina to Maastricht*. Ithaca: Cornell University Press.

Moravcsik, Andrew. 2012. "Europe after the Crisis: How to Sustain a Common Currency." *Foreign Affairs*, May/June

Padoa-Schioppa, Tommaso. 2004. *The Euro and Its Central Bank: Getting United After the Union*. Cambridge, MA: MIT Press.

Pierson, Paul. 1996. "The Path to European Integration: a Historical Institutionalist Analysis." *Comparative Political Studies* 29(2): 123–63.

Rector, Chad. 2009. *Federations: The Political Dynamics of Cooperation*. Ithaca: Cornell University Press.

Riker, William H. 1964. *Federalism: Origin, Operation, Significance*. New York: Little, Brown and Company.

Sadeh, Tel, and Amy Verdun. 2009. "Explaining Europe's Monetary Union: A Survey of the Literature." *International Studies Review* 11(2): 277–301.

Sala-i-Martin, Xavier, and Jeffrey Sachs. 1992. "Fiscal Federalism and Optimum Currency Areas: Evidence for Europe from the United States." Chap. 7 in *Establishing a Central Bank: Issues in Europe and Lessons from the U.S.*, edited by Matthew Canzoneri, Vittorio Grilli, and Paul Masson. Cambridge University Press.

Sandholtz, Wayne, and Alec Stone Sweet, eds. 1998. *European Integration and Supranational Governance*. Oxford University Press.

"Schuman Declaration." 1950. Presented by French Foreign Minister Robert Schuman, May 9. Text of declaration available at European Union website: http://europa.eu/about-eu/basic-information/symbols/europe-day/schuman-declaration/index_en.htm.

Spiegel. 2009. "Germany's Lisbon Treaty Ruling: Brussels Put Firmly in the Back Seat." July 6, online version in English: http://www.spiegel.de/international/germany/germany-s-lisbon-treaty-ruling-brussels-put-firmly-in-the-back-seat-a-634506.html.

Spiegel. 2010. "The Price of Unity: Was the Deutsche Mark Sacrificed for Reunification?" September 30, online version in English: http://www.spiegel.de/international/germany/the-price-of-unity-was-the-deutsche-mark-sacrificed-for-reunification-a-719940.html.

Spolaore, Enrico. 2006. "National Borders and the Size of Nations." In *Oxford Handbook of Political Economy*, edited by Barry R. Weingast and Donald A. Wittman, 778–98. Oxford University Press.

Spolaore, Enrico. 2012. "National Borders, Conflict and Peace." In *Oxford Handbook of the Economics of Peace and Conflict*, edited by Michelle R. Garfinkel and Stergios Skaperdas, 763–86. Oxford University Press.

Spolaore, Enrico, and Romain Wacziarg. 2012, "War and Relatedness." http://sites.tufts.edu/enricospolaore/files/2012/08/War-and-Relatedness.pdf.

Spolaore, Enrico, and Romain Wacziarg. 2013. "How Deep Are the Roots of Economic Development?" *Journal of Economic Literature* 51(2): 325–69.

Stone Sweet, Alec. 2000. *Governing with Judges: Constitutional Politics in Europe*. Oxford University Press.

"Treaty of Rome." 1957. March, 25. http://ec.europa.eu/economy_finance/emu_history/documents/treaties/rometreaty2.pdf.

Van Overtveldt, Johan. 2011. *The End of the Euro: The Uneasy Future of the European Union*. Chicago: Agate Publishing.

Weber, Eugen. 1976. *Peasants into Frenchmen: The Modernization of Rural France, 1870–1914*. Stanford University Press.

Weiler, J. H. H. 1991. "The Transformation of Europe." *Yale Law Journal* 100(8): 2403–83.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 145–166

Political Credit Cycles: The Case of the Eurozone

Jesús Fernández-Villaverde, Luis Garicano, and Tano Santos

Before the eurozone came into existence on January 1, 1999, the conventional wisdom was that it would cause its least productive members—Portugal, Spain, and Ireland, and, later, Greece—to modernize their economies. In the past, these peripheral European countries had used devaluations to recover from adverse business cycle shocks, but without correcting the underlying imbalances of their economies. The arrival of the euro was expected to force a sound fiscal policy, eliminate the bias toward inflation, and encourage widespread structural reforms.

For example, Lucas Papademos, who later became Prime Minister of Greece in 2011 and 2012, but who was Governor of the Central Bank back in 2001, stated at a conference to mark Greece's entry to the euro:

> After entry into the euro area, the Bank of Greece will be implementing the single monetary policy decided by the Governing Council of the European Central Bank and it will certainly be impossible to improve the economy's international competitiveness by changing the exchange rate of our new currency, the euro. The objectives of higher employment and output growth will

■ *Jesús Fernández-Villaverde is Professor of Economics, University of Pennsylvania, Philadelphia, Pennsylvania; Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts; and Research Affiliate, Centre for Economic Policy Research, London, United Kingdom. Luis Garicano is Professor of Economics and Strategy, London School of Economics; Research Fellow, Centre for Economic Performance and Research Fellow, Center for Economic Policy Research (CEPR), all in London, United Kingdom. Tano Santos is the David L. and Elsie M. Dodd Professor of Finance, Columbia Business School, New York City, New York. Their email addresses are jesusfv@econ.upenn.edu, l.garicano@lse.ac.uk, and js1786@columbia.edu.*

http://dx.doi.org/10.1257/jep.27.3.145 doi=10.1257/jep.27.3.145

> therefore have to be pursued through structural reforms and fiscal measures aimed at enhancing international competitiveness by increasing productivity, improving the quality of Greek goods and services and securing price stability (Papademos 2001, p. xxxvii).

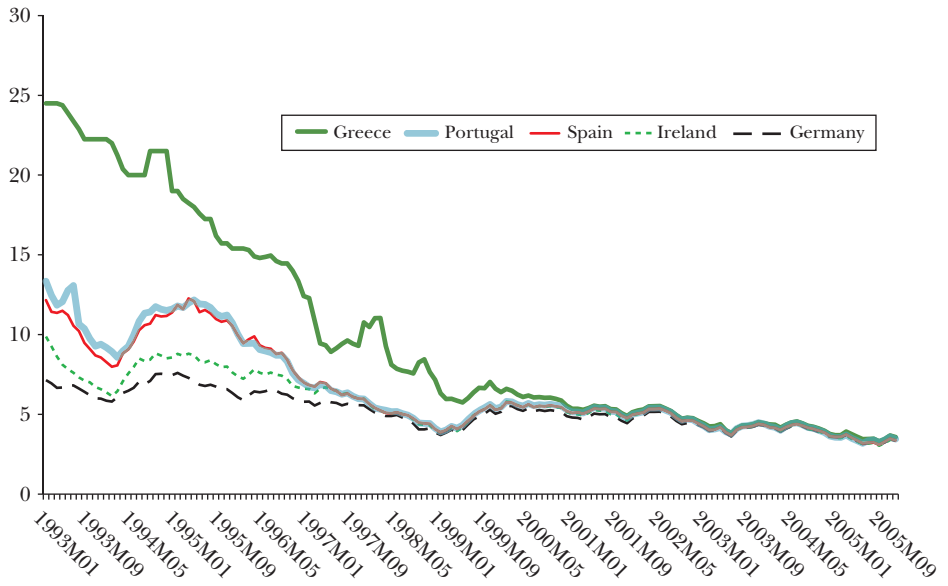
A number of academic authors made similar predictions. For instance, Bentolila and Saint-Paul (2000) wrote: "Indeed the conventional wisdom is that EMU [economic and monetary union] will eventually remove some barriers to reform." Bean (1998) argued that, once monetary and fiscal policies were out of the hands of governments, they would have no alternative but to carry out reforms.

The elimination of exchange rate risk, an accommodative monetary policy, and the worldwide ease in financial conditions resulted in a large drop in interest rates and a rush of financing into the peripheral countries, which traditionally had been deprived of capital. Figure 1 shows the convergence in interest rates, which resulted in much lower interest rates for Ireland, Greece, Spain, and Portugal—indeed, they became able to borrow at German-level interest rates. This paper argues that, as the euro facilitated large flows of capital and a financial bubble in peripheral countries, economic reforms were abandoned, institutions deteriorated, the response to the credit bubble was delayed, and the growth prospects of these countries declined.[1]

In the next section, we explore the two main channels through which these large inflows of capital led to the abandonment of economic reforms. First, these capital inflows relaxed the economic constraints under which agents were acting, thus reducing the pressure for reforms. Second, they made it harder for principals to extract signals about who was performing well or poorly. When all banks are delivering great profits, all managers look competent; when all countries are delivering the public goods demanded by voters, all governments look efficient. As a result, bad agents are not fired, incompetent managers keep their jobs, and inefficient governments are reelected. The efforts to reform key institutions that burden long-run growth, such as rigid labor markets, monopolized product markets, failed educational systems, or hugely distortionary tax systems plagued by tax evasion, were abandoned or even reversed. It is often argued that the inflow of capital to the peripheral countries led to a number of difficulties, such as a debt overhang from excessive borrowing. But in our view, the reform reversal and institutional deterioration suffered by these countries are likely to have the largest negative consequences for growth.

[1] Although there are alternative explanations for the euro crisis, the view that the credit bubble itself is the source of the disturbance is hard to counter. Lane and McQuade (2012) report a strong correlation between net debt flows and domestic credit: the ability of banks to raise external finance was crucial in allowing lending to increase faster than deposits, helping to finance construction booms and public debt. Similarly, Lane (2012) documents how the nontraded sector expanded strongly in the deficit countries, such as Greece, Spain, and Ireland, while it contracted in surplus countries, such as Germany. Our reading of the evidence is thus that the causality mainly runs from the credit bubble to the real changes and not in the opposite direction.

Figure 1
Convergence in Yield for Government Bonds
(10 year yields in percent; monthly from 1993M01 to 2005M12)



Source: Eurostat.

We discuss how these dynamics played out in distinctive ways in five specific countries. The first four countries—Spain, Ireland, Greece, and Portugal—are the four countries with some bailout programs from the European Union as of early 2013. In these countries, instead of the euro leading to a modernization of peripheral Europe, it became the sedative against any reform. While we believe that similar dynamics operate in other countries, such as Italy and France, we will not explore those examples here. By way of contrast, we then turn to a discussion of Germany, which did not enjoy a loosening of its financing conditions as a result of the euro. Faced with a limited margin of maneuver and a stagnant economy, Germany chose the path of structural reforms, and as a result, the underlying divergence in economic policies and institutions between Germany and the other four countries discussed here increased, rather than diminished, as a result of the dynamics induced by the euro.

The Political Economy of Reforms, Institutions, and Monetary Unions

The euro project had four goals (see James, 2012, for a historical narrative): 1) to build a unified European identity; 2) to eliminate nominal exchange rate fluctuations and the imbalances that those could create (and in particular, to channel the export dynamism that Germany had displayed since the 1960s); 3) to create a

monetary authority isolated from political pressures; and 4) to broaden support for structural, supply-side reforms to improve Europe's growth rate. This fourth goal is the main focus of this paper. The main channel through which a monetary union was thought to affect the political economy of reform was by imposing additional constraints on monetary and fiscal policy. In fact, the steep drop in interest rates in the peripheral countries allowed by the euro meant that the budget constraints that these countries faced were loosened, rather than tightened. Moreover, the resulting financial bubble fueled the deterioration of governance and of the institutional arrangements on the euro periphery. Because of this deterioration, the euro may have led to a persistently negative impact on those peripheral countries.

How does an irrevocably fixed exchange rate regime affect the political economy of reform? How does financial integration, and the ensuing credit boom, alter this logic? And how persistent are these effects?

The Arguments 15 Years Ago: Reforms under Fixed Exchange Rates

The debate on the euro typically focused on how the new currency would affect trade, macroeconomic performance, and international finance, but it largely ignored the political economy channel, which in retrospect proved to be crucial. The pre-euro literature that touched on these issues (summarized in Bean 1998) presented two political economy reasons why the euro would facilitate structural reforms. First, governments with less ability to use demand-side policies to lower unemployment would have no choice but to use structural reforms as a substitute. Second, the euro would increase the market discipline on government borrowing because investors would be able to compare investment opportunities across countries without concern for exchange rate risk. The 1989 Delors Report that informed the creation of the euro expected this market discipline to be even more formidable than the formal constraints of the Maastricht Treaty.

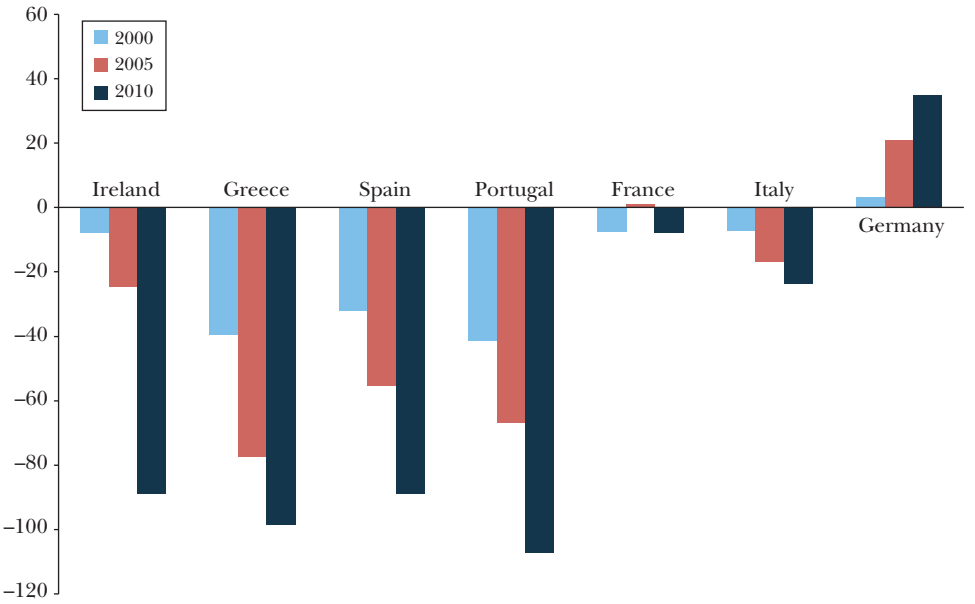
Some researchers did worry that the opposite effect might occur. The absence of an accommodating monetary and fiscal policy would mean that structural reforms would have to be undertaken "without anesthesia," increasing the pain that must be endured by losers and making it less likely that the reforms could be implemented. Chari and Kehoe (2008) also pointed out the danger of "free-riding" in a monetary union. Given that the effects of labor market policies, bank supervision, or fiscal policy of an individual country could negatively affect the welfare of the entire union, the monetary authority could be forced, by the uncoordinated action of its members acting individually, to generate high inflation.

However, during its first years, the euro played a very different role than the ones predicted by much of the literature: instead of tightening government budget constraints, it loosened them—and, thus, gave national governments a tool to avoid painful reforms.

Booms, Reforms, and Information Extraction: Selection and Incentives

The euro caused a gigantic credit inflow to the peripheral countries. As Figure 2 shows, while Ireland, Greece, Spain, and Portugal all started the millennium with

Figure 2
External Indebtedness
(net international investment position as a percentage of GDP)



Source: Eurostat.

sustainable external debt positions, by 2010 all four countries had reached net external debt (the value of the domestic assets owned by foreigners less the value of the assets that nationals owns abroad) close to 100 percent of GDP, through the accumulation of either public (Greece and Portugal) or private (Spain and Ireland) debt. These unprecedented financial booms allowed these countries to expand their public budgets, paying for this either directly through historically cheap debt issuance, as in Greece or Portugal, or through the tax revenue related to the real estate bubble, as in Spain and Ireland.

The consequences for economic reform of such a windfall would not have surprised researchers studying foreign aid. Alesina and Drazen (1991) have argued that the political decision process for economic reform is a war of attrition in which all groups try to delay the reform (with a cost to all) until one group has no more "budget" and gives up, bearing the largest cost. Casella and Eichengreen (1996) show that, in this context, foreign aid will delay concessions and reforms. Svensson (1999), in a game-theoretic model, shows that any windfall (including aid) increases rent-seeking and reduces productive public spending, and he presents empirical evidence (see also Drazen 2000) consistent with the proposition that aid delays reforms. Vamvakidis (2007) extends these arguments to the case of financial booms: he uses a panel of 81 developing and emerging countries to show that increases in external debt are correlated with slowdowns in economic reforms.

We propose here a second channel linking a financial boom to the political economy of reform, unrelated to the "tightness" of the budget constraints: through its effect on the ability of principals to extract performance information on both financial institutions and governmental agencies.[2] It is hard to obtain good signals of performance in a bubble. As Warren Buffet famously put it, "You never know who's swimming naked until the tide goes out." During a bubble, accountability is lost. A manager of a savings bank (a *caja*) in Spain, or of a Greek pension fund, can make bad decisions without negative short-run consequences, because rising asset prices hide their mistakes. Managers and politicians understand that, thanks to the bubble, they can extract more rents without fear of punishment. Consequently, governance deteriorates and weak institutions become weaker.

To understand this mechanism, one can view the quality of governance as a stock of intangible capital. Voters, shareholders, lenders, and other interested parties invest in this stock of governance capital when they imperfectly observe a sequence of actions by the agent and a sequence of outcomes and then infer how good the agent is. Without such investment, the stock of governance depreciates. Governance can be thought of as a stock because it has a persistent effect. For example, bad decisions lead to more bad decisions: naming someone to a bank board without background in banking but who is politically well-connected leads to persistent low governance as he stuffs the board with like-minded individuals more interested in repaying the favor than in monitoring the bank's financial statements.

In this context, a financial boom makes signal extraction harder because all observed outcomes are positive. This increased difficulty of signal extraction has negative consequences for selection as bad agents are not fired. When a crisis hits and there is an acute need for quality leadership, it is less likely to be available. Negative consequences for incentives also arise. When there is a lower probability of underperformance being detected, agents exert less effort.

There are three additional factors that will amplify these mechanisms. First, when downside risk is perceived as being capped by quasi–sovereign-guarantees by the other member states of the monetary union, both on states (Greece and Portugal) and on banks or savings institutions (Spain and Ireland), voters, shareholders, and investors worry less about losses and decrease their investment in monitoring. Second, during the boom times, agents have considerable discretion over the timing of payoffs and can choose to generate large positive payoffs up front and postpone the negative ones. For instance, bank managers can issue highly risky loans that deliver high yields in the short run and that will only become nonperforming years later. Politicians can implement popular spending programs that, while initially cheap, have costs that will escalate over time. Third, signal extraction may be harder when economic activity is concentrated in real estate or finance, rather than in manufacturing, because measuring the fundamentals of output and productivity is harder in both of these fields. Finally, behavioral biases

[2] In a working paper, we present a fuller description of the basic equations of such a model and of the underlying mechanisms (Fernández-Villaverde, Garicano, and Santos 2013).

also contribute to the difficulty in providing appropriate incentives during booms. Agents observe positive outcomes and become overconfident about their abilities and more likely to overreach.

In short, booms cause the quality of the available signals of performance to deteriorate for a number of reasons—statistical, strategic, and behavioral—and, as a result, governance deteriorates as well. Our next step is to argue how this deterioration makes the effects of a negative macroeconomic shock more persistent.

Persistence of the Effects of Bubbles on Governance and Performance: Causes and Channels

It is not immediately obvious why a country's ability to borrow at low nominal interest rates would lead to persistently lower growth rates. But here are four possible reasons.

First, governments that can borrow freely are more likely to waste resources on investments such as airports in the middle of nowhere. Some classic examples are the grandiose investments of oil producers in the 1970s (Gelb 1988). These unproductive expenditures create persistently lower growth since they involve multiyear commitments that must be funded through future distortionary taxation.

Second, countries with easy access to capital suffer a variant of the "Dutch disease" (Sachs and Warner 1995). The credit bubble induces relative price changes that shift the allocation of physical and human capital toward activities such as construction investment and away from the production of tradable goods. While some of the physical inputs can be moved back to the tradable goods sector after the bubble explodes, others are sector-specific and have little scrap value. Human capital investments (or the lack thereof) are sticky, too. In the European periphery, a large part of the work force is simply too poorly prepared to function in a knowledge economy.

Third, the literature on financial frictions has argued that the recovery from financial crisis is inherently slow because agents suffer from a debt overhang in which all sectors of the economy need to deleverage (Bernanke, Gertler, and Gilchrist 1999; Reinhart and Rogoff 2009).

Fourth, a bubble can lead, as we argued above, to a deterioration in economic policy and institutions so that the political economy of finance booms itself becomes a drag on recovery. Debt buys time and can be used to postpone reforms. Low-quality agents are in place throughout the economy, and they will employ every trick imaginable to stay in place (after all, for some of them, staying in power may be the only realistic alternative to a prison term). When politicians dismantle the human capital of a central bank to better make it a servant of their own interests, it takes years to rebuild the institution. Similarly, bad management at the top of a firm damages the quality of middle management.

Moreover, weakened institutions affect the political-economic equilibrium by strengthening the forces against reform and providing few rewards for those in favor of reform. Normally, a group's political success reflects its economic success: if a group grows, its lobby power will be larger and it will be able to push for institutions

favorable to its interests (North, 1990). In a real estate bubble, money flows into the coffers of developers and builders, allowing them to increase their political power. At the same time, the agents in the tradable goods sector have less income and employ fewer workers, reducing their political influence. That is, the bubble creates its own constituency that is only interested in the bubble continuing. And even after the bubble has burst, the constituency is reluctant to accept the required changes in policy.

In the next few sections, we apply this framework to the experiences of Spain, Portugal, Ireland, and Greece and discuss how the relaxation of the credit constraints delayed either the transition to a different path for economic growth or the adoption of reforms. We then argue that the same circumstances that allowed delays in the periphery actually forced reform on a reluctant Germany.

Spain: The Infernal Triangle of Local Governments, Developers, and *Cajas*

The years before the euro were auspicious ones for reform efforts in Spain. Its fiscal position was consolidated, a wave of privatizations created strong multinationals such as Teléfonica, and global, competitive companies such as Inditex (Zara) and Iberdrola emerged. The financial system was strong and well capitalized.

But Spain's real estate bubble ended the reform impulse. Between 1999 and 2007, Spain experienced a period of rapid growth, averaging an annual rate of 3.6 percent. This growth was fostered by the adoption of the euro: real interest rates dropped by 10 percentage points between 1990 and 2005. During this expansion, the grave problems in the Spanish labor market, education system, and institutional design went untouched or worsened. Meager attempts at reforming the sclerotic labor market in 2002 were abandoned, the educational system suffered an increase in the dropout rate, and local governments were infected by the pervasive corruption engendered by the real estate boom (Juan 2011).

The drop in interest rates had a distinctive effect in Spain because Spaniards have historically held a large share of their wealth in real estate: 83 percent of households live in dwellings they own and 80 percent of Spaniards' wealth is invested in real estate, a significantly larger share than in other countries (Bover 2011). Moreover, most mortgages have variable interest rates. Finally, Spain received large immigrant inflows: foreign-born residents went from 2 percent of the Spanish population to 12 percent between 1999 and 2009 (González and Ortega 2009).

During the years of the economic boom, observers noted some disturbing underlying patterns. First, total factor productivity was stagnant: all of Spain's growth between 1995 and 2007 was due to using more labor and capital. Second, external imbalances were rising. Exports of goods and services grew at an annual rate of 8.5 percent from 1995 to 2008, but imports grew at an annual rate of 10.1 percent. The consequent large current account deficits meant that, from 2000 to 2009, Spain required 520 billion euros of external financing (in undiscounted terms). The end

result was a severe deterioration of Spain's net international investment position. Third, a real estate boom was apparent: at the peak, 25 percent of all male Spanish workers were employed in construction (Bonhomme and Hospido 2012). Fourth, the real estate boom and bust led to a similar pattern in government revenues. From 1998 to 2007, government revenues from income and value-added taxes rose by about 140 percent; then from 2007 to 2009, they fell by about 25 percent and have remained low since. A large share of this change was due to the spike in housing transactions (Fernández-Villaverde and Rubio-Ramírez 2009). During the revenue boom, Spain's government committed itself to expenditure programs that, given the cyclical nature of the revenue, were not sustainable. Observers and policymakers were aware of these patterns but no credible measures to correct this situation were undertaken.[3]

While some factors like the drop in real interest rates, favorable demographics, and the surge of immigration contributed to the start of the real estate boom in Spain, political economy factors added fuel to the fire with a self-reinforcing triangle of regional governments, developers, and a type of savings banks called *cajas*.

A key aspect of the transition from dictatorship to democracy in Spain was an ambitious decentralization process with the creation of 17 autonomous regions. In 1997, the Spanish Constitutional Court gave regions nearly complete control over zoning, which had before relied on urban development plans approved decades in advance. The new rules adopted in most regions let private developers or landowners present to the city council detailed plans to build a whole area of a township. The township would usually receive payment in terms of lots or cash. Eminent domain clauses could be used to force land owners to sell to the developer at some "fair price." If the city council approved the plan, the developers and landowners were not bound by any previous zoning restriction. With these changes in place, an entrepreneur could make millions of euros developing areas that had never been on the market before, with the approval of only a city council and the signature of a mayor. Widespread corruption followed. Moreover, since the city would also receive "legal" payments from the developer in cash or in lots, land development became an important revenue source for local authorities, which could use the new-found riches to finance public programs.

For well-connected individuals, this path to unprecedented wealth only needed someone to finance the whole operation. The Spanish financial system was divided into two more-or-less equal parts between the nonprofit *cajas* sector and the for-profit bank segment. The *cajas* were originally created to provide

[3] For example, in 2003, Miguel Angel Fernández Ordóñez, the Governor of the Bank of Spain from 2006 to 2012, warned of increasing debt levels and of the lack of productivity growth ("El legado de Rato," *El País*, September 11th 2003). Miguel Sebastián, an economist who became the main economic advisor of Prime Minister Rodríguez Zapatero, wrote numerous op-eds warning of the imbalances building in the Spanish economy. The banking supervision staff of the Bank of Spain took the unprecedented step in 2006 of writing a memo to the minister of finance denouncing the complacency of the then Governor of the Bank of Spain, Jaime Caruana, in regard to the imbalances building in the Spanish financial sector.

local banking services to the middle-class and working population often ignored by traditional banks, and traditionally had had a strong territorial basis and a conservative outlook.

Two key aspects of the regulation of *cajas* changed with the arrival of democracy. First, the control of the *cajas* was transferred to the regions in 1985, opening the door to their capture by local politicians. Second, the *cajas* were allowed to expand territorially outside their original small area of activity. As a result, the *cajas* engaged in a relentless geographic diversification and the number of branches skyrocketed. By January 1, 2008, Spain had almost 25,000 *caja* branches, one for every 1,800 inhabitants. Not surprisingly, over this period the *cajas* were continuously gaining market share versus banks: back in the 1960s, Spain's banks had about 90 percent of the banking market, while the *cajas* had 10 percent; by around 2004, the *cajas* had a larger market share in Spain than the banks. The 1985 law did not clarify the procedure to be followed for the recapitalization of an insolvent *caja*, which turned out to be a fatal flaw when the crisis came and there was considerable uncertainty regarding this matter.

The *cajas* started channeling lending in an indiscriminate manner to real estate developers. Between 1995 and 2005, lending for construction and development went from 8 percent to 29 percent of GDP, and lending to households for housing purchases grew from 17 percent of GDP to 49 percent (Beltrán et al. 2010). This lending boom was accompanied by a boom in construction. The number of housing units built every year went steadily upward from 150,000 in 1995 to 600,000 in 2007. Prices also increased quickly: according to data from the Spanish Ministry of Housing, between 1998 and the peak of the boom in 2008, nominal housing prices increased by 175 percent, compared to a 61.5 percent increase in Spain's consumer price index.

Because the growth in deposits was not enough to cope with the lending boom, the *cajas* resorted to wholesale funding on an unprecedented scale. Because the loans were euro-denominated and against physical collateral (real estate assets), international institutions were able and willing to lend. The growth of this sector was not accompanied by improvements in *cajas*' governance. *Cajas* did not have shareholders: instead, they were governed by a board selected by the regional and local governments, employees, and clients. These boards were the perfect target for takeovers by low-human-capital managers with the right political alliances and who could finance politically motivated projects. Cuñat and Garicano (2009, 2010) show that the human capital of managers in the *cajas* was low and that those *cajas* where human capital was particularly low had the highest amounts of real estate lending and nonperforming loans.

On May 9, 2012, the collapse of Bankia—a banking giant that was the product of the hurried-up merger of Caja Madrid, Bancaja, and several other smaller *cajas* with assets equal to 33 percent of Spanish output—led directly to Spain's request for a bailout from the European Union. The way in which Bankia came to an ignominious end is a vivid story of institutional deterioration triggered by the real estate boom.

Caja Madrid was one of the oldest *cajas*. For the first decade after the passage of the 1985 law discussed earlier, Caja Madrid was run with the consensus of the main political parties of the high-income region of Madrid. Its head, Jaime Terceiro, managed the entity professionally and made Caja Madrid a fierce competitor in the credit market. However, in 1996, the conservative party pooled its votes with those of a trade union and took control of Caja Madrid. Terceiro was replaced by Miguel Blesa, a close friend of the newly elected prime minister.[4] Starting in 1996, Caja Madrid expanded aggressively, not just in real estate but also in strategic segments of corporate Spain. Eventually, as this complex web of politics, finance, and business interests thickened, Blesa was forced to step down by the head of the regional government of Madrid, who unsuccessfully nominated a close political ally with no experience in banking as the head of Caja Madrid. After some infighting, another powerful politician and former IMF managing director, Rodrigo Rato, was appointed. He ran the entity until its nationalization with a board composed entirely of politically connected appointees.

The other half of what was to become Bankia was Bancaja, the main *caja* in the region of Valencia. It was born in 1878 and had stayed local for most of its history. Only around 1997, with the real estate bubble in its incipient stage, did it start a breakneck expansion after the Valencia regional government modified the law in a way that essentially handed control of the *caja* to the local government. In addition this law vested supervisory authority in a local entity, the "Valencian Institute of Finance," which was an arm of the regional government without any supervisory capability at the time. Few anecdotes could illustrate the unhealthy connection between politics and finance better than the fact that the person appointed as president of Bancaja when the real estate bubble got going in earnest was José Luis Olivas—the same politician who, as Valencia's finance minister, drafted the l997 law regulating the local *cajas* (and who, in the meantime, had also been president of the regional government). Olivas had no experience whatsoever in banking. Bancaja soon became an instrument of the region's political aims in several areas such as housing, energy, telecommunications, and entertainment. Over the next decade, Bancaja would participate in financing all of the major infrastructure projects of the Valencia government, including the Formula 1 in Valencia (at a cost of €244 million), the Castellón Airport (€200 million, although a plane has yet to land there), and Terra Mitica (€300m, an amusement park that entered bankruptcy in 2004). The bursting of the real estate bubble has brought to light numerous corruption scandals in this otherwise wealthy region of Spain.

When problems started in 2009, Caja Madrid and Bancaja were merged into a large systemic institution, Bankia, dominated by the same political interests that had

[4] The agreement between the conservatives (Popular Party or PP) and the trade union (Comisiones Obreras or CC.OO.) was published and openly discussed in the Spanish press. For example, see "El PP modificará la Ley de Cajas de Madrid para cumplir el compromiso con CC.OO." or "CC.OO. y el PP rubrican el acuerdo para que Blesa presida Cajamadrid," *El País*, September 7, 1996.

been running both entities. Two bad *cajas* do not make a good bank, and Bankia was effectively nationalized in the spring of 2012.

Ireland: A Procyclical Fiscal and Regulatory Policy

After a deep recession and huge budget deficits in the late 1970s and early 1980s, Ireland introduced important economic policy reforms in the second half of the 1980s. Also, a consensus emerged among the political parties for reducing budget deficits and tax rates. Reforms in labor market institutions—combined with persistent high unemployment—kept real wage growth below that of Ireland's major trading partners. Strategic sectors of the economy were liberalized, such as air transport (Barrett 1997) and the telecommunication system (Burnham 2003), which at the time was reputed to be the worst in Western Europe. These reforms helped to deliver real annual output growth that averaged more than 6 percent from 1987 to 2000.

However, Ireland's growth relied mostly on an increase in hours worked, while productivity was growing at a rate similar to that of other European countries.[5] In 1989, Ireland had the lowest employment/population ratio in the OECD at 31 percent (Whelan 2010) due to high unemployment and a late baby boom. But by 2000, additional labor as a source of growth was essentially exhausted.

However, real interest rates dropped throughout the 1990s, from roughly 4 percent in the mid 1990s to negative values from 1998 to 2002—that is, for the early years of the euro's existence. Not surprisingly, this led to an increase in valuations and a higher private investment in housing. In the 1990s, Ireland combined a high incidence of owner occupation with the smallest number of dwellings relative to its population in the European Union (Somervile 2007). Thus, the "Celtic Tiger" years started with an abnormally low stock of housing. Construction accelerated, with house completions going from 19,000 in 1990 to 50,000 in 2000 and to 93,000 in 2006. The ratio of house prices to disposable income remained stable until the second half of the 1990s, when it started a growth spurt that would take the ratio from 7 to 12 in less than a decade (Whelan 2010, figure 8). Soon, Ireland was the country with the highest share of housing investment in gross capital formation of any country in the eurozone, and construction became the dominant sector driving growth and employment. By 2007, 13.3 percent of all employment was in the construction sector; for comparison, in the United Kingdom and the United States, that same number never rose above 8 percent. Thus, instead of transitioning from growth based on increased employment to growth based on productivity gains, Ireland instead embarked on a massive speculative cycle in the construction sector.

[5] The measure of productivity growth referred to here controls for the effect of multinationals that book a large fraction of their international profits in Ireland to benefit from low taxation (Honohan and Walsh 2002, figure 13).

In the meantime, rather than seeking to counterbalance the bubble, governmental policy in Ireland accentuated it through a procyclical fiscal policy and regulatory and tax changes that made real estate development even more attractive. We interpret these changes as evidence of a deterioration in institutions and governance.

First, fiscal policy was markedly procyclical: government expenditures in Ireland doubled in real terms between 1995 and 2007. The income tax was cut several times until Ireland reached a stunning income tax and employee contribution average rate of 6.7 percent of gross wage earnings for a single-earning married couple with two children. Ireland already had highly generous tax provisions for owner-occupied housing, being the only OECD country combining a tax deduction for mortgage interest payments with no property tax, capital gains tax, or imputed rent tax (Rae and van de Noord 2006, p. 8). Even so, tax incentives for the real estate sector increased (Honohan 2010). Stamp duties (a sale tax on homes) were lowered in 2001, 2002, 2003, 2005, and 2007, while the ceiling on income tax deductibility of mortgage interest was increased in 2000, 2003, and 2008. Tax concessions were granted for urban renewal, multistory car parks, student accommodations, nursing homes, hotels, and holiday camps. A special incentive tax rate for developers between 2000 and 2007 sought to free up land for development by taxing the proceeds at 20 percent rather than at the higher 42 percent that had prevailed before (Byrne 2012).

Second, several major legislative changes worsened the quality of financial supervision. The 2003 act that established the Central Bank and Financial Services Authority of Ireland (CBFSAI) divided supervisory responsibilities between the newly created Irish Financial Services Regulatory Authority (IFSRA) and the Central Bank of Ireland. This reorganization contributed to the lax banking supervision that characterized this period and which forced the (re)establishment of a single fully integrated regulatory institution in June 2009. Bertie Ahern, Ireland's former prime minister, pointed to this regulatory overhaul as the main culprit in the crisis (Brown 2009).

This new regulatory framework perniciously interacted with a particular development in the Irish banking sector: the emergence of Anglo Irish Bank. In 1999, total assets of Anglo were less than 10 percent of Ireland's GDP; by 2007, total assets were 55 percent of Ireland's GDP, transforming Anglo into a systemic risk for Ireland. This phenomenal expansion was rooted in a business model that emphasized speed in loan approval and a disregard of applicable bank rules. A customer could apply to Anglo for a loan of several million euros for a property development project on a Monday and receive approval by the end of the week (Carswell 2011). Anglo raised funds in international wholesale markets and loaned heavily to a small number of borrowers in the property development sector. Other Irish banks reacted to Anglo by loosening standards. The problems at Anglo Irish were in plain sight for the regulators. However, as documented in the Nyberg Report (Nyberg 2011), neither the management of Anglo Irish nor its Board could recall a meaningful engagement with the regulators on prudential issues. As Whelan (2010)

emphasizes, the collapse of the Irish banking system was not related to financial innovations nor to regulatory arbitrage, but to a failure of supervisory oversight over credit concentration risk and fragile funding. It was, simply, low-quality governance (for accounts of the rise and fall of Anglo Irish Bank, see Carswell, 2011; Lyons and Carey, 2011; Nyberg, 2011).

Why the tolerant government policy toward the boom? Given that underlying problems had been diagnosed by international organizations such as the IMF and the OECD and by Irish economists (Honohan and Walsh 2002), why did the government add more fuel to the fire? The Irish political class, confronted with an unpleasant growth slowdown, preferred to delay any corrective actions. From 1997 to 2007, 35 percent of disclosed donations to Fianna Fail—the party in government—were from property developers and the construction industry (Byrne 2012). Adding hotels (9 percent) and banks and insurance companies (5 percent) shows that 49 percent of disclosed donations were from parties that had a direct interest in the real estate bubble. The coalition of interest groups and an electorate demanding easier access to housing was too powerful to resist.

The extraordinarily close relationship between Ireland's bankers, developers, and government was at the heart of the unprecedented decision to provide a blanket government guarantee of all Irish bank debt on September 30, 2008. This policy included *all* existing and new debt and all deposits, including corporate and even interbank deposits, covered bonds, senior debt, and some subordinated debt. Just why and how this decision was made remains shrouded in mystery (Honohan 2010, chap. 8).[6] While this decision was perhaps just a mistake by an exhausted cabinet taking a hard decision at the worst moment, it is true that as Byrne (2012, p. 202) points out, the "[k]ey political decisions were insulated from critical debate because they were executed within a closed and cartelized system which facilitated regulatory capture." In Ireland, the political economy factors helped to foster a policy in which the government bailed out private creditors from their mistakes during the boom and subsequent crisis, with enormous consequences for the welfare of future Irish taxpayers.

Greece: Sustaining the Unsustainable

In the decade after accession to the euro, Greece enjoyed growth rates over 2 percent in every year from 2000 to 2008, peaking at almost 6 percent in the pre-Olympic year of 2003. This growth was based on financial liberalization coupled with membership in the monetary union, strong export growth, and the fiscal stimulus associated with the Olympic Games. Mitsopoulos and Pelagidis (2012) add to these

[6] The rating companies did not see it as an error. For instance, Fitch affirmed the AAA rating on Ireland following the guarantee decision, stating: "This proactive measure should help buttress confidence in the Irish financial system and limit the risks of a deeper and more-prolonged-than necessary recession at a time of unusual stress in global banking markets" (Bloomberg 2008).

factors improvements in the regulation of some segments of the product markets, such as telecommunications (transportation and energy remained noncompetitive).

Yet the imbalances building in the Greek economy were there for all to see. The current account deficit, already at almost 8 percent in 2000, reached 15 percent in 2008 and was still 9.8 percent in 2011. As a result, net external debt rose from 42.7 percent of output in 2000 to 82.5 percent in 2009. This current account deficit was not, as in Ireland and Spain, the counterpart of large inflows of money into the private sector. The entirety of Greece's net external debt is accounted for by the public sector, which by 2009 had debt that exceeded GDP.

Greece's unsustainable situation had been developing since 1980, with yearly average government deficits over 8 percent of output in the 1980s and 1990s and current account deficits of over 10 percent in both the 1990s and the 2000s. In 2004, the Greek electorate gave a strong mandate to the New Democracy party (after 11 years of PASOK rule) to tackle the many problems afflicting the Greek economy, but the results were disappointing at best. In this sense, Greece is the poster child for postponed adjustment. Greece's curse, more than any of the other peripheral countries, was the problem of an unreformed economy. A report on Greek governance in June 2012 by the OECD, an organization that (like other international organizations) tends to pull its punches in public, wrote: "The combination of these factors—a weak Centre of Government, legal formalism, the absence of basic data, the lack of evidence-based policy making and an undeveloped HR [human resources] strategy—has created an environment conducive to rent seeking" (OECD 2011). Mitsopoulos and Pelagidis (2012, p. 131) argue that Greece is a country with almost first-class per capita output but second-class governance, institutions, business environment, and corruption. The evidence of institutional deterioration in Greece is widespread, from the decreasing reliability of government statistics to a drop in the corruption ranking from Transparency International from position number 35 in 2000 to number 78 by 2010.

How can a country in the heart of the European Union, under pressure from a common currency, avoid the most basic reforms? For no country was the arrival of the euro as large a boon as for Greece. In 1994, the interest on the 10-year bond had reached almost 22 percent. By June 2003, the combination of the global lending boom and the perceived disappearance of currency and default risk meant that Greece was paying an unprecedented 3.6 percent.

Although the examples of arrested reforms that followed are many, one of the clearest is the pension system. Greece's pension system was, by 2009, designed to replace 95.7 percent of the final income level, the highest replacement ratio among high-income countries (OECD 2009). Public pension expenditures were 12 percent of GDP and are projected to rise to an incredible 24 percent of GDP by 2050 (compared to a rise from 8 percent of GDP in 2010 to 12 percent of GDP by 2050 in the OECD). Moreover, the system was extremely fragmented, with 236 separate funds in 2003 (O'Donnell and Tinios 2003). Beyond the multiple inefficiencies and duplications this fragmentation caused, it also had a negative effect on labor mobility, as moving jobs often meant losing previous entitlements.

Finally, the pension system was extremely unequal, with large privileges handed to the professions and the public-sector employees. As a result, despite the high spending levels, the poverty risk for pensioners in Greece was 2.3 times larger than for the general population (versus 1.2 times in the European Union as a whole). In short, Greece's pension problem was considerably larger than that in other countries.

Some reforms had taken place in 1992 when the budget was under serious strain, but they did not tackle the long-term imbalances. Then, in the run-up to joining the euro, the government was under pressure to undertake these reforms and in 2001 tried to pass a reform package that had first been proposed in 1958 and was already considered at that time "extremely urgent" (Börsch-Supan and Tinios 2001). This proposal involved hard choices: "The retirement age was to be raised; the required insurance period for a seniority pension increased; the replacement rate reduced to 60 percent of reference earnings; the minimum pension raised but means-tested; and the lower retirement age for mothers of younger children replaced" (Featherstone 2005). However, in the face of massive protests, with the country booming and all sense of urgency gone, the proposals were withdrawn. A new reform package, characterized by creative accounting and little real reform, sailed through Parliament in 2002. Pension and other economic reforms were abandoned and not taken up again until it was too late to avert the crisis.

Portugal: Neither Demand for Reform nor Supply of Reform

After 15 years of economic growth that followed its accession to the European Union, Portugal's economy started to stagnate around 2000. Shockingly, in 2012, Portugal's output was lower than in 2001. For comparison, Spain's output was still nearly 17 percent higher in 2012 than in 2001 and Ireland's output was 19 percent higher. Portugal's total factor productivity fell in every year between 1999 and 2005 (see the KLEMS data set described in O'Mahony and Timmer 2009). The prime suspects for this drop include restrictions to competition in many sectors, the dominant position of large firms in several key industries, the difficulties for foreign management that sought to take over low-productivity Portuguese firms, and a dysfunctional labor market.

Portugal faced major macroeconomic imbalances, too. The headline government budget deficit never fell below 2.9 percent of GDP and the primary balance (government net borrowing or lending) was constantly in deficit, even after controlling for the effects of the business cycle and one-off and temporary adjustments (Marinheiro 2006, updated 2011). Public debt accumulated as a result, from 51.2 percent of output in 2001 to 92.4 percent in 2010. The private sector responded to the stagnant economic outlook by reducing its saving rate and heavily borrowing from abroad to finance current consumption, while investment actually fell as a percentage of national demand. This translated into persistent current account deficits of between 6 to 12 percent of output, an acute deterioration in the

real exchange rate, and an increasingly negative net asset position, mostly held by banks that had borrowed abroad to lend to local households.

In short, the behavior of both the public and the private sector was unsustainable in the middle run, a point well-recognized by many economists at the time. However, accession to the euro allowed both the public and the private sector to postpone the day of reckoning. In particular, the euro brought historically low nominal interest rates. For example, the yield on the 10-year government bond fell from 12 percent in 1995 to slightly less than 4 percent by early 2005. Consequently, while government debt as a share of GDP rose by 41 percentage points, interest paid on the debt barely budged: it was 2.9 percent of GDP in 2000 and 3.0 percent of GDP in 2010. The amount of private debt, and the interest payments on that debt, followed a similar pattern. Again, accession to the euro allowed Portugal's political-economic equilibrium to be sustained in the middle run by the large capital inflows from the rest of the world, even if a correction was eventually unavoidable.

There was no push for reform in Portugal because there was no "demand" for it, even less a "supply." On the demand side, a broad coalition that cut across traditional party lines supported the status quo. Large firms were reluctant to accept the liberalization of the markets for goods and services, entrenched managers were unwilling to be replaced by newcomers, inside workers resisted attempts to introduce more efficient labor regulations, and many low-income households benefited from increased social transfers (a rise of 4 points of GDP from 2000 to 2005) that succeeded in reducing Portugal's large income inequality and poverty rates. The inheritance from Portugal's historical pattern of inward development and the constraints created by the sudden change to democracy in 1974 made this coalition especially powerful and limited the scope of a more dynamic export sector that could have supported reforms.[7]

From the supply side, Portugal's parliamentary system created by the 1976 constitution provides little incentive for cooperation among the main political agents and makes decisive reforms difficult to approve. First, Portugal divides executive power between the president and the prime minister to a larger extent than other European countries, thus lacking the virtues of either purer presidential systems, such as France, or parliamentarian systems, such as Germany. As one example of this dysfunction, President Jorge Sampaio called for an early parliamentary election in 2005, an election centered to a large extent on the economic policies that Portugal needed to reactivate its economy, despite the fact that the government at the time held a solid parliamentary majority. Second, the electoral law, based on proportional representation, makes it hard for a single party to win an outright majority and forces coalition governments. Third, the power of the prime

[7] See Bermeo (2002) for the strong support of Portuguese voters for aggressive redistribution policies; Costa, Lains, and Miranda (2011) for a discussion of Portugal's historical pattern of growth; Fishman (2005) for the long-run political-economic consequences of the Revolution of Carnations of 1974; and Torres (2006), for the reluctance of important sectors of Portuguese elites to adopt the euro.

minister has been curtailed by the need to placate widely different constituencies (for example, in the Social Democratic Party ranging from right to center left).

When the economic crisis hit Portugal in 2008, private capital flows largely stopped. Portuguese banks, deeply exposed to sovereign debt of their own government, cut loans to firms, and the feedback loop from lower economic activity into lower tax revenue and higher sovereign risk left Portugal in a deep recession and with a banking sector in urgent need of recapitalization.

Even though exports have increased, the fixed exchange rate has prevented a faster adjustment and the current account still presents a substantial deficit that requires fresh external financing. At the same time, the institutional barriers we identified above, including the lack of a broad coalition supporting reform and the constitutional arrangements, have not been removed.

Germany: Without Financial Bubbles, Reforms Are Possible

In the years after the introduction of the euro, Germany undertook painful reforms of its welfare state. Why did the euro not have the same effect in Germany as in the peripheral countries like Spain, Ireland, Greece, and Portugal—namely, to postpone reforms? The answer is implied by the paths of interest rates shown earlier in Figure 1: neither the euro nor the financial boom changed financial conditions in Germany. For Germany, the euro meant tighter budgetary and fiscal constraints, not looser financial conditions. Absent the leeway provided by a financial boom, politicians had no choice but to act.

A decade ago, Germany was the "sick man" of Europe. The average growth rate in the second half of the 1990s and first years of the euro was barely above 1 percent. Unemployment in Germany stayed stubbornly high and had reached 11 percent in 2005. In addition, Germany was aging more rapidly than countries like Ireland and Spain; in Germany, the share of the population between 15 and 64 years of age peaked in 1987 at slightly above 70 percent and then started a steady decline. The sorry state of the East German economy and the difficulties of unification only added to the challenges (Akerlof, Rose, Yellen, and Hessenius 1991); indeed, while Spain and Ireland were enjoying real estate booms, Germany's housing prices were declining by about 10 percent from 1996 up through 2006.

This mediocre economic performance, the negative demographic trends, and the costs of the reunification shock put the German welfare state under severe strain. Compared with other countries, Germany's labor market policies were characterized by high expenditures and long duration of programs. Since social insurance schemes were essentially paid by employees, a decline in hours worked made the situation dire (Jacobi and Kluve 2007). Germany's unification exacerbated an already problematic state of affairs. Indeed, between 1990 and 1998 social insurance contribution rates for unemployment, health care, and pensions increased from 35.5 percent to 42.1 percent; German unification accounted for about half of that increase (Streeck and Trampusch 2005, p. 176). Even with the rise in contribution

rates, additional and growing federal subsidies from general tax revenue were required to finance up to one-quarter of the programs.

German politicians faced severe constraints. Monetary policy was under the control of the European Central Bank, which was establishing its reputation by setting a monetary policy for a newly created euro area and was unwilling to concede to the wishes of German politicians. The political fragmentation associated with German federalism prevented expansionary demand policies (Manow and Seils 2000). Because of the wide coverage of unions, real wage flexibility was limited. Reforms had long been slow in coming (Hassel 2010). In 1997, Chancellor Helmut Kohl introduced reforms aimed at stabilizing contribution rates to social insurance programs by including the use of demographic factors to account for increases in life expectancy, but these measures led to him being voted out of office in 1998, and the changes were reversed. Gerhard Schröder's first term as Chancellor was characterized by policies similar to those of other countries confronted with unsustainable welfare states: that is, a commitment to maintain benefits.

Unpopular reform was the only road left open. As Streeck and Trampusch (2005, p. 181) emphasize, "[h]aving stretched the federal budget to its limits, the measures of 1999 unintentionally forced the government to consider structural reform that went beyond short-term fiscal remedies." Schröder launched the Agenda 2010 program, the core of which was the Hartz I–IV reforms that constitute the greatest overhaul of the German welfare state since World War II. The Hartz reforms came only after much resistance—and a serious corruption scandal that finally forced the issue on the sitting cabinet—and probably cost Schröder the 2005 election (Helms 2007). The reforms changed a core principle of the German welfare state: whereas the system prevailing prior to these reforms was meant to preserve the social status of workers through retraining and public work schemes, the new system emphasized instead quick and sustainable job placement (see Bruttel and Sol, 2006, for the historical evidence on the adoption of "work first" approaches). In particular, job seekers were required to accept any offer of "suitable" work, where the definition of suitable was considerably broadened.

The long-run effects of the Hartz reforms are still being debated (Jacobi and Kluve 2007). But it is undeniable that, for Germany, joining the eurozone did not loosen budget constraints, but it did led to sweeping labor market reforms.

Conclusions

Many observers expected the arrival of the euro to lead to economic reform: when national governments lacked monetary autonomy and had only limited fiscal autonomy, they would face greater pressure to adopt structural reforms that they previously had refused to implement. Instead, nations in the periphery of Europe experienced a financial boom derived from the drop in interest rates and exchange rate risk. The budget constraints for these countries were loosened, rather than tightened. Moreover, both public and private accountability was diminished during

the boom because the consequences of bad decisions are largely imperceptible, at least in the short run, when rising asset prices hide all mistakes.

The line of argument we have pursued suggests several avenues for future research. First, while case studies can help to analyze the mechanisms at play, a more systematic empirical analysis of public and private governance during changing financial conditions is necessary to test our theory. Second, our hypothesis on signal extraction during financial booms needs to be formalized more fully. A final issue concerns the broader applicability of our analysis. Are all situations where financing is plentiful and cheap conducive to the lowering of standards, the deterioration of governance, and the postponement of needed economic reforms? If so, this situation is currently the one the United States, at the zero lower bound, is facing—in which case our analysis suggests that America's citizens and policymakers should be especially vigilant about the evolution of public and private governance.

■ *We sincerely thank the editors and Costas Arkolakis, Markus Brunnermeier, Manolis Galenianos, Philip Lane, Stavros Panageas, Elias Papaioannou, Canice Prendergast, Ricardo Reiss, Waltraud Schalke, Dimitri Vayanos, and Pierre Yared for their generosity with their time in discussing this paper with us. Luis Garicano declares that he is a board member of Liberbank.*

References

Akerlof, George A., Andrew K. Rose, Janet L. Yellen, and Helga Hessenius. 1991. "East Germany in from the Cold: The Economic Aftermath of Currency Union." *Brookings Papers on Economic Activity*, no. 1, pp. 1–106.

Alesina, Alberto, and Allan Drazen. 1991. "Why Are Stabilizations Delayed?" *American Economic Review* 81(5): 1170–88.

Barrett, Sean D. 1997. "The Implications of the Ireland–UK Airline Deregulation for an EU Internal Market." *Journal of Air Transport Management* 3(2): 67–73.

Bean, Charles. 1998. "Monetary Policy under EMU." *Oxford Review of Economic Policy* 14(3): 41–53.

Beltrán, Alejandro, Michele Boldrin, Nicolás Borges, João Castello-Branco, J. Ignacio Conde-Ruiz, Enrique García, Luís Garicano, and Pablo Vázquez. 2010. *A Growth Agenda for Spain*. Madrid: McKinsey and Co and FEDEA.

Bentolila, Samuel, and Gilles Saint-Paul. 2000. "Will EMU Increase Eurosclerosis?" CEPR Discussion Paper 2423, Center for Economic Policy Research.

Bermeo, Nancy. 2002. "Portuguese Democracy in Comparative Perspective." *Portuguese Studies Review* 10(1): 1–11.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." Chap. 21 in *Handbook of Macroeconomics*, Vol. 1, Part C, edited by John B. Taylor and Michael Woodford. Elsevier.

Bloomberg. 2008. "Ireland Guarantees Bank Deposit, Debt for Two Years." September 30.

Bonhomme, Stéphane, and Laura Hospido. 2012. "The Cycle of Earnings Inequality: Evidence from Spanish Social Security Data." Banco de España Working Papers 1225.

Börsch-Supan, Axel, and Platon Tinios. 2001. "The Greek Pension System: Strategic Framework for Reform." In *Greece's Economic Performance and Prospects*, edited by Ralph C. Bryant, Nicholas C. Garganas, and George S. Tavlas, 361–442. Bank of Greece and Brookings Institution.

Bover, Olimpia. 2011. "Efecto de la Rentabilidad de la Vivienda como Inversión sobre las Decisiones de Compra: Evidencia Empírica a partir de la EFF." Bank of Spain, Boletín Económico, May 2011.

Brown, John Murray. 2009. "Ahern Admits to Part in Irish Crisis." *Financial Times,* October 15.

Bruttel, Oliver, and Els Sol. 2006. "Work First as a European Model? Evidence from Germany and the Netherlands." *Policy and Politics* 34(1): 69–89.

Burnham, James B. 2003. "Why Ireland Boomed." *Independent Review* 7(4): 537–56.

Byrne, Elaine A. 2012. *Political Corruption in Ireland, 1922–2012: A Crooked Harp?* Manchester University Press.

Carswell, Simon. 2011. *Anglo Republic: Inside the Bank That Broke Ireland.* Penguin Books, London.

Casella, Alessandra, and Barry Eichengreen. 1996. "Can Foreign Aid Accelerate Stabilization?" *Economic Journal* 106(436): 605–619.

Chari, Varadarajan V., and Patrick J. Kehoe. 2008. "Time Inconsistency and Free-Riding in a Monetary Union." *Journal of Money, Credit and Banking* 40(7): 1329–56.

Costa, Leonor Freire, Pedro Lains, and Susana Münch Miranda. 2011. *História Económica de Portugal, 1143–2010.* Lisbon: Esfera dos Livros.

Cuñat, Vicente, and Luis Garicano. 2009. "Did Good Cajas Extend Bad Loans? The Role of Governance and Human Capital in Cajas' Portfolio Decisions." FEDEA Annual Conference at Bank of Spain, 2009. http://www.crisis09.es/monografia2009/cajas.html.

Cuñat, Vicente, and Luis Garicano. 2010. "¿Concedieron las cajas 'buenas' créditos 'malos'? Gobierno corporativo, capital humano y carteras de créditos." Chapter 6 in *La Crisis de la Economía Española,* Fedea Monograph, edited by Samuel Bentolila, Michele Boldrin, Javier Díaz-Giménez, and Juan J. Dolado. Descargar. http://www.crisis09.es/cajas.html.

Drazen, Allan. 2000. *Political Economy in Macroeconomics.* Princeton University Press.

Featherstone, Kevin. 2005. "'Soft' Co-ordination Meets 'Hard' Politics: The European Union and Pension Reform in Greece." *Journal of European Public Policy* 12(4): 733–50.

Fernández Ordóñez, Miguel Angel. 2003 "El legado de Rato." *El País,* September 11.

Fernández-Villaverde, Jesús, and Juan F. Rubio-Ramírez. 2009. "Una Propuesta de Política Fiscal ante la Crisis." In *La Crisis de la Economía Española: Lecciones y Propuestas.* Madrid: FEDEA.

Fernández-Villaverde, Jesús, Luis Garicano, and Tano Santos. 2013. "Political Credit Cycles: The Case of the Euro Zone." NBER Working Paper 18899.

Fishman, Robert M. 2005. "Legacies of Democratizing Reform and Revolution: Portugal and Spain Compared." Working Paper 1-05, Instituto de Ciências Sociais, Lisbon.

Gelb, Alan H. 1988. *Oil Windfalls: Blessing or Curse?* Oxford University Press and World Bank Research Publications.

González, Libertad, and Francesc Ortega. 2009. "Immigration and Housing Booms: Evidence from Spain." CReAM Discussion Paper 0919, Center for Research and Analysis of Migration, University College London.

Hassel, Anke. 2010. "Twenty Years after German Unification: The Restructuring of the German Welfare and Employment Regime." *German Politics and Society* 28(2): 102–15.

Helms, Ludger. 2007. "The German Federal Election, September 2005." *Electoral Studies* 26(1): 223–27.

Honohan, Patrick. 2010. "The Irish Banking Crisis, Regulatory and Financial Stability Policy, 2003–2008: A Report to the Minister for Finance by the Governor of the Central Bank." Available at: http://www.bankinginquiry.gov.ie/Preliminary_Reports.aspx.

Honohan, Patrick, and Brendan Walsh. 2002. "Catching up with the Leaders: The Irish Hare." *Brookings Papers on Economic Activity* 33(1): 1–78.

Jacobi, Lena, and Jochen Kluve. 2007. "Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany." *Zeitschrift für ArbeitsmarktForschung – Journal for Labor Market Research* 40(1): 45–64.

James, Harold. 2012. *Making the European Monetary Union.* Cambridge: Belknap Press.

Juan, Jorge. 2011. *Nada es Gratis: Cómo Evitar la Década Perdida tras la Década Prodigiosa.* Barcelona: Ediciones Destino.

Lane, Philip R. 2012. "The European Sovereign Debt Crisis." *Journal of Economic Perspectives* 26(3): 49–68.

Lane, Philip R., and Peter McQuade. 2012. "Domestic Credit Growth and International Capital Flows." http://www.bde.es/investigador/papers/sie1210.pdf.

Lyons, Tom, and Brian Carey. 2011. *The FitzPatrick Tapes: The Rise and Fall of One Man, One Bank, and One Country.* London: Penguin Books.

Manow, Philip, and Eric Seils. 2000. "The Employment Crisis of the German Welfare State." *West European Politics* 23(2): 137–60.

Marinheiro, Carlos. 2006. "The Sustainability of Portuguese Fiscal Policy from a Historical Perspective." *Empirica* 33(2): 155–179 (updated in 2011: http://www4.fe.uc.pt/carlosm/research/pdf/Data-CMarinheiro2006-Update2011.xlsx).

Mitsopoulos, Michael, and Theodore Pelagidis.

2012. *Understanding the Crisis in Greece*, Revised edition. London: Palgrave MacMillan.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance*. Cambridge University Press.

Nyberg, Peter [Commission of Investigation into the Banking Sector in Ireland]. 2011. "Misjudging Risk: Causes of the Systemic Banking Crisis in Ireland." Available at: http://www.bankinginquiry.gov.ie/. [The author listed on the document is "Commission of Investigation into the Banking Sector in Ireland" but according to the webpage, "Peter Nyberg was appointed by the Minister as the Sole Member of the Commission and is fully responsible for the contents of the Report."]

O'Donnell, Owen, and Platon Tinios. 2003. "The Politics of Pension Reform: Lessons from Public Attitudes in Greece." *Political Studies* 51(2): 262–81.

OECD. 2009. *Pensions at a Glance 2009: Retirement-Income Systems in OECD Countries*. Organisation for Economic Co-operation and Development.

OECD. 2011. *Greece: Review of the Central Administration*. OECD Public Governance Reviews. Organisation for Economic Co-operation and Development.

O'Mahony, Mary, and Marcel P. Timmer. 2009. "Output, Input and Productivity Measures at the Industry Level: the EU KLEMS Database." *Economic Journal* 119(538): F374–F403.

Papademos, Lucas D. 2001. "Opening Address: The Greek Economy: Performance and Policy Challenges." In *Greece's Economic Performance and Prospects*, edited by Ralph C. Bryant, Nicholas C. Garganas, and George S. Tavlas, xxxiii–xxxix. Bank of Greece and Brookings Institution.

Rae, David, and Paul van de Noord. 2006. "Ireland's Housing Boom: What Has Driven It and Have Prices Overshot?" OECD Economics Department Working Paper 492.

Reinhart, Carmen M., and Kenneth Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Sachs, Jeffrey D., and Andrew Warner. 1995. "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity* no. 1, pp 1–95.

Somerville, R. Andrew. 2007. "Housing Tenure in Ireland." *Economic and Social Review* 38(1): 107–134.

Streeck, Wolgang, and Christine Trampusch. 2005. "Economic Reform and the Political Economy of the German Welfare State." *Politics of Economic Reform in Germany*, Special Issue, pp. 174–95.

Svensson, Jakob. 1999. "Foreign Aid and Rent-Seeking." *Journal of International Economics* 51(2): 437–461.

Torres, Francisco. 2006. "A Convergência Monetária: Portugal e a União Europeia." Working Papers de Economia 39, Departamento de Economia, Gestão e Engenharia Industrial, Universidade de Aveiro.

Vamvakidis, Athanasios. 2007. "External Debt and Economic Reform: Does a Pain Reliever Delay the Necessary Treatment." IMF Working Paper 07/50.

Whelan, Karl. 2010. "Policy Lessons from Ireland's Latest Depression." *Economic and Social Review* 41(2): 225–54.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 167–192

Cross of Euros[†]

Kevin H. O'Rourke and Alan M. Taylor

The eurozone is in trouble. Unemployment is over 12 percent and is getting worse. Youth unemployment is 24 percent. In Spain and Greece, total unemployment exceeds 26 percent, and youth unemployment exceeds 55 percent. These two countries, together with Ireland, Cyprus, and Portugal, are in official bailout programs. If Slovenia eventually joins the club, as seems possible, then a third of the eurozone's 17 members will be in such programs. Capital controls imply that Cypriot euros are no longer convertible into euros elsewhere. There has been a sharp decline in eurozone citizens' confidence in European institutions: reported "distrust" in the European Union exceeds "trust" in 15 out of 17 countries and by an average 28 percentage points overall. In Greece, the fascist Golden Dawn party entered Parliament in 2012 and is gaining in opinion polls. Economies, societies, and political systems are fraying at the seams.

In the Cypriot, Irish, and Spanish cases, banking crises caused economic collapse and loss of political sovereignty. Pre-crisis cross-border flows of capital pushed up wages, prices, and asset prices in recipient countries, implying major adjustment problems that now have to be faced, but the fact that many of these flows have been

■ *Kevin H. O'Rourke is Chichele Professor of Economic History and Fellow of All Souls College, both at the University of Oxford, Oxford, United Kingdom. Alan M. Taylor is Professor of Economics and Finance, Department of Economics and Graduate School of Management, University of California, Davis, California. O'Rourke is Director, Economic History Programme and Taylor is a Research Fellow, Centre for Economic and Policy Research, London, United Kingdom. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are kevin.orourke@all-souls.ox.ac.uk and amtaylor@ucdavis.edu.*

[†] To access the disclosure statements, visit
http://dx.doi.org/10.1257/jep.27.3.167

doi=10.1257/jep.27.3.167

channeled through banks has led to some of the eurozone's most intractable problems. The costs of dealing with banking crises has worsened governments' fiscal positions, putting further strain on banks' balance sheets, in turn crimping credit creation, thus leading to a further deterioration in the economy and governments' fiscal positions, and so on.

The eurozone crisis gives rise to three questions. First what macroeconomic policy mix is consistent with running a diverse monetary union involving 17 independent nation states? More specifically, what policy mix will be required to ensure that the adjustment problems that countries like Greece and Spain now face can be successfully overcome and that the eurozone does not collapse in the short to medium run? Second, what is the minimum institutional framework consistent with the survival of the eurozone in the medium to long run? If macroeconomic adjustment problems cannot be overcome in the shorter run, and if necessary institutional reforms cannot be delivered in the longer run, then a third question becomes potentially relevant: how can the costs of a eurozone break-up be minimized?

Although there are no convincing historical analogies for the eurozone, which is a unique experiment (Eichengreen 2008), we begin with an overview of some of the historical analogies that are often mentioned. Nonetheless, history can provide us with lessons regarding all three questions. The gold standard provides lessons regarding what short-run adjustment strategies the eurozone should be pursuing today (Eichengreen and Temin 2010). The history of American monetary union provides lessons regarding what institutions the eurozone will need in the longer run to survive, and, perhaps more pessimistically, the circumstances in which these are likely to come about, if they ever do. Finally, history also provides lessons relevant to the break-up of the eurozone, should it come to that.

Why Previous "Monetary Unions" Offer a Poor Analogy

In this section, we offer some comparisons and contrasts between the euro and previous arrangements that superficially appear somewhat similar, but were actually very different. Table 1 summarizes the other arrangements and the key points of comparison.

European monetary union has eliminated exchange rate variability among eurozone members by replacing national currencies with a single currency, the euro. The euro is managed by a common European Central Bank whose primary objective is price stability, defined in practice as involving inflation less than 2 percent. If this goal is satisfied, the central bank is also supposed to support "general economic policies in the Union" with a view to achieving objectives such as full employment. The eurozone members are all members of the European Union, but remain independent states. Under the original architecture, national authorities handle banking supervision, resolution, and deposit insurance; there is no banking union. Neither is there a common eurozone fiscal authority, nor anything approaching a

Table 1
Comparing Currency Unions

	EMU	*Latin Monetary Union*	*Scandinavian Monetary Union*	*Anglo-Irish monetary union*	*Currency boards*	*United States*	*Gold standard*
Does it eliminate exchange rate variability?	**Yes**	No	No	Yes	Yes	Yes	No
Does it eliminate national currencies?	**Yes**	No	No	No	No[d]	Yes	No
Is exit easy?	**No**	Yes	Yes	Yes	Yes	No	Yes
Is there a temporary escape clause?	**No**	Yes[a]	Yes[a]	No	No	No	Yes
Is there a common central bank?	**Yes**	No	No	No	No[e]	Yes	No
Are high-denomination coins mutually acceptable?	**N/A**	Yes	Yes	N/A	N/A	N/A	No
Are low-denomination coins mutually acceptable?	**Yes**	No	Yes	Yes[c]	Varies	Yes	No
Is paper currency mutually acceptable?	**Yes**	No	Yes[b]	Yes[c]	Varies	Yes	No
Is there a fiscal union?	**No**	No	No	No	No	Yes	No
Is there a political union?	**No**	No	No	No	No	Yes	No
Is there a banking union?	**No**	No	No	No	No	Yes	No
Is the union symmetric?	**Yes**	Yes	Yes	No	No	Yes	Yes
High labor mobility?	**No**	Partial	Partial	Yes	Varies	Yes	Partial

[a] Could suspend convertibility of paper currency into specie.
[b] Occurred in stages; agreements in 1894 and 1901.
[c] British currency accepted in Ireland.
[d] Except in cases involving dollarization.
[e] There are two common central banks in the CFA franc zone, corresponding to the West African and Central African Currency Unions.

eurozone government. Bailouts of member states are supposedly prohibited, and a series of fiscal rules were (unsuccessfully) adopted to make this credible. There is no legal means of leaving the eurozone, even temporarily, aside from leaving the European Union altogether.

The Latin Monetary Union was created in 1865, initially involving Belgium, France, Italy, and Switzerland. Its purpose was to harmonize the gold and silver content of the coins of the four countries (Redish 1993); it was a coinage agreement, not a monetary union. There was no common unit of account, no common political framework, and no common central bank. The Scandinavian Monetary Union involved Denmark, Norway, and Sweden (Henriksen and Kærgård 1995). Gold coins, and token silver and bronze coins, were legal tender in all three countries. In 1885 the three central banks opened current accounts with each other, gaining the right to draw drafts in each other's currencies at par. Nevertheless, the central banks remained independent of each other, there was no economic policy coordination in other areas, and countries retained their own currencies.

There have been a number of small-scale "monetary unions," typically between a larger and a smaller state, in which each state's currency is legal tender in the other (or the larger state's currency has legal tender status in the smaller state). The Anglo-Irish monetary union which lasted from Irish independence in 1922 until 1979 is one such example: Honohan (1994) considers this to have been *de facto* a currency board arrangement throughout virtually the entire period. History records more than 60 currency board systems, mostly occurring in colonial situations where dependent states and territories issued their own currency (Hanke and Schuler 1998). The key requirements are that the issuer must freely exchange local for foreign currency at par and must hold enough foreign-denominated safe assets to cover its entire monetary base liabilities. Currency boards typically do not involve a common central bank or currency, and countries can choose to leave at any time.

The United States has a *true* monetary union, not simply a more or less hard exchange rate peg between state currencies. As we will see, it gradually developed a common central bank, a banking union, and a fiscal union. The obvious difference between the United States and the eurozone is that in the American case political union preceded monetary union, while the European gamble has been to try to develop a monetary union in the absence of political (and fiscal and banking) union.

Finally, the gold standard was not even formally speaking an exchange rate agreement. Rather, it was a series of country-by-country monetary regimes linking the value of currencies to the price of gold, obliging central banks or their equivalents to hold sufficient reserves to be able to make this commitment credible. It only became a quasi-fixed exchange rate regime as a by-product of free trade in gold, which led to gold prices being almost (not entirely as arbitrage was costly) equalized in different countries. Countries retained their own currencies, central banks, and political and financial sovereignty and could sever the link between their currencies and gold whenever they wished.

The Adjustment Problem: What Can We Learn from the Gold Standard?

Most of the debate in the run-up to the adoption of the euro was couched in terms of traditional optimum currency area theory (Mundell 1961; Kenen 1969). If the benefits of a common currency are that it increases trade, then benefits should be increasing in the extent of trade integration within the currency area. What about the costs? Here, optimum currency area theory focuses on how the regions within the area are able to adjust to macroeconomic shocks. If shocks are symmetric across regions, then a common monetary policy response is appropriate, reducing the cost of a common currency. If shocks are asymmetric, then labor flows from depressed to booming regions will help adjustment, as will wage and price flexibility, or a central fiscal authority that can smooth shocks across regions; but if these

alternative adjustment mechanisms are absent, the regions may be better off with separate currencies so that the exchange rate can be used as a tool of adjustment. The history of the gold standard is a rich source of lessons on how adjustment across countries takes place in modern economies and polities.

The Gold Standard, the Trilemma, and Adjustment in Theory and Practice

The gold standard was supposed to ensure aggregate price stability by making it impossible for governments to engage in inflationary policies. It also offered the prospect of microeconomic benefits by encouraging international integration, and indeed it boosted trade by more than the euro has done (Estevadeordal, Franz, Taylor 2003; Mitchener and Weidenmier 2008; Baldwin 2006; Santos Silva and Tenreyro 2010). But how did it cope with macroeconomic adjustment?

The gold standard operated as a straitjacket on macroeconomic policy, according to the macroeconomic policy *trilemma* which says that a country cannot simultaneously choose three policies: 1) a fixed exchange rate, 2) open capital markets, and 3) monetary policy autonomy. It must pick two. If a country chooses open capital markets, "uncovered interest parity" must hold; that is, since arbitrage equalizes expected returns at home and abroad, the domestic interest rate must equal the foreign interest rate plus the expected appreciation of the foreign currency. If a country chooses open capital markets *and* fixed exchange rates, domestic interest rates have to equal the base-country interest rate, ruling out monetary policy autonomy. If a country chooses open capital markets *and* wishes to set domestic interest rates at levels suitable to domestic conditions, then exchange rates can no longer be fixed. However, a country can choose an autonomous monetary policy and a fixed exchange rate if it imposes capital controls. While the trilemma is a simplification, ample historical evidence supports its key predictions (Obstfeld, Shambaugh, and Taylor 2004, 2005). It provides a useful organizing framework for international macroeconomic history (Eichengreen 1996; Obstfeld and Taylor 2004) as the essential historical plot lines revolve around which leg of the trilemma countries have chosen to sacrifice.

The gold standard, like the eurozone and the US monetary union, offers lessons about what happens when the exchange rate across an area is fixed and capital markets are open, implying that monetary policy is the same across gold standard adherents, eurozone members, or US states, as the case may be. Many scholars have stressed the potential for economic instability in this setting, and the paucity of tools available in response. With a flexible exchange rate, a local demand boom would drive up interest rates, drawing in capital from abroad and appreciating the currency. This in turn would moderate the boom. With a fixed exchange rate, local interest rates cannot rise. Indeed, local central banks must prevent interest rates from rising by expanding money supplies, amplifying the boom. The opposite follows in slumps. As Ford wrote (1962, p. 188), "It is easy to understand the dislike of some Argentines for a system which dictated that a slump must be aggravated by monetary reactions, although, doubtless, they had forgotten that the same system served to enhance booms."

Under the gold standard, the key goal for policymakers was to avoid losing gold. When a country found itself with a balance-of-payments deficit (that is, a current account surplus insufficient to finance capital outflows, or borrowing insufficient to finance a current account deficit), it needed a mechanism to staunch the resulting outflow of gold and eventually to reverse it. Such a mechanism involved, among other things, lowering the prices of domestic goods relative to those of foreign goods (depreciating the real exchange rate), thus improving the trade balance. Such "real depreciation" can be achieved by depreciating the nominal exchange rate—that is, lowering the value of the currency in which domestic prices are expressed—but this step was ruled out under the gold standard. Real depreciation had to be achieved by lowering the domestic price level, a strategy sometimes referred to as "internal devaluation" in the eurozone context.

In Hume's (1742) formulation of the "price-specie-flow mechanism," adjustment was supposed to be automatic. He argued that gold outflows, which were needed to pay for trade deficits, would lower the money supply, since the latter was tied to gold reserves. This in turn would lower the internal price level, depreciate the country's real exchange rate, and improve the trade balance. Conversely, surplus countries would experience gold inflows, inflation, and real exchange rate appreciation.

The late nineteenth century world was more complicated than the world envisaged by Hume. Rapidly expanding international financial markets meant that trade deficits could be financed by borrowing rather than by gold exports. Yet interwar observers, looking back at the pre-1914 experience, believed that adjustment under the gold standard had been smooth as a result of monetary authorities following the "rules of the game." Central banks in deficit countries were supposed to raise discount rates, thus shrinking money supplies and allowing for Humean price reductions and real depreciations. In surplus countries, they were supposed to lower discount rates, implying symmetric adjustment. Economic historians have known for a long time that central banks did *not* follow the rules of the game: adjustment was far from automatic, much less symmetric (Bloomfield 1959; Morgenstern 1959; Eichengreen 1992, Chapter 2). And yet the classical gold standard worked fairly smoothly during the late nineteenth century, at least in core economies such as Britain, France, and Germany.

One reason for this was that key economies such as Britain did not suffer from severe balance-of-payment imbalances so not much adjustment was required.[1] In addition, five specific pre-1914 economic and political conditions meant that insofar as macroeconomic adjustment was needed, it was easier to achieve it than it would become in the interwar period (or in the eurozone today).

[1] For example, Britain's payments remained fairly well balanced because sterling's role as a "vehicle currency" facilitating international transactions meant that long-term capital outflows were in part matched by short-term capital inflows, as borrowers placed money on deposit in Britain; and they were in part matched by exports of British capital goods (Eichengreen 1992, 2008).

First, wages and prices were more flexible then than subsequently: Hanes and James (2003) find no evidence of downward nominal wage rigidity in the United States between 1841 and 1891. This nominal flexibility was already declining before World War I: Hanes (1993, 2000) finds a decline in flexibility from the 1890s onwards associated with the spread of large-scale, capital-intensive, concentrated industry. Cross-country analysis by Basu and Taylor (1999) and Chernyshoff, Jacks, and Taylor (2009) provides further evidence of greater nominal flexibility in the pre-1914 era. By contrast, the escape route of internal devaluation via downward nominal price adjustment appears elusive in today's world, where most significant real depreciations have come through nominal exchange rate adjustments (Shambaugh 2012).

Second, even in cases where macroeconomic adjustment increased unemployment (for example, because falling prices and downwardly sticky wages implied rising real product wages), typical nineteenth century limits on who was allowed to vote meant that the interests of the workers who suffered most could be largely ignored by policymakers with few adverse repercussions (Eichengreen 1992). Modern democracies work differently.

Third, this period was also one of international mass migration, whose timing was influenced by business cycle conditions, and which therefore relieved labor market pressures during periods of stress (Hatton 1995; Hatton and Williamson 1998, chap. 4).

Fourth, limited political opposition to the gold standard, and ample gold reserves that were spread out among the core countries, implied that policymakers' commitment to the gold standard was usually regarded as credible. Credibility implied that capital flows tended to be stabilizing: that is, if an exchange rate started depreciating, it was expected that it would soon appreciate, meaning that private investors would buy the currency—thus helping to bring about the needed appreciation and in the process reversing gold outflows (Eichengreen 1992).

Fifth, when these mechanisms did not suffice, international cooperation between core central banks willing to lend to each other, or intervene together, could be relied upon to stabilize the situation.

It was a different story for countries in southern and eastern Europe, Latin America, and Asia. Trade with these countries amounted to two-thirds of core European trade, and more than 40 percent of US trade. Some pegged to silver, others had inconvertible currencies, and still others tried to peg to gold with only sporadic success. Catão and Solomou (2005) find evidence of large nominal and real depreciations in peripheral economies vis-à-vis the core during time of crisis, such as the late 1870s and early 1890s, and also find that trade balances improved when real exchange rates depreciated.

Summing up, adjustment under the classical gold standard was, in principle, supposed to involve "internal devaluation." Such a strategy was easier before World War I in both economic and political terms than it was during the interwar period. Nevertheless, in the core economies adjustment typically happened in other ways, and only limited adjustment was required in the first place. In the periphery, where

more adjustment was required at times of stress in the international economy, and where countries did not benefit from the same international cooperation that core economies enjoyed, countries frequently adjusted via nominal depreciation. Even in the heyday of the gold standard, the "internal devaluation" strategy was nowhere near as ubiquitous as is sometimes thought.

The Gold Standard and the Great Depression

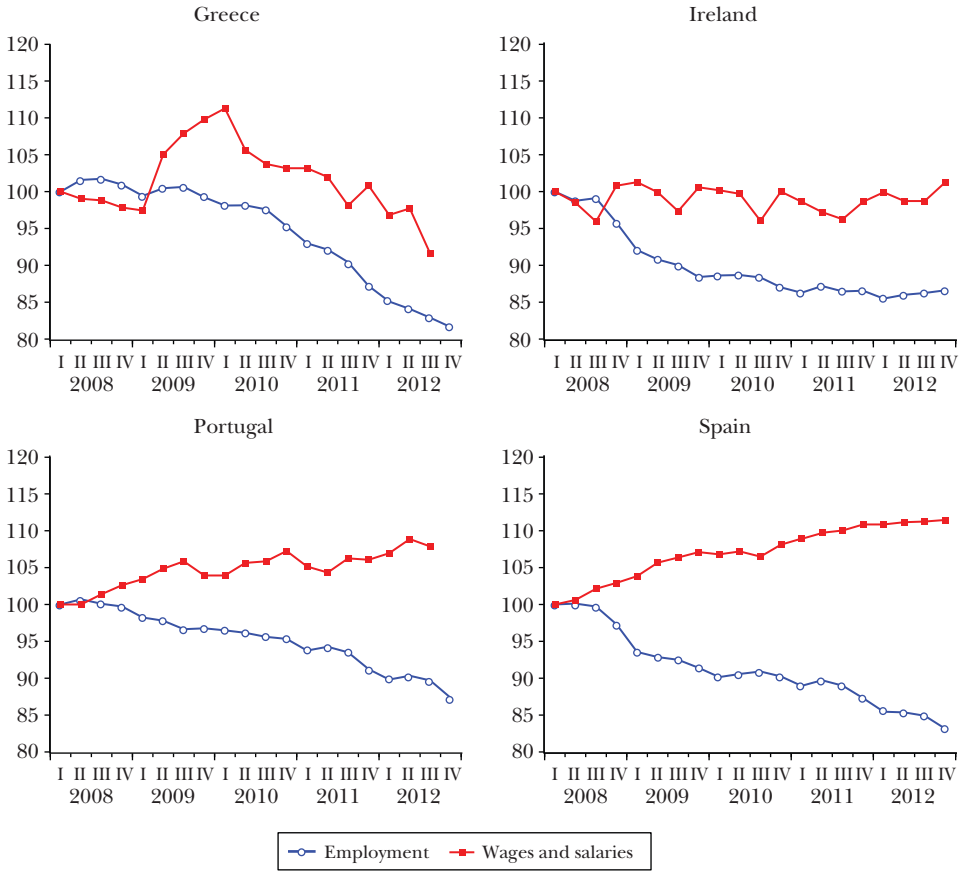
The economic and political environment was very different after World War I, implying that the gold standard worked much less smoothly than it had before (Kindleberger 1973; Temin 1989; Eichengreen 1992).

First, the underlying imbalances facing core economies and requiring adjustment became much larger than previously. Britain's balance-of-payments position was more fragile. Its net international asset position had been greatly weakened by the war, implying less investment income, while war debts were another drain on the economy. Its trade position had been weakened as a result of competitors seizing overseas markets during the conflict. The pound went back onto the gold standard at the pre-war rate in 1925, implying an overvalued exchange rate, particularly vis-à-vis the US dollar and the French franc. Britain experienced large balance-of-payments deficits from about 1925 onwards and a deep industrial slump. Moving beyond the British case, the United States was now emerging as the largest net lender, especially to Latin America and countries such as Germany. When the Federal Reserve raised interest rates in 1928 in an attempt to halt a runaway stock market, these borrowing countries faced a sudden stop in capital imports, and with it, a need to adjust. Many countries abandoned the gold standard soon thereafter.

Second, the process of adjustment to economic shocks became less smooth than it had been before 1914 (Eichengreen 1992). Nominal wages were more rigid. Voting rights had broadened. The macroeconomic and political costs associated with adjustment based on internal devaluation were thus higher than before. These changes reduced the credibility of governments' commitment to the gold standard, implying that capital flows were now potentially destabilizing (Obstfeld and Taylor 2003). Credibility also suffered because World War I had brought general distrust, periods of capital controls, devaluations, and later some hyperinflations, trade barriers, and other shocking forms of economic and political uncertainty. International cooperation between central banks was less effective given political frictions and different views on what constituted appropriate economic policy.

Perhaps most damaging was the asymmetric nature of international adjustment under the interwar gold standard. Countries with a balance-of-payments deficit (like Britain) had an incentive to raise discount rates to prevent gold outflows, while surplus countries (like France) who were experiencing gold inflows had an incentive to "sterilize" them—that is, to adjust the money supply in ways that prevented the inflow of gold from causing inflation—so that they would continue accumulating gold. World gold reserves rose steadily between 1925 and 1932, but with the United States not lowering its gold holdings, and with France rapidly increasing its reserves, there was insufficient gold elsewhere (Irwin 2010). Deflation in countries

Figure 1
Average Nominal Wages and Salaries and Total Employment, 2008–2012



Source: Eurostat.
Note: The wage data are adjusted for working days; no seasonal adjustment in the case of Ireland.

like Germany was not matched by inflation elsewhere, making macroeconomic adjustment all the more difficult.

Several lessons from this disastrous interwar experience are directly relevant for today's Europe.

First, nominal wages were sticky downward during the Great Depression (Bernanke and Carey 1996), implying that deflation led to rising real wages, and falling employment and output. Wages were not unusually rigid during this period (Hanes 2000); rather, downward stickiness is a fact of life in modern economies. Figure 1 shows indices of wages and salaries between 2008 and 2012 in Greece, Ireland, Portugal, and Spain, four countries currently trying to achieve nineteenth-century style internal devaluations. As can be seen, wages have been steadily rising in Portugal and Spain, despite very high levels of unemployment there. Even in

Ireland, a country widely regarded as having unusually flexible labor markets and as having successfully accomplished an "internal devaluation," there is no sign of wages falling, although they have managed to avoid rising. In all four countries, by contrast, employment levels have been continually falling, although the Irish decline came to an end in the second half of 2012.[2]

The one important eurozone exception to the general conclusion that nominal wages are rigid downwards is Greece, where manufacturing wages declined by more than 10 percent in the three years starting in 2010. The impact of the depression on the fabric of Greek society has been particularly harsh: if this is what it takes to produce nominal wage declines, prudence might suggest alternative adjustment mechanisms, such as rising wages and prices in surplus countries. As in the interwar period, however, eurozone countries running current account surpluses are reluctant to accept temporarily higher inflation rates.

Second, deflation during the interwar period was dangerous in other ways. It increased the real value of debts, placing indebted households, businesses, and financial institutions under pressure (Fisher 1933). It weakened bank balance sheets in the financial crisis, with knock-on effects for businesses reliant on bank lending. It increased real interest rates and induced households to postpone expensive purchases. Deflation helped deepen the Depression; even if internal devaluation were possible in modern economies, deflation would not be desirable.

Third, large public debts are difficult or impossible to stabilize when deflation is increasing the real value of the debt and slowing economic growth.[3] During the interwar period, Britain ran primary budget surpluses of 7 percent of GDP during the 1920s. Despite these efforts, the deflationary low-growth environment meant that the British debt-to-GDP ratio increased substantially over the decade. The IMF's (2012, p. 112) conclusion is that this episode is "an important reminder of the challenges of pursuing a tight fiscal and monetary policy mix, especially when the external sector is constrained by a high exchange rate."

Fourth, as the interwar period wore on, more countries (such as Germany) attempted to adjust based not only on internal devaluation, but also with fiscal austerity. This strategy was costly, since fiscal multipliers were high in the 1930s, given weak economies and interest rates affected by the zero lower bound. Almunia, Bénétrix, Eichengreen, O'Rourke, and Rua (2010) find multipliers well in excess of one in a sample of 27 countries between 1925 and 1939; thus, fiscal austerity policies amplified the Great Depression.

Fifth, countries only started to recover from the Great Depression once they left the gold standard (Eichengreen and Sachs 1985; Campa 1990). Revaluing countries' gold reserves as they exited made it possible to boost the money supply. In leaving

[2] Constant nominal wages are consistent with falling unit labor costs if labor productivity increases. Irish unit labor costs fell in the initial stages of the crisis, but the effect is partly a statistical illusion due to a shift in the composition of the Irish workforce, with low-productivity workers being laid off (Darvas 2012).

[3] Flandreau, Le Cacheux, and Zumer (1998) show that European public debts tended to rise as a percentage of GDP in the deflation of 1873–96, before falling thereafter as prices rose.

gold, expectations of deflation were replaced by expectations of inflation (Temin and Wigmore 1990; Romer 1992; della Paolera and Taylor 1999; Eggertsson 2008). There were transitory competitiveness gains for early movers who depreciated first. Countries also tended to do better when they embraced capital controls and used the policy space so liberated, even if their exchange rate remained officially pegged to gold (Obstfeld and Taylor 2004). Regaining monetary independence, one way or another, was the route to recovery.

Sixth, the Depression had calamitous political consequences. Voting for extremism was negatively related to GDP growth during this period, at least in countries that had not been inoculated by a history of democracy stretching back to before World War II (de Bromhead, Eichengreen, and O'Rourke 2013). Ponticelli and Voth (2011) find a strong correlation between fiscal austerity and political chaos (as measured by riots and other disturbances) over the last 100 years or so, and the result is robust when restricted to the interwar sample. It is foolish to ignore the potential political consequences of internationally lopsided and deflationary adjustment strategies.

The experience of the 1930s is not only a cautionary tale of the limitations of adjustment strategies based on internal devaluation and fiscal austerity, but an illustration of the power of monetary policy and of the value of macroeconomic policy flexibility. It is a useful reminder that Keynes' short run is the time frame within which politics occurs, for good or ill.

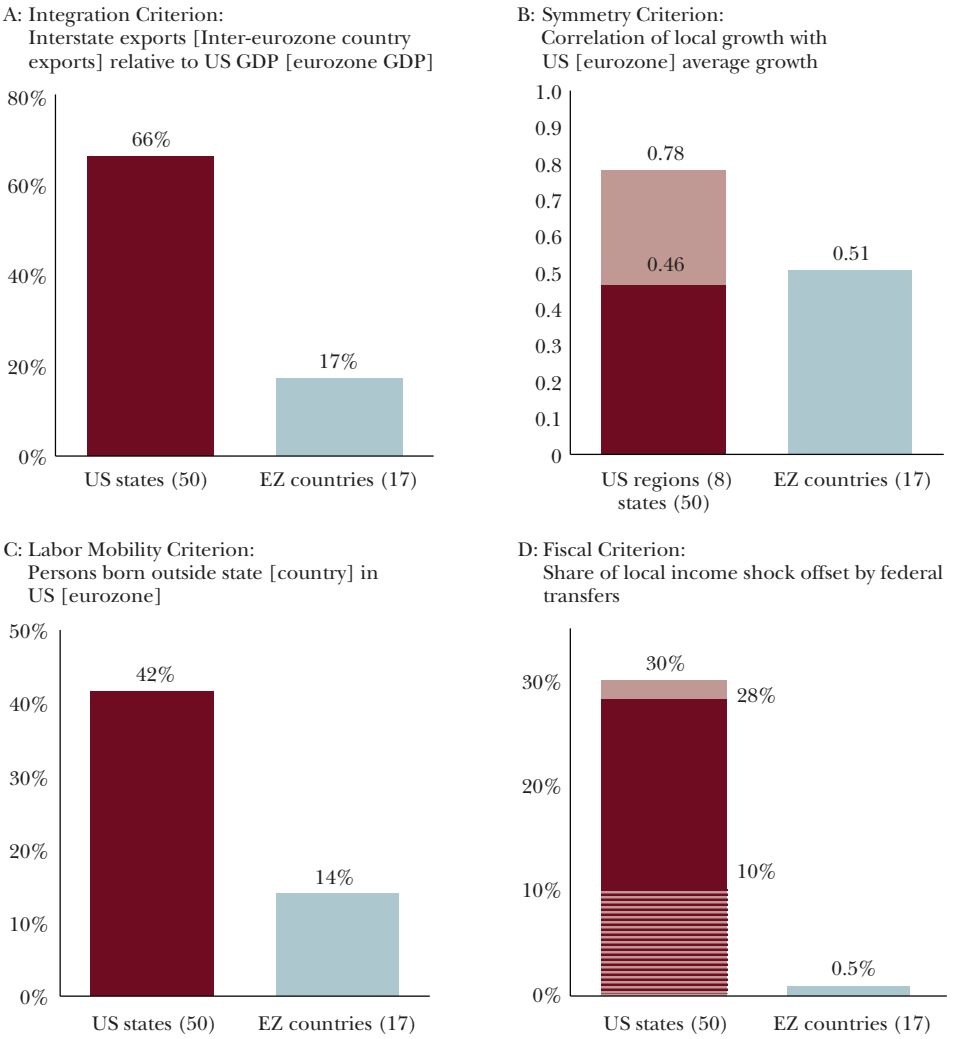
From Optimum Currency Area Theory to the Fiscal and Banking Nexus: Lessons from the United States

The United States is more likely than the eurozone to satisfy the three Mundell-style optimal currency area criteria regarding the integration of product markets, symmetry of shocks, and labor mobility, as well as Kenen's criterion regarding the ability of a central fiscal authority to smooth shocks across regions. Figure 2 illustrates some of the key differences organized around these four criteria.

Regarding market integration, Panel A shows that cross-border interstate trade amounts to 66 percent of GDP in the United States; in the 17-country eurozone, such trade amounts to only 17 percent of eurozone GDP. The US economy is strongly ahead on this criterion.

For the symmetry criterion, we look at the correlation between local growth and growth in the monetary union as a whole. Panel B indicates that the average correlation between real GDP growth in the eight US Census regions and national real GDP growth is 0.78. In the eurozone, the average correlation between GDP growth in the eurozone countries and GDP growth across the whole eurozone is about 0.5. Thus, on the symmetry criterion, the eurozone has lower average correlations between the shocks in its constituent countries than we see in the United States (and the eurozone correlations are also far more varied). This difference may reflect aggregation in large Census regions: it disappears if we take the sample

Figure 2
Optimum Currency Area Criteria: Eurozone versus the United States



Notes and Sources: Panel A: Data for 2007; intra-US trade volumes from the 2007 Commodity Flow Survey, minus 2007 US imports from the Bureau of Economic Analysis, National Income and Product Accounts (BEA NIPA), all divided by GDP from BEA NIPA; eurozone trade with eurozone, and eurozone GDP, from Eurostat. Panel B: Annual data for 1997–2007; US, state, and census region real GDP growth rates from BEA (the higher correlation statistic is for the regions); eurozone and eurozone country growth rates from Eurostat. Panel C: Data for 2012; US data from Statistical Abstract; eurozone data from Eurostat. Panel D: Upper and lower range of US estimates and euro point estimate taken from multiple older sources in HM Treasury (2003); US 28 percent figure based on recent Federal income tax elasticities alone, from Auerbach (2009, figure 2).

of 50 US states, but these are smaller and more diverse units than the 17 eurozone countries. The US economy has perhaps a minor advantage as a single currency area based on the symmetry criterion.

As regards the labor mobility criterion, in Panel C, the average share of people in a US state who were born outside that state is 42 percent. The equivalent index for the eurozone, people born outside the country where they currently reside, is only about 14 percent. This difference is deep-seated: the US economy attained something approaching a single labor market sometime in the nineteenth century. Elastic flows of population from Europe, and then across North America to the open frontier, ensured that labor markets were very fluid, and they have remained so ever since. Over two centuries, US regional real wage gaps have never exceeded 10–30 percent (Margo 1998; Rosenbloom 1996, 2010). Such levels of mobility and integration remain a distant prospect for most of Europe, given language and other barriers.

The US economy also has a central federal fiscal authority, implying national fiscal taxes and transfers that vary with the local business cycle and operate as intra-union automatic stabilizers. Such cross-border automatic stabilizers are absent in both the eurozone and the European Union, and there seems to be no appetite to create them. On the fiscal criterion, Panel D shows a wide range of measures of these federal fiscal stabilizer effects for the United States, but a recent estimate based on income tax alone shows an offset of 28 cents for a state-level $1 income loss, while among eurozone countries, the corresponding figure is effectively nil.

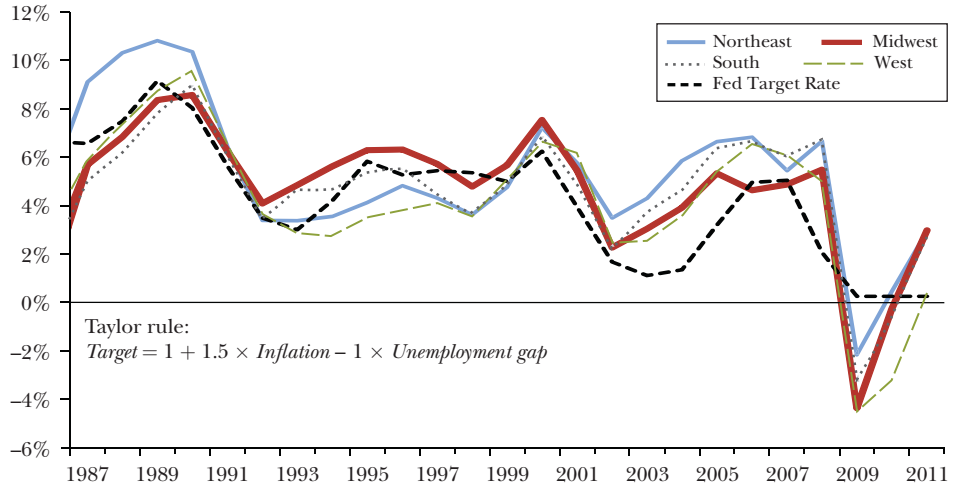
Figure 3 presents a different and arguably more direct take on the question of whether the US economy is better suited for a common currency than the eurozone. It plots estimates of optimal monetary policy responses based on rates of inflation and unemployment (Taylor 1993) for four US regions versus the whole country since 1987, and for the eurozone core and periphery versus the whole since 1999. The contrasts are striking. The estimated rule for the four US regions considered separately indicates small gaps between their "desired" policy rates and the national Fed target rate: divergences are usually between 0 and 200 basis points, a little more after the crisis. The "desired" policy-rate gaps between the eurozone and its core and periphery regions are much larger and more persistent. Prior to the crisis, the target interest rate for the periphery was consistently 300 basis points *above* the core; afterwards it was between 500 and 700 basis points *lower*. By this metric, a one-size-fits-all monetary policy appears more tenable in the United States than in the eurozone.

Banking Union

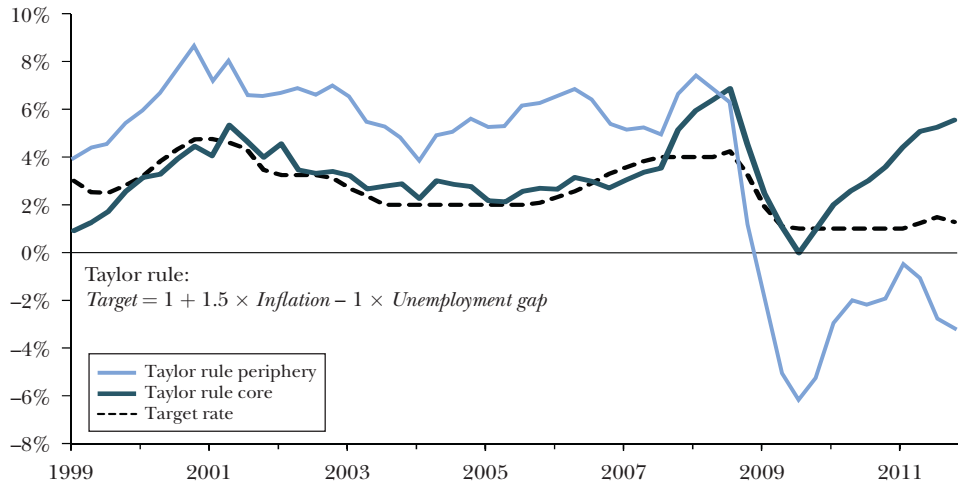
That said, optimal currency area theory neglects some of the most important issues facing the eurozone since the crisis exploded, in particular those having to do with banking and financial stability (Obstfeld 2013). The United States has since the 1930s had an effective national banking union, fully backstopped by a combination of the Federal Reserve and the US Treasury, and augmented by deposit insurance and other collective programs. As of now, the eurozone has virtually nothing in

Figure 3
Monetary Policy Taylor Rules: Eurozone versus the United States

A: Taylor rule by US census region



B: Taylor rule in the euro area: periphery versus core



Source: Malkin and Nechio (2012).
Notes: The two scales are commensurate (1800 basis points). The Taylor rule used is: *Target* = 1 + 1.5 × *Inflation* – 1 × *Unemployment gap*, where the last term is the difference between the measured unemployment rate and the natural rate (the unemployment rate where inflation neither decelerates nor accelerates).

place, except for ad hoc measures offered by the European Central Bank directly or, when collateral is weak, via emergency liquidity assistance through national central banks. Absent banking union, these are ultimately national burdens on taxpayers (as in Ireland) or depositors (as in Cyprus).

America had neither a banking union nor an economically meaningful fiscal union when it gained its independence in the late eighteenth century. However, the United States has maintained a single currency area since the beginning (apart from a brief time during the Civil War when three currencies were in use). How did American monetary union function in the absence of these institutions? And what did it take—and when, and how—to eventually deepen the US institutional architecture?

For a long period, until the US Civil War, bank notes in the United States—that is, promissory notes issued by banks that could be used to transfer funds or to make payments—did not uniformly trade at par with currency (Gorton 2012). Even after this, bank deposits did not always trade at par with currency: that is, if you deposited currency in a bank account, and later withdrew it, the amounts might not match. "Free banking" was mostly the norm, there was no monetary authority, and in this decentralized system only gold functioned as a fixed reference value for money, or as true (par) money itself.

Yet payments frictions were in many respects the least of America's problems. Throughout this time, the US economy suffered asymmetric shocks at the regional level which states felt they couldn't or didn't want to offset given fiscal orthodoxies, and which centralized monetary policy, such as it was, was designed to ignore (Rockoff 2003). How did the economy adjust? States and localities suffered, and defaulted if necessary, banks went under, and labor emigrated to more prosperous towns nearby or states far away.

Eventually the United States experienced a sequence of crises sufficiently intense to spur change. The first shock came at the time of the US Civil War (Gorton 2012). The need for union war finance spurred the National Banking Acts, creating a new standardized national currency, with these uniform notes backed by banks' holdings of US Treasury debt. The Acts also set up a Comptroller to regulate the new form of nationally chartered banks. The new structure placed a large quantity of US Treasury debt on bank balance sheets and not just as a wartime expedient; it remains there to this day as the US banking system's reference safe and liquid asset. Yet no central bank or lender of last resort appeared at this time, and pockets of "non-par" banking survived, especially in rural areas. Bank runs and crises remained, and recessions recurred frequently, but in a political-economic equilibrium where macroeconomic management was not expected to play a role.

Still, by the time, in the early twentieth century, that US banks had become large enough that they constituted systemic risks, they were holding US government securities as their safe and liquid assets, rather than state and local debt: Illinois banks do not hold much, if any, Illinois debt, for example. As a result, defaults by state and/or local governments did not entail a systemic threat to the financial system as a whole. By contrast, in today's eurozone, each "subsidiary sovereign"

nation's banks largely hold national debts of their own country, implying a national sovereign-bank doom loop (Goodhart forthcoming).

Despite the National Banking Acts and even with the emergence of banks' collective self-regulation by clearinghouses, the US economy was characterized in the late nineteenth and early twentieth centuries by increasingly frequent and serious financial crises (1893 and notably 1907) and by deep recessions and depressions (the 1880s and 1990s). During the Panic of 1907, only a large privately coordinated intervention led by J. P. Morgan restored calm. This raised the fear that without a true central bank, with lender of last resort capability backed by unlimited balance-sheet capacity, the system was becoming increasingly fragile, and unacceptably so. In response to this sense of rising systemic risk, the Federal Reserve was established in 1913, creating a full-fledged monetary union with monopoly note issue, par clearing for all member banks, and a national payments system.

However, in its early years the new Federal Reserve System suffered regional tensions, and proved ineffective in halting the 1930s banking panics. It was politically hobbled at first by a mindset suspicious of central banking. In a small but revealing example, the Fed's attempt to eliminate non-par banking was halted by Supreme Court action in the 1920s in the case of *American Bank v. Federal Reserve Bank* (262 U.S. 643 [1923]). Non-par banks clung on until the Monetary Control Act of 1980. In the depths of the Great Depression, some regional Federal Reserve banks threatened not to lend gold to one another through the interregional Gold Settlement Account (set up in 1915). Doubts as to whether Federal Reserve banks would lend to each other were decisively squashed by changes to the Federal Reserve Act in 1935. From then on, the subtly renamed Interdistrict Settlement Account cleared balances in infinitely suppliable fiat money via the System Open Market Account and mutual imbalances are periodically reset to zero reflecting a common pool approach. Today's eurozone equivalent of the Interdistrict Settlement Account is TARGET2, with balances that net out among the separate national central banks and the European Central Bank. These within-system balances represent a small fraction of banking system assets in both the United States (1.9 percent) and the eurozone (2.6 percent); but while they represent only 1.7 percent of US GDP, they represent a much larger 9.4 percent share of eurozone GDP (due to the eurozone's larger banking system). Crucially, since the 1930s there has been absolutely no concern that the US intersystem balances might be limited in size or be subject to settlement (exit/redenomination) risk.[4]

The US financial collapse of the 1930s was utterly devastating, consistent with a pattern of more common and destructive crises in modern, credit-fueled, highly financialized economies where levered balance sheets imply "financial acceleration"

[4] For more on the Interdistrict Settlement Account versus TARGET2, see Koning (2012), Bijlsma and Lukkezen (2012), and Cour-Thimann (2013). The last paper stresses that nothing like the exploding TARGET imbalances can occur in the United States because the Interdistrict Settlement Account is mutualized: it is zeroed out on a regular basis, effectively by treating the Fed's SOMA securities as a "common pool" of assets shared by the system and owned by the national political entity.

(Schularick and Taylor 2012; Jordà, Schularick, and Taylor 2012). In good times, banks, firms, and households feed off wealth effects, borrow, and drive up asset prices after positive shocks and create more wealth and leverage; but after negative shocks this process goes into reverse, producing a vicious circle of contraction (Bernanke 1983; Bernanke and Gertler 1989). The US policy response was to legislate a new prudential architecture, built around the Banking Act of 1935 and other measures. Central regulation and insurance edicts were promulgated; deposit insurance was instigated with concomitant supervision; the Glass–Steagall Act of 1933 separated commercial and investment banking; and the role of the Federal Reserve Board of Governors was upgraded, permitting them to impose uniform monetary policy in all regions without opposition. A century and a half after its founding, the United States finally had a strong central bank and lender of last resort, with substantial powers, especially if exigent circumstances should recur.

Fiscal Union

Just as banking union progressed gradually in the United States, so too did fiscal union. Initially the central government left states to themselves under a "no bailout" constitutional settlement brokered by Alexander Hamilton: the US central government enacted a once-and-for-all debt mutualization, assuming all state-level Revolutionary War debts, but then expected each of the states to stand on its own fiscally, observe near–budget balance, and, if need be, default (Sargent 2012).[5] These rules survive to the present, and many states have been through fiscal distress and even default. The ability to default provided a dimension of flexibility at times of crisis, while protecting federal taxpayers from moral hazard risk and bailout burdens entailing higher taxes and/or inflation on the collective. State-level debts are typically modest in size, and, as noted, that paper is largely kept off US banks' balance sheets. But such arrangements also implied a potential bias toward procyclical fiscal policies at the state-and-local level, a destabilizing feature witnessed again in today's Great Recession.

Under these constitutional arrangements, US fiscal union was weak or nonexistent in the nineteenth century. An important step was the Sixteenth Amendment to the US Constitution in 1913 allowing a federal income tax. After World War I and the Depression, federal expenditures and taxes grew consistently large enough to provide substantial and elastic fiscal transfers (unemployment insurance, agricultural support, and later Social Security) and steady components of spending (like defense spending) that were shared between states and so helped smooth out asymmetric shocks. The federal system also leads to longer-run transfers between states, reflecting persistent cross-state imbalances in incomes, defense activity, the location of retirees on Social Security and so on.

Thus, the US fiscal and banking union has gradually evolved an interesting and, so far, durable mix of hard long-run rules, such as the "no bailout" setup allowing state

[5] Some haircuts or debt "reprofiling" were involved under Hamilton's scheme; the net-present-value losses were severe for the continental debts, not so much for the states' debts (Hall and Sargent 2013).

default, and institutional innovations that have made the system more stable. Since the Civil War, the system has been insured from a state-level financial doom loop by the primary role of US Treasury securities as liquid bank assets; since the Great Depression, it has provided an even more elastic short-run policy regime, embedded in intra-union fiscal stabilizers and union-level banking sector backstop and oversight. In contrast, since the current eurozone crisis began, the authorities involved have been unable to decide on whether there should be bailouts or not, defaults or not, automatic stabilizers or not, or bank backstops and oversight or not. Indeed, at various times they have veered towards almost all of these positions.

With US history in mind, an optimist might argue that since the eurozone project is barely 10 years old, and the United States took perhaps 140 years to fully develop an appropriate institutional structure, we should be impressed rather than concerned by how far the eurozone has come already. The evolution of US monetary and fiscal institutions was a fitful and crisis-ridden process, from the fights (ongoing) over the role, if any, for central banking and for large fiscal transfers; the longstanding political obstacles to government deposit insurance, at the state and then federal levels; and the conflict over a hard monetary regime, only resolved by the gigantic disaster of the 1930s, but not before a series of disputes, notably the bimetallism and "cross of gold" arguments of the 1890s, had posed deep questions about the desirability of a deflation-prone and asymmetrically adjusting regime. There were times when the monetary regime, sometimes even the monetary union itself, were deeply unpopular and the subject of national or regional tensions.[6]

Recent developments in Europe, such as the 2012 decision in principle to move towards a banking union, offer some hope of eventual institutional reform. Unfortunately, more rapid change may be required today in the eurozone than was the case in nineteenth-century America. The nature of modern economies, and of politics in the independent democracies that comprise the eurozone, is such that Europe may not have the luxury of experimenting for 140 years before finding workable arrangements. Popular calls for public goods, social insurance, countercyclical macroeconomic policy, and financial stability cannot be brushed aside so easily as in the less-democratic era of the nineteenth-century classical gold standard.

The United States began with a secure political union from which exit (although tried once during the Civil War) is now unthinkable and this provided a stage on which economic and monetary union developments could be slowly constructed. The US national constitution embodied key assumptions about the existence and permanence of the national debt (a key collective safe asset), federal

[6] The Democratic Party Platform adopted at Chicago, July 9, 1896, was "Opposed to the Gold Standard": "Gold monometallism is a British policy, and its adoption has brought other nations into financial servitude to London. It is not only un-American, but anti-American . . ." (http://projects.vassar.edu/1896/chicagoplatform.html.) On political cleavages over gold in the Populist era, see Rockoff (1990). Calls to end, audit, or otherwise change the Federal Reserve, or to return to gold, echo today.

taxing power (ultimate central fiscal capacity), as well as the common currency and the commerce clause (truly free interstate trade). Onto this, after major crises, a banking union and an economically meaningful fiscal union were later grafted. In comparison, neither the eurozone nor the European Union comprise a political union; exit is conceivable from both and openly discussed; there is no central fiscal authority in either, nor any common debt, and there seems to be no appetite on the part of creditor nations to go down that route. Recognizing these limits means that what is desirable for the eurozone may not be feasible, a glum thought to which we will return in the conclusion.

Costs of Exit: Lessons from Past Break-ups

What if the eurozone ultimately fails? History can speak of past cases in which common currencies split up, although again the analogies are imperfect.

Austro-Hungarian Empire

Following the end of World War I in 1918, the Austro-Hungarian empire was rapidly divided into successor states: the Kingdom of Serbs, Croats and Slovenes (Yugoslavia after 1929); Czechoslovakia; Austria; Hungary; and Romania. Initially the monetary union based on the krone continued, with banknotes for the entire region being printed in Budapest and Vienna, but this arrangement proved unsustainable (Garber and Spencer 1994). The separate states decided to introduce their own currencies from early 1919 onwards.

In one way, this process proved straightforward: countries typically overstamped existing banknotes and converted bank deposits into the new currency at a prearranged parity, imposing levies or forced loans in the process as needed. However, Garber and Spencer (1994) note a feature of the process with obvious implications for any eurozone break-up: the fact that these measures were enacted at different times in different countries led to large flows of currency across borders, despite attempts by the authorities to block them, as people sought to move their currency holdings to wherever they thought they would be most valuable. These decisions were based partly on assumptions about where the conversion would take place at the most favorable rate and partly on assumptions about future rates of inflation; the old currency also tended to flow to where it remained legal tender the longest (in this instance, Hungary). The obvious implication is that any break-up of the eurozone would work best if it happened quickly and in a coordinated manner, with "temporary" capital controls being essential. Another implication is that any suggestion of a future break-up could prove extremely destabilizing as investors and households anticipate the capital gains and losses that it would imply (Eichengreen 2010).

The Austro-Hungarian example does not imply that hyperinflation is a necessary consequence of a currency break-up, as is sometimes suggested: the hyperinflation experienced in both Austria and Hungary reflected the inflationary financing of

large budget deficits that had helped precipitate the break-up of monetary union in the first place. Czechoslovakia put in place an institutional framework prohibiting such policies, and suffered deflation rather than inflation, as it attempted to rejoin the gold standard at the pre-war parity.

Argentine Currency Board

Argentina has exited currency board experiments three times, in 1914, 1929, and 2002. The most recent exit illustrates some problems that could occur in the event of a eurozone break-up. A recession was followed by a sudden stop in lending by foreign creditors in 2001, including (eventually) even the IMF; fiscal space was gone. In the endgame especially, Argentina had borrowed large amounts from local banks. When the government defaulted, the banks became insolvent as well, leading to a textbook "triple crisis": a banking crisis, a sovereign debt crisis, and a currency crash (della Paolera and Taylor 2003). In the aftermath came the problem of who bore the losses.

In 2002, Argentina declared that dollar loans would be repaid in pesos. This reduced the cost to the government of bailing out Argentina's banks, but also led to a plethora of costly legal disputes (Roubini and Setser 2004). These included disputes about "pesification" itself and about the asymmetric pesification values attached to different claims; these dragged on for many years, generating considerable uncertainty. The external default led to an eventual bond exchange where, although a majority of creditors accepted a write down, there remain minority "vulture" holdouts who are still fighting in New York courts a decade later. Capital controls (*corralito*) had to be imposed immediately to prevent arbitrage and have never been fully dismantled. Costly side effects thus have to be set against the benefits of devaluation and default to Argentina—costs which must likewise be weighed up by any states looking at a possible eurozone exit.

Conclusions

Europe's current depression drags on. The jury is still out on whether the eurozone can achieve the minimal collective institutions needed to sustain deep integration and macro-financial stability of the kind that the US economy can take for granted. Drawing on the lessons of history, what do we think these institutions might be?

The fact that the eurozone scores so poorly on optimal currency area grounds suggests a need for mechanisms allowing smoother and more symmetric adjustment between its members. Moves to enhance labor mobility, for example by improving pension or health insurance portability, can help—but, we suspect, only to a limited extent.[7] A stronger fiscal center as in the United States is desirable, but there seems

[7] Indeed, emigration reduces the tax bases of the countries concerned, making it more difficult to pay for the fixed costs of running states and the retirements of those left behind; and it makes it more difficult to pay back large debts, which is another argument in favor of debt restructuring in these countries (see O'Rourke, 2010; Coppola, 2013).

little prospect of this; thus, member states will have to engage in countercyclical fiscal policy, if at all, by themselves. For some countries, the size of their existing debts means that debt restructuring will be required for them to regain the ability to do this (Wyplosz 2012).

The difficulty of developing eurozone-wide automatic stabilizers should focus attention on the design and policies of the European Central Bank. Since asymmetric adjustment based on internal devaluation is so costly and ineffective, the European Central Bank should allow a higher rate of inflation for the eurozone as a whole at times of economic and financial stress to facilitate relative price adjustment. This could be embedded in various policy regimes, like the much-debated nominal GDP target or the "Evans rule" of the US Federal Reserve, which promises to keep interest rates low until certain unemployment targets are reached. A shift to such a regime need not be viewed as incompatible with the price stability mandate of the European Central Bank. If these kinds of changes are politically impossible, pessimism about the euro's survival becomes more justifiable.

The institutional architecture of the eurozone needs to be deepened if a recurrence of the present crisis is to be avoided. A banking union seems essential. This would involve common banking supervision, common resolution procedures, and common deposit insurance—and in consequence at least some elements of a fiscal union (Pisani-Ferry, Sapir, Véron, and Wolff 2012; Goyal, Brooks, Mahmood, Tressel, Dell'Ariccia, and Pazarbasioglu 2013). However, Europe's banking system will remain fragile as long as national banks hold national debt of the "subsidiary sovereign" (Goodhart forthcoming); in contrast, a "safe" eurozone asset would allow governments to default or restructure their debts without collateral damage to financial systems in their own countries and potentially, via contagion, across the entire eurozone (Brunnermeier et al. 2011). The creation of such a safe European asset is particularly important since the ability of national governments to default is also essential, being the logical corollary of the no-bailout clause which has worked well in the US context and which seems consistent with the requirements of national democracies. By contrast, the eurozone attempt to avoid fiscal free riding by legally constraining national governments has been an intrusive failure, mimicking the German "centralized-federal" approach, which has not even worked well within Germany (Mody 2013; Wyplosz 2012).

So where the eurozone needs to go in the long run, we argue, is towards a genuine banking union; a eurozone-wide safe bond to break the sovereign-bank doom loop; a central bank that is more flexible and willing to act as a true lender of last resort against such bonds and other assets as necessary; and a fiscal union at least sufficient to support the above. But the short-run problems facing countries in the periphery of Europe are now so great that politicians may never get a chance to solve these long-run problems because the eurozone may well have collapsed in the meantime. The history of the gold standard tells us that an asymmetric adjustment process involving internal devaluation in debtor countries, with no corresponding inflation in the core, is unlikely to be economically or politically sustainable. A more flexible and countercyclical macroeconomic policy mix, involving some combination

of looser monetary policy, a higher inflation rate, a weaker euro, debt restructuring, and fiscal stimulus by core governments (or some European Union–level institution such as the European Investment Bank) is currently needed in order to make the adjustment process less asymmetrical and lessen the risk of a eurozone collapse.

US experience suggests that major institutional reforms tend to follow major political and economic crises, such as the Civil War, the Panic of 1907, and the Great Depression. But these crises occurred within the context of a pre-existing state. It is one thing to develop deeper US federal institutions at times of crisis within what is already one country, but another thing to do so in a union of 17 independent states. There, a sufficiently major crisis may lead to countries deciding to abandon the euro project altogether, which is why the possibility of a eurozone break-up cannot be excluded. In 1978, Chancellor Helmut Schmidt, seeking to reassure a Bundesbank nervous that the proposed "European Monetary System" would not pose excessive demands on Germany, quoted a Latin legal phrase, "*Clausula rebus sic stantibus. . . . Ultra posse nemo obligatur.*" This roughly translates as "a fundamental change of circumstances could make a treaty inapplicable . . . no one [country] is obligated to do more than they can do."[8]

A number of countries across Europe may eventually ask themselves if fundamental circumstances have changed in a way that renders their previous commitment to the eurozone inapplicable. A eurozone break-up would involve the redenomination of assets and liabilities, and in all likelihood sovereign defaults in some cases as well. This would imply large cross-border redistributive effects with substantial official-sector claims in dispute. An even larger plethora of private contracts would be affected, involving not only eurozone banks and firms; the scope for legal chaos seems clear. If the eurozone is destined to break up, then speed and cooperation are essential if both destabilizing capital flows and years of costly litigation and uncertainty are to be avoided. Such a benign scenario may seem fanciful in the extreme, but if eurozone policymakers do not rapidly move towards a different macroeconomic policy mix, and at the very least a meaningful banking union, then Europe may ultimately find itself clutching at such straws.

■ *We are grateful to David Autor, Barry Eichengreen, Chang-Tai Hseih, Ulrike Malmendier, Maurice Obstfeld, and Timothy Taylor for their comments on a previous draft. O'Rourke has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no. 249546, and expresses his gratitude. The usual disclaimer applies.*

[8] "EMS: Bundesbank Council meeting with Chancellor Schmidt (assurances on operation of EMS) *[declassified 2008]*." Bundesbank Archives (N2/267). Translation from the Margaret Thatcher Foundation, http://www.margaretthatcher.org/document/111554. Schmidt noted that such considerations had already led Germany to unilaterally and without prior notification contravene international treaty law on the Bretton Woods intervention commitments and withdraw support for the US dollar in 1973, precipitating a crisis; the same issues would arise with Germany's decision to terminate support for the Exchange Rate Mechanism in 1992, precipitating another crisis.

References

Almunia, Miguel, Agustín Bénétrix, Barry Eichengreen, Kevin H. O'Rourke, and Gisela Rua. 2010. "From Great Depression to Great Credit Crisis: Similarities, Differences and Lessons." *Economic Policy* 25(62): 219–65.

Auerbach, Alan J. 2009. "Implementing the New Fiscal Policy Activism." *American Economic Review* 99(2): 543–49.

Baldwin, Richard E. 2006. "The Euro's Trade Effects." ECB Working Paper 594, European Central Bank.

Basu, Susanto, and Alan M. Taylor. 1999. "Business Cycles in International Historical Perspective." *Journal of Economic Perspectives* 13(2): 45–68.

Bernanke, Ben S. 1983. "Nonmonetary Effects of the Financial Crisis in the Propagation of the Great Depression." *American Economic Review* 73(3): 257–76.

Bernanke, Ben S., and Kevin Carey. 1996. "Nominal Wage Stickiness and Aggregate Supply in the Great Depression." *Quarterly Journal of Economics* 111(3): 853–83.

Bernanke, Ben, and Mark Gertler. 1989. "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review* 79(1): 14–31.

Bijlsma, Michiel, and Jasper Lukkezen. 2012. "Why is There No TARGET2 Debate in the US?" Brugel blog. http://www.bruegel.org/nc/blog/detail/article/729-why-is-there-no-target2-debate-in-the-us/.

Bloomfield, Arthur I. 1959. *Monetary Policy under the International Gold Standard: 1880–1914.* Federal Reserve Bank of New York.

Bordo, Michael D. 2004. "The United States as a Monetary Union and the Euro: A Historical Perspective." *Cato Journal* 24 (1–2): 163–70.

Brunnermeier, Markus, Luis Garicano, Philip R. Lane, Marco Pagano, Ricardo Reis, Tano Santos, Stijn van Nieuwerburgh, and Dimitri Vayanos. 2011. "European Safe Bonds (ESBies)." September 26. Available at: http://euro-nomics.com/2011/european-safe-bonds/.

Campa, José Manuel. 1990. "Exchange Rates and Economic Recovery in the 1930s: An Extension to Latin America." *Journal of Economic History* 50(3): 677–82.

Catão, Luis A. V., and Solomos N. Solomou. 2005. "Effective Exchange Rates and the Classical Gold Standard Adjustment." *American Economic Review* 95(4): 1259–75.

Chernyshoff, Natalia, David S. Jacks, and Alan M. Taylor. 2009. "Stuck on Gold: Real Exchange Rate Volatility and the Rise and Fall of the Gold Standard, 1875–1939." *Journal of International Economics* 77(2): 195–205.

Coppola, Frances. 2013. "The Creeping Desert." *Coppola Comment*, June 6, 2013. http://coppolacomment.blogspot.com/2013/06/the-creeping-desert.html.

Cour-Thimann, Philippine. 2013. "TARGET Balances and the Crisis in the Euro Area." CESifo Forum, vol. 14, Special Issue, April.

Darvas, Zsolt. 2012. "Compositional Effects on Productivity, Labour Cost, and Export Adjustments." Bruegel Policy Contribution, Issue 2012/11, June 22.

de Bromhead, Alan, Barry Eichengreen, and Kevin H. O'Rourke. 2013. "Political Extremism in the 1920s and 1930s: Do German Lessons Generalize?" *Journal of Economic History* 73(2): 371–406.

della Paolera, Gerardo, and Alan M. Taylor. 1999. "Economic Recovery from the Argentine Great Depression: Institutions, Expectations, and the Change of Macroeconomic Regime." *Journal of Economic History* 59(3): 567–99.

della Paolera, Gerardo, and Alan M. Taylor. 2003. "Gaucho Banking Redux." *Economía* 3(2): 1–34.

Eggertsson, Gauti B. 2008. "Great Expectations and the End of the Depression." *American Economic Review* 98(4): 1476–1516.

Eichengreen, Barry. 1992. *Golden Fetters: The Gold Standard and the Great Depression 1919–1939.* Oxford University Press.

Eichengreen, Barry. 1996. *Globalizing Capital: A History of the International Monetary System.* Princeton, N.J.: Princeton University Press.

Eichengreen, Barry. 2008. "Sui Generis EMU." NBER Working Paper 13740.

Eichengreen, Barry. 2010. "The Breakup of the Euro Area." In *Europe and the Euro* edited by Alberto Alesina and Francesco Giavazzi, pp. 11–51. University of Chicago Press.

Eichengreen, Barry, and Jeffrey Sachs. 1985. "Exchange Rates and Economic Recovery in the 1930s." *Journal of Economic History* 45(4): 925–46.

Eichengreen, Barry, and Peter Temin. 2010. "Fetters of Gold and Paper." *Oxford Review of Economic Policy* 26(3): 370–84.

Estevadeordal, Antoni, Brian Frantz, and Alan M. Taylor. 2003. "The Rise and Fall of World Trade, 1870–1939." *Quarterly Journal of Economics* 118(2): 359–407.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1(4): 337–57.

Flandreau, Marc, Jacques Le Cacheux, and Frédéric Zumer. 1998. "Stability without a Pact?

Lessons from the European Gold Standard, 1880–1914." *Economic Policy* 13(26): 115–62.

Ford, Alec G. 1962. *The Gold Standard, 1880–1914: Britain and Argentina*. Oxford: Clarendon Press.

Garber, Peter M., and Michael G. Spencer. 1994. "The Dissolution of the Austro-Hungarian Empire: Lessons for Currency Reform." Essays in International Finance no. 191, International Finance Section, Department of Economics, Princeton University.

Goodhart, Charles A. E. Forthcoming. "Global Macroeconomic and Financial Supervision: Where Next?" In *Globalization in an Age of Crisis: Multilateral Economic Cooperation in the Twenty-First Century,* edited by Robert C. Feenstra and Alan M. Taylor. University of Chicago Press.

Gorton, Gary B. 2012. *Misunderstanding Financial Crises: Why We Don't See Them Coming*. Oxford University Press.

Goyal, Rishi, Petya Koeva Brooks, Pradhan Mahmood, Thierry Tressel, Giovanni Dell'Ariccia, and Ceyla Pazarbasioglu. 2013. "A Banking Union for the Euro Area." IMF Staff Discussion Note SDN/13/01, February.

Hanes, Christopher. 1993. "The Development of Nominal Wage Rigidity in the Late 19th Century." *American Economic Review* 83(4): 732–56.

Hanes, Christopher. 2000. "Nominal Wage Rigidity and Industry Characteristics in the Downturns of 1893, 1929, and 1981." *American Economic Review* 90(5): 1432–46.

Hanes, Christopher, and John A. James. 2003. "Wage Adjustment under Low Inflation: Evidence from U.S. History." *American Economic Review* 93(4): 1414–24.

Hanke, Steve H., and Kurt Schuler. 1998. "Currency Boards and Free Banking." In *Money and the Nation State: The Financial Revolution and the World Monetary System,* edited by Kevin Dowd and Richard H. Timberlake, Jr., 403–21. New Brunswick, N.J., and London: Transaction Publishers.

Hatton, Timothy James. 1995. "A Model of U.K. Emigration, 1870–1913." *Review of Economics and Statistics* 77(3): 407–15.

Hatton, Timothy J., and Jeffrey G. Williamson. 1998. *The Age of Mass Migration: Causes and Economic Impact*. New York: Oxford University Press.

Henriksen, Ingrid, and Niels Kærgård. 1995. "The Scandinavian Currency Union 1875–1914." In *International Monetary Systems in Historical Perspective,* edited by Jaime Reis, pp. 91–112. London: Palgrave Macmillan.

HM Treasury. 2003. *The United States as a Monetary Union: EMU Study*. Norwich: TSO.

Honohan, Patrick. 1994. "Currency Board or Central Bank? Lessons from the Irish Pound's Link with Sterling, 1928–79." CEPR Discussion Paper 1040.

Hume, David. 1742. "Of the Balance of Trade." Part 2, Essay 5 of *Essays, Moral, Political, and Literary*. http://www.econlib.org/library/LFBooks/Hume/hmMPL28.html.

IMF. 2012. "The Good, the Bad, and the Ugly: 100 Years of Dealing with Public Debt Overhangs." Chap. 3 of *World Economic Outlook: October 2012*. International Monetary Fund.

Irwin, Douglas A. 2010. "Did France Cause the Great Depression?" NBER Working Papers 16350.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2012. "When Credit Bites Back: Leverage, Business Cycles, and Crises." NBER Working Paper 17621. Forthcoming in *Journal of Money, Credit, and Banking*.

Kenen, Peter. 1969. "The Theory of Optimum Currency Areas: An Eclectic View." In *Monetary Problems in the International Economy,* edited by Robert A. Mundell and Alexander K. Swoboda, pp. 41–60. University of Chicago Press.

Kindleberger, Charles P. 1973. *The World in Depression, 1929–1939*. Berkeley: University of California Press.

Koning, J. P. 2012. "The Idiot's Guide to the Federal Reserve Interdistrict Settlement Account." A blog, February, 5: http://jpkoning.blogspot.ca/2012/02/idiots-guide-to-federal-reserve.html.

Malkin, Israel, and Fernanda Nechio. 2012. "U.S. and Euro-Area Monetary Policy by Regions." *FRBSF Economic Letter*, Federal Reserve Bank of San Francisco, February 27.

Margo, Robert A. 1998. "Labor Market Integration before the Civil War." NBER Working Paper 6643.

Mitchener, Kris James, and Marc Weidenmier. 2008. "Trade and Empire." *Economic Journal* 118 (533): 1805–34.

Mody, Ashoka. 2013. "A Rest Stop for Europe." Project Syndicate, March 5.

Morgenstern, Oskar. 1959. *International Financial Transactions and Business Cycles*. Princeton University Press.

Mundell, Robert A. 1961. "A Theory of Optimum Currency Areas." *American Economic Review* 51(4): 657–65.

Obstfeld, Maurice. 2013. "Finance at Center Stage: Some Lessons of the Euro Crisis." http://elsa.berkeley.edu/~obstfeld/Revisiting%20EMU.pdf.

Obstfeld, Maurice, Jay C. Shambaugh, and Alan M. Taylor. 2004. "Monetary Sovereignty, Exchange Rates, and Capital Controls: The Trilemma in the Interwar Period." *IMF Staff Papers* 51 (Special Issue): 75–108.

Obstfeld, Maurice, Jay C. Shambaugh, and Alan M. Taylor. 2005. "The Trilemma in History: Tradeoffs among Exchange Rates, Monetary Policies, and Capital Mobility." *Review of Economics and Statistics* 87(3): 423–38.

Obstfeld, Maurice, and Alan M. Taylor. 2003. "Sovereign Risk, Credibility and the Gold Standard: 1870–1913 versus 1925–31." *Economic Journal* 113(487): 241–75.

Obstfeld, Maurice, and Alan M. Taylor. 2004. *Global Capital Markets: Integration, Crisis, and Growth*. Cambridge University Press.

O'Rourke, Kevin H. 2010. "Migration, the Limits of Internal Devaluation, and the Bailout." At "The Irish Economy" website, November 28. http://www.irisheconomy.ie/index.php/2010/11/28/migration-the-limits-of-internal-devaluation-and-the-bailout/.

Pisani-Ferry, Jean, André Sapir, Nicolas Véron, and Guntram B. Wolff. 2012. "What Kind of European Banking Union?" Bruegel Policy Contribution, Issue 2012/12, June 25.

Ponticelli, Jacopo, and Hans-Joachim Voth. 2011. "Austerity and Anarchy: Budget Cuts and Social Unrest in Europe, 1919–2009." CEPR Discussion Paper 8513.

Redish, Angela. 1993. "The Latin Monetary Union and the Emergence of the International Gold Standard." In *Monetary Regimes in Transition*, edited by Michael D. Bordo and Forrest Capie, 68–85. Cambridge University Press.

Rockoff, Hugh. 1990. "The 'Wizard of Oz' as a Monetary Allegory." *Journal of Political Economy* 98(4): 739–60.

Rockoff, Hugh. 2003. "How Long Did It Take the United States to Become an Optimal Currency Area?" In *Monetary Unions: Theory, History, Public Choice,* edited by Forrest H. Capie and Geoffrey E. Wood, 70–103. London: Routledge.

Romer, Christina D. 1992. "What Ended the Great Depression?" *Journal of Economic History* 52(4): 757–84.

Rosenbloom, Joshua L. 1996. "Was There a National Labor Market at the End of the Nineteenth Century? New Evidence on Earnings in Manufacturing." *Journal of Economic History* 56(3): 626–56.

Rosenbloom, Joshua L. 2010. "The History of American Labor Market Institutions and Outcomes." EH.net, http://eh.net/encyclopedia/article/Rosenbloom.LaborInst.

Roubini, Nouriel, and Brad Setser. 2004. *Bailouts or Bail-ins? Responding to Financial Crises in Emerging Economies*. Washington, DC: Institute for International Economics.

Santos Silva, J. M. C., and Silvana Tenreyro. 2010. "Currency Unions in Prospect and Retrospect." *Annual Review of Economics* vol. 2, pp. 51–74.

Sargent, Thomas J. 2012. "United States Then, Europe Now." *Journal of Political Economy* 120(1): 1–40.

Schularick, Moritz, and Alan M. Taylor. 2012. "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review* 102(2): 1029–61.

Shambaugh, Jay C. 2012. "The Euro's Three Crises." *Brookings Papers on Economic Activity* no 1, pp. 157–211.

Taylor, John B. 1993. "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy,* vol. 39, pp. 195–214.

Temin, Peter. 1989. *Lessons from the Great Depression*. Cambridge: MIT Press.

Temin, Peter, and Barrie A. Wigmore. 1990. "The End of One Big Deflation." *Explorations in Economic History* 27(4): 483–502.

Wyplosz, Charles. 2012. "Europe's Quest for Fiscal Discipline." http://www.wyplosz.eu/fichier/Quest%20for%20fiscal%20discipline%20Nov2012.pdf.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 193–212

Downward Nominal Wage Rigidity and the Case for Temporary Inflation in the Eurozone

Stephanie Schmitt-Grohé and Martín Uribe

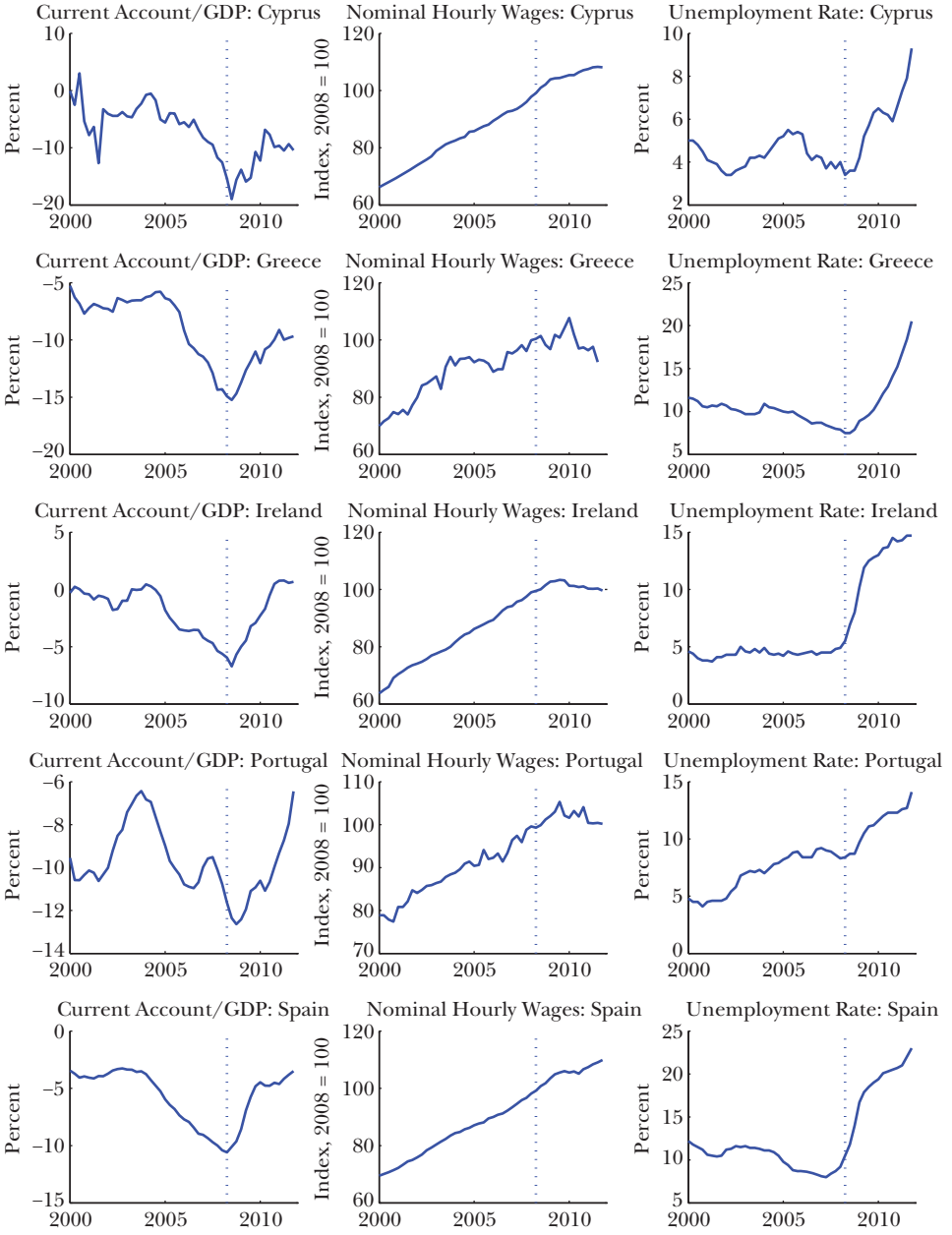
Since 2008, the periphery of Europe has been suffering an economic contraction of a magnitude that in several countries is comparable to the US Great Depression. During the early 2000s, the periphery of Europe enjoyed rapid growth in domestic demand, wages, and employment. Much of this bonanza was fueled by large international capital inflows. Figure 1 displays the current account, nominal hourly wages, and the rate of unemployment in five peripheral eurozone countries between 2000 and 2011. In all five countries, current accounts sharply deteriorated between 2000 and 2008. During this period, some countries increased their external debt position by more than 50 percent of GDP. This large amount of external borrowing financed a boom in domestic demand and was accompanied by increases in nominal wages of about 50 percent.

With the arrival of the international financial crisis of 2008, external credit to peripheral Europe suddenly dried up, causing a sharp contraction in aggregate demand. However, nominal hourly wages, shown in the second column of Figure 1, far from falling, remained largely unchanged from the high levels they had reached during the boom years. The combination of weak aggregate demand and high labor costs was associated with widespread unemployment, shown in the third column of Figure 1.

■ *Stephanie Schmitt-Grohé and Martín Uribe are Professors of Economics, Columbia University, New York City, New York. Schmitt-Grohé is a Research Affiliate, Centre for Economic Policy Research, London, United Kingdom, and both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are stephanie.schmittgrohe@columbia.edu and martin.uribe@columbia.edu.*

http://dx.doi.org/10.1257/jep.27.3.193 doi=10.1257/jep.27.3.193

Figure 1
Boom–Bust Cycle, Downward Wage Rigidity, and Unemployment in the Eurozone



Source: Eurostat.

The observed failure of nominal wages to adjust downward after 2008 despite sizable increases in unemployment suggests that downward nominal wage rigidity played an important role in the current unemployment crisis in the euro area. In Schmitt-Grohé and Uribe (2010, 2011a, 2011b, 2012), we develop a dynamic stochastic general equilibrium model of a small open economy with tradable and nontradable goods, downward nominal wage rigidity, and a fixed nominal exchange rate that captures these dynamics. The fundamental intuition behind that model is as follows. Suppose a member of the euro area suffers a negative external shock, such as an increase in the country risk premium, which causes aggregate demand to decline. The efficient adjustment to an external shock of this type is a decline in real wages. However, downward nominal wage rigidity and a fixed exchange rate together imply that real wages measured in terms of tradable goods are downwardly rigid. As a result, the labor market fails to clear and involuntary unemployment emerges. This narrative is well-known and goes back at least to Keynes (1925) and Friedman (1953).

If downward nominal wage rigidity is persistent, the current crisis in Europe is likely to be a protracted one, unless policymakers intervene. The affected eurozone countries, however, find themselves with limited room for national monetary and fiscal policy action. For instance, devaluation of the domestic currency would foster employment by reducing firms' real labor costs. But breaking away from the European monetary union is, at least thus far, off the table. Expansionary domestic fiscal policies, such as labor or sales subsidies, could be effective in remedying the distortions in the labor market. However, this policy option also faces obstacles insofar as expansionary fiscal policy is discouraged by international institutions, which make fiscal austerity a precondition for financial assistance. It is conceivable therefore that putting the periphery of Europe back on the path to recovery might at least in part require the involvement of supranational European institutions. The fact that the eurozone is a monetary but not a fiscal union rules out Europe-wide fiscal policy interventions.

In this article, we argue that a natural and practical remedy to the problems of the eurozone is monetary in nature. Specifically, a one-time rise in the overall price level in the euro area, a temporary period of inflation, would go a long way toward restoring full employment in the periphery of Europe. This monetary policy should be geared toward deflating the real value of wages in the periphery countries back to their pre-boom levels. We estimate that this policy could be implemented by raising the euro area annual rate of inflation to about 4 percent for the next five years.

However, this line of argument does not also hold that inflation should be higher in the long run. In fact, loose monetary policy should cease as soon as employment returns to normal levels. As such, our arguments here are not inconsistent with the long-run inflation target of the European Central Bank of slightly below 2 percent per annum. Therefore our argument is also different from that presented in related work suggesting that current inflation targets in developed countries are too low and should be permanently raised to deal with crises (Summers 1991; Blanchard, DellAriccia, and Mauro 2010).

Involuntary Unemployment Due to Wage Rigidity

The large capital inflows experienced by countries in the periphery of the eurozone during the boom years of 2000 to 2008 lead to sizeable increases in nominal hourly wages and the relative price of nontradables in terms of tradables. To capture these price dynamics, we adopt a framework in which the economy produces and consumes traded and nontraded goods. In this framework, large capital inflows raise the demand for nontraded goods. The supply of nontraded goods is relatively more inelastic than the supply of traded goods because nontraded goods cannot be imported and must be produced domestically. To satisfy the increased demand, producers of nontradables will increase production, given cost, only if the relative price of nontradables increases. Thus a model with traded and nontraded goods has the potential to explain why the relative price of nontradables appreciates when a country experiences large capital inflows.

The adjustment friction that motivates the policy interventions we suggest in this article is downward rigidity in nominal wages. Specifically, we model downward nominal wage rigidity as a lower bound on the growth rate of nominal wages such that nominal wages in a given time period do not decline, and might rise, from the previous period. We embed the assumption of downward nominal wage rigidity into a small open economy with a tradable and a nontradable sector. For a formal derivation of the underlying theoretical framework see Schmitt-Grohé and Uribe (2010). Here we focus on a graphical representation.

There is extensive empirical evidence on downward nominal wage rigidity. One branch of the empirical literature studies asymmetries in the distribution of observed hourly wage changes of individual workers. This literature documents that far more workers receive a wage increase than a wage cut. Importantly, almost half of all workers experience no changes in their hourly compensations. For examples, see Fortin (1996), Kuroda and Yamamoto (2003), Fehr and Goette (2005), Gottschalk (2005), Dickens et al. (2007), Barattieri, Basu, and Gottschalk (2012), Kaur (2012), and Daly, Hobijn, and Lucking (2012).

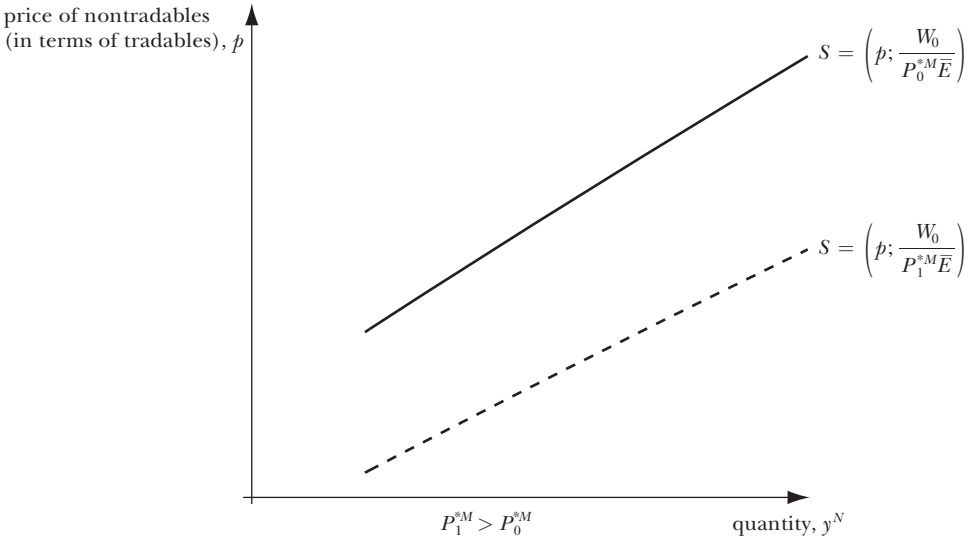
A second branch of the empirical literature provides evidence for downward nominal wage rigidity using aggregate data. For example, Holden and Wulfsberg (2008) document downward nominal wage rigidity using data for hourly nominal wages at the industry level in 19 OECD countries over the period 1973 to 1999. In Schmitt-Grohé and Uribe (2010), we provide empirical evidence for a number of emerging countries, including those in the periphery of Europe.

Figure 1 also suggests that nominal wages are downwardly rigid. Specifically, it shows that since the onset of the Great Recession in 2008Q2, even though unemployment rose sharply in the periphery of Europe, nominal wages in Cyprus, Ireland, Portugal, and Spain did not decline.

The Supply of Nontradables

Consider the supply schedule of nontradables, which is shown with a solid upward sloping line in Figure 2. The horizontal axis measures the quantity of

Figure 2
The Supply Schedule for Nontradables



Notes: On the *y*-axis, p is the relative price of nontradables in terms of tradables. On the *x*-axis, y^N is the quantity of nontradables produced. P_t^M is the nominal price of traded goods. W_t is the nominal wage rate. W_t/P_t^M is the real wage in terms of tradables. P_t^{*M} is the foreign currency price of traded goods. $P_t^M = P_t^{*M} E_t$, where E is the exchange rate.

nontradables produced, y^N. The vertical axis measures the relative price of nontradables in terms of tradables (or the real exchange rate), denoted p. The supply schedule is upward sloping because an increase in the relative price of nontradables induces firms to supply more nontraded goods.

The location of the supply schedule depends on the cost of production. If costs of production decrease, the supply schedule shifts down and to the right. This is because at lower costs, profit-maximizing firms are willing to supply more goods at the same price. Costs of production depend on labor cost, and labor costs in turn depend on the wage rate. The wage rate relevant for the supply schedule drawn in the space (y^N, p) are wages in terms of traded goods. Let P_t^M denote the nominal price of traded goods. Then the location of the supply schedule will depend on the real wage in terms of tradables, or W_t/P_t^M, where W_t denotes the nominal wage rate in period t.

We assume that the law of one price holds for traded goods. This means that the price of traded goods is the same in the home country and the foreign country when expressed in a common currency. Formally, when the law of one price holds, the domestic currency price of traded goods must satisfy $P_t^M = P_t^{*M} E_t$, where P_t^{*M} denotes the foreign-currency price of traded goods and E_t denotes the nominal exchange rate defined as the domestic-currency price of one unit of foreign currency. Throughout our analysis, we will assume that the exchange rate is credibly

and permanently fixed at $\bar{E}$, that is, $E_t = \bar{E}$ for all periods $t \geq 0$. This assumption captures the situation of a country in the periphery of the eurozone for whom breaking away from the euro is not an option.

To reflect the dependence of the supply schedule on labor costs, we express it as $y^N = S(p; W/(P^{*M}\bar{E}))$.[1] Holding constant the nominal wage, W, and the nominal exchange rate, $\bar{E}$, an increase in the international price of tradable goods, P^{*M}, lowers labor costs by lowering real wages in terms of traded goods. Graphically, an increase in the foreign price level from P_0^{*M} to $P_1^{*M} > P_0^{*M}$ results in a shift in the supply schedule down and to the right. This shift is shown with a dashed line in Figure 2. An increase in the foreign price level, holding constant nominal wages, lowers the real labor cost faced by domestic firms and provides an incentive for increased production. This is the key economic mechanism through which foreign inflation can have an expansionary effect in economies with downwardly rigid nominal wages and a fixed exchange rate.

The Demand for Nontradables

Nontraded goods are goods that are domestically produced and cannot be easily or cheaply shipped abroad. Therefore, the law of one price does not hold for nontraded goods. An example of a nontraded good is housing, and hence the construction sector is a sector that produces nontraded goods.

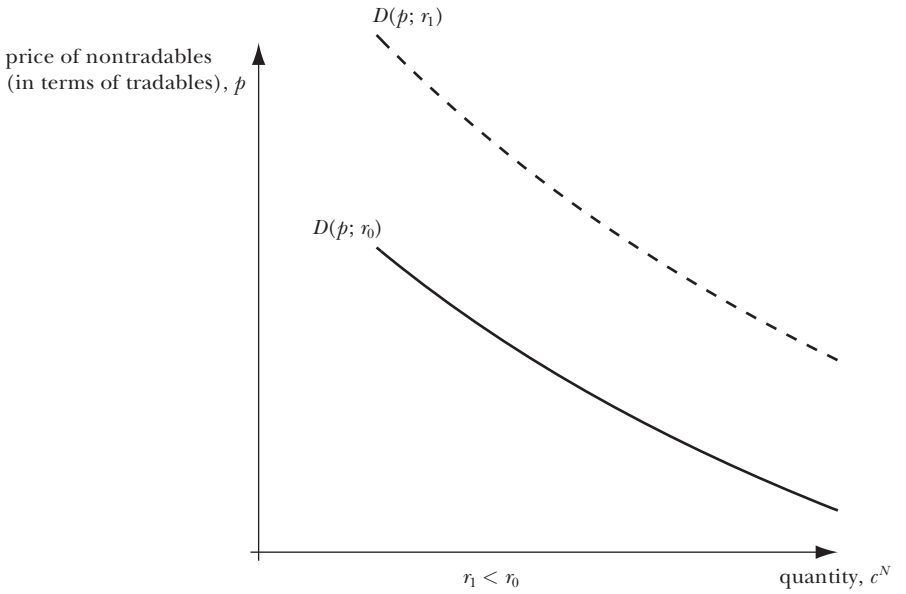
We denote the aggregate demand function for nontradables by $c_t^N = D(p_t; r_t, tot_t, \ldots)$, where c_t^N denotes consumption of nontradables, r_t denotes the country interest rate, and tot_t denotes the terms of trade.[2] The demand function is decreasing in the relative price of nontradables, p_t, because, all other things equal, as nontradables become relatively more expensive, households substitute tradables for nontradables in their consumption basket.

The shifters of the demand schedule are variables that affect either the intertemporal price of consumption, or households' wealth, or both. Among the variables that affect both the intertemporal price of consumption and wealth is the real interest rate

[1] The supply schedule can be derived as follows. Assume that the production function is $y^N = F(h)$, where h denotes hours, and $F(.)$ is an increasing and concave production function. Profits of firms are given by $P^N y^N - Wh$, where P^N is the nominal price of nontraded goods. Firms take prices and wages as given. Profits are maximized when the value of the marginal product of labor equals marginal costs, that is, when $P_t^N F'(h_t) = W_t$. Suppose the production function is of the form $F(h) = h^\alpha$, where $\alpha \in (0, 1)$ is a positive coefficient. Then the supply schedule takes the form $S(p; W/(P^{*M}\bar{E})) = (\alpha p/(W/(P^{*M}\bar{E}))^{\alpha/1-\alpha}$.

[2] Formally, one can derive the demand for nontradables as follows. Suppose households have preferences over an aggregate consumption good c_t composed of traded and nontraded consumption, c_t^T and c_t^N, respectively. Assume further that the technology for aggregating tradable and nontradable consumption goods into a composite consumption good (denoted c) is of the constant elasticity of substitution (CES) form $c = [a(c^T)^{1-1/\xi} + (1-a)(c^N)^{1-1/\xi}]^{1/(1-1/\xi)}$, where $a \in (0, 1)$ denotes the share of traded goods in total consumption, and ξ denotes the intratemporal elasticity of substitution between traded and nontraded goods. Utility-maximizing households choose consumption of traded and nontraded goods so as to make the marginal rate of substitution between traded and nontraded goods, which is given by $(1-a)/a(c^T/c^N)^{1/\xi}$, equal to the relative price, p. This implies that $c^N = c^T((1-a)/(ap))^\xi$. Consumption of traded goods c^T, in turn, is a function of the interest rate, r_t, the terms of trade, tot_t, and other measures of household wealth.

Figure 3
The Demand Schedule for Nontradables



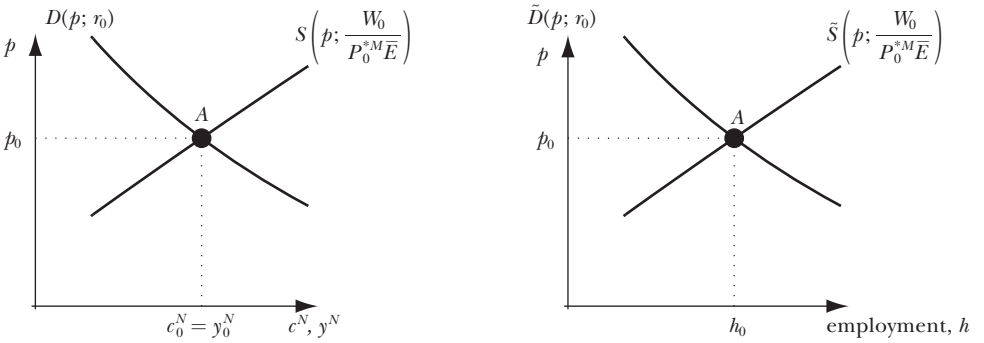
Notes: On the *y*-axis, p is relative price of nontradables in terms of tradables. On the *x*-axis, c^N is the consumption of nontradables. Demand for nontradables is a function of p, and the country interest rate, r.

that the rest of the world charges the country for borrowing in international bond markets. Country interest rate shocks have been shown to be an important driver of business cycles in emerging economies (Neumeyer and Perri 2005; Uribe and Yue 2006). Assuming that the country is a net debtor to the rest of the world, which is the case of empirical interest in the application considered here, an increase in the country interest rate r_t causes a negative wealth effect and a negative intertemporal substitution effect. Both of these effects tend to depress the demand for nontraded goods, shifting the demand schedule down and to the left.

Figure 3 displays the demand for nontradables in the space (c_t^N, p_t) for two values of the interest rate. The solid line shows the demand schedule for a high value of the interest rate, r_0, and the dashed line shows the demand schedule for a lower value of the interest rate $r_1 < r_0$. A decrease in the interest rate, from r_0 to r_1, would shift the demand schedule up and to the right. Conversely, an increase in the interest rate from r_1 back to the level of r_0 (the situation we were discussing above), shifts the demand schedule down and to the left (from the dashed line to the solid line).

Another shifter of the demand schedule is the terms of trade, denoted tot_t, and defined as the relative price of exportable goods, P_t^{*X}, in terms of importable goods,

Figure 4
Equilibrium in the Nontraded Goods Market and the Demand for Labor



Notes: p is relative price of nontradables in terms of tradables, c^N is the consumption of nontradables, y is the quantity of nontradables, and r is the country interest rate. W is the nominal wage rate, P^{*M} is the foreign currency price of traded goods, E is the exchange rate, and $W/P^{*M}E$ is the real wage. Output of nontradables is increasing in hours worked, so in the graph on the right along the x-axis, we replace quantity consumed and produced by employment in hours worked, h. Figure 4 represents a partial equilibrium analysis because it takes the nominal wage rate as given.

$tot_t \equiv P_t^{*X}/P_t^{*M}$. An improvement in the terms of trade (that is, an increase in tot_t) has a positive wealth effect that pushes the demand schedule up and to the right (not shown in Figure 3).

Equilibrium in the Nontraded Sector and the Demand for Labor

In equilibrium, the market for nontraded goods must clear. This means that the quantity supplied of nontradables must equal the quantity demanded of nontradables.[3] The left panel of Figure 4 indicates with point A the intersection of the demand schedule with the supply schedule. Given the real wage, $W/(P^{*M}\bar{E})$, and given the country interest rate, r_0, the intersection of the demand and supply schedule determines the value of the relative price of nontradables. At point A, the relative price is equal to p_0 and consumption and production of nontraded goods is equal to $c_0^N = y_0^N$. These values depend of the level of exogenous variables such as the country interest rate, the foreign price level, the exchange rate, and on endogenous variables, such as the nominal wage rate. Figure 4 represents a partial equilibrium analysis because it takes the nominal wage rate, W_0, as given.

The nominal wage rate is an endogenous variable that adjusts, to the extent possible, to clear the labor market. To understand how nominal wages are determined we need to introduce the supply and demand for labor, h, into our graphical

[3] Formally, the equilibrium value of the relative price of nontraded goods must solve $S(p_t; W_t/(P_t^{*M}\bar{E})) = D(p_t; r_t)$.

analysis. We can use the fact that demand and supply of nontraded goods must be equal to each other to derive the demand for labor in the nontraded sector.

Note that nontraded goods are produced with labor so that output of nontradables is increasing in hours worked in the nontraded sector. Thus, we can redraw the left panel of Figure 4 replacing output and consumption of nontradables by hours worked in the nontraded sector. The horizontal axis of the right panel of Figure 4 measures hours worked in the nontraded sector. Because the horizontal axis measures employment rather than output, we change the label of the supply curve from $S(p;\ W_0/(P_0^{*M}\bar{E}))$ to $\tilde{S}(p;\ W_0/(P_0^{*M}\bar{E}))$. Similarly, we change the label of the demand function from $D(p;\ r_0)$ to $\tilde{D}(p;\ r_0)$. The intersection of the demand and supply schedule can now be interpreted as determining the demand for labor given the real wage, $W_0/(P_0^{*M}\bar{E})$.

One can use the right panel of Figure 4 to find the demand for labor for different values of the country interest rate or the real wage. For example, an increase in the country interest rate would shift the demand schedule $\tilde{D}(p;\ r)$ down and to the left, so that the intersection of the demand and supply schedule occurs for a lower value of h. Therefore, an increase in the country interest rate lowers demand for labor in the nontraded sector given wages. Or consider an increase in the foreign price level, P^{*M}. Such an increase lowers the real wage and shifts the supply schedule $\tilde{S}(p;\ W/P^{*M}\bar{E})$ down and to the right. The new intersection of demand and supply would occur at a larger value of h, indicating that an increase in foreign inflation raises the demand for labor in the nontraded sector.

Involuntary Unemployment

We assume that workers supply inelastically $\bar{h}$ hours of work to the market.[4] Suppose that for a given nominal wage, the intersection of the demand and supply for nontradables occurs at a point that requires more labor than workers are willing to supply at that wage. In terms of Figure 4, this would correspond to a case in which $h_0 > \bar{h}$. In that case, we would expect wages to rise. That is, W_0 could not be the equilibrium wage rate. Alternatively, suppose that at point A in Figure 4, the amount of labor needed to produce the equilibrium level of nontradables is less than the amount workers would like to supply at that wage, that is, suppose that $h_0 < \bar{h}$. If wages were fully flexible, this excess supply of labor should result in a decline in the wage rate to a value below W_0. But if wages are downwardly rigid, then the excess supply of labor remains because wages cannot fall. The labor market is in disequilibrium because labor demand falls short of labor supply, and involuntary unemployment emerges.

[4] One can show that our analysis is robust to introducing elastic labor supply or production in the traded sector.

Boom Years Followed by External Crisis

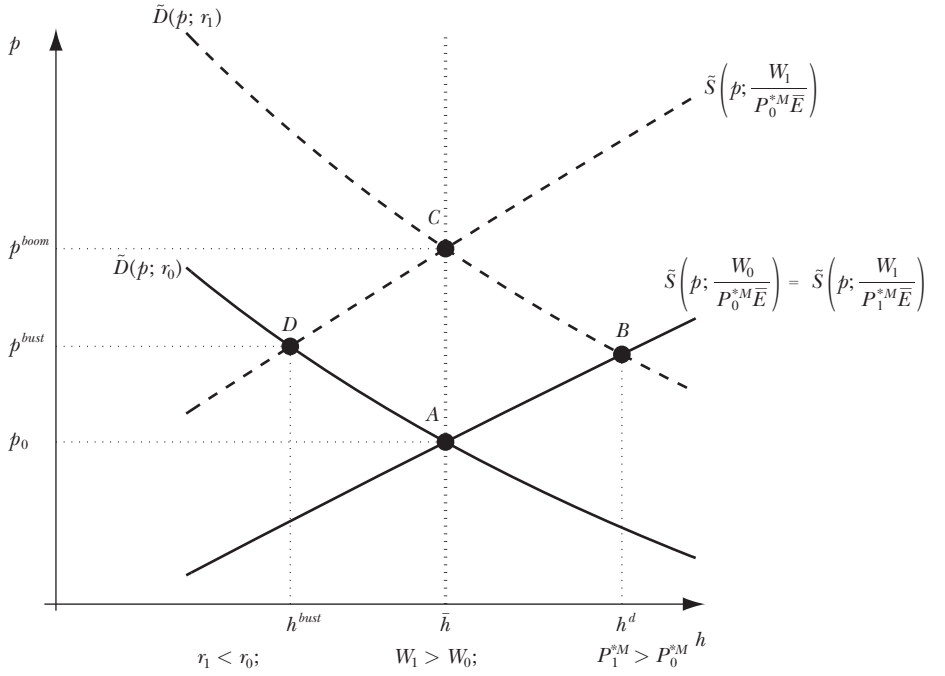
Having introduced the key elements of our framework, we can now use it to interpret the boom–bust episode that afflicted the periphery of Europe over the past decade. Suppose the economy starts at point A in Figure 5. The figure is drawn such that when the international price of tradables equals P_0^{*M}, the country interest rate equals r_0, and the nominal wage equals W_0, the economy enjoys full employment—that is, $h = \bar{h}$, and the equilibrium relative price of nontradables is p_0. Notice that now the figure also shows the labor supply schedule, which, because of our assumption of an inelastic labor supply, is represented by a vertical dashed line. Think of point A in Figure 5 as the location of peripheral Europe in 2000, the year the euro was adopted. With the adoption of the euro, borrowing rates dropped sharply for most peripheral countries. The decline in rates led to large capital inflows.

The decline in interest rates shifts the demand of nontradables up and to the right. Specifically, in Figure 5 we assume that interest rates fall from r_0 to $r_1 < r_0$. The demand schedule associated with r_0 is shown with a downward-sloping solid line and the demand schedule associated with r_1 is indicated with a dashed line. The new intersection of demand and supply for nontradables is at point B. However, at point B, labor demand, indicated with h^d in the figure, exceeds labor supply, $\bar{h}$, and thus point B cannot be supported as an equilibrium. Nominal wages begin rising to eliminate the excess demand for labor. The increase in the nominal wage shifts the supply schedule up and to the left, as shown by the dashed upward-sloping line in Figure 5. The new equilibrium is at point C, where wages are higher (and equal to W_1), relative prices are higher (and equal to p_{boom}), and the economy enjoys full employment. Think of point C as the state of the economy at the peak of the business cycle in 2008Q2. The data underlying Figure 1 indicate that nominal hourly wages in the five countries considered increased about 50 percent over the period 2000 to 2008.

In 2008, the Great Recession hits and borrowing conditions for countries in the periphery of Europe deteriorate rapidly. For simplicity, we represent this deterioration in borrowing conditions as interest rates rising back to $r_0 > r_1$. This rise in interest rates shifts the demand curve back down to its original position, indicated with a solid downward sloping line in Figure 5. However, the economy does not return to point A because nominal wages cannot fall. The economy therefore settles at point D, where labor demand is equal to h_{bust}, which is lower than labor supply, and involuntary unemployment in the amount $\bar{h} - h_{bust}$ emerges. If nominal wages were downwardly flexible, full employment would be restored by a decline in wages that shifts the supply schedule down and to the right. The resulting equilibrium would be at point A, where $h_t = \bar{h}$.

We note that the adjustment of the economy from point C to point D could be viewed as displaying insufficient depreciation of the real exchange rate: the equilibrium real exchange rate, given by p_{bust}, is too appreciated relative to the full employment real exchange rate, given by p_0. The real exchange rate gap shown on Figure 5 as $p_{bust} - p_0 > 0$ means that nontradables are too expensive relative to

Figure 5
Boom Followed by Crisis and Unemployment



Notes: Recall that p is the relative price of nontradables in terms of tradables, h is hours worked, W is the nominal wage rate, P^{*M} is the foreign currency price of traded goods, E is the exchange rate, and $W/P^{*M}E$ is the real wage. Assume that workers inelastically supply $\bar{h}$ hours of work to the market. Suppose the economy starts at point A. Think of this as the location of peripheral Europe in 2000, the year the euro was adopted. Next assume that interest rates fall from r_0 to $r_1 < r_0$. The demand schedule associated with r_1 is the dashed line. The new intersection of demand and supply for nontradables is point B, where labor demand, h^d, exceeds labor supply, $\bar{h}$. Nominal wages begin rising to eliminate the excess demand for labor, which shifts the supply schedule up and to the left. The new equilibrium is at point C, where wages are higher, relative prices are higher, and the economy again enjoys full employment. Think of point C as the state of the economy at the peak of the business cycle in 2008Q2. In 2008, the Great Recession hits and borrowing conditions for countries in the periphery of Europe deteriorate rapidly. For simplicity, we represent this as interest rates rising back to $r_0 > r_1$. This rise in interest rates shifts the demand curve down (to its original position), however, the economy does not return to point A because of downward nominal wage rigidity. Therefore, the economy settles at point D, and involuntary unemployment emerges.

tradables, preventing households from engaging in an expenditure switch away from tradables and toward nontradables. This expenditure switch is necessary to restore full employment in the nontraded sector.

The apparent downward rigidity in the real exchange rate could be interpreted as stemming from nominal rigidities in the price of nontradables. But in the present framework, nominal product prices are fully flexible. The reason why firms in the nontraded sector are reluctant to lower nominal prices further is that their nominal

labor cost, determined by the nominal wage, are too high. Lowering nominal prices would lead to losses that would force firms out of the market. It follows that nominal wage rigidity causes nominal product prices to behave as if they themselves were downwardly rigid.

Domestic Policy Options

The adjustment friction created by downward nominal wage rigidity can be overcome by a number of domestic policy interventions. One such domestic policy would be to devalue the currency—that is, to increase E_t. (Recall that in the above analysis E_t is credibly and permanently fixed.) By deflating the real labor cost faced by firms, a devaluation causes a shift in the supply schedule down and to the right. A sufficiently large devaluation would result in an equilibrium with full employment, indicated by point *A* in Figure 5.

In a theoretical model in Schmitt-Grohé and Uribe (2010), we show that the optimal exchange rate policy consists in large devaluations during external crisis. We also show that the optimal exchange rate policy can restore full employment in all periods and increases welfare, relative to a peg, by more than 5 percent of consumption per period. This numerical value of the welfare benefits is based on a number of modeling assumptions, and thus might not be representative. In the calibration we use, the large welfare benefits arise because the optimal exchange rate policy would lower the unemployment rate in crisis between 5 and 10 percentage points.

Of course, no country in the periphery of the eurozone has (yet) chosen to abandon the euro and devalue its currency. However, other countries that suffered from large adverse external shocks and had pegged their currency have responded to the adverse shock by devaluing. One example is Argentina in 2001. After pegging the peso at a one-to-one parity with the US dollar for more than a decade, Argentina devalued by 250 percent in December of 2001: that is, the peso price of one US dollar went from 1 Argentine peso to 3.5 Argentine pesos. The devaluation followed three years of adverse external shocks that had caused large increases in involuntary unemployment. The Argentine subemployment' rate, which measures the fraction of the workforce that is either unemployed or working part-time involuntarily, increased from 26 percent to 39 percent between 1998 and 2002. As in the recent experience of peripheral Europe, this increase in unemployment did not result in a downward adjustment in nominal wages. On the contrary, nominal hourly compensation costs in manufacturing increased from 7.87 pesos in 1998 to 8.14 pesos in 2001. During this period Argentina experienced deflation of 0.86 percent per year, so real wages actually rose at a time when unemployment increased sharply. When Argentina finally devalued at the end of 2001, real wages declined, falling 65 percent in real terms in 2002. Unemployment also fell rather swiftly and by 2005 subemployment stood again at 25 percent, 15 percentage points below the pre-devaluation peak.

Another historical example of the effectiveness of devaluations in reducing real wages and unemployment has been documented by Eichengreen and Sachs (1985) for the Great Depression. These authors show that countries that left gold early enjoyed more rapid recoveries than countries that stayed on gold. This difference in performance was associated with earlier reflation of price levels in the countries leaving gold. Specifically, Eichengreen and Sachs compare the change in real wages and industrial production from 1929 to 1935 observed in the sterling bloc countries, consisting of the United Kingdom, Sweden, Finland, Norway, and Denmark, and the gold bloc countries, which are France, Belgium, the Netherlands, and Italy. The sterling bloc countries left gold beginning in 1931, and the gold block countries stayed on gold much longer, some until 1935. Eichengreen and Sachs show that, relative to their respective 1929 levels, real wages in the sterling bloc countries were lower than real wages in the gold bloc countries. And industrial production in the sterling bloc countries in 1935 exceeded their respective 1929 levels, whereas industrial production in the gold bloc countries was below their respective 1929 levels.

Certain domestic fiscal policies can mimic the effect of a devaluation of the nominal exchange rate. These policies can, for instance, take the form of a wage subsidy or of a sales subsidy in the nontraded sector. The key characteristic of such policies is that they lower the effective cost of production of firms operating in the nontraded sector. Graphically, the effect of these policies is to shift the supply schedule in Figure 5 down and to the right. If the subsidies are large enough, they can bring the economy from point *D* to point *A* and restore full employment. In Schmitt-Grohé and Uribe (2010, 2012), we discuss this avenue in more detail. As in the case of devaluations, countries in the periphery of the eurozone have not adopted expansionary fiscal policies of the type described here. On the contrary, the observed changes in the fiscal policy stance over the past four years have been toward fiscal austerity. For example, the change in real government spending during the Great Recession in peripheral Europe was negative, whereas in the previous three recessions, government spending rose—by about 25 percent (for discussion, see Box 1.1 of the April 2013 World Economic Outlook).

In the absence of domestic policy intervention, the current situation in the periphery of the euro area is described by point *D*, where unemployment is high and persistent.

Eurozone-wide Policy Options

Given the difficulties that policymakers in the periphery of the eurozone face in adopting domestic policies that stimulate employment, a natural question is whether the required stimulus could originate from supranational European policy institutions. On the fiscal side, the eurozone is not a fiscal union. The revenues and spending of the European Union are quite small relative to the size of the EU economy, and they are not set up in a way to create an effective supranational tax or transfer system.

However, on the monetary policy side, there exists a well-defined institutional framework, embodied in the European Central Bank, to implement eurozone-wide monetary policy. The central bank of the monetary union could ameliorate the unemployment problem created by the combination of downward nominal wage rigidity and a fixed exchange rate. The required policy takes the form of a temporarily higher rate of union-wide inflation. In terms of the notation used here, the central bank of the monetary union should induce a temporary increase in the general price level, P_t^{*M}.

Figure 5 shows how this policy would affect employment in the peripheral countries. The economy starts at point D in which the labor market suffers involuntary unemployment, the nominal wage rate is W_1 and the general price level is P_0^{*M}. Suppose now the central bank creates inflation and this lifts the price level from P_0^{*M} to $P_1^{*M} > P_0^{*M}$. Graphically, the increase in P_t^{*M} shifts the supply schedule (shown with a dashed line) down and to the right, back to the upward sloping solid line in Figure 5.[5] If the increase in the general price level P_t^{*M} is sufficiently large, the equilibrium will be at point A, where the economy displays full employment.

The present analysis suggests that emerging countries of the eurozone could benefit significantly from a temporary increase in union-wide inflation. But what about countries in which the lower bound on nominal wages was not binding during the crisis—arguably Germany, Austria, and the Benelux countries? In these countries, the increase in the general price level P_t^{*M} would result in an increase in nominal wages without affecting employment or production. Graphically, starting from point A in Figure 5 and assuming that for these countries the borrowing conditions did not change significantly (so their interest rate does not change), the increase in the general price level P_t^{*M} shifts the supply schedule down and to the right, and an increase in nominal wages W_t shifts it back to its original position, so the economy never leaves point A. The increase in wages occurs automatically because at the original wage level W_0, the increase in the general price level P^{*M} causes an excess demand for labor. Because wages are assumed to be fully flexible upwardly, the excess demand in the labor market is eliminated spontaneously by an increase in nominal wages.

An implication is that the required level of price increases in the monetary union as a whole is larger than the one that will take place in the peripheral countries in which the lower bound on wages is binding. The reason is that adjustment in the latter countries requires a real depreciation, that is, a fall in the relative price of nontradables, p_t. To see this effect in graphical terms, refer to Figure 5. In the crisis equilibrium, point D, the relative price of tradables p_t equals p_{bust}, which is higher than the full employment equilibrium relative price p_0 associated with point A. Since $p_t = P_t^N/(P_t^{*M}\overline{E})$, in order for p_t to fall, the nominal price of nontradables P_t^N must increase proportionally less than the general increase in the price level P_t^{*M}. The situation is different for the members of the monetary union in which the lower bound on

[5] The demand schedule is unaffected by the increase in foreign prices. In particular, because both P_t^{*M} and P_t^{*X} are assumed to increase in the same proportion, the terms of trade, tot_t, are unchanged.

wages is not binding. There, the real exchange rate is at its full employment level and therefore does not need to change. Consequently, the general increase in the price level P_t^{*M} should be met with an increase in the nominal price of nontradables of the same proportion. It follows that the increases in the overall price level are proportionally larger in countries in which the nominal wage rigidity is not binding.

How Much Union-wide Inflation Is Needed?

How large is the eurozone-wide price increase necessary to eliminate the unemployment problem in the periphery? We present two alternative ways of calculating the size of the required inflationary stimulus. One uses wage data and the other employment data.

A Wage-based Calculation

We assume that real wages observed in the periphery of Europe in 2000Q1 were at trend full employment values. The average cumulative increase in nominal hourly wages observed in the five peripheral countries listed in Figure 1 over the period 2000Q1 to 2011Q2 was 48.7 percent. We proxy foreign inflation by the Consumer Price Index inflation rate in Germany. Over the period 2000Q1–2011Q2, the German price index rose by 20.6 percent.

To take into account by how much full employment real wages rose, we need a measure of total factor productivity growth. It turns out that total factor productivity in the euro area was roughly constant in the past decade. Barkbu, Rahman, and Valdés (2012, figure 2) compute total factor productivity growth for each decade since the 1960s. For the 2000s, they report, for the euro area as a whole, annual total factor productivity growth of less than 0.2 percent. The EU KLEMS Growth and Productivity Accounts project, at http://www.euklems.net, offers valued-added estimates of total factor productivity growth for Spain and Ireland. It reports that by this measure, productivity in Spain fell by 4 percent between 2000 and 2007 and productivity in Ireland fell by 1 percent over the same period. Therefore, we assume that productivity stayed constant, and we only adjust for foreign inflation but not for productivity growth.

It then follows that in 2011Q2 real wages in the five peripheral European countries were on average 23.3 percent above their full employment levels. (This figure results from dividing 1.487 by 1.206; these numbers come from the German price index changes, mentioned above.) Therefore, the implied increase in the eurozone price level necessary to restore full employment is 23.3 percent.[6] If the required price

[6] A further assumption implicit in this calculation is that the full employment real exchange rate (that is, the full employment value of p_t), is the same in 2011Q2 as it was in 2000Q1. This assumption is conservative. It is reasonable to assume that in order to service the increases in external debt observed between 2000 and 2011, the full employment real exchange rate would have to depreciate relative to its 2000 level.

increase were to be implemented in a period of five years, the annual rate of inflation in the eurozone would have to be raised to 4.3 percent. This figure is more than twice as large as the current area-wide inflation target of 2 percent.

An Employment-based Calculation

An alternative way to determine how much euro-area-wide inflation is needed to restore full employment in the periphery of the eurozone is to use data on employment. This methodology does not use data on wages or foreign inflation.

Suppose that current employment in the nontraded sector is 10 percent below full employment. This figure is in line with the unemployment increases observed in Greece, Ireland, and Spain since 2008, but higher than those seen in Portugal and Cyprus. To find by how much real wages exceed full employment real wages, we will use the fact that firms set employment so as to equate the value of the marginal product of labor to the marginal cost of labor, which is the real wage. We can calculate by how much the value of the marginal product of labor falls if employment were to increase back to full employment, that is, if employment were to increase by 10 percent. Then using the fact that in equilibrium the value of the marginal product of labor has to equal the real wage, we can obtain an estimate of the required decline in real wages.

The value of the marginal product of labor in the nontraded sector is the product of the relative price of nontradables and the physical marginal product of labor. Assuming that the (physical) marginal product of labor is proportional to labor productivity, which is true for a large class of commonly used production functions, and assuming a labor share of income of 75 percent, an increase in employment by 10 percent lowers labor productivity by $-(0.75 - 1) \times 10 = 2.5$ percent. That is, a 10 percent increase in employment lowers the physical marginal product of labor by 2.5 percent. If the relative price was unchanged, we would conclude that real wages would have to fall by 2.5 percent to restore full employment. However, the economy will only consume (absorb) more nontraded goods if the relative price of nontraded goods falls. Thus we need to take into consideration the effect of a 10 percent increase in employment on the relative price of nontradables. First note that the percentage increase in nontraded output is equal to the labor share times the percentage increase in labor. With a labor share of 75 percent, a 10 percent increase in employment then increases output of nontradables by 7.5 percent. To find the change in the relative price of nontradables that would induce agents to absorb 7.5 percent more nontraded goods, we can use estimates on the intratemporal elasticity of substitution between traded and nontraded goods.

Existing empirical estimates (for examples, see the survey by Akinci 2011) indicate that the intratemporal elasticity between traded and nontraded consumption is about 0.4. This means that if the relative price of nontradables falls by 1 percent, then the relative consumption of nontraded goods increases by 0.4 percent. From here it follows that if output and hence consumption of nontraded goods increases by 7.5 percent, then the relative price of nontraded goods must decline by $1/0.4 \times 7.5 = 18.75$ percent. It follows that an increase in employment of

10 percent will lower the value of the marginal product of labor in the nontraded sector by 18.75 plus 2.5 percent, which is 21.25 percent. In sum, the employment-based calculation indicates that eliminating 10 percentage points of unemployment necessitates a decrease in the real wage of 21.25 percent.

Because the lower bound on nominal wages is binding and the exchange rate is fixed, all of the decrease in the real wage must come from an increase in the general price level P_t^{*M}. That is, the monetary authority of the Euro area must engineer an increase in the overall price level of 21.25 percent, or about 4 percent per year for five years. This level of inflation is essentially the same as the one obtained in the previous subsection using a methodology that relies on observations of wages, total factor productivity growth, and foreign inflation.

Discussion and Conclusion

Clearly, downward nominal wage rigidity is an adjustment friction that is less likely to have negative real effects—that is, it is less likely to be associated with inefficiently high real wages, inefficiently high real exchange rates, and inefficiently low employment—in recessions in which inflation remains strong. Specifically, if inflation is sufficient to make the full employment real wage attainable without having to cut nominal wages, then downward nominal wage rigidity should not matter. Thus, an important aspect of the current recession in Europe is that it occurs in the context of low inflation targets of the monetary authority. The monetary policy objective of the European central bank is to keep inflation below but close to 2 percent per year. Moreover, the monetary authority has been undershooting its inflation target slightly over the course of the recession. In April 2013, year-over-year euro-area inflation, as measured by the harmonized index of consumer prices, was 1.2 percent, 0.8 percentage points below target.

Downward nominal wage rigidity is, of course, not the only possible explanation for the observation that nominal wages have failed to decline in peripheral Europe. For example, some explain the observation that aggregate wages have failed to decline by a composition effect, according to which during a recession low-productivity workers lose their jobs more frequently than do high-productivity workers. Under this hypothesis, average wages do not decline because the workforce has a larger share of high-wage workers. But even so, we would still expect to see wage cuts for those workers (whether low- or high-productivity) who do keep their jobs; however, micro evidence on wage changes of individual workers shows that nominal wage cuts are rarely observed in low-inflation recessions.

For example, Fortin's (1996) study of downward nominal wage rigidity chronicles price, wage, and unemployment during the great Canadian slump of the 1990s, when inflation was low, around 1.4 percent. Fortin shows that the distribution of nominal wage changes during this time had a large mass at zero and otherwise looks like the right tail of a normal distribution. Similarly, Daly, Hobijn, and Lucking's (2012) study of downward nominal wage rigidity in the United States documents

wage changes during the Great Recession of 2008, a period characterized by low inflation (about 1.7 percent on average) and high unemployment. Daly et al. show that the distribution of nominal wage changes of individuals in 2011 has a large spike at zero, and much fewer wage cuts than wage increases. Likewise, the study of Fehr and Goette (2005) is motivated by the great slump that affected Switzerland in the 1990s in which nominal wage cuts were almost never observed in a sample of 35,000 wage changes taken from firms and average inflation was below 1 percent per year. Finally, perhaps the most well-known, recent example of a recession coupled with low inflation is that of Japan in the 1990s. During that decade the Japanese unemployment rate more than doubled, but Kuroda and Yamamoto (2003) document downward nominal wage rigidity in Japan over the period 1993–1998, a time when nominal price inflation was not only low but negative. Taken together, this empirical evidence seems consistent with the assumption that in low-inflation recessions, observed nominal wages do not decline because of downward nominal wage rigidity.

In this essay, we have investigated the consequences of downward nominal wage rigidity for the adjustment of employment in the periphery of Europe to the large negative external shock caused by the 2008 global Great Recession. We show that in an environment of low inflation, downward nominal wage rigidity can prevent a decline in real wages. Absent an adjustment in real wages, the relative price of nontradables cannot fall—that is, the real exchange rate cannot depreciate—causing involuntary unemployment in the economy. We discuss a number of alternative policy options to address the distortions created by downward nominal wage rigidity in a currency union. We argue that, while there are several options, a particularly simple and practical solution lies in increasing inflation in the currency union as a whole. Specifically, our findings suggest that a policy of 4 percent inflation for five years would go a long way towards bringing unemployment down to pre-crisis levels.

Our proposed temporary inflation policy is not of the beggar-thy-neighbor type. In our framework, increasing euro-area-wide inflation is beneficial even if it has no effect on the terms of trade, that is, even if it has no effect on the relative price of exported and imported goods, and even if it has no effect on the trade balance of the country whose employment will benefit from euro-area-wide inflation. Thus our argument is not one of increasing the competitiveness of the periphery relative to the core of the currency union. In this sense, our argument is related to that put forth by Eichengreen and Sachs (1985) for how recovery from the Great Depression could have been hastened. These authors argue that during the Great Depression of the 1930s, devaluations taken by a group of countries were mutually beneficial to those countries and had those policies been more widely adopted and internationally coordinated, they would have sped up the recovery from the Great Depression. Furthermore, the effectiveness of our proposed policy does not hinge on the euro depreciating relative to other currencies, because the success of the policy depends on changing the domestic price of nontraded goods in terms of traded goods and not on changing the terms of trade of Europe relative to the rest of the world.

■ *We would like to thank Pablo Ottonello for outstanding research assistance, Nils Gottfries, Ed Nelson, and the editors of this journal for very helpful comments.*

References

Akinci, Ozge. 2011. "A Note on the Estimation of the Atemporal Elasticity of Substitution between Tradable and Nontradable Goods." Unpublished paper, Columbia University, February 2.

Barkbu, Bergljot, Jesmin Rahman, and Rodrigo Valdés. 2012. "Fostering Growth in Europe Now." With a staff team. IMF Staff Discussion Note, SDN/12/07, June 18.

Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk. 2010. "Some Evidence on the Importance of Sticky Wages." NBER Working Paper 16130, June.

Blanchard, Olivier, Giovanni DellAriccia, and Paolo Mauro. 2010 "Rethinking Macroeconomic Policy." IMF Staff Position Note 2010-03, February 12.

Daly, Mary, Bart Hobijn, and Brian Lucking. 2012. "Why Has Wage Growth Stayed Strong?" FRBSF Economic Letter 2012-10, Federal Reserve Bank of San Francisco, April 2.

Dickens, William T., Lorenz Goette, Erica L. Groshen, Steinar Holden, Julian Messina, Mark E. Schweitzer, Jarkko Turunen, and Melanie E. Ward. 2007. "How Wages Change: Micro Evidence from the International Wage Flexibility Project." *Journal of Economic Perspectives* 21(2): 195–214.

Eichengreen, Barry, and Jeffrey Sachs. 1985. "Exchange Rates and Economic Recovery in the 1930s." *Journal of Economic History* 45(4): 925–46.

Fehr, Ernst, and Lorenz Goette. 2005. "Robustness and Real Consequences of Nominal Wage Rigidity." *Journal of Monetary Economics* 52(4): 779–804.

Fortin, Pierre. 1996. "The Great Canadian Slump." *Canadian Journal of Economics* 29(4): 761–87.

Friedman, Milton. 1953. "The Case For Flexible Exchange Rates." In *Essays in Positive Economics*, edited by Milton Friedman, 157–203. University of Chicago Press.

Gottschalk, Peter. 2005. "Downward Nominal-Wage Flexibility: Real or Measurement Error?" *Review of Economic and Statistics* 87(3): 556–68.

Holden, Steinar, and Fredrik Wulfsberg. 2008. "Downward Nominal Wage Rigidity in the OECD." *B. E. Journal of Macroeconomics: Advances* 8(1): Article 15.

Kuroda, Sachiko, and Isamu Yamamoto. 2003. "Are Japanese Nominal Wages Downwardly Rigid? (Part I): Examinations of Nominal Wage Change Distributions." *Monetary and Economic Studies*, Bank of Japan, 21(2): 1–29.

Kaur, Supreet. 2012. "Nominal Wage Rigidity in Village Labor Markets." Unpublished manuscript, Columbia University, January 15.

Keynes, John Maynard. 1925 [1972]. "The Economic Consequences of Mr. Churchill." In The *Collected Writings of John Maynard Keynes*, Vol. IX, edited by Donald Moggridge, 207–230. New York: St. Martin's Press.

Neumeyer, Pablo A., and Fabrizio Perri. 2005. "Business Cycles in Emerging Markets: The Role of Interest Rates." *Journal of Monetary Economics* 52(2): 345–80.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2010. "Pegs and Pain." Unpublished manuscript, Columbia University.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2011a. "Pegs, Downward Wage Rigidity, and Unemployment: The Role of Financial Structure." Unpublished manuscript, Columbia University, November 2011a.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2011b."Prudential Policy for Peggers." Unpublished manuscript, Columbia University, December.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2012. "Managing Currency Pegs." *American Economic Review* 102(3): 192–97.

Summers, Lawrence. 1991. "How Should Long-term Monetary Policy Be Determined?" *Journal of Money, Credit and Banking* 23(3): 625–31.

Uribe, Martín, and Z. Vivian Yue. 2006. "Country Spreads and Emerging Countries: Who Drives Whom?" *Journal of International Economics* 69(1): 6–36.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 213–228

Retrospectives
John Maynard Keynes, Investment Innovator

David Chambers and Elroy Dimson

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please write to Joseph Persky of the University of Illinois at Chicago at jpersky@uic.edu.

Introduction

When John Maynard Keynes managed the endowment of King's College at Cambridge University, the actively managed part of his portfolio beat the performance of the British common stock index by an average of 8 percentage points per year from 1921 to 1946. Little wonder that such modern investment giants as Warren Buffett (2013), George Soros (1987, 2011), and David Swensen (2001, 2005) have invoked Keynes in support of their investment beliefs and strategies. However, Keynes' actual investment strategy has been largely unexplored. An almost-complete record of Keynes' stock trading has until now remained dormant in the King's College Archives (Cox 1995), and we utilize this rich resource to reconstruct Keynes' investment decision making.

■ *David Chambers is Tenured Lecturer in Finance, Cambridge Judge Business School, United Kingdom, and (at the time of writing) Thomas McCraw Fellow, Harvard Business School, Cambridge, Massachusetts. Elroy Dimson is Emeritus Professor of Finance, London Business School, and Visiting Professor, Cambridge Judge Business School, United Kingdom. Their email addresses are d.chambers@jbs.cam.ac.uk and edimson@london.edu.*

http://dx.doi.org/10.1257/jep.27.3.213 doi=10.1257/jep.27.3.213

Keynes was an investment innovator. He traded currencies at the very inception of modern forward markets (Accominotti and Chambers 2013), commodity futures (Fantacci, Marcuzzo, and Sanfilippo 2010), and stocks—which are the focus of this paper. Most importantly, Keynes was among the first institutional managers to allocate the majority of his portfolio to the then-alternative asset class of equities. In contrast, most British (and American) long-term institutional investors of a century ago regarded ordinary shares or common stocks as unacceptably risky and shunned this asset class in favor of fixed income and real estate.

In addition, Keynes independently championed value investing in the United Kingdom at around the same time as Benjamin Graham was doing so in the United States. Both Keynes' public statements and his economic theorizing (Keynes 1936, chap. 12) strongly suggest that he did not believe that "prices of securities must be good indicators of value" (Fama 1976). Beginning as a top-down portfolio manager, seeking to time his allocation to stocks, bonds, and cash according to macroeconomic indicators, he evolved into a bottom-up investor from the early 1930s onwards, picking stocks trading at a discount to their "intrinsic value"—terminology he himself employed. Subsequently, his equity investments began to outperform the market on a consistent basis.

The result of such innovative thinking can be observed in his unconventional portfolios with pronounced tilts towards smaller companies and stocks with a high-dividend yield. His early confidence in an equity risk premium, and his development of small-stock and value-driven investment strategies, anticipated approaches employed by some of the better performing institutional investors in modern times (Lewellen 2011).

The King's College Endowment: 1921–1946

John Maynard Keynes invested on his own account as well on behalf of several institutions (Pierce 1993), among which King's College, Cambridge, was closest to his heart. In 1911, he had been elected to the Estates Committee of his college, and in 1912 to the Council, its governing body. He was appointed Second Bursar just after World War I and had primary responsibility for investments from 1921. In 1924, he became First Bursar, the senior financial administrator of his college, and from that point he had full and apparently unchallenged discretion over investment policy until his death in 1946. Indeed, his annual "Chancellor of the Exchequer" speech became a not-to-be-missed fixture in the College calendar.

Oxford and Cambridge colleges are perhaps the ultimate long-horizon investors and King's, founded in 1441, was by no means the oldest. Traditionally, their assets were largely invested in real estate (Dunbabin 1975; Acharya and Dimson 2007), and the bursar collected rents from a predominantly agricultural portfolio, managed the expenditures, and drew up the college books (Neild 2008, p. 100). This pattern was reinforced by Trustee Acts from the mid-nineteenth century intended to ensure that trust funds were managed conservatively. The Acts together with college

statutes severely restricted the freedom of a bursar to undertake financial investments, requiring that they be primarily in such high-quality fixed income securities as UK and colonial government securities, UK railway securities, water company securities, and local authority housing bonds and mortgages (*Stock Exchange Official Intelligence* 1926, pp. 1922–23).

However, in 1921 Keynes persuaded his College Fellows to permit a part of the endowment to be excluded from these restrictions. Accordingly, we divide the King's endowment into two composite funds which we call the "Restricted Portfolio," subject to the Trustee Acts, and the "Discretionary Portfolio," comprising those funds where Keynes had full discretion. While Keynes was responsible for both portfolios, our focus will be on the Discretionary Portfolio, since it offers the purest expression of his views. The Discretionary Portfolio initially comprised a fund known as the "Chest" and then from September 1933 also included Fund B, a pooled vehicle for a myriad of small endowed funds which had previously been managed on a segregated basis. The two accounts were managed in a similar style. Although attention has tended to concentrate on the Chest, we look at Keynes' trading record for both these discretionary accounts.

The Discretionary Portfolio grew through a combination of performance and cash inflows from 8 percent of the £285,000 in total securities held by the College in 1921 to 68 percent of £1,222,000 in 1946. To put these valuations in a contemporary context, £285,000 invested in the UK equity market in 1921 would, by the start of 2013, have appreciated to £54 million, and £1,222,000 invested in the UK equity market in 1946 would, by the start of 2013, have appreciated to £141 million. Both estimates assume that the investment would have been in a capitalization-weighted index of the largest 100 companies, as published in Dimson, Marsh, and Staunton (2013), and that all income would have been disbursed.[1] As our starting point, we revisit Keynes' performance by taking information from his annual investment review of the College endowment, the *Report to Inspectors of Accounts*, which he prepared from 1922 until his death (although the report for 1926 is missing) for the investment committee, known as the Estates Committee. In the *Reports*, Keynes reviewed separately each of the four main accounts that made up the Restricted Portfolio and the two accounts making up the Discretionary Portfolio. For each, he provided separate lists of year-end holdings at market values as well as annual capital appreciation and income figures. This reporting format therefore suggests that each account was managed on a segregated basis.

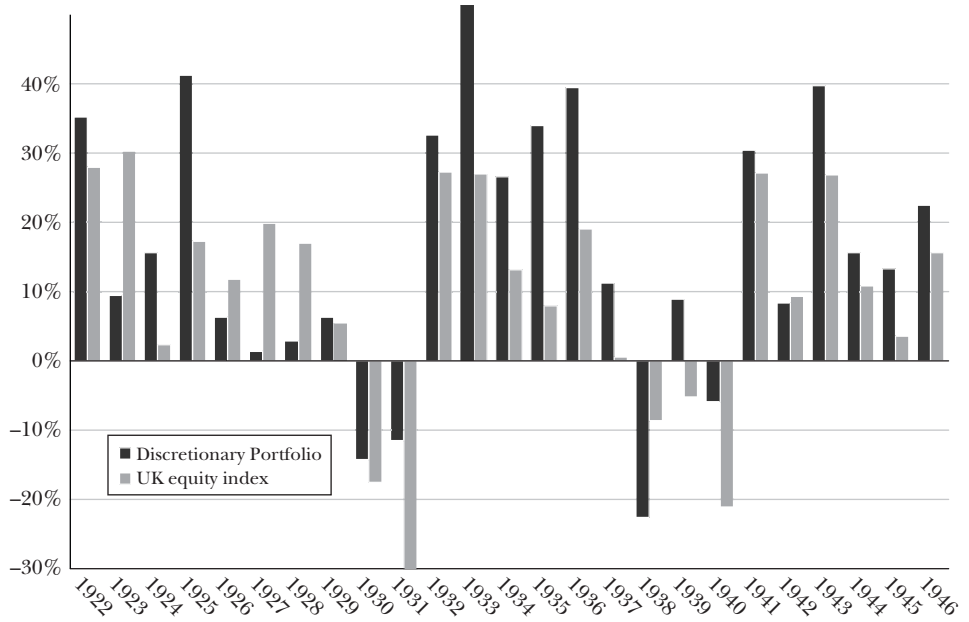
From the *Reports,* we have calculated the returns achieved by the Discretionary Portfolio (taking The Chest and Fund B together[2]), the Restricted Portfolio, and the Total Fund excluding real estate.[3] The returns are based on the estimates made

[1] We refer throughout to this index of the 100 largest UK equities, which is estimated by Dimson, Marsh, and Staunton, as the DMS Index.

[2] The Discretionary Portfolio returns also include a third fund, Fund C, established in 1933, which on average represented less than 1 percent of the total market value of the assets we analyze.

[3] Both the *Reports* and the College Accounts exclude any valuation of real estate holdings in the endowment throughout this period.

Figure 1
King's College Investment Returns 1922–1946



Sources: Chambers, Dimson, and Foo (2013). The total returns of the Discretionary Portfolio for the financial years ended August 1922 to August 1946 are estimated from the annual *Report to Inspectors of Accounts*, from the King's College Archives for financial years ended August. The UK equity index is based on the equally-weighted Dimson, Marsh, and Staunton (2002) total return index.
Note: Figure 1, plots the total return (including dividends) for the Discretionary Portfolio, which was managed by Keynes, and for the UK equity index.

by Keynes of the appreciation or depreciation for each year as a percentage of the start-year market value. To this capital gain or loss is added the income return for the year, which is the reported investment income divided by the average of the beginning and end portfolio values. As a general rule, all endowment income was spent by King's College rather than retained in the endowment.

Our estimates are illustrated in Figure 1, which plots the total return (including dividends) for the Discretionary Portfolio and for the UK equity index. According to these estimates, over the whole period, from end August 1922 to end August 1946, the annual performance of the Discretionary Portfolio averaged +16.0 percent, as compared to +10.4 percent for the market index. In contrast, the annual arithmetic mean total return of the Restricted Portfolio was only +6.8 percent, which did not compare favorably with total returns from UK government bonds of +7.1 percent.

The year-by-year results also show that Keynes underperformed in only six out of the 25 financial years and that four of those years occurred in the first eight years of his management of the Discretionary Portfolio. By August 1929, he was lagging the UK equity market by a cumulative 17.2 percent since inception. In addition, he

failed to foresee the sharp fall in the market the following month. We discuss the significance of this period of underperformance below, arguing that it was the catalyst for the fundamental change in his investment approach sometime in the early 1930s.

To what extent is this higher return in the Discretionary Portfolio across the whole period generated by assuming higher risk? One measure of risk, known as "tracking error," is the standard deviation of returns measured relative to a stock market index. Keynes' tracking error, measured relative to our UK equity index, is 13.9 percent, a high level of portfolio risk.[4] The high tracking error of the Discretionary Portfolio was attributable to Keynes' focus on stock selection, which we discuss in the following section.

Keynes' risk-taking was rewarded with superior investment performance. The excess return on a portfolio is defined as the return on the portfolio minus the risk-free rate of interest. The reward-to-risk ratio is the excess return divided by the standard deviation of portfolio returns. This reward-to-risk ratio, usually termed the Sharpe Ratio, was 0.73 for the Discretionary Portfolio, which compares favorably with the Sharpe Ratio for the Restricted Portfolio (0.57) and the UK equity market (0.49).[5]

Keynes' Stock Portfolio Characteristics

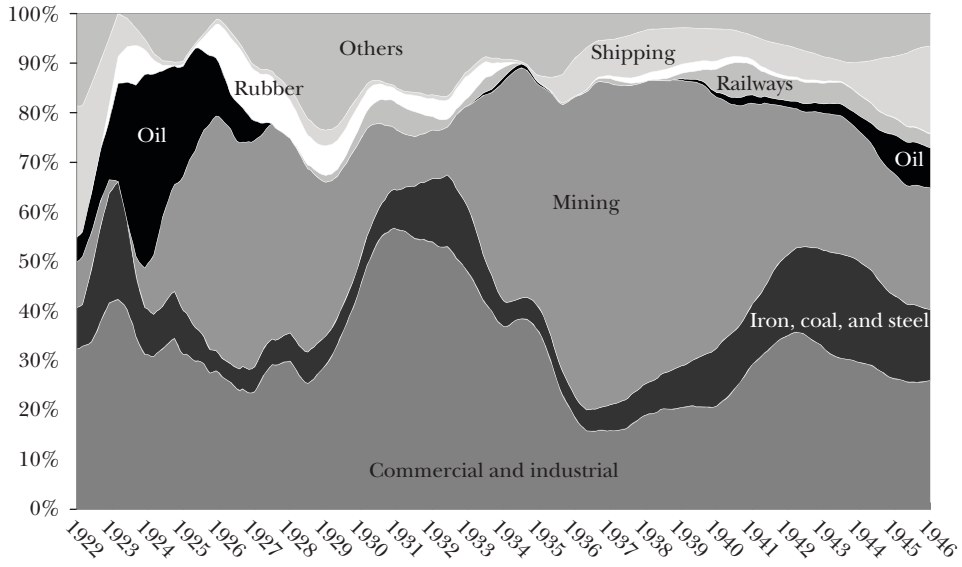
We describe the Discretionary Portfolios on four important dimensions. First, how diversified were Keynes' portfolios? Second, how "active" were his sector weightings, or, in other words, to what extent did his sector weightings differ from those of the market? Third, measured in terms of stock market capitalization, was he more inclined to invest in small or large firms? Last, were his portfolios more tilted to high-dividend or low-dividend yield stocks? The characteristics data in this section are estimated from UK security prices collected from the *Stock Exchange Daily Official List* and capital and dividend histories from the *Stock Exchange Daily Official List* and the *Stock Exchange Official Yearbooks*.

One measure of portfolio diversification is the proportion of the market value of the Discretionary Portfolio's UK equity securities allocated to the largest five (C5) or 20 (C20) shareholdings. The C20 measure started at 100 percent in the early 1920s when the total portfolio value was relatively modest, and although it declined, it remained at 80 percent in the 1940s as the portfolio value expanded. The C5 annual averages rose from 46 percent in 1921–29 to 49 percent in 1930–39, and then fell back to 33 percent in 1940–46. Keynes ran quite concentrated portfolios,

[4] For comparison, in the post-2000 period, the time-series tracking error for US university endowment funds averages only 3.4 percent (Brown, Dimmock, Kang, and Weisbenner forthcoming), and even for individual mutual funds it averages less than 6 percent (Cremers and Petajisto 2009).

[5] The only prior statistics on Keynes' investment performance were reported by Donald Moggridge who edited *The Collected Writings of John Maynard Keynes* (1983) and analyzed by Chua and Woodward (1983), who suggested that Keynes was a very extreme performer. As we explain in Chambers, Dimson, and Foo (2013), the Chua–Woodward study suffered from a number of data limitations, and the performance of Keynes' transactions was more nuanced than was previously believed.

Figure 2
UK Discretionary Portfolio Sector Weights: 1922–1946



Notes: Weights within the UK equity portfolio are estimated over rolling twelve-month periods from August 1922 to August 1946. Sector definitions follow the London Stock Exchange classification. Others include Breweries, Electric, Lighting & Power; Financial, Trusts, Land, etc.; Insurance; Investment Trusts; Rubber; Telegraphs & Telephones.

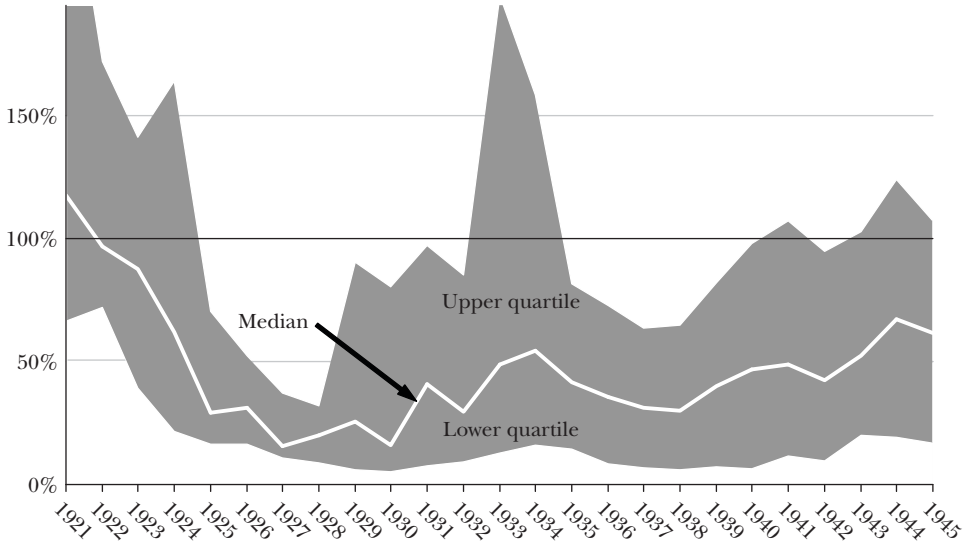
but there is no evidence that concentration measured in this way rose over time. We return to this point in the next section.

In looking at sectors, the majority of his UK equity holdings were concentrated in just two sectors, metal mining—tin mining stocks in the 1920s and gold mining stocks in the following decade—and commercial and industrial firms, as shown in Figure 2. While the latter was the largest sector represented in the stock market, representing at least 40 percent of the market index, mining only accounted for between 5 and 10 percent of the market index (Chambers, Dimson, and Foo 2013). Keynes' large overweighting of mining relative to the market was similar in magnitude to his underweighting of the second-largest sector, banking. Banking carried an index weight of 20 percent, and Keynes had little or no exposure in this sector.

As a director, Keynes contributed to the detailed statistical analysis of the *London and Cambridge Economic Service* which concluded in the mid-1920s that a combination of price-inelastic supply and strong underlying demand would cause the tin price to appreciate sharply (Moggridge, *The Collected Writings of John Maynard Keynes* XII, pp. 373–378, 416–421).[6] This analysis underpinned his decision to move into

[6] We refer throughout to *The Collected Writings of John Maynard Keynes* (1983, edited by Donald Moggridge) as *Collected Writings*.

Figure 3
Size Distribution of UK Discretionary Portfolio Holdings: 1921–1945



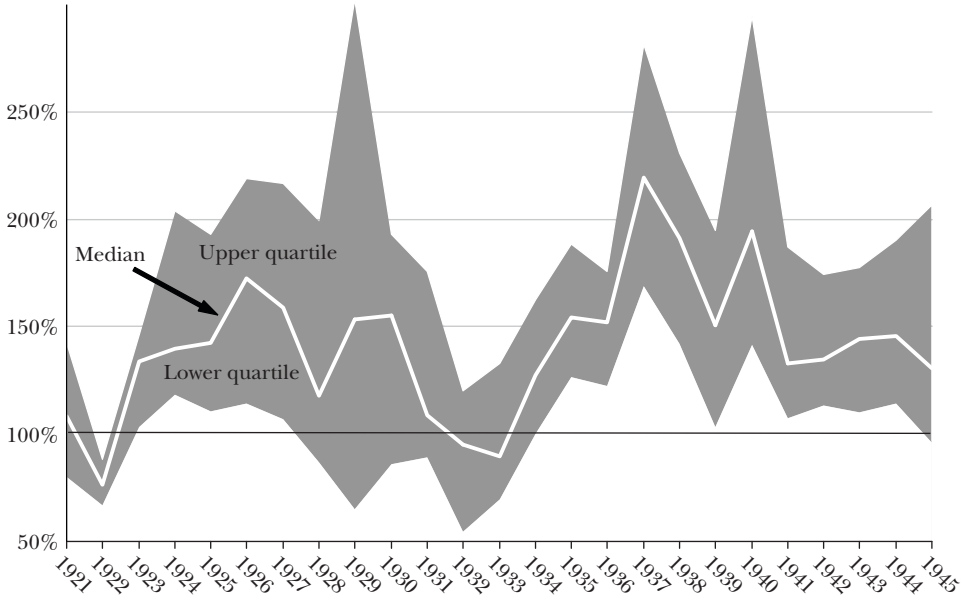
Notes: Each calendar year-end from 1921 to 1945, the size of each company in the Discretionary Portfolio is defined as its stock market capitalization expressed as a percentage of the market capitalization of the smallest firm in the DMS 100-share index. We plot this measure of relative size for the 25th percentile, median, and 75th percentile of firms in the Discretionary Portfolio. The inter-quartile range is shaded gray, and the median size of an equity holding is shown as a white line. The horizontal black line represents the size of the smallest constituent of the index. In most years, three-quarters of holdings are in companies that are smaller than the smallest firm in the index.

tin stocks. In the 1930s, he accumulated South African gold mining shares based upon his early realization that the devaluation of the South African currency in 1933 would have a favorable impact on their earnings (*Collected Writings* XXI, pp. 225–29). Keynes' substantial weighting in commercial and industrial stocks began in the early and mid-1920s with a diversified portfolio of industrial names. However, soon thereafter he concentrated his exposure in this sector on the two leading British automobile stocks, Austin Motors and Leyland Motors. In the context of the time, these would have been viewed as "technology" stocks.

In terms of firm size, Keynes had a decided tilt towards mid-cap and small-cap stocks. Compared to the ordinary market capitalization of the top 100 firms listed on the UK market at that time, the majority of his UK stocks were smaller firms. For each stock held in the Discretionary Portfolio at each calendar year-end from 1921 to 1945, we estimated its relative size, defined as its equity market capitalization expressed as a percentage of the market capitalization of the smallest firm in the DMS 100-share index.

Figure 3 plots this measure of relative size for the 25th percentile, median, and 75th percentile of firms in the Discretionary Portfolio. The interquartile range is shaded gray, and the median size of an equity holding is shown as a white line; the

Figure 4
Yield Distribution of UK Discretionary Portfolio Holdings: 1921–1945



Notes: Each calendar year-end from 1921 to 1945, the dividend yield of each dividend-paying company in the Discretionary Portfolio is expressed as a percentage of the dividend yield of the DMS 100-share index. We plot this measure of relative dividend yield for the 25th percentile, median, and 75th percentile of firms in the Discretionary Portfolio. The inter-quartile range is shaded gray, and the median dividend yield of an equity holding is shown as a white line. The horizontal black line represents the dividend yield of the index. In most years, three-quarters of holdings are in companies that have a dividend yield higher than the index.

horizontal black line represents the size of the smallest index constituent. It can be seen that, in most years, three-quarters of holdings are in companies that are too small to enter the list of the top 100 firms by market capitalization.

Turning to dividend yield, Keynes had a preference for high-dividend yield firms. For each dividend-paying stock held in the Discretionary Portfolio at each calendar year-end from 1921 to 1945, we express its dividend yield as a percentage of the DMS 100-share index. In Figure 4, we plot this measure of relative dividend yield for the 25th percentile, median, and 75th percentile of firms in the Discretionary Portfolio. The interquartile range is shaded gray, and the median dividend yield of an equity holding is shown as a white line; the horizontal black line represents the dividend yield of the index. In most years, three-quarters of holdings were in companies with a dividend yield higher than the index of the largest 100 UK firms. The weighted average dividend yield of the Discretionary Portfolio (not reported) dips slightly below the dividend yield on the index only twice (in 1922 and in 1932) and is higher than the median for the Discretionary Portfolio.

However, Keynes was willing to be flexible in this area. Hence, he substantially increased his exposure to mining companies and to distressed businesses that offered the potential for recovery between 1927 and 1936, even though these stocks had, at that time, stopped paying dividends; in total, they represented one-in-three of his equity holdings. Thereafter, the proportion of zero-dividend paying stocks fell to less than one-in-five of his holdings through a combination of his rebalancing decisions and of companies reinstating their dividends. As a result, the Discretionary Portfolio readopted its distinctive high-dividend yield tilt.

In summary, Keynes constructed portfolios that were very different from the overall market. He adopted very active sector weightings, selected small-cap and mid-cap stocks, and rotated between high-dividend and low-dividend stocks relative to the market.

Keynes' Investment Philosophy

The core of Keynes' investment philosophy throughout the two-and-a-half decades he managed his college fund was a belief in the attractions of equities as a new asset class for long-horizon investors such as endowments.

During Keynes' time, UK institutional investor portfolios remained dominated by fixed income securities (Hannah 1986; Burton and Corner 1968; Baker and Collins 2003). To our knowledge, no other Oxbridge colleges made a substantial allocation to equities until the second half of the twentieth century. In the United States, the largest university endowments allocated less than 10 percent to common stock in the 1920s (on a historical cost-weighted basis), and this total only rose above 20 percent in the late 1930s (Goetzmann, Griswold, and Tseng 2010).

At this time when equities were rarely perceived as an institutional asset, Keynes made a radical shift to equities starting in 1921, and can justly be regarded as among the first institutional equity investors. Several years later, Keynes (1925) wrote a very positive review of Smith's (1924) book on *Common Stocks as Long Term Investments*, in which he extolled the virtues of US common stocks as residual claims on industrial growth. The annualized US equity risk premium over government bonds had been 4.28 percent during 1900–1920, and according to Dimson, Marsh, and Staunton (2013) over the same period it had been a near-identical 4.26 percent in the United Kingdom. Seeing the same potential in British ordinary shares as in American common stocks, Keynes identified their capacity to deliver a return premium over bonds.

History proved Keynes right, and between 1921 and 1946, the period during which he moved King's into equities, equities provided a real (inflation adjusted) return of 8.01 percent in the US and of 8.28 percent in the UK. Keynes' allocation of the Discretionary Portfolio to UK ordinary shares averaged 75 percent over the financial years 1922–29, 46 percent over 1930–39, and 69 percent over 1940–46. When we include the US common stocks, which he added to the portfolio from

1933 onwards, his total combined ordinary share and common stock weighting averaged 57 percent over 1930–39 and 73 percent over 1940–46.

Keynes (1925) also claimed a second advantage for equities, namely, the availability of an income premium over bonds. Over the period 1921–46, the Discretionary Portfolio was indeed able to earn an average dividend yield on its UK equities, including non-dividend-paying stocks, consistently higher than either the dividend yield on the UK equity market or the income return on government bonds.

Keynes was not a believer in market efficiency. In his Chairman's Speech to the 1938 National Mutual Insurance Company annual meeting, he argued: "[Markets] are governed by doubt rather than conviction, by fear more than forecast, by memories of last time and not by foreknowledge of next time. The level of stock prices does not mean that investors *know*, it means they do *not* know. Faced with the perplexities and uncertainties of the modern world, market values will fluctuate more widely than will seem reasonable in the light of after-events" (*Collected Writings* XII, p. 238, emphasis in original). Among modern economists, Shiller (2003) notably echoes these sentiments. In chapter twelve of *The General Theory*, Keynes (1936, p. 154) writes:

> It might have been supposed that competition between expert professionals, possessing judgment and knowledge beyond that of the average private investor, would correct the vagaries of the ignorant individual left to himself. It happens, however, that the energies and skill of the professional investor and speculator are mainly occupied otherwise. For most of these persons are, in fact, largely concerned, not with making superior long-term forecasts of the probable yield of an investment over its whole life, but with foreseeing changes in the conventional basis of valuation a short time ahead of the general public. They are concerned, not with what an investment is really worth to a man who buys it "for keeps", but with what the market will value it at, under the influence of mass psychology, three months or a year hence. Moreover, this behaviour is not the outcome of a wrong-headed propensity. It is an inevitable result of an investment market organised along the lines described. For it is not sensible to pay 25 for an investment of which you believe the prospective yield to justify a value of 30, if you also believe that the market will value it at 20 three months hence.

Given his views on stock market efficiency, Keynes pursued an active investment approach. However, this approach changed radically during the course of his investment career. He began managing the Discretionary Portfolio employing a top-down investment approach using monetary and economic indicators to market-time his switching between equities, fixed income, and cash. This approach called the "credit cycle theory of investment" was described in the prospectus of the Independent Investment Company, a closed-end fund he cofounded and floated on the London Stock Exchange in 1924 (*Collected Writings*, XII, p. 33).

However, after disappointing performance in the late 1920s, Keynes discarded his top-down market-timing approach. Our formal statistical tests fail to find any evidence that Keynes was successful at timing the stock market (Chambers, Dimson, and Foo 2013). Keynes himself confessed, when reflecting on his investment record for King's in a 1938 internal memorandum to his College investment committee, that: "Credit cycling means in practice selling market leaders on a falling market and buying them on a rising one and, allowing for expenses and loss of interest, it needs phenomenal skill to make much out of it. . . . We have not proved able to take much advantage of a general systematic movement out of and into ordinary shares as a whole at different phases of the trade cycle" (*Collected Writings* XII, pp. 100, 106).

Consistent with Keynes' rejection of his market-timing investment approach, there is a marked contrast between how he reacted to the major equity market falls in 1929–30 and in 1937–38. When the UK market began its fall in October 1929, he sold one-fifth of his UK equities by value over the following 12 months and switched into government bonds. In contrast, when the UK market began to decline sharply a second time in August 1937, he added modestly to his UK equity positions and maintained his equity allocation at over 90 percent.

Whilst maintaining his commitment to equities as a new asset class during his tenure as manager of the King's endowment, Keynes radically switched his investment approach to a bottom-up, buy-and-hold stock-picking approach. Detailed analysis of his investment correspondence as well as the statistical analysis of his performance in event-time strongly suggests the early 1930s as the most likely inflection point in the evolution of his investment approach (Chambers, Dimson, and Foo 2013).

In August 1934, Keynes outlined this new approach as follows: "As time goes on, I get more and more convinced that the right method in investment is to put fairly large sums into enterprises which one thinks one knows something about and in the management of which one thoroughly believes" (*Collected Writings* XII, p. 57).

In the same 1938 internal memorandum to his investment committee, Keynes justified his success in managing the College investments as being due to his "careful selection of a few investments" as judged by their "*intrinsic* value" (*Collected Writings* XII, p. 107, emphasis in original). Keynes' value-oriented approach is reminiscent of the framework favored by Graham and Dodd in their influential 1934 book, *Security Analysis*. The two Columbia finance professors advocated the use of careful fundamental analysis of corporate financial statements in order to identify stocks undervalued by the market. While Graham had applied this approach to investment beginning in 1923 (Carlen 2012, pp. 142–3), Keynes' focus on investment value was developed independently (Woods 2013). Our searches in the King's College Archives reveal no indication of any contact between the two men.

A good example of Keynes' value-oriented stock-picking is provided by one of his largest core holdings, Union Corporation, the South African mining company, which accounted on average for 51 percent of his gold mine exposure over the period 1933–46. In June 1934, Keynes outlined the key reasons he liked the stock, namely, the fact that the share price was trading at a 30 percent

discount to his conservatively estimated break-up value and that he evaluated and trusted the management very highly (*Collected Writings* XII, pp. 54–57). In thinking about the intrinsic value of a stock, Keynes sometimes thought in a novel way about equity valuation. In the case of Austin Motors, another of his core holdings, he valued the shares in terms of market capitalization per car produced and estimated that Austin traded at a 67 percent discount to General Motors in October 1933 (King's Archives JMK/PC/1/221-2).

One last question remains. If, as Keynes claimed, he shifted his approach to concentrating on a few core investments, then why does this shift not become apparent in the trend in the C5 and C20 Discretionary Portfolio concentration measures described earlier? The reason is, although the number of portfolio holdings expanded into the 1930s, Keynes put larger amounts into his favorite stocks, as noted by Boyle, Garlappi, Uppal, and Wang (2012). A better measure of the effective level of diversification is provided by the inverse of the Herfindahl index of individual equity weights, a measure first used in financial economics by Sharpe (1970). The inverse of the Herfindahl index indicates the number of uncorrelated holdings, each with the same risk attributes as the average portfolio constituent and held in equal proportions, that would have the same volatility as the Discretionary Portfolio.

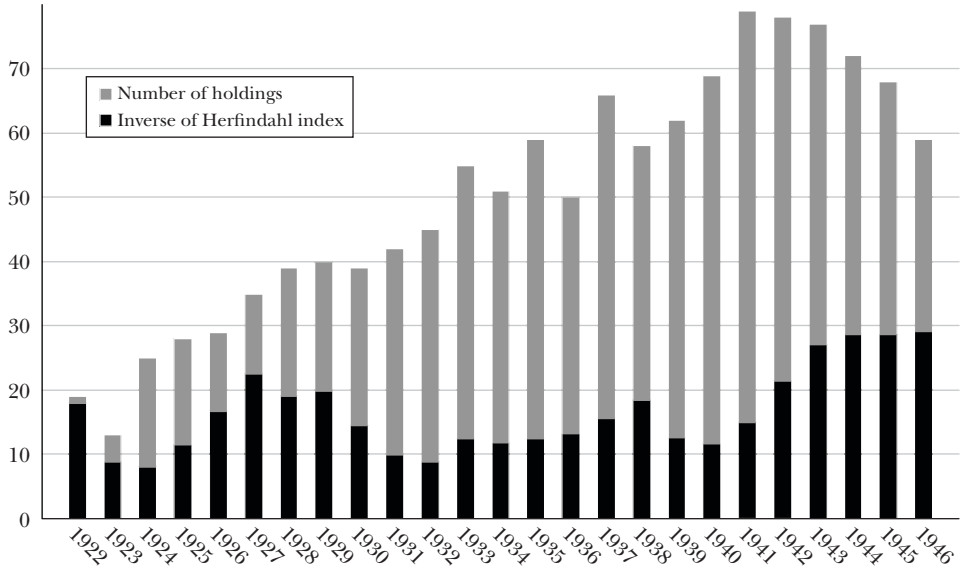
Looking at Figure 5, we see that in the early 1920s, when Keynes held a portfolio with almost identical weights in each constituent, the inverse of the Herfindahl index was close to the number of portfolio holdings. But over time the Discretionary Portfolio became less balanced and the inverse of the Herfindahl index peaked at 29 in the early 1940s. By that time, the effective diversification of Keynes' portfolio was markedly lower than the number of equity holdings might suggest. Moreover, Keynes' tendency to emphasise companies from a few industry sectors further reduced the effective diversification of his portfolios.

Was Keynes an Inside Trader?

Judging from his correspondence, Keynes' stock-picking was for the most part the product of fundamental security analysis based on reading the financial press and on sell-side research received mainly through a selection of London and provincial stockbrokers. Keynes supplemented his fundamental analysis by making use of his considerable network of City contacts. For example, he counted among his closest friends Rupert Trouton—his former pupil, his broker, and a director of the Norwegian whaling firm Hector Whaling, one of his core holdings—and Henry Strakosch, who was chairman of Union Corporation, another of his core holdings. Keynes made particular use of Strakosch and his staff, when undertaking due diligence on mining stocks (King's Archives JMK/KC/5/3). However, he was not always personally connected to his core holdings and there was no apparent personal connection in the case of Austin Motors.

When the 7,632 potential personal contacts from Keynes' time at Eton College, Cambridge University, the Treasury during World War I, and from public

Figure 5
UK Discretionary Portfolio Concentration: 1922–1946



Notes: For each financial year ended August, we present the number of holdings (the gray bars) and the inverse of the Herfindahl index for the Discretionary Portfolio (the black bars). The inverse of the Herfindahl index indicates the number of uncorrelated holdings, each with the same risk attributes as the average portfolio constituent and held in equal proportions, that would have the same volatility as the Discretionary Portfolio.

life are matched with the directors of the 247 firms in which he invested, Keynes was ultimately connected to 46 of those firms (Eldridge 2012). His connections proved particularly influential in the mining sector. The existence of a connection to a director at the time of investment led Keynes to allocate on average four times the weighting to stocks of mining firms as compared to nonmining firms, and furthermore this benefited performance (Chambers, Dimson, and Foo 2013).

Given his extensive contacts, the question arises as to whether Keynes was an insider. It would be surprising if he was not sometimes the recipient of what today would be deemed price-sensitive information. For example, he became aware of an imminent change in the Bank of England interest rate in 1925 (Mini 1995). However, the question is ultimately anachronistic given that insider trading by investors was not regulated in the United Kingdom until 1980, with the main exception being directors who owed fiduciary duties to their company not to trade on price-sensitive information (Cheffins 2008, pp. 39–40).

It is not possible with our data to discover how frequently and the extent to which Keynes exploited such information in his trading. However, if Keynes was using his

insider status, he was doing so not for gains from short-term trading, but instead for systematically accumulating long-term positions in his favorite shares such as Union Corporation, Hector Whaling, and Austin Motors. In addition, Keynes rarely engaged in flipping new issues—an activity that might plausibly be linked to insider status. Only 5 percent of his purchase transactions for King's College were initial public offerings and in only a very few cases did he sell out in the days immediately following the start of trading.

Conclusion

Investing is an important and largely overlooked part of Keynes' professional life. His experiences trading in the stock market influenced his economic theorizing as evidenced by chapter 12 of the *General Theory*. Here he reflected on the possible adverse effect on corporate investment and the macroeconomy of stock market fluctuations resulting from the ebb and flow of investor sentiment. Indeed, his initial setbacks as an investor led him to bemoan the seeming inability of the "serious-minded" investor "to purchase investments on the best genuine long-term expectations he can frame" (Keynes 1936, p. 156). Whilst impressive, his investment performance for King's College was not the uninterrupted success that has previously been believed. As Keynes himself documented, early attempts at a macro market-timing style of investing proved disappointing and led to a fundamental overhaul of his investment approach and his becoming a more patient, bottom-up stock-picker after the early 1930s.

The most significant of Keynes' contributions to professional investment management was his path-breaking and strategic allocation to equities together with his early adoption of value-based investment strategies. By the 1940s, the weight of common stocks had increased to represent over half of the whole King's College endowment's security portfolio as a result of his equity-focused strategy. Institutional managers in general did not mimic this strategy until the second half of the twentieth century.

■ *We wish to acknowledge the important work of the King's College Cambridge archivist, Patricia McGuire. Valued comments were received from Sir John Bradfield, John Campbell, Keith Carne, Oliver Dawson, Charley Ellis, Justin Foo, Walter Friedman, William Goetzmann, Chang-Tai Hsieh, Ulrike Malmendier, Tom Nicholas, Nigel Prentice, Raghavendra Rau, William Sharpe, Timothy Taylor, seminar participants at Boston, Harvard (HBS), London (LSE), Maastricht, Oxford (Saïd), Rutgers, Tilburg, Vienna, Virginia (Darden School), Warwick, and Yale (School of Management), and conference participants at the ASSA Meetings (New Orleans), CFA Society of the UK (London), Economic History Society Annual Conference (Nottingham), Gutmann Center for Portfolio Management (Vienna), NBER Conference on the Great Recession and Higher Education (Boston), Russell Academic Advisory Board, Summit on Investing for the Long Run (Oslo), World Economic History Congress (Stellenbosch), and*

World Investment Forum (Georgia). We especially thank Stephen Dimmock, Paul Marsh, and Mike Staunton for data and ideas. The authors were supported by the Robert Brown Foundation, a JM Keynes Fellowship in Financial Economics (Chambers) and a Leverhulme Trust Emeritus Fellowship (Dimson).

References

Accominotti, Olivier, and David Chambers. 2013. "The Returns to Currency Trading: Evidence from the Interwar Period." Available at SSRN: http://ssrn.com/abstract=2293684.

Acharya, Shanta, and Elroy Dimson. 2007. *Endowment Asset Management: Investment Strategies in Oxford and Cambridge*. Oxford University Press.

Baker, Mae, and Michael Collins. 2003. "The Asset Portfolio Composition of British Life Insurance Firms, 1900–1965." *Financial History Review* 10(2): 137–164.

Boyle, Phelim, Lorenzo Garlappi, Raman Uppal, and Tan Wang. 2012. "Keynes Meets Markowitz: The Trade-off between Familiarity and Diversification." *Management Science* 58(2): 253–72.

Brown, Jeffrey R., Stephen G. Dimmock, Jun-Koo Kang, and Scott J. Weisbenner. Forthcoming. "How University Endowments Respond to Financial Market Shocks: Evidence and Implications." *American Economic Review*.

Buffett, Warren. 2013. *Berkshire Hathaway Letters to Shareholders 1965–2012*. Omaha, NE: Berkshire Hathaway Inc.

Burton, Harry, and Desmond Corner. 1968. *Investment and Unit Trusts in Britain and America*. London: Elek.

Carlen, Joe. 2012. *Einstein of Money: The Life and Timeless Financial Wisdom of Benjamin Graham*. New York: Prometheus Books.

Chambers David, Elroy Dimson, and Justin Foo. 2013. "Keynes the Stock Market Investor: The Inception of Institutional Equity Investing." Available at SSRN: http://ssrn.com/abstract=2023011. (Forthcoming in *Journal of Financial and Quantitative Analysis*)

Cheffins, Brian R. 2008. *Corporate Ownership and Control: British Business Transformed*. Oxford University Press.

Chua, Jess H., and Richard S. Woodward. 1983. "J. M. Keynes' Investment Performance: A Note." *Journal of Finance* 38(1): 232–35.

Cox, Jacqueline. 1995. "Keynes: An Archivist's View." *History of Political Economy* 27(5): 163–175.

Cremers, Martijn, and Antti Petajisto. 2009. "How Active is Your Fund Manager? A New Measure That Predicts Performance." *Review of Financial Studies* 22(9): 3329–65.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton University Press.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2013. *Global Investment Returns Sourcebook*. Zurich: Credit Suisse Research Institute.

Dunbabin, John. 1975. "Oxford and Cambridge College Finances, 1871–1913." *Economic History Review* 28(4): 631–47.

Eldridge, Matthew. 2012. *Keynes' Personal Investor Network*. Cambridge University MPhil dissertation.

Fama, Eugene F. 1976. *Foundations of Finance: Portfolio Decisions and Securities Prices*. New York: Basic Books.

Fantacci, Luca, Maria Cristina Marcuzzo, and Eleonora Sanfilippo. 2010. "Speculation in Commodities: Keynes' 'Practical Acquaintance' with Future Markets." *Journal of the History of Economic Thought* 32(3): 397–418.

Goetzmann, William N., John Griswold, and Yung-Fang (Ayung) Tseng. 2010. "Educational Endowments in Crises." *Journal of Portfolio Management* 36(4): 112–23.

Graham, Benjamin, and David Dodd. 1934 [2008]. *Security Analysis*. New York: Whittlesey House (a division of McGraw-Hill).

Hannah, Leslie. 1986. *Inventing Retirement: The Development of Occupational Pensions in Britain*. Cambridge University Press.

Keynes, John Maynard. 1925. "An American Study of Shares versus Bonds as Permanent Investments." (In *The Collected Writings of John Maynard Keynes* (1983), edited by Moggridge vol. 12, pp. 247–52.)

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest, and Money*. Cambridge University Press.

Lewellen, Jonathan. 2011. "Institutional Investors and the Limits of Arbitrage." *Journal of Financial Economics* 102(1): 62–80.

Mini, Piero V. 1995. "Keynes' Investments: Their Relation to the General Theory." *American Journal of Economics and Sociology* 54(1): 47–56.

Moggridge, Donald, ed. 1983. *The Collected Writings of John Maynard Keynes*, 30 vols. Cambridge University Press. [A 2012 edition of *The Collected Writings* is also available from Cambridge University Press.]

Neild, Robert. 2008. *Riches and Responsibility: The Financial History of Trinity College, Cambridge*. Cambridge: Granta Editions.

Pierce, Tomas. 1993. "Keynes' Personal Investing: Activities and Beliefs." *Social Science Journal* 30(1): 13–22.

Sharpe, William. 1970. *Portfolio Theory and Capital Markets*. New York: McGraw-Hill.

Shiller, Robert J. 2003. "From Efficient Markets Theory to Behavioral Finance." *Journal of Economic Perspectives* 17(1): 83–104.

Smith, Edgar Lawrence. 1924 [2012]. *Common Stocks as Long Term Investments*. New York: Ferris Printing Company. [Whitefish, MT: Montana: Kessinger Publishing.]

Soros, George. 1987. *The Alchemy of Finance: Reading the Mind of the Market*. New York: Wiley.

Soros, George. 2011. "My Philanthropy." In *The Philanthropy of George Soros: Building Open Societies*, edited by Chuck Sudetic, 1–60. Philadelphia, PA: Perseus Books.

Stock Exchange Official Intelligence, The. 1926. London: Spottiswoode, Ballantyne.

Swensen, David F. 2001 [2009]. *Pioneering Portfolio Management: An Unconventional Approach to Institutional Investment*. New York, NY: Free Press.

Swensen, David F. 2005. *Unconventional Success: A Fundamental Approach to Personal Investment*. New York, NY: Simon & Schuster.

Woods, John E. 2013. "On Keynes as an Investor." *Cambridge Journal of Economics* 37(2): 423–42.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 229–236

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 email at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., Saint Paul, Minnesota, 55105.

Smorgasbord

The 2013 *Global Monitoring Report*, jointly published by the World Bank and the IMF, discusses "Rural–Urban Dynamics and the Millennium Development Goals." "In the past two decades, developing countries have urbanized rapidly, with the number of people living in urban settlements rising from about 1.5 billion in 1990 to 3.6 billion (more than half of the world's population) in 2011. . . . Nearly 50 percent of the population in developing countries was urban in 2011, compared with less than 30 percent in the 1980s. Urban dwellers are expected to double between 2000 and 2030, from 2 billion to 4 billion people . . ." "Cities and towns are hubs of prosperity—more than 80 percent of global economic activity is produced in cities by just over half of the world's population." "Location remains

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives*, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

http://dx.doi.org/10.1257/jep.27.3.229. doi=10.1257/jep.27.3.229

important at all stages of development, but it matters less in rich countries than in poor ones. Estimates from more than 100 Living Standard Surveys indicate that households in the most prosperous areas of developing countries such as Brazil, Bulgaria, Ghana, Indonesia, Morocco, and Sri Lanka have an average consumption almost 75 percent higher than that of similar households in the lagging areas of these countries. In comparison, the disparity is less than 25 percent in developed countries such as Canada, Japan, and the United States." "Slums are the urban face of poverty and emerge when cities are unable to meet the demand for basic services and to supply the expected jobs. A likely 1 billion people live in urban slums in developing countries, and their numbers are projected to grow by nearly 500 million between now and 2020." "Urbanization is largely a natural process, driven by the opportunities cities offer. Unregulated markets are unlikely to get densities right, however, and spontaneous development of cities can create negative side effects such as congestion or, alternatively, excessive sprawl. . . . While a market-driven process could possibly gradually increase densities through shifting land values over time, the long-lived and lumpy nature of urban investment often inhibits such a process. A city's physical structures, once established, may remain in place for more than 150 years." http://siteresources.worldbank.org/INTPROSPECTS/Resources/334934-1327948020811/8401693-1355753354515/8980448-1366123749799/GMR_2013_Full_Report.pdf.

The US Postal Service Office of Inspector General has published "The Untold Story of the ZIP Code," which is now 50 years old. "The ZIP Code was established as an open use product publicly accessible from the outset. In fact, the Postal Service only filed a trademark for the 'ZIP Code' name in 1973. The openness of the ZIP Code as a platform for economic activity is part of the reason for its immense success far beyond its initial conception. Unlike most commodities, the ZIP Code is not rivalrous; use by one party does not exclude its use by any other. The Post Office took no steps to make the ZIP Code exclusive but rather provided it as a public good for use by any party, free of charge. . . . Other organizations and businesses soon realized the ZIP Code possessed an elegant simplicity for efficiently organizing data by geography. The U.S. Census Bureau, for example, uses the ZIP Code to organize its statistics. Other industries, like real estate and target marketing companies, redefined the way they do business by basing their informational structure on the ZIP Code. The ZIP Code is solicited or used in a variety of transactions, such as buying gas with a credit card at an automated pump. Today, a ZIP Code and physical mailing address are widely recognized attributes of an individual's identity." The study estimated that the ZIP code produces about $10 billion per year in economic value in the United States. "Current estimates show as many as 4 billion people worldwide are unaddressed and approximately sixty Universal Postal Union countries have no postal code system." "There is strong evidence that implementing addressing systems in impoverished neighborhoods can actually increase the overall quality of life by allowing basic infrastructure, such as electricity, water, communication, and government services to be delivered to the area. This was seen in the slums of Calcutta, for example, where spray-painting unique addressing numbers on

houses yielded significant positive effects on overall quality of life in the city's neighborhoods. This effort has allowed the local government to organize the delivery of water and electrical utilities to the slums and residents now have the legal identities required to apply for bank accounts and jobs." April 1, 2013. Research Paper RARC -WP-13-006. https://www.uspsoig.gov/foia_files/rarc-wp-13-006.pdf.

The *Global Wage Report 2012/13* of the International Labour Organization considers the theme of "Wages and Equitable Growth." As one example of many topics discussed, Chapter 5 examines "The Fall in the Labour Income Share." "The OECD has observed, for example, that over the period from 1990 to 2009 the share of labour compensation in national income declined in 26 out of 30 developed economies for which data were available, and calculated that the median labour share of national income across these countries fell considerably from 66.1 per cent to 61.7 per cent . . . Looking beyond the advanced economies, the ILO World of Work Report 2011 found that the decline in the labour income share was even more pronounced in many emerging and developing countries, with considerable declines in Asia and North Africa and more stable but still declining wage shares in Latin America." 2013. http://www.ilo.org/wcmsp5/groups/public/—dgreports /—dcomm/—publ/documents/publication/wcms_194843.pdf.

Douglas J. Elliott, Greg Feldberg, and Andreas Lehnert discuss "The History of Cyclical Macroprudential Policy in the United States." "We provide a simple taxonomy and economic model of the countercyclical macroprudential tools that the Federal Reserve and other agencies have used since the First World War. The key distinction is between tools that operate on the demand for credit, such as limits on loan-to-value ratios and loan maturities, and those that operate on the supply of credit, such as limits on deposit rates (and therefore the supply of funds to lend), limits on lending rates, restrictions on banks' portfolios, reserve requirements, capital requirements, and supervisory pressure." "Many of these tools appear to have succeeded in their short-term goals; for example, limiting specific types of bank credit or liability and impacting terms of lending. It is less obvious that they have improved long-term financial stability or, in particular, successfully managed an asset price bubble, and this is fertile ground for future research. Meanwhile, these tools have faced substantial administrative complexities, uneven political support, and competition from nonbank or other providers of credit outside the set of regulated institutions." Federal Reserve Board, Finance and Economics Discussion Series staff working paper 2013-29. May 15, 2013. http://www.federalreserve .gov/pubs/feds/2013/201329/201329pap.pdf.

A Selection of Symposia

The Spring 2013 issue of the *International Productivity Monitor* features a group of papers on the prospects for future US productivity growth. Martin Neil Baily, James Manyika, and Shalabh Gupta lead off with "U.S. Productivity Growth: An Optimistic Perspective." "[D]igital technology and the digital revolution are proceeding apace,

and we also agree that this will eliminate many traditional jobs in manufacturing and elsewhere. But the offset is that innovation-led growth can create new jobs, new lines of business and new profit opportunities." Robert J. Gordon responds in: "U.S. Productivity Growth: The Slowdown Has Returned after a Temporary Revival." "I have often posed the following set of choices. Option A is to keep everything invented up until ten years ago, including laptops, Google, Amazon, and Wikipedia, while also keeping running water and indoor toilets. Option B is to keep everything invented up until yesterday, including Facebook, iphones, and ipads, but give up running water and indoor toilets; one must go outside to take care of one's needs; one must carry all the water for cooking, cleaning, and bathing in buckets and pails. Often audiences laugh when confronted with the choice between A and B, because the answer seems so obvious. But running water and indoor toilets were not the only inventions between 1870 and 1970 that made it possible for U.S. labour productivity to grow at the 2.48 per cent rate . . . The list is endless—electric light, elevators that made possible the vertical city, electric machine tools and hand tools, central heating, air conditioning, the internal combustion engine that replaced the horse, commercial aviation, phonographs, motion pictures, radio, TV, and many others including fundamental medical inventions ranging from aspirin to penicillin. By comparison the computer revolution kick-started productivity growth between 1996 and 2004 for only eight years, compared to the 81 years propelled by the second Industrial Revolution of the late nineteenth century." David M. Byrne, Stephen D. Oliner, and Daniel E. Sichel consider possible time lags in their paper, "Is the Information Technology Revolution Over?" "Just as a long lag transpired from the development of the PC in the early 1980s to the subsequent pickup in labour productivity growth, there could be a lagged payoff from the development and diffusion of extensive connectivity, handheld devices, and ever-greater and cheaper computing power. In 1987, Robert Solow famously said 'You see the computer revolution everywhere except in the productivity data.' . . . [C]omputers comprised too small a share of the capital stock in 1987 to have made a large contribution to overall productivity growth. But, several years later, the imprint of the revolution became very evident. In a parallel vein, one could now say: 'You see massive connectivity and ever-cheaper computing power everywhere but in the productivity data.' Subsequently, those contributions could become evident in aggregate data." http://www.csls.ca/ipm/ipm25.asp.

Lucrezia Reichlin and Richard Baldwin have edited an e-book, *Is Inflation Targeting Dead? Central Banking after the Crisis*, with 14 crisp and readable essays. As one example, Michael Woodford contributes: "Inflation Targeting: Fix It, Don't Scrap It." He writes: "It is important, first of all, to recognise that proponents of inflation targeting do not actually have in mind a commitment by the central bank to base policy decisions purely on their consequences for inflation, and to act so as to keep the inflation rate as close as possible to the target rate at all times. Mervyn King (1997) memorably referred to this as the 'inflation nutter' position, and distinguished the 'flexible' inflation targeting that he advocated from it . . ." "I thus believe that it would be possible to avoid the problems with inflation targeting as currently practised, that have been the focus of recent criticism of inflation targeting as such,

while retaining the essential features of an inflation targeting regime: not only a public commitment to a fixed numerical target for the medium-run rate of inflation, and a commitment to regularly explain how policy decisions are consistent with that commitment, but the use of a forecast-targeting procedure as the basis both for monetary-policy deliberations and for communication with the public about the bank's decisions and their justification." 2013. A VoxEU.org book. http://www.voxeu.org/sites/default/files/file/P248%20inflation%20targeting%282%29.pdf.

In a background paper for an IMF conference on "Rethinking Macro Policy," Olivier Blanchard, Giovanni Dell'Ariccia, and Paolo Mauro seek to focus the discussion with "Rethinking Macro Policy II: Getting Granular." "The 2008–09 global economic and financial crisis shook the consensus on how to run macroeconomic policy. . . . This prompted a healthy reconsideration of what worked and what did not, and a debate on how to fix things, ranging from nitty-gritty technical points to broad-based institutional design questions." They set the stage with a dozen live issues about monetary, fiscal, and macroprudential policies: 1) Should Central Banks Explicitly Target Activity? 2) Should Central Banks Target Financial Stability? 3) Should Central Banks Care about the Exchange Rate? 4) How Should Central Banks Deal with the Zero Bound? 5) To Whom Should Central Banks Provide Liquidity? 6) What Are the Dangers of High Public Debt? 7) How to Deal with the Risk of Fiscal Dominance? 8) At What Rate Should Public Debt Be Reduced? 9) Can We Do Better Than Automatic Stabilizers? 10) How to Combine Macroprudential Policy and Microprudential Regulation? 11) What Macroprudential Tools Do We Have and How Do They Work? 12) How to Combine Monetary and Macroprudential Policies? The conference was April 16–17, 2013. Video of panel discussions full of high-powered participants, along with conference papers, are available at http://www.imf.org/external/np/seminars/eng/2013/macro2/index.htm.

The Spring 2013 issue of *Future of Children* contains a symposium on "Postsecondary Education in the United States," with articles on costs, returns, student support, for-profit education, financial aid, and other issues. My eye was caught, particularly, by "E-learning in Postsecondary Education," by Bradford S. Bell and Jessica E. Federman. "During the fall 2010 term 31 percent of U.S. college students took at least one online course." "These meta-analytic studies . . . as well as the findings of earlier reviews not discussed here, suggest that e-learning is at least as effective as, and in some cases more effective than, classroom instruction. But taking into account various methodological and instructional factors can change the findings—typically not reversing them but rather weakening or eliminating the observed benefits of e-learning. . . . Such variability suggests that other explanations—such as aspects of the instruction, teacher effectiveness, or student characteristics—account for the relative effectiveness of e-learning in the studies." http://futureofchildren.org.

The Summer 2013 issue of *Regulation* magazine has five articles under the cover headline: "Gambling on Global Warming: How Much Should We Tax an Unknown Risk?" As one example, Robert S. Pindyck writes on "Pricing Carbon When We Don't Know the Right Price: Despite the Unknowns, We Should Begin to Tax Carbon." "Some would argue that any increases in global temperatures will be moderate,

will occur in the far distant future, and will have only a small impact on the economies of most countries. If that's all true, it would imply that the SCC [social cost of carbon] is small, perhaps only around $10 per ton of CO_2, which would justify a very small (almost negligible) tax on carbon emissions, e.g., something like 10 cents per gallon of gasoline. Others would argue that without an immediate and stringent GHG abatement policy, there is a reasonable possibility that substantial temperature increases will occur and might have a catastrophic effect. That would suggest the SCC is large, perhaps $100 or $200 per ton of CO_2, which would imply a substantial tax on carbon, e.g., as much as $2 per gallon of gas. So who is right, and why is there such wide disagreement?" "In fact, we know very little about the effects of climate change, in part because of the possibility of adaptation (climate change will occur slowly, over decades) and in part because we know little about technological change that might reduce GHG [greenhouse gas] emissions and/or facilitate adaptation. Finally, there is disagreement about the framework that should be used to evaluate the benefits from GHG abatement, including the social welfare function and discount rate to be used to value benefits that will occur in the distant future." http://www.cato.org/regulation/summer-2013.

Interviews

Dani Rodrik offers his perspective on economic methods and diverse conclusions in answering questions from the *World Economics Association Newsletter.* "I have never thought of neoclassical economics as a hindrance to an understanding of social and economic problems. To the contrary, I think there are certain habits of mind that come with thinking about the world in mainstream economic terms that are quite useful: you need to state your ideas clearly, you need to ensure they are internally consistent, with clear assumptions and causal links, and you need to be rigorous in your use of empirical evidence. Now, this does not mean that neoclassical economics has all the answers or that it is all we need. Too often, people who work with mainstream economic tools lack the ambition to ask broad questions and the imagination to go outside the box they are used to working in. But that is true of all 'normal science.' Truly great economists use neoclassical methods for leverage, to reach new heights of understanding, not to dumb down our understanding. Economists such as George Akerlof, Paul Krugman, and Joe Stiglitz are some of the names that come to mind who exemplify this tradition. Each of them has questioned conventional wisdom, but from within rather than from outside. . . . "Pluralism on policy is already a reality, even within the boundaries of the existing methods . . . And there are certainly some areas, for example international trade, where economists' views are much less diverse than public opinion in general. But economics today is not a discipline that is characterized by a whole lot of unanimity." April 2013. http://www.worldeconomicsassociation.org/files/newsletters/Issue3-2.pdf.

Renee Haltom interviews Christopher Carroll, with a focus on savings and consumption behavior. "The theory in every textbook says that if you know you're

going to be richer in the future because you're a fast-growing country, why in the world would you save now, when you're poor, making your future rich self better off? It makes much more sense to borrow now since it'll be easy for you to pay off that debt in the future when you're richer. The latest example that's on everybody's minds is, of course, China, a country that has grown very fast for the last 20 years and has had a saving rate that just seems to get higher every year. . . . But what China is doing right now actually looks virtually identical to Japan 30 years ago. Japan didn't have a particularly high saving rate in the 1950s, and by the 1970s it had the highest saving rate in the world, and that was a period of high growth in Japan. It's also true in South Korea. It grew at a very rapid rate starting from the early 1960s, and its saving rate went up and up. We also see this in Taiwan, Singapore, and Hong Kong. And it's not just East Asian countries; the same is true of Botswana and Mauritius. It's also true in the opposite direction for European countries, which were growing pretty fast after World War II. That fast growth came to an end in the early 1970s, and afterward the saving rate declined . . ." So it seems to be a pretty pervasive, large effect that is really very much the opposite of what you'd expect from the standard off-the-shelf models. . . . In fact, what I really think is the right story is one that combines habit formation and a precautionary motive, such that they intensify each other. If I have these habits, then a good reason to resist spending when my income goes up is uncertainty over whether the factory that I'm working for will close down and I'll have to go back to my rural peasant roots." *Econ Focus:* Federal Reserve Bank of Richmond. 2013. First Quarter, pp. 30–34. http://www.richmondfed.org/publications/research/econ_focus/2013/q1/pdf/interview.pdf.

Douglas Clement interviews Susan Athey. "In fact, the need for theory is in some ways magnified by having large amounts of data. When you have a small amount of data, you can just look at the data and build your intuition from it. When you have very large amounts of data, just taking an average can cost thousands of dollars of computer time. So you'd better have an idea of what you're doing and why before you go out to take those averages. The importance of theory to create conceptual frameworks to know what to look for has never been larger . . . I think what *is* true is that when you have large amounts of data, if you ask it the right questions, you have a greater ability to let the data speak, and so you can be much less reliant on assumptions. But you still need a strong conceptual framework to understand what's coming out. . . . I think that the data scientists should take a little more economics. That would help; economics puts a lot of emphasis on the conceptual framework. And I also think that economics should be paying a lot more attention to the statistics of big data. Right now, economics as a profession has *very little* market share in the business analysis of this big data. It's mostly statisticians. We're just not training our undergraduates to be qualified for these jobs. Even our graduate students, even someone with a Ph.D. from a very good economics department really doesn't have the right skills to analyze the kinds of data sets that big Internet firms are creating." *The Region*, Federal Reserve Bank of Minneapolis, June 2013, pp. 16–28. http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=5112.

Discussion Starters

Steve Henderson looks at "Spending on Pets: 'Tails' from the Consumer Expenditure Survey." "Nearly three-quarters of U.S. households own pets. There are about 218 million pets in the United States, not counting several million fish. . . . Americans spent approximately $61.4 billion in total on their pets in 2011. On average, each U.S. household spent just over $500 on pets. This amounts to about 1 percent of total spending per year for the average household. . . . Expenditures on pets include pet food, pet purchases, supplies and medicine, pet services, and veterinarian services." US Bureau of Labor Statistics, *Beyond the Numbers*, May 2013, vol. 2, no. 16. http://www.bls.gov/opub/btn/volume-2/spending-on-pets.htm.

The International Energy Administration has published *Tracking Clean Energy Progress 2013*. "The IEA Energy Sector Carbon Intensity Index (ESCII) tracks how many tonnes of CO_2 are emitted for each unit of energy supplied. It shows that the global aggregate impact of all changes in supply technologies since 1970 has been minimal. Responses to the oil shocks of the 1970s made the energy supply 6% cleaner from 1971 to 1990. Since 1990, however, the ESCII has remained essentially static, changing by less than 1%. . . ." http://www.iea.org/publications/TCEP_web.pdf.

Rebecca Ray, Milla Sanes, and John Schmitt lay out some international comparisons on the legal right to paid vacation in "No-Vacation Nation Revisited." "The United States is the only advanced economy in the world that does not guarantee its workers paid vacation. European countries establish legal rights to at least 20 days of paid vacation per year, with legal requirements of 25 and even 30 or more days in some countries. Australia and New Zealand both require employers to grant at least 20 vacation days per year; Canada and Japan mandate at least 10 paid days off. The gap between paid time off in the United States and the rest of the world is even larger if we include legally mandated paid holidays, where the United States offers none, but most of the rest of the world's rich countries offer at least six paid holidays per year." Center for Economic and Policy Research. May 2013. http://www.cepr.net/documents/publications/no-vacation-update-2013-05.pdf.

Journal of Economic Perspectives—Volume 27, Number 3—Summer 2013—Pages 237–238

Notes

***For additional announcements, check out the continuously updated JEP online Bulletin Board**, (http://www.aeaweb.org/bulletinboard.php). Calls for papers, notices of professional meetings, and other announcements of interest to economists should be submitted to Ann Norman, at jep@jepjournal.org, in one or two paragraphs containing the relevant information. These will be posted at the JEP online Bulletin Board. Given sufficient lead time (at least one month before an issue goes online), we will also print a shorter, one-paragraph version of your notice in the "Notes" section of the* Journal of Economic Perspectives. *We reserve the right to edit material received.*

The Annual Meeting of the American Economic Association will be held in Philadelphia, PA, January 3–5, 2014. The headquarters, registration and exhibits will be in the Philadelphia Marriott Downtown. Information and procedures for employers and job seekers are in the registration material at (www.vanderbilt.edu/AEA). There is no on-site interview arrangement service, nor will there be an on-site message exchange center; all correspondence, including interview scheduling, should take place over the Internet prior to arrival in Philadelphia. The location of the interview tables will be the Pennsylvania Convention Center. Registration will open mid-September; for additional information or to register for the meeting, please go to (www.vanderbilt.edu/AEA).

The Association's 2014 Continuing Education Program will feature three concurrent programs on January 5–7, 2014, in Philadelphia, PA. Participants can choose from among: Cross-Section Econometrics, Economic Growth, and Education and the Economy.

New Disclosure Policy for AEA journal submissions. Beginning July 1, 2012, all submissions to AEA journals, including revisions of previously submitted papers, must be accompanied by a Disclosure Statement. Please see the complete Disclosure Policy at (http://www.aeaweb.org/aea_journals/AEA_Disclosure_Policy.pdf).

John Bates Clark Medal. The American Economic Association is pleased to announce that Raj Chetty was awarded the John Bates Clark Medal for 2013.

2013 Distinguished Fellows. The Association is pleased to announce that the Distinguished Fellows for 2013 are Harold Demsetz, Stanley Fischer, Jerry Hausman, and Paul Joskow.

Nominations of AEA Officers for 2014. The slate of nominees for Association offices is available at the AEA website (www.vanderbilt.edu/AEA). If you do not have Internet access, you may request this information by fax (615–343–7590) or by mail (American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203).

EconLit now starts with 1886. The AEA has added EconLit records for journal articles from 1886–1968 that were previously in the Index of Economic Articles, Vols. 1–10. EconLit on library websites now includes older articles from 146 journals, 95 of which are currently indexed. Many of these journals are available through libraries' full text subscriptions and may be linked to/from EconLit.

The Committee on the Status of Women in the Economics Profession (CSWEP) seeks nominations for the 2013 Carolyn Shaw Bell Award, given annually to an individual who has furthered the status of women in the economics profession, through example, achievements, increasing our understanding of how women can advance in the economics profession, or mentoring others. Nominations should include a nomination letter, current CV, and at least two supporting letters. More information on this award, including past winners, is at (http://www.aeaweb.org/committees/cswep/awards/). Nominations for this award may be sent to: Marjorie McElroy, CSWEP Chair, Duke University, Department of Economics, 219A Social Sciences, Campus Box 90097, Durham, NC 27708-0097; phone: (919) 660-1840; FAX: (919) 684-8974;

email: (cswep@econ.duke.edu). **Due date for nominations is September 20, 2013.**

CSWEP will hold its national mentoring workshop, CeMENT 2014, immediately following the ASSA meetings in Philadelphia on January 5–7, 2014. The workshop is geared toward female junior faculty in tenure-track positions at PhD-granting economics departments in the US, or at institutions with similar research expectations. Application and registration material are at http://www.aeaweb.org/committees/cswep/mentoring/. The **application deadline is September 20, 2013.**

The 2013 Julius Shiskin Memorial Award for Economic Statistics goes to John C. Haltiwanger, Distinguished University Professor of Economics at the University of Maryland, and Maurine Haver, President and founder of Haver Analytics Inc. The award recognizes unusually original and important contributions in the development of economic statistics or in the use of statistics in interpreting the economy.

Submissions are invited for the 2013 TIAA-CREF Paul A. Samuelson Award for outstanding scholarly writing on lifelong financial security, which includes a cash prize of $10,000. The research must be in a book or article form and must have been published between January 1, 2012, and June 30, 2013. **Deadline for submissions: September 10, 2013.** For more information visit: http://www.tiaa-crefinstitute.org/awards/samuelson.html.

Submissions invited for the Warren Samuels Prize: The Association for Social Economics (ASE) together with the *Review of Social Economy* invite submissions for the 2014 Warren Samuels Prize. This prize is awarded to a paper, preferably one presented at the January ASSA meetings, that best exemplifies scholarly work that: is of high quality, is important to the project of social economics, and has broad appeal across disciplines. Submit your paper, as a word or pdf attachment, to Wilfred Dolfsma, Editor-in-Chief, *Review of Social Economy*, **before December 5, 2013** at w.a.dolfsma@rug.nl.

***Strategic Change: Briefings in Entrepreneurial Finance* invites papers for a special issue on microfinance.** Please submit your full papers to the Editorial Assistant Gemma Valler, gvaller@wiley.com, with a copy to Prof. Arvind Ashta, arvind.ashta@escdijon.eu, **by January 15, 2014.** For more, see: http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1099-1697/homepage/ForAuthors.html.

The first Winter School on the Analytics and Policy Design of Migration will be held at Georgetown University's School of Foreign Service in Doha, Qatar (GU-Q) on January 7–11, 2014. The school will address various approaches to the modeling of migration choices and consequences and how this could improve the spectrum of informed migration policies. **Application deadline: October 15, 2013.** For details, see: http://qatar.sfs.georgetown.edu/academics/migrationpolicy/.

The 5th International Conference on the Institutional and Technological Environment for Microfinance, ITEM 5, organized by Banque Populaire Chair in Microfinance of the Burgundy School of Business, will be held March 19–20, 2014, in Casablanca, Morocco, and will focus on the how microfinance recovers from its own image crisis. Submit **abstracts by September 30th, 2013,** to: microfinancechair@escdijon.eu. Updates at: http://burgundy-microfinance.weebly.com/annual-events-item.html.

CoreLogic Academic Research Council (CLARC) encourages and supports research on real estate markets, mortgage finance, and housing policy, and is **inviting grant proposals for the fall of 2013** based on the CoreLogic data. For more information on available data assets, the CLARC grant program, and the Award of Excellence, visit CLARC at http://www.corelogic.com/about-us/researchtrends/academic-research-council.aspx.

Errata:

In the reference list of "Latin America's Social Policy Challenge: Education," by Santiago Levy and Norbert Schady, which appeared in the Spring 2013 issue, the references

Pagés, Carmenk. 2011. *Social Strategy for Equity and Productivity in Latin America and the Caribbean*. Washington, DC: Inter-American Development Bank.
Pagés, Carmen. 2012. *The World of Forking Paths: Latin America and the Caribbean in the Face of Global Economic Risks*. Washington, DC: Inter-American Development Bank.

should read:

Inter-American Development Bank. 2011. *Social Strategy for Equity and Productivity in Latin America and the Caribbean*. Washington, DC: Inter-American Development Bank.
Inter-American Development Bank. 2012. *The World of Forking Paths: Latin America and the Caribbean in the Face of Global Economic Risks*. Washington, DC: Inter-American Development Bank.

This was an editing mistake and not the fault of the authors. The online version of the paper has been corrected.

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary *JEP* for your e-reader, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $20.00, $30.00, or $40.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of

Economic Perspectives

Summer 2013, Volume 27, Number 3

Symposia

The Top 1 Percent

The Euro

Features

AMERICAN ECONOMIC ASSOCIATION