The Journal of
Economic Perspectives

A journal of the American Economic Association

The Journal of
Economic Perspectives

Contents
Volume 26 • Number 3 • Summer 2012

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.

A Search and Matching Approach to Labor Markets: Did the Natural Rate of Unemployment Rise?

Mary C. Daly, Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta

The increase in the U.S. unemployment rate associated with the 2007–2009 recession is unprecedented during the post–World War II era. The unemployment rate rose by 5.6 percentage points from a low of 4.4 percent in late 2006 and early 2007 to 10.0 percent in October 2009; this exceeds the net increase of 5.2 percentage points between mid-1979 and late 1982 (which spans two recessionary episodes). Moreover, in contrast to relatively rapid labor market recoveries following prior, deeper, post–World War II recessions, two and a half years into this recovery, as of early 2012, the unemployment rate had declined by only about 1.7 percentage points. Such persistently anemic labor market conditions are partly a reflection of the sluggish overall economic recovery, a common occurrence following financial crises (Reinhart and Rogoff 2009; Jordà, Schularick, and Taylor 2011). The lackluster pace of job creation has barely kept up with trend labor force growth and therefore has not generated enough jobs to make a significant dent in the unemployment rate or substantially reduce unemployment duration. Moreover, as we will discuss in more detail later, the unemployment rate has remained high relative to its historical relationship with other business cycle indicators, such as job vacancy rates.

The disconnect between the unemployment rate and other indicators of aggregate economic conditions has raised the concern that rather than being purely cyclical, U.S. unemployment now contains a substantial structural component that will persist even after the U.S. economy has fully recovered from the recession. In

■ *Mary C. Daly is Associate Director of Research and Group Vice President, Bart Hobijn is Senior Research Adviser, and Robert G. Valletta is Research Adviser, all at the Federal Reserve Bank of San Francisco, San Francisco, California. Ayşegül Şahin is Assistant Vice President, Federal Reserve Bank of New York, New York City, New York. Daly is the corresponding author at ⟨mary.daly@sf.frb.org⟩.*

turn, this concern implies that the full employment or "natural" rate of unemployment is now higher than it was before the recession. The distinction is crucial from a policy perspective, because in general, short-run monetary and fiscal stabilization policies are designed to address a cyclical shortfall in labor demand rather than structural factors in the labor market; such structural factors may include a mismatch between workers' skills or geographic locations and employers' labor needs, or the effects of changes in the generosity of social welfare programs. However, understanding how much of the sustained high level of unemployment is cyclical or structural is a challenging task, a point emphasized by Diamond (2011) in his recent Nobel Prize Lecture and illustrated by the wide span of views on the topic held by economists and policymakers.

In this paper, we use a search and matching framework to assess the degree to which the natural rate of unemployment has changed and the reasons underlying any changes. In the first section, we discuss the implications of a standard textbook model of frictional unemployment based on a search and matching framework (Pissarides 2000, chap. 1). This model specifies two curves—the Beveridge curve and the job creation curve—that capture labor supply and labor demand factors, as reflected in the unemployment and job vacancy rates, and that interact to determine equilibrium frictional unemployment. Using this framework, we estimate that the natural rate of unemployment has increased over the recession and recovery, but by far less than unemployment has risen. Our preferred estimate indicates an increase in the natural rate of unemployment of about one percentage point during the recession and its immediate aftermath, putting the current natural rate at around 6 percent. Importantly, even at the maximum of our range of plausible estimates, we find the natural rate increased by only about one and a half percentage points, which would boost the current natural rate to about 6.6 percent. For context, the highest natural rate in the last few decades as estimated by the Congressional Budget Office (2011) was 6.3 percent in 1978.

In the second part of the analysis, we focus on the three primary factors that economists have offered that may account for an increase in the natural rate of unemployment: 1) a *mismatch* between the characteristics of job openings, such as skill requirements or location, and the characteristics of the unemployed; 2) the availability of *extended unemployment insurance* benefits, which may reduce the intensity of job search or cause some recipients to claim they are looking for work, which is a requirement for benefits receipt, when they have in fact effectively left the workforce; and 3) *uncertainty about economic conditions*, which may have induced firms to focus their efforts on raising productivity and output without extensive hiring of new employees.[1] We argue that the increase in mismatch has been quite limited. We find a larger contribution arising from extended unemployment insurance benefits, although this effect is likely to disappear when such benefits are allowed to expire.

---

[1] More broadly, this evidence will draw upon our previous work on the labor market during the recession and recovery: Daly and Hobijn (2010), Elsby, Hobijn, and Şahin (2010), Kwok, Daly, and Hobijn (2010), Valletta and Kuang (2010a, b), as well as Wilson (2010).

Finally, we provide speculative evidence that the unusual degree of uncertainty may be contributing to elevated unemployment through the resulting suppression of hiring; again, this factor would be expected to ease as these uncertainties are resolved.

Overall, we conclude that although the natural rate of unemployment has risen to a moderate degree over the last few years, substantial slack remains in the labor market and is likely to persist for several years. Moreover, since most of the increase in the natural rate appears to be transitory, we expect that as the cyclical recovery in the labor market proceeds, the natural rate will fall back to a value close to its pre-recession level of around 5 percent.

The Equilibrium Natural Rate of Unemployment in a Search Model

The equilibrium natural rate of unemployment is the average rate of unemployment that would prevail in the absence of business cycle fluctuations (Brauer 2007). Underlying the natural rate is frictional unemployment, which reflects the normal time that the unemployed spend in job search, and structural unemployment, which reflects mismatches between employers' labor demand and the skills and geographic location of the unemployed; the two terms are often used synonymously. The natural rate is conceptually similar but not identical to the non-accelerating inflation rate of unemployment, or NAIRU, which defines equilibrium unemployment as the rate at which price inflation is not changing.

The concept of a natural rate was originally introduced by Milton Friedman (1968) and Edmond Phelps (1968) as a way to distinguish between cyclical swings in unemployment that monetary policy can affect and structural changes that it cannot. Under the standard neoclassical assumption of fully flexible prices for factors of production and output, the natural rate is primarily determined by the characteristics of workers and the efficiency of the labor market matching process. These factors affect the rate at which jobs are simultaneously created and destroyed, the rate of turnover in particular jobs, and how quickly unemployed workers are matched with vacant positions. Given the severe shock to labor markets in the most recent recession and the unusually vigorous expansion in a narrow set of housing-related and financial sectors that preceded the downturn, it is reasonable to ask whether some of these noncyclical factors have been altered in a way that increases the natural rate of unemployment in either the short or the long term.

Frictional Unemployment in Equilibrium

To assess the factors affecting the unemployment rate in the short run as well as its longer-run level, we rely on the model of equilibrium frictional unemployment from Pissarides (2000, chap. 1). In this model, equilibrium unemployment is determined by the intersection of two curves: the Beveridge curve (BC), which depicts a negative relationship between vacancies (job openings) and the unemployment rate, and the job creation curve (JCC), which reflects employers' decisions to create job openings and can loosely be interpreted as an aggregate labor demand

Figure 1
Determinants of Shifts in Equilibrium Unemployment
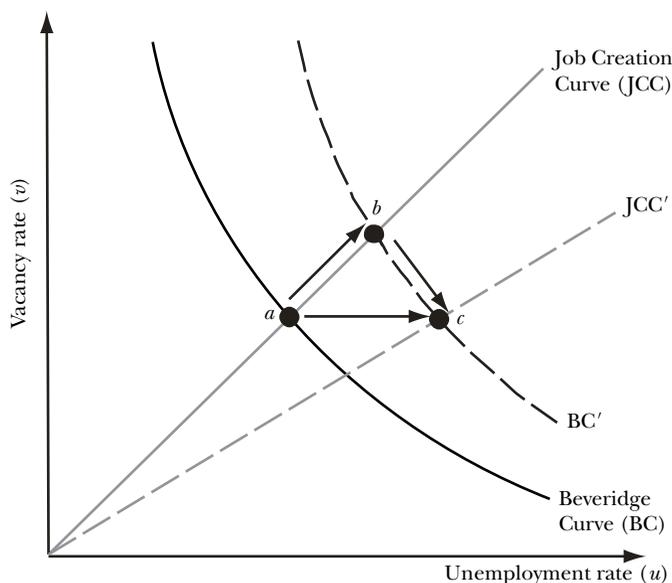
Source: Authors.

curve. We use this framework to analyze the potential increase in the natural rate of unemployment. Daly, Hobijn, and Valletta (2011) offer further details and a formal presentation of the underlying model.

In frictionless models of the labor market, wages adjust to equate labor demand to labor supply in a spot market, which excludes the existence of unemployment as an equilibrium outcome. However, in labor market models with search frictions, not every employer that is looking to hire finds a worker, and not every job searcher finds an employer. Therefore, the labor market does not fully clear in each period, and some job openings remain unfilled at the same time that some job seekers remain unemployed. Because employers and job seekers each benefit from a job match, wages are determined by the bargain between employers and employees over the surplus generated by the match, which occurs after the match.[2] So the equilibrium in this model is defined in terms of vacancies and unemployment—the intersection of the BC and JCC—rather than wages and the equilibrium level of employment. Figure 1 (based on Figure 1.2 in Pissarides 2000, p. 20) depicts a typical BC and JCC relationship.

To understand how the Beveridge curve and job creation curve interact to produce equilibrium vacancy and unemployment rates, it is useful to review the determinants of each curve separately. As implied in the original research of its

---

[2] Assumptions about the type of wage bargaining are important for the cyclical properties of the model (Pissarides 2009) but are not important for the equilibrium concept we focus on here.

namesake (Beveridge 1944) and formalized in subsequent research, the Beveridge curve is essentially a production possibility frontier for the job matching capabilities of the labor market, where the rate at which job seekers are matched to job openings depends primarily on the ratio of the vacancy rate to the unemployment rate (Dow and Dicks-Mireaux 1958; Blanchard and Diamond 1989; Petrongolo and Pissarides 2001). The job vacancy rate is constructed (analogous to the unemployment rate) as a ratio of the number of vacancies to the sum of the total employed plus the number of vacancies. It measures the incidence of open but unfilled jobs in the economy. Movement along the BC reflects cyclical changes in aggregate labor demand: for example, as labor demand weakens, vacancies decline and the unemployment rate rises, causing movement towards the lower right in the diagram. By contrast, an outward shift in the overall position of the BC reflects a decline in the efficiency of the job matching process: for a given level of vacancies, workers have more trouble finding acceptable jobs, and for a given level of unemployment, firms have more trouble finding suitable workers.[3] All else equal, reduced matching efficiency will raise the frictional or structural level of unemployment, hence the natural rate of unemployment.

As Figure 1 shows, the Beveridge curve by itself does not determine an equilibrium combination of vacancies and unemployment. This requires the job creation curve, which is determined by firms' recruiting behavior. Firms hire workers to produce output and will create vacancies up to the point where the expected value of a job match equals the expected search cost to fill the vacancy. The expected value of a job match is determined by the marginal product of labor. The expected search cost combines firms' direct recruiting expenses with the probability that a job is filled.

In the basic model we discuss here, the probability of filling a job rises with the unemployment rate. Thus, the job creation curve is upward sloping, implying that firms create more job openings when unemployment is higher (as depicted in Figure 1). The exact degree of upward slope is affected by other factors that may change over time or across the business cycle, such as the job separation rate, the level of recruiting costs, and the value of jobs (as reflected in worker productivity and the value of output). More generally, the slope of the JCC depends on the structure of the product and labor markets in which firms operate and how they bargain over wages, as well as external factors such as the discount or interest rate.

Changes in the expected value of a job associated with changes in the marginal product of labor can shift the job creation curve (as depicted in Figure 1). This is a channel through which shifts in aggregate demand can affect the unemployment rate even when the efficiency of the job matching process is unchanged. For example, in recessions, declines in aggregate demand reduce the marginal product of labor, which reduces the value of creating jobs. This causes the JCC to rotate

---

[3] Petrongolo and Pissarides (2001) describe the derivation of the Beveridge curve from an underlying job matching technology and discuss functional forms for the job matching function. The BC is typically depicted as convex to the origin, which is consistent with job matching functions that have constant returns to scale in unemployment and vacancies (and hence diminishing returns to either factor with the other one fixed).

down, resulting in a higher unemployment rate with no shift in the Beveridge curve. Although this decline in aggregate demand increases measured unemployment, it does not raise the natural rate of unemployment. Theoretically, the JCC can also shift in response to changes in firm search costs. If the probability of filling a vacancy falls, for example due to a rise in skill mismatch, the JCC will rotate down, indicating a lower rate of vacancies posted for a given job value.

The key implication of this model is that the equilibrium unemployment rate is determined jointly by the intersection of the Beveridge curve and the job creation curve. In this framework, changes in the equilibrium unemployment rate, point *a* in Figure 1, can occur due to an outward shift in the BC, a downward shift in the JCC, or a combination of both. The example in Figure 1 illustrates how shifts in these curves can affect equilibrium unemployment. In the illustration, an outward shift in the BC from BC to BC' shifts equilibrium unemployment from *a* to *b*. Since the JCC is upward sloping, equilibrium unemployment increases by less than the outward shift in the BC. In this model, the unemployment rate can only increase by the same amount as the rightward shift in the BC if the JCC is flat or shifts outward (or down) as in JC'. In these cases, equilibrium unemployment would move from *b* to *c*.

One main message from this graphical illustration is that knowledge of the Beveridge curve is not sufficient to draw conclusions about equilibrium unemployment. As the figure shows, it is not possible to infer, for a given shift in the BC, how much the unemployment rate changes without knowing the shape of the job creation curve. This point may seem obvious, but it has been overlooked in policy discussions in which shifts in the BC are interpreted as one-for-one increases in the natural rate of unemployment.

The insights from this model of equilibrium frictional unemployment point to two directions for empirical analysis. First, to understand the driving forces of the rise in the unemployment rate, one must consider not only what is shifting the Beveridge curve and by how much, but also what is affecting firms' incentives for job creation. Second, to distinguish what part of the rise in the unemployment rate reflects purely cyclical fluctuations in labor demand and what parts are due to other factors, either transitory or permanent, that raise the natural rate, one has to consider what is driving the shifts in the BC and the JCC and how long these effects are likely to last. We consider these in turn.

Empirical Estimates of the Beveridge Curve

Policymakers and analysts who have posited a rise in structural unemployment have largely focused on movements in the empirical Beveridge curve (for example, Benson 2011; Bernanke 2010; Kocherlakota 2010). As noted earlier, this focus is problematic on theoretical grounds, and here we show that empirically, there are at least two difficulties with relying on simple plots of the BC to make inferences about changes in equilibrium unemployment. First, estimates of the shift in the empirical BC suggest that the horizontal shift is not uniform but instead is larger at lower levels of the vacancy rate (as will be demonstrated below). Second, consistent with an upward-sloping job creation curve, past horizontal shifts in the BC have

Figure 2
The U.S. Beveridge Curve, December 2000–November 2011
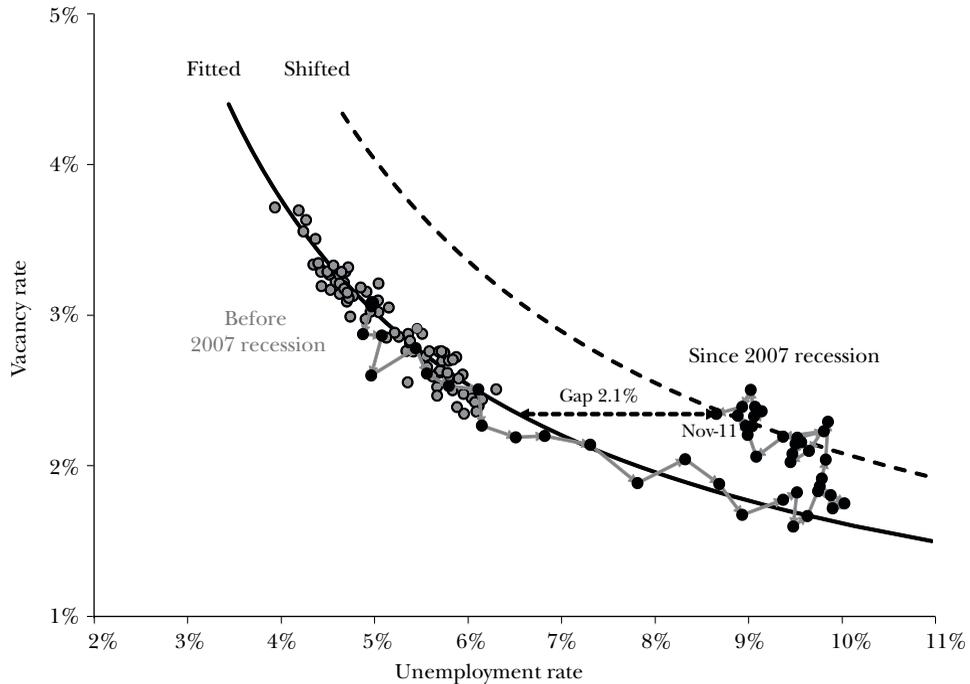
Sources: Job Openings and Labor Turnover Survey (JOLTS), Current Population Survey, and authors' calculations.
Notes: Data are monthly observations. The fitted Beveridge curve is constructed using pre-2007-recession data. The fitted and shifted Beveridge curves are calculated using the methodology introduced in Barnichon, Elsby, Hobijn, and Şahin (2010). The black dots indicate data since the 2007 recession.

been found upon later analysis to coincide with much smaller movements in the estimated natural rate of unemployment (or the twin concept of the NAIRU).[4] This finding underscores the empirical relevance of the other curve in the model—the JCC—which we also analyze empirically below.

Figure 2 displays the empirical Beveridge curve based on data from the U.S. Bureau of Labor Statistics. It combines the official unemployment rate formed from the Bureau's monthly household survey (the Current Population Survey) with data on job openings from the Job Openings and Labor Turnover Survey (JOLTS). The JOLTS is a monthly survey of about 16,000 establishments that was established relatively recently, with data first becoming available in December 2000.

[4] The natural rate is not an observable entity. Rather it is estimated using historical information on the unemployment patterns of demographic subgroups or in the case of the NAIRU, generated from various Phillip's curve models. Real-time estimates of the natural rate are frequently revised as a consensus forms regarding cyclical versus more structural adjustments in the data.

It focuses on job turnover, collecting information on job openings, hires, quits, layoffs and discharges, and other separations. (JOLTS survey information is available online at ⟨http://www.bls.gov/jlt/home.htm⟩.) The Beveridge curve in Figure 2 uses data for December 2000 through November 2011. The data are divided into pre- and post-recession groups with observations occurring prior to the 2007–2009 recession labeled as lighter points and the observations occurring since the recession labeled as darker points. The solid line represents an estimate of the empirical relationship between the unemployment and vacancy rates prior to the onset of the recession, based on the underlying transition rates between the labor force states of employment, unemployment, and out of the labor force (updated from Barnichon, Elsby, Hobijn, and Şahin 2010).

As noted previously, the position of the post-recession points relative to the fitted curve has been interpreted by some observers as evidence of a substantial rightward shift in the Beveridge curve, indicating a higher unemployment rate for a given rate of job vacancies. The logic underlying this interpretation is illustrated in Figure 2. In November 2011, the last point in our sample, the unemployment rate was 8.7 percent. Drawing a horizontal line from that point to the pre-recession, fitted BC produces an unemployment gap of 2.1 percentage points.[5]

However, inferring an outward shift in the Beveridge curve in this manner is problematic for three reasons. First, as the BC shifts right, it also tilts, so that the horizontal shift is not uniform across all levels of the vacancy rate. To see this, consider the shifted curve plotted as the dashed line in Figure 2, which is an update of the estimated shifted BC in Barnichon, Elsby, Hobijn, and Şahin (2010). In this figure, the size of the horizontal shift in the BC varies across the vacancy rate. At the November 2011 vacancy rate of 2.3 percent, the horizontal shift equals 2.1 percentage points; at a vacancy rate of 3.0 percent along the same curve, the horizontal shift is only 1.6 percentage points. The concurrent shifting and tilting implies that estimates of the rightward shift in the BC at low vacancy rates will overstate the size of the shift at higher levels of the vacancy rate.

Second, estimating real-time movements in the Beveridge curve is difficult because the size of the implied shift depends heavily on the specific month chosen. In 2010 and 2011, estimates of the shift obtained in this manner varied from about 1.5 to over 3 percentage points. This large variation in the implied shift occurs because the recently observed points are near a very flat part of the BC, which combines large changes in the unemployment rate with small changes in vacancy rates. In other words, the concavity or flattening of the BC at high unemployment rates means that the outward shift implied by a specific increase in the vacancy rate is not uniform, but rather increases as unemployment rises.

Finally, as noted in earlier research, the movement of vacancies and unemployment back to the stable Beveridge curve following a labor market shock typically follows

---

[5] The size of the imputed current gap is not very sensitive to the estimation method applied. The nonlinear ordinary least squares estimate in Valletta and Kuang (2010b) yields a similar size gap, as does the recalibrated version of Shimer's (2007) Beveridge curve model presented by Kocherlakota (2010).

Figure 3
Historical Shifts in the Beveridge Curve, 1951–2011
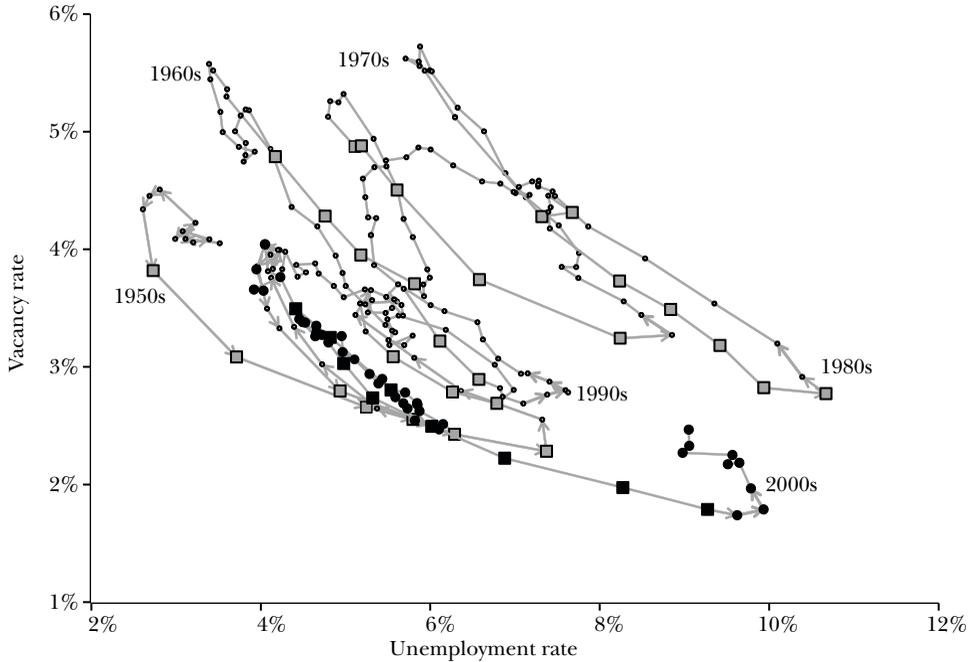
Sources: BLS, Conference Board, Barnichon (2010), and authors' calculations.
Notes: Data are quarterly averages. Recession quarters are squares. The black dots are the 2000s and correspond to the part of the Beveridge curve displayed in Figure 2.

a counterclockwise adjustment pattern (for example, Bowden 1980; Blanchard and Diamond 1989). This pattern occurs because firms can adjust their targeted hiring (job openings) rapidly when labor market conditions improve, but the matching process that will effectively reduce the unemployment rate lags behind the increase in labor demand. As such, the unemployment-vacancy combinations observed in the aftermath of a recession may primarily represent the labor market adjustment process back to a stable BC rather than an outward shift in the BC. For all of these reasons, it is difficult to draw definitive conclusions about shifts in the BC from the pattern of unemployment and vacancy rates observed in the aftermath of the recent severe recession.

Even if we were to take the range of monthly estimates as information about recent shifts in the Beveridge curve, historical comparisons suggest that the recent rightward shift in the BC does not necessarily imply a similarly sized increase in the NAIRU. Using a constructed series on job vacancies created by Barnichon (2010) that combines data from JOLTS with the Help-Wanted Index published by the Conference Board, Figure 3 plots historical BCs for the past six decades. Several facts are worth highlighting. First, the counterclockwise dynamics noted above, in which vacancies adjust more quickly than does unemployment in the aftermath of

recessions, are evident in various past cycles. We labeled this pattern with arrows for each of the recessions in the figure. Second, as can be seen from the groupings of data points by decade as labeled in the figure, the BC has shifted considerably over time. The BC shifted rightward about 4 percentage points between the 1960s and the early 1980s and then shifted back about 2.5 percentage points between 1984 and 1989. Based on this history, the current outward shift of the BC is not unusual, falling within the range of shifts that occurred during past business cycles. Most importantly perhaps, based on estimates now available, the variation in the NAIRU over these periods was much smaller than the horizontal movement in the Beveridge curve. Indeed, estimates of the NAIRU over these earlier periods suggest that it may have changed by about half as much as the shift in the Beveridge curve (for example, Brauer 2007; Orphanides and Williams 2002).

An Estimate of the Long-Run Job Creation Curve

To our knowledge, there are no existing estimates of the historical U.S. job creation curve. We therefore provide a rudimentary estimate here. Although the job creation curve can exhibit short-run movements, we focus on estimating its long-run shape, since we are mainly interested in establishing empirically the relationship between unemployment and job vacancies in the absence of cyclical fluctuations. Our estimate is based on the theoretical relationships discussed earlier which showed that the intersection of the BC and JCC gives us the equilibrium level of frictional unemployment, or the natural rate. Based on this relationship we can use information about the average vacancy rate at various values of the natural rate of unemployment to estimate the natural rate of vacancies, or the vacancy rate in the absence of cyclical fluctuations. This approach essentially takes the historical shifts in the Beveridge curve plotted in Figure 3 and translates them into a long-run job creation curve via our current estimates of the natural rate that prevailed at those times.

The results of this exercise are shown in Figure 4, which plots quarterly observations from the historical vacancy rate series used in Figure 3 against the estimates of the natural rate of unemployment from the Congressional Budget Office. Each of the vertical stacks of points on Figure 4 correspond to one of the Beveridge curves plotted (and given decade labels) in Figure 3. So each point in the vertical stack represents the normal cyclical movements along a given Beveridge curve, or alternatively, the cyclical fluctuations in labor demand for a given natural rate of unemployment. The dotted line shows the relationship between the average level of vacancies and the natural rate of unemployment in the U.S. economy over the sample period (1951:Q1 through 2011:Q3). The line comes from a regression of the historical vacancy rate series on the natural rate of unemployment, using data points observed prior to the recent recession. The regression shows a statistically significant upward-sloping relationship between the average vacancy rate and the natural rate of unemployment. We interpret this relationship as support for the view that the long-run job creation curve is upward sloping.

Of course, our estimate is rudimentary and, like the empirical Beveridge curve, based on data that could itself be mismeasured. For example, there is some

Figure 4
Estimated Long-Run Job Creation Curve
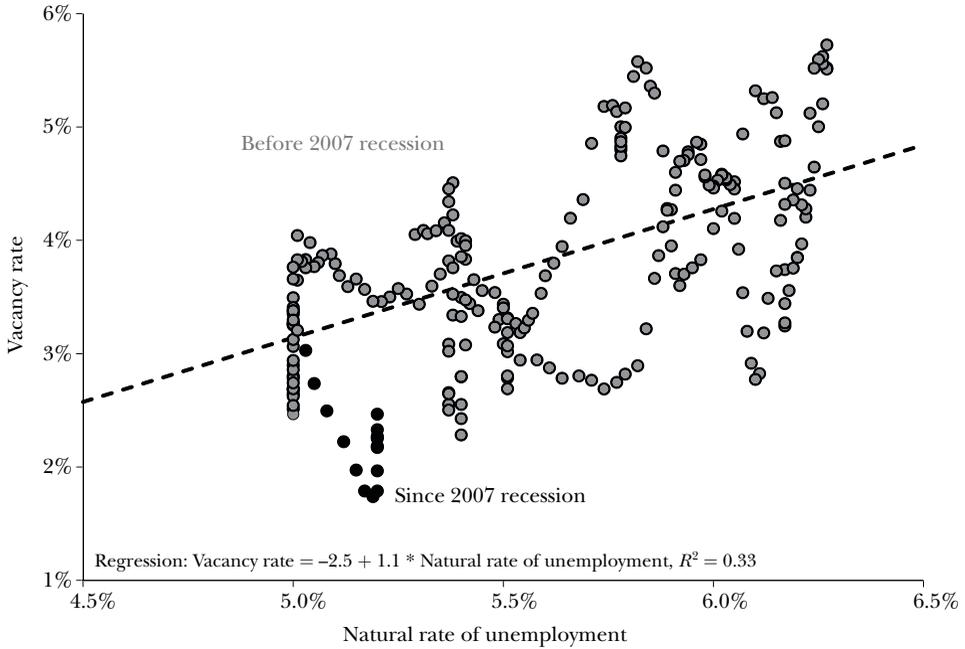
Sources: Bureau of Labor Statistics, Congressional Budget Office, and authors' calculations.
Notes: Seasonally adjusted quarterly data. Regression based on pre-2008 data. The black dots indicate data since the 2007 recession.

disagreement on whether our data accurately capture the historical vacancy rate. Abraham (1987) points out that some of the variation in the Help-Wanted Index data used for the construction of the historical vacancy rate time series reflect a longer-run trend due to the occupational mix of job openings, the consolidation in the newspaper industry, and the increased requirements to post job openings for Equal Employment Opportunity purposes. These factors likely drove up the index relative to the actual number of vacancies during the period of rising unemployment in the 1970s and 1980s, which might lead to an overestimate of the slope of the long-run job creation curve. Estimates of the natural rate of unemployment also can vary. Alternative estimates such as those computed by Orphanides and Williams (2002), which allow for greater time variation in the natural rate than the Congressional Budget Office estimate, produce a slightly flatter JCC.

Putting the Empirical Beveridge Curve and Job Creation Curve Together
    To estimate the potential increase in the natural rate of unemployment following the most recent recession, we combine our estimates of the pre-recession and shifted Beveridge curves from Figure 2 with the estimated long-run job creation

Figure 5
Estimated Job Creation and Beveridge Curves
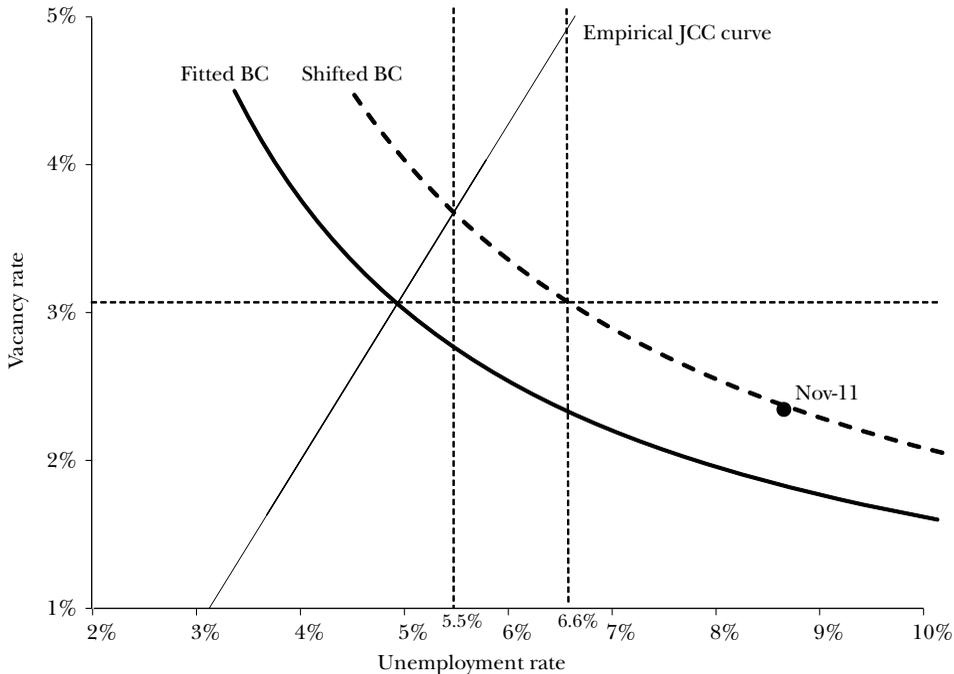
Sources: Job Openings and Labor Turnover Survey (JOLTS), Current Population Survey, Congressional Budget Office, and authors' calculations.
Notes: "BC" is "Beveridge curve." "JCC" is "job creation curve."

curve from Figure 4. The results of this combination are presented in Figure 5. As the figure shows, the empirical long-run JCC intersects the fitted pre-recession Beveridge curve at slightly below a 5 percent unemployment rate, which is very close to the Congressional Budget Office estimate of the pre-recession level of the natural rate. The vacancy rate that coincides with this unemployment rate on the fitted Beveridge curve is 3.1 percent. The shifted BC and the empirical long-run JCC intersect at an unemployment rate of 5.5 percent. If one were to use alternative time-varying estimates of the natural rate of unemployment, the estimated job creation curve would be flatter, and the estimate of the natural rate would increase. For this reason, we interpret the 5.5 percent estimate as a lower bound on the current natural rate of unemployment. Conversely, if the job creation curve is flat, then the increase in the natural rate is given by the 1.6 percentage point horizontal shift in the Beveridge curve at the 3.1 percent vacancy rate. So our upper-bound estimate of the natural rate of unemployment is 6.6 percent. Thus, we find that if the currently estimated shift in the Beveridge curve is permanent and the economy returns to its long-run job creation curve, then the long-run natural rate of unemployment has increased from its 5 percent level in 2007 to somewhere between 5.5

and 6.6 percent as of November 2011.[6] In the absence of additional evidence to pin down its exact value, we regard 6 percent, the approximate midpoint of this range, as our preferred estimate of the current long-run natural rate of unemployment. According to our estimate of the shifted BC, this 6 percent natural rate of unemployment corresponds to a new natural vacancy rate of 3.3 percent. This is substantially higher than the 3.0 percent natural vacancy rate associated with the pre-recession natural rate of unemployment and fitted Beveridge curve.

Implications for Potential GDP

In November 2011, the unemployment rate was 8.7 percent, 2.7 percentage points above our estimate of the new natural rate, while the vacancy rate was 2.3 percent, 1.0 percentage point lower than the new natural vacancy rate. This 2.7 percentage point unemployment gap, by definition, reflects an ongoing cyclical shortfall in the demand for labor associated with the recent recession.

This shortfall in labor demand, or unemployment gap, can be mapped into a measure of the shortfall in the level of overall economic activity relative to the level that would have occurred in the absence of the business cycle. The latter measure is known as potential GDP and the percentage difference between actual GDP and potential GDP is known as the output gap. The relatively stable statistical relationship between these unemployment and output gaps over a long historical period is known as Okun's Law, named after Arthur Okun (1962). Figure 6 shows Okun's Law based on the output and unemployment gaps implied by the Congressional Budget Office's (2011) historical estimates of potential GDP and the natural rate of unemployment. The figure suggests that, as a reasonable rule of thumb, for every percentage point that the unemployment rate exceeds its natural rate, GDP drops two percentage points below its potential.

During the recession and in 2009 and 2010, we saw historically large deviations from the unemployment/GDP relationship implied by Okun's Law, with the unemployment rate being as much as a percentage point higher than implied by the GDP gap. This was reflected in high average labor productivity growth during that period (Daly and Hobijn 2010). However, the decline in the unemployment rate in the first quarter of 2011, combined with revised and slower GDP growth, has brought the unemployment and output gaps back in line with the historical Okun's Law relationship.

Our analysis above suggested that the natural rate of unemployment is likely to be about a percentage point higher than the Congressional Budget Office estimate used in Figure 6. In that case, the unemployment gap would be a percentage point lower as well and, to be in line with Okun's Law, potential GDP would be about

---

[6] As the recovery has proceeded, our estimated range for the natural rate has been falling. For example, in January 2011 we estimated the range to be bounded at 6.9 percent rather than the 6.6 percent we find currently. We interpret this decline as partially reflecting data revisions to JOLTS and partially reflecting the evolution of unemployment and vacancies in 2011. The latter of these two points highlights the transitory nature of recent changes in the natural rate, which we discuss in more detail below.

Figure 6
Okun's Law
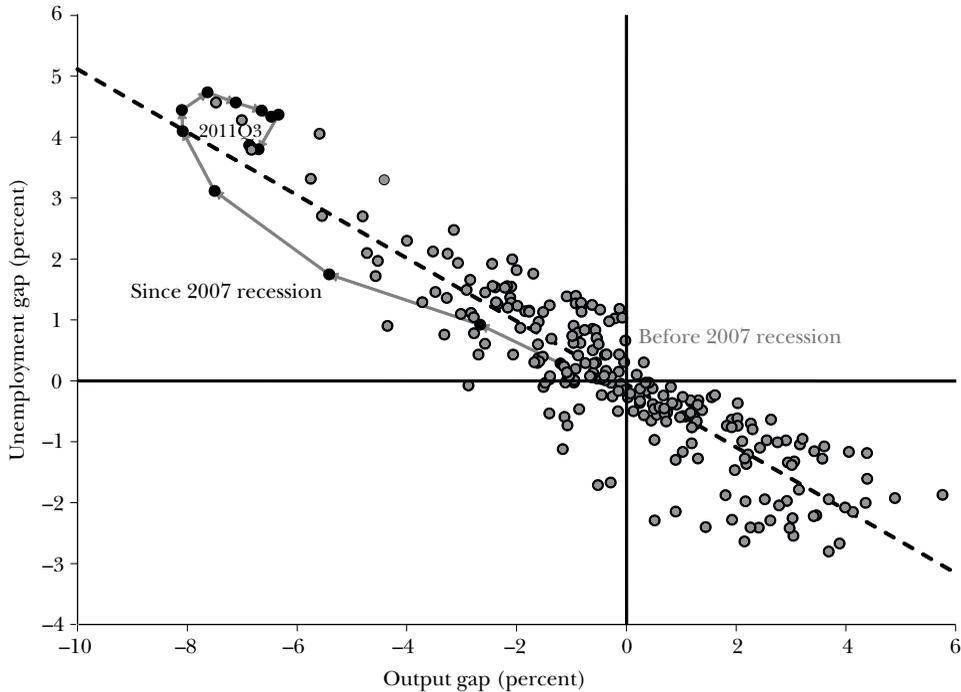
Note: The black dots indicate data since the 2007 recession.

2 percent less than the current CBO estimate, which amounts to $332 billion of annual GDP. The corrected output gap in 2011Q3 would be 4.9 percent rather than 6.9 percent.

In the context of the IS-LM/AS-AD framework that is often used in textbooks on macroeconomics (for example, Abel, Bernanke, and Croushore 2011, chap. 9), this conclusion implies that the shortfall of actual GDP relative to its full employment level, often referred to as "economic slack," is less than the CBO estimates. However, even if the output gap since the beginning of the recession was 2 percentage points lower than currently estimated by the Congressional Budget Office, this recession would still be the second-deepest of the last 60 years, after that of the early 1980s.

What is Shifting the Beveridge and Job Creation Curves?

What factors affect the positions of the Beveridge and job creation curves? Table 1 lists the five factors we consider. The factors are divided into two groups: cyclical factors that cause a shortfall in aggregate demand and higher layoff rates,

Table 1
Factors that Move the Beveridge Curve (BC) and the Job Creation Curve (JCC)

Shifter	JCC	BC	Transitory or Permanent
Cyclical factors			
Shortfall in aggregate demand	⇓		Transitory
Elevated layoffs rate	⇓	⇒	Transitory
Structural/Noncyclical factors			
Decrease in match efficiency (mismatch)	⇓	⇒	Mostly transitory
Increased generosity of unemployment insurance	⇓	⇒	Transitory
Uncertainty	⇓	⇒	Transitory

and structural or noncyclical factors that have a persistent effect on the natural rate of unemployment. In our discussion below, we ignore the first two (cyclical) factors in the table and focus on the other factors because our intent is to identify changes in the natural rate of unemployment that are independent of persistent shortfalls in aggregate demand. Weak aggregate demand drives the shortfall in labor demand that depresses job creation and, for a given Beveridge curve, generates cyclical movements along a fixed Beveridge curve. Elevated layoffs are closely related to weak aggregate demand since layoffs typically rise when aggregate demand weakens. An increase in the rate of layoffs can cause an outward shift in the Beveridge curve, suggesting that an increase in layoffs contributed to the shifts that we estimated above. However, variation in layoff rates tend to play an important role during the onset of recessions, when labor demand plunges, but a limited role during recoveries, when labor demand improves.[7] Since the measured layoffs rate from the JOLTS data has returned to its pre-recession levels, we do not consider them a concern for the labor market recovery going forward and therefore do not address them here.

In the remainder of this section we provide recent empirical evidence on the potential importance of each of the structural/noncyclical factors in Table 1 and discuss whether these factors are likely to be transitory or permanent.

Mismatch

The mismatch argument for sustained increases in the unemployment rate and the natural rate of unemployment is predicated on imbalances in labor supply and demand across industry sectors, geographic areas, or skill groups. Of course, labor markets always display a certain degree of mismatch—or else all job vacancies would fill immediately. However, any rise in mismatch above its usual level makes it harder than usual for workers to find a job and more expensive for firms to fill a vacancy. The result is a decline in match efficiency that both shifts the Beveridge

---

[7] For U.S. evidence on this point, see among others, Darby, Haltiwanger, and Plant (1985, 1986) and Fujita and Ramey (2009). Elsby, Hobijn, and Şahin (2008) show that this is also true across countries.

Figure 7
Industry Mismatch
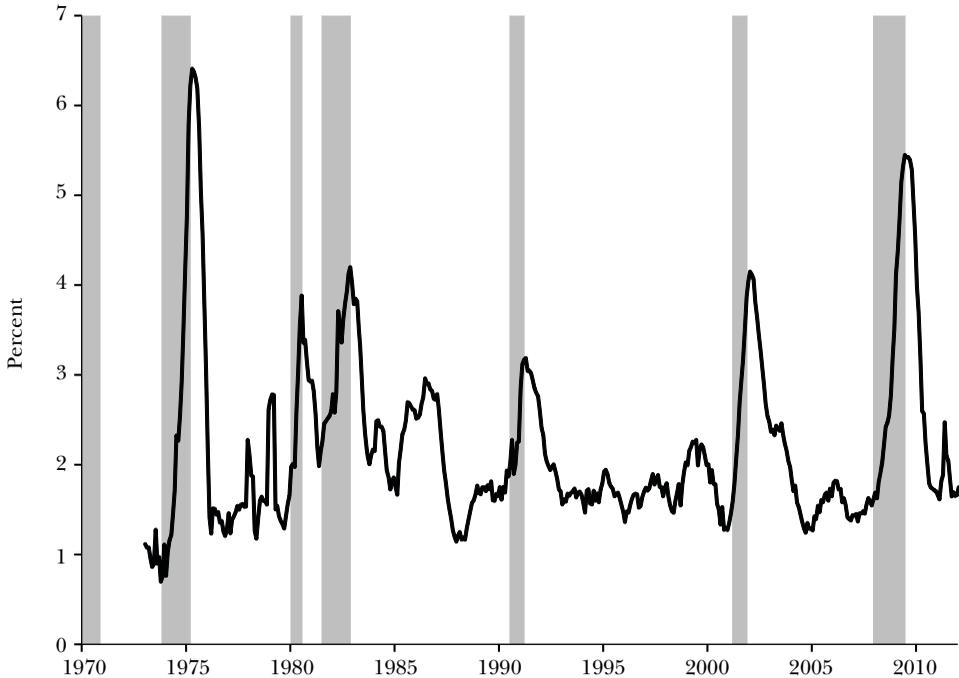
Sources: Bureau of Labor Statistics and authors' calculations.
Notes: The *y*-axis shows the standard deviation of payroll employment growth (12-month change) across 13 major industry categories, weighted by industry employment shares. The grey bars indicate recessions.

curve out and the job creation curve down. In the case of the BC, mismatch makes it harder to form matches, moving the curve out. In the case of the JCC, mismatch increases the search cost to firms for a given job value, pushing the curve down. Mismatch is often regarded as a main potential cause of a long-run increase in the natural rate since training or relocating workers and jobs take a substantial amount of time.

A highly uneven distribution of job gains and losses across industry sectors and states is an indication of mismatch in the sense that it suggests that those who are unemployed did not work in industries and regions where hiring is taking place. As shown in Figure 7, which displays the standard deviation of the rate of payroll employment growth across 13 broad industry sectors that span the complete work-force, the dispersion of employment gains and losses across industries and states spiked in the most recent recession. This pattern is similar to past recessions, and in fact the dispersion of employment gains and losses peaked at a lower level in the recent recession than in the recession of the mid-1970s. Moreover, as aggregate employment stabilized and has started to grow again, the dispersion of employment gains and losses across industries and states has returned to its pre-recession

level. This suggests very little sectoral imbalance in employment growth during the nascent recovery. Valletta and Kuang (2010b) show that dispersion in employment growth *across states* also has returned to its pre-recession level.

Even though the dispersion of employment gains and losses across industries and states has declined substantially, a large number of unemployed workers remain who previously held jobs in sectors like construction and financial activities. Since these sectors will probably take a fair amount of time to return to their pre-recession employment levels, these workers might suffer from prolonged spells of unemployment due to skill mismatch.

To address this possibility, Şahin, Song, Topa, and Violante (2011) introduce mismatch indices that combine measures of both labor demand and labor supply. For labor demand, they use vacancy data from the Job Openings and Labor Turnover Survey and the Conference Board's Help Wanted OnLine database, while for labor supply they rely on unemployment measures from the Current Population Survey. These indices show that both sectoral as well as occupational mismatch increased during the recession but geographic mismatch across states remained relatively low. At the industry level, this increase can be traced back to the construction, durable goods manufacturing, health services, and education sectors. Occupational mismatch rose mostly due to the construction, production work, healthcare, and sales-related occupations. Şahin, Song, Topa, and Violante (2011) also quantify how much of the recent rise in U.S. unemployment is due to an increase in mismatch and find that higher mismatch across industries and occupations accounts for 0.6 to 1.7 percentage points of the recent rise in the unemployment rate. Geographical mismatch turns out to be quantitatively insignificant.[8]

These findings linking mismatch to the observed unemployment rate do not necessarily imply that the underlying natural rate of unemployment increased by the same amount as the contribution of mismatch to the rise in the unemployment rate. Just like the dispersion measures considered in Valletta and Kuang (2010b), the mismatch indices constructed by Şahin, Song, Topa, and Violante (2011) rose during the recession and then started to decline in 2010. Thus far, the evidence suggests that mismatch has had a pronounced cyclical component, moving together with the unemployment rate. While mismatch has contributed to the increase in the unemployment rate, its current path suggests that it is not likely to cause a large long-lasting increase in the natural rate of unemployment.

We do expect a modest increase in the natural rate due to the contraction of the construction sector. A simple back-of-the-envelope calculation also supports this view. The seasonally adjusted unemployment rate for construction workers has been hovering in the range of 15 to 20 percent during the recovery, compared

[8] This result for geographic mismatch is consistent with recent empirical papers, most notably Molloy, Smith, and Wozniak (2011) and Schulhofer-Wohl (2010), and Valletta (2010) which all find a very limited role for geographic immobility of unemployed individuals whose home values have fallen below the amount owed on their mortgages ("house lock"). Recent theoretical work by Sterk (2010) suggests that although house lock will lead to an outward shift in the Beveridge curve, the likely shift is much smaller than the one depicted in Figure 2.

with a more typical rate from 2003 to 2006 of about 7 to 8 percent. This represents about 1.25 million more unemployed construction workers in the current recovery than was typical during the preceding expansion. Assuming that half of them are re-employable in other industries—a plausible estimate given the recent evidence on industry mobility of workers (for example, Bjelland, Fallick, Haltiwanger, and McEntarfer 2010)—the decline of construction would cause structural unemployment to increase by only about 0.4 percentage point. Because most construction workers are not hired through formal job openings, we expect the effect of this type of mismatch on the long-run job creation curve to be limited. Instead, we regard this effect of mismatch on the natural rate of unemployment as mainly due to the persistent part of the contribution of the construction sector to the outward shift of the Beveridge curve calculated by Barnichon, Elsby, Hobijn, and Şahin (2010).

Extended Unemployment Benefits

Extensions of unemployment insurance are a standard policy response to elevated cyclical unemployment, and the sharp increase in the unemployment rate during the 2007–2009 recession resulted in an unprecedented increase in the potential duration of receipt for unemployment benefits. Beginning in June 2008, the maximum duration of unemployment insurance benefits was extended multiple times, reaching 99 weeks for most job seekers eligible for unemployment insurance as of late 2009.[9] Congress has allowed the primary extension program to expire twice, most notably for nearly two months in June–July of 2010, but in each case renewed the extensions, which as of this writing are effective through early March 2012.

In the context of the job matching function described earlier, increased availability of unemployment insurance benefits is likely to increase the duration of unemployment through two primary behavioral channels. First, the extension of unemployment insurance benefits, which represents an increase in their value, may reduce the intensity with which unemployed individuals eligible for these benefits search for work, along with their likelihood of accepting a given job offer. This occurs because the additional unemployment insurance benefits reduce the net gains from finding a job and also serve as an income cushion that helps households maintain acceptable consumption levels in the face of unemployment shocks (Chetty 2008). Alternatively, the measured unemployment rate may be artificially inflated because some individuals who are not actively searching for work are identifying themselves as active searchers in order to receive unemployment insurance benefits (a "reporting effect," in the language of Card, Chetty, and Weber 2007). These behavioral effects on job search will increase the noncyclical or structural

---

[9] The joint federal-state unemployment insurance program provides up to 26 weeks of normal benefits. The recent benefit extensions reflect the impact of two federally funded programs: the permanently authorized Extended Benefits program, which provides up to 20 additional weeks of benefits, and the special Emergency Unemployment Compensation, which provides up to 53 weeks of benefits, depending on the unemployment rate in the recipient's state of prior employment (which causes the share of unemployed workers eligible for the 99-week maximum to change over time). The previous maximum eligibility was 65 weeks under the Federal Supplemental Benefits program in the mid-1970s.

component of the unemployment rate during the period over which extended benefits are available.[10]

Assessing the magnitude of the extended unemployment insurance effect is challenging—and the challenge is even more difficult under the unusually weak labor market conditions of the last few years, which by themselves have led to unusually long unemployment durations. Based on existing empirical research using U.S. data, Chetty (2008) noted that a 10 percent increase in the overall value of unemployment insurance benefits increases unemployment durations by 4–8 percent. Other estimates, particularly those that focus on extension periods rather than the dollar value of benefits, lie below this range (for example, Card and Levine 2000). Thus, a wide range of uncertainty exists around the impact of unemployment insurance extensions on the duration of unemployment. Moreover, as noted by others (for example, Katz 2010), the effect of unemployment insurance benefits on job search likely was higher in the 1970s and 1980s than it is now, due to the earlier period's greater reliance on temporary layoffs and the corresponding sensitivity of recall dates to unemployment insurance benefits. As such, reliance on past estimates of the effects of the generosity of unemployment insurance benefits on the duration on unemployment is likely to lead to overestimates in the current economic environment.

Our own empirical assessment, reported in Daly, Hobijn, and Valletta (2011) and based on the methodology introduced in Valletta and Kuang (2010a), focuses on direct calculation and comparison of the duration of unemployment for individuals who are eligible or not eligible for unemployment insurance benefits, as reflected in their reported reason for unemployment. The receipt of unemployment insurance benefits generally is restricted to individuals who are unemployed through "no fault of their own," to quote U.S. Labor Department eligibility guidelines, and have recent employment history that allows them to meet a base earnings test. In terms of the data on cause of unemployment from the Current Population Survey, individuals who are eligible for unemployment insurance are concentrated among the unemployed who classify themselves as "job losers," while those who are ineligible tend to be voluntary job leavers and labor force entrants. The eligible (job losers) group accounted for about two-thirds of the unemployed as of late 2009, which was very close to the share of actual unemployment insurance recipients in overall unemployment.

During the recent recession and its aftermath, unemployment durations rose substantially from their pre-recession baseline levels, both for those who are eligible for unemployment insurance and for others. According to the measure of expected completed duration used by Valletta and Kuang (2010a), duration approximately doubled, from about 18 weeks to about 35 weeks. However, the increase in duration

---

[10] Our narrow focus on the direct behavioral effects of extended unemployment insurance ignores the potential aggregate demand stimulus provided by such benefits, which reduces the cyclical component of the unemployment rate but does not affect the level of structural unemployment. Some recent research suggests that multiplier effects of normal and extended unemployment insurance payments may be quite large (for example, Vroman 2010). It is possible that the reduction in cyclical unemployment from this channel may exceed the increase in the structural component from the micro-behavioral channel.

of unemployment was larger for those eligible for unemployment benefits, by about 3½ weeks. If one attributes all of this difference to eligibility for extended unemployment insurance benefits, which is the primary factor that differentially affected eligible and ineligible individuals during the recession, then the extension of unemployment insurance raised the unemployment rate by about 0.8 percentage points. These results are relatively insensitive to alternative assumptions about the relationship between the stated reason for unemployment in the Current Population Survey data and likely eligibility for unemployment insurance. Other recent formal estimates of the effect of extended unemployment insurance benefits on the natural rate of unemployment are generally smaller, ranging from about 0.1 to 0.7 percentage points (Aaronson, Mazumder, and Schecter 2010; Farber and Valletta 2011; Rothstein, forthcoming) up to a maximum of 1.2 percentage points (Fujita 2011).

The effect of extended unemployment insurance on the unemployment rate is expected to dissipate as labor market conditions improve and the extended unemployment insurance provisions expire. As a result, the extensions of unemployment insurance do not affect the long-run job creation curve.[11] As extended unemployment insurance provisions expire, the shifted Beveridge curve is expected to move back inwards.

Uncertainty

Extensive anecdotal evidence suggests that the severity and persistence of the recession and associated financial crisis, combined with significant changes in federal government policies such as the Dodd–Frank financial reform bill, the Patient Protection and Affordable Care Act, along with potential changes in energy policy have increased firms' uncertainty about the environment in which they are operating. In a labor market search framework with fixed hiring and firing costs, such uncertainty about the future state of aggregate demand lowers the option value of hiring new workers, thereby putting downward pressure on job creation (Bentolila and Bertola 1990; Bloom 2009). According to such models, firms may choose to incur the fixed cost of investing in workers based on the option value of using them when they are needed for production. As uncertainty about future demand increases, the option value of making this up-front investment is reduced and fewer workers are demanded. In this way, uncertainty about economic conditions and policy might contribute to the outward shift in the Beveridge curve and, more importantly, to the low number of vacancies firms are posting.

Theoretical models of jobless recoveries, like Van Rens (2004) and Koenders and Rogerson (2005), suggest that firms might postpone hiring by temporarily boosting productivity growth. In Van Rens (2004), faster productivity growth comes from moving workers from the production of intangibles to the production of

---

[11] Extended unemployment insurance might raise current reservation wages and thus suppress job creation in the short run. No quantitative analysis of this short-run effect exists. However, we expect this short-run effect of extended unemployment insurance on the job creation curve to be small, because it is offset by the aggregate demand stimulus provided by unemployment insurance payments.

measured output. In Koenders and Rogerson (2005), firms choose to adopt organizational changes that improve productivity but were temporarily shelved during the prior expansion. In either case, the reorientation of production activity reduces the rate of hiring but raises productivity growth. Effects of this sort might have driven the significant deviation from Okun's Law in Figure 6 during 2009 and 2010 (Daly and Hobijn 2010). However, such temporary measures only raise productivity growth in the short run. If uncertainty remains elevated, the effect of these measures on productivity growth is likely to diminish and uncertainty will mainly reduce job creation. This pattern is consistent with the combination of low productivity growth and low job creation in the first half of 2011.

Uncertainty might also cause firms that create vacancies to become more selective about filling them. Such a change in firms' hiring decisions would cause a decline in the number of hires per vacancy; this is consistent with a reduction in recruiting intensity identified by Davis, Faberman, and Haltiwanger (2010).

Though the high level of uncertainty is a possible explanation for the joint weakness in vacancy creation and vacancy yields (hires per vacancy) relative to the strong productivity growth during the first part of the recovery, we know of no studies that have tried to quantify this effect. Since we expect firms' uncertainty about the economic environment to dissipate as the recovery persists and gains momentum, we anticipate that any increase in the natural rate of unemployment arising from uncertainty is likely to be temporary rather than permanent.

Conclusion

The stubbornly high rate of unemployment in the face of ongoing GDP growth and rising job openings has raised concerns that the level of structural unemployment, or the natural rate of unemployment, has risen over the past few years in the United States. This possibility raises important policy issues since short-run monetary and fiscal stabilization policies are not designed to alleviate structural unemployment and can be costly if misapplied. Our estimates suggest that the natural rate of unemployment has risen from its pre-recession level of 5.0 percent to a value between 5.5 and 6.6 percent, with our preferred estimate lying at the midpoint of approximately 6 percent. This value implies an unemployment gap of over 2.7 percentage points in late 2011, which remains quite high. Thus, even with a higher natural rate of unemployment, considerable slack remains in the labor market.

There are a number of unanswered questions for researchers in this area. Our analysis relied on a rudimentary formulation and estimation of the job creation curve that relates firms' decisions about job vacancies to the level of unemployment. There may be factors that we have not identified or measured well that permanently restrain vacancy and job creation going forward. Perhaps most vexing, the 2007–2009 recession is now the third successive one in which the U.S. economy has experienced a jobless recovery (that is, the rate of unemployment has remained

high for years after the recession is deemed to have ended). It is not yet clear how to apply the common terminology of cyclical and structural unemployment to this phenomenon. Is the U.S. economy now experiencing greater fluctuations in structural unemployment than in the 1960s, 1970s, and 1980s? Or is it experiencing longer-run bouts of cyclical unemployment than in those decades? A better understanding of the determinants of job creation in the aftermath of recession is crucial for improving the empirical analysis of equilibrium models of frictional unemployment, and it also holds promise for improving labor market policies aimed at combating jobless recoveries.

References

Aaronson, Daniel, Bhashkar Mazumder, and Shani Schechter. 2010. "What is Behind the Rise in Long-Term Unemployment?" Federal Reserve Bank of Chicago *Economic Perspectives* 34(2nd Quarter): 28–51.

Abel, Andrew B., Ben S. Bernanke, and Dean Croushore. 2011. *Macroeconomics,* 7th edition. New York: Pearson.

Abraham, Katherine G. 1987. "Help Wanted Advertising, Job Vacancies and Unemployment." *Brookings Papers on Economic Activity,* no. 1, 207–243.

Barnichon, Regis. 2010. "Building a Composite Help-Wanted Index." *Economics Letters* 109(3): 175–78.

Barnichon, Regis, Michael Elsby, Bart Hobijn, and Ayşegül Şahin. 2010. "Which Industries are Shifting the Beveridge Curve?" Federal Reserve Bank of San Francisco Working Paper 2010-32.

Benson, David. 2011. "Macroeconomic Policy and Labor Markets: Lessons from Dale Mortensen's Research." *Chicago Fed Letter,* August 2011.

Bentolila, Samuel, and Giuseppi Bertola. 1990. "Firing Costs and Labour Demand: How Bad is Eurosclerosis?" *The Review of Economic Studies* 57(3): 381–402.

Bernanke, Ben S. 2010. "Monetary Policy Objectives and Tools in a Low-Inflation Environment." Remarks at the "Revisiting Monetary Policy in a Low-Inflation Environment" Conference sponsored by the Federal Reserve Bank of Boston, October 15, 2010.

Beveridge, William. 1944. *Full Employment in a Free Society.* London: George Allen and Unwin.

Bjelland, Melissa, Bruce Fallick, John C. Haltiwanger, and Erika McEntarfer. 2010. "Employer-to-Employer Flows in the United States: Estimates Using Linked Employer-Employee Data." Center for Economic Studies Working Paper 10-26.

Blanchard, Olivier Jean, and Peter Diamond. 1989. "The Beveridge Curve." *Brookings Papers on Economic Activity,* no. 1, pp. 1–76.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77(3): 623–685.

Bowden, Robert J. 1980. "On the Existence and Secular Stability of u–v Loci." *Economica* 47(185): 35–50.

Brauer, David. 2007. "The Natural Rate of Unemployment." CBO Working Paper 2007-06.

Card, David, Raj Chetty, and Andrea Weber. 2007. "The Spike at Benefit Exhaustion: Leaving

the Unemployment System or Starting a New Job?" *American Economic Review* 97(2): 113–18.

Card, David, and Phillip B. Levine. 2000. "Extended Benefits and the Duration of UI Spells: Evidence from the New Jersey Extended Benefit Program." *Journal of Public Economics* 78(1–2): 107–138.

Chetty, Raj. 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy* 116(2): 173–234.

Congressional Budget Office. 2011. *The Budget and Economic Outlook: An Update.* August 2011.

Daly, Mary C., and Bart Hobijn. 2010. "Okun's Law and the Unemployment Surprise of 2009." *FRBSF Economic Letter 2010-07,* Federal Reserve Bank of San Francisco.

Daly, Mary C., Bart Hobijn, and Rob Valletta. 2011. "The Recent Evolution of the Natural Rate of Unemployment." Federal Reserve Bank of San Francisco Working Paper 2011-05.

Darby, Michael R., John C. Haltiwanger, and Mark W. Plant. 1985. "Unemployment Rate Dynamics and Persistent Unemployment under Rational Expectations." *American Economic Review* 75(4): 614–37.

Darby, Michael R., John C. Haltiwanger, and Mark W. Plant. 1986. "The Ins and Outs of Unemployment: The Ins Win." NBER Working Paper 1997.

Davis, Steven J., Jason Faberman, and John C. Haltiwanger. 2010. "The Establishment-Level Behavior of Vacancies and Hiring." NBER Working Paper 16265.

Diamond, Peter. 2011. "Unemployment, Vacancies, Wages." Lecture presented for the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, Stockholm, Sweden. http://econ -www.mit.edu/files/6574.

Dow, J. C. R., and L. A. Dicks-Mireaux. 1958. "The Excess Demand for Labour: A Study of Conditions in Great Britain, 1946–56." *Oxford Economic Papers* 10(1): 1–33.

Elsby, Michael, Bart Hobijn, and Ayşegül Şahin. 2008. "Unemployment Dynamics in the OECD." NBER Working Paper 14617.

Elsby, Michael, Bart Hobijn, and Ayşegül Şahin. 2010. "The Labor Market in the Great Recession." *Brookings Papers on Economic Activity*, no. 1, pp. 1–48.

Farber, Henry S., and Robert G. Valletta. 2011. "Extended Unemployment Insurance and Unemployment Duration in the Great Recession: The U.S. Experience." Unpublished paper, Federal Reserve Bank of San Francisco and Princeton University, June.

Friedman, Milton. 1968. "The Role of Monetary Policy." *American Economic Review* 58(1): 1–17.

Fujita, Shigeru. 2011. "Effects of the UI Benefit Extensions: Evidence from the Monthly CPS." FRB of Philadephia Working Paper 10-35/R, Federal Reserve Bank of Philadelphia, January.

Fujita, Shigeru, and Garey Ramey. 2009. "The Cyclicality of Job Loss and Hiring." *International Economic Review* vol. 50, pp. 415–430.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2011. "When Credit Bites Back: Leverage, Business Cycles, and Crises." NBER Working Paper 17621.

Katz, Lawrence. 2010. "Long-Term Unemployment in the Great Recession." Testimony to the Joint Economic Committee, U.S. Congress, April 29.

Kocherlakota, Narayana. 2010. "Inside the FOMC." Speech at Marquette, Michigan, August 17.

Koenders, Kathryn, and Richard Rogerson. 2005. "Organizational Dynamics over the Business Cycle: A View on Jobless Recoveries." Federal Reserve Bank of St. Louis *Review* 87(4): 555–79.

Kwok, Joyce, Mary C. Daly, and Bart Hobijn. 2010. "Labor Force Participation and the Future Path of Unemployment." *FRBSF Economic Letter* 2010-27, Federal Reserve Bank of San Francisco.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25(3): 173–96.

Okun, Arthur M. 1962. "Potential GNP: Its Measurement and Significance." In *Proceedings of the Business and Economics Statistics Section of the American Statistical Association*, 98–104. Available as Cowles Foundation Paper 190, http://cowles .econ.yale.edu/P/cp/p01b/p0190.pdf.

Orphanides, Athanasios, and John C. Williams. 2002. "Robust Monetary Policy Rules with Unknown Natural Rates." *Brookings Papers on Economic Activity*, no. 2, pp. 63–118.

Petrongolo, Barbara, and Christopher A. Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature* 39(2): 390–431.

Phelps, Edmund S. 1968. "Money-Wage Dynamics and Labor-Market Equilibrium." *Journal of Political Economy* 76(4, Part 2): 678–711.

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

Pissarides, Christopher A. 2009. "The Unemployment Volatility Puzzle: Is Wage Stickiness the Answer?" *Econometrica* 77(5): 1339–69.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. "The Aftermath of Financial Crises," *American Economic Review* 99(2): 466–72.

Rothstein, Jesse. Forthcoming. "Unemployment Insurance and Job Search in the Great Recession." *Brookings Papers on Economic Activity.*

Şahin, Ayşegül, Joseph Song, Giorgio Topa, and Gianluca Violante. 2011. "Measuring Mismatch in the U.S. Labor Market." http://www.ny.frb.org/research/economists/sahin/USmismatch.pdfMimeo.

Schulhofer-Wohl, Sam. 2010. "Negative Equity Does Not Reduce Homeowners' Mobility." Federal Reserve Bank of Minneapolis Working Paper 682.

Shimer, Robert. 2007. "Mismatch." *American Economic Review* 97(4): 1074–1101.

Sterk, Vincent. 2010. "Home Equity, Mobility, and Macroeconomic Fluctuations." DNB Working Paper 265, Netherlands Central Bank.

Valletta, Robert G. 2010. "House Lock and Structural Unemployment." Unpublished paper, Federal Reserve Bank of San Francisco.

Valletta, Robert G., and Katherine Kuang. 2010a. "Extended Unemployment and UI Benefits." *FRBSF Economic Letter* 2010-12, Federal Reserve Bank of San Francisco.

Valletta, Robert G., and Katherine Kuang. 2010b. "Is Structural Unemployment on the Rise?" *FRBSF Economic Letter* 2010-34, Federal Reserve Bank of San Francisco.

Van Rens, Thijs. 2004. "Organizational Capital and Employment Fluctuations." http://www.crei.cat/~vanrens/orgcap/jmp_tvr.pdf.

Vroman, Wayne. 2010. "The Role of Unemployment Insurance as an Automatic Stabilizer during a Recession." Report prepared under a subcontract between Urban Institute and IMPAQ International, July.

Wilson, Daniel J. 2010. "Is the Recent Productivity Boom Over?" *FRBSF Economic Letter* 2010-28, Federal Reserve Bank of San Francisco.

Who Suffers During Recessions?[†]

Hilary Hoynes, Douglas L. Miller, and Jessamyn Schaller

The Great Recession generated large reductions in employment, earnings, and income for workers and families in the United States. The seasonally adjusted unemployment rate increased from 5 percent in December 2007 to 9.5 percent in June 2009, the start and end of the recession according to the National Bureau of Economic Research (NBER, at ⟨http://www.nber.org/cycles.html⟩). From 2007 to 2010, median real family income fell by 6 percent and the poverty rate increased from 12.5 percent to 15.1 percent (DeNavas-Walt, Proctor, and Smith 2011). The recovery since June 2009 has been slow relative to historical averages. In the more than two and a half years since the official start of the recovery, the unemployment rate has fallen by just over a percentage point, reaching 8.3 percent in February 2012. The effects of the Great Recession, however, are not experienced equally by all workers. National statistics can obscure dramatic differences in the severity of the cyclical impacts for different groups. For example, men experienced significantly larger job loss in the Great Recession compared to women, but during the recovery, male employment is picking up more rapidly (Kochhar 2011).

■ *Hilary Hoynes is Professor of Economics and Douglas L. Miller is Associate Professor of Economics, both at the University of California at Davis, Davis, California. Jessamyn Schaller is Assistant Professor at the University of Arizona, Tucson, Arizona. Hoynes is a Research Associate and Miller is a Faculty Research Fellow, both at the National Bureau of Economic Research, Cambridge, Massachusetts. During the 2011–12 academic year, Miller was Visiting Research Scholar, Center for Health and Wellbeing, Princeton University, Princeton, New Jersey. Their email addresses are ⟨hwhoynes@ucdavis.edu⟩, ⟨dlmiller@ucdavis.edu⟩, and ⟨jschaller@eller.arizona.edu⟩.*

We begin this paper with an overview of cyclical fluctuations in unemployment rates and employment from 1979 through 2011. Using national time-series data, we compare the Great Recession to earlier recessions in terms of its severity, duration, and subsequent recovery. We then go on to use individual-level data from January 1979 through December 2011 from the Current Population Survey, Merged Outgoing Rotation Group (CPS-MORG) to measure and illustrate how unemployment and employment have changed in the Great Recession for persons of different ages, educational attainment, race, and gender. After establishing the basic descriptive findings, we estimate a state panel data model to measure the responsiveness of different groups to the state-month unemployment rate. The labor market outcomes we analyze are the groups' employment and unemployment.

Our findings are summarized as follows: First, the labor market decline in the Great Recession is both deeper and longer than the early 1980s recession. Second, the impacts of the Great Recession have been felt most strongly for men, black and Hispanic workers, youth, and low-education workers. Third, these dramatic differences in the cyclicality across demographic groups are remarkably stable across three decades of time and across recessionary periods versus expansionary periods. Fourth, the differences across demographic groups during the 2007 recession are explained to a large extent by variation in the groups' exposure to cycles across industry-occupation.

Our study builds on a large existing literature in labor economics and macroeconomics on how business cycles affect outcomes for workers and families, including our own prior work (Bitler and Hoynes 2010; Hoynes 2000; Hines, Hoynes, and Krueger 2001; Stevens, Miller, Page, and Filipski 2011). Our study makes several contributions to this existing literature. First, our primary focus is identifying differences in the cyclicality across demographic groups. Second, we present the results of statistical tests for differences in the cyclicality both across groups (for a given time period) and over time (for a given group). Third, by using data through the end of 2011, we highlight the results for the Great Recession and compare them to the early 1980s recession. Finally, we compare the recovery periods following the two most severe recessions in our time frame: the recession(s) of the earlier 1980s (counted as one recession) and the 2007–09 recession.

Overview of Labor Market Fluctuations Since 1979

The U.S. economy from 1979 to 2011 has seen five recessions: six months from January 1980 to July 1980; 16 months from July 1981 to November 1982; eight months from July 1990 to March 1991; eight months from March 2001 to November 2001; and 19 months from December 2007 to June 2009. We follow a common practice of combining the back-to-back 1980 and 1981 recessions, and the graphs therefore compare four cycles, designated by the starting years of the recessions as 1980, 1990, 2001, and 2007. To put the labor market dimension of these recessions in context, consider Figures 1 and 2. Following the standard definitions, the percent

Figure 1
U.S. Seasonally Adjusted Unemployment Rate, Months since Peak
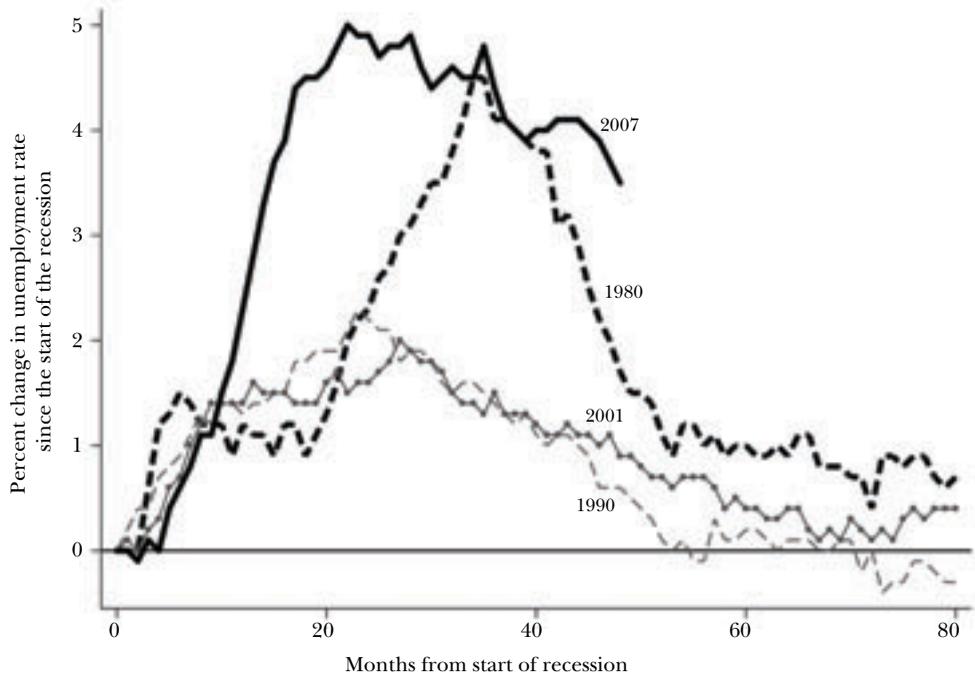
Sources: Current Population Survey (Bureau of Labor Statistics 2012a). Labor market peaks come from NBER (2011).
Note: For the 1980 recession, the recessions beginning in January 1980 and July 1981 are combined into one cycle.

unemployed is among those in the labor force, while the percent employed is among the entire population. We analyze both to capture different margins of behavior.

When discussing the monthly unemployment rate for this and all subsequent analyses in the paper, we present seasonally adjusted measures, which remove the typical variation that takes place within a calendar year. In Figure 1, we plot the percentage point increase in the unemployment rate for these four business cycles by the number of months since the official start of the recession. The paths of the unemployment rate after the 1991 and 2001 recessions were quite similar. After the 1980–82 recessions, unemployment was slower to rise (which may be the result of combining two back-to-back recessions), but after about 48 months, the unemployment rate had dropped sharply. In contrast, the 2007 recession exhibits the steepest and largest increase in the unemployment rate among the four recessions. The unemployment rate rose from 5 percent in December 2007 to a high of 10.1 in October 2009. While the recession officially ended in July 2009, the unemployment rate has remained high. As of December 2011 (the last data point), unemployment rates remain almost 2 points higher, relative to the peak, than at a similar point in the double-dip recession of the early 1980s; however, by comparison with either the January 1980 recession or the

Figure 2
U.S. Seasonally Adjusted Employment, Months since Peak
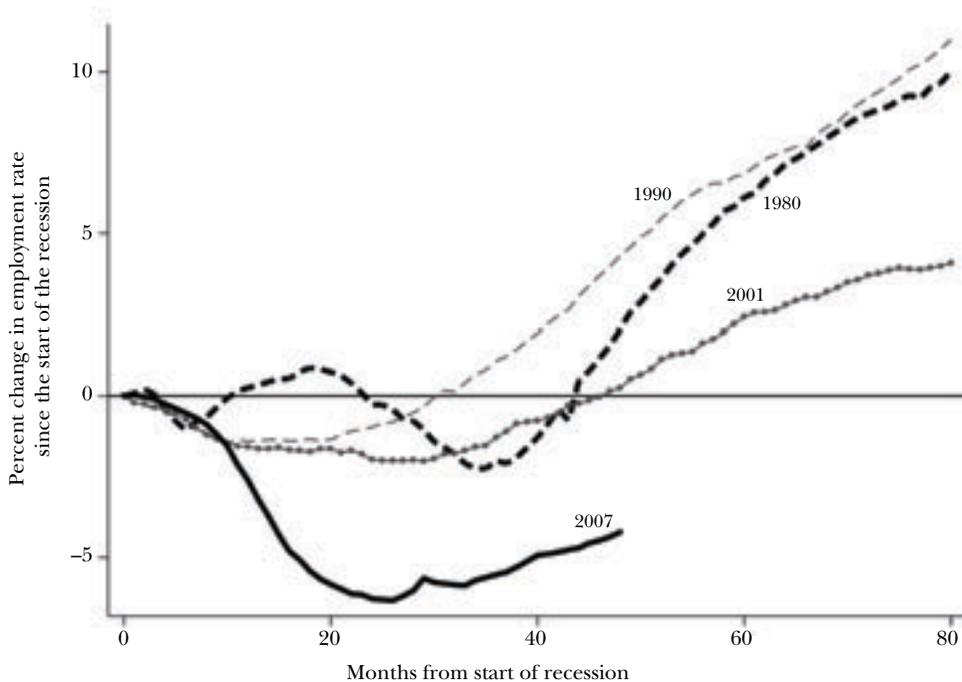
Sources: Current Employment Statistics (Bureau of Labor Statistics 2012b). Labor market peaks come from NBER (2011).
Note: For the 1980 recession, the recessions beginning in January 1980 and July 1981 are combined into one cycle.

July 1981 recession, the increases in unemployment in the current recession appear far more dramatic and long-lasting.

Figure 2 highlights the relatively weak recovery of 2010–2011 by looking at aggregate monthly employment (seasonally adjusted). This figure shows the percent change in employment compared with the employment level of the first month of each of the four recessions. The magnitude of the fall in the employment level is comparable in the 1980, 1991, and 2001 recessions; and employment falls much more severely in the 2007 recession. In the timing of the recovery of job growth, by 48 months since the beginning of the 2007 recession (where our data end), employment had returned to its prerecession level in the three previous cycles. We are far from that in the Great Recession.

Many earlier studies have examined the effect of business cycles on labor market outcomes. Research on the Great Recession has confirmed that, across demographic groups, the decline in labor market outcomes since 2007 has been worse than any other recession in the postwar period (Goodman and Mance 2011). As in previous recessions, evidence suggests that the effects of the recent downturn have been born disproportionately by racial and ethnic minorities and by male,

younger, and less-educated workers (Elsby, Hobijn, and Şahin 2010; Farber 2011; Kochhar, Fry, and Taylor 2011; Sierminska and Takhtamanova 2011; Verick 2009). However, by contrast with previous recoveries, employment growth patterns have favored men since the official end of the recession in June 2009 (Kochhar 2011). Since the recent recovery has been sluggish relative to previous recoveries, much attention has been paid to the possibility of increased structural unemployment due to job mismatch and the unprecedented extension of unemployment insurance benefits to 99 weeks (as discussed in this symposium by Daly, Hobijn, Şahin, and Valletta; see also Howell and Azizoglu 2011; Reich 2010; Rothstein 2011).

In this paper, we investigate the differential impacts of these factors across demographic groups. The approach we take is most similar to that of Hines, Hoynes, and Krueger (2001) who use annual data from the March Current Population Survey for 1976–96 to examine the impact of cycles on employment, hours, earnings, and income. They adopt a state panel approach where the effects of the business cycle are identified by variation in the timing and severity of cycles across states. They explore differences across education groups (finding greater sensitivity for the less educated) and test for a structural break in sensitivity in 1990 (finding none), as well as examining effects of business cycles on wage growth, health and work injuries, and government finances. As described below, we also use a state panel model in our analysis. We expand on their work by examining monthly data through December 2011, which enables a detailed analysis of the Great Recession and the start of the current recovery. Further, we examine differences across race, gender, age, and education groups and test for differences across groups and over time.

Raw Changes by Group and Comparisons across 1980 and 2007 Recessions

We begin with a snapshot of the labor market outcomes by demographic group in May 2007, on the eve of the Great Recession. Table 1 shows the employment, unemployment, hours, and earnings of individuals by age, race, sex, and education. Employment, hours, and earnings are higher for men, whites, prime-age workers, and those with higher education levels. The opposite pattern, for most groups, is found for unemployment. These differences can be substantial. For example, less than half of individuals with no high school degree are working at the peak of the business cycle in 2007, compared to 86 percent of college graduates. Fifty-nine percent of black women are working, compared to 71 percent of white women.

For this comparison, and for much of what follows, we utilize individual-level data from the Current Population Survey (CPS) Merged Outgoing Rotation Group (MORG) covering the period from January 1979 to December 2011.[1] The CPS is a representative monthly household survey conducted by the U.S. Bureau of Labor

---

[1] We obtain the CPS-MORG extracts from the National Bureau of Economic Research: ⟨http://www .nber.org/morg/annual/⟩. Our sample includes individuals aged 16 to 60. We drop those over age 60

Table 1

Labor Market Outcomes by Race, Gender, Education, and Age, May 2007

	Employment rate (%)	*Unemployment rate (%)*	*Usual weekly earnings (2010$)*	*Hours last week*
White men	81	3.6	830	34
White women	71	3.2	499	25
Black men	66	9.1	448	26
Black women	59	6.5	401	24
Hispanic men	79	6.2	524	32
Hispanic women	58	4.9	298	20
Age 16 to 19	33	14.4	69	8
Age 20 to 24	68	6.4	306	23
Age 25 to 44	81	3.7	679	32
Age 45 to 60	75	3.3	707	30
Less than high school	48	10.1	187	16
High school graduate	72	5.4	306	28
Some college	76	3.6	545	29
College graduate	86	1.6	1,039	35

Source: Authors' tabulations of Current Population Survey Merged Outgoing Rotation Group (CPS-MORG) data.
Note: May 2007 was the eve of the Great Recession.

Statistics that collects information on unemployment, labor force participation, and demographic characteristics of the population. The MORG is a subset of the full CPS sample, with detailed information for 25,000 or more individuals per month, including their employment status, weekly work hours, and usual weekly earnings, as well as the age, education, race, ethnicity, and gender of each recipient. We collapse the MORG into cells based on state, year-month, and demographic group. Our demographic groups are defined by single year of age, gender, race/ethnicity (white, black, Hispanic, other),[2] and education (less than high school, high school, some college, college graduate or more).[3] For each cell, we calculate the percent employed and the percent unemployed using the CPS-provided weights.

---

to abstract from retirement decisions; we also drop the small number of observations missing ethnicity, which are all pre-2002.

[2] White, black, and other races are all non-Hispanic. Because of small population shares, we do not present results for the "other" race group. For the remainder of the paper we will refer to these as "race" groups even though they are more accurately race/ethnicity groups. By "single year of age" we mean, for example, that 18 year-olds are a separate group from 19 year-olds.

[3] Beginning in January 1992, the Bureau of Labor Statistics and the Census Bureau changed the focus of the Current Population Survey educational-attainment question from years of attainment to degree-receipt. We follow the matching procedure outlined in Jaeger (1997) to create categories that are comparable over time. However, the redesign of the education question creates a discontinuity in the categorization of educational attainment for which we cannot fully correct.

Next we turn to exploring the "raw" changes in labor market outcomes for these groups during the 2007 recession and comparing them to the changes in the recessionary episodes of the early 1980s. Here, we define the recessions by identifying the low and high points of the seasonally adjusted national unemployment rate; the subsequent high to low points of the unemployment rate identify the recovery. Our qualitative conclusions are unchanged if we use the NBER dating. However, for present purposes we prefer using the unemployment rates to date the cycles because the NBER dating depends in substantial part on GDP growth, and labor market measures tend to lag changes in GDP. Thus for 2007, we have the recession of May 2007 to October 2009 and the recovery of October 2009 to December 2011 (the last month in our data). For the 1980 cycle, we have the recession of May 1979 to November 1982 and the recovery of November 1982 to January 1985 (we use 27 months of recovery because that matches the data availability for the current recovery).

In the first two columns of Table 2, we show peak-to-trough changes in the unemployment rate for the race/sex, age, and education subgroups over the 1980 and 2007 recessions. To construct this table (and all subsequent calculations using the Current Population Survey), we first compute monthly unemployment rates for each demographic group from the CPS-MORG data. We then carry out a seasonal adjustment to this data, regressing each time series on a set of month dummies (with December omitted), and using the constant and residuals from this regression to create the adjusted series. Bold typeface in the table indicates groups for which the difference between peak-to-trough changes in labor market outcomes in the two recessions is statistically significant at the 5 percent level.

In the 2007 recession, the demographic groups who have high baseline unemployment rates (Table 1) also had the greatest increase in unemployment (Table 2) over the recession. Men had larger increases than women; blacks and Hispanics had larger increases than whites; youth had larger increases than the middle aged; and low education groups were also hit the hardest.

Comparing the 2007–2009 recession to the 1980s recession, several patterns emerge. First, for most groups, the increase in unemployment is greater in the more recent recession (although only statistically significantly different for high school graduates and college graduates). The largest increases (relative to the 1980s recession) are for Hispanic women and those with a high school degree. The exceptions include black men, Hispanic men, and those with less than a high school degree, all groups that experienced a smaller increase in unemployment rates compared to the 1980s recession. However, over time the educational distribution has shifted toward the higher educational categories so, although all but the least-educated did worse, on average people have moved into the better-faring groups.

The final two columns of Table 2 show results focusing on changes (in percentage point terms) in the employment rate. The patterns across groups are fairly similar to those of the unemployment rate: men, black and young workers, and low education groups all experienced greater reductions in employment in the current recession. However, comparing the two recessions presents a noticeably different pattern than the one for the unemployment rate. For all groups

Table 2

Peak-to-Trough Percentage Point Changes in Unemployment and Employment Rates by Group, 1980 and 2007 Recessions

(percentage points)

	Peak-to-trough changes in unemployment rate		Peak-to-trough changes in employment rate	
	May 1979 to Nov. 1982	*May 2007 to Oct. 2009*	*May 1979 to Nov. 1982*	*May 2007 to Oct. 2009*
White men	5.79	6.47	**−4.79**	**−7.34**
White women	3.73	3.59	**1.92**	**−2.81**
Black men	11.91	9.50	−8.41	−9.02
Black women	4.79	5.73	−0.85	−6.14
Hispanic men	10.23	6.09	−10.94	−6.25
Hispanic women	3.63	6.46	−0.56	−4.97
Age 16 to 19	10.55	10.86	−6.99	−7.79
Age 20 to 24	8.05	8.76	−5.39	−8.69
Age 25 to 44	5.29	5.78	**−2.05**	**−5.90**
Age 45 to 60	3.57	3.89	−0.82	−2.93
Less than high school	10.83	8.12	−5.95	−8.72
High school graduate	**5.96**	**8.28**	**−3.37**	**−7.99**
Some college	3.64	5.17	**−0.02**	**−4.72**
College graduate	**1.75**	**2.84**	−1.35	−2.15

Source: Authors' tabulations of Current Population Survey Merged Outgoing Rotation Group (CPS-MORG) data.
Notes: Peak-to-trough dated using minimum and maximum seasonally adjusted U.S. unemployment rates. Bold typeface indicates groups for which the difference between peak-to-trough changes in labor market outcomes in the two recessions is statistically significant at the 5 percent level. This significance test was implemented by a simple difference-in-differences regression as follows: using data for the four time periods 5/79, 11/82, 5/07, and 10/09, we regressed group-specific employment for each major demographic group on indicator variables for 1) 2007 recession (5/07 *or* 10/09), 2) trough periods (11/82 or 10/09), and 3) 2007 trough (10/09). The test is based on the statistical significance (at the 5 percent level) of indicator "3" for the 2007 trough.

except Hispanic men, the employment rate fell more during the recent recession than during the 1980s (and statistically significantly so for whites, those ages 25–44, high school graduates, and those with some college). One possible reason for this difference is that the 1980s recession occurred while women's labor force participation rates were undergoing a secular increase; that increase leveled out (and even slightly reversed) at the start of the twenty-first century (as discussed in this journal by Juhn and Potter 2006). For example, white women experienced *increases* in employment rates during the 1980s recession, but decreases during the current recession. The 1980s increase in employment for white women (and relatively small decreases in employment rates for black and Hispanic women) were likely driven by the secular increase in women's labor force participation rates, thus masking any business cycle sensitivity.

A Regression Approach for Potentially Confounding Factors

The crude changes over time across recessions are informative about the cross-group and cross-recession patterns, but they are also limited. Cross-group comparisons may be confounded by changes in other determinants of labor market success. For example, if the composition of low education groups is shifting over time to racial, ethnic, or age groups that fare worse in the labor market, then the measured change over time for low education groups will be confounded with those changes. If there are nonrecession-based changes in labor market patterns over time—like the increase in women's labor market participation—then these will also be wrapped up in the measured changes. To address these issues, we turn to a regression-adjusted measure of sensitivity to business cycles. We seek to use differences in the timing and intensity of state-level movement in unemployment rates to estimate how different demographic groups are affected by business cycle swings.

Again, we use the Current Population Survey (CPS) Merged Outgoing Rotation Group (MORG) data from January 1979 to December 2011. As noted already, we collapse the MORG into cells based on state, year-month, and the demographic groups (race/sex × age × education) described earlier. Also, we supplement this data with national and state unemployment statistics compiled from the Current Population Survey by the Bureau of Labor Statistics (2011a, b).

As a starting point, we estimate a regression in which the dependent variable y is the unemployment rate for a particular group, defined by the demographic cell g for that group, (race/sex × age × education), state s, and time (year-month) t. Our regression equation takes the form:

$$y_{gst} = \beta_{\text{major-group}} UN_{st} + RaceSex_g + Age_g + Educ_g + \alpha_s + \delta_t + year_t \gamma_s + \varepsilon_{gst}$$

We estimate this equation for each major demographic group, such as black men, white women, those without a high school degree, those with a college degree or more, those 18 years of age, 19 years of age, and so on. On the right-hand side of each equation, UN_{st} is the state unemployment rate in month-year t, $RaceSex_g$, Age_g, and $Educ_g$ are group-specific intercepts, and we include state (α_s) and year-month (δ_t) fixed effects and state-specific time trends (γ_s). The coefficient of interest is $\beta_{\text{major-group}}$, which gives the sensitivity of the group (for example, white men) to the state unemployment rate.[4] We use the Current Population Survey population weights for each cell, and we conduct statistical inference clustering on U.S. states.[5]

---

[4] We estimate this model separately by major demographic group, with the unit of observation being subgroup by state by year-month cells. For example, when we estimate the model for white men, there are 180 observations (45 age categories × 4 education categories) for each state-year-month. In this example, the $RaceSex_g$ dummies are dropped from the regression; the Age_g and $Educ_g$ dummies control for compositional shifts within white men.

[5] Our approach is similar in spirit to equation 7 and Table 2 in Blanchard and Katz (1992), who examine the responsiveness of U.S. states to the overall U.S. business cycle. We differ from their approach in that we use state-year-demographic group variation, and we examine responsiveness by specific demographic

This regression analysis embodies several changes to our analysis, compared to the raw differences presented in the previous section. First, it changes the source of variation used to estimate the sensitivity to the business cycle. The raw changes in the previous section were driven by national changes over time; specifically, comparing labor market outcomes by group between the peak and trough of a recession. Instead, here our coefficients are based on panel fixed effect estimates. We include state fixed effects, which remove variation that is purely driven by cross-state differences. We also include time fixed effects, which remove variation common to a point in time and control for flexible national time trends. Doing so protects our estimates from being driven by secular changes in demographic patterns such as changes in women's attachment to the labor market. After controlling for the fixed effects, we are left with variation that is driven by how the timing and severity of the business cycle affects states differently. When a state enters a recession (or recovery) earlier than the national average, or when a state's change in overall unemployment is greater than the national average, that variation is used to identify the coefficients in our regression.

Another feature of the regression analysis, compared to the raw changes above, is that we can control for demographic characteristics, thereby statistically adjusting for any differences in the composition of groups. For example, the group of workers with less than a high school degree is becoming more Hispanic over time. The raw differences for education groups, shown above, may in part reflect such changes in composition. A final important difference between the two approaches is that the regression results are not only estimated over the recession periods, but instead are estimated using data from both contractions and expansions.

To begin, we estimate this regression separately for each of our major demographic groups. For example, we estimate it for all 16 year-olds, and preserve the coefficient β_{16} (for this regression, the age dummies are excluded from the estimation). We then estimate the regression for 17 year-olds, and so on. After estimating for each age, we re-estimate the equation separately for each of our six main race/sex groups, and for our four education categories. The results are presented in Figures 3 and 4.

Figure 3 shows the results of the series of regression estimates for each single year of age. Each point on the graph represents estimates from a separate regression: the *x*-axis gives the person's age, and on the *y*-axis, we plot the estimated coefficient and the 95 percent confidence interval. For example, the first point on the graph is interpreted as "when a state-year experiences a percentage point higher unemployment rate, 16 year-olds in that state experience a 2.8 percentage point higher unemployment rate." Figure 3 shows that the labor market cycle hits especially hard

---

groups to overall state-year variation. By regressing state-specific labor market outcomes on the overall unemployment rate, Blanchard and Katz note, "Here, obviously, the proper weighted average of coefficients [across states] is equal to one; of interest is the distribution of [group-specific coefficients] across [groups]." Here, we too are interested in the descriptive findings for the differences in effects of cycles across demographic groups.

Figure 3

Effect of State Unemployment Rate on Group Unemployment Rate, by Single Year of Age

(percentage points)

Source: Authors' tabulations of the Current Population Survey, Merged Outgoing Rotation Group (CPS-MORG) for 1/1979–12/2011.

Notes: Each point is the estimate on state unemployment rate from a separate regression (along with the 95 percent confidence interval) for a given demographic group. The model also includes fixed effects for demographic group, state, and year-month, as well as state linear time trends. See text for details.

for youth, with responsiveness for 16–19 year-olds more than twice that of those in their mid-20s. The coefficients continue to decline, at a more modest rate, until ages in the mid-50s.

In Figure 4 we present results from stratifying on race-sex demographic groups and on education. The results here suggest that the unemployment rate of men is more responsive to business cycle movements than the unemployment rate for women; that the response for blacks is greater than for Hispanics, for whom in turn the labor market response is higher than for whites; and that low education groups are more responsive than high education groups. The differences are large: an increase of one percentage point in the state unemployment rate leads to almost a two percentage point increase in unemployment for workers with less than a high school degree compared to less than half a percentage point increase for those with a college degree. The responsiveness of the unemployment rate of black men to the business cycle is almost double the responsiveness of white men's unemployment

Figure 4

Effect of State Unemployment Rate on Group Unemployment Rate, by Race/Sex and Education

(percentage points)

Notes: Each point is the estimate on state unemployment rate from a separate regression (along with the 95 percent confidence interval) for a given demographic group. The model also includes fixed effects for demographic group, state, and year-month, as well as state linear time trends. See text for details.

rate, and the responsiveness of the unemployment rate of black women is more than double the responsiveness of that of white women.

These results are qualitatively similar to the raw changes presented earlier. This correspondence is remarkable, for three reasons: First, the two models are estimated using fundamentally different sources of variation, that is, state versus national cycles. Second, the model controls for time trends for each subgroup. Third, the regression model is estimated over the full 1979–2011 time period, rather than just during the 2007 or 1980s recessions.

We then carry out a parallel regression exercise, but this time using the employment rate, rather than the unemployment rate, as our left-hand-side dependent variable. In Figure 5 we consider the sensitivities of the age-specific employment rate to the overall state-month unemployment rate. The interpretation of the coefficients is similar to that discussed above; for example, a one percentage point increase in the state unemployment rate leads to a 1.7 percentage point reduction

Figure 5

Effect of State Unemployment Rate on Group Employment Rate, by Single Year of Age

(percentage points)

Notes: Each point is the estimate on state unemployment rate from a separate regression (along with the 95 percent confidence interval) for a given demographic group. The model also includes fixed effects for demographic group, state, and year-month, as well as state linear time trends. See text for details.

in the employment rate for 16 year-olds. The patterns here are similar to Figure 3: the youngest are the most responsive, and by-age cyclicality declines with age.

Figure 6 shows estimates of the impact of the overall state-month unemployment rate on the employment rate by race-sex and education groups. The patterns for the race/sex groups are somewhat different for employment compared to the unemployment rate in Figure 4. It is still the case that white individuals are less responsive than their black counterparts, women are less responsive than are men, and higher education groups, less responsive than lower education groups. (In reading the graph, note that as one moves *up* the *y*-axis the sensitivity gets closer to 0 and is therefore *less* sensitive). However, in contrast to Figure 4, here the gender differences in responsiveness are much greater. The gender differences are large enough to dominate the race differences, so that the three least-responsive groups (among the six race/sex groups) are the three groups of women. For this measure, Hispanic women are the least responsive of all the demographic groups. It can be shown that the cyclical responsiveness of the employment rate, the unemployment

Figure 6

Effect of State Unemployment Rate on Group Employment Rate, by Race/Sex and Education

(percentage points)

Source: Authors' tabulations of the Current Population Survey, Merged Outgoing Rotation Group (CPS-MORG) for 1/1979–12/2011.
Notes: Each point is the estimate on state unemployment rate from a separate regression (along with the 95 percent confidence interval) for a given demographic group. The model also includes fixed effects for demographic group, state, and year-month, as well as state linear time trends. See text for details.

rate, and the labor force participation rate are related through an adding-up identity. The larger gender differences for cyclical responsiveness of the employment rate is consistent with women being more likely to act as added workers (labor force increasing in recessions) and men being more likely to act as discouraged workers (labor force decreasing in recessions). Hispanic women, with their high rates of marriage (compared to the other groups) may be most likely to behave as added workers; hence the very large widening for Hispanics.

One potential limitation of our specification is that the time dummies throw away a large portion of the national macro cycle. We have re-estimated our first regression equation without the year-month dummies, and present results from this in figures in an online appendix available with this paper at ⟨http://e-jep .org⟩. The results are very similar to those in Figures 3–6. The exception to this is that women's employment appears to be more responsive when the year-month

dummies are omitted. This is exactly the demographic group and the outcome variable that reflects the concerns discussed above about bias due to long-run demographic trends.

Taken as a whole, these regression results largely reinforce the simple over-time patterns: men, nonwhites, youth, and those with lower education levels are the most responsive to cycles. Given the important differences in these two methodological approaches discussed above, we are impressed by the similarity of the findings. We interpret this as evidence of the robustness of the patterns that we document.

Did Cyclical Responses Differ in the Great Recession?

We can use a variation of our regression model to explore whether the Great Recession is different from earlier business cycle patterns. In particular, as above in our analysis of raw changes, we compare the Great Recession to the early-1980s recession. In so doing, we focus on two additional questions: First, for each demographic group, is the pattern of business cycle responsiveness in the Great Recession similar to what it was in the back-to-back recessionary episodes of the early 1980s? Second, how do the responses to the recoveries compare across the demographic groups?

To investigate these questions, we again implement a regression model. We start with our original regression equation, but instead of estimating separate models for each major demographic group (for example, less than a high school education), we pool all observations from all groups together. We then run three regressions on this pooled data set, with each regression focusing on different categories of major groups: race/sex groups, age groups, and education groups.[6] In each model, we allow for the responsiveness of each major demographic group in the category under consideration to vary depending on the time period. The time periods cover the 1980s recession, the 2007 recession, and all other time periods. For example, in the regression focusing on education categories, we estimate 12 key coefficients, one for each of the three time periods times four education groups ($\beta_{education-group}^{time-period} \, UN_{st}$). One coefficient in this regression would measure the responsiveness for high school graduates in the 1980s recession and another for college graduates in the 2007 recession, and so on. For each major demographic group (for example, high school graduates), we then test for equality of coefficients across the two recessions $\left(\text{testing whether } \beta_{major-group}^{1980} = \beta_{major-group}^{2007}\right)$.

We repeat this exercise focusing on the recovery periods for each recession. This leads to a total of six regression models, covering three major group categories (age, race/sex, and education) and two phases of the cycle (recessions and recoveries). To implement these regressions, we use data from May 1979 to November 1982

---

[6] We consider six race/sex demographic groups (white men, white women, black men, black women, Hispanic men, and Hispanic women); four age groups (16–19, 20–24, 25–44, and 45–60); and four education groups (less than high school, high school grad only, some college, and college graduate).

for the recessionary period of the early 1980s and from May 2007 to October 2009 for the 2007 recession, based on the minimum to maximum of the national (seasonally adjusted) unemployment rate. For the recovery periods, we use November 1982 through January 1985 and October 2009 to December 2011.

The detailed findings of these regressions, along with some additional statistical tests, are presented in an online appendix available with this paper at ⟨http://e-jep .org⟩.[7] Here, we summarize the main qualitative conclusions.

When comparing the responsiveness of unemployment rates for different major demographic subgroups in the recession of the 1980s with the Great Recession, the across-group patterns are similar to those of the stratified regression (shown earlier in Figures 3 and 4). The responsiveness of the unemployment rates of men, Hispanics, youth, and those with lower education levels are higher in both recessions, while the unemployment rates of women, prime-aged workers, and higher education groups are less responsive. For each of the race/sex groups, the cyclical responsiveness is very similar across recession periods, and we cannot reject the hypothesis of equality across the 1980 and 2007 recessions for any of these groups.[8] We do find that the Great Recession has statistically significantly larger impacts for older workers, and for each education category. The magnitude of the change is small, however: for example, the coefficient for those aged 45–60 increases from 0.70 in the 1980s recession to 0.85 in the 2007 recession. Our main punch-line is thus reinforced: the Great Recession is deeper than previous recessions, but otherwise is affecting groups more or less similarly.

The story is somewhat different when we consider the responsiveness of the unemployment rate for different demographic groups in the recoveries following the 1980s and the 2007–2009 recessions. The cyclicality for the race/sex groups is significantly lower for the Great Recession, suggesting a weaker responsiveness to the recovery. For example, for black women the coefficient is 1.58 in the 1980s recovery and 1.34 in the 2007 recovery. For the age and education comparisons, the patterns for the 1980s recovery and the current recovery are relatively comparable.[9]

---

[7] The pooled regression presented in the electronic Appendix is more restrictive than the stratified regressions behind Figures 3–6 because it imposes identical time dummies and state fixed effects for all demographic groups. In order to preserve the flexibility of the pooled regressions, we include as control variables group-specific quadratic time trends and group-specific state fixed effects. These controls allow us to recover similar coefficients to the stratified models.

[8] In an alternative specification in which we pooled together all men as a group, we did find that the cyclicality in the Great Recession for men is statistically significantly higher than in the 1980s recession, although the magnitude of the over-time differences is fairly small.

[9] Given that we are regressing group-specific unemployment rates on the state aggregate unemployment rate, one might expect that the average across demographic subgroups (appropriately weighted by population shares) should average to 1. This is not necessarily the case because our group outcome measures come from our MORG sample where we limit the sample to those aged 16–60. The cycle measure, the state unemployment rate, is the aggregate unemployment rate published by the Bureau of Labor Statistics (2011b).

Figure 7
Percent Change in Employment over 2007 Recession versus Share Male, by Industry
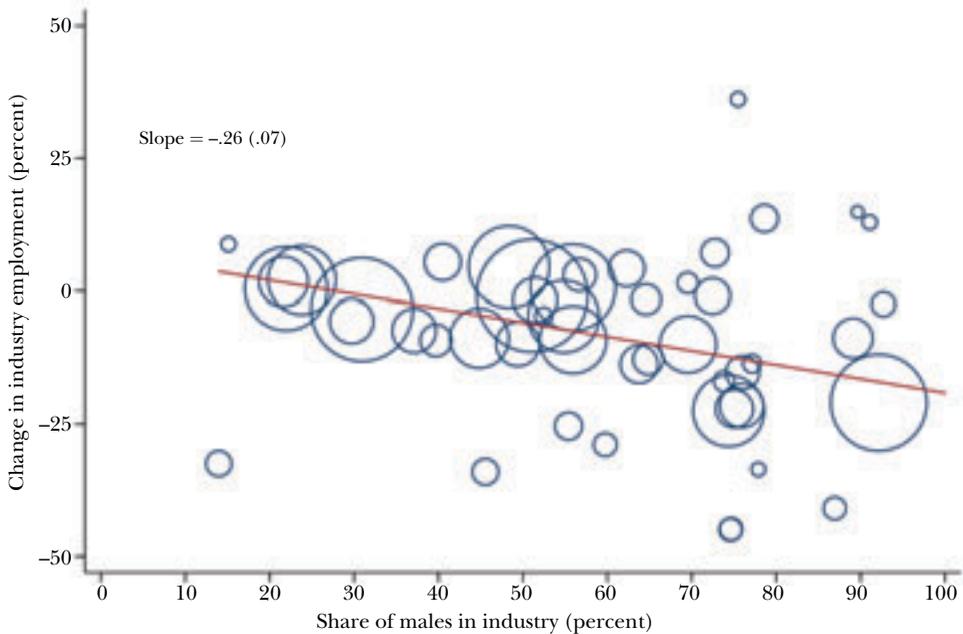
Source: Authors' tabulations of Current Population Survey, Merged Outgoing Rotation Group (CPS-MORG) data.
Notes: Observations are weighted by total industry employment in May 2007 (the start of the recession). Industry classification is based on 2-digit sectors from the 2002 North American Industry Classification System (NAICS).

What Explains the Differences across Demographic Groups?

One likely explanation for these persistent differences in the impacts of cycles across demographic groups derives from the variation in cyclicality across industries. Construction and manufacturing are more-cyclical industries while services and government are less cyclical. Furthermore, many of the demographic groups that exhibit larger cyclical variation (men, those with lower education levels, minorities) are more likely to be employed in the industries with greater exposure to cycles. As an illustration of the importance of industry in the context of demographic comparisons, Figure 7 presents a scatterplot of the percent decline in industry employment between the peak and trough of the current recession (for 52 industry groups). We show the difference in the severity of the labor market shock on the *y*-axis, and on the *x*-axis is the share male in the industry (measured at the peak). We have added a bivariate regression line for guidance.[10] As the figure shows, the higher the share

---

[10] The percent change in employment is calculated between May 2007 and October 2009, and we collapse the data to industry using the "2-digit" NAICS industry codes. The regression line is calculated

male in the industry, the larger the employment decline in the current recession. This appears to be an industry effect (as opposed to a "male" effect), because the employment pattern persists if we decompose the employment loss into the loss for women and the loss for men.

To explore this further, we create a "predicted" peak-to-trough change in the employment rate (May 2007 to October 2009) for each demographic group. Specifically, we follow Bartik (1991) and create predicted changes in the employment level for each demographic group by multiplying the group's share of total employment in 30 industry-occupation cells at the peak (May 2007) by the U.S.-wide peak-to-trough change in total industry-occupation employment and summing across industry-occupations. The difference between the actual and predicted changes can be interpreted as the group-specific component of employment loss that operates above (or below) the direct effect of being in cyclical industry-occupations. The 30 cells are defined by ten industries times three occupations (managerial, clerical/services, and "blue collar").[11]

We present the results in Table 3, setting out the predicted change in the employment rate in column 2, the actual change in the employment rate in column 3, and the employment rate at the peak in column 1 (repeated from Table 1).

The results in Table 3 show that the difference in the cyclicality between men and women is explained almost completely by the gender differences in the industry-occupation of employment. The male employment rate is predicted to decrease by 7.4 percentage points, slightly larger than the observed decline of 7.1 percentage points. The female employment rate is predicted to drop by 3.0 percentage points, just below the observed 3.4 percentage point decline. Interestingly, the Great Recession has larger impacts than predicted for blacks, young workers, and more-educated workers. On the other hand, whites, older workers, and less-educated workers experienced smaller declines than predicted. For example, older workers (45–60) experienced a 3.3 percentage point reduction in their employment rate, two percentage points lower than their predicted decline. College-educated workers experienced a 4.6 percentage point decline in their employment rate compared to the predicted decline of 3.2 percentage points. The largest discrepancies between predicted and actual change are for youth, especially for teens. For this group, their industry/occupation mix predicts a loss of 1.6 percentage points of employment; the actual loss was 7.3 points. We speculate that this finding may reflect the dynamics of hiring and separations during the recession. Workers with job tenure were able to lower their rate of quits, but those starting without jobs (such as youth)

---

using a weighted regression, with industry employment at the peak as the weights. There are a total of 52 industries and while we include all observations to calculate the regression line, in the figure we drop the few observations outside the –50 percent, +50 percent range on the *y*-axis to improve the scaling.

[11] Using detailed industry codes in the CPS-MORG, we group observations into 10 major industries: 1) Agriculture, Forestry and Fishing, 2) Mining, 3) Construction, 4) Manufacturing, 5) Transportation, Warehousing, and Utilities, 6) Wholesale Trade, 7) Retail Trade, 8) Finance, Insurance, Real Estate, and Information, 9) Services, and 10) Public Administration. We create regression-based seasonally adjusted data series for each group-industry-occupation prior to performing this analysis.

Table 3

Actual and Predicted Percentage Point Change in Employment Rate, by Group

(predictions based on industry-occupation mix)

		May 2007 to October 2009	
	Employment rate (%) May 2007	*Predicted change*	*Actual change*
Men	81	−7.4	−7.1
Women	71	−3.0	−3.4
White	66	−5.1	−4.7
Black	59	−4.8	−6.9
Hispanic	58	−6.4	−6.3
Age 16 to 19	33	−1.6	−7.3
Age 20 to 24	68	−5.0	−8.3
Age 25 to 44	81	−6.0	−5.5
Age 45 to 60	85	−5.3	−3.3
Less than high school	48	−5.7	−4.8
High school graduate	**72**	−7.1	−6.7
Some college	76	−4.9	−6.6
College graduate	86	−3.2	−4.6

Source: Authors' tabulations of Current Population Survey, Merged Outgoing Rotation Group (CPS-MORG) data.

Notes: We create predicted changes in the employment level for each demographic group by multiplying the group's share of total employment in 30 industry-occupation cells at the peak (May 2007) by the U.S.-wide peak-to-trough change in total industry-occupation employment and summing across industry-occupation. The difference between the actual and predicted changes can be interpreted as the group-specific component of employment loss that operates above (or below) the direct effect of being in cyclical industry-occupations. The 30 cells are defined by 10 industries times three occupations (managerial, clerical/services, and "blue collar").

may have been hit hardest by the large drop in hiring rates (Davis, Faberman, and Haltiwanger 2012).

Conclusion

The labor market decline during the Great Recession and its aftermath has been both deeper and longer than the early 1980s recession—indeed, the longest and deepest since the Great Depression. The labor market effects of the Great Recession have not been not uniform across demographic groups. Men, blacks, Hispanics, youth, and those with lower education levels experience more employment declines and unemployment increases compared to women, whites, prime-aged workers, and those with high education levels. However, these dramatic differences in the cyclicality across demographic groups have been remarkably stable since at least the late 1970s and across recessionary periods versus expansionary periods. These

gradients persist despite the dramatic changes in the labor market over the past 30 years, including the increase in labor force attachment for women, Hispanic immigration, the decline of manufacturing, and so on.

The general tone of these findings might be surprising given much emphasis in the press on the "man-cession"—that is, the greater effect that the Great Recession has had on men (for examples of newspaper accounts, see Rampell 2009, Irwin and Dennis 2011). Our analysis shows that men, across recessions and recoveries, experience more cyclical labor market outcomes. This is largely the result of the higher propensity of men to be employed in highly cyclical industries such as construction and manufacturing, while women are more likely to be employed in less-cyclical industries such as services and public administration. More generally, much of the difference in the cyclical effect across groups during the 2007 recession is explained by differing exposure to fluctuations due to the industries and occupations in which the groups are employed.

Although overall the 2007–2009 recession appears similar to the 1980s recession, reponsiveness by women's employment and by that of the youngest and oldest workers was somewhat greater in the more recent recession. Further, we do find evidence of a "he-covery;" and the extent to which the current recovery is being experienced more by men than women (compared to the 1980s recovery) is largely due to a drop in women's cyclicality during the current recovery.

Despite these various distinctions, the overarching picture is one of stability in the demographic patterns of response to the business cycle over time. Who loses in the Great Recession? The same groups who lost in the recessions of the 1980s and who experience weaker labor market outcomes even in the good times. Viewed through the lens of these demographic patterns across labor markets, the Great Recession is different from business cycles over the three decades earlier in size and length, but not in type.

References

Bitler, Marianne P., and Hilary W. Hoynes. 2010. "The State of the Safety Net in the Post-Welfare Reform Era." Brookings Papers on Economic Activity, no 2 (Fall), pp. 71–127.

Blanchard, Olivier, and Lawrence F. Katz. 1992. "Regional Evolutions." Brookings Papers on Economic Activity, no. 1, pp. 1–76.

Bureau of Economic Analysis. 2011a. Personal Consumption Expenditures: Chain Type Price Index. Obtained from the Federal Reserve

Economic Data (FRED) on 9/9/11. http://research.stlouisfed.org/fred2/series/PCECTPI.

Bureau of Labor Statistics. 2011b. Local Area Unemployment Statistics from the Current Population Survey. http://www.bls.gov/lau/ (accessed 8/15/11).

Bureau of Labor Statistics. 2012a. Labor Force Statistics from the Current Population Survey. http://www.bls.gov/cps/ (accessed 2/20/12).

Bureau of Labor Statistics. 2012b. Current Employment Statistics - CES (National). http://www.bls.gov/ces/ (accessed 2/20/12).

Davis, Steven J., R. Jason Faberman, and John C. Haltiwanger. 2012. "Labor Market Flows in the Cross Section and Over Time." *Journal of Monetary Economics* 59(1): 1–18.

DeNavas-Walt, Carmen, Bernadette D. Proctor, and Jessica C. Smith. 2011. *Income, Poverty, and Health Insurance Coverage in the United States: 2010.* Current Population Reports P60-239, U.S. Census Bureau. Washington, D.C.: U.S. Government Printing Office.

Elsby, Michael, Bart Hobijn, and Ayşegul Şahin. 2010. "The Labor Market in the Great Recession." Brookings Papers on Economic Activity, no. 2: 1–48.

Farber, Henry. 2011. "Job Loss in the Great Recession: Historical Perspective from the Displaced Workers Survey, 1984-2010." NBER Working Paper 17040.

Goodman, Christopher J., and Steven M. Mance. 2011. "Employment Loss and the 2007–09 Recession: An Overview." *Monthly Labor Review* 134(4): 3–12.

Hines, James R., Hilary W. Hoynes, and Alan B. Krueger. 2001. "Another Look at Whether a Rising Tide Lifts All Boats." Chap. 10 in *The Roaring Nineties: Can Full Employment Be Sustained,* edited by Alan Krueger and Robert Solow. Russell Sage Foundation: New York.

Howell, David R., and Bert M. Azizoglu. 2011. "Unemployment Benefits and Work Incentives: The U.S. Labor Market in the Great Recession." Working Paper 257, Political Economy Research Institute.

Hoynes, Hilary W. 2000. "The Employment and Earnings of Less Skilled Workers over the Business Cycle." In *Finding Jobs: Work and Welfare Reform,* edited by Rebecca M. Blank and David Card, 23–71. Russell Sage Foundation: New York.

Irwin, Neil, and Brady Dennis. 2011. "Jobs Market: Men, Hit Hardest in Recession, Are Getting Work Faster Than Women." *The Washington Post,* July 6. http://www.washingtonpost.com/business/economy/jobs-market-men-hit-hardest-in-recession-are-getting-work-faster-than-women/2011/07/06/gIQAbGxH1H_story.html.

Jaeger, David A. 1997. "Reconciling the New Census Bureau Education Questions: Recommendations for Researchers." *Journal of Business and Economic Statistics* 15(4): 300–309.

Juhn, Chinhui, and Simon Potter. 2006. "Changes in Labor Force Participation in the United States." *Journal of Economic Perspectives* 20(3): 27–46.

Kochhar, Rakesh. 2011. "In Two Years of Economic Recovery, Women Lost Jobs, Men Found Them." Pew Social & Demographic Trends, Washington, D.C. http://pewsocialtrends.org/files/2011/07/Employment-by-Gender_FINAL_7-6-11.pdf.

Kochhar, Rakesh, Richard Fry, and Paul Taylor. 2011. "Wealth Gaps Rise to Record Highs Between Whites, Blacks, and Hispanics." Pew Social & Demographic Trends, Washington, D.C.

National Bureau of Economic Research (NBER). 2011. "U.S. Business Cycle Expansions and Contractions." http://www.nber.org/cycles.html (accessed 9/1/2011).

Rampell, Catherine. 2009. "As Layoffs Surge, Women May Pass Men in Job Force." *The New York Times,* February 6. http://www.nytimes.com/2009/02/06/business/06women.html?pagewanted=all.

Reich, Michael. 2010. "High Unemployment after the Great Recession: Why? What Can We Do?" Center on Wage and Employment Dynamics Policy Brief.

Rothstein, Jesse. 2011. "Unemployment and Job Search in the Great Recession." NBER Working Paper 17534.

Sierminska, Eva, and Yelena Takhtamanova. 2011. "Job Flows, Demographics, and the Great Recession." In *Who Loses in the Downturn? Economic Crisis, Employment, and Income Distribution, Research in Labor Economics,* vol. 32, edited by H. Immervoll, A. Peichl, and K. Tatsiramos, 115–54. Emerald.

Solon, Gary, Robert Barsky, and Jonathan Parker. 1994. "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?" *Quarterly Journal of Economics* 109(1): 1–26.

Stevens, Ann Huff, Douglas L. Miller, Marianne Page, and Mateusz Filipski. 2011. "The Best of Times, the Worst of Times: Understanding Pro-cyclical Mortality." NBER Working Paper 17657.

Verick, Sher. 2009. "Who is Hit Hardest during a Financial Crisis? The Vulnerability of Young Men and Women to Unemployment in an Economic Downturn." IZA Discussion Paper 4359, Institute for the Study of Labor.

The European Sovereign Debt Crisis[†]

Philip R. Lane

The capacity of the euro-member countries to withstand negative macroeconomic and financial shocks was identified as a major challenge for the success of the euro from the beginning (in this journal, for example, see Feldstein 1997; Wyplosz 1997; Lane 2006). By switching off the option for national currency devaluations, a traditional adjustment mechanism between national economies was eliminated. Moreover, the euro area did not match the design of the "dollar union" of the United States in key respects, since the monetary union was not accompanied by a significant degree of banking union or fiscal union. Rather, it was deemed feasible to retain national responsibility for financial regulation and fiscal policy.

On the one side, the ability of national governments to borrow in a common currency poses obvious free-rider problems if there are strong incentives to bail out a country that borrows excessively (Buiter, Corsetti, and Roubini 1993; Beetsma and Uhlig 1999). The original design of the euro sought to address the over-borrowing incentive problem in two ways. First, the Stability and Growth Pact set (somewhat arbitrary) limits on the size of annual budget deficits at 3 percent of GDP and the stock of public debt of 60 percent of GDP. Second, the rules included a "no bailout" clause, with the implication that a sovereign default would occur if a national government failed to meet its debt obligations.

On the other side, the elimination of national currencies meant that national fiscal policies took on additional importance as a tool for countercyclical macroeconomic policy (Wyplosz 1997; Gali and Monacelli 2008; Gali 2010). Moreover, since

■ *Philip R. Lane is Whately Professor of Political Economy at Trinity College Dublin, Dublin, Ireland, and Research Fellow, Centre for Economic Policy Research, London, United Kingdom. His email address is ⟨plane@tcd.ie⟩.*

banking regulation remained a national responsibility, individual governments continued to carry the risks of a banking crisis: both the direct fiscal costs (if governments end up recapitalizing banks or providing other forms of fiscal support) and also the indirect fiscal costs since GDP and tax revenues tend to remain low for a sustained period in the aftermath of a banking crisis (Honohan and Klingebiel 2003; Reinhart and Rogoff 2009).

There are three phases in the relationship between the euro and the European sovereign debt crisis. First, the initial institutional design of the euro plausibly increased fiscal risks during the pre-crisis period. Second, once the crisis occurred, these design flaws amplified the fiscal impact of the crisis dynamics through multiple channels. Third, the restrictions imposed by monetary union also shape the duration and tempo of the anticipated post-crisis recovery period, along with Europe's chaotic political response and failure to have institutions in place for crisis management. We take up these three phases in the next three major sections of this article, and then turn to reforms that might improve the resilience of the euro area to future fiscal shocks.

As will be clear from the analysis below, the sovereign debt crisis is deeply intertwined with the banking crisis and macroeconomic imbalances that afflict the euro area. Shambaugh (2012) provides an accessible overview of the euro's broader economic crisis. Even if the crisis was not originally fiscal in nature, it is now a full-blown sovereign debt crisis and our focus here is on understanding the fiscal dimensions of the euro crisis.

Pre-Crisis Risk Factors

Public debt for the aggregate euro area did not, at least at first glance, appear to be a looming problem in the mid 2000s. During the previous decade, the euro area and the United States shared broadly similar debt dynamics. For example, the ratio of gross public debt to GDP in 1995 was about 60 percent for the United States and 70 percent for the set of countries that would later form the euro area, based on my calculations with data from the IMF Public Debt Database. In both the United States and the euro area, the debt/GDP ratios declined in the late 1990s, but had returned to mid 1990s levels by 2007. The debt/GDP ratios then climbed during the crisis, growing more quickly for the United States than for the euro area.[1]

However, the aggregate European data mask considerable variation at the individual country level. Figure 1 shows the evolution of public debt ratios for seven key euro area countries over 1982–2011. These countries were chosen because Germany, France, Italy, and Spain are the four largest member economies, while the fiscal crisis so far has been most severe in Greece, Ireland, and Portugal (of course, Italy

[1] For a detailed country-by-country breakdown of the evolution of public sector debt across these seven countries from 1992–2011, see the Appendix available online with this paper at ⟨http://e-jep.org⟩.

Figure 1
The Evolution of Public Debt, 1982–2011
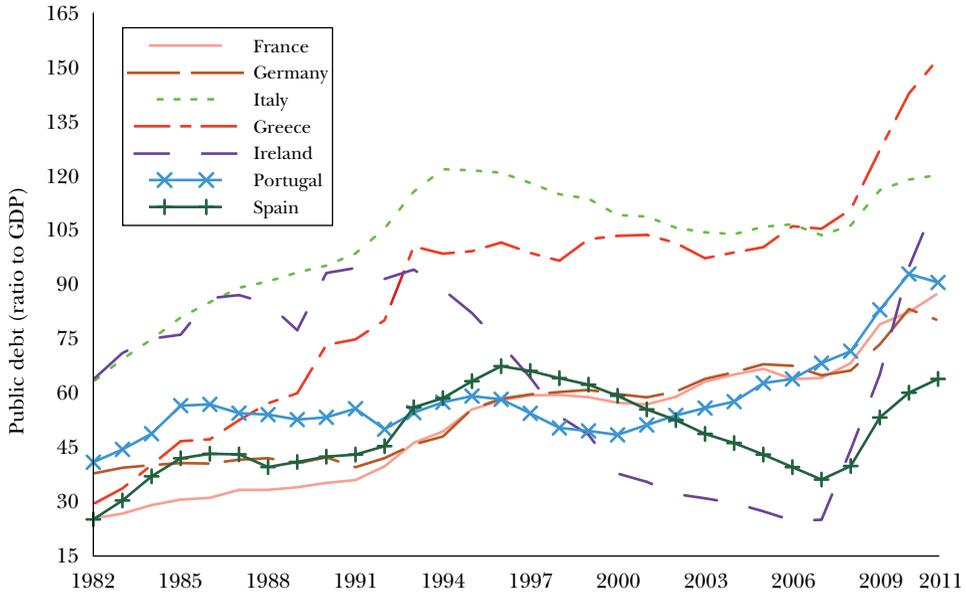
Source: Data from IMF Public Debt Database.

and Spain have also been flagged as fiscally vulnerable countries during the crisis). Clearly, these countries have quite different debt histories.

In one group, both Italy and Greece had debt/GDP ratios above 90 percent since the early 1990s; these countries never achieved the 60 percent debt/GDP limit specified in the European fiscal rules. Ireland, Portugal, and Spain each achieved significant declines in debt ratios in the second half of the 1990s, dipping below the 60 percent ceiling. While the Portuguese debt ratio began to climb from 2000 onwards, rapid output growth in Ireland and Spain contributed to sizable reductions in debt–output ratios up to 2007. Finally, France and Germany had stable debt/GDP ratios at around 60 percent in the decade prior to the onset of the crisis; indeed, their debt ratios were far above the corresponding values for Ireland and Spain during 2002–2007. Thus, circa 2007, sovereign debt levels were elevated for Greece and Italy, and the trend for Portugal was also worrisome, but the fiscal positions of Ireland and Spain looked relatively healthy. Moreover, the low spreads on sovereign debt also indicated that markets did not expect substantial default risk and certainly not a fiscal crisis of the scale that could engulf the euro system as a whole.

However, with the benefit of hindsight, 1999–2007 looks like a period in which good growth performance and a benign financial environment masked the accumulation of an array of macroeconomic, financial, and fiscal vulnerabilities (Wyplosz 2006; Caruana and Avdjiev 2012).

Table 1
Private Credit Dynamics

	Loans to private sector from domestic banks and other credit institutions (percent of GDP)		
	1998	*2002*	*2007*
Greece	31.8	56.5	84.4
Ireland	81.2	104.4	184.3
Portugal	92.1	136.5	159.8
Spain	80.8	100.1	168.5
Italy	55.7	77.3	96.5
Germany	112.2	116.7	105.1
France	81.0	85.6	99.3

Source: World Bank Financial Database.

Financial Imbalances and External Imbalances

A key predictor of a banking crisis is the scale of the preceding domestic credit boom (Gourinchas and Obstfeld 2012). Table 1 shows the evolution of credit/GDP ratios for the seven euro area countries. The European periphery experienced strong credit booms, in part because joining the euro zone meant that their banks could raise funds from international sources in their own currency—the euro—rather than their previous situation of borrowing in a currency not their own (say, U.S. dollars or German marks or British pounds) and then hoping that exchange rates would not move against them. In related fashion, lower interest rates and easier availability of credit stimulated consumption-related and property-related borrowing (Fagan and Gaspar 2007).

A related phenomenon was the increase in the dispersion and persistence of current account imbalances across the euro area. Table 2 shows that current account imbalances were quite small in the pre-euro 1993–1997 period. But, by the 2003–2007 period, Portugal (–9.2 percent of GDP), Greece (–9.1 percent), and Spain (–7.0 percent) were all running very large external deficits. Conversely, Germany ran very large external surpluses averaging 5.1 percent of GDP, while the overall euro area current account balance was close to zero.

To the extent that current account imbalances accelerated income convergence by reallocating resources from capital-abundant high-income countries to capital-scarce low-income countries, this would be a positive gain from monetary union (Blanchard and Giavazzi 2002). Similarly, current account deficits might have facilitated consumption smoothing by the catch-up countries to the extent that current income levels were perceived to be below future income levels. However, if capital inflows rather fueled investment in capital that had little effect on future productivity growth (such as real estate) and delayed adjustment to structural shocks (such as increasing competition from Central and Eastern Europe and emerging Asia in the production of low-margin goods),

Table 2
Current Account Balances
(percent of GDP)

	1993–1997	*1998–2002*	*2003–2007*	*2008–2011*
Greece	–2.0	–5.9	–9.1	–11.1
Ireland	3.4	–0.2	–2.6	–1.6
Italy	2.1	0.2	–1.8	–2.9
Portugal	–2.4	–9.0	–9.2	–10.5
Spain	–0.6	–3.1	–7.0	–5.8
France	1.1	2.0	–0.2	–1.9
Germany	–0.9	–0.3	5.1	5.7

Source: International Monetary Fund's World Economic Outlook database.

then the accumulation of external imbalances posed significant macroeconomic risks (Blanchard 2007; Giavazzi and Spaventa 2011; Chen, Milesi-Ferretti, and Tressel forthcoming).

For countries running large and sustained external deficits, Blanchard (2007) identifies several risk factors. In terms of medium-term growth performance, a current account deficit can be harmful if increased expenditure on nontradables squeezes the tradables sector by bidding up wages and drawing resources away from industries that have more scope for productivity growth. This is especially risky inside a currency union, because nominal rigidities mean that the downward wage adjustment required once the deficit episode is over can only be gradually attained through a persistent increase in unemployment.

In addition, a large current account deficit poses short-term risks, if there is a sudden stop in funding markets such that the deficit must be narrowed quickly. Large and sudden capital flow reversals have often proven costly in terms of output contractions, rising unemployment, and asset price declines (Freund and Warnock 2007). A reversal in capital flows is also associated with a greater risk of a banking crisis, especially if capital flows have been intermediated through the domestic banking system.

The 2003–2007 Boom

The most intense phase of the dispersion in credit growth and current account imbalances did not occur at the onset of the euro in 1999. Rather, there was a discrete increase during 2003–2007 (Lane and Pels 2012; Lane and McQuade 2012). A complete explanation for the timing of this second, more intense phase of current account deficits and credit booms is still lacking, but the simultaneous timing with the securitization boom in international financial markets, the U.S. subprime episode, and the decline in financial risk indices suggest that the answer may be found in the underlying dynamics of the global financial system and the unusually low long-term interest rates prevailing during this period.

The credit boom in this period was not primarily due to government borrowing. For Ireland and Spain, the government was not a net borrower during 2003–2007. Rather, households were the primary borrowers in Ireland and corporations in Spain, with the property boom fueling debt accumulation in both countries. In Portugal and Greece, the government and corporations were both significant borrowers, but these negative flows were partly offset during this period by significant net accumulation of financial assets by the household sector in these countries.

Failure to Tighten Fiscal Policy

Looking back, the failure of national governments to tighten fiscal policy substantially during the 2003–2007 was a missed opportunity, especially during a period in which the private sector was taking on more risk. In some countries (Ireland and Spain), the credit and housing booms directly generated extra tax revenues, since rising asset prices, high construction activity, and capital inflows boosted the take from capital gains taxes, asset transaction taxes, and expenditure taxes. Faster-growing euro member countries also had inflation rates above the euro area average, which also boosted tax revenues through the non-indexation of many tax categories. Finally, low interest rates meant that debt servicing costs were below historical averages. However, these large-scale revenue windfalls were only partially used to improve fiscal positions, with the balance paid out in terms of extra public spending or tax cuts. Overall, fiscal policy became less countercyclical after the creation of the euro, undoing an improvement in cyclical performance that had been evident in the 1990s (Benetrix and Lane 2012).

A contributory factor in the failure to tighten fiscal policy was the poor performance of the analytical frameworks used to assess the sustainability of fiscal positions. In evaluating the cyclical conduct of fiscal policy from 2002–2007, domestic authorities and international organizations such as the IMF, OECD, and European Commission primarily focused on point estimates of the output gap in order to estimate the "cyclically adjusted" budget balance, without taking into account the distribution of macroeconomic, financial, and fiscal risks associated with the expansion in external imbalances, credit growth, sectoral debt levels, and housing prices. A more prudential and forward-looking approach to risk management would have suggested more aggressive actions to accumulate buffers that might help if or when the boom ended in a sudden and disruptive fashion (Lane 2010).

For the euro periphery, the 2008 global financial crisis triggered a major reassessment among investors of the sustainability of rapid credit growth and large external deficits. In turn, this took the form of significant private sector capital outflows, the tightening of credit conditions, and a shuddering halt in construction activity, with national banking systems grappling with the twin problems of rising estimates of loan losses and a liquidity squeeze in funding markets. In turn, the combined impact of domestic recessions, banking-sector distress, and the decline in risk appetite among international investors would fuel the conditions for a sovereign debt crisis.

The Financial Crisis and the Sovereign Debt Crisis

August 2007 marked the first phase of the global financial crisis, with the initiation of liquidity operations by the European Central Bank. The high exposure of major European banks to losses in the U.S. market in asset-backed securities has been well documented, as has the dependence of these banks on U.S. money markets as a source of dollar finance (McGuire and von Peter 2009; Acharya and Schnabl 2010; Shin 2012). The global crisis entered a more acute phase in September 2008 with the collapse of Lehman Brothers. The severe global financial crisis in late 2008 and early 2009 shook Europe as much as the United States.

From Financial Shock to Sovereign Debt Crisis

Through 2008 and 2009, there was relatively little concern about European sovereign debt. Instead, the focus was on the actions of the European Central Bank to address the global financial shock. In tandem with the other major central banks, it slashed short-term interest rates, provided extensive euro-denominated liquidity, and entered into currency swap arrangements to facilitate access by European banks to dollar-denominated liquidity.

But the global financial shock had asymmetric effects across the euro area. Cross-border financial flows dried up in late 2008, with investors repatriating funds to home markets and reassessing their international exposure levels (Milesi-Ferretti and Tille 2011). This process disproportionately affected countries with the greatest reliance on external funding, especially international short-term debt markets. Inside the euro area, Ireland was the most striking example: the high dependence of Ireland's banking system on international short-term funding prompted its government at the end of September 2008 to provide an extensive two-year liability guarantee to its banks (Honohan 2010; Lane 2011).

More generally, the global financial crisis prompted a reassessment of asset prices and growth prospects, especially for those countries that displayed macroeconomic imbalances. For instance, Lane and Milesi-Ferretti (2011) show that the pre-crisis current account deficit and rate of domestic credit expansion are significant correlates of the scale of the decline in output and expenditure between 2007 and 2009, while Lane and Milesi-Ferretti (forthcoming) show that "above-normal" current account deficits during 2005–2008 were associated with sharp current account reversals and expenditure reductions between 2008–2010. The cessation of the credit boom was especially troubling for Ireland and Spain, since the construction sectors in these countries had grown rapidly. The decline in construction was a major shock to domestic economic activity, while abandoned projects and falling property prices indicated large prospective losses for banks that had made too many property-backed loans.

Still, euro area sovereign debt markets remained relatively calm during 2008 and most of 2009. During this period, the main focus was on stability of the area-wide banking system, with country-specific fiscal risks remaining in the background. Furthermore, the relatively low pre-crisis public debt ratios of Ireland and Spain

gave some comfort that these countries could absorb the likely fiscal costs associated with a medium-size banking crisis. Demand for sovereign debt of euro area countries was also propped up by banks that valued government bonds as highly rated collateral in obtaining short-term loans from the European Central Bank (Buiter and Sibert 2006).

In late 2009, the European sovereign debt crisis entered a new phase. Late that year, a number of countries reported larger-than-expected increases in deficit/GDP ratios. For example, fiscal revenues in Ireland and Spain fell much more quickly than GDP, as a result of the high sensitivity of tax revenues to declines in construction activity and asset prices. In addition, the scale of the recession and rising estimates of prospective banking-sector losses on bad loans in a number of countries also had a negative indirect impact on sovereign bond values, since investors recognized that a deteriorating banking sector posed fiscal risks (Mody and Sandri 2012).

However, the most shocking news originated in Greece. After the general election in October 2009, the new government announced a revised 2009 budget deficit forecast of 12.7 percent of GDP—more than double the previous estimate of 6.0 percent.[2] In addition, the Greek fiscal accounts for previous years were also revised to show significantly larger deficits. This revelation of extreme violation of the euro's fiscal rules on the part of Greece also shaped an influential political narrative of the crisis, which laid the primary blame on the fiscal irresponsibility of the peripheral nations, even though the underlying financial and macroeconomic imbalances were more important factors.

These adverse developments were reflected in rising spreads on sovereign bonds. For example, the annual spread on ten-year sovereign bond yields between Germany and countries such as Greece, Ireland, Portugal, Spain, and Italy was close to zero before the crisis. Remember that sovereign debts from these countries are all denominated in a common currency, the euro, so differences in expected yield mainly represent perceived credit risks and differences in volatility.

Figure 2 shows the behavior of country-level ten-year bond yields for seven euro area countries from October 2009 through June 2012. Three particularly problematic periods stand out. First, the Greek yield began to diverge from the group in early 2010, with Greece requiring official assistance in May 2010. Second, there was strong comovement between the Irish and Portuguese yields during 2010 and the first half of 2011 (Ireland was next to require a bailout in November 2010, with Portugal following in May 2011). Third, the yields on Italy and Spain have moved together, with these spreads at an intermediate level between the bailed out countries and the core countries of Germany and France. For Italy and Spain, the spread against Germany rose above 300 basis points in July 2011 and remained at elevated levels thereafter. In 2011, a visible spread also emerges between the French and

---

[2] See also Gibson, Hall, and Tavlas (2012). These authors also point out that the Greek announcement was coincidentally soon followed by the surprise request from Dubai World for a debt moratorium, such that the climate in international debt markets markedly deteriorated in October/November 2009.

Figure 2
Yields on Ten-Year Sovereign Bonds, October 2009 to June 2012
(percent)
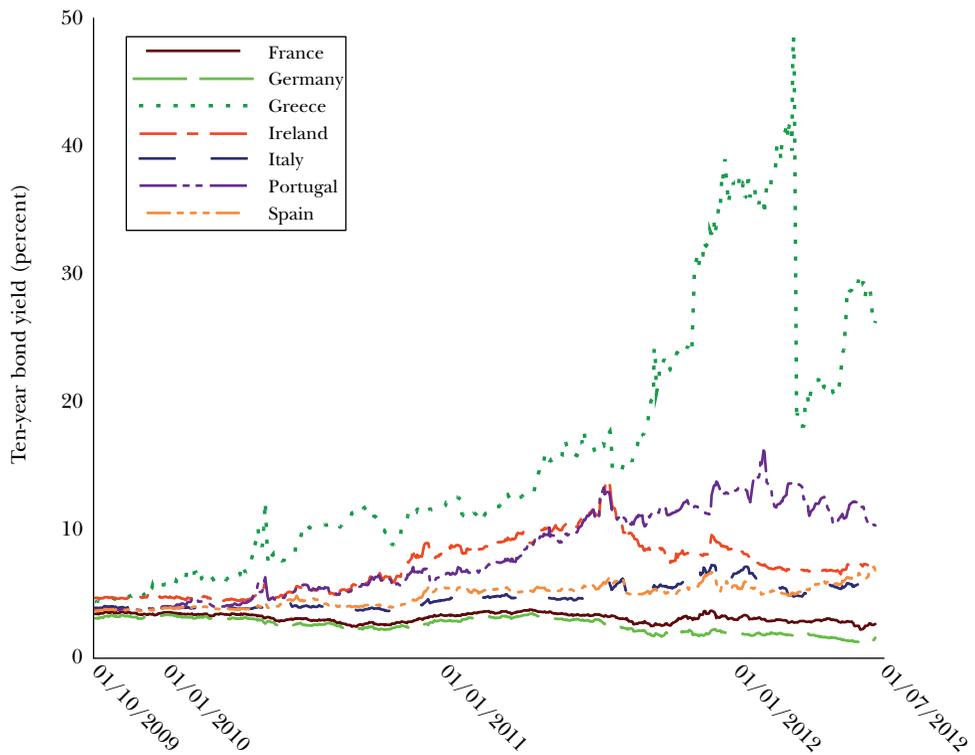
Source: Author's calculations based on data from Datastream.

German yields, although the greater relative vulnerability of France is not pursued in this paper.

Cobbling Together a Response to the Sovereign Debt Crisis

   Greece was the first country to be shut out of the bond market in May 2010, with Ireland following in November 2010, and Portugal in April 2011. (In June 2012, Spain and Cyprus also sought official funding. At the time of writing, it is unclear whether Spain will require only a limited form of official funding to help it recapitalize its banking system or a larger-scale bailout.)

   In each of the three bailouts, joint European Union/IMF programs were established under which three-year funding would be provided on condition that the recipient countries implemented fiscal austerity packages and structural reforms to boost growth (especially important in Greece and Portugal) and recapitalized and deleveraged overextended banking systems (especially important in Ireland). The scale of required funding far exceeded normal IMF lending levels, so the European

Union was the major provider of funding. At that time, it was also decided to set up a temporary European Financial Stability Facility that could issue bonds on the basis of guarantees from the member states in order to provide official funding in any future crises. In addition, the pre-existing European Stability Mechanism, which had previously only been used for balance-of-payments foreign currency support for non–euro member countries, was adapted to also provide funding for euro member countries.

In principle, a temporary period of official funding can benefit all parties. For the borrower, it can provide an opportunity for a government to take the typically unpopular measures necessary to put public finances on a trajectory that converges on a sustainable medium-term path, while also implementing structural reforms that can boost the level of potential output. For the lender, avoiding default can benefit their creditor institutions (especially banks), while guarding against possible negative international spillovers from a default.

The details of the funding plans for Greece, Ireland, and Portugal largely copied standard IMF practices, but they faced a number of potential problems. Here are six issues, in no particular order.

First, given the scale of macroeconomic, financial, and fiscal imbalances, the plausible time scale for macroeconomic adjustment was longer than the standard three-year term of such deals. In particular, fiscal austerity by individual member countries cannot be counterbalanced by a currency devaluation or an easing in monetary conditions, which is especially costly if a country has to simultaneously close both fiscal and external deficits. By June 2011, it was clear that Greece would need a second package, while it is also likely that Ireland and Portugal will not be able to obtain full market funding after the expiry of their current deals. The slow pace of adjustment was also recognized in Summer 2011 through the extension of the repayment period on the official debt from 7.5 years to 15–30 years.

Second, in related fashion, excessively rapid fiscal consolidation can exacerbate weaknesses in the banking system. Falling output and a rising tax burden shrinks household disposable income and corporate profits, increasing private sector default risk. This was identified as an especially strong risk in the Irish program in view of the scale of household debt.

Third, the fiscal targets were not conditional on the state of the wider European economy. As growth projections for the wider European economy declined throughout 2011, the country-specific targets looked unobtainable for external reasons.

Fourth, the original bailouts included a sizable penalty premium of 300 basis points built into the interest rate, which is standard IMF practice. A penalty rate discourages countries from the moral hazard of taking such loans when not really needed and also compensates the funders for the nontrivial default risk. However, it also makes repaying the loans harder and gives an appearance that the creditor EU countries are profiteering at the expense of the bailed-out countries. This penalty premium on the European component of the official loans was eliminated in July 2011, although the interest rate on the IMF-sourced component of the funds continued to include a penalty premium.

Fifth, the bailout funds have been used to recapitalize banking systems, in addition to covering the "regular" fiscal deficits. So far, this element has been most important in the Irish bailout, but it was also a feature of the Greek and Portuguese bailouts; it is also the primary element in the official funding requested by Spain in June 2012. While publicly funded recapitalization of troubled banks can ameliorate a banking crisis, this strategy is problematic if it raises public debt and sovereign risk to an excessive level (Acharya, Drechsler, and Schnabl 2010). Moreover, excessive levels of sovereign debt can amplify a banking crisis for several reasons: domestic banks typically hold domestic sovereign bonds; a sovereign debt crisis portends additional private sector loan losses for banks; and a highly indebted government is likely to lean on banks to provide additional funding (Reinhart and Sbrancia 2011). Furthermore, the generally poor health of major European banks and the cross-border nature of financial stability inside a monetary union means that national governments are under international pressure to rescue failing banks in order to avoid the cross-border contagion risks from imposing losses on bank creditors.[3] Despite these international externalities, at least until mid 2012, the only type of European funding for bank rescues was plain-vanilla official loans to the national sovereign, with fixed repayment terms. Under this approach, the fates of national sovereigns and national banking systems remain closely intertwined.

The sixth issue involves a standard IMF principle that funding is only provided if the sovereign debt level is considered to be sustainable. If it is not sustainable, the traditional IMF practice has been to require private sector creditors to agree to a reduction in the present value of the debt owed to them. Under the joint EU–IMF programs, such "private sector involvement" was not initially deemed necessary in the three bailouts of 2010 and 2011.

The argument against requiring private sector involvement is that it can spook an already nervous sovereign debt market. For example, when the prospect of requiring private sector involvement was broached in October 2010 (in the Franco-German "Deauville Declaration"), interest rate spreads immediately increased, especially for Greece, Ireland, and Portugal. Ireland's efforts to avoid a bailout came to a halt soon thereafter in early November 2010. European banks also had increased difficulties in raising funds, especially the local banks in the troubled periphery, in line with the increase in the perceived riskiness of their home governments.

The March 2012 agreement to provide Greece with a second bailout package did require that private sector creditors accept a haircut, which eventually turned out to be about 50 percent of value, which is equal to 47 percent of Greek GDP.[4] But

---

[3] The poor design of European bank resolution regimes has also increased the fiscal cost of rescuing banks, since it is difficult to shut down failing banks and impose losses on holders of the senior bonds issued by banks.

[4] Although the plausibility of this projection has been disputed by many commentators, the second bailout package is officially projected to deliver a Greek debt/GDP ratio of 120 percent by 2020, which is a shade above the debt ratios of the some of the other troubled euro member countries. See also Ardagna and Caselli (2012) for an account of the Greek crisis.

as this requirement was discussed during the course of 2011, it contributed to the sharp widening of the spreads on Spanish and Italian debt.

Listing some of the difficulties in this way may make the European response to its sovereign debt issues appear more coherent than it has actually been. Instead, it may be fair to characterize Europe's efforts to address its sovereign debt problem as makeshift and chaotic, at least through the middle of 2012.

Risks of Multiple Equilibria when Sovereign Debt is High

A significant factor during the crisis has been the increased volatility in euro area sovereign debt markets. A country with a high level of sovereign debt is vulnerable to increases in the interest rate it pays on its debt (Calvo 1988; Corsetti and Dedola 2011). This risk can give rise to self-fulfilling speculative attacks: an increase in perceptions of default risk induces investors to demand higher yields, which in turn makes default more likely. In contrast, if default risk is perceived to be low, interest rates remain low, and default does not occur. This multiple equilibria problem may have greater force in the context of a multicountry currency union, since a small adverse shift in the fundamentals of one individual country can trigger a large decline in demand for the sovereign debt of that country as investors "run for the exit" and switch to sovereign debt of other safer euro area countries.

What policies might encourage the "good" equilibrium? One option is to create a firewall through the availability of an official safety net. This would reduce the risk of the "bad" equilibrium arising because investors would not need to fear that a country will be pushed into involuntary default by an inability to rollover its debt. As of mid 2012, the available funding through the European Financial Stability Facility and its successor, the European Stability Mechanism, was only enough to address the bailouts of Greece, Ireland, and Portugal—and thus not nearly sufficient to offer substantial support to Spain and/or Italy. Proposals to create a large firewall fund are politically unpopular in creditor countries for many reasons, including fear of taking losses, and concerns that such a fund would tempt politicians in at-risk countries to postpone or avoid tough fiscal and structural reform decisions.

The European Central Bank's program to purchase sovereign bonds can also be viewed as a way to reduce the risk of the "bad" equilibrium. Between May 2010 and October 2010, about 65 billion euro of bonds were bought by the ECB; a further 125 billion euro were committed during the market turmoil between August 2011 and November 2011 such that the cumulative bond holdings grew to over 200 billion euros (about 2 percent of euro area GDP). The ECB has taken pains to emphasise that these purchases are not monetizing debt because liquidity created is canceled out through offsetting sterilization operations. Instead, the program seeks to provide liquidity and depth when certain sovereign debt markets are troubled. A useful analogy here is to the modern argument for currency market interventions. Such interventions do not try to fix asset values; instead, limited intervention by a central bank can be temporarily stabilizing by breaking momentum dynamics.

There have also been calls for the European Central Bank to take further steps to stabilize the sovereign debt market (for example, De Grauwe 2012). At

one level, it could increase the firepower of the European Stability Mechanism by allowing it to borrow from the ECB. Going further, the ECB could announce a ceiling to the interest rate it would tolerate on the sovereign debt of countries that meet certain fiscal criteria (such as taking credible steps to ensure debt declines to a safe level over the medium term), and guarantee to buy the debt at that price if needed.

Even more controversially, outright debt monetization might be viewed in some quarters as preferable to outright default by large member countries if it becomes clear that solvency concerns are so great that market funding will not be available for an extended period. While debt monetization exceeds the current legal mandate of the European Central Bank, debate over these proposals might heat up if a more acute and severe phase of the crisis were to take hold. At least for now, it is hard to envisage that such a change would be supported by all member countries of the euro area. However, it is also important to appreciate that the reserve capacity to monetize debt is commonly cited as the reason why highly indebted governments such as Japan, the United Kingdom, and the United States are still able to borrow at low interest rates.

Prospects for Post-Crisis Reduction in Sovereign Debt

The legacy of the euro area sovereign debt crisis is that a number of countries will have dangerously elevated public debt ratios, while others will have debt levels that are lower by comparison but still high relative to long-term normal values. Even if current austerity programs are sufficient to stabilize debt ratios, there remains the post-crisis adjustment challenge of gradually reducing government debt to safer levels. This medium-term challenge is viewed with trepidation in European circles. Consider four reasons why the underlying fundamentals for reducing the debt/GDP ratio are not promising.

First, growth in nominal GDP is likely to be low. Debt/GDP ratios are stickier in high-income countries than in emerging economies in part because there is less scope for rapid output growth in the former group of countries. There is nothing to suggest that real growth rates for advanced economies should exceed a long-term annual average of about 2 percent. Indeed, real annual growth of 2 percent may be optimistic given several factors: the erosion of human capital from the prolonged unemployment of the last few years (DeLong and Summers 2012); the likelihood of tax increases and reduced public investment; and the historical pattern that output growth can be compromised for a decade in the aftermath of a banking crisis (Reinhart and Rogoff 2010). For the most-indebted countries, nominal GDP is unlikely to grow much faster than real GDP. The European Central Bank has a 2 percent aggregate inflation target (approximately), and the most indebted member countries are likely to have average inflation substantially below that level in view of the correlation between domestic demand and the price level of nontradables (Lane and Milesi-Ferretti 2004).

Second, the political economy environment is likely to be challenging. The highly indebted countries will need to be led by governments that must impose spending cuts and tax increases with no short-term prospect of fiscal relaxation. Adjustment fatigue can set in, making it difficult to sustain long-term fiscal austerity.

Third, the possibilities for financing at least some of the sovereign debt through "financial repression" are limited. This approach uses tight regulations on domestic financial institutions—including banks, pension funds, and others—so that these institutions are pressured to put a greater portion of their assets than they would otherwise choose into sovereign debt (Reinhart and Sbrancia 2011). However, the principle of open capital markets across the European Union means that countries have fairly limited scope for financial repression in comparison to what was historically possible.

Fourth, risk premia will likely remain nontrivial for most indebted member countries. The large losses experienced by private sector investors in Greek sovereign debt underline that the sovereign debt of euro area member countries can no longer be categorized as risk-free investments. Indeed, the historical evidence suggests that further rounds of debt restructuring will form part of the adjustment process (see also the discussion by Reinhart, Reinhart, and Rogoff in this issue).

Accordingly, the medium-term outlook suggests that sovereign debt is likely to pose significant policy challenges for the euro area over the next few years. The next section outlines some possible reforms that could help to alleviate the situation and avoid a similar disaster in the future.

Reforms to Address Sovereign Debt Concerns

The high outstanding sovereign debt levels and the importance of avoiding future fiscal crises in the euro area have induced reforms to the fiscal rules for the euro area, with a new Fiscal Compact Treaty that is scheduled to go into effect at the start of 2013 (if it is ratified by 12 members of the euro area by then). The Fiscal Compact requires that the new fiscal principles be embedded in each country's national legislation. These fiscal governance reforms are based on two principles: first, high public debt levels pose a threat to fiscal stability; and, second, the fiscal balance should be close to zero "over the cycle."

The operation of the pre-crisis fiscal rules focused on the overall budget balance, with a maximum annual budget deficit set at 3 percent of GDP, while there was no strong pressure on highly indebted countries (such as Greece and Italy) to reduce debt levels below the specified 60 percent ceiling. Even on its own terms, this approach had two main defects: it did not adequately allow for cyclical variation in budget positions, and it did not provide much discipline for countries inside the limit.

In contrast, the new system focuses on the structural budget balance, thus stripping out cyclical effects and one-off items. A structural budget balance

target encourages a government to bank cyclical revenue gains during upturns in exchange for a greater slippage in the overall budget balance during recessions. Under the new system, there is a specified time frame for reducing public debt below the ceiling of 60 percent of GDP, with the excess above the ceiling eliminated at an average rate of "one twentieth" each year.

This new approach faces several implementation problems. For example, a fiscal framework based on structural budget balance faces knotty measurement problems because it requires macroeconomic forecasters to differentiate between cyclical fluctuations and trend fluctuations in output almost in real time. For this reason, the Fiscal Compact requires that governments enact a mechanism that requires adjustments if the forecast errors for the structural budget balance cumulate over several years to a significant level. In the German fiscal law, for example, a cumulative overshoot above 1.5 percent of GDP requires a gradual correction by running tighter structural budgets until the excess is eliminated (Bundesbank 2011).

Another potential issue is that, in contrast to the original Stability and Growth Pact, the primary source of fiscal discipline is intended to be national. The Fiscal Compact requires that the fiscal rules are written into domestic legislation and that national independent fiscal councils be created to monitor the compliance with the specified fiscal rules. The hope is that national-level discipline will be more effective, since it should have greater political legitimacy than external surveillance. However, external surveillance and the threat of external sanctions remain as a "second line of defense" against fiscal misbehavior.

In recognition that fiscal stability can be quickly undone by financial and macroeconomic shocks, the Fiscal Compact is accompanied by new European regulations that go beyond narrow fiscal governance in monitoring "excessive imbalances." A wide range of risk indicators will be tracked, including credit growth, house price indices, and external imbalances. The intention is that a country experiencing severe imbalances should respond with policy interventions to mitigate crisis risks and improve resilience. However, it remains unclear whether national governments have the capacity to identify excessive imbalances accurately or to deploy policy instruments that can be effective in managing such risk factors.

Given the limited nature of these initiatives, more extensive reforms are also under discussion. A partial list of such proposals includes the following. First and foremost is the creation of a banking union, since the diabolic loop between national banking systems and national sovereigns has been central to the fiscal crisis. The ingredients of banking union are well-known and include European-level regulatory responsibility, deposit insurance, bank resolution policies, and a joint fiscal backstop in the event that fiscal resources were deemed necessary to stabilize the banking system (Allen, Beck, Carletti, Lane, Schoenmaker, and Wagner 2011; Brunnermeier et al. 2011; Marzinotto, Sapir, and Wolff 2011). A partial move in this direction was announced at the June 2012 European Council meeting, which also opened up the possibility of the European Stability Mechanism making direct equity injections into troubled banks. However, the details of these new plans have yet to be ironed out.

A second step is the introduction of common areawide "eurobonds," with the goal of avoiding the disruptive impact of destabilizing speculative attacks on national sovereign debt markets inside the euro area (Favero and Missale 2012). Fiscally stronger member states might support eurobonds if it is cheaper than the alternatives for reducing default risk, for instance with bigger bailout funds. To prevent fiscally weaker member states from using eurobonds to overborrow, these could be restricted in various ways. One option is to limit eurobonds to short maturities, so that ill-disciplined countries could quickly be denied access to funding (Philippon and Hellwig 2011); another option is to limit eurobond funding only for sovereign debt up to 60 percent of GDP, with the excess still requiring funding through the issuance of national bonds (Delpla and Von Weizsäcker 2011); or eurobonds could be limited to countries that satisfy certain criteria for good macroeconomic and fiscal fundamentals (Muellbauer 2011).[5]

Alternatively, Brunnermeier et al. (2011) point out that many of the advantages of eurobonds can be obtained even if sovereign debt remains a national responsibility. In particular, a European Debt Agency could be established that would buy up large quantities of national sovereign bonds (up to a limit of 60 percent of GDP in each case). This agency would be funded by the issuance of two tranches of bonds—European Safe Bonds and European Junior Bonds—with the latter having the primary exposure in the event of defaults on the underlying portfolio of national sovereign bonds. Accordingly, the senior European Safe Bonds should be safe assets, which in turn should make them preferred collateral for central bank liquidity operations. Since this proposal does not require joint backing of sovereign debt issues, it avoids the moral hazard problems that plague the eurobond proposals.

Third, Europe might seek *a deeper level of fiscal union*, agreeing to share certain tax streams or spending programs in a way that would be delinked from fluctuations in national-level output. In related fashion, *enhanced coordination of national fiscal policies* would also be helpful, thereby enabling the collective fiscal position of the euro area to be appropriately calibrated in relation to the prevailing macroeconomic conditions.

Many of these policy proposals would require changes in the treaties governing the European Union and imply a transformative increase in the level of political integration. Paradoxically, the European crisis has generated severe political tensions across the member states, while at the same time prompting much discussion of the desirability of more extensive types of political union. In this debate, the parallels with the historical development of fiscal federalism in the United States have been well-flagged (Henning and Kessler 2012; Sargent 2012).

---

[5] A temporary type of eurobond has been suggested by the German Council of Economic Experts (Bofinger, Feld, Franz, Schmidt, and Weder di Mauro 2011). Under this proposal, a jointly-backed Debt Redemption Fund would refinance the excess debt above 60 percent of GDP, thereby relieving the roll-over pressures facing highly-indebted countries. Once debt levels fall back to the 60 percent ceiling, the Debt Redemption Fund would be wound up.

In conclusion, the origin and propagation of the European sovereign debt crisis can be attributed to the flawed original design of the euro. In particular, there was an incomplete understanding of the fragility of a monetary union under crisis conditions, especially in the absence of banking union and other European-level buffer mechanisms. Moreover, the inherent messiness involved in proposing and implementing incremental multicountry crisis management responses on the fly has been an important destabilizing factor throughout the crisis.

The most benign perspective on the European sovereign debt crisis is that it provides an opportunity to implement reforms that are necessary for a stable monetary union but that would not have been politically feasible in its absence. A more modest hope is that the unfolding reform process will deliver a monetary union that can survive, even if it remains vulnerable to recurring crises. However, the alternative scenario in which the single European currency implodes is no longer unthinkable, even if it would unleash the "mother of all financial crises" (Eichengreen 2010). The stakes are high.

References

Acharya, Viral V., and Philipp Schnabl. 2010. "Do Global Banks Spread Global Imbalances? The Case of Asset-Backed Commercial Paper during the Financial Crisis of 2007–09." *IMF Economic Review* 58(1): 37–73.

Acharya, Viral V., Itamar Drechsler, and Philipp Schnabl. 2010. "A Pyrrhic Victory? Bank Bailouts and Sovereign Credit Risk." *CEPR Discussion Paper 8679.*

Allen, Franklin, Thorsten Beck, Elena Carletti, Philip R. Lane, Dirk Schoenmaker, and Wolf Wagner. 2011. *Cross-Border Banking in Europe: Implications for Financial Stability and Macroeconomic Policies.* CEPR Report. Center for Economic Policy Research.

Ardagna, Silvia, and Francesco Caselli. 2012. "The Political Economy of the Greek Debt Crisis: A Tale of Two Bailouts." Centre for Economic

Performance Special Paper No. 25, London School of Economics.

Beetsma, Roel, and Harald Uhlig. 1999. "An Analysis of the Stability and Growth Pact." *Economic Journal* 109(458): 546–71.

Benetrix, Agustin, and Philip R. Lane. 2012. "Fiscal Cyclicality and EMU." Unpublished paper, Trinity College Dublin.

Blanchard, Olivier. 2007. "Current Account Deficits in Rich Countries." *IMF Staff Papers* 54(2): 191–219.

Blanchard, Olivier, and Francesco Giavazzi. 2002. "Current Account Deficits in the Euro Area: The End of the Feldstein–Horioka Puzzle?" *Brookings Papers on Economic Activity,* no. 2, pp. 147–86.

Bofinger, Peter, Lars P. Feld, Wolfgang Franz, Christoph M. Schmidt, and Beatrice Weder di

Mauro. 2011. "A European Redemption Pact." VoxEU, November 9.

Brunnermeier, Markus, Luis Garicano, Philip R. Lane, Marco Pagano, Ricardo Reis, Tano Santos, David Thesmar, Stijn Van Neiuwerburgh, and Dimitri Vayanos. 2011. "European Safe Bonds (ESBies)." http://euro-nomics.com/wp-content /uploads/2011/09/ESBiesWEBsept262011.pdf.

Buiter, Willem, Giancarlo Corsetti, and Nouriel Roubini. 1993. "'Excessive Deficits': Sense and Nonsense in the Treaty of Maastricht." *Economic Policy* 8(16): 57–100.

Buiter, Willem, and Anne Sibert. 2006. "How the Eurosystem's Treatment of Collateral in its Open Market Operations Weakens Fiscal Discipline in the Eurozone (and What to Do about It)." In *Fiscal Policy and the Road to the Euro,* 29–60. National Bank of Poland.

Bundesbank. 2011. "The Debt Brake in Germany: Key Aspects and Implementation." *Monthly Bulletin,* October, pp. 15–39.

Caruana, Jaime, and Stefan Adjiev. 2012. "Sovereign Creditworthiness and Financial Stability: An International Perspective." *Banque de France Financial Stability Review,* no. 16, April, pp. 71–85.

Calvo, Guillermo. 1988. "Servicing the Public Debt: The Role of Expectations." *American Economic Review* 78(4): 647–61.

Chen, Ruo, Gian Maria Milesi-Ferretti, and Thierry Tressel. Forthcoming. "Euro Area Debtor Countries: External Imbalances in the Euro Area." *Economic Policy.*

Corsetti, Giancarlo, and Luca Dedola. 2011. "Fiscal Crises, Confidence and Default: A Bare-Bones Model with Lessons for the Euro Area." https://sites.google.com/site/giancarlocorsetti /main/calvo-code.pdf?attredirects=0.

De Grauwe, Paul. 2012. "Fragile Eurozone in Search of a Better Governance." *Economic and Social Review* 43(1): 1–30.

DeLong, J. Bradford, and Lawrence H. Summers. 2012. "Fiscal Policy in a Depressed Economy." *Brooking Papers on Economic Activity,* forthcoming.

Delpla, Jacques, and Jakob Von Weizsäcker. "Eurobonds: The Blue Bond Concept and Its Implications." Bruegel Policy Contribution 2011/02.

Eichengreen, Barry. 2010. "The Breakup of the Euro Area." In *Europe and the Euro,* edited by Alberto Alesina and Franceso Giavazzi, 11–56. University of Chicago Press.

Fagan, Gabriel, and Vitor Gaspar. 2007. "Adjusting to the Euro." *ECB Working Paper 716.*

Favero, Carlo, and Alessandro Missale. 2012. "Sovereign Spreads in the Euro Area: Which Prospects for a Eurobond?" *Economic Policy* 27(70): 231–73.

Feldstein, Martin. 1997. "The Political Economy of an European Economic and Monetary Union: Political Sources of an Economic Liability." *Journal of Economic Perspectives* 11(4): 23–42.

Freund, Caroline, and Frank Warnock. 2007. "Current Account Deficits in Industrial Countries: The Bigger They Are, The Harder They Fall?" In *G7 Current Account Imbalances: Sustainability and Adjustment,* edited by Richard Clarida, 133–68. University of Chicago Press.

Gali, Jordi. 2010. "Notes on the Euro Debt Crisis." http://www.crei.cat/people/gali/debt%20 crisis%2003%2006.pdf.

Gali, Jordi, and Tommaso Monacelli. 2008. "Optimal Monetary and Fiscal Policy in a Currency Union." *Journal of International Economics* 76(1): 116–32.

Giavazzi, Francesco, and Luigi Spaventa. 2011. "Why the Current Account Matters in a Monetary Union." In *The Euro Area and The Financial Crisis,* edited by Miroslav Beblavy, David Cobham, and L'udovit Odor, 59–80. Cambridge University Press.

Gibson, Heather D., Stephen G. Hall, and George S. Tavlas. 2012. "The Greek Financial Crisis: Growing Imbalances and Sovereign Spreads." *Journal of International Money and Finance* 31(3): 498–516.

Gourinchas, Pierre-Olivier, and Maurice Obstfeld. 2012. "Stories of the Twentieth Century for the Twenty-First." *American Economic Journal: Macroeconomics* 4(1): 226–65.

Henning, C. Randall, and Martin Kessler. 2012. "Fiscal Federalism: US History for Architects of Europe's Fiscal Union." *Bruegel Essays and Lectures Series,* Brussels: Bruegel.

Honohan, Patrick. 2010. "The Irish Banking Crisis: Regulatory and Financial Stability Policy 2003–2008." Report for the Commission of Investigation into the Banking Sector in Ireland. Available at: http://www.bankinginquiry.gov.ie /Preliminary_Reports.aspx.

Honohan, Patrick, and Daniela Klingebiel. 2003. "The Fiscal Cost Implications of an Accommodating Approach to Banking Crises." *Journal of Banking and Finance* 27(8): 1539–60.

Lane, Philip R. 2006. "The Real Effects of European Monetary Union." *Journal of Economic Perspectives* 20(4): 47–66.

Lane, Philip R. 2010. "Some Lessons for Fiscal Policy from the Financial Crisis." *Nordic Economic Policy Review* 1(1): 13–34.

Lane, Philip R. 2011. "The Irish Crisis." In *The Euro Area and The Financial Crisis,* edited by Miroslav Beblavy, David Cobham, and L'udovit Odor, 59–80. Cambridge University Press.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2004. "The Transfer Problem Revisited: Real Exchange Rates and Net Foreign Assets." *Review of Economics and Statistics* 86(4): 841–57.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2007. "Europe and Global Imbalances." *Economic Policy* 22(51): 519–73.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2011. "The Cross-Country Incidence of the Global Crisis." *IMF Economic Review* 39(1): 77–110.

Lane, Philip R., and Gian Maria Milesi-Ferretti. Forthcoming. "External Adjustment and the Global Crisis." *Journal of International Economics.*

Lane, Philip R., and Barbara Pels. 2012. "Current Account Imbalances in Europe." *Moneda y Credito,* forthcoming.

Lane, Philip R., and Peter McQuade. 2012. "Domestic Credit Growth and International Capital Flows." Unpublished paper, Trinity College Dublin.

Marzinotto, Benedicta, Andre Sapir, and Guntram B. Wolff. 2011. "What Kind of Fiscal Union?" *Bruegel Policy Brief No. 2011/06.*

McGuire, Patrick, and Goetz von Peter. 2009. "The US Dollar Shortage in Global Banking and the International Policy Response." *BIS Working Paper No. 291.*

Milesi-Ferretti, Gian Maria, and Cedric Tille. 2011. "The Great Retrenchment: International Capital Flows during the Global Financial Crisis." *Economic Policy* 26(66): 285–342.

Mody, Ashoka, and Damiano Sandri. 2012. "The Eurozone Crisis: How Banks and Sovereigns Came to be Joined at the Hip." *Economic Policy* 27(70): 199–230.

Muellbauer, John. 2011. "Resolving the Eurozone Crisis: Time for Conditional Eurobonds." *CEPR Policy Insight No. 59.*

Philippon, Thomas, and Christian Hellwig. 2011/ "Eurobills, Not Eurobonds." VoxEU, December 2. http://www.voxeu.org/article/eurobills-not-euro-bonds.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time Is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010. "Growth in a Time of Debt." *American Economic Review* 100(2): 573–78.

Reinhart, Carmen M., and M. Belen Sbrancia. 2011. "The Liquidation of Government Debt." *NBER Working Paper 16893.*

Sargent, Thomas J. 2012. "United States Then, Europe Now." *Journal of Political Economy* 120(1): 1–40.

Shambaugh, Jay. 2012. "The Euro's Three Crises." *Brookings Papers on Economic Activity,* forthcoming.

Shin, Hyun Song. 2012. "Global Banking Glut and Loan Risk Premium." *IMF Economic Review,* forthcoming.

Wyplosz, Charles. 1997. "EMU: Why and How It Might Happen." *Journal of Economic Perspectives* 11(4): 3–22.

Wyplosz, Charles. 2006. "European Monetary Union: The Dark Sides of a Major Success." *Economic Policy* 21(46): 207–261.

Public Debt Overhangs: Advanced-Economy Episodes Since 1800[†]

Carmen M. Reinhart, Vincent R. Reinhart, and Kenneth S. Rogoff

The recent financial crisis and recession has left a legacy of historically high and rising level of public indebtedness across the advanced economies. The central policy debate across Europe, Japan, and the United States now centers on how fast to stabilize soaring public debt/GDP ratios, given that post-crisis growth remains fragile. We bring evidence to bear on the issue by identifying the major public debt overhang episodes in advanced economies since the early 1800s. Following Reinhart and Rogoff (2010), we select stretches where gross public debt exceeds 90 percent of nominal GDP on a sustained basis. Such public debt overhang episodes are associated with lower growth than during other periods. Even more striking, among the 26 episodes we identify, 20 lasted more than a decade. The long duration belies the view that the correlation is caused mainly by debt buildups during business cycle recessions. The long duration also implies that the cumulative shortfall in output from debt overhang is potentially massive. These growth-reducing effects of high public debt are apparently not transmitted exclusively through high real interest rates, in that in eleven of the episodes, interest rates are not materially higher.

■ *Carmen M. Reinhart is Minos Zombanakis Professor of the International Financial System, Harvard Kennedy School of Government, Cambridge, Massachusetts. She is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and Research Fellow, Centre for Economic Policy Research, London, England. Vincent R. Reinhart is Chief U.S. Economist at Morgan Stanley, New York City, New York. Kenneth S. Rogoff is Professor of Economics and Thomas D. Cabot Professor of Economics, Harvard University, Cambridge, Massachusetts, and Research Associate, National Bureau of Economic Research. Their email addresses are ⟨Carmen_Reinhart@harvard.edu⟩, ⟨vincent .reinhart@morganstanley.com⟩, and ⟨krogoff@harvard.edu⟩.*

In this paper, we use the long-dated cross-country data on public debt developed by Reinhart and Rogoff (2009) to examine the growth and interest rates associated with prolonged periods of exceptionally high public debt, defined as episodes where public debt to GDP exceeded 90 percent for at least five years. (The basic results here are reasonably robust to choices other than 90 percent as the critical threshold, as in Reinhart and Rogoff 2010a, b).[1] Over the years 1800–2011, we find 26 such episodes across the advanced economies. Previous studies of high public debt episodes have typically focused on the very small number of cases, including mainly the post-1970 or post-1980 cases. While data limitations may have prevented us from including every episode of high public debt in advanced economies since 1800, we are confident that this list encompasses the preponderance of such episodes. To focus on the association between high debt and long-term growth, we only cursorily treat shorter episodes lasting under five years, of which there turn out to be only a few. The long length of typical public debt overhang episodes suggests that even if such episodes are originally caused by a traumatic event such as a war or financial crisis, they can take on a self-propelling character.

Consistent with a small but growing body of research, we find that the vast majority of high debt episodes—23 of the 26—coincide with substantially slower growth. On average across individual countries, debt/GDP levels above 90 percent are associated with an average annual growth rate 1.2 percent lower than in periods with debt below 90 percent debt; the average annual levels are 2.3 percent during the periods of exceptionally high debt versus 3.5 percent otherwise. Of course, public debt overhang and slow growth are surely a simultaneous relationship: countries experiencing a period of slower growth may be more vulnerable to ending up with very high levels of public debt, and once the public debt overhang arises, countries with slower growth are going to take longer to escape it. As we shall discuss, a number of recent studies have concluded that the relationship cannot be entirely from low growth to high debt, and that very high debt likely does weigh on growth. Those who view the correlation from high debt to slower growth as mainly due to the cyclical effects of slowdowns on public finances will need to address certain aspects of the data. For example, why does the typical episode of high public debt last far beyond any plausible business cycle frequency—decades, not years? Also, if the debt-to-growth correlation is driven by business cycles, then why so little correlation between debt and growth below the 90 percent debt/GDP threshold, yet such a pronounced correlation above it?

Another contribution of this paper is to provide, to our knowledge, the first systematic evidence on the association between public debt overhang and real

---

[1] In Reinhart and Rogoff (2010a), the annual observations are grouped into four categories, according to the ratio of debt to GDP during that particular year, as follows: years when debt to GDP levels were below 30 percent (low debt); and years where debt/GDP was 30 to 60 percent (medium debt), 60 to 90 percent (high), and above 90 percent (very high). The main finding is that across both advanced countries and emerging markets, high debt/GDP levels (90 percent and above) are associated with notably lower growth outcomes. Much lower levels of external debt/GDP (60 percent) are associated with adverse outcomes for emerging market growth.

interest rates. The modern policy debate often presumes that the main cost of high public debt ultimately comes from sovereign default, with all its attendant disruptions and dislocations. However, we find that countries with a public debt overhang by no means always experience either a sharp rise in real interest rates or difficulties in gaining access to capital markets. Indeed, in 11 of the 26 cases where public debt was above the 90 percent debt/GDP threshold, real interest rates were either lower, or about the same, as during the lower debt/GDP years. This result is, for instance, consistent with the classic friction identified in Barro (1979) who, using a model where the government always pays in full, showed how ultimate debt stabilization requires raising distorting taxes or (in principle) adjusting expenditures, both of which potentially affect output.

We begin with a brief tour of the concept of debt overhangs in the advanced economies, including both public and private debt, both in historical context and relative to developments in emerging market economies. We then look more closely at the 26 debt overhang episodes we identify. In the background of this discussion, of course, lurks the rapid growth in public debt that many advanced economies have experienced in the last few years in the aftermath of financial crisis and recession. The high level of public debt in Greece has already sparked a broader crisis in the European Union, with the public debt/GDP ratios of several other European economies also a cause for concern, especially when the imputed costs of future bank bailouts are taken into account. The U.S. government surpassed a 90 percent ratio of gross federal debt/GDP in 2010, with Japan at debt/GDP levels more than twice as high.

Our work suggests that the long-term secular costs of high debt need to be weighed against the short-term expediency of Keynesian fiscal stimulus. Our work also highlights the historical importance of default, debt restructuring, and a variety of debt conversions (encompassing both voluntary and involuntary episodes) in coping with debt overhangs. "Credit events" are not just an emerging market phenomenon; these were commonplace among the advanced economies prior to World War II.

Preamble: Varieties of Debt Overhangs

Although our primary focus here is on public debt overhangs, today's high debt burdens also extend to private debt, external debt (including both government and private debt owed to foreigners), and the actuarial debt implicit in underfunded old age pension and medical care programs. Although the data for these broader debt measures is far less comprehensive across time and countries than for public debt, it seems clear that the overall magnitude of the debt burdens facing the advanced economies as a group is in many dimensions without precedent. The interaction between the different types of debt overhang is extremely complex and poorly understood, but it is surely of great potential importance. For example, the lines between public and private debt often become blurred in a

Figure 1

Gross Central Government Debt as Percent of GDP: Advanced and Emerging Market Economies, 1860–2011

(unweighted averages)
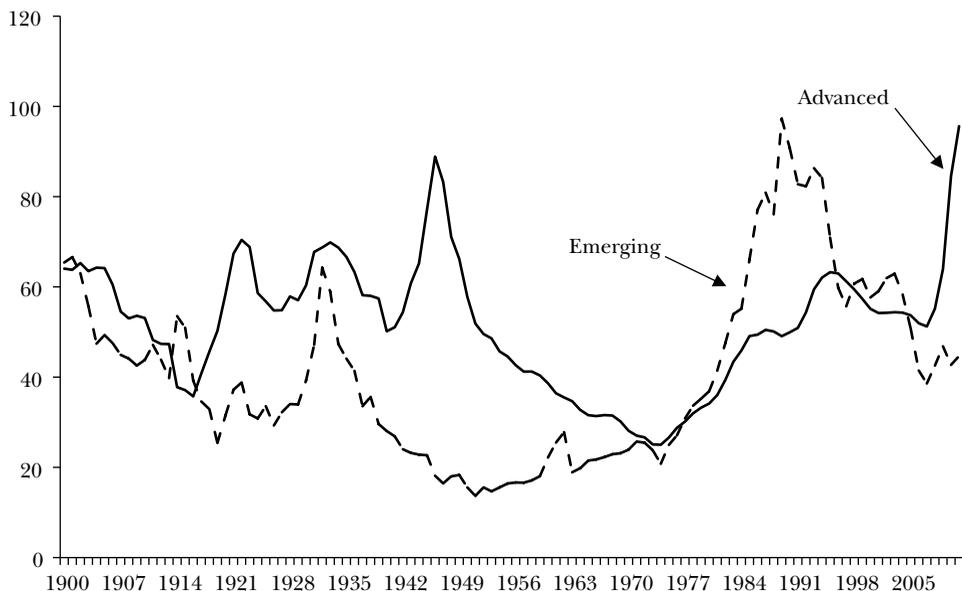
Sources: Reinhart and Rogoff (2009) and sources cited therein.

crisis, as for example in Ireland where the government took on massive quantities of bank debt shortly after the collapse of Lehman Brothers in September 2008.

Figure 1 presents average gross central government debt as a percent of GDP for 70 countries aggregated into subgroups consisting of 22 advanced and 48 emerging market economies from 1900 to 2011. The lines show the simple unweighted arithmetic averages presented for the two groups. The average for emerging market economies topped out at a debt/GDP ratio of about 100 percent in the late 1980s and early 1990s. The 22 advanced economies averaged a debt/GDP ratio around 90 percent in the years just after World War II, and in 2010 are just above the 90 percent benchmark. Of course, this benchmark should not be taken as a law of nature, like the boiling point of water at sea level, but it suggests that numerous countries are in the neighborhood of experiencing a public debt overhang.[2]

---

[2] Of course, focusing on gross debt issued by the central government has its shortcomings. For example, it would be desirable to have long-dated measures of general government debt that include states and municipalities. However, for long-dated historical data, the Reinhart–Rogoff (2009) database only contains central government debt. There is also the issue of net debt versus gross debt, with the main difference being government debt held by government-run, old-age support trust funds. This distinction has become much more important recently as the trust funds have massively expanded. Again, net debt data is not available on a long-dated cross-country basis. However, per our arguments in the conclusions,

Figure 2

Gross Total (Public plus Private) External Debt as a Percent of GDP: 22 Advanced and 25 Emerging Market Economies, 1970–2011
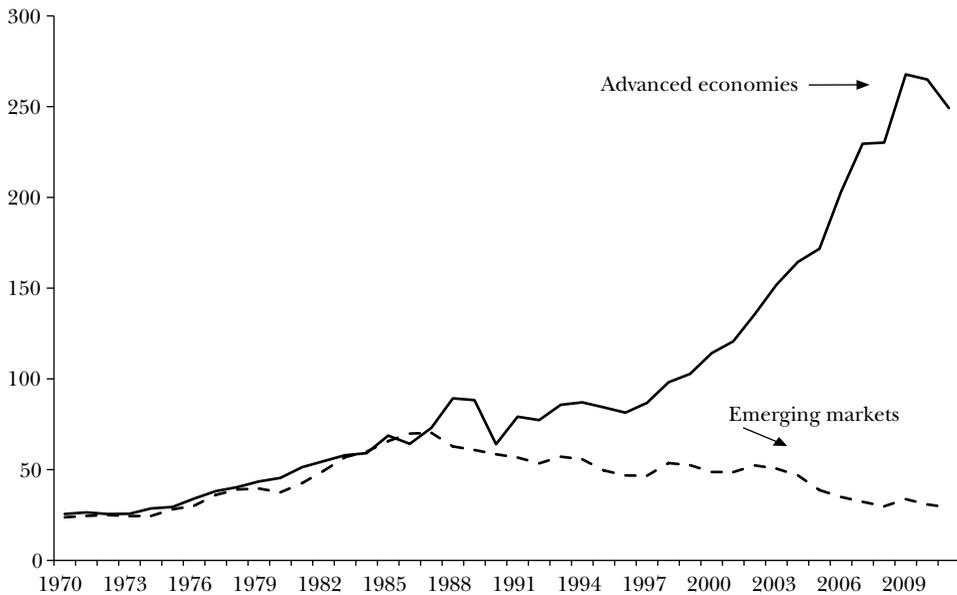
Sources: Lane and Milesi-Ferretti (2010), Reinhart and Rogoff (2009), and sources cited therein; *Quarterly External Debt Statistics*, World Bank, various years; *Global Development Finance*, World Bank, various years.

Figure 2 traces the trajectory of the sum of gross public and private external debt/GDP since 1970 for the same sample of 22 advanced and 48 emerging market economies. The overlap and interaction between different types of debt is particularly acute when it comes to external debt. As Reinhart and Rogoff (2009, 2011) note, the historical record indicates that private external debts are often absorbed by the sovereign during a debt crisis.

Led by European countries, the surge in external debts of the advanced economies since the early 2000s is unprecedented; for example, it dwarfs the late 1970s to early 1980s lending boom to emerging markets.[3] For Europe as a whole, public and private external debts are already more than double the 90 percent threshold and constitute a considerable source of uncertainty.

---

the fact that net public debt today tends to be significantly lower than gross public debt would do little to reverse our conclusions since by and large the trust funds are woefully underfunded, and implicit tax liabilities in most pension systems are hugely positive. In other words, these trust funds are hardly sources of future revenues to offset gross government deficits.

[3] Of course, this recent rise in external debt arises partly because we (and others including the IMF) label debt across euro-zone countries as external. This is clearly a plausible approximation given the weakness of euro-wide institutions, but as euro institutions are still stronger than many international counterparts, it may also be regarded as an exaggeration.

Figure 3
Private Domestic Credit as a Percent of GDP
(22 advanced and 28 emerging market economies, 1950–2011)
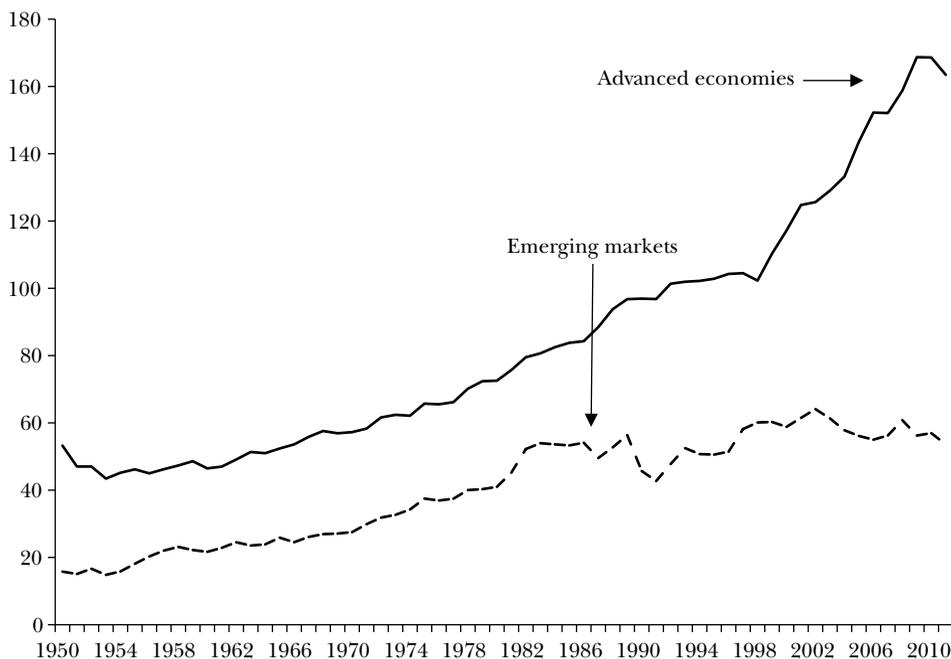
Figure 3 plots private domestic credit—the data is essentially bank loans. Although this measure of private credit is incomplete, particularly for the United States with its highly sophisticated capital market, this measure is most easily compared across time and countries. For the 48 emerging markets, the average unweighted value of this measure of private domestic credit has been roughly constant at around 40 percent of GDP since the 1980s. For the 22 advanced economies, there is a steady rise in credit going back to the 1950s, which through much of this time probably reflects the development and deepening of the financial sector. But there is also a rapid rise in the rate of credit growth starting around 2000. Schularick and Taylor (2012) have calculated the rise in private debt by looking at bank assets, and find a similar pattern of a steady increase from the 1950s up through the 1990s followed by a more rapid expansion after about 2000. This general similarity of the pattern in Figure 2 showing the rise in external debt and the pattern in Figure 3 showing the rise in private debt should not come as a surprise. The literature on domestic credit booms (for example, Mendoza and Terrones 2011) links these booms to capital inflow surges—to borrowing from the rest of the world.

In short, our focus on public debt in this paper should not obscure the reality that many advanced economies are facing quadruple debt overhang problems: public, private, external, and pension. Nor have we paid attention here to the likely possibility of significant "hidden debts," especially in the public sector, which Reinhart and Rogoff (2009) find to be a significant factor in many debt crises, and as documented in detail in the Reinhart (2011) chartbook. Although we focus here on exceptionally high public debt episodes, the topic of multidimensional debt overhang is a critical topic for future research.

Features of Episodes of High Public Debt Since 1800

As noted earlier, we focus on public debt overhang episodes where the gross public debt/GDP ratio exceeds 90 percent for five years or more. We identify 26 public debt overhang episodes in 22 advanced economies from the early 1800s through 2011. This tally does not include the unfolding cases in Belgium, Iceland, Ireland, Portugal, and the United States, where the beginnings of the debt over-hangs date to the financial crisis and recession in 2008 or later, and thus do not meet our five-year minimum criterion. We suspect all these will eventually reach the five-year mark as, indeed, episodes in advanced countries lasting only one to four years appear very infrequently in our data set. Among more recent public debt overhang episodes, our sample does include the cases of Greece, Italy, and Japan, where the beginnings of the debt overhangs (as defined above) date back to 1993, 1988, and 1995, respectively.

The 26 Episodes
Tables 1 and 2 provide information on the 26 episodes that fulfilled the criteria on magnitude and duration of our definition of public debt overhang. Table 2 also provides information on four shorter spells of high debt (with the duration marked with an asterisk) lasting less than five years, that were largely associated with war or cyclical downturn in the Depression of the 1930s.

The first column of Table 1 is categorized by country. As noted, our tabulation covers 22 advanced economies. Of these, nine countries have no episodes that meet our criteria of a public debt overhang: Austria, Denmark, Finland, Germany, Iceland (not until 2009), Norway, Portugal (not until 2010), Sweden, and Switzerland. The fact that many countries do not have any history of public debt/GDP above 90 percent helps explain the finding in Reinhart and Rogoff (2010a) that fewer than 10 percent of the post–World War II annual observations of public debt/GDP for all advanced economies are above the 90 percent cutoff. The remaining 13 countries record one or more debt overhang episodes, as shown in Tables 1 and 2. Table 1 presents the averages for growth and real interest rates across the debt overhang episodes listed individually in Table 2.

The sample coverage (in the second column in Table 1) is determined by data availability and varies by country. The next six columns provide averages for real GDP

Table 1

Features of Public Debt Overhang Episodes: Advanced Economies, 1800–2011

| | | Average real GDP growth | | Average real interest rates | | | | Share of years above 90% |
| | | | | Short-term | | Long-term | | |
Country	Sample	Below 90%	Above 90%	Below 90%	Above 90%	Below 90%	Above 90%	
Australia	1852–2011	4.0	3.5	1.7	−0.4	3.2	1.6	6.1
Belgium	1836–2011	2.5	2.7	2.5	2.4	2.9	3.6	20.5
Canada	1871–2011	3.6	3.2	0.6	2.4	2.3	4.5	10.6
France	1880–2011	3.2	1.9	0.7	2.1	2.1	2.5	28.0
Greece	1848–2011	4.7	3.0	−1.8	4.7	−6.0	12.5	56.1
Ireland	1924–2011	3.4	2.5	−0.6	6.1	2.3	6.5	15.5
Italy	1861–2011	3.9	1.1	0.4	4.1	2.2	4.3	48.0
Japan	1872–2011	4.2	0.8	2.1	0.3	2.7	1.4	12.1
Netherlands	1816–2011	3.3	2.1	2.4	3.1	3.4	4.3	45.6
New Zealand	1861–2011	4.8	3.1	1.9	2.7	2.1	3.0	48.0
Spain	1850–2011	2.9	2.1	2.18	2.52	2.39	9.05	18.6
United Kingdom	1830–2011	2.1	1.8	2.42	2.57	2.74	3.68	45.3
United States	1791–2011	3.6	−1.0	1.75	−4.45	3.72	−2.73	3.2

Memorandum items:
Countries where debt/GDP exceeded 90% for 1 to 4 years (not meeting the debt overhang criteria)

Austria	1880–2011
Finland	1914–2011
Iceland	1908–2011
Portugal	1850–2011

Countries where debt/GDP did not exceed 90% in any year over the sample

Denmark	1880–2011
Germany	1880–2011
Norway	1880–2011
Sweden	1719–2011
Switzerland	1880–2011

Note: For Belgium, real rate averages exclude 1926, when inflation hit an all-time peak of 40 percent and real ex-post interest rates were about −34 percent.

growth, real (inflation adjusted) short-term interest rates, and real long-term interest rates. For each of these three variables, we provide the averages for debt/GDP below and above 90 percent. Details on the interest rate and other data used are provided in a Data Appendix available online with this paper at ⟨http://e-jep.org⟩. The final column of Table 1 provides a calculation of the share of years in the total sample (shown for each country in column 2) where debt/GDP was above 90 percent. For example, since 1848 (when the public debt data is available), Greece leads the way with 56 percent of the debt/GDP ratio observations above 90 percent.

In Table 2, we list each of the 26 episodes meeting our criteria of a public debt overhang (plus the four shorter episodes, marked with an asterisk). The last column provides some commentary on each debt overhang episode. The episodes are grouped along the lines of whether the debt arises primarily from specific wars,

financial crises and economic depression, domestic turmoil, or other factors. Owing to their multidecade span, several episodes incorporate several wars and a multitude of business cycles. In the comment entries, we indicate features such as peak levels of debt and interest rates and whether there were other related events or arrangements in financial markets, such as a debt conversion or financial repression.[4] It is noteworthy that most pre–World War II episodes involved credit events ranging from default on all debt and selective default on some debt (such as World War I debts to the United States) to a variety of conversions (voluntary and otherwise).

As the commentary in the final column of the tables highlights, many debt overhangs result from costly wars. There are distinct clusterings around World War II and, to a lesser extent, World War I, which then merges with the Depression era debt buildup. Back in Figure 1, this sequence of World War I, the Great Depression, and World War II shows up as the three nearly consecutive peaks in the advanced economies' aggregate debt ratios. Greece and Italy are tied for first place in the number of debt overhang episodes: each has four episodes, and the percent of years in the total sample where they had an overhang is 56 and 48 percent, respectively. It is perhaps more surprising that the two previous world powers, the Netherlands and the United Kingdom, have so few debt overhang episodes—just three and two, respectively. However, the few episodes that did happen in these nations lasted for a long time. The Napoleonic wars of the early nineteenth century, in particular, left a deep mark on the finances of both countries. It took a longer time to work down debt ratios in the nineteenth century (Reinhart and Sbrancia 2011). In those days before fiat currency, inflation was not as prevalent as it would later become. Thus, the "liquidation" of government debt via a steady stream of negative real interest rates was not as easily accomplished in the days of the gold standard and relatively free international capital mobility as in the decades after World War II.

In addition, governments in the second half of the twentieth century often used policies of "financial repression" to reduce the cost of the public debt, by limiting capital flows and regulating financial institutions in such a way that alternative investments were blocked and financing for government debt would flow more cheaply. The modern tools of financial repression were not as available to advanced economies in the nineteenth century, but other forms of economic repression were available. In particular, there were substantial transfers from the colonies to finance debts and facilitate debt reduction. During much of the 1800s, the Netherlands, for example, earmarked Indonesian revenues for deficit reduction (Bos 2007).

---

[4] "Financial repression" includes directed lending to the government by captive domestic audiences (such as pension funds or domestic banks), explicit or implicit caps on interest rates, regulation of cross-border capital movements, and a tighter connection between government and banks, either explicitly through public ownership of some of the banks or through heavy "moral suasion." It is often associated with relatively high reserve requirements (or liquidity requirements), securities transaction taxes, prohibition of gold purchases (as in the United States from 1933 to 1974), or the placement of significant amounts of government debt that is nonmarketable. In principle, "macroprudential regulation" need not be the same as financial repression, but in practice, one can often be a prelude to the other.

Table 2

"Types" of Public Debt Overhang Episodes, Advanced Economies, 1800–2011

Country	Debt overhang	Years duration	Comments on factors contributing to debt build
Shorter post-WWI and WWII episodes			
Australia	1945–1950	6	Significantly negative real interest rates (−7%); financial repression; growth is below average.
Belgium	1920–1926	7	Postwar boom and reconstruction; inflation spike and sharply negative real rates.
Belgium	1946–1947	*	Too short to define as a debt overhang.
Canada	1944–1950	6	Debt peaked at 136% in 1946, Real short rates and long rates averaged 0.39 and 2.69%.
Finland	1943–1945	*	Too short to define as a debt overhang.
Italy	1940–1944	5	Default during 1940–1946; inflation peaks at 344% liquidating debts by 1947 debt/GDP is 25%.
United States	1944–1949	6	Federal gross debt peaks at 121% in 1946. Deployment and output decline of 11% in 1946. Era of financial repression worldwide under Bretton Woods agreement; negative real interest rates.
Longer WWI/banking crises 1930s depression/WWII episodes			
France	1920–1945	26	1922 debt is 262%; 1932 WWI debt to the U.S. is in default. 1932 WWI debt to U.S. is in default.
Italy	1917–1936	20	Several debt conversions in 1920s.
Netherlands	1932–1954	25	Strong post WWII recovery; negative real interest rates.
United Kingdom	1917–1964	48	Default on WWI debts to U.S. in 1932. Post WWII debt 248%. Financial repression era; short and long rates −1.1% and 0.5%.
Banking crisis and economic depression			
Australia	1931–1934	*	Too short to define as a debt overhang.
Greece	1928–1939	12	Banking crisis in 1931; default 1932–1964.
Italy	1881–1904	24	Severe banking crisis in early 1890s.
Japan	1995–2012	18, ongoing	1989 equity market crash, severe banking crisis in 1991; large private sector debt "overhang" by any measure since 1980s.

(Continued)

There were also "usury laws" that were the ancestors to the interest-rate ceilings that accompanied financial repression after World War II (Homer and Sylla 1996).

The relatively modern peacetime episodes of public debt overhang in the advanced economies are comprised of Belgium, Canada, Greece, Ireland, Italy, and Japan. Of these six, the shortest were Canada and Ireland, lasting 8 and 11 years, respectively. Japan's mounting public debts had their origins in the systemic banking crisis of 1991 and asset (equity and real estate) collapse that began somewhat earlier. It can be conjectured that Greece, Ireland, and Italy's debt build-ups may have been in part connected to their efforts in joining the euro zone; in effect, these countries had been using high rates of inflation to manage their debt/GDP ratios, but when

Table 2—continued

Longer episodes, other wars, and internal conflicts

France	1880–1905	26	Franco-Prussian War, 1870–1871 legacy of reparations payments to Germany.
Netherlands	1816–1872	57	Napoleonic War debts; 1830s war with Belgium debt rises to 280% followed by several conversions.
Spain	1868–1882	15	1868–1876, Third Carlist Wars. Real bond yields around 25%. Default in 1877–1882.
Spain	1896–1909	14	1879 external public debt peak 52%. Wars and loss of last colonies.
United Kingdom	1830–1863	34	Debt peaks at 260% in 1819–1821 after Napoleonic Wars. (no real GDP data prior to 1830) There are several debt conversions.

Modern peacetime episodes often involving inflation stabilization

Belgium	1982–2005	24	Growth is below average; inflation declines from over 8%.
Canada	1992–1999	7	Real bond rates were as high as 9%; shortest peacetime episode.
Ireland	1983–1993	11	Inflation near 20%. Real rates on the long bond peak at 10% in 1986; real short-term rates averaged 15% during 1992 ERM crisis.
Greece	1993–2012	20, ongoing	Inflation near 15% in 1993; real bond yields about 4% in episode, lower than pre-war; boom followed by banking crisis and restructuring.
Italy	1988–2012, ongoing	25, ongoing	Lower real interest rates than pre-war; lower reliance on external debt.

Other episodes

Austria	1882–1883	*	Too short to define as a debt overhang.
Greece	1848–1883	36	Nation-building. Pre-WWII real long-rates were over 15%.
Greece	1887–1913	27	Defaults in 1843–1878 and 1894–1897.
Netherlands	1886–1898	13	Shrunken revenues from Indonesia added to debt buildup.
New Zealand	1881–1951	71	Severe banking crisis in 1893. Debt peaks at 226% in 1932 amid collapsing commodity prices; debt conversion in 1933.

*Too short to define as a debt overhang.

joining the euro zone required them to hold down their inflation rate, their debts continued to rise. In effect, debt financing supplanted inflation finance.

Public Debt Overhang and Slow Growth

The standard textbook discussion of connections between public debt and economic growth emphasizes two potential channels. The first channel operates through a quantity effect on private sector investment and savings. When public debt is very high, it will tend to soak up the available investment funds and thus to crowd out private investment. If the government at the same time is imposing policies that attempt to reduce its debt burden with higher taxes, a burst of unexpected

inflation, or various types of financial repression, then investment may well be discouraged further. The second channel involves a rising risk premium on the interest rates for government debt. Sufficiently high levels of public debt call into question whether the debt will be repaid in full, and can thus lead to a higher risk premia and its associated higher long-term real interest rates, which in turn has negative implications for investment as well as for consumption of durables and other interest-sensitive sectors, such as housing. Our long-run data on public debt and output does not include sectoral data on investment and savings, so we cannot examine the possible mechanisms underlying public debt and growth. But in this section, we look at some of the evidence connecting a public debt overhang with lower growth rates. In the next section, we consider the link from public debt overhang to real interest rates.

As a starting point, we observe that in the countries that have one or more episodes of public debt overhang listed in Table 1, real GDP growth averages 3.5 percent per annum over the full period for which debt/GDP is less than 90 percent and data is available. The comparable average for all debt overhang episodes is 2.3 percent (or 1.2 percent lower than the lower debt periods). Similarly, Reinhart and Rogoff (2009) show that periods where public debt is over 90 percent of GDP are associated with roughly 1 percent lower growth, while at lower debt thresholds, the correlation of the public debt/GDP ratio with growth is small. Three episodes of public debt overhang, however, are associated with higher GDP growth. One of these, an outright boom, is associated with post–World War I rebuilding in Belgium.

But obvious concerns arise here about cause and effect. Is the public debt overhang causing the slower growth? Or is an exogenous shock that causes slower growth either helping to generate the public debt overhang or else prolonging the escape from that debt overhang? This endogeneity conundrum has not been fully resolved. However, a number of recent studies have tackled the problem. The common finding from a number of approaches is that the relationship between public debt and growth is nonlinear, but at high levels, often at a debt/GDP ratio around 90 percent of GDP, public debt overhang does seem to have a negative effect on growth.

As one approach, Kumar and Woo (2010) look at a panel of 38 advanced and emerging market economies with population over five million from 1970–2007. Using a variety of estimation strategies and subsamples within the context of an endogenous growth model, they find an inverse relationship between initial debt and subsequent growth, after controlling for a number of other determinants of growth. On average, they find that an increase of 10 percentage points in the initial debt/GDP ratio is associated with a slowdown of around 0.2 percentage points per year, with some evidence that this effect is only significant at a debt/GDP ratio above about 90 percent. Along similar lines, Balassone, Francese, and Page (2011) seek to deal with endogeneity in their study of Italy from 1861–2010 by fitting the data to an endogenous growth model and then using a variety of estimation strategies.

Another method of attempting to control for possible feedback from economic growth to public debt is to use five-year averages of growth that are a

function of regressors that are predetermined (and thus not subject to feedback effects). Cecchetti, Mohanty, and Zampolli (2011) take this approach in examining public debt and growth in 18 OECD economies (none in emerging markets) from 1980–2010. They find that government debt begins to reduce economic growth once it crosses a threshold of about 85 percent. Arcand, Berkes, and Panizza (2012) also work with five-year growth averages and find threshold results similar to most other studies for a group of 44 advanced and emerging market economies over the 1976–2005 period.

Yet another approach to address endogeneity problems is to use instrumental variables. For example, in the Checherita and Rother (2010) study of 12 euro area economies from 1970–2008, the authors use the lagged value of debt and average debt in the euro area as instruments. Patillo, Poirson, and Ricci (2011) use robust general method of moments (GMM) estimation as a way of controlling for endogeneity in their study of how external debt (public and private) affects growth in 93 developing countries from 1969–1998. Both approaches find that public debt reduces growth above a certain threshold.

Finally, using a similar estimation strategy as Cechetti, Mohanty, and Zampolli (forthcoming) and Kumar and Woo (2010), Panizza and Prebistero (2012) use the share (in total debt) of foreign currency debt as their key instrument to deal with the potential endogeneity of debt. They conclude that there is little evidence that high public debt levels hurt future growth in advanced economies, but suggest that things are different in developing countries, where a significant fraction of debt is external and the debt overhang argument has more bite. Of course, as the authors discuss, their critical instrument has an important drawback—it has nearly zero mean and variance in France, Germany, Japan, the Netherlands, and the United States, as these five countries do not have public debt denominated in foreign currency.

In investigating the potential transmission mechanisms, Elmeskov and Sutherland (2012) argue that high public debt overhang affects growth through a number of channels including the cost of capital. An Appendix available with this paper at ⟨http://e-jep.org⟩ gives summary information for recent studies, including those mentioned here, that examine the connections between public debt, private debt, external debt, and economic growth.

We would not claim that the cause-and-effect problems involved in determining how public debt overhang affects economic growth have been definitively addressed. But the balance of the existing evidence certainly suggests that public debt above a certain threshold leads to a rate of economic growth that is perhaps 1 percentage point slower per year. In addition, the 26 episodes of public debt overhang in our sample had an average duration of 23 years, so the cumulative effect of annual growth being 1 percentage point slower would be a GDP that is roughly one-fourth lower at the end of the period. This debt-without-drama scenario is reminiscent for us of T.S. Eliot's (1925) lines in "The Hollow Men": "This is the way the world ends/Not with a bang but a whimper." Last but not least, those who are inclined to the belief that slow growth is more likely to be causing high debt, rather

Figure 4

Growth and Real Interest Rate Outcomes for 26 High-Debt Episodes in Advanced Economies, 1800–2011

	Higher REAL INTEREST RATES	Lower REAL INTEREST RATES
Higher GROWTH		Belgium, *1920–1926* Netherlands, *1932–1954*
	Interest rates about the same UK, *1830–1868*	
Lower GROWTH	Belgium, *1982–2005* Canada, *1992–1999* France, *1880–1905* Greece, *1848–1883* Greece, *1887–1913* Greece, *1928–1939* Greece, *1993–2011* Ireland, *1983–1993* Italy, *1881–1904* Italy, *1917–1936* Italy, *1940–1944* Italy, *1988–2011* Netherlands, *1816–1862* Spain, *1868–1882* Spain, *1896–1909*	Australia, *1945–1950* **(–0.1/–7.3)** U.S., *1944–1949* France, *1920–1945* Japan, *1995–2012* *Interest rates about the same* Canada, *1944–1950* Netherlands, *1886–1898* New Zealand, *1881–1951* UK, *1917–1964*

Source: Authors' calculations based on data sources listed in the Data Appendix.

than vice versa, need to better reconcile their beliefs with the apparent nonlinearity of the relationship, in which correlation is relatively low at low levels of debt but rises markedly when debt/GDP ratios exceed the 90 percent threshold. Overall, the general thrust of the evidence is that the cumulative economic losses from a sustained public debt overhang can be extremely large compared with the level of output that would otherwise have occurred, even when these economic losses do not manifest themselves as a financial crisis or a recession.

Will Interest Rates Sound the Alarm?

Higher real interest rates are more common than not during periods of high debt as we see these for 15 of the 26 episodes shown in Figure 4. Figure 4 places the individual episodes into a two-by-two matrix. The rows divide the episodes into 1) those where the average growth during the period of debt overhang is associated with higher average growth than the country's average growth across all the years in which debt/GDP was below 90 percent (upper row) and 2) those episodes where the growth during the debt overhang is lower (bottom row). The columns perform a comparable division for episodes where real interest rates (long bond) were

higher (left column) and those where rates were lower. The middle insets represent the cases where there was little differential in interest rates between the high and lower debt periods. As Figure 4 illustrates, a nontrivial share of the 26 episodes are characterized by both lower growth *and* lower or comparable real interest rates (relative to the period without a debt overhang). This potential outcome is left largely unexplored in textbooks.

Furthermore, there is little to suggest a systematic mapping between the largest increases in average interest rates and the largest (negative) differences in growth during the individual debt overhang episodes. Belgium's post–World War I debt overhang from 1920–1926 is associated with a rebuilding boom in which average annual GDP growth during this period was 6.2 percent—that is, 3.7 percent above the long term-growth average of 2.5 percent (for all years in which debt/GDP is below 90 percent). A rare (for Belgium) postwar inflation spike also produced what turned out to be very negative real interest rates (minus 8 percent). At the other end, average post–World War II GDP growth during the six-year debt overhang (1944–1949) in the United States is sharply lower (there was no need to rebuild entire cities, as in Europe and Japan). More germane to the current situation are the longer peacetime debt overhangs—for example, Belgium, Canada, Greece, Ireland, and Italy in the 1980s and 1990s, and Greece, Italy, the Netherlands, and New Zealand in an earlier era. With the exception of the United Kingdom at the height of its colonial powers in the nineteenth century, these long peacetime debt overhangs are consistently associated with lower growth (in varying degrees), irrespective of whether real interest rates rose, declined, or remained about the same.

The relationship between debt and alternative measures of sovereign external default risk is similarly highly nonlinear as discussed in Reinhart, Savastano, and Rogoff, (2003) as well as Reinhart and Rogoff (2009). Up to a critical level, in the neighborhood of 60 percent of GDP but lower for some countries, market measures of default risk are relatively invariant to total external debt, but they spike at some point when debt rises above that level.

Conclusion

We identified 26 episodes since 1800 of public debt overhang in advanced economies: that is, cases where the ratio of gross public debt to GDP exceeded 90 percent in a given country for more than five years.[5] Taken as a whole, these episodes suggest several lessons about public debt overhang. First, once a public debt overhang has lasted five years, it is likely to last 10 years or much more (unless the debt was caused by a war that ends). The average duration of our debt overhang

---

[5] Reinhart and Rogoff (2010a) point out that a threshold substantially above the 90 percent debt/GDP ratio would leave relatively few observations. For example, on a yearly basis since World War II, just over 1 percent of all gross central government debt-to-GDP ratios among advanced countries have exceeded 120 percent.

episodes was 23 years. Second, it is quite possible to have a "no drama" public debt overhang, which doesn't involve a rise in real interest rates or a financial crisis. Indeed, in 11 of our 26 public debt overhang episodes, real interest rates were on average comparable, or lower, than at other times. Third, the weight of the evidence suggests that a public debt overhang does slow down the annual rate of economic growth, and given the length of these episodes of public debt overhang, losing even 1 percentage point per year from the growth rate will produce a substantial decline in the level of output, and a massive cumulative loss.

The advanced world has entered an era characterized by massive overhang of public and private debt. The average level of gross public debt to GDP in advanced countries as a whole already exceeds our 90 percent threshold. To what extent should we be concerned that the lessons we have just outlined will apply in the next decade or two? Of course, there are always reasons why lessons drawn from a collection of historical episodes may be less or more pertinent to the problems of today.

For example, one possible reason for minimizing concerns about public debt overhang is to argue that financial globalization in the early twenty-first century has made it easier to carry high public debt burdens. However, we see no compelling evidence that this is the case for advanced countries as a whole. Indeed, one might argue that financial globalization has created the possibility of greater volatility and sharper crises in sovereign debt markets. Moreover, one should not underestimate the sophistication and interconnection of national markets in the nineteenth century, which is half the timespan covered in this study. Another line of reasoning for dismissing concerns about public debt overhangs is the view that causality mostly runs from growth to debt. However, we discussed a body of evidence which argues that causality does indeed run from the public debt overhang to slower growth. There are counterexamples where a public debt overhang was accompanied by rapid growth, like the immediate period after World War II for the United States and United Kingdom, but these exceptions to the typical pattern do not seem to be the most relevant parallels for the modern world economy. At the very least, the multidecade-long duration of past public debt overhang episodes suggests that the association between public debt overhang and slower growth is not due to recessions at business cycle frequencies.

Of course, new developments in technology and globalization might conceivably provide such a remarkable reservoir of growth that today's public debt burdens will prove to be quite manageable. Barring such a growth resurgence, the public debt overhang problem that already affects some advanced economies, and has the potential to affect many others including the United States sometime soon, could have consequences at least as large as those seen in the 26 historical episodes that have been our focus here. There are three reasons to worry this could happen. First, public debt is projected over the next decade or two to rise from its already high levels in many advanced economies, as the contingent liabilities now built into old-age programs come to pass. At present, the momentum is for public debt to become substantially worse over time, even when or if more sustained and rapid economic growth resumes. Second, many advanced economies are in fact facing a

quadruple debt overhang of public, private, external, and pension debt. Third, we have not paid attention here to the likely possibility of significant "hidden debts," especially in the public sector, which Reinhart and Rogoff (2009; see also Reinhart 2011) find to be a significant factor in many debt crises.[6]

This paper should not be interpreted as a manifesto for rapid public debt deleveraging exclusively via fiscal austerity in an environment of high unemployment. Our review of historical experience also highlights that, apart from outcomes of full or selective default on public debt, there are other strategies to address public debt overhang including debt restructuring and a plethora of debt conversions (voluntary and otherwise). The pathway to containing and reducing public debt will require a change that is sustained over the middle and the long term. However, the evidence, as we read it, casts doubt on the view that soaring government debt does not matter when markets (and official players, notably central banks) seem willing to absorb it at low interest rates—as is the case for now.

References

Arcand, Jean Louis, Enrico Berkes, and Ugo Panizza. 2012. "Too Much Finance?" IMF Working Paper 12/161, June.

Barro, Robert. 1979. "On the Determination of the Public Debt." *Journal of Political Economy* 87(5): 940–71.

Cecchetti, Stephen, M. S. Mohanty, and Fabrizion Zampolli. 2011. "The Real Effects of Debt." Presented at the "Achieving Maximum Long-Run Growth" symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, 25–27 August 2011.

Checherita, Christina, and Philipp Rother. 2010. "The Impact of High and Growing Government Debt on Economic Growth: An Empirical Investigation for the Euro Area." Working Paper 1237, European Central Bank.

Balassone, Fabrizio, Maura Francese, and Angelo Pace. 2011. "Public Debt and Growth in Italy." Quaderni di Storia Economica Banca D'Italia No. 11, October.

Bos, Frits. 2007. "The Dutch Fiscal Framework: History, Current Practice and the Role of the CPB." CPB Document 150, July.

Elmeskov, Jørgen, and Douglas Sutherland. 2012. "Post-Crisis Debt Overhang: Growth and Implications across Countries." http://www.oecd.org/dataoecd/7/2/49541000.pdf.

---

[6] Hidden debt can involve contingent liabilities of the public sector, payment arrears, or other off-balance-sheet items.

Homer, Sidney, and Richard Sylla. 1996. *A History of Interest Rates,* 3rd edition. New Jersey: Rutgers University Press.

IMF. Various years/issues. *International Financial Statistics.* International Monetary Fund.

IMF. Various years/issues. *World Economic Outlook,* International Monetary Fund.

Kumar, Mohan, and Jaejoon Woo. 2010. "Public Debt and Growth." IMF Working Paper WP/10/174. July.

Lane, Philip, and Gian Maria Milesi-Feretti. 2010. Updated and Extended "External Wealth of Nations" Dataset, 1970–2007. http://www.philiplane.org/EWN.html.

Mendoza, Enrique G., and Marco E. Terrones. 2011. "An Anatomy of Credit Booms and their Demise." Unpublished paper, University of Maryland, November.

Panizza, Ugo, and Andrea Prebistero. 2012. "Public Debt and Economic Growth: Is There a Causal Effect?" Mo. Fi. R. Working Papers 65, Money and Finance Research group, Univ. Politechnica Marche.

Pattillo, Catherine, Hélène Poirson, and Luca Antonio Ricci. 2011. "External Debt and Growth." *Review of Economics and Institutions* 2(3): Article 2.

Reinhart, Carmen M. 2011. "Chartbook of Country Histories of Debt, Default, and Financial Crises." Chap. 2 in *A Decade of Debt,* Policy Analyses in International Economics 95, by Carmen M. Reinhart and Kenneth S. Rogoff. Washington DC: Peterson Institute for International Economics.

Reinhart, Carmen M., and Vincent R. Reinhart. 2010. "After the Fall." In Federal Reserve Bank of Kansas City Economic Policy Symposium "Macroeconomic Challenges: The Decade Ahead" at Jackson Hole, Wyoming, on August 26–28, 2010. Available at: http://www.kc.frb.org/publications/research/escp/escp-2010.cfm.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly.* Princeton University Press.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010a. "Growth in a Time of Debt" *American Economic Review* 100(2): 573–78.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2010b. "Debt and Growth Revisited." Vox EU, August 11.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2011. "From Financial Crash to Debt Crisis." *American Economic Review* 101(5): 1676–1706.

Reinhart, Carmen M., Miguel A. Savastano, and Kenneth S. Rogoff. 2003. "Debt Intolerance." *Brookings Papers on Economic Activity,* no. 1, pp. 1–74.

Reinhart, Carmen M., and M. Belen Sbrancia. 2011. "The Liquidation of Government Debt." NBER Working Paper 16893, March.

Schularick, Moritz, and Alan Taylor. 2012. "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review* 102(2): 1029–61.

World Bank. Various years. *Global Development Finance.*

World Bank. Various years. *Quarterly External Debt Statistics.*

The Economics of Spam

Justin M. Rao and David H. Reiley

The term "spam," as applied to unsolicited commercial email and related undesirable online communication, is derived from a popular Monty Python sketch set in a cafe that includes the canned meat product SPAM in almost every dish. As the waitress describes the menu with increasing usage of the word "spam," a group of Vikings in the cafe start singing, "Spam, spam, spam, spam, spam," drowning out all other communication with their irrelevant, repetitive song. The analogy to unsolicited commercial solicitations jamming one's inbox seems apt. Every day about 100 billion emails are sent to valid email addresses around the world; in 2010 an estimated 88 percent of this worldwide email traffic was spam (Symantec 2010; MAAWG 2011). Almost all of this spam is illegal under current laws.

How does spam differ from legitimate advertising? If you enjoy watching network television, using a social networking site, or checking stock quotes online, you know that you will be subjected to advertisements, many of which you may find irrelevant or even annoying. Google, Yahoo!, Microsoft, Facebook, and others provide valuable consumer services, such as search, news, and email, supported entirely by advertising revenue. While people may resent advertising, most consumers accept that advertising is a price they pay for access to content and services that they value. By contrast, unsolicited commercial email imposes a negative externality on consumers without any market-mediated benefit, and without the opportunity to opt out.

This negative externality makes spam particularly useful for teaching purposes. When asked for an example of an externality, most economists think of environmental

■ *Justin M. Rao is a Research Scientist, Microsoft Research, New York, New York. David H. Reiley is a Research Scientist, Google, Inc., Mountain View, California. When this paper was written, both authors were working at Yahoo! Research, Santa Clara, California.*

pollution: groundwater toxins, acid rain, air pollution, global warming, and so on. Indeed, given the great linguistic generality of the term "pollution" (including noise pollution, light pollution, and others), it can be difficult for economists to find an example of a negative externality that *cannot* be described as a form of pollution. Our two favorite nonpollution externalities for teaching are traffic congestion and spam.

Of course, a similar externality has been present for decades in other forms of unsolicited advertising, including junk mail, telemarketing, and billboards. These intrusive activities also impose claims on consumer attention without offering compensation or choice. However, email spam is breathtakingly larger in magnitude, with quantities in the absence of automated spam filters equal to hundreds of emails per user per day—if our email inboxes stood unguarded, they would quickly become totally useless. (In contrast, junk mail has not yet reduced our unguarded postal mailboxes to this fate.) One can purchase unsolicited email delivery on the black market for a price at least a thousand times less than that to send bulk postal mail. Spam has become such a widespread phenomenon that trademark holder Hormel finally stopped objecting to the use of the term to refer to unsolicited email (Templeton, undated).

Spam also seems to be an extreme externality in the sense that the ratio of external costs to private benefits is quite high. We estimate that American firms and consumers experience costs of almost $20 billion annually due to spam. Our figure is more conservative than the $50 billion figure often cited by other authors, and we also note that the figure would be much higher if it were not for private investment in anti-spam technology by firms, which we detail further on. On the private-benefit side, based on the work of crafty computer scientists who have infiltrated and monitored spammers' activity (Stone-Gross, Holz, Stringhini, and Vigna 2011; Kanich et al. 2008; Kanich et al. 2011; Caballero, Grier, Kreibich, and Paxson 2011), we estimate that spammers and spam-advertised merchants collect gross worldwide revenues on the order of $200 million per year. Thus, the "externality ratio" of external costs to internal benefits for spam is around 100:1.

In this paper, we start by describing the history of the market for spam, highlighting the strategic cat-and-mouse game between spammers and email providers. We discuss how the market structure for spamming has evolved from a diffuse network of independent spammers running their own online stores to a highly specialized industry featuring a well-organized network of merchants, spam distributors (botnets), and spammers (or "advertisers"). Indeed, email service provision has become more concentrated in part because the high fixed costs and economies of scale of filtering spam offer a significant advantage to large service providers. We then put the spam market's externality ratio of 100 into context by comparing it to other activities with negative externalities, such as pollution associated with driving an automobile, for which we estimate a ratio of about 0.1, and for nonviolent property crime such as automobile theft, for which we estimate a ratio of 7–30. Lastly, we evaluate various policy proposals designed to solve the spam problem, cautioning that these proposals may err in assuming away the spammers' ability to adapt.

The History of Spam: Cat-and-Mouse Games

Email is sent via a "sender push" technology called Simple Mail Transfer Protocol (SMTP). Other examples of sender-push transfer include postal mail, text messaging, and voice mail. (In contrast, the Hypertext Transfer Protocol (HTTP) used in web browsing is "receiver pull"—nothing shows up in your web browser until you specify a URL.) SMTP was designed in the early 1980s when the trust level across what was then called the "Arpanet" was quite high. Accordingly, senders were not required to authenticate their emails. SMTP servers all over the world were programmed to cooperate in relaying messages. In many respects, SMTP replicates the "transfer protocol" of the U.S. Postal Service. Anyone in the United States can anonymously drop a letter in a mailbox and, provided it has proper postage, have it delivered, without any requirement for the sender to provide an authentic return address.

Spammers first developed technology to automate the sending of bulk email in the mid 1990s by opportunistically tapping into mail relay servers and anonymously floating a deluge of spam from phony domains (Goodman, Cormack, and Heckerman 2007). In 1994, the attorneys Canter and Siegel hired a programmer to automate a posting to every USENET newsgroup in existence, so that thousands of discussion groups devoted to every topic from *Star Trek* to board games were inundated with advertisements for services to help immigrants apply for the green-card lottery. This software soon evolved into the first automated bulk emailer (Zdziarski 2005, pp. 10–13). In 1995, the first commercial "spamware," aptly titled Floodgate, appeared for sale at a price of $100. Floodgate advertised its ability to harvest email addresses from a variety of sources including newsgroups, CompuServe classified ads, AOL Member Directory, and other sources. Then, via the included companion software Goldrush, it promised an ability to send out thousands of emails per hour (Zdziarski 2005, p. 16; Everett-Church 1999). Such software, crude by today's standards, enabled spammers to send email at a cost on the order of $0.0001 per message. Since then, the spam market has been shaped by the technological cat-and-mouse game between spammers and email service providers.

Anti-spam Filtering Techniques

As an early response to spam, Internet administrators developed authentication protocols: where previously one only had to type a password to collect one's incoming mail, now most had to authenticate themselves by providing a password to send outgoing mail. To prevent domain spoofing—using the domain of a well-known company to make an email seem more legitimate—domain authentication routines check that the IP address listed in the Domain Name System matches the sending IP. However, many SMTP servers remained unauthenticated for a long time, and the default mail delivery protocol is still to deliver email from any sending IP address.

After authentication, the arsenal of filtering technologies consists of machine learning, crowdsourcing, and IP blacklisting. Such screening devices detect suspected spam messages and either reject them from being delivered, or send them to a junk-mail folder.

The machine-learning approach dates from the late 1990s (Sahami, Dumais, Heckerman, and Horvitz 1998; Androutsopoulos, Koutsias, Chandrinos, Paliouras, and Spyropoulos 2000). A typical machine-learning implementation uses "ground truth" data on a subset of observations to learn rules to classify the remaining data. With spam, the ground truth is given by human-labeled examples of spam versus non-spam emails and the algorithm is trained to recognize features of the email that predict whether it is spam. For example, one can have the classifier examine the predictive power of all words and word pairs in the subject lines of the emails, which might lead to dummy variables for the presence of "Viagra," "Nigeria," and "Free Money" being included as key predictors. Other examples of heavily weighted features include unusual punctuation common to spam, such as "!!!", and nouns associated with spam-advertised products, such as "Rolex". Every URL contained in a message could also be treated as a possible predictor as spam emails nearly always include the URL of the website to place orders for the advertised product. Any machine-learned filter can have false positives—that is, legitimate mail that is filtered to the junk folder (for instance, spam filters might make it hard to converse legitimately about bank transfers in Africa).

Spammers responded with creative misspellings designed to avoid the filters, such as "V1agra," created many unique URLs all mapping to the same order form, and included attachments of graphical images of text messages, which became popular with spammers when they realized that text-based classifiers could not find the text in the form of a GIF or JPEG image. Spammers also include irrelevant text passages, such as excerpts of news stories that are common in legitimate conversations, or create random permutations of words from one email to the next to throw classifiers off track. Of course, over time the anti-spam classifiers continued to improve and adapt, too. Goodman, Cormack, and Heckerman (2007) present a nice introduction to such anti-spam technology for nonexperts.

Crowdsourcing represents a way to collect additional data to improve the predictive power of machine-learning models. Large webmail providers such as Yahoo! Mail collect data when users press the "mark as spam" button to move email from the inbox to the junk mail folder. Data from such marked spam can be used, as soon as that same day, to retrain the spam classifier. However, most users just delete irrelevant emails rather than marking them as spam. We took a random sample of six months of mail activity for 1.3 million active Yahoo! Mail users and found that only 6 percent of users ever marked any email as spam, but the vast majority deleted messages without reading them.

Spammers have developed a strategic response to the spam voting system. In addition to the "spam" button in the inbox, webmail services also provide a "not spam" button to mark false-positive messages in the junk mail folder. In four months of 2009 Yahoo! Mail data, our Yahoo! colleagues found that (suspiciously) 63 percent of all "not spam" votes were cast by users who never cast a single "spam" vote. After examining additional data on these accounts, such as IP address, position in the network of users, and repeatedly casting not-spam votes on a variety of emails that were receiving multiple "spam" votes from legitimate users, the authors concluded

that the vast majority of these accounts were created by spammers to cast strategic votes in order to help their campaigns beat the spam filters (Cook, Hartnett, Manderson, and Scanlan 2006; Ramachandran, Dasgupta, Feamster, and Weinberger 2011). The researchers discovered 1.1 million of these sleeper accounts, and Yahoo! inserted a detection algorithm to mitigate the effects of this strategic voting.

The single most effective weapon in the spam-blocking arsenal turns out to be blacklisting an email server (Cook, Hartnett, Manderson, and Scanlan 2006; Ramachandran, Feamster, and Vempala 2007). In 2011, 80 percent of all emails received by Yahoo! Mail were rejected by their servers through IP blacklisting. Fortunately, just as the postmark from the sending post office limits the ability to spoof one's return address, Transmission Control Protocol (TCP) makes it impossible to spoof the IP address of the mail server from which the message was sent. Therefore, if email administrators noticed that their users were receiving tremendous amounts of mail from one server, they could "blacklist" such a server. Sharing blacklist information enables multiple organizations to shut down spam activity more quickly. For example, the Spamhaus Block List, founded in 1998 by Steve Linford, now protects nearly 1.8 billion email inboxes from spam (⟨http://www.spamhaus.org /organization/index.lasso⟩, accessed February 9, 2012).

An unintended side effect of blacklisting occurs when a single user starts sending spam and causes their email server to be blacklisted. At that point, everyone else using the same email server will suddenly find their outbound emails being blocked. This situation could arise within any large organization, like a college, a corporation, or a shared Internet service provider (ISP). Of course, information technology professionals can then sort out the problem, and organizations such as Spamhaus strive to act quickly in unblocking any email server who was falsely accused or who corrects the problem with its users, but blacklists still routinely cause reliability problems for users trying to send email.

The larger email services such as Yahoo! Mail, Microsoft Hotmail, and Google Gmail have large, dedicated anti-spam and customer support teams. The high fixed costs of anti-spam technologies and benefits of crowdsourced data have made it difficult for small email providers to compete, which has contributed to significant increases in concentration in email provision since the mid 1990s. We obtained market share data (from comScope) for the top 50 largest consumer web-based email services (including home Internet service providers) for the period 2006–2012. The data show that webmail provision has become increasingly concentrated in the "Big Three" of Hotmail, Yahoo! Mail, and Gmail. The three-firm concentration ratio in this market has increased from 55 percent to nearly 85 percent over the last six years; we believe that spam is a significant contributor to this increase in concentration.

Botnets

Blacklists gradually made it impossible for spammers to use their own servers (or others' open relay servers) with fixed IP addresses. Spammers responded with a "Whack-a-Mole" strategy, popping up with a new computer IP address every time the old one got shut down. This strategy was observed and named as early as 1996, and

eventually became considerably cheaper with another major innovation in spam: the botnet.

A botnet is a network of "zombie" computers infected by a piece of malicious software (or "malware") designed to enslave them to a master computer. The malware gets installed in a variety of ways, such as when a user clicks on an ad promising "free ringtones." The infected computers are organized in a militaristic hierarchy, where early zombies try to infect additional downstream computers and become middle managers who transmit commands from the central "command and control" servers down to the frontline computers (John, Moshchuk, Gribble, and Krishnamurthy 2009; Caballero, Poosankam, Kreibich, and Song 2009; Cho, Caballero, Grier, Paxson, and Song 2010).

The first spamming botnets appeared in 2003. Static blacklists are powerless against botnets. In a botnet, spam emails originate from tens of thousands of IP addresses that are constantly changing because most individual consumers have their IP addresses dynamically allocated by Dynamic Host Control Protocol (DHCP). Dynamic blacklisting approaches have since been developed; Stone-Gross, Holz, Stringhini, and Vigna (2011) document that 90 percent of zombie computers are blacklisted before the end of each day. However, if the cable company assigns a zombie computer a new IP address each day, that computer gets a fresh start and can once again successfully send out spam.

In response to botnets, many Internet service providers, such as Comcast, began to prevent their users' computers from operating as send-mail servers. This meant that individuals and small businesses could no longer run their own mail servers, as in the original, decentralized vision of the Internet, and now had to rely on larger commercial email vendors.

A second generation of botnets makes use of accounts at large commercial email providers. For example, a zombie could be programmed to sign up for hundreds of thousands of free email accounts at Gmail, and then send spam email through these accounts. Email providers have implemented sending thresholds designed to detect and prevent this sort of spamming. If a user exceeds these limits, the system may refuse to send out the email, or it may ask the user to solve a CAPTCHA (as discussed in the next subsection). Such rules cut down on outbound spam, but also impose negative side effects on users who happen to be high-volume senders of legitimate email. In 2011, Yahoo! Mail experienced an average of 2.5 million sign-ups for new accounts each day. The anti-spam team deactivated 25 percent of these immediately, because of clearly suspicious patterns in account creation (such as sequentially signing up account names JohnExample1, JohnExample2, . . .) and deactivated another 25 percent of these accounts within a week of activation due to suspicious outbound email activity.

In 2009, six botnets accounted for over 90 percent of botnet spam (Symantec 2010; John, Moshchuk, Gribble, and Krishnamurthy 2009). The largest botnet on record, known as Rustock, infected over a million computers and had the capacity to send 30 billion spam emails per day before it was taken down in March 2011. Microsoft, Pfizer, FireEye network security, and security experts at the University

of Washington collaborated to reverse engineer the Rustock software to determine the location of the command servers. They then obtained orders from federal courts in the United States and the Netherlands allowing them to seize Rustock's command-and-control computers in a number of different geographic locations. (Microsoft financially supported the operation presumably because Rustock sent its emails through Windows Live Hotmail accounts, while Pfizer participated because a Rustock spam often advertised counterfeit versions of Pfizer's patent-protected Viagra.) If the servers had been located in less-friendly countries, it is not clear whether the takedown could have been successful. The takedown of this single botnet coincided with a one-third reduction in global email spam—and hence a one-quarter reduction in global email traffic (Thonnard and Dacier 2011; Microsoft 2011). Thus, the efforts of these private firms produced a remarkably large *positive* externality.

CAPTCHA: Screening Humans from Bots

To avoid spammers setting up many commercial email accounts, services like Yahoo! Mail have implemented a screening device called a CAPTCHA, which is an acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart." This test will be familiar to most readers as a set of twisty, distorted text characters. Spammers turned to visual-recognition software to break CAPTCHAs, and in response email providers have created progressively more difficult CAPTCHAS, to the point where many legitimate human users struggle to solve them.

However, the big breakthrough in CAPTCHA breaking arose when spammers figured out how to employ human labor to break CAPTCHAs for them. In this idea's first incarnation, a spammer would set up a pornography site, offering to display a free photo to any user who could successfully type in the text characters in a CAPTCHA image. In the background, their software had applied for a mail account at a site like Hotmail, received a CAPTCHA image, and relayed it to the porn site; they would obtain text from a user interested in free porn and relay this back to the Hotmail site (Kotadia 2004).

More formal labor markets subsequently developed for CAPTCHA breaking (Motoyama, Levchenko, Kanich, McCoy, Volker, and Savage 2010). A market maker typically operates one website for interacting with buyers of CAPTCHA-breaking services, and another for interacting with workers who sell their labor. For example, one can purchase CAPTCHA-breaking services from the DeCaptcher website, which transmits each CAPTCHA to a worker at the PixProfit website for breaking, then back to the customer at DeCaptcher. The customer may use a separate piece of software (such as GYCAutomator, which specializes in Gmail, Yahoo! Mail, and Craigslist CAPTCHAs) to transmit the CAPTCHA and its solution. The entire process takes less than 30 seconds. The market wage advertised for CAPTCHA-breaking laborers declined from nearly $10 per thousand CAPTCHAs in 2007 to $1 per thousand in 2009. These labor markets started with Eastern European labor and then moved to locations with lower wages: India, China, and Southeast Asia.

In February 2012, Kotalibablo.com was advertising to workers that they could earn wages starting at $0.35 per thousand. The same company operates the buyer-facing website Antigate.com, which at that time advertised a price of $0.70 per thousand to customers wanting to break CAPTCHAs. Motoyama, Levchenko, Kanich, McCoy, Voelker, and Savage (2010) measured typical response times of around 10–15 seconds per CAPTCHA, with accuracy rates around 90 percent. (During one peak-load period, they experimentally measured a labor supply elasticity of approximately one: increasing their bid amount from $2 per thousand to $5 per thousand increased quantity solved from 8 to 18 per second.) Several websites can provide more than ten CAPTCHAs per second, putting total industry capacity (at a price of $1 per thousand) at over a million broken CAPTCHAs per day. These services market themselves as "Image to Text" providers and operate in the light of day—as of 2012 U.S. law, there does not appear to be anything illegal about the services they offer.

CAPTCHAs are also used to authenticate senders in what are known as "challenge-response systems." Such a service will intercept messages from anyone not in a preset contact list, sending an autoreply before allowing the message to be delivered. The autoreply requires that the sender solve a CAPTCHA, thus authenticating the sender as human. Such systems have been available for at least seven years, but the market has for the most part rejected this technology, and with good reasons (Isacenkova and Balzarotti 2011). First, the autoreply "challenge" itself often gets caught in a spam filter because it contains stock text and a link, and is sent frequently from the same sender—which are all strong signals in machine-learned spam filters. Second, it requires that receivers maintain a continually updated contact list. Third, spammers can use the challenge-response system to spoof messages from unsuspecting "senders," who receive the spammers' message as "backscatter spam" when they fail the challenge and get bounced to the apparent sender.

Hijacking Accounts from Legitimate Users

Another recent strategy of botnets has been to hijack existing email accounts from legitimate users. (These same techniques can be used for even more nefarious purposes, such as hijacking a bank account; for more details about this form of online crime, we refer readers to Moore, Clayton, and Anderson 2009, in this journal.) For example, "phishing" occurs when the culprit sends an email posing as a legitimate institution—say, "Hotmail user account services," often including the actual logo of the institution being spoofed—and asks the victim to visit a website to "verify your account password." In the practice of "keylogging," a type of malware records keystrokes and transmits information (especially suspected passwords) to the spammer. The practice of "packet sniffing" takes advantage of small companies and colleges who still transmit user passwords over the Internet in unencrypted text, and so a spammer "listening" at a login page can not only hijack that account, but also any other accounts (such as Yahoo! Mail) for which the user has conveniently chosen the exact same password. This technique was recently used to obtain access to 93,000 accounts on the Sony Playstation Network (Gross 2011).

In 2005, an industry consortium established a technology standard called Domain Keys Identified Mail (DKIM) as a new weapon in the war against both spamming and phishing. Now adopted by a number of firms, including Yahoo! Mail, Gmail, PayPal, and eBay, this standard creates a digital signature that email senders can adopt. For example, if a phisher pretends to be PayPal asking a user to verify their account password, Yahoo! Mail will immediately notice that the message does not have the correct digital signature (based on public-key encryption) and will therefore reject the forged email without delivering it. Unfortunately, spammers have already responded to this strategy by trying to hijack the account of a corporate user that has been "whitelisted" via DKIM. In March 2011, a number of accounts became compromised at Epsilon, an email service provider who handles the sending of legitimate bulk email for a number of corporate clients, such as TiVo, Capital One, U.S. Bank, and the Kroger grocery chain (Moyer 2011).

On the whole, anti-spam efforts at large companies have mitigated the nuisance of spam to customers. However, the cat-and-mouse moves seem certain to continue.

Spammers and the Field of Dreams

From a spammer's perspective, any online platform delivering eyeballs is a natural target. In other words, as Kevin Costner's character in *Field of Dreams* famously heard, "If you build it, they will come."

Spam is prevalent on social bookmarking sites[1] (Krause, Schmitz, Hotho, and Stumme 2008) and online classifieds (Tran, Hornbeck, Ha-Thuc, Cremer, and Srinivasan 2011). On Twitter, spam takes the form of inserting a spammy link to an ongoing conversation between users (Yardi, Romero, Schoenebeck, and boyd 2009), using Twitter's hashtag feature. Twitter spam also occurs when an ostensible fan of a celebrity writes a message including the characters "@LadyGaga," in hopes of getting it exposed to her fans. Facebook suffers relatively less from spam because of the way it requires users to verify connections with each other, but spammers continue to invent new techniques, from malicious apps to friend requests from fictitious identities, that keep Facebook's anti-spam team quite busy (Warren 2011; Cohen 2012; Ghiossi 2010). Text-messaging spam has become a serious problem in certain countries: one source estimated that 30 percent of text messages in China are now spam. However, in the United States the relatively high price of SMS messaging (often $0.10 per message, orders of magnitude higher than in China) has kept text message spam rates below 1 percent (Gómez Hidalgo, Bringas, Sánz, and García 2006). Text spam is aggressively filtered by cell phone providers, especially for text messages from a computer to a phone through a webmail client (Almeida, Gómez, and Yamakami 2011). Providers of online instant message software also struggle to block spam.

Next to email spam, the most prominent form of spam is known as "web spam" or "black-hat search-engine optimization." A typical web-spam implementation mines

---

[1] Social bookmarking, also known as "tagging," is a way to share webpages with a community of users.

news feeds for headlines and automatically creates pages with snippets of popular stories. The article snippet is used under a "fair use" exception to copyright law, and the remainder of the page is typically saturated with advertisements. Such web spam can deceive search engines into featuring these ad-laden pages prominently in search results about popular topics, thereby annoying users, but it is not illegal. It differs fundamentally from all the other forms of spam discussed in this paper in that it is not sender-push: One only sees a web spam page if one voluntarily clicks. Web spam has been combated through machine learning about the credibility of potential links and the downgrading of low-credibility links in search results (Caverlee and Liu 2007; Zhou, Burges, and Tao 2007; see also Ntoulas, Najork, Manasse, and Fetterly 2006, and Castillo, Donato, Gionis, Murdock, and Silvestri 2007).

Market Structure

Most spam is illegal under the United States CAN-SPAM Act of 2003, which requires unsolicited emails to have valid return addresses and opt-out provisions. While many people use "spam" to refer to the (sometimes annoyingly frequent) messages they receive from businesses with which they have previously transacted, for the purposes of this paper we define spam to be messages from economic agents who do not have a previous relationship with the customer and who do not offer opt-out provisions.

The spam market does have some similarities to the market for legitimate online advertising (whose institutions have been described in this journal by Evans 2009) in the sense that spam attempts to generate a sale. However, while in legitimate advertising the whole point is to promote awareness (of a firm or a product), spam typically uses obfuscation to get its message through. Spam-based advertising is dominated by "affiliate marketing," in which a merchant recruits intermediaries known as affiliates (a.k.a. spammers) to advertise on its behalf, in return for a share of the final purchase amount (Levchenko et al. 2011; Samosseiko 2009; Kanich et al. 2011; Kanich et al. 2008). Thus, a merchant advertising via spam generally shrouds its identity, hiding behind an array of cookie-cutter storefronts, in order to increase the chances of getting its offer through to users.

The supply (or "publishing") side of the spam market has become dominated by botnets, as discussed earlier. Several teams of computer scientists have demonstrated that botnets are distinct economic entities from the merchants on the demand side of the spam market (John, Moschuk, Gribble, and Krishnamurthy 2009; Kanich et al. 2011; Stone-Gross, Holz, Stringhini, and Vigna 2011). Major merchants are advertised by multiple botnets, and botnets compete with each other for clients (Thonnard and Dacier 2011). A botnet may either rent out its services to independent spammers, or send its own spam while acting as an affiliate for a merchant. Both business models appear to be widely practiced (John, Moschuk, Gribble, and Krishnamurthy 2009; Kanich et al. 2011; Stone-Gross, Holz, Stringhini, and Vigna 2011). The market structure appears to be an oligopoly (Zhao et al. 2009). The

Table 1

Breakdown of the Spam Supply Chain

Stage	Pharmacy	Software	Replicas	Total
Unique URLs	346,993,046	3,071,828	15,330,404	365,395,278
Domains	54,220	7,252	7,530	69,002
Store-front styles	968	51	20	1,039
Merchants	30	5	10	45

Source: From a study of 45 merchants tracked by Levchenko et al. (2011).

Stone-Gross team infiltrated the Cutwail botnet and documented its offerings, which range from a bare bones rental of computation time on the compromised machines all the way to a user-friendly interface allowing a customer to create a mass mailing and test it against open-source spam filters before sending. Like publishers in the legitimate advertising market, botnets invest in significant fixed costs of ad serving, match advertisers with potential customers, and offer large reach.

To probe the demand side of the spam market, Levchenko et al. (2011), a team of 14 coauthors based at the University of California at San Diego and the University of California at Berkeley, developed spam feeds to identify examples of spam, a web crawler to follow advertised URLs, and botnet infiltration and botnet detection algorithms (see also John, Moschuk, Gribble, and Krishnamurthy 2009) to monitor botnet activity. Table 1 presents statistics on the merchants tracked through this technique. The first row shows that spam for only 45 merchants included 365 million distinct URLs during the data collection period. The second row of the table shows that there are more than 5,000 times as many URLs as domain names used by spammers. For example, a spammer might register the domain pharma.com and then host thousands of identical pages with different URLs on the same domain: "pharma.com/buy123.html," "pharma.com/purchase01.html," and so on. There are also more than 1,000 domain names per merchant. A merchant may be represented by several affiliate spammers, each of whom might register multiple domains. Large, reputable registrars generally reject applications for spammy-sounding domain names, such as those containing "med" or "pharm" (Kanich et al. 2008), but hundreds of registrars are willing to look the other way (Levchenko et al. 2011). Row three gives the number of "store-front styles" that represent individual user interfaces, each with a distinct look and feel. When law enforcement tries to shut down illegal sales, they look for identical storefronts and try to take them down all at once, so store-front variation helps a merchant avoid complete shutdown. For each pharmaceutical merchant, there are approximately 30 distinct store fronts; this figure is much lower for software and replicas. The final row of the table shows the number of merchants anchoring the market. Despite the large numbers of domains, URLs, and store fronts, only 100 merchants had a measurable market share of spam activity, and fewer than ten merchants account for over 80 percent of the market (Levchenko et al. 2011; Kanich et al. 2011).

After tracking the merchants via the botnets, Levchenko et al. (2011) placed 120 orders for the advertised goods, spread across the 100 identified merchants. The affiliated spammer usually hosts the entire consumer storefront experience; that is, the spammer generally collects payment information and then hands the transaction to the merchant before credit card authorization. Payment processing services for these merchants are quite concentrated: a total of only 17 banks serve the 100 merchants, with just three banks (from Latvia, Azerbaijan, and St. Kitts and Nevis) processing the payments for more than 75 percent of the transactions. Postage stamps on the packages revealed the physical locations where the goods originated: nearly all the pharmaceuticals came from India, for example, while replica watches generally came from China.

Overall, while spammers have nearly free entry in registering domains and renting services from botnets, merchants appear to face more significant fixed costs, especially in obtaining payment processing services. Only a small number of banks appear willing to take the risk of associating with gray-market merchants. This may explain why a relatively small number of merchants supply most of the market for these spam-advertised goods.

Assessing the Externality

What are the costs of spam to users, and how does it compare with the return to spammers? A widely cited report from Ferris Research (2005) placed the world-wide cost of spam in 2005 at $50 billion; Ferris raised its estimate to $100 billion in 2007 and $130 billion in 2009 (Jennings 2009). However, the Ferris reports did not describe how they estimated such key parameters as the amount of time per worker spent deleting spam; indeed, one of the authors of that report indicated to us that their work was "not a scientific survey," but that it attempted to be a lower-bound estimate. Regarding the returns to spammers, the most common estimate of profits involves the phrase "millions of dollars a day," which in turn apparently originated in a widely cited IBM press release.[2] In this next section, we find these widely cited estimates of user costs and spammer profits are somewhat exaggerated, but of the right order of magnitude.

Measuring the Diffuse Costs of Spam

The negative externalities imposed by spam include wasted time for consumers: both wading through irrelevant advertisements in one's inbox and missing an important message that went to the junk mail folder. They also include the costs

---

[2] See Malik (2008), "IBM Says Storm Worm Creators Making Millions Daily," ⟨http://gizmodo.com /354741/ibm-says-storm-worm-creators-making-millions-daily⟩. Phishing for account information in order to steal money is a form of online crime representing less than 0.3 percent of all email traffic. Researchers at Microsoft found that conventional wisdom was an overestimate by 50 of the true profits to phishing (Herley and Florêncio 2009).

of server hardware, which requires more than five times as much capacity as would be required in the absence of spam, as well as the costs of spam prevention services provided by firms to reduce the burden on end users.

The chief challenge in totaling up the social cost is credibly estimating the number of hours lost by people dealing with spam. Estimating the amount of spam that beats spam filters is difficult—after all, if we knew it was spam, we would have filtered it. We choose to examine success rates of spam in influencing consumer behavior, and use these to infer how many spam messages must have gotten through. Here we rely on the work of Kanich et al. (2008), who observed that out of 347 million attempted mailings for an online pharmacy, about 83 million were accepted for delivery rather than bounced; our question is how many arrived in the inbox versus the spam box. The 83 million messages accepted for delivery resulted in 10,500 clicks by consumers; we can estimate the number of spam messages reaching the inbox with an educated guess about the clickthrough rate for spam campaigns. We know legitimate email marketing for medical products has a clickthrough rate of about 1.1 percent (Email Marketing Metrics Report, 2011), while untargeted display advertising on Yahoo! usually has clickthrough rates of 0.1 percent or less. The clickthrough rate for spam email should be lower than the former but higher than the latter, because spam targets consumers more indiscriminately than legitimate email marketing, while email that reaches the inbox attracts more attention than the average web graphical ad. Using a clickthrough rate of 0.25 percent for spam, we estimate that about 4,200,000 messages (10,500 clicks divided by 0.0025 clicks/ message) reached inboxes, out of 347 million messages sent. That is, we estimate that only about 1.2 percent of sent spam messages actually reach user inboxes.

As a consistency check on this estimate, we look at spammers' costs and revenues. Given the free entry of spammers (as opposed to botnets or merchants), we should expect them to earn zero profits. Stone-Gross, Holz, Stringhini, and Vigna (2011) estimate that spammers pay around $30 per million unblocked message deliveries (or five million emails sent, 80 percent of which were blocked by blacklisting). Spammers appear to earn about $50 per purchase (Kanich et al. 2011), so to break even each spammer will have to generate 8.3 million email sends.[3] Knowing that spammers earn about one purchase per 375 clicks (Kanich et al. 2008) and assuming as before a click-through rate of 0.25 percent, we estimate that 150,000 emails must reach inboxes in order to generate one purchase. That gives us an estimate of 1.8 percent of attempted spams reaching user inboxes (0.15 million out of 8.3 million messages). This estimate is slightly higher than our original estimate, but in the same ballpark.

Given this figure, we can arrive at an estimate of the total user cost of spam. Ninety billion spam messages were sent each day worldwide in 2010 (Symantec 2010; MAAWG 2011); we just estimated that 1.2 percent of these 90 billion get through to the consumer. (These 2010 figures ignore the subsequent 30 percent decrease in global spam due to the Rustock botnet takedown described above, though anecdotal

---

[3] The Kanich group found 1 conversion per 12.3 million emails sent, which is in the right ballpark.

evidence suggests that other botnets have been growing to fill the void.) A large fraction of this spam targets the United States: more than 90 percent of this spam was in English (Symantec, 2010), and Kanich et al. (2008) observe nearly 100 times as much spam going to the United States as to any other country. Suppose, then, that the average value of a user's time is $25 per hour, and that each piece of spam takes an average of five seconds to deal with. (False positives in the spam box are more costly, but so rare that we ignore them in this estimate.) This brings the total worldwide end-user cost of spam to nearly $14 billion per year.

As to the costs of anti-spam technology and hardware, Jennings (2009) in a report published by Ferris Research estimated the costs at approximately $6.5 billion worldwide, based on surveys of firms purchasing anti-spam solutions. This seems roughly correct, given that the largest anti-spam service provider, Symantec, had $6.2 billion in annual revenues in 2011, although it is hard to know exactly how much of the revenue for this firm was due to spam as opposed to network security. Other firms providing anti-spam services to corporate clients include McAfee, Trend Micro, and Barracuda. Our total should also include the labor costs of the staff who install and maintain the anti-spam solutions, and the costs of additional server capacity required by spam email. We believe $6.5 billion is a reasonable estimate for the total, which represents approximately $30 per user for just over a billion users.[4]

If firms were not investing in anti-spam technology, end users would be receiving 100 times as much spam, which given our estimate of the current time loss due to spam, would put the total economic loss at over $1 trillion. However, without any spam filtering it is unlikely that email would be a popular means of communication, so while one cannot take this number literally, it does give a feel for the magnitude of user time savings resulting from private investment in anti-spam technology.

Taken together, the total costs of spam worldwide today appear to be approximately $20 billion, in round numbers. Our estimate is half that of the widely cited Ferris Research (2005) number, because we use a lower value of end-user time, we ignore help-desk support for users struggling with spam, and we use a lower estimate of the number of spams that reach user inboxes.

Measuring the Private Returns to Spam

Researchers have used three tactics to estimate the revenues of botnets and merchants: 1) monitor botnet activity and infiltrate spot markets for spam services, 2) hijack a botnet to estimate the number of purchases generated by a merchant through a spam campaign, and 3) estimate order volume through periodically placing one's own orders and examining the gaps in the sequential order ID numbers.

As an example of the first approach, Stone-Gross, Holz, Stringhini, and Vigna (2011) infiltrated the (then prolific) Cutwail botnet. They were able to monitor every advertising campaign run on the botnet, recording message volume, purpose, and associated merchants. Next, the team infiltrated a private web forum operated

---

[4] For reference, Yahoo! Mail incurs anti-spam costs of approximately $55 million per year for 500 million active email accounts, a cost of $0.10 per account per year.

Cost of Spam Advertising Relative to Other Advertising Media

(cost per thousand impressions (CPM))

Advertising vector	CPM	Breakeven conversion with marginal profit = $50.00	
		Percent	*Per 100,000 deliveries*
Postal direct mail	$250–1,000	2–10%[a]	2000
Super Bowl advertising	$20	0.04%	40
Online display advertising	$1–5	0.002–0.006%	2
Retail spam	$0.10–0.50	0.001–.0002%	0.3
Botnet wholesale spam	$0.03	0.00006%	0.06
Botnet via webmail	$0.05[b]	0.0001%	0.1

Sources: For direct mail, U.S. Postal Service website, For Super Bowl advertising, ⟨http://money.cnn.com /2011/02/03/news/companies/super_bowl_ads?index.hrm⟩. For retail spam and botnet wholesale spam, Stone-Gross, Holz, Stringhini, Vigna (2011); for botnet via webmail, Motoyama, Levchenko, Kanich, McCoy, Voelker, and Savage (2010).

Notes: Cost per thousand impressions (CPM) is the standard unit of measurement in the advertising industry. For spam email, an impression will be a "successful connection"—an email that is not screened out by IP blacklisting and so lands in either the inbox or spam folder.

[a] Direct Marketing Association (2012) reports 2.2%.

[b] Assumes botnet rental is delivery method.

by the botnet masters as a market for spam services. The authors document two ways to publish spam through the Cutwail botnet. Retail spam services were offered at $100 to $500 per million emails in this market. "Wholesale" spam service involves separately acquiring email address lists and renting time on the botnet's spam-send infrastructure. A monthly rental, capable of pumping out 10 million unblocked messages per day, was priced at $10,000, or about $33 per million emails. A more premium wholesale spam product, sending all messages through webmail accounts (and therefore incurring the higher cost of having to break CAPTCHAs), cost about one-third more. The authors estimate that the Cutwail botnet earned $1.7–4.2 million in profit during the 14-month period of study.

In Table 2, we convert the cost estimates for spam from the first research technique (Stone-Gross, Holz, Stringhini, Vigna 2011; Motoyama, Levchenko, Kanich, McCoy, Voelker, and Savage 2010) to the standard unit used in the advertising industry: cost per thousand impressions (CPM). For spam email, an impression will be a "successful connection"—an email that is not screened out by IP blacklisting and so lands in either the inbox or spam folder. To put the figures in perspective, we also include estimates for the cost of sending consumers messages via direct mail, Super Bowl advertising, or legitimate online advertising. We next suppose the average transaction, or "conversion" in online-advertising parlance, to produce profits of $50. Given this assumption, Column 3 gives the conversion rate necessary to break even on each form of advertising. For legibility, Column 4 restates the breakeven conversion rates in units of conversions per 100,000 ads.

Direct mail is the most expensive form of advertising, due to printing and postage costs; this medium thus requires high breakeven conversion rates of at least 2 percent. For the case of $50 profit per sales, standard online display advertising can be profitable down to a conversion frequency of 2 per 100,000 ads, while "premium display" would require 10 per 100,000 ads. Retail spam is profitable down to 0.2 conversions per 100,000. Bulk spam through wholesale botnet rental is sustainable with a mere 0.06 conversions per 100,000 ads, or about 1 in 2,000,000. Clearly, spam can be orders of magnitude less effective than traditional forms of advertising and still remain profitable.

The second research technique, hijacking a botnet, appears in the influential 2008 "Spamalytics" paper (Kanich et al. 2008), in which the researchers co-opted a portion of the Storm botnet by modifying the software instructions given to a set of downstream zombie computers. The modified instructions replaced the link to the spammer's storefront with a link to their own replica storefront. Users could place an order at the replica storefront, but would then receive an error message. The researchers could thus measure how many conversions would have been generated by the spam emails with their modified instructions.

In total, the group modified 345 million pharmaceutical emails sent from botnet zombies. Three-quarters of these were blocked through blacklisting, and the remaining 82 million emails led to a scant 28 conversions, or about 1 in 3,000,000. This conversion rate is far lower than what could be profitable for a retail spam campaign. We suspect that the reason for this lack of success is that a large portion of this major spam campaign went to large email providers like Yahoo! and Gmail and failed to evade their spam filters. We hypothesize that small-scale spammers can beat spam filters more easily and can spend time crafting creatively targeted campaigns; meanwhile, large-scale bulk campaigns spray email like a firehose, but the vast majority of it is blocked by filters.

The same research group also introduced the third estimation technique: placing sequential orders and drawing inferences from order ID numbers (Kanich et al. 2011). They began by making multiple purchases only a few seconds apart. Ten merchants were determined to use simple ascending rules for order IDs; for these merchants, the researchers placed a series of orders spaced over a period of six weeks. The order IDs fully revealed the quantity of other orders placed in the intervening time periods. The researchers also learned that one spammer hosted the images for its storefronts on a server belonging to someone else, which the spammer had hijacked through malware. The researchers notified the server's owner, who in turn gave them permission to monitor requests for the relevant image URLs, which provided reliable data on average order size and the basket of goods purchased. Each of these ten large spam-oriented merchants earned between $500,000 and $1.5 million per month in revenue—of course, profits would be lower. The researchers project that, in total, spam-oriented merchants receive gross revenues of about $180–360 million dollars annually.

We can check this revenue estimate using estimates of the prices and quantities of spam emails sent. As noted earlier, Symantec (2010) estimates the volume of spam at 90 billion attempted connections per day, 80 percent of which are refused

due to blacklisting. Similarly, the Yahoo! Mail team told us that in October 2011, they received approximately 30 billion attempted connections per day, 80 percent of which were bounced, just under 10 percent of which went to the spam folder, and just over 10 percent of which went to a user's inbox. If the unblocked 20 percent of spam is priced at $50 per million ("premium bulk" rates), this would amount to $600,000 worth of spam being sent to Europe and North America each day—so perhaps $750,000 worldwide. This figure seems a bit high given our previous estimate of just under $1 million per day in revenues for the entire supply chain (which must also include the cost of goods sold), but it is of the right order of magnitude.

Overall, we feel comfortable with an estimate of total industry revenue for spam-advertised goods on the order of $300 million per year. One might, in principle, want to include consumer surplus in a calculation of the total benefits of spam. However, because consumers who wanted these goods would likely be able to find them via online searches in the absence of spam, we assume that the consumer benefits are less than the total revenues earned by the spam industry. Since we have estimated the revenues rather than the profits of the spam industry, and we know there are marginal costs to the goods sold, we will assume for convenience that the revenues represent approximately the total surplus generated by spam, including both producer and consumer surplus.

The "Externality Ratio" of Spam in Context

Spam to end users costs around $20 billion annually, compared with approximately $200 million in surplus generated by the spam to these same users. The ratio of the cost of this externality to society relative to the ratio of private benefits it generates is about 100:1.

To put this magnitude into context, Table 3 provides estimates for the externality ratios associated with 1) the air pollution from driving a vehicle, and 2) the (nonviolent) stealing of automobiles. For driving, we use a low value for the benefit accrued to a driver, a figure just above the operation cost per mile. In reality, people make many inframarginal trips, valued by the consumer well over the marginal cost. The cost estimate comes from Delucchi (1998), who does a nice job of accounting for the social cost of the various air pollutants emitted by an automobile; time congestion externalities are not measured so this estimate should be viewed as the cost of driving on an uncongested roadway. (Interested readers are directed to Parry, Walls, and Harrington 2007, who survey the literature more broadly, including the matter of congestion costs.) Delucchi's preferred estimate for the social cost per mile was $0.06; using this figure gives an externality ratio of about 0.1, three orders of magnitude less than the value we obtain for spam. By contrast, stealing automobiles has a much higher externality ratio, as demonstrated by Field (1993). The societal costs include uninsured losses to victims, insurance premiums, law enforcement patrol costs, and the cost of prosecuting and incarcerating offenders who are caught. Adding it all up, the costs imposed on society by auto thieves are a whopping 7 to 30 times the revenue extracted from the vehicle theft.

In certain ways, nonviolent auto theft turns out to be a fairly close analogue to spam. The costs of both auto theft and spam are high, and are distributed diffusely

Table 3
Extracted Revenue, Imposed Costs, and Externality Ratios

Activity	Revenue/benefit	Cost	Externality ratio
Driving automobiles	$0.60 per mile	$0.02–0.25 per mile[a]	0.03–0.41
Stealing automobiles	$400–1200 million per year	$8–12 billion per year	6.7–30.3
Email spam	$160–360 million per year	$14–18 billion per year[b]	39–112

Sources: The source for the first row is Delucchi (1997), for the second row, Field (1993). (The *FBI Uniform Crime Report* (2010) places the vehicle value extracted by criminals in the same range as Field 1993.) The final row is based on the authors' calculations.
[a] Air pollution costs.
[b] Cost to end users.

across the majority of the population (because insurance rates and law enforcement costs account for the bulk of the costs of auto theft, as in Field 1993). Relative to other types of crime with poor insurance coverage, both have particularly diffuse costs. Unlike most crime, spam has no specifically identifiable victim, no especially wronged persons inspiring law enforcement to vigorously bring spammers to justice. In fact, some of the "victims" of spam, those who voluntarily make purchases from illegal advertising, arguably exert large negative externalities on the rest of society. Accounting for how much spam actually reaches the inbox, we estimate that only about 1 in 25,000 people needs to succumb to the temptation to make a grey-market purchase to make it profitable for spammers to inundate everyone with advertisements at current levels. From an economic perspective, one could view a law enforcement system as providing disincentives to make such purchases.

While the externality ratio of spam is large, the cost comes in the form of attention and time, not disease and death as in the case of air pollutants. We are not aware of any estimates of the externality ratio of violent crime, but if we use $10,000,000 as the value of a life and say that a victim of a violent crime has one chance in 1,000 of death, then the expected value of the loss of life would be $10,000. For comparison, the gains to an armed robber may have greater utility than the losses to a victim because of differences in the marginal utility of income, but any plausible estimate of this social welfare gain from this typical armed robbery places the externality ratio far higher than our estimates for spam. Thus, there are examples of externality ratios higher than that of spam, though these tend to have their harm concentrated in a small number of people. Various forms of air pollution are similarly diffuse to spam, and may have much larger social costs than spam, but their externality ratios are much smaller.

Policy Proposals

Considerable effort has gone into anti-spam measures. We already discussed many of the private (and cooperative) technological solutions that have been

adopted by firms in an attempt to reduce the social cost of spam. Here we consider public policy proposals from the legal and economic perspectives.

Legal Interventions

American spam legislation began in earnest with the Telephone Consumer Protection Act (TCPA) of 1991, which, as a response to rising fax machine spam, required fax marketing to be opt-in.[5] The legislation also required phone telemarketers to offer an opt-out. In 2003, a consumer challenge to unsolicited email was unsuccessful; the Pennsylvania Superior Court ruled in *Aronson vs. Bright-Teeth Now* (2003 Pa. Super 187, 824 A.2d 320) that email transmission, without the tangible costs of paper and toner, was legally different from fax transmission. The TCPA did little to stop telemarketing, especially with the *Aronson* decision, because opting out on a firm-by-firm basis was difficult and time consuming. However, the National Do-Not-Call Registry adopted in 2003 allowed consumers to opt out of all telemarketing (with some exemptions for nonprofits and politicians) by filling out a single form.

The first national legislation directed at email spam was the Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003. The cumbersome title created the catchy acronym "CAN-SPAM." The law requires unsolicited email to have a valid return address, to offer a simple opt-out option, and to identify itself as advertising in the subject line. The CAN-SPAM Act does not appear to have markedly affected the illegal advertising market (Sipior, Ward, and Bonner 2004). One reason is that much of spamming activity was already illegal, including the sale of counterfeit goods infringing on trademarks and intellectual property rights, or pharmaceuticals that are illegal to dispense without a prescription in many jurisdictions (or even to ship across state lines to a consumer with a valid prescription). In addition, jurisdictional boundaries hamper spam prosecutions. A spammer may be based in Latvia, work for a merchant in Moscow, send spam to the United States from a botnet with zombie computers all over the world, and have the final goods shipped from India. Governments around the world have not been willing to strain diplomatic relations with other countries over spammers.

A different legal tactic has been proposed by Levchenko et al. (2011). Recall that they found a potential choke point for spammers: the small number of banks willing to process payment for the merchants. American authorities might seek penalties for U.S. banks who transact with spammers in places like Azerbaijan, Latvia, and St. Kitts & Nevis. Indeed, some of the basis for such legislation could come from the war on drugs, since a fair number of spam purchases are for controlled narcotic substances such as oxycodone.

Economic Policy Proposals

To correct problems created by a negative externality, the standard solution in the economist's toolkit is to levy a Pigouvian tax on the externality-causing activity.

---

[5] Some illegal fax spam continued. Horror stories from recipients are documented at ⟨http://www .junkfax.org/fax/stories/Kirsch.html⟩.

In the case of spam, the popular economic solution is to require a "postage stamp," costing perhaps a tenth of a cent, for delivery of an unsolicited email advertisement, and transfer that postage amount to the receiver to compensate them for their attention (for example, Kraut, Morris, Telang, Filer, Cronin, and Sunder 2002; and Bill Gates at the World Economic Forum in Davos, Switzerland, in January 2004 as reported in Jesdanun 2004). However, pricing all email in order to disincentivize the irrelevant material is highly inefficient: many legitimate and useful emails, such as flight reminders and nonprofit newsletters, might well cease to exist.

A related option would be to levy penalties on consumers who purchase goods from spammers, on the grounds that every purchase goes a long way toward increasing the profitability of spam to U.S. consumers. However, enforcing such a law would be quite difficult without severe restrictions on privacy—like giving government the ability to monitor purchase receipts sent to webmail clients.

Instead, economic authors generally prefer the a variant of the Pigouvian tax called "attention bonds" (Loder, Van Alstyne, and Walsh 2004). The idea is to have the sender of an email pay the receiver for attention. The sender sends a bond—for example, say five cents–along with each email. When the recipient reads the mail, the recipient gets the sender's five cents deposited in a bank account (or the recipient can choose to accept the email without payment, returning the money to the sender). The recipient can also "whitelist" a sender to receive all of future emails from that sender, even those with zero posted bond price. This whitelisting option is designed to avoid penalizing useful automatic emails like newsletters and flight-change notifications: solicited (whitelisted) emails have a zero price, which is efficient given the near-zero cost of transmission, while unsolicited emails have a positive price designed to compensate recipients for the imposition on their attention. Any (non-whitelisted) messages without the required bond never reach the recipient's mailbox. Ideally, consumers would have the ability to set individual thresholds for the price of their attention; for example, a high-school student might be willing to look at any unsolicited email whose bond exceeded half a cent, while a busy lawyer might require at least $20. Internalizing the attention externality in this way would give advertisers incentives to make sure they were targeting their emails only to those consumers most likely to be interested in the advertised products, thus increasing economic efficiency.

While we admire the elegance of attention bonds, we wish to sound a note of caution. No method currently exists to link email accounts with payment mechanisms. Should adoption of the attention bond proposal eventually become feasible, how might spammers respond? With attention bonds, a cybercriminal could earn the size of the bond per email, say $0.05 (a figure often suggested, see for example Van Alstyne 2007), by hijacking a legitimate account and sending mail to his own account to collect the bond. Account hijacking is already a serious problem, and the incentives to hijack would increase by at least three orders of magnitude if one could steal $500 by sending 10,000 emails from a hijacked account. Of course, countermeasures could then be taken, but our point is that the attention bond system will surely produce attempts to exploit the new system for profit. By the time one takes into account the transactions costs of setting up an attention bond system, along

with a much heightened incentive to hijack accounts, the overall welfare effects of such a change are unclear to us.

There are two key inefficiencies at work with the sender-push property right of SMTP. This paper has thus far focused on the first: unsolicited email imposes an externality on user attention. The second is that spam has arguably created a stigma for legitimate email marketers, destroying potential surplus that could be created by legitimate players who would, in the absence of such stigma, offer some well-targeted emails to consumers who would mostly appreciate them. This inefficiency has presented an arbitrage opportunity for middlemen, including "daily deal" sites like Groupon and LivingSocial. A daily deal site collects email addresses via consumers opting in. If the deals turn out not to be of sufficiently high quality, consumers can easily opt out with a single action (which is much easier than opting out of unsolicited emails from hundreds of individual merchants). Merchants reach consumers through the transmission rights of the middleman, and pay a substantial fee to do so. As of this writing in mid 2012, Groupon's market valuation exceeds $5 billion and it has about $1.8 billion in annual revenue, which gives one an idea of the size of this second inefficiency. The other major daily deal site, LivingSocial, is a private company, so revenue figures are not available, but the company controls a comparable market share to Groupon. There are many other competitors in this space, ranging from big players such as Google, to local newspapers. We view $5 billion as a reasonable estimate of the daily deal market.

In contrast to the high-level market-design interventions that have been proposed, we feel the most promising economic interventions are those that raise the cost of doing business for the spammers by cutting into their margins and thus making many campaigns unprofitable. As mentioned, one fruitful avenue is to put legal pressure on domestic banks that process payments from foreign banks known to act on behalf of spam merchants. This could put downward pressure on conversion rates and with them, profits. Another proposal comes from our colleague Randall Lewis, who imagines "spamming the spammers" by identifying spam emails and placing fake orders on spam-advertised stores. This step would increase the merchants' costs dramatically, as they would find it much more difficult to fill orders, and their banks may raise their fees if they submit many invalid payment authorization requests. Of course, an unintended consequence is that from time to time a legitimate merchant will be inundated with bogus product orders.

Email-spam advertising has evolved over the past 15 years from a handful of independent spam kings to a well-organized, sophisticated market. The spam supply chain includes merchants at the top, affiliate spammers downstream, and a relatively concentrated market of botnets producing the majority of the spam emails. Nearly 40 trillion spam emails per year advertise a variety of products, including pharmaceuticals, gambling, counterfeit watches, gray-market job opportunities, pornography, software, and dating services. The costs of spam to consumers outweigh the social benefits by an enormous margin, on the order of 100:1. While we admire high-level economic proposals to introduce Pigouvian taxes on spam, our research on the cat-and-mouse games played by spammers leads us to be cautious about the possible

unintended consequences of these proposals. Instead, we advocate supplementing current technological anti-spam efforts with lower-level economic interventions at key choke points in the spam supply chain, such as legal intervention in payment processing, or even spam-the-spammers tactics. By raising spam merchants' operating costs, such countermeasures could cause many campaigns no longer to be profitable at the current marginal price of $20–50 per million emails. These proposals are no panacea, but could bring about a significant reduction in spam.

References

Almeida, Tiago A. Almeida, Gómez, José María, and Yamakami, Akebo. 2011. "Contributions to the Study of SMS Spam Filtering: New Collection and Results." In *Proceedings of the 11th ACM Symposium on Document Engineering,* pp. 259–62. New York, NY: ACM.

Androutsopoulos, Ion, John Koutsias, Koutsias V. Chandrinos, George Paliouras, and Constantine D. Spyropoulos. 2000. "An Evaluation of Naive Bayesian Anti-spam Filtering." arXiv preprint cs/0006013. http://arxiv.org/abs/cs.CL/0006013.

Caballero, Juan, Chris Grier, Christian Kreibich, and Vern Paxson. 2011. "Measuring *Pay-per-Install*: The Commoditization of Malware Distribution." In *Proceedings of the 20th USENIX Security Symposium.* http://static.usenix.org/events/sec11/tech/full_papers/Caballero.pdf.

Caballero, Juan, Pongsin Poosankam, Christian Kreibich, and Dawn Song. 2009. "Dispatcher: Enabling Active Botnet Infiltration Using Automatic Protocol Reverse-Engineering." In *Proceedings of the 16th ACM Conference on Computer and Communications Security,* pp. 621–34. New York, NY: ACM.

Castillo, Carlos, Debora Donato, Aristides Gionis, Vanessa Murdock, and Fabrizio Silvestri. 2007. "Know Your Neighbors: Web Spam Detection Using the Web Topology." In *SIGIR'07: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* pp. 423–30. New York, NY: ACM.

Caverlee, James, and Ling Liu. 2007. "Countering Web Spam with Credibility-based Link Analysis." In *Proceedings of the Twenty-Sixth Annual ACM Symposium on Principles of Distributed Computing,* pp. 157–66. New York, NY: ACM.

Cho, Chia Yuan, Juan Caballero, Chris Grier, Vern Paxson, and Dawn Song. 2010. "Insights from the Inside: A View of Botnet Management from Infiltration." In *Proceedings of the Third USENIX Workshop on Large-scale Exploits and Emergent Threats (LEET'10).* http://www.icsi.berkeley.edu/pubs/networking/insightsfrom10.pdf.

Cohen, David. 2012. "Busted: Fake Facebook Friend Requests." *AllFacebook: The Unofficial Facebook Blog,* February 28. http://allfacebook.com/facebook-fake_b79558.

Cook, Duncan, Jacky Hartnett, Kevin Manderson, and Joel Scanlan. 2006. "Catching Spam before It Arrives: Domain Specific Dynamic Blacklists." In *ACSW Frontiers '06: Proceedings of the 2006 Australasian Workshops on Grid Computing and E-Research,* Volume 54, pp. 193–202. Darlinghurst, Australia: Australian Computer Society.

Delucchi, Mark A. 1998. *The Annualized Social Cost of Motor-Vehicle Use in the US, 1990–1991:*

Summary of Theory, Data, Methods, and Results. UCD-ITS-RR-96-3(1), Institute of Transportation Studies, University of California, Davis. http://www.fhwa.dot.gov/scalds/delucchi.pdf.

Direct Marketing Association. 2010 "Response Rate Trend Reports." *DMA White Paper.* Available at: http://www.the-dma.org/cgi/dispannouncements?article=1451.

Evans, David S. 2009. "The Online Advertising Industry: Economics, Evolution, and Privacy." *Journal of Economic Perspectives* 23(3): 37–60.

Everett-Church, Ray. 1999. "The Spam That Started It All." *Wired Magazine,* April 13. http://www.wired.com/politics/law/news/1999/04/19098.

FBI. 2010. *Uniform Crime Report: Crime in the United States, 2010: Motor Vehicle Theft.* http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/property-crime/mvtheftmain.pdf.

Ferris Research. 2005. "The Global Economic Cost of Spam." Report #409.

Field, Simon. 1993. "Crime Prevention and the Costs of Auto Theft: An Economic Analysis." *Crime Prevention Studies,* Volume 1, edited by Ronald V. Clarke, 69–91. Lynne Rienner Publishers.

Ghiossi, Caroline. 2010. "Explaining Facebook's Spam Prevention Systems." *The Facebook Blog,* June 29. https://blog.facebook.com/blog.php?post=403200567130.

Gómez Hidalgo, José María, Guillermo Cajigas Bringas, Enrique Puertas Sánz, and Francisco Carrero García. 2006. "Content Based SMS Spam Filtering." In *Proceedings of the 2006 ACM Symposium on Document Engineering,* pp. 107–114. New York, NY: ACM.

Goodman, Joshua, Gordon V. Cormack, and David Heckerman. 2007. "Spam and the Ongoing Battle for the Inbox." *Communications of the ACM* 50(2): 24–33.

Gross, Doug. 2011. "Again? Sony's Playstation Network Hit with Another Attack." *CNN Tech,* October 12. http://articles.cnn.com/2011-10-12/tech/tech_gaming-gadgets_sony-playstation-network-attack_1_lulzsec-passwords-sony-pictures?_s=PM:TECH.

Herley, Cormac, and Dinei Florêncio, D. 2009. "A Profitless Endeavor: Phishing as Tragedy of the Commons." In *NSPW'08: Proceedings of the New Security Paradigms Workshop,* pp. 59–70. ACM. http://www.nspw.org/papers/2008/nspw2008-herley.pdf.

Isacenkova, Jelena, and Davide Balzarotti. 2011. "Measurement and Evaluation of a Real World Deployment of a Challenge-Response Spam Filter." In *Proceedings of ACM ICM 2011,* http://conferences.sigcomm.org/imc/2011/docs/p413.pdf.

Jennings, Richi. "Cost of Spam is Flattening—Our 2009 Predictions. Ferris Research, http://email-museum.cm/2009/01/28/cost-of-spam-is-flattening-our-2009-predictions/.

Jesdanun, Anick. 2004. "Is Metered E-mail a Viable Anti-spam Tactic?" *USA Today,* March 5. http://www.usatoday.com/tech/news/techpolicy/2004-03-05-metering-email_x.htm.

John, John P., Alexander Moshchuk., Steven D. Gribble, and Arvind Krishnamurthy. 2009. "Studying Spamming Botnets Using Botlab." In *NSDI'09: Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation,* pp. 291–306. Berkeley, CA: USENIX Association.

Kanich, Chris, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M. Voelker, Vern Paxson, and Stefan Savage. 2008. "Spamalytics: An Empirical Analysis of Spam Marketing Conversion." In *Proceedings of the 15th ACM Conference on Computer and Communications Security.* ACM. http://www.icsi.berkeley.edu/pubs/networking/spamalytics.pdf.

Kanich, Chris, Nicholas Weaver, Damon McCoy, Tristan Halvorson, Christian Kreibich, Kirill Levchenko, Vern Paxson, Geoffrey M. Voelker, and Stefan Savage. 2011. "Show Me the Money: Characterizing Spam-advertised Revenue." In *Proceedings of the 20th USENIX Security Symposium.* http://static.usenix.org/events/sec11/tech/full_papers/Kanich.pdf.

Kotadia, Munir. 2004. "Porn Gets Spammers Past Hotmail, Yahoo Barriers." *CNET News,* May 6. http://news.cnet.com/2100-1023-5207290.html.

Krause, Beate, Christoph Schmitz, Andreas Hotho, and Gerd Stumme. 2008. "The Anti-social Tagger: Detecting Spam in Social Bookmarking Systems." In *AIRWeb'08: Proceedings of the 4th International Workshop on Adversarial Information Retrieval on the Web,* pp. 61–68. New York, NY: ACM.

Kraut, Robert E., James Morris, Rahul Telang, Darrin Filer, Matt Cronin, and Shyam Sunder. 2002. "Markets for Attention: Will Postage for Email Help?" In *CSCW '02 Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work,* pp. 206–215. New York, NY: ACM.

Levchenko, Kirill, et al. 2011. "Click Trajectories: End-to-End Analysis of the Spam Value Chain." *IEEE Symposium on Security and Privacy 2011,* pp. 431–46.

Loder, Thede, Marshall Van Alstyne, and Rick Wash. 2004. "An Economic Answer to Unsolicited Communication." In *EC '04 Proceedings of the 5th ACM Conference on Electronic Commerce,* pp. 40–50.

Mailer Mailer, LLC. 2011. "Email Marketing Metrics Report." Corporate White Paper. Available at: http://www.mailermailer.com/resources/metrics/2011/click-rates.rwp.

Malik, Haroon. 2008. "IBM Says Storm Worm Creators Making Millions Daily." *Gizmodo,*

February 10. http://gizmodo.com/354741/ibm-says -storm-worm-creators-making-millions-daily.

Messaging Anti-Abuse Working Group (MAAWG). 2011. *Email Metrics Program: The Network Operator's Perspective.* Report 14. http:// www.maawg.org/sites/maawg/files/news /MAAWG_2010_Q3Q4_Metrics_Report_14.pdf.

Microsoft. 2011. "Battling the Rustock Threat." In *Microsoft Security Intelligence Report, Special Edition.* Microsoft Corporation.

Moore, Tyler, Richard Clayton, and Ross Anderson. 2009. "The Economics of Online Crime." *Journal of Economic Perspectives* 23(3): 3–20.

Motoyama, Marti, Kirill Levchenko, Chris Kanich, Damon McCoy, Geoffrey Voelker, and Stefan Savage. 2010. "Re: CAPTCHAs—Understanding CAPTCHA-solving Services in an Economic Context." In *Proceedings of the 19th USENIX Security Symposium,* Volume 10.

Moyer, Edward. 2011. "Breach Exposes Chase, Capital One, TiVo Customers." *CNET News,* April 2. http://news.cnet.com/8301-1009_3-20050068-83 /breach-exposes-chase-capital-one-tivo-customers/.

Ntoulas, Alexandros, Marc Najork, Mark Manasse, and Dennis Fetterly. 2006. "Detecting Spam Web Pages through Content Analysis." In *WWW '06: Proceedings of the 15th International Conference on World Wide Web,* pp. 83–92. New York, NY: ACM.

Parry, Ian W. H., Margaret Walls, and Winston Harrington. 2007. "Automobile Externalities and Policies." *Journal of Economic Literature* 45(2): 373–99.

Ramachandran, Anirudh, Anirban Dasgupta, Nick Feamster, and Kilian Weinberger. 2011. "Spam or Ham? Characterizing and Detecting Fraudulent 'Not Spam' Reports in Web Mail Systems." In *Proceedings of ACM CEAS 2011: 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference.* ACM Digital Library.

Ramachandran, Anirudh, Nick Feamster, and Santosh Vempala. 2007. "Filtering Spam with Behavioral Blacklisting." In *CCS '07: Proceedings of the 14th ACM Conference on Computer and Communications Security,* pp. 342–51. ACM Digital Library.

Sahami, Mehran, Susan Dumais, David Heckerman, and Eric Horvitz. 1998. "A Bayesian Approach to Filtering Junk E-mail." In *Learning for Text Categorization: Papers from the 1998 Workshop,* vol. 62, pp. 98–105. Madison, Wisconsin: AAAI Press.

Samosseiko, Dimitry. 2009. "The Partnerka: What is It, and Why Should You Care?" In *VB2009: Proceedings of Virus Bulletin Conference.* http:// www.sophos.com/security/technical-papers /samosseiko-vb2009-paper.pdf.

Sipior, Janice C., Burke T. Ward, and P. Gregory Bonner. 2004. "Should Spam Be on the Menu?" *Communications of the ACM* 47(6): 59–63.

Stone-Gross, Brett, Thorsten Holz, Gianluca Stringhini, and Giovanni Vigna. 2011. "The Underground Economy of Spam: A Botmaster's Perspective of Coordinating Large-scale Spam Campaigns." Presented at *LEET'11: 4th USENIX Workshop on Large-Scale Exploits and Emergent Threats.* http://iseclab.org/papers/cutwail-LEET11.pdf.

Symantec. 2010. *Messagelabs Intelligence: 2010 Annual Security Report.*

Symantec. 2012. "Intelligence Report, May 2012." May, 2012. Available at: http://www.symanteccloud .com/globalthreats/overview/r_mli_reports.

Templeton, Brad. Undated. "The Origin of the Term 'Spam' to Mean Net Abuse." http://www .templetons.com/brad/spamterm.html.

Thonnard, Oliver, and Marc Dacier. 2011. "A Strategic Analysis of Spam Botnets Operations." In *CEAS '11: Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference,* pp. 162–171. AMC Digital Library.

Tran, Hung, Thomas Hornbeck, Viet Ha-Thuc, James Cremer, and Padmini Srinivasan. 2011. "Spam Detection in Online Classified Advertisements." In *WebQuality '11: Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality,* pp. 35–41. AMC Digital Library.

Van Alstyne, Marshal. 2007. "Curing Spam: Rights, Signals & Screens." *Economists' Voice* 4(2).

Warren, Christina. 2011. "How to: Avoid and Prevent Facebook Spam." March 28. http:// mashable.com/2011/03/28/facebook-spam-tips/.

Yardi, Sarita, Daniel Romero, Grant Schoenebeck, and danah boyd. 2009. "Detecting Spam in a Twitter Network." *First Monday* 15(1). http:// firstmonday.org/htbin/cgiwrap/bin/ojs/index .php/fm/article/view/2793/2431.

Zdziarski, Jonathan A. 2005. *Ending Spam: Bayesian Content Filtering and the Art of Statistical Language Classification.* No Starch Press.

Zhao, Yao, Yinglian Xie, Fang Yu, Qifa Ke, Yuan Yu, Yan Chen, and Eliot Gillum. 2009. "Botgraph: Large Scale Spamming Botnet Detection." In *NSDI '09: Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation,* pp. 321–34. Berkeley, CA: USENIX Association.

Zhou, Dengyong, Christopher J. C. Burges, and Tao Tao. 2007. "Transductive Link Spam Detection." In *AIRWeb '07: Proceedings of the 3rd International Workshop on Adversarial Information Retrieval on the Web,* pp. 21–28. New York, NY: ACM.

Identifying the Disadvantaged: Official Poverty, Consumption Poverty, and the New Supplemental Poverty Measure[†]

Bruce D. Meyer and James X. Sullivan

Few economic indicators are more closely watched or more important for policy than the official poverty rate. It is used to gauge the extent of deprivation in the United States and to determine how economic well-being has changed over time. The poverty rate is often cited by policymakers, researchers, and advocates who are evaluating social programs that account for more than half a trillion dollars in government spending. Eligibility for some means-tested transfer programs is determined based on the poverty thresholds, and local poverty rates affect the allocation of billions of dollars in federal funds.

The methods for calculating the current poverty measure, largely unchanged since the 1960s, have been criticized by many researchers. In response, the Census Bureau has led a two-decade process of research and discussion of poverty measurement with an eye to revising the official measure. The process has involved hundreds of papers, dozens of official Census Bureau publications (U.S. Census 2010), and two National Academy of Sciences reports (Citro and Michael 1995; Iceland 2005). We will not summarize this vast literature here. Rather, we will examine the properties of three measures of poverty: the official U.S. poverty rate; the new Supplemental Poverty Measure first released by the U.S. Census Bureau in fall 2011; and a consumption-based measure of poverty. We will focus on two fundamental goals of these measures: to identify the most disadvantaged and to assess changes

■ *Bruce D. Meyer is McCormick Foundation Professor, Harris School of Public Policy Studies, University of Chicago, Chicago, Illinois. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. James X. Sullivan is Associate Professor of Economics, University of Notre Dame, Notre Dame, Indiana. Their email addresses are ⟨bdmeyer@uchicago.edu⟩ and ⟨sullivan.197@nd.edu⟩.*

over time in disadvantage. These goals accord very closely with those stated in the National Academy of Sciences report *Measuring Poverty*: "The panel proposes a new measure that will more accurately identify the poor population today. . . . Equally important, the proposed measure will more accurately describe changes in the extent of poverty over time that result from new public policies and further social and economic change" (Citro and Michael 1995, pp. 1–2).

We start by describing these three approaches to measuring poverty. We then compare these measures of poverty by looking at the demographic and material circumstances of who they define as poor. A measure of poverty can, of course, produce a higher or lower poverty rate depending on how high the cutoffs that define poverty are set. However, two different measures of poverty that include the same overall number of poor people will be made up of overlapping but different groups. By looking at the characteristics of those who a given poverty measure would include, or would leave out, we can provide evidence on whether that measure does a better job of capturing the disadvantaged. For example, we find that, compared to the official poverty measure, the Supplemental Poverty Measure adds to poverty individuals who are more likely to be college graduates, own a home and a car, live in a larger housing unit, have air conditioning, health insurance, and substantial assets, and have other favorable characteristics than those who are dropped from poverty. On the other hand, we find that a consumption measure, compared to the official measure or the Supplemental Poverty Measure, adds to the poverty rolls individuals who appear worse-off. We then examine how each of the poverty measures assesses changes in disadvantage over time. The Supplemental Poverty Measure uses a complex and convoluted way of determining changes in poverty over time that we argue makes it difficult to interpret.

Our results present strong evidence that a consumption-based poverty measure is preferable to both the official income-based poverty measure and to the Supplemental Poverty Measure for determining who are the most disadvantaged. Our findings also raise the question as to whether a flawed measure of income, even when modified to be conceptually closer to consumption, can reliably be used to measure poverty.

Three Ways of Measuring Poverty

The broader literature on measuring poverty proposes a wide variety of approaches for identifying who is poor. Some approaches are multidimensional, emphasizing functional capabilities, social inclusion, relationships, the environment, and other components of well-being (Atkinson, Cantillon, Marlier, and Nolan 2002; Stiglitz, Sen, and Fitoussi 2009). In this article, we will focus on three single-dimensional, resource-based poverty measures.

Single-dimensional poverty measures are typically constructed by making a set of eight choices: 1) How should the resources available to people be defined? Typically, resources are measured using income or consumption, but there is debate

about how to define income and consumption. 2) Is an annual measure about right for measuring poverty, or should poverty be measured over shorter or longer time periods? 3) Should the resource-sharing unit that is pooling income and making joint purchases be a group of related family members or another unit such as a group of people sharing a residence? 4) Should the measure count the number of people with resources below a cutoff or threshold (a head count measure), or should it specify the total resources needed to raise all of the poor up to the poverty threshold (a poverty gap measure)? 5) Should the poverty threshold be set as an absolute level of resources or relative to some standard, such as the median level of income? For example, the European Union focuses on a measure of poverty defined as the fraction below 60 percent of median income. 6) Where should the poverty line, or thresholds, be drawn, recognizing that this essentially arbitrary choice will have a large effect on the estimated poverty rate? 7) Should poverty thresholds be adjusted over time using the rise in the cost of living or the rise in income levels, and should they be adjusted for geographic price differences or other factors? 8) How should the "equivalence scale" be determined to set poverty thresholds for families that differ in size or composition? In describing the three poverty measures, we will touch upon each of these issues, although we will leave a full discussion of the adjustment of the thresholds over time until later. For now, we focus on the determinants of poverty at a point in time.

The Official Poverty Measure

The official poverty rate in the United States is determined by comparing the pretax money income of a family or a single unrelated individual to poverty thresholds that vary by family size and composition. For example, in 2011, the poverty threshold for a one-parent, two-child family was $18,106 (for current and past poverty thresholds, see the U.S. Census Bureau data at ⟨http://www.census .gov/hhes/www/poverty/data/threshld/index.html⟩). The underlying data on pretax money income come from the Current Population Survey Annual Social and Economic Supplement. If a family has income below the poverty threshold for that size family, all family members are classified as poor. In terms of the eight choices needed to define a poverty measure, the resources are pretax money income, the time period is one year, and the resource sharing unit is the family (or those related by blood or marriage). Official poverty is a discrete, head count measure. The original thresholds were based on the cost of a food plan—a nutritionally balanced, low-cost diet for families of different size and composition. For most families, the cost of the food plan was multiplied by three because 1955 survey data on expenditures (the data available when this poverty line was first defined in the early 1960s) suggested that the average family of three or more people allocated about a third of their after-tax income for food. Variation in the cost of the plan by family size and composition provided an implicit equivalence scale that accounts for different food needs across these families. Except for a few minor changes, the only adjustment to these thresholds over the past five decades has been for inflation, using the Consumer Price Index for all Urban Consumers. There is no geographic adjustment. For a more detailed summary, see Citro and Michael (1995), Blank (2008), and Blank and Greenberg (2008).

The official poverty measure has a number of widely recognized flaws. Here, we focus on two of them. First, it defines resources as pretax money income, failing to reflect the full resources at a family's disposal. Pretax money income does not subtract tax liabilities (even poor workers must pay payroll taxes for Social Security and Medicare), nor does it include the Earned Income Tax Credit and other tax credits or noncash benefits such as food stamps, housing or school lunch subsidies, or public health insurance. Thus, many of the major antipoverty initiatives of the last few decades are not reflected in the poverty rate, because policies like a rise in the Earned Income Tax Credit, a more generous Child Tax Credit, and expansions of Medicaid and food stamps do not show up as pretax money income.

Second, the equivalence scale implicit in the official poverty thresholds—that is, the relationship between poverty thresholds for families with different numbers and ages of people—has been criticized. These thresholds reflect the economies of scale in food, but not in other goods. In addition, the scale implicit in the official poverty thresholds suggests children are more costly than adults in some cases and does not exhibit diminishing marginal increments for additional individuals over the whole range of family sizes (Ruggles 1990). For example, the second child in a two-parent family adds much more to the poverty thresholds than the first or third child.

The Supplemental Poverty Measure

In November 2011, the U.S. Census Bureau released the Supplemental Poverty Measure for the first time. It indicated a poverty rate of 16.0 percent for 2010, instead of the 15.1 percent estimated by the official poverty measure. However, as noted earlier, the selection of a poverty cutoff is inherently arbitrary, so the finding that the poverty rate as calculated by the Supplemental Poverty Measure exceeds the official rate is a subjective or political decision, not a scientific one. The release of this new poverty measure reflects the culmination of more than three decades of research on poverty measurement; in particular, this measure is largely based on a 1995 National Academy of Sciences report (Citro and Michael 1995) and follow-up workshop (Iceland 2005). According to the Census Bureau, the Supplemental Poverty Measure is intended to "be an additional macroeconomic statistic providing further understanding of economic conditions and trends" (Short 2011, p. 3). It is designed to complement the current official measure, not to replace it, and it will be published in the future alongside the official rate, funding permitting. There has been a parallel effort to produce poverty measures similar to the Supplemental Poverty Measure for certain states and localities.[1]

---

[1] These efforts include New York City estimates from researchers at the Center for Economic Opportunity, Minnesota estimates from the Urban Institute, Wisconsin estimates from researchers at the University of Wisconsin, and estimates for other states (Levitan, D'Onofrio, Krampner, Scheer, and Seidel 2010; Zedlewski, Giannarelli, Wheaton, and Morton 2010; Chung, Isaacs, Smeeding, and Thornton 2012). While these studies calculate alternative poverty rates using procedures similar to those for the Supplemental Poverty Measure, some differences do exist. For example, the state-level studies do not use income data from the Current Population Survey. Instead, to obtain a large sample, they employ the American Community Survey which lacks information on certain income sources such as food stamp amounts and receipt of housing subsidies.

The Supplemental Poverty Measure differs from the official poverty measure in a number of ways. Perhaps most important, it uses a definition of income that is conceptually closer to resources available for consumption. In addition, it includes a more defensible adjustment for family size and composition, and an expanded definition of the family unit that includes cohabitors.

Recall that the official poverty measure is based on pretax money income. The Supplemental Poverty Measure resource definition includes not only money income, but also tax credits like the Earned Income Tax Credit and the Child Tax Credit, as well as the value of some noncash benefits. In addition, the measure of resources subtracts several categories of expenses from income, including tax liabilities, payments for child support, child care and other work expenses, and out-of-pocket medical expenses.[2] Thus, this measure of resources more closely approximates resources available for consumption than does pretax money income. Also, by including tax credits and in-kind transfers, the Supplemental Poverty Measure is intended to gauge more accurately the effectiveness of antipoverty efforts.

The official poverty measure treats the resource-sharing unit as those related by family ties; in contrast, the sharing unit in the Supplemental Poverty Measure also includes cohabitors and their children, who are treated in the official measure as a separate family unit within the household even though they live together and may share resources. Analytically, the sharing unit should be, well, those who share resources. Information on resource sharing across cohabitors is not collected in the Current Population Survey, although resources or cost-sharing provided to a family by cohabitors may be substantial. The treatment of cohabitors has become more important in recent years as the fraction of households with cohabitors present has risen.

The Supplemental Poverty Measure thresholds are based on expenditure data for food, clothing, shelter, and utilities from the Consumer Expenditure Interview Survey.[3] To arrive at the thresholds, the first step is to pool all consumer units with exactly two children from the past five years of data. Because these families will differ in the number of adults in the unit, a three-parameter equivalence scale is used to convert spending for these families into spending for the reference family of two adults and two children. The overall three-parameter equivalence scale is of the following form (A is the number of adults and C is the number of children): $A^{0.5}$ for one- and two-adult units; $[A + 0.8 + 0.5(C - 1)]^{0.7}$ for single-parent families; and $[A + 0.5C]^{0.7}$ for all other families. The parameter in front of C represents the child proportion of an adult, the exponent is the economies of scale factor, and 0.8 allows for a separate adjustment for single-parent families to reflect the fact that the first child in such families consumes less in total resources than an adult but more than the first child in two-parent families.

---

[2] The Current Population Survey recently added questions so that it could estimate these expenses subtracted from income, but this information is not available historically.

[3] The thresholds for the Supplemental Poverty Measure are provided to the Census Bureau by the Bureau of Labor Statistics. See Garner and Hokayem (2011) and Garner (2010) for more details on these thresholds.

To specify the threshold levels, the Supplemental Poverty Measure then focuses on consumer units who are between the 30th and 36th percentiles of equivalence-scale-adjusted spending on food, clothing, shelter, and utilities (FCSU) for this pooled two-child sample. The measure relies on a moving average of spending over five years, with the data for different years indexed using the Consumer Price Index. Separate poverty thresholds are calculated for three different housing status groups: renters, homeowners with a mortgage, and homeowners without a mortgage (those in public housing are included in this last group). Mean overall shelter and utility expenses are subtracted from the mean FCSU spending for each housing status group, and then the mean shelter and utility expenses within each of these groups is added back. The resulting adjusted mean is then multiplied by 1.2 (to account for "additional basic needs") to determine the reference threshold for each housing status group.

The thresholds for other size families are then calculated from these reference thresholds for the three groups of families using the three-parameter equivalence scale. This equivalence scale offers several important improvements over the scale implicit in the official thresholds. In particular, it is a more transparent and consistent adjustment for differences in needs across families of different sizes and composition. Unlike the scale adjustment in the official measure, it exhibits diminishing marginal cost with each additional child or adult.

Finally, the Supplemental Poverty Measure makes an additional adjustment to the poverty thresholds to reflect geographic variation in the cost of living. This adjustment is based on American Community Survey estimates over five years of median gross rent for a typical apartment for the 264 metropolitan statistical areas observed in the Current Population Survey. For those outside of metropolitan statistical areas, state-level medians for nonmetropolitan areas are estimated. There is considerable geographic price variation in housing. This adjustment is controversial. Rents vary across locations, but at least part of this variation reflects geographical differences in amenities and wages.[4]

Consumption-Based Poverty Measures

Both the official poverty measure and the Supplemental Poverty Measure use income as the measure of resources. However, annual income will not capture the standard of living of individuals who smooth consumption by drawing upon savings. Also, income-based measures of well-being will not capture differences over time or across households in wealth accumulation, ownership of durable goods such as houses and cars, or access to credit. In addition, many antipoverty programs provide an insurance value to households that will not be reflected in their income. These conceptual limitations have influenced a large literature that looks at consumption-based measures

---

[4] While most of the features of the Supplemental Poverty Measure follow the recommendations of the 1995 National Academy of Sciences report, there are differences. For example, the Supplemental Poverty Measure uses a different equivalence scale than recommended in the 1995 report; it specifies thresholds that vary by housing status; and it determines thresholds using a five-year moving average of expenditures. Hutto, Waldfogel, Kaushal, and Garfinkel (2011) provide more details on this point.

of well-being and discusses their advantages (Cutler and Katz 1991; Poterba 1991; Slesnick 1993, 2001; Meyer and Sullivan 2003, 2011, 2012).

Another advantage of consumption is that it appears to be a better predictor of deprivation than income; in particular, material hardship and other adverse family outcomes are more severe for those with low consumption than for those with low income (Meyer and Sullivan 2003, 2011).

Yet another advantage is that consumption appears to be more accurately reported than income for the most disadvantaged families. Income in the Current Population Survey appears to be substantially underreported, especially for categories of income important for those with few resources, and the extent of underreporting has worsened over time. For example, the share of dollars received from means-tested transfer programs that are reported in the Current Population Survey is low and declining (Meyer, Mok, and Sullivan 2009; Meyer and Goerge 2011). The shares reported have fallen below 0.6 for food stamps and 0.5 for Temporary Assistance for Needy Families in recent years. In the most recent Current Population Survey data for 2010, only 36 percent of food stamp dollars paid out to families are directly reported in the survey. Another 20 percent of the dollars paid out are imputed to those who did not report receiving food stamps, leaving 44 percent neither reported nor imputed.[5] Comparisons of survey microdata to administrative microdata for the same individuals also indicate severe underreporting of government transfers in other household surveys such as the American Community Survey (which has been used to implement state and local versions of the Supplemental Poverty Measure).

Comparisons of income and consumption at the bottom of the distribution provide additional evidence that income is underreported. Reported consumption exceeds reported income at the bottom of the distribution, even for those with little or no assets or debts (Meyer and Sullivan 2003, 2011). For recent years, the 5th percentile of the expenditures distribution in the Consumer Expenditure Survey is more than 40 percent higher than the 5th percentile of the income distribution in the Current Population Survey. For families in the Consumer Expenditure Survey in the bottom 5 percent of the income distribution, expenditures exceed income by more than a factor of seven (Meyer and Sullivan 2011).[6]

---

[5] The Current Population Survey, in its current form, also lacks important information for imputing some in-kind benefits. For example, the value of housing subsidies is imputed for each household in the survey that reports receipt of such subsidies. However, because the size of the housing unit is not observed in the Current Population Survey, this must be imputed based on family composition. A reasonable estimate of housing subsidies can be computed using the Consumer Expenditure Survey because the survey provides information on out-of-pocket rent and the characteristics of the housing unit, including the number of rooms, bathrooms and bedrooms, and appliances such as a washer and dryer.

[6] While comparisons of survey data on aggregate expenditures to National Income and Product Accounts (NIPA) consumption indicate underreporting of expenditures as well, the poor consume a different bundle of goods than the general public, so that the typical comparisons do not reflect the composition of consumption for the poor. In fact, key components of spending match up well with national income and product account (NIPA) aggregates, and these components account for a large fraction of total spending for the poor—about 70 percent of consumption for those near the poverty line (Meyer and Sullivan 2012). For food at home, on average the Consumer Expenditure Survey/NIPA ratio is over 0.85, and for rent plus utilities, the ratio is nearly 1.00 (Bee, Meyer, and Sullivan forthcoming).

In terms of the choices at the beginning of this section, we construct a consumption measure of poverty in the following way. Our resource measure is expenditures, excluding human capital investments such as educational and medical expenses. We also exclude purchases of vehicles and mortgage and property tax payments by homeowners, which we replace with a flow value of car- and homeownership. We annualize expenditures, which are reported for a three-month period in the survey. The underlying source of our data is the Consumer Expenditure Interview Survey, which asks respondents if they share resources and uses that information to define the unit of analysis. We use a headcount measure of poverty, as does the official measure and the Supplemental Poverty Measure. We also use the same three-parameter equivalence scale as the Supplemental Poverty Measure. We set the poverty thresholds so that the same share of people is below the poverty line as with the other poverty measures. For more detail, see Meyer and Sullivan (2012).

Who Do the Poverty Measures Identify as Poor?

While many alternative poverty measures have been proposed, surprisingly little research has been done to assess how well these measures identify the disadvantaged. The 1995 National Academy of Sciences Report *Measuring Poverty* includes a table of mean demographic characteristics of those who are poor under the official definition and the proposed alternative measure. A similar table can be found in Short (2011). Both sources do not venture much beyond this analysis. Choices about an appropriate poverty measure are rarely decided by empirical tests of their implications for the characteristics of the poor. In this section, we seek to place the choice of a poverty measure on a firmer footing by presenting empirical evidence on how well different poverty measures capture deprivation.

A typical comparison of the poor under alternative definitions can be seen in Table 1, which reports mean characteristics of the poor in 2010 for three different measures: official poverty, the Supplemental Poverty Measure, and consumption poverty. To ensure that differences in mean characteristics are not simply the result of looking at different cutoffs in the distribution of resources, we keep the baseline poverty rate constant at the estimated Supplemental Poverty Measure rate in 2010 in the Consumer Expenditure Survey (16.5 percent).[7] Thus, each of the three measures of poverty in Table 1 designates the same number of people as poor, but as Table 1 clearly shows, the three poverty measures differ considerably in who is designated as poor. Those categorized as "poor" by the Supplemental Poverty Measure appear less disadvantaged than the official poor: they have higher consumption, are much more likely to have private health insurance, are more likely to own a home and various appliances, are slightly more educated, and have accumulated more assets.

---

[7] For example, for the official measure, we find the 16.5 percentile of the distribution of the official income-to-poverty threshold ratio and then report mean characteristics for those with a ratio below that percentile.

Table 1

Mean Characteristics of the Official, Supplemental Poverty Measure (SPM), and Consumption Poor, Consumer Expenditure Survey, 2010

	Official income poor (1)	SPM poor (2)	Consumption poor (3)
Consumption	$26,886	$29,140	$18,000
Head employed	48%	47%	57%
Number of earners	.91	.97	1.50
Any health insurance	62%	63%	57%
Private health insurance	27%	34%	27%
Homeowner	37%	41%	35%
Single family home	27%	32%	26%
Own a car	73%	75%	74%
Service flows from vehicles	$398	$502	$277
Service flows from owned homes	$1,998	$2,442	$1,012
Total service flows	$2,395	$2,944	$1,289
Family size	3.72	3.51	4.51
# of children	1.70	1.37	1.88
# over 64	0.19	0.26	0.21
# of rooms	6.06	6.34	5.08
# of bedrooms	3.02	3.13	2.60
# of bathrooms	1.64	1.73	1.33
Appliances and amenities			
Microwave	92%	93%	91%
Disposal	33%	35%	30%
Dishwasher	41%	44%	36%
Any air conditioning	74%	75%	72%
Central air conditioning	48%	49%	45%
Washer	70%	73%	71%
Dryer	63%	66%	61%
Television	96%	96%	94%
Computer	63%	64%	61%
Education of head			
Less than high school	34%	33%	40%
High school degree	32%	31%	32%
Some college	26%	26%	21%
College graduate	9%	10%	7%
Race of head			
White	72%	73%	73%
Black	22%	21%	21%
Asian	4%	4%	4%
Other	2%	3%	3%
Hispanic origin	27%	24%	33%
Family type			
Single parent families	31%	28%	29%
Married parent families	32%	25%	38%
Single individuals	20%	22%	14%
Married without children	6%	9%	8%
Head 65 and over	12%	16%	10%
Total financial assets			
75th percentile	$260	$500	$300
90th percentile	$2,400	$4,000	$2,502
Unweighted number of families	4,893	5,085	3,704

Notes: The official income and consumption poverty measures are anchored to the SPM poverty rate for this sample, or 16.5 percent. Consumption poverty is calculated using the three-parameter equivalence scale. Financial asset statistics come from samples of families in their fifth Consumer Expenditure Survey interview. Rooms and total consumption are equivalence-scale adjusted and set equal to a family with two adults and two children. All characteristics are for the family but are weighted by family size.

Conversely, those categorized as "poor" by the consumption measure appear more disadvantaged than the official poor: they have much lower consumption, are less likely to have health insurance, are less likely to own most appliances, and are less educated.

The means for those in poverty reported in Table 1 mask some important differences across these measures. A comparison of mean characteristics of the poor under different definitions does not distinguish between those added and subtracted from poverty. The means are also silent on how many people have their poverty status altered by the change of measure. In comparing any two measures of poverty, there will be some people identified as poor under both measures, some poor under neither measure, and some that are poor under one measure but not the other. Thus, a useful way to compare two measures of poverty is to focus on the characteristics of those whose poverty status is altered in moving from one measure to another. A poverty measure that more accurately identifies the disadvantaged would add to poverty individuals who are worse off in other dimensions than those who are subtracted. We attempt to look at all measures of well-being that are available in the datasets we use. One can think of the process as determining which single measure of material well-being is most correlated with other measures of well-being.[8]

The analyses that follow rely on data from the Consumer Expenditure Survey. These data include both income and consumption, as well as information on ownership of durables and assets that is not available in the Current Population Survey. Another advantage of the Consumer Expenditure Survey data is that information is available to calculate a historical series for a Supplemental Poverty Measure. Such calculations cannot be made using the Current Population Survey data because many of the expenses subtracted from income are only available in recent years. Our results are not sensitive to our choice of dataset. In fact, for variables available in both surveys, our analyses line up very closely. For example, our estimate of the Supplemental Poverty Measure poverty rate for 2010 using the Consumer Expenditure Survey, 16.5 percent, is very close to Census estimates of the Supplemental Poverty Measure poverty rate using Current Population Survey data, 16.0 percent.[9] In what follows, we hold the poverty rate constant across measures, as we did in Table 1.

---

[8] This process draws from the social indicator literature. A version of this line of work looks at "social inclusion" (Atkinson, Cantillon, Marlier, and Nolan 2002), which, in practice, is taken to include material well-being, education, health, housing, labor market outcomes, and the ability to participate in society. An even broader set of measures is argued for in Stiglitz, Sen, and Fitoussi (2009), which includes social connections and relationships, the environment, and physical and economic insecurity. While these multidimensional approaches offer certain advantages, an evaluation of this much broader set of indicators is beyond the scope of this paper.

[9] The estimates of the Supplemental Poverty Measure differ due to small definitional differences. For example, the estimate based on the Consumer Expenditure Survey does not include some noncash benefits—WIC (the Special Supplemental Nutrition Program for Women, Infants and Children), school lunch subsidies, and energy assistance—because receipt of these benefits is not observed in this survey.

Mean Characteristics of the Official and Supplemental Poverty Measure (SPM) Poor by Poverty Status, Consumer Expenditure Survey, 2010

	Both SPM poor and official poor (1)	*SPM poor only* (2)	*Official poor only* (3)	*Neither SPM nor official poor* (4)	*+ favors SPM*
Consumption	$27,159	$37,030	$25,799	$51,699	–
Any health insurance	61%	68%	65%	78%	–
Private health insurance	28%	55%	20%	70%	–
Homeowner	37%	55%	36%	76%	–
Single family home	28%	46%	24%	66%	–
Own a car	71%	89%	78%	94%	–
Service flows from vehicles	$415	$849	$330	$1,363	–
Service flows from owned homes	$2,099	$3,809	$1,594	$6,380	–
Total service flows	$2,514	$4,658	$1,924	$7,743	–
Family size	3.582	3.205	4.268	3.387	–
# of rooms	6.19	6.92	5.57	7.58	–
# of bedrooms	3.08	3.31	2.76	3.59	–
# of bathrooms	1.68	1.94	1.48	2.15	–
Appliances and amenities					
Microwave	92%	95%	93%	97%	–
Disposal	33%	44%	33%	57%	–
Dishwasher	40%	57%	42%	75%	–
Any air conditioning	73%	82%	77%	83%	–
Central air conditioning	47%	58%	51%	67%	–
Washer	70%	82%	70%	90%	–
Dryer	63%	79%	62%	88%	–
Television	96%	97%	95%	98%	–
Computer	63%	69%	63%	88%	–
Head is a college graduate	9%	14%	7%	34%	
Total financial assets					
75th percentile	$300	$3,000	$200	$14,000	–
90th percentile	$2,502	$20,000	$1,400	$97,000	–
Share of people	13%	3%	3%	80%	
Unweighted number of families	4,085	1,000	808	22,322	

Notes: Official income poverty is anchored at the SPM poverty rate for this sample, 16.5 percent. Official poverty is calculated using the official scale and pretax money income. The sample includes all families in the Consumer Expenditure Survey. Rooms and total consumption are equivalence-scale adjusted and set equal to a family with two adults and two children. All characteristics are for the family but are weighted by family size. Financial asset statistics come from samples of families in their fifth Consumer Expenditure Survey interview.

Comparing Characteristics of Those Added to or Removed from Poverty across Measures

In Table 2, we examine 25 indicators of well-being including consumption, health insurance coverage, home and car ownership, housing characteristics such as number of rooms, number of bathrooms, air conditioning, appliance ownership, education of head, and percentiles of total financial assets. The first column shows the characteristics for those identified as "poor" by both the official poverty measure and the Supplemental Poverty Measure. The second column shows characteristics of

those who would be added to poverty by using the Supplemental Poverty Measure, but who would not be counted as "poor" under the official measure. The third column shows the reverse: that is, the characteristics of those who would be counted as poor by the official measure, but not by the Supplemental Poverty Measure. Finally, the fourth column shows the characteristics of those who are not poor by either the official measure or the Supplemental Poverty Measure.

When comparing the Supplemental Poverty Measure and the official poverty measure, poverty status is classified differently for 6 percent of individuals. Quite strikingly, those added to poverty by the Supplemental Poverty Measure (column 2) appear to be better off than those removed (column 3) according to all 25 indicators. For example, those added to poverty are: consuming nearly 50 percent more; 3 percentage points more likely to be covered by health insurance and 34 percentage points more likely to be covered by private health insurance; 19 percentage points more likely to be a homeowner; 11 percentage points more likely to own a car; living in a house or apartment with nearly 1.4 more rooms; twice as likely to be in a family headed by a college graduate; and wealthier, with more than ten times the assets at the 75th or 90th percentiles (assets are generally zero at lower percentiles). All nine types of appliances or amenities we consider are more common among those added to poverty, even though these families are on average much smaller. In an online Appendix available with this paper at ⟨http://e-jep.org⟩, we present results from the Current Population Survey that are very similar to the results from the Consumer Expenditure Survey reported in Table 2.[10]

In the same spirit, Table 3 compares consumption poverty to the official poverty measure by looking at who it adds to and removes from poverty in 2010. For this comparison, a much larger fraction of individuals, 16 percent, are classified differently. Those added to poverty by switching to a consumption measure appear to be worse off than those removed for 21 out of 25 indicators. For example, compared to those subtracted from poverty, those added to poverty are: consuming about half as much; 10 percentage points less likely to be covered by health insurance, but slightly more likely to be covered by private health insurance; 3 percentage points less likely to be a homeowner; owning cars with half the value (though slightly more likely to own a car at all); living in homes with about two fewer rooms; 3 percentage points less likely to be in a family headed by a college graduate; and similar in terms of financial assets. Eight of the nine types of appliances or amenities we assess are less common among those added to poverty even though these families are on average much bigger. While the consumption poor will have lower consumption by construction, the full set of indicators overwhelmingly show that the consumption poor are

---

[10] The similarity of the results across these two data sources is striking, especially given that we are looking at a subtle feature of the data that we can only examine after cross-tabulating poverty calculated two different ways in the different datasets. Among the 17 indicators available in both datasets, only two indicators in the Current Population Survey have a different sign for the difference between those added and subtracted from poverty than in the Consumer Expenditure Survey. Most of the magnitudes are similar as well. These results confirm that our main results are unlikely to be due to something unique to the Consumer Expenditure Survey.

Table 3

Mean Characteristics of the Official and Consumption Poor by Poverty Status, Consumer Expenditure Survey, 2010

	Both consumption poor and official poor (1)	*Consumption poor only* (2)	*Official poor only* (3)	*Neither consumption nor official poor* (4)	*+ favors consumption measure*
Consumption	$17,068	$18,956	$36,959	$54,593	+
Any health insurance	59%	55%	65%	80%	+
Private health insurance	20%	35%	34%	73%	–
Homeowner	26%	45%	48%	78%	+
Single family home	17%	36%	38%	68%	+
Own a car	65%	83%	80%	95%	–
Service flows from vehicles	$194	$362	$607	$1,449	+
Service flows from owned homes	$666	$1,368	$3,364	$6,808	+
Total service flows	$859	$1,730	$3,971	$8,257	+
Family size	4.320	4.696	3.103	3.237	+
# of rooms	5.08	5.09	7.04	7.82	+
# of bedrooms	2.61	2.58	3.41	3.69	+
# of bathrooms	1.31	1.36	1.96	2.23	+
Appliances and amenities					
Microwave	90%	92%	95%	98%	+
Disposal	26%	35%	40%	58%	+
Dishwasher	31%	40%	50%	78%	+
Any air conditioning	71%	73%	77%	84%	+
Central air conditioning	42%	48%	53%	69%	+
Washer	65%	77%	75%	91%	–
Dryer	55%	68%	72%	90%	+
Television	95%	94%	97%	99%	+
Computer	56%	66%	70%	90%	+
Head is a college graduate	4%	10%	13%	36%	+
Total financial assets					
75th percentile	$100	$800	$700	$16,025	–
90th percentile	$800	$3,600	$4,200	$109,000	+
Share of people	8%	8%	8%	75%	
Unweighted number of families	2,072	1,632	2,821	21,690	

Notes: Both measures are anchored at the Supplementary Poverty Measure (SPM) poverty rate for this sample, 16.5 percent. Consumption poverty is calculated using the three-parameter equivalence scale. Official poverty is calculated using the official scale and pretax money income. The sample includes all families in the Consumer Expenditure Survey. Rooms and total consumption are equivalence-scale adjusted and set equal to a family with two adults and two children. All characteristics are for the family but are weighted by family size. Financial asset statistics come from samples of families in their fifth Consumer Expenditure Survey interview.

worse off along many dimensions than the official poor as defined by income.[11] Fisher, Johnson, Marchand, Smeeding, and Torrey (2009) find a similar result

---

[11] We verify that the results in Tables 2 and 3 are not unique to 2010. In the online Appendix we provide versions of Tables 2 and 3 for a pooled sample from 2004–2010. The results for this much larger sample are very similar to those reported here.

for the assets of the elderly—comparing consumption poverty to income poverty for people age 65 to 74, they show that median assets for those who are income-poor but not consumption-poor are nearly nine times greater than median assets for the consumption-poor but not income-poor.

We also examine how the characteristics of those in deep poverty—having resources below half the poverty line—differ across our three poverty measures. Specifically, we conducted analyses similar to those in Tables 1, 2, and 3, but fix the poverty rates at 5.4 percent rather than 16.5 percent. We chose 5.4 percent because that is the Supplemental Poverty Measure deep poverty rate in 2010 based on Consumer Expenditure Survey data. In general, the results for deep poverty, which are in an online Appendix available with this paper at ⟨http://e-jep.org⟩, are very similar to those discussed above: compared to the official measure, individuals added to deep poverty by the Supplemental Poverty Measure appear better off than those subtracted based on all 25 indicators—consumption for those added to deep poverty is nearly double that for those subtracted from deep poverty. In addition, compared to the official measure, those added to deep poverty by a consumption-based measure appear worse off than those subtracted for all but three indicators. In fact, using the Supplemental Poverty Measure, those below 50 percent of the poverty line appear better off than the larger group below 100 percent of the poverty line. This finding is consistent with other research that has shown that many families with extremely low reported income in surveys are actually well off in consumption terms, suggesting significant underreporting of income for these families (Meyer and Sullivan 2011).

Decomposing Differences between the Measures

Table 4 decomposes differences between measures to isolate the effects of the components of the change from one poverty measure to another. The decomposition allows us to isolate the extent to which the differences in characteristics reported above are a result of changing the equivalence scale or the resource measure, or varying the thresholds by housing status.

Row 1 reports average consumption, education of the head of the family, share of family covered by health insurance, and number of rooms in home for those classified as "poor" based on the official measure of poverty (using pretax money income, the official poverty thresholds, and the equivalence scale implicit in these thresholds), but fixing the baseline poverty rate at the estimated Supplemental Poverty Measure rate in 2010 in the Consumer Expenditure Survey (16.5 percent). As in Tables 1–3, for all the poverty measures reported in Table 4, we fix the poverty rate at 16.5 percent so the same number of people are considered poor regardless of how poverty is measured. The means in row 1 are also reported in Table 1.

Row 1a of Table 4 indicates how the switch from the (implicit) equivalence scale used in the official poverty measure to the three-parameter equivalence scale used in the Supplemental Poverty Measure affects the mean characteristics of those designated as poor. For example, mean consumption is $287 lower for those labeled as poor using the three-parameter equivalence scale in the Supplemental Poverty

Table 4

Decomposition of Differences in Poverty Measures as Captured by their Effects on the Mean Characteristics of the Poor in 2010

		Mean for the poor		
	Consumption	Education of head at least a high school degree (percent)	Share covered by health insurance (percent)	Number of rooms in home
1) Official poverty: official scale and resources, single tenure threshold	**$26,886**	**65.7**	**61.9**	**6.1**
1.a Official scale to 3-parameter SPM Scale	−$287	−0.8	−0.2	−0.1
1.b Official resources (pretax money income) to SPM Resources	$2,844	2.2	1.5	0.5
1.b.i Pretax to after-tax income	$531	0.1	−0.4	0.2
1.b.ii Add non-cash benefits to income	$356	0.2	−1.4	0.1
1.b.iii Subtract from income child care, work expenses, and child support paid	−$25	0.7	−0.6	−0.1
1.b.iv Subtract from income medical out-of-pocket spending	$1,982	1.2	3.9	0.4
1.c Single threshold to ones that vary by housing tenure	−$303	0.3	−0.7	−0.1
2) SPM poverty: SPM scale and resources, thresholds vary with tenure	**$29,140**	**67.3**	**62.5**	**6.3**
2.a Thresholds that vary by housing tenure to single threshold	$303	−0.3	0.7	0.1
2.b SPM resources to consumption as resources	−$11,444	−7.5	−6.1	−1.4
3) Consumption poverty: SPM scale, consumption as resources	**$18,000**	**59.5**	**57.1**	**5.1**
4) SPM poverty – Official poverty	$2,254	1.7	0.6	0.3
5) SPM poverty – Consumption poverty	$11,141	7.8	5.5	1.3
6) Consumption poverty – Official poverty	−$8,886	−6.1	−4.9	−1.0

Notes: All data are from the Consumer Expenditure Survey. Bolded rows are from Table 1. The other rows denote how much the mean for each characteristic for the poor changes as a result of changing one component of a poverty measure. For example, $2,844 in the first column indicates that mean consumption is $2,844 higher for those labeled as poor using the SPM definition of resources as compared to those labeled as poor using the official poverty measure's definition of resources. All poverty measures are anchored to the SPM rate in 2010, so that the fraction poor for each measure is 16.5 percent.

Measure as compared to the poor using the scale implicit in the official measure. All of the changes in row 1a are negative, which indicates that using the three-parameter equivalence scale results in those classified as poor being more deprived, suggesting that this step leads to a more accurate identification of the disadvantaged.

Row 1b shows how changing from the measure of income used for official poverty (pretax money income) to the measure of income used in the Supplemental Poverty Measure affects the average characteristics of those designated as poor. For example, mean consumption is $2,844 higher for those labeled as poor using the Supplemental Poverty Measure definition of resources as compared to those designated as poor using the official poverty measure's definition of resources. Overall, the Supplemental Poverty Measure resource definition does poorly—all entries in row 1b are positive and most are substantial. The change in the resource definition leads the average poor person to have more than 10 percent higher consumption, to live with a head who is 2 percentage points more likely to have a high school degree, and to live in a home with 0.5 more rooms. This happens despite a small share of the population having their classification changed.

To determine why the Supplemental Poverty Measure resource definition does so poorly, we look at how mean characteristics of the poor change as we move, in steps, from the official poverty measure's definition of resources to the Supplemental Poverty Measure definition of resources. This breakdown shows that the change from pretax to after-tax income and the addition of noncash benefits to income have counterproductive or mixed effects as seen by the mostly positive signs in rows 1bi and 1bii. Accounting for child care, work expenses, and child support payments has the desired (but small) effect as indicated by the mostly negative entries in row 1biii. The biggest impact comes from the subtraction of out-of-pocket medical spending from income (row 1biv). This subtraction raises average consumption among the poor by $1,982, accounting for more than two-thirds of the rise in mean consumption of the poor when moving from the official poverty measure's definition of resources to the Supplemental Poverty Measure definition of resources.

It is troubling that this change has such a large impact, because subtracting out-of-pocket medical spending is probably the most controversial of these adjustments on a priori grounds. On the one hand, large out-of-pocket medical expenses resulting from poor health can drain family resources. On the other hand, these expenses can arise because families choose to allocate resources towards health, purchasing expensive health insurance or electing to have procedures that are not fully covered by insurance. It is difficult a priori to determine whether most out-of-pocket medical spending reflects those with lower health status or those who have greater resources and make choices to spend more on out-of-pocket health care costs. While our analysis does not directly address the connection between health status and health spending, our findings point out that when out-of-pocket medical expenses are subtracted from income to calculate poverty, those identified as "poor" have higher consumption, more education, more rooms in their home and are more likely to be covered by health insurance. This pattern is consistent with a belief that many families with large medical out-of-pocket expenses have the resources to support such spending, and they are making a choice to spend as much as they do on medical care. The importance of this issue, and its substantial impact on who is defined as poor, suggests a need for more research on the relationship between health spending and health status.

Another perhaps surprising result that runs contrary to long-held beliefs among poverty researchers is that when the Supplemental Poverty Measure accounts for noncash benefits and taxes, it is designating a better-off group as poor. Conceptually, in a world without defects in data and measurement, there is a strong argument for including noncash benefits and taxes in the measure of income. However, as already noted, one of the largest noncash benefits, food stamps, is more likely to be omitted than reported by a recipient in the Current Population Survey. In addition, taxes are typically imputed in surveys.[12] In the Current Population Survey, even when 100 percent take-up of the Earned Income Tax Credit is assumed, the imputed dollars amount to only two-thirds of what the IRS actually pays out to the working poor for some large demographic groups such as single parents. Given that those most likely to take up government benefits such as food stamps and Temporary Assistance for Needy Families are those who are in greatest need (Blank and Ruggles 1996) and those most likely to report them are the worst-off recipients (Meyer and Goerge 2011), it may be that accounting for the benefits may remove from the poverty count those who are among the worst off, distorting the ability of the measure to identify the disadvantaged. Similarly, tax credits may be particularly well targeted to the disadvantaged, leading to a situation where the credits are accounted for but other sources of income are not, so that those raised above the poverty line by tax credits are in fact more needy than those who are left behind. Providing firmer answers to the puzzle of why the after-tax and noncash transfer income adjustment performs so poorly should be a high priority.

A fundamental problem with income-based poverty measures is that income misses the rental value of homeownership. Someone who owns a home outright receives a flow of services and does not have to pay high housing expenses. Shelter expenses are by far the largest expenditure for most families, and this share has been rising over time—in 2008 they accounted for about 36 percent of expenditures in the bottom income quintile, up from about 28 percent in 1980. The Supplemental Poverty Measure attempts to address this problem by setting different thresholds by three housing status groups: homeowners with a mortgage, homeowners without a mortgage, and renters. Row 1c shows the effect of specifying different thresholds by housing status. The change in characteristics supports this step; adjusting thresholds by housing status results in a group designated as "poor" that has slightly lower consumption and is slightly less likely to be covered by health insurance. However, this adjustment is only a partial solution to the problem that income misses the value of homeownership. The split of households by housing status only accounts for about 25 percent of the actual variation in housing costs, based on our own regressions of housing expenses on indicators for housing status. The Supplemental Poverty Measure treats as the same a small mortgage payment on a loan taken out 25 years ago and a large payment on one taken out in the last year.

---

[12] Taxes are imputed in the Current Population Survey. We impute taxes when using the Consumer Expenditure Survey because the tax information that is collected appears to be significantly underreported.

This last step completes the transition from the official poverty measure to the Supplemental Poverty Measure. Taken together the Supplemental Poverty Measure performs worse than the official measure, in the sense that all four indicators have higher values for the Supplemental Poverty Measure than official poverty (as shown in row 4).

Two important components of the Supplemental Poverty Measure are not addressed in Table 4: the effect of changing from a resource-sharing unit based on those related by blood or marriage to one that includes cohabitors and other individuals who may be sharing resources; and the effect of adjusting thresholds for geographic variation in prices. In separate analyses, we examine the impact of these two changes using the Current Population Survey, which has a more limited set of indicators of well-being. These results, which are in an online Appendix available with this paper at ⟨http://e-jep.org⟩, are mixed. For example, moving to the Supplemental Poverty Measure unit slightly increases the likelihood that the poor are covered by health insurance while it slightly decreases the fraction living with heads with at least a high school degree. Adjusting the thresholds for geographic price variation decreases the likelihood that the poor are covered by health insurance and the fraction living with heads with at least a high school degree, but both of these changes are small.

The remaining part of Table 4 takes us from the Supplemental Poverty Measure to our consumption-based measure of poverty in two steps. First, we undo the step that adjusts the thresholds by housing status. We then shift from the income-based measure of resources used in the Supplemental Poverty Measure to a consumption-based measure. This change lowers the average characteristics of those designated as poor significantly (row 2b). Not surprisingly, average consumption is substantially lower (39 percent) for the consumption poor. But the other characteristics also indicate greater deprivation for the consumption poor: the family head is 7.5 percentage points less likely to have a high school degree; the family is 6.1 percentage points less likely to be covered by health insurance; and their homes have 1.4 fewer rooms.

Capturing Differences in Well-Being across Age Groups

One of the most noticeable differences between the Supplemental Poverty Measure and the official measure is that poverty rates by age change sharply. In 2010, the official poverty rate for children was 22.5 percent while the Supplemental Poverty Measure rate was 18.2 percent. For those 65 or older, the official poverty rate was 9 percent while the Supplemental Poverty Measure rate was 15.9 percent. A range of other evidence shows that the economic circumstances of the elderly are better (and the poverty rate is much lower) than that of other groups, which is inconsistent with the estimates of who is poor from the Supplemental Poverty Measure.

The major reason for these differences by age traces back to the subtraction of medical out-of-pocket expenses from income when calculating the Supplemental Poverty Measure. Short (2011) reports that subtracting medical out-of-pocket

expenses raises overall poverty by 3.3 percentage points, while no other incremental change has more than a 1.9 percentage point effect. This adjustment dispropor-tionately affects the elderly; subtracting medical out-of-pocket expenses raises their poverty rate from 8.5 percent to 15.5 percent, nearly doubling it.

Further complicating income-based poverty measures for the elderly is the fact that these measures will understate the well-being of elderly Americans, because older Americans are more likely to be spending out of savings and using assets (like homes and cars) that they own. In the 2000s, two-thirds of those in the bottom income quintile of the elderly owned a home; conversely, for the bottom income quintile of children, 35 percent lived in an owned home. The elderly as a group also have considerably more assets than those in the bottom income quintile for other groups. In recent years, the financial assets of the low-income elderly were 19 times greater than those for children in low-income families, and 3.5 times greater than those of low-income nonelderly adults. Income surveys such as the Current Popula-tion Survey also seem to have difficulty in capturing retirement income sources. For example, in 2006, of $125 billion in taxable IRA withdrawals, $6 billion was reported in the Current Population Survey (Investment Company Institute 2009).

Our own calculations, using a consumption-based measure of poverty, find that those 65 and older have much lower poverty rates than most other demographic groups and that these rates have fallen sharply over time: over the past three decades elderly poverty has fallen by more than 60 percent, while child poverty has fallen by about 25 percent (Meyer and Sullivan 2012). Aguiar and Hurst (2005) argue that even consumption may understate the well-being of the aged, because the prices that the elderly pay are lower than what others pay. In addition, while our consump-tion measures capture the largest durables (vehicles and homes), the stock of other durables such as furniture and appliances owned by the elderly is greater than that of others, providing a flow of resources that exceeds that of other age groups.

To examine the possible effects of age on poverty measures, we re-did the calcu-lations behind Tables 1–4 separately for children, nonelderly adults, and the elderly. Our general results continue to hold: that is, when classifying by age group the Supplemental Poverty Measure typically identifies as "poor" people who are better off by the characteristics we look at compared to the official poverty measure, while a consumption-based poverty measure typically identifies as "poor," people who are worse off by the characteristics we look at compared to the official poverty measure. Again, the detailed results are available in an online Appendix available with this paper at ⟨http://e-jep.org⟩.

Changes in Measures of Poverty over Time

How one adjusts poverty thresholds over time will determine how a poverty measure assesses changes in disadvantage over time. Recall that assessing changes in poverty over time was one of two main goals for a poverty measure. One needs to decide whether the poverty thresholds should be absolute cutoffs or be relative

to some standard. With an absolute poverty measure the thresholds are adjusted for inflation, so that the real value of the thresholds remains unchanged over time. With a relative poverty measure, the real value of the thresholds can rise or fall over time. An absolute measure of poverty is particularly useful for understanding changes in the material circumstances of the population or for evaluating policy changes that aim to reduce the number of people with very few resources. However, an important concern with an absolute measure is that societal views on what it means to be poor change, particularly over longer periods. Goods that are viewed as luxuries for one generation, such as televisions or cars, may be viewed as necessities by future generations. Even some of those involved in President Johnson's War on Poverty who advocated an absolute measure of poverty acknowledged that antipoverty goals should be updated, albeit infrequently, to reflect rising living standards (Lampman 1971).

Relative poverty measures provide another way of characterizing the extent of deprivation in a population. The most common type of relative poverty measure sets the thresholds as a given percentage of median income or consumption. For example, the European Union focuses on a measure of poverty defined as the fraction of the population below 60 percent of median income. However, relative poverty measures have a number of important limitations. A relative measure keeps adjusting the standard for overcoming poverty, which makes understanding what the poverty measure captures much more difficult. This characteristic is particularly problematic for evaluating policy. Antipoverty policies that affect incomes around the median as well as at the bottom might reduce the extent of deprivation but have no impact on a poverty measure defined relative to median income. As one example, Ireland grew rapidly in recent years with real growth in incomes throughout the income distribution, including the bottom. However, because the middle grew a bit faster than the bottom, a relative poverty measure shows an increase in poverty while an absolute measure shows a sharp decrease in poverty (Nolan, Munzi, and Smeeding 2005). Another troubling example occurs during a recession in which median income or consumption falls. With a recent period of falling officially measured median income in the United States, we could have relative poverty falling despite a decline in incomes at low percentiles.

How Well Do the Three Poverty Measures Assess Changes in Disadvantage over Time?

The official measure of poverty is often advertised as an absolute measure, but this characterization is not quite right, because the poverty lines are adjusted upwards over time to account for inflation using the Consumer Price Index, which overstates the true rise in the cost of living. The price index has this bias because it does not take into account sufficiently the arrival of new goods in the market, quality improvements in existing goods, and possibilities for substitution between goods. In Meyer and Sullivan (2012), we provide an extensive discussion of the

evidence for and implications of the overstatement of inflation in setting the official poverty thresholds.

This bias has a considerable effect on changes in the poverty rate over time. Between 1980 and 2010, the official poverty rate rose by 2 percentage points. If one corrects for the overstatement in inflation, however, the poverty rate would have fallen by more than 2 percentage points. If, in addition, poverty is calculated using income that more closely approximates resources available for consumption, then the poverty rate would have fallen by more than 5 percentage points over the past three decades. If a consumption-based measure of poverty was used, the decline would have been more than 8 percentage points.[13] Clearly, how one measures poverty has a considerable impact on our understanding of how poverty has changed over time.

The Supplemental Poverty measure is not a pure absolute measure of poverty, because the value of the poverty thresholds will change in real terms over time. It is also not a pure relative measure of poverty, because the value of the poverty thresholds do not change one-for-one with a change in a point in the distribution of income (like the median). As a result, interpreting changes in the poverty rate as calculated by the Supplemental Poverty Measure will be challenging.

For example, in a deep recession during which the 30th to 36th percentiles of spending on food, clothing, shelter and utilities fall, the poverty rate as calculated by the Supplemental Poverty Measure indicate estimate that poverty fell, even at a time when absolute deprivation rose. Likewise, if we were to expand programs that provide for those around the 33rd percentile of the distribution of spending (or cut the rates in the lowest income tax brackets), then the rise in incomes for those around the 33rd percentile would lead to higher poverty thresholds—and likely lead to a conclusion that these policies raised the poverty rate. It will be unclear whether changes in the poverty rate generated by the Supplemental Poverty Measure are due to family incomes changing or the thresholds changing, making it difficult to determine whether antipoverty policies are effective at reducing deprivation.

As an illustration of this point, we use data from the Consumer Expenditure Survey, along with the Supplemental Poverty Measure definition of poverty, to create a data series of what the changes in poverty thresholds would have looked like. In Table 5, we report the level and decadal changes, adjusted for inflation, in the Supplemental Poverty Measure thresholds along with several benchmark series. We find that the changes in the Supplemental Poverty Measure thresholds have been very different than the changes in other benchmarks of well-being like changes in median consumption, expenditures, or after-tax and transfer income. For example, in the 1980s, the Supplemental Poverty Measure thresholds would have risen by

Table 5

Official and Supplemental Poverty Measure Thresholds and Median Consumption and Income, 1980–2010

	Official poverty thresholds (1)	*SPM thresholds* (2)	*Median consumption* (3)	*Median expenditures* (4)	*Median after-tax income plus noncash benefits* (5)
1980	$16,567	$16,793	$30,218	$31,399	$40,129
1985	$17,328	$16,879	$32,139	$34,299	$42,620
1990	$18,327	$18,095	$33,260	$35,861	$47,014
1995	$19,519	$19,139	$34,420	$37,180	$49,050
2000	$20,441	$20,725	$37,887	$39,830	$57,793
2005	$21,268	$22,202	$41,288	$44,418	$58,005
2010	$22,113	$24,457	$39,993	$43,197	$54,540
% Change: 1980–1990	10.6%	7.8%	10.1%	14.2%	17.2%
% Change: 1990–2000	11.5%	14.5%	13.9%	11.1%	22.9%
% Change: 2000–2010	10.0%	18.0%	5.6%	8.5%	−5.6%
% Change: 1980–2010	35.8%	45.6%	32.3%	37.6%	35.9%

Notes: All numbers are in 2010 dollars using the adjusted CPI-U-RS price index from Meyer and Sullivan (2012). Official thresholds are those reported by the U.S. Census Bureau for a family with two adults and two children. The Supplemental Poverty Measure (SPM) thresholds are for a family with two adults and two children. Consumption and income are equivalence-scale adjusted using the three-parameter scale, and set equal to a family with two adults and two children. Columns 2–4 are calculated using Consumer Expenditure Survey data, while column 5 is calculated using the Current Population Survey. Resources are measured at the family level but individual weighted. Income includes all money income less tax liabilities plus tax credits, food stamps, and CPS-imputed measures of housing and school lunch subsidies.

7.8 percent, while median after-tax income plus noncash benefits rose 17.2 percent. In the 2000s, on the other hand, while the Supplemental Poverty Measure thresholds would have risen 18 percent, median after-tax income plus noncash benefits fell by 5.6 percent. In short, it is difficult to get an intuitive sense of exactly what any change in the Supplemental Poverty Measure would capture.

Conclusion: Goals for a Poverty Measure

Constructing a measure of deprivation is inherently difficult. The Census Bureau's new Supplemental Poverty Measure, released for the first time last fall, has some conceptual advantages over the official poverty measure, including a more defensible adjustment for family size and composition, an expanded definition of the family unit that includes cohabitors, and a definition of income that is conceptually closer to resources available for consumption. However, when we compare those added to and dropped from the poverty rolls by the alternatives to the current official measure, we find that the Supplemental Poverty Measure adds to poverty

individuals who have higher consumption levels and are more likely to be college graduates; to own a home and a car; to live in a larger housing unit; and to have other more favorable characteristics than those who are dropped from poverty. On the other hand, we find that a consumption-based poverty measure compared to either official poverty or the Supplemental Poverty Measure adds to the poverty rolls individuals who are more disadvantaged than those who are dropped. Even if the Supplemental Poverty Measure did not subtract out-of-pocket medical spending from income, it would perform slightly worse than the official measure, and much worse than a consumption-based measure of poverty, in terms of identifying the disadvantaged. Our results present strong evidence that a well-constructed consumption-based poverty measure would be preferable to income-based measures of poverty, like the official income measure and the Supplemental Poverty Measure, for determining the most disadvantaged.

We have also discussed how a poverty measure captures changes in disadvantage over time due to public policies and social and economic change. The official poverty resource measure that misses taxes and in-kind transfers is clearly ill-suited to analyze program effects. However, the Supplemental Poverty Measure resource measure may not perform well, either. Of particular concern is the high and sharply increasing rate of underreporting of government transfers in the Current Population Survey. Furthermore, because the Supplemental Poverty Measure poverty thresholds change in an opaque and unintuitive way over time, it will be hard to determine if changes in poverty are due to changes in income or changes in thresholds. In comparison, consumption-based poverty measures with thresholds that are periodically revised in real terms could have many of the advantages of the Supplemental Poverty Measure, but fewer disadvantages.

We have focused in this paper on the use of a poverty measure to determine who is disadvantaged at a point in time and over time, but there are other uses for a poverty measure. Given the limits on data, a consumption-based measure of poverty will work better for some of these uses than others. For example, the current sample sizes in the Consumer Expenditure Survey are not sufficient for useful comparisons across states or localities. Also, while a consumption-based measure of poverty may be used to set overall standards for program eligibility, individual consumption data are not suitable for determining eligibility for antipoverty programs. Given that at least some components of income, such as formal earnings and transfer income, are easier to collect and validate, income will typically be more appropriate for determining program eligibility for individuals or families.

Our results raise the question as to whether income, even when modified to be conceptually closer to consumption, can reliably be used to measure well-being for the most disadvantaged. Our results also suggest that some largely untested but common presumptions may turn out to be wrong. For example, many researchers have argued that income after accounting for taxes and noncash benefits more closely reflects material well-being than pretax money income. While this may be true conceptually, in practice accounting for taxes and noncash benefits may not help if they are imprecisely measured in income data sources.

References

Aguiar, Mark, and Erik Hurst. 2005. "Consumption versus Expenditure." *Journal of Political Economy* 113(5): 919–48.

Atkinson, Tony, Bea Cantillon, Eric Marlier, and Brian Nolan. 2002. *Social Indicators: The EU and Social Inclusion.* Oxford: Oxford University Press.

Bee, Adam, Bruce D. Meyer, and James X. Sullivan. Forthcoming. "Micro and Macro Validation of the Consumer Expenditure Survey." In *Improving the Measurement of Consumer Expenditures,* edited by C. Carroll, T. Crossley, and J. Sabelhaus. University of Chicago Press.

Blank, Rebecca. 2008. "Presidential Address: How to Improve Poverty Measurement in the United States." *Journal of Policy Analysis and Management* 27(2): 233–54.

Blank, Rebecca, and Mark Greenberg. 2008. "Improving the Measurement of Poverty." Discussion Paper 2008-17, Brookings Institution and the Hamilton Project. http://www.brookings.edu/~/media/research/files/papers/2008/12/poverty%20measurement%20blank/12_poverty_measurement_blank.pdf.

Blank, Rebecca M., and Patricia Ruggles. 1996. "When Do Women Use AFDC & Food Stamps? The Dynamics of Eligibility vs. Participation." *Journal of Human Resources* 31(1): 57–89.

Chung, Yiyoon, Julia B. Isaacs, Timothy M. Smeeding, and Katherine A. Thornton. 2012. "Wisconsin Poverty Report: How the Safety Net Protected Families from Poverty in 2010." The Fourth Annual Report of the Wisconsin Poverty Project. Institute for Research on Poverty Working Paper, University of Wisconsin-Madison. http://www.irp.wisc.edu/research/WisconsinPoverty/pdfs/WIPovSafetyNet_Apr2012.pdf.

Citro, Constance F., and Robert T. Michael, eds. 1995. *Measuring Poverty: A New Approach.* Washington, D.C.: National Academy Press.

Cutler, David M., and Lawrence F. Katz. 1991. "Macroeconomic Performance and the Disadvantaged." *Brookings Papers on Economic Activity* no. 2, pp. 1–74.

Fisher, Jonathan D., David S. Johnson, Joseph T. Marchand, Timothy M. Smeeding, and Barbara B. Torrey. 2009. "Identifying the Poorest Older Americans." *Journal of Gerontology: Social Sciences* 64B(6): 758–66.

Garner, Thesia. 2010. "Supplemental Poverty Measure Thresholds: Laying the Foundation." Paper prepared for the Allied Social Science Associations Annual Meeting, Denver, CO. http://www.bls.gov/pir/spm/spm_pap_thres_foundations10.pdf.

Garner, Thesia, and Charles Hokayem. 2011. "Supplemental Poverty Measure Thresholds: Imputing Noncash Benefits to the Consumer Expenditure Survey Using Current Population Survey—Parts I and II." September. Posted at the Bureau of Labor Statistics website.

Hutto, Nathan, Jane Waldfogel, Neerej Kaushal, and Irwin Garfinkel. 2011. "Improving the Measurement of Poverty." *Social Science Review* 85(1): 39–52.

Iceland, John (*Rapporteur***, Planning Group for the Workshop to Assess the Current Status of Actions Taken in Response to** *Measuring Poverty: A New Approach***).** 2005. *Experimental Poverty Measures: Summary of a Workshop.* National Research Council Washington, DC: National Academy Press.

Investment Company Institute. 2009. "The Evolving Role of IRAs in U.S. Retirement Planning." *Research Perspective* vol. 15, no. 3.

Jefferson, Philip. Forthcoming. "A New

Statistic: The U.S. Census Bureau's Supplemental Poverty Measure." *The Oxford Handbook of the Economics of Poverty,* edited by Philip N. Jefferson. Oxford: Oxford University Press.

Lampman, Robert J. 1971. *Ends and Means of Reducing Income Poverty.* Chicago: Markham.

Levitan, Mark, Christine D'Onofrio, John Krampner, Daniel Scheer, and Todd Seidel. 2010. *The CEO Poverty Measure, 2005–2008.* A Working Paper by the NYC Center for Economic Opportunity, March. http://www.nyc.gov/html/ceo /downloads/pdf/ceo_poverty_measure_v5.pdf.

Meyer, Bruce D., and Robert Goerge. 2011. "Errors in Survey Reporting and Imputation and their Effects on Estimates of Food Stamp Program Participation." October 5. http://harrisschool .uchicago.edu/faculty/web-pages/10-2011%20 -%20Food%20Stamp%20Survey%20Error.pdf.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2009. "Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences." NBER Working Paper 15181.

Meyer, Bruce D., and James X. Sullivan. 2003. "Measuring the Well-Being of the Poor Using Income and Consumption." *Journal of Human Resources* 38(S): 1180–1220.

Meyer, Bruce D., and James X. Sullivan. 2011. "Viewpoint: Further Results on Measuring the Well-Being of the Poor Using Income and Consumption." *Canadian Journal of Economics* 44(1): 52–87.

Meyer, Bruce D., and James X. Sullivan. 2012. "Five Decades of Consumption and Income Poverty." NBER Working Paper 14827. Revised February 2012.

Nolan, Brian, Teresa Munzi, and Timothy Smeeding. 2005. "Two Tales of Irish Poverty." Box 3 in "Note on Statistics in the Human Development Report" in the *Human Development Report 2005,* p. 334. New York: United Nations Human Development Office.

Poterba, James M. 1991. "Is the Gasoline Tax Regressive?" In *Tax Policy and the Economy,* vol. 5, edited by David Bradford, 145–64, Cambridge, MA: MIT Press.

Ruggles, Patricia. 1990. *Drawing the Line–Alternative Poverty Measures and Their Implications for Public Policy.* Washington, DC: The Urban Institute Press.

Short, Kathleen. 2011. "The Research Supplemental Poverty Measure: 2010." Current Population Report P60-241, U.S. Census Bureau. November. http://www.census.gov/prod/2011pubs/p60-241 .pdf.

Slesnick, Daniel T. 1993. "Gaining Ground: Poverty in the Postwar United States." *Journal of Political Economy* 101(1): 1–38.

Slesnick, Daniel T. 2001. *Consumption and Social Welfare: Living Standards and their Distribution in the United States.* Cambridge University Press.

Stiglitz, Joseph E., Amartya Sen, and Jean-Paul Fitousi. 2009. *Report by the Commission on Measurement of Economic Performance and Social Progress.* http://www.stiglitz-sen-fitoussi.fr/documents /rapport_anglais.pdf.

U.S. Census Bureau. 2010. "Poverty—Experimental Measures." http://www.census.gov/hhes /povmeas/publications/index.html.

Zedlewski, Sheila R., Linda Giannarelli, Laura Wheaton, and Joyce Morton. 2010. "Measuring Poverty at the State Level." Low-Income Working Families, Paper 17, Urban Institute.

The New Demographic Transition: Most Gains in Life Expectancy Now Realized Late in Life[†]

Karen N. Eggleston and Victor R. Fuchs

The original "demographic transition" describes a process that began in Europe by the early 1800s with decreases in mortality followed, usually after a lag, by decreases in fertility (Davis 1945; for an overview in this journal, see Lee 2003). According to Lee and Reher (2011, p. 1), "this historical process ranks as one of the most important changes affecting human society in the past half millennium." The increase in life expectancy associated with this demographic transition has been accompanied by rising levels of per capita output, which have in turn spurred further improvements in population health through better nutrition and living standards (Fogel 1994; Barker 1990) and, especially since World War II, through advances in medical care (in this journal, Cutler, Deaton, and Lleras-Muney 2006). At the same time, increases in life expectancy have resulted in a higher proportion of each cohort living long enough to participate in the production of goods and services. Reductions in fertility are also closely linked to higher labor force participation rates among women (Galor and Weil 1996; Costa 2000; Guinnane 2011).

During the original demographic transition, mortality decline prior to fertility decline often led to larger cohorts concentrated in working ages; this transitional change in the age structure of the population provided a boost to income that has

■ *Karen N. Eggleston is Director of the Stanford Asia Health Policy Program and Center Fellow at the Shorenstein Asia-Pacific Research Center, and Victor R. Fuchs is Henry J. Kaiser, Jr., Professor Emeritus in the Departments of Economics and of Health Research and Policy, and Senior Fellow, Stanford Institute of Economic Policy and Research, both at Stanford University, Stanford, California. Their email addresses are ⟨karene@stanford.edu⟩ and ⟨vfuchs@stanford.edu⟩.*

been called a "demographic dividend" (Bloom, Canning, and Sevilla 2003). Swift (2011) documents a significant two-way positive relationship between life expectancy and GDP per capita between 1820 and 2001 for 13 high-income countries.

Now, the United States and many other countries are experiencing a new kind of demographic transition. Instead of additional years of life being realized early in the lifecycle, they are now being realized late in life. At the beginning of the twentieth century, in the United States and other countries at comparable stages of development, most of the additional years of life were realized in youth and working ages; and less than 20 percent was realized after age 65. Now, more than 75 percent of the gains in life expectancy are realized after 65—and that share is approaching 100 percent asymptotically. The choice of age 65 to illustrate this new demographic transition is somewhat arbitrary, but if we used 60 or 70 instead, the results would be qualitatively similar.

The new demographic transition is a *longevity* transition: How will individuals and societies respond to mortality decline when almost all of the decline will occur late in life? This issue is broader and more far-reaching than the issue of cohort size in each age group, with its usual focus on the prospective retirement of the unusually large "baby boomer" cohort, and has important socioeconomic implications independent of patterns of fertility.

When the gains in life expectancy occur mainly towards the end of life, they contribute more to the age bracket that is traditionally mostly retired rather than to the age bracket in prime working years. Retirees are highly dependent on transfers from the working population for living expenses, including large consumption of medical care. Thus, gains in life expectancy concentrated at the end of life can unsettle an economy's balance between production and consumption in ways that pose a long-run challenge for public policy. The obvious changes needed (at least "obvious" to many economists) would be to raise productivity, the savings rate, and the age of retirement, but how to accomplish such goals is controversial and uncertain.

This paper covers the years 1900–2007 for the United States and 16 other "developed countries," chosen for the continuity of their mortality data: Australia, Belgium, Canada, Denmark, England and Wales, Finland, France, Iceland, Italy, Netherlands, Northern Ireland, Norway, Scotland, Spain, Sweden, and Switzerland. We focus on demographic statistics including life expectancy at birth and at age 65, the percent of each birth cohort expected to survive to age 65, and the share of the increase in life expectancy at birth realized after age 65. For the U.S. economy, we also calculate expected labor force participation for each birth cohort, which allows us to investigate how changes in mortality affect labor force participation and work-life as a share of life expectancy. Results on the longevity transition and expected labor force participation for the United States and other high-income countries are followed by consideration of economic and social changes in China and other countries that are experiencing an earlier stage of the original demographic transition. The paper concludes with a brief discussion of the long-run implications of the new demographic transition.

The Longevity Transition

To examine long-term trends in life expectancy at birth, we draw upon the life tables in the Human Mortality Database, which offers high-quality demographic data for selected countries and regions compiled by a respected group of demographers at ⟨http://www.mortality.org⟩. We first extract data on life expectancy at birth; in particular, we calculate "period" life expectancy, which is the projected average age of death for a cohort if it experienced the age-specific death rates prevailing at the year of birth. We also look at rates of survival from birth to age 65 and life expectancy at age 65. We use the five-year period life tables since 1900 (or earliest available year) for each of the 17 countries or regions in the Human Mortality Database that have data extending back at least 70 years. The five-year intervals help to smooth annual fluctuations in demographic trends.

We calculate changes for nine overlapping 20-year intervals: 1907–1927, 1917–1937, and so on up to 1987–2007.[1] (The years ending in "7" are chosen to represent mid points of each of our five-year intervals.) To calculate the change in years lived past 65, we first multiply survival to 65 by life expectancy at age 65 for each five-year period and then take differences across 20-year intervals. Finally, we calculate the change in years lived past 65 as a percentage of change in life expectancy at birth for each country for each of the nine 20-year intervals.

Figure 1A shows that life expectancy at birth has increased almost continuously for well over a century in high-income countries. Much of this rise in life expectancy was due to a particularly large fall in death rates for infants, children, and young adults, resulting in a sharp rise in the percentage of a cohort surviving to age 65, as indicated in Figure 1B. Survival rates from birth to age 65 more than doubled over the twentieth century from 40.9 percent in 1900–04 to 83.3 percent in 2005–09 in the United States. Similarly, survival rates from birth to age 65 in 16 high-income comparators increased from 42.0 to 87.8 percent over the same period.

The other major demographic change that contributes to the longevity transition is an increase in life expectancy at age 65, an increase which has become larger in recent decades as shown in Figure 2A. The interaction between the increase in life expectancy at age 65 and the increase in the percentage of the cohort that survives to age 65 has resulted in an exceptionally large increase in the share of the gain in life expectancy that is realized after age 65. As can be seen in Figure 2B, that share was only about 20 percent during each 20-year period at the beginning of the twentieth

---

[1] For our detailed underlying data on the five-year averages for each country, see the online Appendix with this paper at ⟨http://e-jep.org⟩. Online Appendix tables 1–3 show the decreases in the coefficient of variation across the 17 high-income countries for the demographic variables portrayed in Figures 1 and 2. To include data for the United States prior to 1933 (when the Human Mortality Database series begins for the United States), we use life table data from U.S. National Vital Statistics Reports, derived from death registration states for the period 1900 to 1928, and for the whole United States thereafter (all races combined). For a small share of observations at the beginning of the century—Australia, Canada, and Northern Ireland in 1900–1919; Spain in 1900; and the United States in 1905, 1915, and 1925—we use imputed values from regressions with year and country fixed effects and country-specific linear time trends.

Figure 1
Life Expectancy at Birth and Survival to Age 65, since 1990
(in the United States and 16 other high-income countries)

A: Life Expectancy at Birth
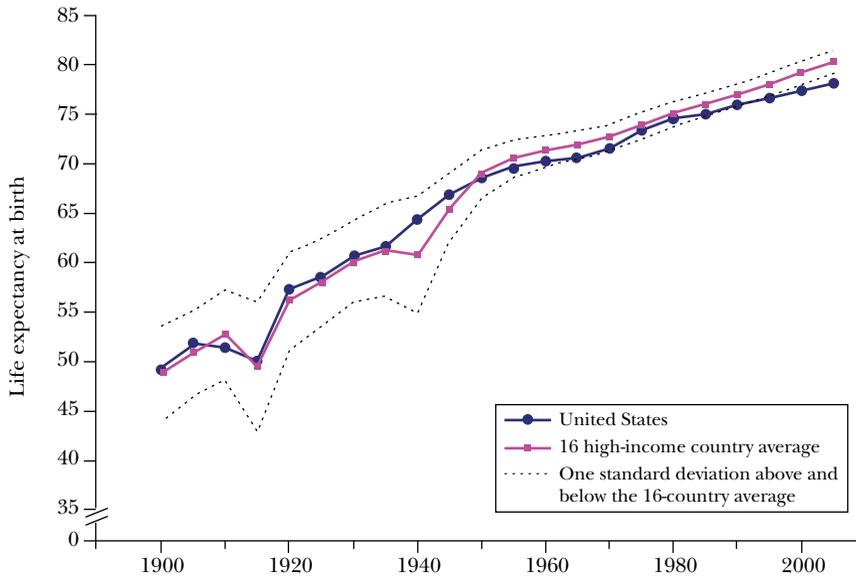
B: Percent of Birth Cohort Expected to Survive to Age 65
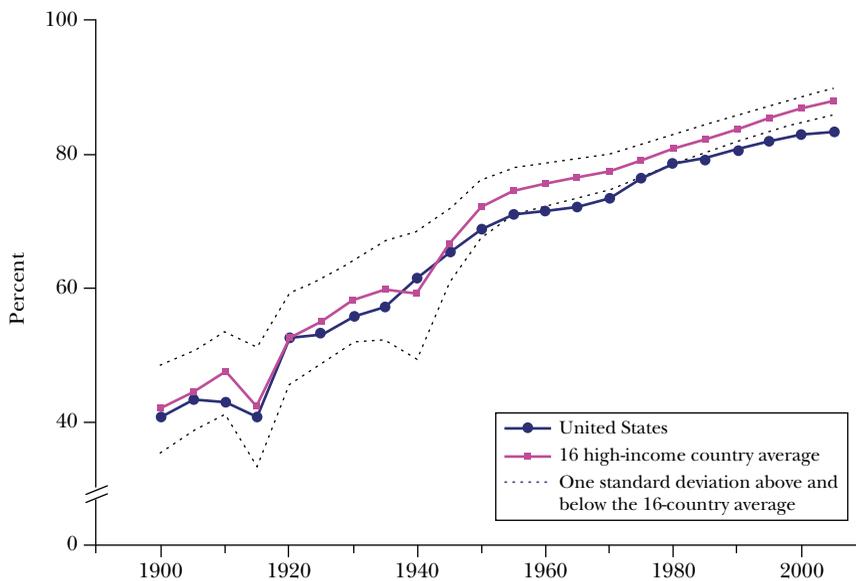
Source: Authors' calculations using data from the Human Mortality Database and other sources as detailed in the online Appendix.

Figure 2

Life Expectancy at Age 65 and Gains in Life Expectancy Realized after Age 65 Since 1990

(in the United States and 16 other high-income countries)
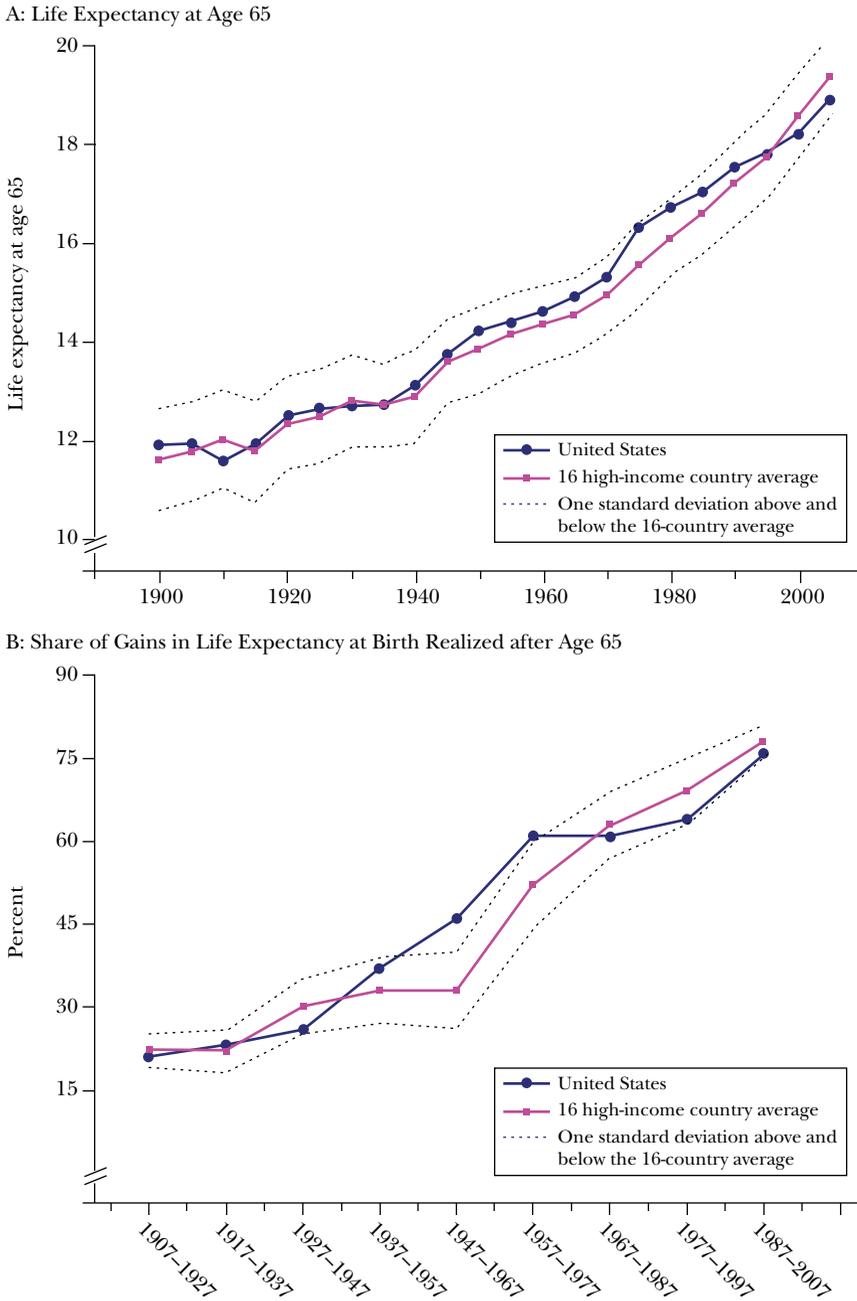
A: Life Expectancy at Age 65

B: Share of Gains in Life Expectancy at Birth Realized after Age 65

Source: Authors' calculations using data from the Human Mortality Database and other sources as detailed in the online Appendix.

Figure 3

Decrease in Death Rates by Age Group in England and Wales, 1900–1904 to 1950–54 and 1950–1954 to 2000–2004
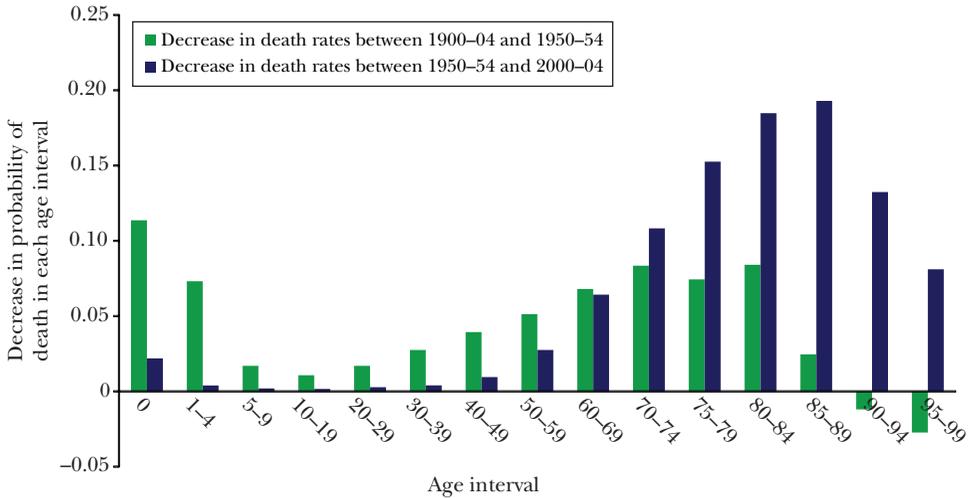
Source: Authors' calculations using data from the Human Mortality Database.

century, but it was 76 percent in the United States and 78 percent for the 16-country mean by the end of the century, and is approaching 100 percent asymptotically. Our results here are quite similar to, and extend over time, those of Lee and Tuljapurkar (1997) based on the 1995 survival profile of the United States.

We can illustrate the shift in survival improvement toward older ages by comparing the age distribution of mortality decline between the first half and second half of the twentieth century for a region with particularly reliable long-run data: England and Wales. Figure 3 shows that between 1900–1904 and 1950–1954, declines in death rates were largest for infants and children, whereas between 1950–54 and 2000–2004, declines were most salient for those over age 70. This pattern of age-specific mortality decline across the twentieth century was similar for Sweden, another country where reliable long-run data is available.[2]

The actual survival of a given birth cohort will differ from the estimates of life expectancy at birth when survival is changing over time. Remember, estimates of life expectancy at birth (what we earlier called "period" life expectancy) are based on the age-specific death rates prevailing at that year of birth. For example, in 1900–04, life expectancy at birth in England and Wales was 48.6 years. In contrast, the cohort born in 1900–1904 had a cohort life expectancy (actual mean age of death) of 53.8 years, since they experienced part of the increase in survival shown

---

[2] For details on Sweden, see the online Appendix. Figure 3 shows a slight increase in death rates for the oldest [90+] age groups between 1900–1904 and 1950–1954, perhaps because of small numbers, less-reliable data, and/or survival of a less-healthy cohort to those ages.

in Figures 1–3. The cohort born only 17 years later experienced a cohort life expectancy of 62.4 years, whereas "period" life expectancy at birth did not reach that level until 1935–1939.[3]

Nevertheless, we find that estimates based on cohort life tables prepared by the Social Security Administration (Bell and Miller 2005) exhibit a similar trend towards survival gains realized late in life: for men, the share of life expectancy increases realized after age 65 was 28 percent between the 1900 and 1920 birth cohorts, rising to a projected 62 percent between the 1980 and 2000 birth cohorts. For women, the share of life expectancy gains realized after age 65 increased from 30 percent (between the 1900 and 1920 birth cohorts) to an estimated 69 percent (between the 1980 and 2000 birth cohorts).

The century-long demographic trends shown in Figures 1 and 2 have been similar in all 17 countries with available data. From a U.S. perspective, the main difference is lagging survival to 65 compared to the other 16 countries (the U.S. line is below the 16-country average in Figure 1B); also, the United States experienced a larger rise in female life expectancy at age 65 between the 1940s and 1970s than the other countries. The relative differences among countries have decreased over time, especially for life expectancy at birth and survival to age 65.

The Longevity Transition and Expected Labor Force Participation

One of the most significant economic effects of the longevity transition is on expected lifetime labor force participation, partly in terms of total years in the workforce and especially in terms of years in the workforce as a fraction of expected years of life. Two factors affecting the connection from life expectancy to years of work are 1) whether the growing numbers of elderly are healthy enough to work and 2) the economic, social, and political pressures for a period of retirement at the end of life.

Greater longevity can have opposing effects on age-specific health status. If improved survival is correlated with reductions in morbidity for the elderly, then illness may be compressed into the end of life, as posited by the "compression of morbidity" hypothesis (Fries 1980). On the other side, medical interventions do tend to keep alive those who are in worse health (Zeckhauser, Sato, and Rizzo 1985), which suggests the possibility that the longer-lived elderly could be sicker for a longer period. The net effect of rising longevity on age-specific morbidity is an empirical question. According to the National Long-Term Care Survey, the share of elderly Americans with severe disabilities decreased from 26.2 to 19.7 percent between 1982 and 1999 (Manton and Gu 2001). Milligan and Wise (2011) find a

---

[3] Survival gains have been so dramatic that period and cohort survival significantly differs. For example, age-specific death rates for England and Wales in 1900–1904 would have led to only 43.7 percent of women and 36.4 percent of men surviving to 65. But of the cohort born in 1900–1904, 61.3 percent of women and 49.6 percent of men actually survived to age 65.

strong within-country correlation between declining mortality and improved self-assessed health for several European countries. Thus, the empirical record suggests that better health in terms of both improved survival and reduced morbidity could tend to raise age-specific rates of labor force participation. Changes in occupational structure which lower the physical demands of work also can increase participation.

Higher incomes tend to increase the demand for leisure, in the form of fewer hours of work per week and, especially recently, as a block at the end of life (Costa 1998; Murphy and Topel 2006). Furthermore, several factors might give rise to a negative interaction between improved survival and employment, at least for some subgroups. For example, the reduced selection effect of mortality might also increase the proportion of the cohort that is less valued in employment (because of less stamina, ambition, education, and the like), reducing age-specific labor force participation. Alternatively, if firms have pyramid-like organizational structures with many jobs at entry and fewer at higher levels in the hierarchy—such as the military's "up or out" policy regarding age and promotion of officers—then increases in survival will lead to crowding at higher levels of the pyramid and lower rates of participation. Moreover, a sharp rise in employment rates for women, at wages that were often below those paid to men, might have led to some decrease in the demand for men's labor.

On net, which of these forces have predominated over the past century, and which are likely to predominate in the future? Estimates of what we call "expected labor force participation" can help answer this question.

Calculating Expected Labor Force Participation

We define "expected labor force participation" (XLFP) as the total years an individual is expected to participate in the labor force, based on period estimates of survival, and labor force participation by gender and age. That is

$$XLFP_{jt} = \sum_{i=1}^{100} \pi_{ijt} \, L_{ijt},$$

where L_{ijt} is the labor force participation rate for age i and gender j in year t, weighted by probability of survival to age i (π_{ijt}). It is necessary to examine men and women separately because of the large upsurge in female labor force participation between the 1950s and 2000 (Goldin 1986, 1990; Costa 2000). Our calculations rely on labor force participation rates from decennial censuses (1900–1930) and the Current Population Survey (1942–2007). As in the earlier estimates of life expectancy, we can calculate both "period" expected labor force participation, which is based on the age-specific labor force participation rates prevailing at a certain point of time, or the actual realized labor force participation rates for a birth cohort; these estimates will differ when age-specific labor force participation rates are changing over time.

Changes in lifetime expected labor force participation can be decomposed into two factors: changes in survival to given ages and changes in age/sex-specific

rates of labor force participation. For example, we calculate the effect of improving survival, holding age-specific labor force participation rates constant at their 2007 values. We also calculate the effect of changing rates of labor force participation, holding survival rates constant.[4]

Our work is related to the literature on expected lifetime work hours (Hazan 2009) and work-life expectancy (Smith 1982), including the work-life estimates for the U.S. population from the 1950s through the early 1980s from the Bureau of Labor Statistics.[5] As far as we are aware, this paper is the first to produce work-life estimates for the United States covering the period 1900 to 2007, decompose those changes into survival and age/sex-specific labor force participation effects, and to estimate work-life expectancy relative to life expectancy at birth for a broader range of countries in recent decades.

U.S. Expected Labor Force Participation since 1900

In the early twentieth century, most of the increase in life expectancy arose from the dramatic decrease in mortality at young ages. This change first increased the years of youth dependency for these cohorts, and then increased expected labor force participation—the expected number of years that an individual will be in the labor force if he or she participates at the average labor force participation rate for each sex and age in a given year.

Figure 4A shows that years of expected labor force participation at birth for U.S. males increased by a third—from about 30 to 40 years—between 1900 and 1950. For the most recent half century, however, increases in survival have been offset by decreasing age-specific labor force participation rates for men, causing expected lifetime labor force participation to be relatively constant at about 40 years. Because life expectancy at birth has continued to increase, male expected labor force participation as a fraction of expected years of life has declined, as shown in Figure 4B. Table 1 shows that in the United States between 1900 and 2000, male labor force participation increased from 30 to 40.5 years, female participation from 6.4 years to 34.4 years, and for the total population from 18.5 to 37.4 years. This increase in years of expected labor participation is two-thirds of the total gain in life expectancy at birth of 28.2 years over the twentieth century.

How much of this change is attributable just to longer life expectancies? If we hold age-specific rates of labor force participation constant but allow survival rates to grow at the actually observed pace, the rise in life expectancy alone would have increased expected labor force participation by 13.3 years for males and by 10.8 years for females since 1900 (as shown in Table 1). The effect of mortality decline was concentrated in the first half of the twentieth century. Indeed, for men,

---

[4] These are decompositions 1B and 2B, respectively, in Table 7 of the online Appendix. Alternative calculations, using 1900 as the base year (decompositions 1A and 2A), show similar results.

[5] In other pre-existing work in this area, Hunt, Pickersgill, and Rutemiller (2001) update worklife estimates for the U.S. based on 1998–1999 labor force participation rates. Millimet, Nieswiadomy, Ryu, and Slottje (2003) use a regression framework. In related research, Hazan (2009) estimates lifetime working hours for U.S. men born between 1840 and 1970 and for the U.S. population born between 1890 and 1970.

Figure 4

U.S. Expected Labor Force Participation since 1990 and as a Share of Life Expectancy at Birth

A: Expected Labor Force Participation (XLFP)
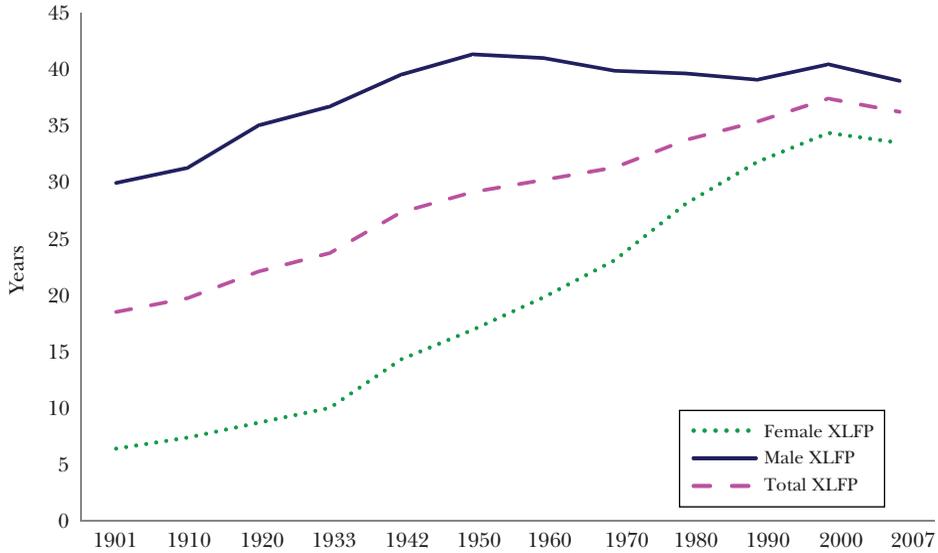
B: Expected Labor Force Participation (XLFP) as a Share of Life Expectancy (LE)
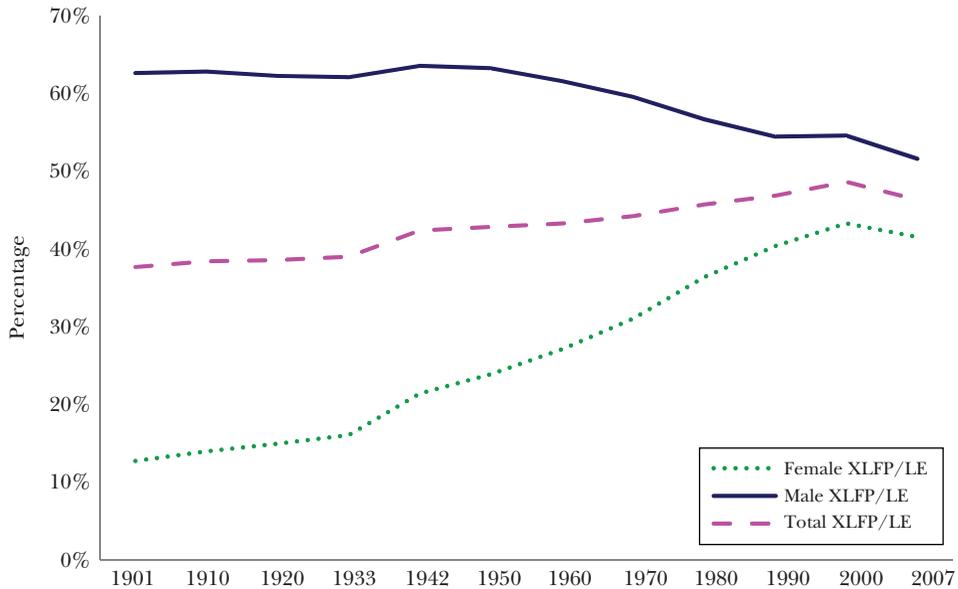
Source: Authors' calculations using data from the Human Mortality Database and other sources as detailed in the online Appendix.

Table 1

Expected Labor Force Participation in the United States, by Sex, 1900–2007

	Men					Women		
Year	*Male XLFP*	*Male XLFP holding LFP constant*	*Male XLFP adjusted for hours worked*	*Male XLFP/ LE$_0$*	*Male XLFP adjusted for hours/ LE$_0$*	*Female XLFP*	*Female XLFP holding LFP constant*	*Female XLFP/ LE$_0$*
1900	30.0	25.7	37.28	62.6%	77.9%	6.4	22.7	12.7%
1910	31.3	27.1	39.96	62.8%	80.2%	7.4	24.1	13.9%
1920	35.1	30.4	37.65	62.2%	66.8%	8.7	26.3	14.9%
1933	36.7	32.3	40.40	62.0%	68.2%	10.0	28.3	16.0%
1942	39.5	34.1	42.66	63.5%	68.5%	14.3	30.1	21.3%
1950	41.3	35.6	38.22	63.2%	58.4%	16.9	31.3	23.8%
1960	41.0	36.3	36.79	61.5%	55.2%	19.8	32.0	27.0%
1970	39.9	36.4	34.67	59.5%	51.7%	23.1	32.2	31.0%
1980	39.6	37.4	n.a.	56.6%	n.a.	28.1	32.8	36.3%
1990	39.1	37.9	n.a.	54.4%	n.a.	31.8	33.1	40.3%
2000	40.5	38.7	n.a.	54.5%	n.a.	34.4	33.3	43.2%
2007	39.0	39.0	n.a.	51.6%	n.a.	33.5	33.5	41.5%
Change, 1900 to most recent	**9.0**	**13.3**	**−2.6**	**−11.0%**	**−26.1%**	**27.1**	**10.8**	**28.8%**

	Total (men and women)				
Year	*Total XLFP*	*Total XLFP holding LFP constant*	*Total XLFP adjusted for hours worked*	*Total XLFP/LE$_0$*	*Total XLFP adjusted for hours/LE$_0$*
1900	18.5	24.2	n.a.	37.6%	n.a.
1910	19.8	25.6	n.a.	38.4%	n.a.
1920	22.1	28.4	n.a.	38.5%	n.a.
1933	23.7	30.3	29.0	39.0%	47.5%
1942	27.4	32.2	29.2	42.3%	45.1%
1950	29.1	33.6	29.0	42.8%	42.5%
1960	30.2	34.2	28.8	43.2%	41.2%
1970	31.3	34.4	28.9	44.2%	40.7%
1980	33.8	35.2	n.a.	45.7%	n.a.
1990	35.4	35.6	n.a.	46.8%	n.a.
2000	37.4	36.0	n.a.	48.6%	n.a.
2007	36.3	36.3	n.a.	46.3%	n.a.
Change, 1900 to most recent	**17.7**	**12.0**	**n.a.**	**8.7%**	**n.a.**

Sources: Author calculations based on survival data from the Human Mortality Database (1933–2007), supplemented by data for death registration states for 1900–1920; and labor force participation rates from decennial censuses (1900–1930) and the Current Population Survey (1942–2007). Adjustments for hours worked draw from Hazan (2009). See the online Appendix for details.

Notes: Expected Labor Force Participation (XLFP) is calculated as the total years an individual is expected to participate in the labor force based on period estimates of labor force participation and survival by gender and age. XLFP for a given year represents the expected number of years that an individual would be in the labor force if he or she participates at the average LFP rate for each age in that given year. LE$_0$ is life expectancy at birth. "XLFP holding LFP constant" uses 2007 age- and sex-specific labor force participation rates, but allows survival to each age to vary as it actually did between 1900 and 2007.

if we hold age-specific labor force participation rates constant but allow survival rates to vary in calculating expected labor force participation ("male XLFP holding LFP constant"), the ratio of years of expected labor force participation to life expectancy at birth was relatively constant at 54 percent from early in the twentieth century until about 1970 (not shown in the table). At that point, it began a slow but seemingly inexorable decline, now falling to about 50 percent.

Actual years of expected labor force participation, reflecting both survival effects and changes in age-specific labor force participation rates, have also begun to decline. As shown in both Table 1 and Figure 4B, the ratio of years of expected labor force participation to life expectancy at birth ($XLFP/LE_0$) has declined for U.S. men from 62.6 percent in 1900 to 51.6 percent in 2007. That same ratio for women increased from 12.7 percent in 1900 to 43.2 percent in 2000, before declining slightly to 41.5 percent by 2007. For the overall U.S. population, years of expected labor force participation divided by life expectancy at birth peaked at 48.6 percent in 2000 and declined slightly to 46.3 percent by 2007.

Since 1950, increases in survival and declines in age-specific participation rates of men tended to offset one another. For example, between 1950 and 2007, labor force participation rates of men ages 45–54 declined from 95.8 percent to 88.2 percent, but survival to age 50 increased from 84.1 to 92.2 percent, so the total expected years in the labor force between ages 45 and 55 remained eight years.[6] For women, increases in years of expected labor force participation mostly reflect increases in age-specific rates of labor force participation, especially after 1950. Accordingly, for women, if we hold age-specific labor force participation rates constant but allow survival rates to vary in calculating expected labor force participation ("female XLFP holding LFP constant"), the ratio of years of expected labor force participation to life expectancy at birth has declined slowly but steadily from about 45 percent in the first few decades of the twentieth century to about 40 percent (not shown in the table).

The increase in female labor force participation since the late 1950s could be considered primarily a one-time substitution from unpaid home production to paid work outside the home (Goldin 1990; Costa 2000). If so, then the decrease in years of expected labor force participation for women in the United States since 2000 would reflect completion of the one-off change and the beginning of a similar trend as seen for men—that is, a decline of years in the labor force as a share of life expectancy at birth.

Taking into account the decrease in the intensive margin—annual hours worked per full-time worker—tends to reinforce the conclusion that expected work life has declined as a fraction of life expectancy at birth. Hazan (2009)

---

[6] For the detailed data behind these calculations across the range of ages, for both men and women, see online Appendix Table 7, which offers alternative decompositions of changes in both male and female labor force participation. Online Appendix Table 7 also shows that holding age-specific labor force participation rates constant (at either their 1900 or 2007 values) would have led to a larger increase in male expected labor force participation than actually observed.

estimated lifetime work hours over the past century conditional on survival to age five. We adapt Hazan's data to life expectancy at birth to calculate years of expected labor force participation adjusted for hours worked and show the results in Table 1 (the online Appendix available with this paper at ⟨http://e-jep.org⟩ has details of our calculations).

Calculation of a century-long trend in expected years of labor force participation in other high-income countries is not possible because there is no reliable source for internationally comparable labor force participation rates before 1980. Given the similarities in trends of both survival and labor force participation across these countries for the available years, we suspect the trend of declining expected labor force participation as a share of life expectancy at birth that we found for the United States reflects a broad and robust trend that countries experience as they reach high life expectancy levels. Indeed, with the sole exception of the Netherlands, the ratio of years of expected labor force participation to life expectancy at birth has declined since 1980 for males in all other high-income countries in our analyses.[7] Adjusting for a decline in work hours would reinforce this trend.

Demographic Transition across Stages of Economic Development

The demographic transition traces out a pathway, with many societies arrayed along earlier phases of the transition roughly and imperfectly in accordance with their per capita incomes. Many developing countries are currently experiencing the original demographic transition. For example, Table 2 shows that between 1990 and 2010, the share of years lived past 65 as a percentage of increase in life expectancy at birth was only a little over one-third in Vietnam and Brazil, and less than one-quarter in Bangladesh—comparable to levels a century earlier in today's high-income countries.

Improving health and increasing life expectancy at birth clearly can contribute to better living standards for the world's poor (World Health Organization 2002). Data on labor force participation for developing countries is not always reliably comparable across countries and over time. Nevertheless, the importance of improved survival for gains in expected labor force participation at early stages of the longevity transition can be illustrated with extant data. For example, in 1980 only 70 percent of Indonesian men survived to age 45; by 2007, 90 percent did. This improved survival added 10 years to expected labor force participation rates for Indonesian males between 1980 and 2007. As a result, expected labor force participation rates for Indonesian males rose to 43.7 years, which was 64.5 percent of life expectancy at birth in 2007.

---

[7] The online Appendix tables provide calculations of expected labor force participation across 15 countries since 1980; see Appendix Table 8 in the online Appendix available with this paper at ⟨http://e-jep .org⟩. Milligen and Wise (2011, p. 17) examine the age at which male mortality was 1.5 percent in 1977 and 2007, finding that at that age almost 90 percent of UK men were employed in 1977, but by 2007, only 30 percent were.

Table 2
The Longevity Transition in Asia and Select Developing Countries

Country	Change in years lived past 65 as a percentage of change in life expectancy at birth, 1990–2010	
	Males	*Females*
Japan	72.7%	87.0%
South Korea	45.4%	57.1%
China	51.9%	40.6%
Philippines	26.2%	36.0%
Indonesia	26.1%	35.7%
Brazil	34.2%	35.0%
Vietnam	32.5%	34.7%
India	23.6%	25.8%
Bangladesh	20.7%	25.4%

Source: Authors' calculations based on the life tables for each country prepared by the International Programs Center of the U.S. Bureau of the Census in its International Data Base.

China and India are especially important cases to consider, given their large populations and relatively rapid economic development. In India, the share of years lived past 65 as a percentage of increase in life expectancy at birth was only one-quarter (as shown in Table 2) in the most recent 20-year period. For China, that share was 52 percent for men and 41 percent for women in the 1990–2010 period.

China's position reflects the rapidity of its demographic transition since the early 1970s and its achievement of relatively high levels of health despite low per capita income by the end of the Mao era (Banister 1987; Wang 2011). Indeed, despite the higher death rates associated with the Great Leap Famine of 1959–1961, China's growth in life expectancy from approximately 35–40 years in 1949 to 65.5 years in 1980 ranks as the most rapid sustained increase in documented global history.[8] These earlier health improvements and growth of the working-age population contributed to China's unprecedented economic growth for the past quarter-century. Wang and Mason (2008) estimate that between 1982 and 2000, about 15 percent of China's rapid growth in output per capita stemmed from the demographic dividend. (Bloom and Williamson (1998) estimate that one-quarter to one-third of the growth rates in the "East Asian miracle" stemmed from the demographic dividend.) Although the pace of mortality decline in China has slowed, it

[8] Miller, Eggleston, and Zhang (2012) assess the relative importance of various explanations proposed for these gains, including better nutrition, widespread public health interventions, improved access to medical care, and increases in educational levels. They find that gains in education and public health campaigns jointly explain 25–32 percent of the crude death rate decline under Mao, and similar proportions of the dramatic reductions in infant and under-five mortality in that period.

continues: Chinese life expectancy increased between 1990 and 2010 from 69.9 to 76.8 years for women and from 66.9 to 72.5 years for men.

With a rapid demographic transition to relatively low mortality and low fertility, China's population is now aging (Peng 2011). Many policy challenges loom as China establishes social and economic institutions commensurate with its transition to a middle-income, market-based economy with a large elderly population (Eggleston and Tuljapurkar 2010; Chen, Eggleston, and Li forthcoming). One additional challenge for China in reducing the growth-slowing potential of the new demographic transition is China's increasing burden of chronic disease. Fueled by rapid urbanization, increases in high-fat and calorie-rich diets, reductions in physical activity, unabated male smoking and other factors, prevalence of chronic disease in China has quickly caught up with that of high-income countries. For example, the age-standardized prevalence of diabetes among adults in China was 9.7 percent in 2007–2008, more than three times reported prevalence in 1994 (Yang et al. 2010), comparable to the U.S. rate of 8.3 percent overall in 2010 and 11.3 percent among adults (CDC 2011), and higher than the OECD average (OECD 2011).

The timing and the rapidity of the longevity transition has varied across countries and regions. For example, in Japan between 1950 and 1970, only 13.1 percent of increase in male life expectancy at birth was realized after age 65; for women, that figure was 17.3 percent. During the 1990 to 2009 period, Japan led the world in the new demographic transition, with the share of gains in life expectancy at birth realized after age 65 reaching 72.7 percent for men and 87 percent for women (again, as shown in Table 2).

The original and the new demographic transitions are inextricably intertwined with the evolution of social and economic institutions (Aoki 2011). Evidence is mounting that no society at an advanced stage of economic development can presume that further gains in longevity will contribute to growth of per capita income under currently prevailing institutions. For example, Lee and Mason (2011) compare the "average age of consumption" to the "average age of labor income"[9] across a large group of countries for which they and their international collaborators have collected detailed generational accounts, including the value of assets and transfer wealth from social support programs (but not including bequests or value of nonmarket labor). They find that for developing countries, net transfers flow strongly downward from older to younger ages. However, in a "sea change" analogous to what we call the new demographic transition, "the direction of intergenerational transfers in the population has shifted from downward to upward, at least in a few leading rich nations" including Germany, Austria, and Japan (Lee and Mason 2011, p. 116). Although the Lee–Mason estimates are cross-sectional, the link to the longevity transition is clear: for the 13 countries that overlap between

---

[9] They construct the average ages of consumption and labor income as follows: "The average age of consumption is calculated by multiplying each age by the aggregate consumption at that age, summing these products over all ages, then dividing by the total amount of consumption at all ages. An equivalent calculation gives the average age of labor income" (Lee and Mason 2011, p. 123).

their dataset and ours, there is a strong negative correlation (–0.89) between the share of gains in life expectancy over the past 20 years that were realized after age 65 and the current number of years by which the average age of income exceeds the average age of consumption. In other words, the more the gains in life expectancy are concentrated in traditional retirement years, the closer the intergenerational transfers are to being upward rather than downward.

For a broader group of 107 countries, Bloom, Canning, and Fink (2010) calculate counterfactual annual growth rates of per capita income between 1960 and 2005, using 2005–2050 projections of demographics. The results vary depending on the level of economic development. They find that in most non-OECD countries, declining youth dependency would more than offset increasing old-age dependency. However, about half of countries would have grown more slowly using 2005–2050 projections of demographics. Among 26 OECD countries analyzed, 25 of them (Turkey is the exception) would have had lower economic growth—averaging 2.1 rather than 2.8 percent per year—under the counterfactual of 2005–2050 demographic change.

Policy Implications of the New Demographic Transition

Historically, adults produced more than they consumed and supported children. With such a pattern in place, the increase in proportion of the population in older years implied by the demographic transition might have been thought to shift out the social budget constraint as people expanded their number of years worked. However, "a funny thing happened along the way: societies invented retirement . . . and the economic consequences of population aging are now viewed with alarm" (Lee and Mason 2011, p. 115).

Retirement, a relatively new phenomenon in human history, can be viewed as a response to many economic and social changes. Contributing factors include the shift from self-employment on farms or small businesses to wage and salary status; more rapid technological change, resulting in more rapid obsolescence of human capital (alongside compensation packages that often underpay at the beginning and overpay at the end of a career, as discussed in Lazear 1981); the introduction of a variety of health and welfare programs which assist the elderly but also discourage work; an income-driven increase in the demand for leisure, with the diminishing marginal value of an even shorter work week overtaken by the efficiency gains of a block of leisure at the end of life; and, in times of high unemployment, public concern about job opportunities for younger workers.

Will the new demographic transition inevitably lead to slower economic growth? As people foresee longer lives, they might choose to work longer, save more, and/or invest in human capital in sufficient amounts and innovative enough ways that longer lives continue to contribute to increased prosperity. In this spirit, Bloom, Canning, and Fink (2010) assert that "the problem of population ageing is more a function of rigid and outmoded policies and institutions than a problem of demographic change per se" (p. 607).

It is not clear, however, that the United States or other high-income countries even further along in the new demographic transition are reshaping their policies and institutions sufficiently in response to the longevity transition. Although both the United States and France have increased the age of retirement or age to qualify for early retirement, social welfare systems across the high-income countries of the world continue to give strong incentives for earlier, rather than later, retirement (Gruber and Wise 1998). Between 1965 and 2005, the correlation between change in male life expectancy at birth and change in retirement age is actually negative: −0.21 (Bloom, Canning, and Fink 2010, p. 591). This trend cannot continue indefinitely: longer and longer retirement lives are not consistent with continued increases in per capita income unless there are significant increases in savings, investment, and productivity. It is ironic that the same phenomenon that led to higher GDP per capita—namely higher life expectancy—could now lead to lower GDP per capita.

Successful navigation of the new demographic transition calls for a combination of policies to give incentives for more savings and investment (including in human capital) earlier in the lifecycle and for additional work later in the lifecycle. Two forces in particular might move the society in that direction: improvement in health, and reductions in the transfers that the elderly can expect to receive from the young.

Public policy should encourage higher labor force participation for the elderly, both by reducing the disadvantages that employers face when employing older workers and by providing enhanced incentives to individuals to continue to work. "People cannot expect to finance 20–25 year retirements with 35-year careers," Shoven noted (as quoted in Haven 2011). "It just won't work. Not in Greece [or] the United States . . . Eventually, we are going to have to increase retirement ages." However, increasing labor force participation for the 65-plus age group alone probably won't make a big difference: even a doubling of those rates from their 2007 levels of 12.6 for women and 20.5 for men would not bring the U.S. ratio of expected labor force participation to life expectancy at birth back to its 2000 level. Increased labor force participation by men in the 50–64 age bracket is also needed.

Public policy might also seek to improve productivity, with an emphasis on education and building human capital early in the lifecycle, and on investment to reduce morbidity and improve the ability to work later in life. Whether compression of morbidity later in life will continue depends on whether improvements in medical technology and in the socioeconomic determinants of health are offset by adverse trends such as increasing obesity. A potentially promising focus here would be to consider investments in public health and medical technologies that reduce morbidity and improve quality of life, as well as more focus on medical innovations that reduce costs of care. (One example of a policy consistent with both objectives would be expansion of palliative care as a substitute for what can otherwise be extremely expensive end-of-life care in a hospital—especially in countries where the concept of hospice services is relatively new, such as China.)

Finally, increased savings, investment, and capital formation could help in fueling endogenous growth (Lucas 1988; Romer 1990). U.S. personal savings rates

have been low for many decades. Increasing the savings rate of individuals before they retire would ameliorate the potential adverse impact of longevity on economic growth. Countries will need to make fiscally realistic structural changes to entitlement programs—such as Medicare and Social Security in the United States—to support acceptable living standards and improvements in health.

High-income societies are now facing a new demographic transition: the longevity transition. They must decide how to respond to mortality decline when almost all of the decline will occur late in life. Additional increases in life expectancy will result in further declines in expected labor force participation as a percentage of life expectancy at birth unless there is a significant rise in labor force participation rates across both middle and older ages. Of course, increased life expectancy has great value independent of its relationship to per capita income (Murphy and Topel 2006). The original demographic transition gave society a "demographic gift" of higher per capita incomes (Bloom and Williamson 1998) without much need for a policy response, but the new demographic transition requires politically difficult policies if societies wish to preserve a positive relationship running from increased longevity to greater prosperity.

References

Aoki, Masahiko. 2011. "The Five-Phases of Economic Development and Institutional Evolution in China and Japan." Presidential Lecture at the XVIth World Congress of the International Economic Association.

Banister, Judith. 1987. *China's Changing Population.* Stanford University Press.

Barker, David J. 1990. "The Fetal and Infant Origins of Adult Disease." *British Medical Journal* 301(6761): 1111.

Bell, Felicitie C., and Michael L. Miller. 2005. *Life Tables for the United States Social Security Asia 1900–2100.* Actuarial Study No. 120, Social Security Administration Office of the Chief Actuary, SSA Pub. No. 11-11536. Available at: http://www.socialsecurity.gov/OACT/NOTES/s2000s.html.

Bloom, David E., David Canning, and Günther Fink. 2010. "Implications of Population Ageing for Economic Growth." *Oxford Review of Economic Policy* 26(4): 583–612.

Bloom, David E., David Canning, and Jaypee Sevilla. 2003. *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change.* Monograph Reports, MR-1274. Monica, CA: RAND Corporation. http://www.rand.org/pubs/monograph_reports/MR1274.

Bloom, David E., and Jeffrey G. Williamson. 1998. "Demographic Transitions and Economic Miracles in Emerging Asia." *World Bank Economic Review* 12(3): 419–55.

Chen, Qiulin, Karen N. Eggleston, and Ling Li. Forthcoming. "Demographic Change, Intergenerational Transfers, and the Challenges to Social Protection Systems in China." *Demographic Transition and Inclusive Growth in Asia.* Edward Elgar.

Centers for Disease Control and Prevention (CDC). 2011. "National Diabetes Fact Sheet, 2011." Atlanta, GA: CDC. http://www.cdc.gov/diabetes /pubs/pdf/ndfs_2011.pdf.

Costa, Dora L. 1998. *The Evolution of Retirement: An American Economic History, 1880–1990.* University of Chicago Press.

Costa, Dora L. 2000. "From Mill Town to Board Room: The Rise of Women's Paid Labor." *Journal of Economic Perspectives* 14(4): 101–122.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The Determinants of Mortality." *Journal of Economic Perspectives* 20(3): 97–120.

Davis, Kingsley. 1945. "The World Demographic Transition." *Annals of the American Academy of Political and Social Science* 237(1): 1–11.

Eggleston, Karen N., and Shripad Tuljapurkar, eds. 2010. *Aging Asia: Economic and Social Implications of Rapid Demographic Change in China, Japan, and South Korea.* Shorenstein APRC; distributed by Brookings Institution Press.

Fogel, Robert W. 1994. "Economic Growth, Population Theory, and Physiology: The Bearing of Long-term Processes on the Making of Economic Policy." *American Economic Review* 84(3): 369–395.

Fries, James F. 1980. "Aging, Natural Death, and the Compression of Morbidity." *New England Journal of Medicine* 303(3): 130–135.

Fuchs, Victor R. 1999. "'Provide, Provide': The Economics of Aging." In *Medicare Reform: Issues and Answers,* edited by Andrew J. Rettenmaier and Thomas R. Saving, 15–36. University of Chicago Press.

Galor, Oded, and David N. Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86(3): 374–87.

Gruber, Jonathan, and David A. Wise. 1998. *Social Security Programs and Retirement around the World.* University of Chicago Press.

Goldin, Claudia. 1986. "The Female Labor Force and American Economic Growth: 1890 to 1980." In *Long-Term Factors in American Economic Growth,* Conference on Income and Wealth, Volume 51, edited by Stanley Engerman and Robert Gallman, 557–604. University of Chicago Press.

Goldin, Claudia. 1990. *Understanding the Gender Gap: An Economic History of American Women.* Oxford University Press.

Guinnane, Timothy W. 2011. "The Historical Fertility Transition: A Guide for Economists." *Journal of Economic Literature* 49(3): 589–614.

Haven, Cynthia. 2011. "Stanford Economist: How Do We 'Get off This Path of Deficits as Far as the Eye Can See?'" *Stanford Report,* August 2. http://news.stanford.edu/news/2011/august /shoven-debt-qanda-080211.html.

Hazan, Moshe. 2009. "Longevity and Lifetime Labor Supply: Evidence and Implications." *Econometrica* 77(6): 1829–63.

Human Mortality Database. University of California, Berkeley; and Max Planck Institute for Demographic Research. Available at: www .mortality.org.

Hunt, Tamorah, Joyce Pickersgill, and Herbert Rutemiller. 2001. "Recent Trends in Median Years to Retirement and Worklife Expectancy for the Civilian U.S. Population (Prepared Using 1998/99 BLS Labor Force Participation Rates)." *Journal of Forensic Economics* 14(3): 203–227.

Lazear, Edward P. 1981. "Agency, Earnings Profiles, Productivity, and Hours Restrictions." *American Economic Review* 71(4): 606–620.

Lee, Ronald D. 2003. "The Demographic Transition: Three Centuries of Fundamental Change." *Journal of Economic Perspectives* 17(4): 167–190.

Lee, Ronald D., and David S. Reher. 2011. "Introduction: The Landscape of Demographic Transition and its Aftermath." *Population and Development Review* 37(Issue Supplement s1): 1–7.

Lee, Ronald D., and Andrew Mason. 2011. "Generational Economics in a Changing World." *Population and Development Review* 37(Issue Supplement s1): 115–142.

Lee, Ronald D., and Shripad Tuljapurkar. 1997. "Death and Taxes: Longer Life, Consumption, and Social Security." *Demography* 34(1): 67–81.

Lucas, Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22(1): 3–42.

Manton, Kenneth G., and Xiliang Gu. 2001. "Changes in the Prevalence of Chronic Disability in the United States Black and Nonblack Population above Age 65 from 1982 to 1999." *PNAS* 98(11): 6354–59.

Miller, N. Grant, Karen N. Eggleston, and Qiong Zhang. 2012. "Understanding China's Mortality Decline under Mao: A Provincial Analysis, 1950–1980." Unpublished paper.

Milligan, Kevin, and David A. Wise. 2011 "Social Security and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms–Introduction and Summary." NBER Working Paper 16719.

Millimet, Daniel L., Michael Nieswiadomy,

Hang Ryu, and Daniel Slottje. 2003. "Estimating Worklife Expectancy: An Econometric Approach." *Journal of Econometrics* 113(1): 83–113.

Murphy, Kevin M., and Robert H. Topel. 2006. "The Value of Health and Longevity." *Journal of Political Economy* 114(4): 871–904.

OECD. 2011. *Health at a Glance 2011: OECD Indicators.* Organization for Economic Cooperation and Development. http://www.oecd.org /dataoecd/6/28/49105858.pdf.

Peng, Xizhe. 2011. "China's Demographic History and Future Challenges." *Science* 333(6042): 581–587.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy*, vol. 98, no. 5, Part 2: The Problem of Development: A Conference of the Institute for the Study of Free Enterprise Systems, pp. S71–S102.

Smith, Shirley J. 1982. "New Worklife Estimates Reflect Changing Profile of Labor Force." *Monthly Labor Review* 105(3): 15–20.

Swift, Robyn. 2011. "The Relationship between Health and GDP in OECD Countries in the Very Long Run." *Health Economics* 20(3): 306–322.

Wang, Feng. 2011. "The Future of a Demographic Overachiever: Long-Term Implications of the Demographic Transition in China." *Population and Development Review* 37(Supplement): 173–90.

Wang, Feng, and Andrew Mason. 2008. "The Demographic Factor in China's Transition." In *China's Great Economic Transformation,* edited by Loren Brandt and Thomas G. Rawski, 136–66. Cambridge University Press.

World Health Organization. 2002. *Macroeconomics and Health: Investing in Health for Economic Development: Report of the Commission on Macroeconomics and Health*: World Health Organization (WHO).

Yang, W. et al. 2010. "Prevalence of Diabetes among Men and Women in China." *New England Journal of Medicine* 362(12): 1090–1101.

Zeckhauser, Richard J., Ryuzo Sato, and John Rizzo. 1985. "Hidden Heterogeneity in Risk: Evidence from Japanese Mortality." In *Health Intervention and Population Heterogeneity: Evidence from Japan and the United States,* pp. 23–131. National Institute for Research Advancement.

Groups Make Better Self-Interested Decisions[†]

Gary Charness and Matthias Sutter

Adecision maker in an economics textbook is usually modeled as an individual whose decisions are not influenced by any other people, but of course, human decision-making in the real world is typically embedded in a social environment. Households and firms, common decision-making agents in economic theory, are typically not individuals either, but groups of people—in the case of firms, often interacting and overlapping groups. Similarly, important political or military decisions as well as resolutions on monetary and economic policy are often made by configurations of groups and committees rather than by individuals. Economic research has developed an interest regarding group decision-making—and its possible differences with individual decision-making—only rather recently. Camerer (2003) concludes his book on *Behavioral Game Theory* with a section on the top ten open research questions for future research, listing as number eight "how do teams, groups, and firms play games?" Potential differences between individual and group decision-making have been studied over the past ten to 15 years in a large set of games in the experimental economics literature.

In this paper, we describe what economists have learned about differences between group and individual decision-making. This literature is still young, and in this paper, we will mostly draw on experimental work (mainly in the laboratory) that has compared individual decision-making to group decision-making, and to individual

■ *Gary Charness is Professor of Economics, Department of Economics, University of California at Santa Barbara, Santa Barbara, California. Matthias Sutter is Professor of Experimental Economics, Department of Public Finance, University of Innsbruck, Innsbruck, Austria, and Professor of Economics, Department of Economics, University of Gothenburg, Göteborg, Sweden. Their email addresses are ⟨charness@econ.ucsb.edu⟩ and ⟨matthias.sutter@uibk.ac.at⟩.*

decision-making in situations with salient group membership.[1] In a nutshell, the bottom line emerging from economic research on group decision-making is that groups are more likely to make choices that follow standard game-theoretic predictions, while individuals are more likely to be influenced by biases, cognitive limitations, and social considerations. In this sense, groups are generally less "behavioral" than individuals. An immediate implication of this result is that individual decisions in isolation cannot necessarily be assumed to be good predictors of the decisions made by groups. More broadly, the evidence casts doubts on traditional approaches that model economic behavior as if individuals were making decisions in isolation.

We focus on three main lessons in this paper. First, the use of rationality as a useful assumption for studying real-world economic behavior may not be as problematic as some have argued. In this context, what we mean by rationality is that cognitive limitations (in the sense of bounded rationality) apply less to groups and that groups engage in more self-interested behavior than do individuals. In fact, we find that such rationality applies pretty well to group decisions, and we argue that groups are at least an element in most decisions. People always belong to some groups (for instance, males or left-handed people, and the like) and their behavior may well be affected when a sense of group membership is present. In addition, many important economic decisions—including decisions where consequences affect individual decision units, such as buying a home or choosing a health insurance plan—are made after some consultations with others, even if they are not explicitly part of a group decision-making process. Thus, while the behavioralist critique of deviations from the rational paradigm is important and has many applications, we should be careful about how we describe economic agents in our models.[2] If we were to specify that most of these agents are acting in social or group contexts, then the claim that they are rational actors would be strengthened.

A second lesson is that, from a social point of view, group decision-making may be a method for individuals to try to protect themselves from the consequences of their own behavioral irrationalities or limitations. Suppose an individual is very present oriented and so has great difficulties in saving for retirement. Perhaps through participation in groups at work or in one's social, political, recreational, or religious life, one can achieve at least a modicum of success in assuring a retirement income. As another example, perhaps one does not have the self-discipline to exercise on one's own but will do so with regularity if one forms or joins a group of people who jog together or meet to play tennis. In a business environment, one might find it personally nearly impossible ever to fire anyone, even if the result is that one's business goes bankrupt. But it might be possible to achieve this end by being part of a committee that makes such decisions. In short, group membership

[1] The evidence from laboratory experiments has the advantage of allowing for a clean and controlled analysis of group decision-making and group membership effects because subjects are randomly assigned to making a choice individually or as a group member. This is more difficult with field data, but not impossible, as Lesson 2 below will confirm.
[2] See Levitt and List (2007) for an account of the behavioralist critique.

and group participation can facilitate people doing things that they wish (on some important level) to do, but might be unable to do without the support of a group.

The third lesson is that in some environments—for example, in cases where trust and cooperation lead to improved social welfare—it might make sense to have individuals making decisions, and in other cases—for example, when deeper levels of insight or analytical problem-solving and coordination are especially valuable—it might make sense to have groups making decisions. For example, a considerable body of experimental literature suggests that, perhaps because individuals are unselfish or socially oriented, they are able to reach welfare, socially efficient outcomes in situations like the prisoner's dilemma or a "trust game."[3] In these settings, group decision-making/membership presumably leads to lower social welfare (in the sense of the total social material payoffs), because the element of trust or cooperation is sharply reduced. However, we will explore a number of other settings where group decision-making is more sophisticated and effective. Thus, researchers can start groping toward a provisional taxonomy concerning where and when it is optimal to have a group process or an individual one.

We discuss these three lessons in the following sections. We intersperse the discussions with evidence, primarily experimental, for the story being told.[4] Building on this evidence, we then discuss the major sources for differences in decisions made by individuals and groups before we conclude with an outlook on promising avenues for future research on group decision-making.

Lesson One: Groups are More Cognitively Sophisticated

We look first at experiments that compare individual and group decisions where each player is only concerned with making the best selfish decision without regard to social considerations.[5] This category includes investment or portfolio decisions, tournaments, and tasks where the ability to reason through the problem is important due to some cognitive limitation or psychological bias that typically affects the outcome.

One well-known example is the beauty-contest game (also known as "the guessing game"). In this simultaneous move game, a set of n decision makers chooses a number from the interval $[0, 100]$, and the winner is the decision maker whose

[3] It may also be possible that an individual's social concerns are directed at one's group in the case of group membership. We shall take up this point later.

[4] While we only discuss a few studies in each section, we present a brief description of the most important other studies supporting our conclusions in an online Appendix available with this paper at ⟨http://e-jep.org⟩.

[5] We focus here on results from experimental economics, with less emphasis on psychological research: see Levine and Moreland (1998) for an account of small-group research in psychology. From our perspective, the research in experimental economics has two particular advantages: 1) the ubiquitous use of financial incentives, a condition that is often not met in experimental research in the field of psychology; and 2) the use of simpler paradigms that allow for benchmarking behavior to standard game-theoretic predictions. Psychological paradigms are often much more complex, thereby making it more difficult to characterize general patterns of behavioral differences between individuals and groups.

Figure 1
Median Number Chosen by Groups and Individuals in a Beauty-Contest Game
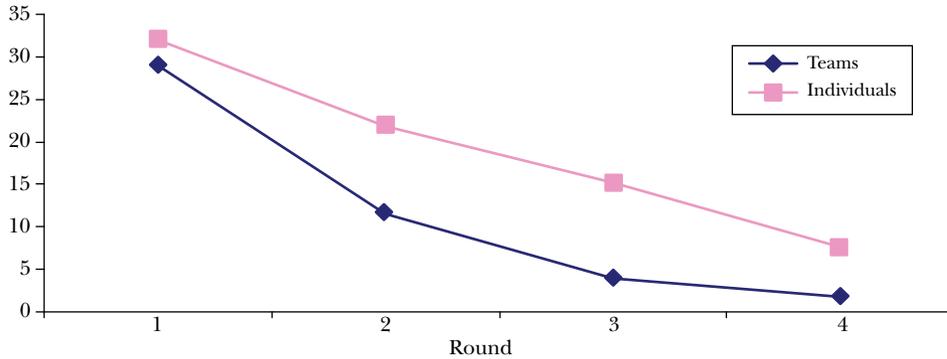
Note: In this simultaneous move game, a set of n decision makers chooses a number from the interval $[0, 100]$, and the winner is the decision maker whose number is closest to p times the average chosen number, with p being some fraction less than 1.

number is closest to p times the average chosen number, with p being some fraction less than 1. The name of the beauty-contest game comes from the Keynes (1936) analogy between beauty contests and financial investing in the *General Theory*: "It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees." Similarly, in a beauty-contest game, the choice requires anticipating what average opinion will be.

However, since $p < 1$, the rational equilibrium choice will be zero. For example, a player might begin by asking what the right choice will be if all other players choose randomly over the interval between 0 and 100 with $p = 2/3$, a standard value in the experimental literature. In this case, the expected value of the average of a random choice would be 50. If one anticipates that people are guessing randomly, the best response (assuming one's own guess does not distort matters) is 33.3. However, if one anticipates that everyone else will anticipate and also best-respond to random choice, the best response is 22.2. Continuing this pattern of inference through multiple iterations, the equilibrium choice is zero.

Several studies show that in the beauty-contest game, groups choose systematically lower numbers, thus suggesting that they are reasoning more deeply about the strategy of the game and are expecting the other parties to reason more deeply as well (Kocher and Sutter 2005; Kocher, Strauss, and Sutter 2006; Sutter 2005). Kocher and Sutter (2005) find that groups think one step ahead of individuals, leading them to quicker convergence towards equilibrium play, as is shown in Figure 1, which presents the median number chosen by groups (of three subjects

Figure 2
An Urn Experiment

	Left Urn	Right Urn
Up ($p = .5$)	●●●● ○○	●●● ●●●
Down ($p = .5$)	●● ○○○○	○○○ ○○○

Source: The experiment is from Charness, Karni, and Levin (2007).

each) and individuals across four rounds. When groups and individuals compete against each other (rather than groups competing against groups, or individuals against individuals), groups outperform individuals significantly by earning under the rules of the game roughly 70 percent more than individuals (Kocher and Sutter 2005; Kocher, Strauss, and Sutter 2006).[6] One possible explanation why groups choose lower numbers is that the groups, in thinking through the situation, also expect other groups to think more deeply than individuals.

Two papers by Charness, Karni, and Levin (2007, 2010) specifically examine deviations from rational behavior (by looking at error) rates in tasks involving violations of first-order stochastic dominance, and in a task involving the well-known conjunction fallacy described in Tversky and Kahneman (1983). In these studies, comparisons are made among the error rates for different group sizes.

Charness, Karni, and Levin (2007) set up a situation (see Figure 2) with a left urn and a right urn, where the state of the world is "up" or "down" with equal probability; this state is fixed for two periods. A person draws a ball, observes the color, and the ball is replaced. In the "up" state, there are four black balls and two white balls in the left urn, and in the "down" state there are two black balls and four white balls in the left urn. The right urn contains six black balls in the "up" state and six white balls in the "down" state. The most interesting case is when the first draw is from the left urn, as is required in some periods. In the original set-up, black balls pay and white balls don't. With a "good" draw (black ball) one should switch to drawing from the right urn, while with a "bad" draw (white ball) one should stay with the left urn.[7] Of course, this violates the common "win-stay, lose-shift" heuristic and is thus counterintuitive. In another treatment, subjects do

---

[6] In all comparisons, the per-capita incentives were kept constant across conditions, meaning that for an identical set of decisions in a particular game, the payoffs per head were identical for individuals and each single group member.

[7] To see this, note that given the draw of a black ball, the probability that the state is "up" is 2/3. If it is "up", then the probability of drawing a black ball is 2/3; if it is "down", the probability of drawing a black ball is 1/3. Since $(2/3 \times 2/3) + (1/3 \times 1/3) = 5/9$ and the probability of drawing a black ball from the right urn is 2/3, one should switch. By the same token, the probability of drawing a black

Table 1

Error Rates in an Urn Experiment in Which One Choice Stochastically Dominates the Other

(ABCD refers to the treatment with affect, Bayesian updating, a compound lottery, and dominance, while BCD, CD, and D drop one condition in turn)

Group size	ABCD	BCD	CD	D
1	.375	.188	.302	.087
2	—	.154	.230	.030
3	—	.075	—	.000

Source: Charness, Karni, and Levin (2007).
Notes: The table shows error rates in an experiment in which the choice to draw from one urn first-order stochastically dominates the choice to draw from the other. (See text for a description of the experiment.) We only consider choices after a successful first draw, as we do not have observations for the CD and D cases after unsuccessful first draws.

not know before drawing which color will pay off, with the first draw (unpaid, informational only) made automatically from the left urn. In this way, there is no sense of success or failure (and corresponding emotions) upon observing the color of the ball drawn. Removing the psychological affect in this way was found to substantially reduce the error rate in Charness and Levin (2005). A third treatment performs the Bayesian updating for the subjects, a fourth treatment eliminates the compound lottery, and a fifth treatment only considers dominance (drawing from an urn with six good balls out of nine or an urn with five good balls out of nine). Table 1 shows the corresponding error rates.

Since first-order stochastic dominance is a very basic principle, it is clear that these refusals to switch are violations of rationality. In all cases, the error rate goes down as the number of people in the decision-making group increases. In the case of dominance, the rate goes to a flat zero.

Charness, Karni, and Levin (2010) consider the Linda paradox, where this question is asked:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable:

(*a*) Linda is a bank teller.
(*b*) Linda is a bank teller and is active in the feminist movement.

---

ball from the left urn given that the first draw was a *white* ball is $(1/3 \times 2/3) + (2/3 \times 1/3) = 4/9$, while the probability of drawing a black ball from the right urn is only $1/3$.

Table 2

Violations of the Conjunction Rule in an Experiment Undertaken with Individuals, Pairs, and Trios

Study	Details	Incorrect answers/ total sample	Error rate (percent)
Individuals			
T&K, 1983	UBC undergrads, no incentives	121/142	85.2
CKL, 2010	UCSB students, singles, no incentives	50/86	58.1
CKL, 2010	UCSB students, singles, incentives	31/94	33.0
CKL, 2010	UCSB students, total singles	**81/180**	**45.0**
Pairs			
CKL, 2010	UCSB students, in pairs, no incentives	27/56	48.2
CKL, 2010	UCSB students, in pairs, incentives	5/38	13.2
CKL, 2010	UCSB students, total in pairs	**32/94**	**34.0**
Trios			
CKL, 2010	UCSB students, in trios, no incentives	10/39	25.6
CKL, 2010	UCSB students, in trios, incentives	5/48	10.4
CKL, 2010	UCSB students, total in trios	**15/87**	**17.2**

Source: Charness, Karni, and Levin (2010); Tversky and Kahneman (1983).
Notes: This question was asked in the experiment: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable: (a) Linda is a bank teller. (b) Linda is a bank teller and is active in the feminist movement. (Since condition *b* imposes an extra restriction, it quite clearly cannot be more probable than *a*.) UBC is the University of British Columbia; UCSB is the University of California, Santa Barbara.

Since condition *b* imposes an extra restriction, it quite clearly cannot be more probable than *a*. And yet Tversky and Kahneman (1983) report that 85 percent of respondents answer *b*. This seems a shocking violation of rational choice, no doubt due to cognitive limitations. The question was asked with and without incentives for a correct answer; people in groups consulted with each other, but then made individual decisions. Table 2 presents the data from the study for singles, pairs, and trios.

Once again, we see a clear pattern of reductions in the error rate as the number of people in the group grows. For example, without incentives the error rate drops from 58.1 percent with singles to 48.2 percent with pairs to 25.6 percent with trios. We also note that people do far better when they are provided with financial incentives, perhaps the more realistic case.

We close this section with two experimental results in games where the issue is cognitive ability. Cooper and Kagel (2005) study the "limit-pricing game" where one player acting as a market incumbent with either high or low costs of production has to decide on an output level before another player acting as a potential entrant makes a decision about market entry. In this setting, game-theoretic considerations suggest that the incumbent should choose the "limit-pricing" output with higher

quantities and thus lower prices than would otherwise prevail, in order to deter market entry of the potential entrant, which could lead to still-lower prices. Indeed, Cooper and Kagel find that groups (of two persons each) play strategically far more often and thus are more successful in deterring market entry. This is particularly true in situations where the market parameters (through cost functions) change, in which cases groups are faster in learning the new "limit-pricing" output to deter market entry.

Finally, another example of how groups often see more deeply into a strategic situation is the two-person "company takeover game." In this game, a seller has a single item to sell. The item has a specific value to the seller, which the seller knows. However, the item will be worth 50 percent more than that to the buyer, but the buyer knows only a distribution of potential values for the seller. If the bid is at least as large as the seller's value, the buyer acquires the company after paying the bid. The optimal bid is zero, yet the vast majority of buyers fail to condition their bids on winning, and so select a positive bid (say, the expected value of that distribution).[8] An insightful bidder will recognize that potential values (seller values) above the bid are irrelevant, and so will condition her bid appropriately. This set-up is effectively a form of the "winner's curse," where the winner of an auction loses money. Casari, Zhang, and Jackson (2010) analyze group and individual behavior in this game. They find that groups fall prey to the "winner's curse" of overbidding significantly less often than individuals do, by a margin of about 10 percentage points. A similar finding of less overbidding by groups (by reducing their bids in a contest by about 25 percent) is reported in Sheremeta and Zhang (2010). In both papers, groups learn to reduce their bids from communication inside the group, indicating that groups are better in learning rational bidding strategies than individuals.

These examples (and others in the online Appendix) are rather compelling in illustrating that group choices in decision-making environments characterized by cognitive limitations (bounded rationality) are closer to the predictions of standard theory than are individual choices. These findings let us conclude that groups are more rational decision makers in the sense that economists have defined.

Lesson Two: Groups Can Help with Self-Control and Productivity Problems

Nearly everyone has self-control problems, such as procrastination, not exercising despite the lasting benefits of doing so, and being unable to control one's spending to save money. A lack of self-control or even motivation is also often found in the workplace, so that productivity is far from optimal. People engage in

---

[8] It is easy to show that the optimal bid is zero. Suppose one bids x from the interval $[0, 100]$. Assuming a uniform distribution, the average relevant seller value is not 50, but is instead $x/2$, since values above x lead to no sale. Thus the expected value to the buyer conditional on acquiring the company is 50 percent more or $3x/4$, so one loses $x/4$ on average, and choosing $x = 0$ is best.

a wide variety of commitment mechanisms to cope with these issues. For example, researchers quite often employ the commitment device known as co-authorship. One does not wish to let down a co-author (who presumably produces!), so one works harder. In a sense, this form of production is enhanced by being in a group. In this section, we present evidence from experimental and empirical studies that suggest that group decision-making and group membership can help to alleviate these self-control problems.

The evidence in this embryonic area is limited. It is difficult to observe self-control problems in the laboratory, so the experimental evidence on this topic comes from field experiments.[9] One such experiment was conducted by Falk and Ichino (2006). They let subjects perform a real-effort task, which was to put letters into envelopes for a mass mailing. In one condition, subjects had to perform the task alone in a room, while in another condition there were two subjects in the room, and both could easily watch the performance of the other. Falk and Ichino find that in the condition where groups of subjects were working, average productivity was 16 percent higher than in the isolated condition, indicating that peer effects in the group had a positive impact on productivity. Mas and Moretti (2009) also report such positive spillovers in a supermarket chain where the introduction of high-productivity workers into shifts increased the average individual productivity. While in the previous two examples, the wages of subjects were independent of their coworkers, Hamilton, Nickerson, and Owan (2003) examined how productivity in a garment factory in California changed when the plant shifted from an individual piece-rate to a group piece-rate production system (where a group member's wage did depend on the other group members' performance). While the problem of free-riding in groups (Holmstrom 1982) might decrease average productivity, Hamilton, Nickerson, and Owan (2003) find that the adoption of a group payment scheme at the plant improved worker productivity by 14 percent on average, even after controlling for systematic selection of high-ability workers into work groups. Interestingly, their data also reveal that an increase in a group's heterogeneity in ability levels increases productivity.

Babcock and Hartman (2011) investigate peer effects at the level of individual connections, and leverage the approach to shed light on peer mechanisms. In a field experiment with college freshmen, they elicited friendship networks and offered monetary incentives in some treatments for using the recreation center. Their main findings are that treated subjects with treated best friends put forth significantly more effort toward the incentivized task than do treated subjects with control best friends. The peer effect is about 20 percent as large as the direct individual effect of the incentive. There is also clear evidence of a mechanism: subjects coordinate with

best friends to overcome pre-commitment problems or reduce effort costs. Their results highlight subtle peer effects and other mechanisms that often go undetected.

In a related paper, Babcock, Bedard, Charness, Hartman, and Royer (2012) find evidence that pairing people helps to overcome problems with exercising and studying. In a field experiment involving studying and a field experiment involving exercise, large team effects operate through social channels. These experiments feature exogenous team formation and opportunities for repeated social interactions over time; one suspects that the effects would be substantially larger with endogenous group formation. In any case, in the pay-for-study intervention, people assigned to the team treatment frequented the study room considerably more often than people assigned to the individual treatment. The team-compensation system induced agents to choose their effort as if they valued a marginal dollar of compensation for their teammate from two-thirds as much to twice as much as they valued a dollar of own compensation. The paper concludes that the social effects of monetary team incentives can be used to induce effort at significantly lower cost than through direct individual payment.

Recent evidence from microfinance suggests that the frequency of meeting with others to discuss micro-loans is positively associated with repayment rates, thus helping to avoid self-control problems due to a wish for immediate gratification (Laibson 1997), which increases default risks. While the effects of group liability—where borrowers are organized in groups in which they are the guarantors of each other's loans—on default rates have been diverse (Armendariz de Aghion and Morduch 2005),[10] Feigenberg, Field, and Pande (2011) show that more frequent meetings of Indian microfinance borrowers lead to substantially lower default rates. People in a group that met once per month were 3.5 times more likely to default on a second loan than people in a group that met once per week. While this study does not provide direct evidence that people who met in groups default less frequently than people who did not (although extrapolation suggests that this is the case), it does appear that these meetings generated a form of economically valuable social capital that promoted more trustworthy behavior. In fact, there was considerably more external social interaction amongst members of the weekly group than amongst members of the monthly group. In this sense, organizing people into groups that meet frequently can enhance responsible behavior.

Lesson Three: Groups May Decrease Welfare Because of Stronger Self-interested Preferences

In the first two lessons, we have argued that decision making in groups leads to choices that are closer to predicted choices under the standard assumptions of

---

[10] In a carefully controlled natural field experiment on group versus individual liability in microfinance credits in the Philippines, Giné and Karlan (2011) do not find a difference in repayment rates between group and individual liability contracts.

Table 3
Social Welfare in a Trust Game
(as a fraction of the maximum possible payoff)

		Second-mover	
		Individual	*Group*
First-mover	Individual	0.77	0.84
	Group	0.69	0.62

Source: This is a trust game described in Kugler, Bornstein, Kocher, and Sutter (2007).
Note: Social welfare is the actual payoff per person divided by the maximum possible payoff.

rationality and that help individuals to overcome or at least contain their behavioral biases. While all of this seems like a desirable influence of group decision-making, we have not yet addressed how group decision-making may affect social welfare as we have defined it above (as total social material payoffs). We attend to this issue here, showing that decision making in groups may, in fact, be detrimental for social welfare in specific situations, whereas it is good for social welfare in others. Because the evidence in this relatively young field of research is still emerging, we are not yet able to provide a definitive taxonomy of when group decision-making is good for welfare and when it is bad, but we can lay some cornerstones upon which such a taxonomy could be built in the future.

We start with evidence from a game originally termed "the investment game" but now more commonly known as "the trust game." In this game, the first player can send an amount $x \leq c$ to a second player. The second player receives $3x$, and can send back any (non-tripled) amount $y \leq 3x$, which finishes the game. In this setting, the standard game-theoretic prediction is that the first player won't expect to get anything back, and so will send nothing. Given that an increase in the amount x is associated with higher social welfare (as the sum of payoffs for both players), the standard prediction is associated with the least efficient outcome.

Kugler, Bornstein, Kocher, and Sutter (2007) have run a trust game where either individuals, or groups of three subjects each, were in the role of first- or second-mover. They find that groups send significantly smaller amounts (by about 20 percentage points) as first-movers, and also return on average smaller amounts (although this second result was statistically insignificant). Hence, group choices are closer to the standard rationality paradigm. Table 3 shows social welfare in the four different conditions in the experiment as a fraction of the maximum possible payoff per subject. If first-movers are groups, social welfare is significantly smaller. Since second-movers are only making redistributive choices, they do not affect social welfare.[11]

---

[11] Cox (2002) finds that groups as second-movers return significantly smaller amounts than individuals do. Again, this does not affect total social payoffs, since second-movers only redistribute money.

Instead of using group decision-making, Song (2008) has studied how group representatives make decisions on behalf of their group in a trust game. This means that the representative had to make a decision that determined the outcome of a three-person group. Song finds that group representatives send about 20 percent less as first-movers and return about 40 percent less as second-movers than individuals who decide only for themselves. These results support the earlier work of Kugler, Bornstein, Kocher, and Sutter (2007) on the negative effect of group decision-making on social welfare when trust is crucial to increase social welfare.

The "centipede game" can be viewed as a multistage version of the trust game. There are two stakes on the table: one large and one small. Players must decide either to pass the stakes to the other player, at which point both stakes increase in size, or end the game by taking the larger stake for themselves and giving the smaller stake to the other player. The payoffs are arranged such that if one passes the stakes in a particular stage and the opponent immediately ends the game in the next stage, one receives less than if one had taken the payoff and not passed the stakes.The centipede game is played for a limited number of rounds. Thus, backward induction suggests that players should end the game earlier, rather than run the risk of getting a lower payoff in the event that the other player "takes" at the next move. Figure 3 displays the centipede game used in a study by Bornstein, Kugler, and Ziegelmeyer (2004) in which they let individuals play against individuals, and groups (of three subjects each) against groups. They find that individuals' median action is to "take" at node 5, while the median action of groups is to "take" at node 4. The difference is statistically significant and yields also significantly smaller payoffs for group members (50 on average) than for individuals (58 on average). Hence, the evidence shows a similar pattern as in the trust game: group play is more likely to conform to the rationality standard of game theory, but as a result, group play is also less likely to reap the potential efficiency gains.

As a final piece of evidence that group behavior may be bad for social welfare, we refer to a classic prisoner's dilemma. Of course, a prisoner's dilemma game is the familiar setting in which each of two players will find it a dominant strategy to defect, but if they can coordinate on cooperation, their combined payoff will be larger. Charness, Rigotti, and Rustichini (2007) study how individuals play this game on behalf of groups: that is, when they are making (individual) choices in front of their group members and when their actions influence the other group members' payoffs (referred to as "payoff commonality"). They find that cooperation rates go down considerably and significantly when individuals play this game against an out-group member in front of their in-group and when payoff commonality applies. Hence, while defection is the self-interested choice here, group membership makes this choice more frequent, but as a consequence social welfare is reduced. In sum, the evidence summarized so far suggests that in trust games, centipede games, and prisoner's dilemma games (all of which share the characteristic that they have a unique and socially inefficient, pure-strategy Nash equilibrium) group

Figure 3
A Centipede Game
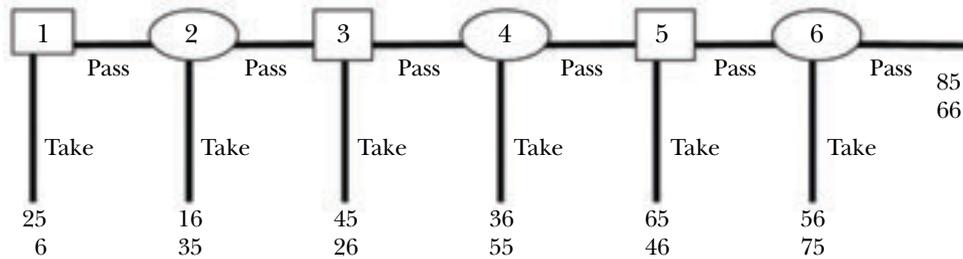
Source: Bornstein, Kugler, and Ziegelmeyer (2004).
Notes: Player 1's decision nodes are denoted by squares, and Player 2's by circles. At the start of this game, the large stake is 25 and the small stake is 6. Each time a player passes, both stakes are increased by 10. At each terminal node, the top number shows the payoff for Player 1 and the bottom for Player 2 if the game ends at that stage.

decision-making and group membership decrease social welfare, because groups show too little trust regarding cooperation from their interaction partners.

This negative effect of groups on social welfare does not generalize to all games, however. In particular, there is strong evidence that in games with multiple pure-strategy equilibria—commonly referred to as coordination games—group decision-making helps achieve efficient coordination, thus increasing social welfare.

Charness, Rigotti, and Rustichini (2007) consider a battle-of-the-sexes game. This is a 2 × 2 game often described with a story like this one: A couple agrees to get together, but they cannot remember where they agreed to meet. Both parties know that the husband preferred to attend a certain sports event and the wife preferred to attend a certain play. Both parties receive higher benefits if they coordinate on a location, yet they cannot communicate with each other. This setting has two pure-strategy equilibria where both parties attend the same location, either the sports event or the play.[12] Efficiency in this game requires successful coordination (avoiding the outcomes in which the couple ends up in different places). Charness, Rigotti, and Rustichini (2007) show that *salient* group membership (one person in the pair plays in front of an audience of one's group members) significantly increases the rate of successful coordination compared to the rate in a situation without salient group membership. In this case, salient group membership leads to better social outcomes.

Some coordination games have multiple equilibria that are Pareto-ranked— that is, some equilibria are more efficient than others. For example, the "weakest link" game studied in Feri, Irlenbusch, and Sutter (2010) shares this feature, and it works like this: There are five players, which can be either individuals or groups with

---

[12] There is also a mixed-strategy equilibrium.

Figure 4

Effort Levels of Individuals and Groups in a Weakest Link Game
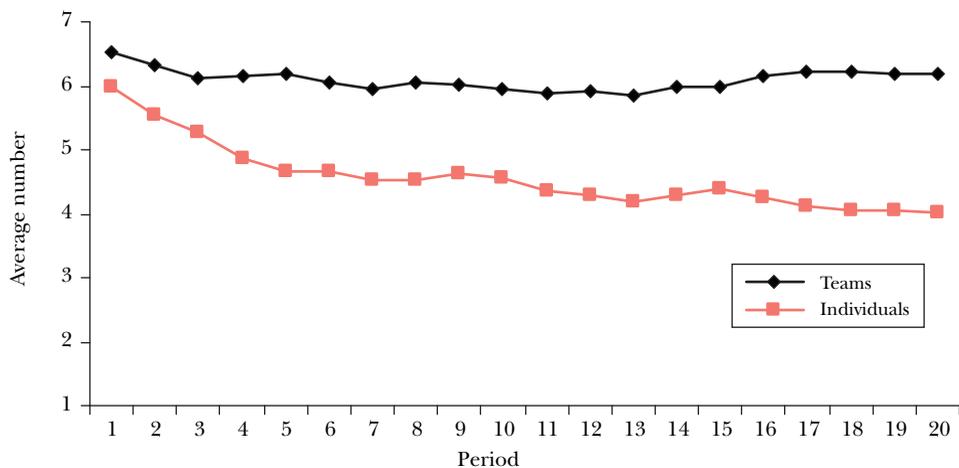
Notes: This game, denoted *WL-BASE*, is described in Feri, Irlenbusch, and Sutter (2010). There are five players, which can be either individuals or groups with three members each. Each player (either an individual or a three-person group) chooses an effort level between 1 and 7. The payoff each player receives gets higher if they all choose to exert more effort, but it also gets lower—at a faster rate—the lower the minimum choice (or "weakest link") of all players.

three members each. Each player (either an individual or a three-person group) chooses an effort level between 1 and 7 (group members may communicate briefly first). The payoff each player receives gets higher if they all choose to exert more effort, but it also gets lower—at a faster rate—the lower the minimum choice (or "weakest link") of all players. In this setting, it turns out that any setting where all the players choose the same level of effort will be an equilibrium. The biggest payoffs for all players together will arise if everyone coordinates on a high level of effort. But the weakest-link dynamic tends to push toward coordinating on a lower level of effort. Feri, Irlenbusch, and Sutter find that the three-player groups not only play more efficient high-effort equilibria more often than individuals, but also are more successful in avoiding miscoordination (which in this case means picking different effort levels). Figure 4 shows the average effort levels across 20 periods for individuals and groups, indicating a large and significant difference in the ability to coordinate on more efficient outcomes. Social welfare is on average 24 percent higher when groups play this coordination game than when individuals make decisions.

In short, the effect of group decision-making on social welfare can go in either direction. The pattern emerging from the evidence seems to indicate that more rational choices of groups decrease social welfare when games have a unique pure-strategy equilibrium (with a dominant strategy, in fact), but that groups are more successful in coordinating on more efficient equilibria when a multiplicity of equilibria exist. The common denominator for these seemingly divergent effects of

group decision-making may be that groups put more weight on own payoffs than do individuals (something discussed also in the next section). Studying the learning of groups and individuals, Feri, Irlenbusch, and Sutter (2010) find that groups are more sensitive to the attractions of different strategies and take into account more strongly the potential payoffs of previously not-chosen strategies. These learning characteristics of groups imply that payoffs play a significantly larger role in determining their choice probabilities than they do for individuals, leading to a higher frequency of choosing dominant strategies in trust games ("do not trust"), centipede games ("take"), or prisoner's dilemma games ("defect"), but also to a higher frequency of choosing more efficient equilibria in coordination games.

Sources of Differences in Individual and Group Decisions

Why might groups behave in a more rational manner than individuals? We explore three possible reasons: 1) multiple brains are better at seeking answers; 2) multiple brains are better at anticipating the actions of other parties and thus better at coordinating behavior with what other parties are likely to do; and 3) groups may be more likely than individuals to emphasize monetary payoffs over alternative concerns, such as fairness or reciprocity towards another player.

Our first possible explanation for differences between individuals and groups is that groups can potentially benefit from having multiple brains. In some cases, this may lead to better decisions in the sense of avoiding errors. In addition to the examples given in Lesson One, consider an information cascade game. Here players receive a private signal and then announce a public belief in sequential order: for example, players might look at one marble drawn from a bag, and then announce their belief as to whether the bag is two-thirds white marbles or two-thirds black marbles. Later players must then compare their own private signal to the public beliefs of others. In an information cascade, players ought to disregard their private information and instead follow the belief being expressed by many others at some stage of the game. Fahr and Irlenbusch (2011) find that groups make fewer mistakes in an information cascade experiment than individuals (and thus earn more money).[13] Evidence from psychology supports the argument that social interaction improves the decision-making process. For instance, in letters-to-numbers problems, where a random coding of the letters A–J to the numbers 0–9 needs to be solved, groups do much better than individuals by taking about 30 percent fewer trials to solve the problems (Laughlin, Bonner, and Miner 2002). Likewise, in the "Wason selection task," developed to test whether individuals employ the rules of formal logic when

---

[13] Also in information cascade experiments by Alevy, Haigh, and List (2007), professional traders were shown to be better able to discern the quality of public signals. One possible explanation for the superiority of professional traders over college students might be that professional traders are more used to being in a group, so they make better decisions, an interpretation that would be consistent with the findings by Fahr and Irlenbusch (2011).

testing conditional statements of the form "if *p*, then *q*," groups have solution rates of 50 percent while individuals have solution rates of 11 percent (Maciejovsky and Budescu 2007). The Wason selection task is an example of a "truth wins" problem: that is, a problem where the solution is difficult to reach without grasping a specific insight, but then the solution is easily explained to another individual. In such cases, groups can be expected to solve the problem with higher probability. Consider that a fraction *p* of all individuals has the specific insight to solve the problem, then the likelihood that a group with *n* members solves the problem is $1 - (1 - p)^n$, which is larger than *p* (if $p < 1$). The likelihood $1 - (1 - p)^n$ is often referred to as the "truth-wins benchmark." While groups typically do better than individuals in such insight problems, they rarely meet or exceed the truth-wins benchmark.[14]

In an interesting experiment from the psychology literature, groups actually beat this benchmark. Michaelson, Watson, and Black (1989) grouped together students in a class (average group size of six) and asked them to answer questions based on assigned reading, with the scores counting towards the course grade. These tasks ranged from recalling specific concepts from the reading to ones requiring higher cognitive ability and a deeper understanding to being able to synthesize concepts. The key comparison was between the highest score of any individual in a group and the average score of the group on the task; the notion behind this comparison is to test the view that, in an organizational context, group decisions will be better than the decisions of the most knowledgeable group member. In fact, a remarkable 97 percent of all groups outperformed their best member. Each person first completed the task individually and then retook the test as a member of a group that could have internal discussions. Group scores were compared with the highest score for any individual in the group. In the economics literature, choices made in the Cooper and Kagel (2005) limit-pricing game and in the Maciejovsky and Budescu (2007) Wason selection task provide examples where groups do better than the truth-wins benchmark.

A second possible reason why groups make more rational decisions than individuals, especially in interactive games, is that group members are better able to put themselves into the shoes of their competitors when discussing their own strategy. It seems that the need to discuss the game with another group member often leads to a discussion regarding how the group members would play the game, making it a salient feature then to consider the other player's available strategies and payoffs more extensively than individuals would do (Cooper and Kagel 2005). For this reason, groups can be better prepared to anticipate the actions of other players. From there, it is only a short step to think about the best reply to one's own expectation about the opponent's most likely strategy. As a consequence, group behavior is pushed towards the standard game-theoretic predictions. This insight is consistent with what has been observed in the limit-pricing game of Cooper and Kagel (2005). Further support is presented in Sutter, Czermak, and Feri (2010).

---

[14] Meaning that their solution rates stay below $1 - (1 - p)^n$ but remain above *p*.

They let individuals and groups make choices in simple two-player games (with unique pure-strategy, Pareto-inefficient Nash equilibria). Groups play the Nash equilibrium in these games about 10 percentage points more frequently than individuals, and the main reason is that they expect their opponent to play the Nash equilibrium more frequently than individuals expect this from individuals. Accordingly, groups more often play the equilibrium as a best response to their own beliefs.

A third reason why groups may behave "less behaviorally" than individuals is that groups may be more concerned with their own group's monetary payoffs and thus disregard more frequently the payoffs of the other player. Communication within groups may change an individual's reference point for optimization. Instead of maximizing own payoffs, individuals may consider the joint payoff (or welfare) of those engaged in the discussion as the appropriate target for optimization. Psychologists have long been emphasizing such an effect of communication: Elster (1986, pp. 112–113), for instance, has suggested that it is "pragmatically impossible to argue that a given solution should be chosen just because it is good for oneself. By the very act of engaging in a public debate . . . one has ruled out the possibility of invoking such reasons. To engage in discussion can in fact be seen as one kind of self-censorship, a pre-commitment to the idea of rational decision." By rational decision, however, Elster (1986) refers to decisions which are advantageous for the *group* of communicating subjects as a whole, but not necessarily aligned with (and sometimes even contrary to) the interests of other players in the opponent group. Such an argument links our discussion to the long-standing literature on in-group/out-group effects. (For an overview from an economic perspective, interested readers might start with Chen and Li 2009.) By design, group decision-making creates an in-group—one's own group—and an out-group—with whom the own-group is interacting. Social psychology has coined the term "discontinuity effect" (for example, Schopler et al. 2001) to describe the fact that, typically, groups act more competitively and more selfishly when interacting with other groups than when individuals interact with individuals.

Conclusion

The existing literature that compares group and individual decision-making provides considerable evidence that groups make choices that are more rational in a standard game-theoretic sense than those of individuals. As a result, group decision-making and being a member of a group can overcome cognitive biases and limitations. However, making decisions in groups does not always lead to increases in social welfare, which raises the question: Under which conditions is individual or group decision-making better for society as a whole? We have identified several games (with unique equilibria) where individual decision-making yields higher welfare, while in coordination games (with multiple equilibria), groups achieve more efficient outcomes.

Since group decision-making is present in a wide variety of economic environments, this issue has considerable practical relevance. Generally, decision making in groups seems to be most effective when there is a good degree of diversity in the group and when the environment is a participatory one in which diverse ideas can be expressed (rather than an environment with a dominant and intimidating personality). For example, any single individual group member could have an insight that sheds light on what would otherwise be a blind spot for the group; it pays to broaden the base. Still, it seems best to have groups of modest size, so that interior coordination problems and "social loafing"—in this case, reduced effort—are manageable. As Surowiecki (2004, pp. 190–91) wrote: "If small groups are included in the decision-making process, then they should be allowed to make decisions. If an organization sets up teams and then uses them for purely advisory purposes, it loses the true advantage that a team has: namely, collective wisdom." It is noteworthy, however, that it remains to be determined what constitutes an ideal group size. A useful starting point here is Forsyth's (2006) work on group size and performance. We suspect that the optimal size of the group will depend on factors such as the complexity of the decision, but more research is clearly needed here.

Some other open issues for future research include the influence of different communication media on group decisions. Do group dynamics change when video calls substitute for face-to-face communication? Another relatively unexplored area is the effect of internal conflicts on the rationality and character of group decisions: that is, what happens when the payoffs to members of a group are not identical? Groups can be a way of diffusing decision-making and avoiding responsibility, but they can also be a powerful force for more careful and productive decisions. Ultimately, the goal of comparing individual and group decision-making is to identify the contexts and types of decisions where each is likely to work best.

References

Alevy, Jonathan E., Michael S. Haigh, and John A. List. 2007. "Information Cascades: Evidence from a Field Experiment with Financial Market Professionals." *Journal of Finance* 62(1): 151–80.

Armendariz de Aghion, Beatriz, and Jonathan Morduch. 2005. *The Economics of Microfinance.* MIT Press.

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer. 2012. "Letting Down the Team: Social Effects of Team Incentives." Unpublished paper.

Babcock, Philip, and John Hartman. 2011.

"Coordination and Contagion: Peer Effects and Mechanisms in a Randomized Field Experiment." Unpublished paper.

Bornstein, Gary, Tamar Kugler, and Anthony Ziegelmeyer. 2004. "Individual and Group Decisions in the Centipede Game: Are Groups More 'Rational' Players?" *Journal of Experimental Social Psychology* 40(5): 599–605.

Camerer, Colin F. 2003. *Behavioural Game Theory: Experiments in Strategic Interaction.* Princeton University Press.

Casari, Marco, Jingjing Zhang, and Christine

Jackson. 2010. "Do Groups Fall Prey to the Winner's Curse?" IEW Working Paper 504, Institute for Empirical Research in Economics, University of Zurich.

Charness, Gary, Edi Karni, and Dan Levin. 2007. "Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violations of First-Order Stochastic Dominance." *Journal of Risk and Uncertainty* 35(2): 129–48.

Charness, Gary, Edi Karni, and Dan Levin. 2010. "On the Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda." *Games and Economic Behavior* 68(2): 551–56.

Charness, Gary, and Dan Levin. 2005. "When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect." *American Economic Review* 95(4): 1300–1309.

Charness, Gary, Luca Rigotti, and Aldo Rustichini. 2007. "Individual Behavior and Group Membership." *American Economic Review* 97(4): 1340–52.

Chen, Yan, and Xin Li. 2009. "Group Identity and Social Preferences." *American Economic Review* 99(1): 431–57.

Cooper, David J., and John H. Kagel. 2005. "Are Two Heads Better Than One? Team versus Individual Play in Signaling Games." *American Economic Review* 95(3): 477–509.

Cox, James C. 2002. "Trust, Reciprocity, and Other-Regarding Preferences: Groups vs. Individuals and Males vs. Females." In *Advances in Experimental Business Research*, edited by Rami Zwick and Amnon Rapoport, 331–50. Dordrecht: Kluwer Academic Publishers.

Elster, Jon. 1986. "The Market and the Forum: Three Varieties of Political Theory." In *Foundations of Social Choice Theory: Studies in Rationality and Social Change,* edited by J. Elster and A. Hylland, 103–132. Cambridge University Press.

Fahr, René, and Bernd Irlenbusch. 2011. "Who Follows the Crowd—Groups or Individuals?" *Journal of Economic Behavior and Organization* 80(2): 200–209.

Falk, Armin, and Andrea Ichino. 2006. "Clean Evidence on Peer Effects." *Journal of Labor Economics* 24(1): 39–57.

Feigenberg, Benjamin, Erica Field, and Rohini Pande. 2011. "The Economic Returns to Social Interaction: Experimental Evidence from Microfinance." http://www.economics.harvard.edu/faculty /field/files/Social_Capital_feb10_ef_rp.pdf.

Feri, Francesco, Bernd Irlenbusch, and Matthias Sutter. 2010. "Efficiency Gains from Team-Based Coordination—Large-Scale Experimental Evidence." *American Economic Review* 100(4): 1892–1912.

Forsyth, Donelson R. 2006. *Group Dynamics,* 4th edition. Belmont, CA: Thomson Higher Education.

Giné, Xavier, and Dean S. Karlan. 2011. "Group versus Individual Liability: Short and Long Term Evidence from Philippine Microcredit Lending Groups." June. http://karlan.yale.edu/p/Group versusIndividualLending.pdf.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy* 111(2): 465–97.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13(2): 324–40.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money.* Macmillan Cambridge University Press for the Royal Economic Society.

Kocher, Martin G., Sabine Strauss, and Matthias Sutter. 2006. "Individual or Team Decision-Making—Causes and Consequences of Self-Selection." *Games and Economic Behavior* 56(2): 259–70.

Kocher, Martin G., and Matthias Sutter. 2005. "The Decision Maker Matters: Individual versus Group Behavior in Experimental Beauty-Contest Games." *Economic Journal* 115(500): 200–223.

Kugler, Tamar, Gary Bornstein, Martin G. Kocher, and Matthias Sutter. 2007. "Trust between Individuals and Groups: Groups are Less Trusting Than Individuals But Just as Trustworthy." *Journal of Economic Psychology* 28(6): 646–57.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics* 112(2): 443–77.

Laughlin, Patrick R., Bryan L. Bonner, and Andrew G. Miner. 2002. "Groups Perform Better Than the Best Individuals on Letter-to-Numbers Problems." *Organizational Behavior and Human Decision Processes* 88(2): 606–620.

Levine, John M., and Robert L. Moreland. 1998. "Small Groups." In *The Handbook of Social Psychology,* 4th edition, vol. 2, edited by Gilbert, D. T., S. T. Fiske, and G. Lindzey, 415–69. McGraw-Hill.

Levitt, Steven, and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21(2): 153–74.

List, John A. 2011. "Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off." *Journal of Economic Perspectives* 25(3): 3–16.

Maciejovsky, Boris, and David V. Budescu. 2007. "Collective Induction without Cooperation? Learning and Knowledge Transfer in Cooperative

Groups and Competitive Auctions." *Journal of Personality and Social Psychology* 92(5): 854–70.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review* 99(1): 112–45.

Michaelson, Larry K., Warren E. Watson, and Robert H. Black. 1989. "A Realistic Test of Individual versus Group Consensus Decision Making." *Journal of Applied Psychology* 74(5): 834–39.

Schopler, John, Chester A. Insko, Jennifer Wieselquist, Michael Pemberton, Betty Witcher, Rob Kozar, Chris Roddenberry, and Tim Wildschut. 2001. "When Groups Are More Competitive Than Individuals: The Domain of the Discontinuity Effect." *Journal of Personality and Social Psychology* 80(4): 632–44.

Sheremeta, Roman M., and Jingjing Zhang. 2010. "Can Groups Solve the Problem of Overbidding in Contests?" *Social Choice and Welfare* 35(2): 175–97.

Song, Fei. 2008. "Trust and Reciprocity Behavior and Behavioral Forecasts: Individuals versus Group-Representatives." *Games and Economic Behavior* 62(2): 675–96.

Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations.* Doubleday.

Sutter, Matthias. 2005. "Are Four Heads Better Than Two? An Experimental Beauty-Contest Game with Teams of Different Size." *Economics Letters* 88(1): 41–46.

Sutter, Matthias, Simon Czermak, and Francesco Feri. 2010. "Strategic Sophistication of Individuals and Teams in Experimental Normal-Form Games." *IZA Discussion Paper 4732.*

Tversky, Amos, and Daniel Kahneman. 1983. "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90(40): 293–315.

Deleveraging and Monetary Policy: Japan Since the 1990s and the United States Since 2007[†]

Kazuo Ueda

The U.S. economy in the aftermath of the Great Recession that started in 2007 has a number of similarities with Japan's experience since the early 1990s, at least on the surface. Both economies experienced an unsustainable boom in real estate prices along with high stock market valuations, and when the bubble burst, many households and financial institutions found themselves in dire straits. One major lesson from this experience is that deleveraging attempts by individual economic agents in the aftermath of large financial imbalances can generate significant negative macroeconomic externalities. In Japan's case, a negative feedback loop developed among falling asset prices, financial instability, and stagnant economic activity. This negative feedback loop has sometimes been called "Japanization." As the U.S. economy works through a sluggish recovery several years after the Great Recession technically came to an end in June 2009, it can only look with horror toward Japan's experience of two decades of stagnant growth since the early 1990s.

Japan's deleveraging became serious because the negative feedback loop was not contained in its early stage of development. The Japanese government did not act promptly to recapitalize banks that were suffering from the erosion of their capital buffer due to their large holdings of stocks. As a result, Japan's banks only slowly recognized bad loans, while stopping lending to promising new projects. Slow but protracted asset sales resulted in a long period of asset price declines. Nonfinancial companies perceived the deterioration of their balance sheets as permanent and cut spending drastically. As Japan's economy stagnated, the total amount of bad loans turned out to be much larger than initially estimated.

■ *Kazuo Ueda is Professor of Economics, University of Tokyo, Tokyo, Japan. His email address is* ⟨*ueda@e.u-tokyo.ac.jp*⟩.

In contrast to Japan, U.S. policy authorities responded to the financial crisis since 2007 more quickly. Surely, they learned from Japan's experience. It is also important to recognize, however, that the market-based nature of the U.S. financial system, as compared to a Japanese financial sector, which is more intertwined with government and less subject to market pressures, meant that the need for government action was more apparent in the U.S. context.

When a national economy is confronted with Japanization, the central bank finds itself on the front line of policy making. As with Japan's other policymakers, the Bank of Japan's response in the 1990s was slow. As a result, the process of deleveraging became overly severe and protracted. This criticism of the Bank of Japan is not a new one: for example, Ben Bernanke (2000, see also 2003), then still a professor at Princeton University, criticized the Bank of Japan for not being more aggressive in its fight against deflation. Krugman (2012), Ball (2012), and others have argued that, in a provocative turnabout, Federal Reserve Chairman Ben Bernake has not been willing to push for the same aggressive monetary remedies for the United States that he earlier prescribed for Japan. Bernanke has responded by making two points: 1) the U.S. economic situation is objectively different, in the sense that Japan faced actual deflation in the late 1990s; and 2) the Fed has indeed pursued aggressively expansive monetary policy in a number of nonstandard ways (Federal Reserve 2012b, p. 9).

This paper does not seek to resolve the debate over the degree of consistency between what Bernanke wrote in the early 2000s and the policies that the Federal Reserve has undertaken since 2007. However, the paper does show that a rapid response by a central bank in a situation of financial crisis and economic stagnation can be a better choice than allowing a process of Japanization to drag on for years. In a weak economy, interest rates are already very low and the zero lower bound on interest rates limits a central bank's ability to stimulate the economy further. Moreover, as I will explain below, nonconventional monetary policy measures work by reducing risk premiums and interest rate spreads between long-term and short-term financial instruments. However, when a long period of economic stagnation occurs, these spreads have a tendency to decline to low levels, which then limits the effectiveness of such measures.

I will begin by describing how Japan's economic situation unfolded in the early 1990s and offering some comparisons with how the Great Recession unfolded in the U.S. economy. I then turn to the Bank of Japan's policy responses to the crisis and again offer some comparisons to the Federal Reserve. I will discuss the use of both the conventional interest rate tool—the federal funds rate in the United States, and the "call rate" in Japan—and nonconventional measures of monetary policy and consider their effectiveness in the context of the rest of the financial system.

The Deleveraging Experience

Japan experienced an enormous bubble in asset prices during the 1980s. For example, the value of the Nikkei 225 stock market index rose from 6,000 in 1980

to about 40,000 in 1989 before falling back to half of that level in 1990, and it fell further back to about 8,000 by 2003, which was still the level in early June 2012. Property values in Japan experienced a huge bubble as well, essentially doubling from the beginning to end of the 1980s, and sliding back to early 1980s levels by the early 2000s. At present, property prices in Japan have fallen about 60 percent since their peak circa 1990. Thus, both stock and property prices in Japan have been on a downward trend for more than two decades. Those interested in a discussion of the causes underlying Japan's asset price bubbles and their bursting might begin with Hoshi and Kashyap (1999) and Ueda (2000). Here, we take up that story in the early 1990s.

The U.S. economy in the 2000s experienced a bubble in housing prices, along with a sharp up-and-down-and-up-again movement in stock prices. U.S. housing prices rose 90 percent from 2000 to 2006, according to the Case–Shiller Index, but then fell by about one-third from 2006 to 2009—leaving them at about 25 percent above the price level of 2000. Although U.S. housing prices seem to have stabilized in the last couple of years, it is of course impossible to know whether they will drop further. The U.S. stock market, as measured by the Dow Jones Industrial Index, nearly doubled from mid 2002 to mid 2007, then fell to slightly below mid 2002 levels during the worst of the financial panic in early 2009, but since then has rebounded back to less than 10 percent away from its mid 2007 high.

These extreme fluctuations in asset values created severe economic problems for overleveraged households and firms, and for the financial institutions that were holding the loans. This financial crisis and the associated economic slowdown brought deflationary tendencies, but although these pressures were stronger in Japan than in the United States, they weren't the central problem. Here, I say a few words about deflation and then turn to the more substantial problem—the process of deleveraging.

Mild Deflation

Inflation in Japan has been in negative territory since 1998, but only modestly so: the cumulative decrease in Japan's consumer price index since the late 1990s (adjusted to purge the effects of the consumption tax rate hikes in 1989 and 1997) has been only about 5 percent. Thus, the classic debt-deflation dynamic—that is, deflation making it harder to repay debts, and the resulting lack of buying power leading to more deflation—has not been a major cause of Japan's economic stagnation.

The U.S. economy has largely escaped deflation since 2007, although four of the five months at the end of 2008 reported negative movements in the Consumer Price Index, and this index fell by 0.4 percent from 2008 to 2009. However, the Consumer Price Index rose 3.2 percent from 2010 to 2011, and seems to be rising at a similar pace in 2012.

Even though these mild deflationary pressures did not throw the economies of Japan or the United States into a deflationary spiral, they have hindered the effectiveness of monetary easing. The real interest rate has stayed at higher levels

than desirable, especially in Japan, and undermined the power of near-zero interest rates to stimulate the economy.

Deleveraging

Declines in asset prices in Japan since around 1990 and in the United States since 2007 have generated serious negative effects on both the Japanese and the U.S. economy. I will focus first on the initial shock the asset price declines created in Japan, then discus how the shock was amplified by economic policy mismanagement, and finally offer some comparisons with what has happened in the U.S. economy during the aftermath of the Great Recession.

At the peak of the bubble, Japan's nonfinancial firms held 34.6 percent of total stock market value in the economy and 24.4 percent of land value. Deposit-taking banks held 11.7 percent of the stocks (according to data from Japan's National Income Accounts and Flow of Funds Statistics). The large holdings of stocks by these two groups are in sharp contrast to the U.S. economy, where the corresponding figures are almost zero. Of course, U.S. banks and the so-called "shadow bank" financial institutions had huge exposures to mortgage loan-related assets and suffered dearly from the sharp fall in real estate prices since 2007. The reconciliation accounts of Japan's national income accounts show that nonfinancial firms lost 133 trillion yen in unrealized capital losses from their land holdings during 1990–91 (20.8 percent of their net worth in 1990) and 201 trillion yen from their stock holdings during 1989–91 (30 percent of their net worth in 1990). Banks lost 36.4 trillion yen from their stock holdings between 1989 and 1991, which was 35.9 percent of their net worth in 1988. Thus, the initial shock to their balance sheets was extremely large. The damage to the balance sheet of nonfinancials is consistent with a model of deleveraging by Eggertsson and Krugman (2011) whereby asset price declines worsen information asymmetries and lead to decreased financial intermediation and spending by borrowers.[1]

The decline in capital ratios in Japan's banks constrained their risk-taking behavior severely. Hanson, Kashyap, and Stein in this journal (2011) point out that generalized asset shrinkage by financial institutions generates two primary macroeconomic costs (or "external diseconomies"): credit crunches and fire sales. Japan also saw these negative effects of asset price declines, but they manifested themselves in peculiar ways. Fire sales did not come in one bout, but were spread out across a long period of time. The credit crunch, a quantitative shortage of credit at the prevailing prices, also involved a significant degree of credit misallocation. Thus, economic stagnation became a prolonged process in Japan, aggravating asset price declines and causing further financial instability.

---

[1] Ogawa and Suzuki (1998) analyze the role played by the use of land as a device to alleviate information asymmetry between lenders and borrowers in Japan. They show that firms increasingly relied on the use of land as collateral in the 1980s as land prices soared, which was one of the reasons for the sharp rise in business fixed investment during the period. Conversely, the decline in land prices since the early 1990s exerted strong negative effects on investment through this route.

Figure 1
Leverage for Japan's Nonfinancial Firms and Deposit-taking Banks
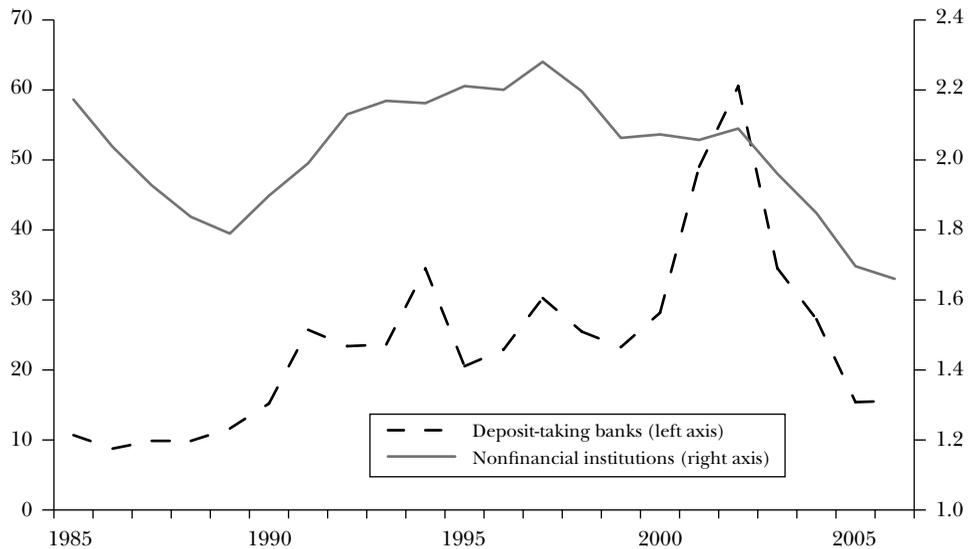
Source: Japan's National Income Accounts and Flow of Funds.
Notes: Leverage is measured by total assets with stocks and real estate at market value divided by net worth. Deposit-taking banks exclude postal savings and agricultural banks.

To gain a sense of how long the deterioration in the balance sheets in Japan lasted, Figure 1 presents estimates of leverage—that is, total assets (including loans and reserves) divided by net worth, with stocks and real estate on the asset side evaluated at market value—for nonfinancials and deposit-taking banks (excluding Japan's "postal savings accounts"). For both sectors, the leverage ratio started to increase around 1990 in response to the collapse in asset prices. The increases in the ratio continued until the late 1990s for nonfinancials and to the early 2000s for banks. It was not until the mid 2000s that the leverage ratios returned to the levels of the late 1980s.

The increases in leverage during the 1990s were largely unintentional; players in both the financial and nonfinancial sectors were attempting to deleverage, but changes in the numerator of their leverage ratios were overwhelmed by further declines in asset prices affecting the denominator of that ratio.[2] Japan's banks were large net sellers of stocks in late 1990 toward early 1991 and then again between 1996 and 2006. Nonfinancial firms were net sellers of land in the mid to late 1990s. Such asset "fire sales" that occurred under pressure to deleverage contributed to further declines in asset prices and made the deleveraging process even more

---

[2] Increases in bank leverage between the mid 1990s and early 2000s were also caused by bad-loan write-offs.

severe—a process observed again during 2007–09 in the United States for many financial instruments.

Regulatory Forbearance

The protracted nature of Japan's deleveraging substantially magnified the cost associated with the process. Decisions by Japan's regulatory authorities were a major factor behind the delay. In line with the international standards established by the Basel Committee at the Bank for International Settlements, capital adequacy regulation was introduced in Japan the early 1990s: specifically, the rules required that the ratio of the bank's capital to its risk-weighted assets must exceed 8 percent. However, the rule was poorly designed. The regulatory minimum was fixed and did not allow for fluctuations in response to the state of the economy. The definition of capital included unrealized capital gains on stocks held by banks. Thus, the sharp fall in stock prices meant a serious erosion of bank capital. Bank of Japan (2001) shows that the risk-based capital adequacy ratios for Japanese banks were barely above 8 percent in 1990–1992. Despite this, Japan's regulatory authority for years delayed making the tough decision to recapitalize the banks. One early attempt at the resolution of the bad loan problem reportedly came in the summer of 1992, when then-prime minister Kiichi Miyazawa discussed the possibility of bank recapitalization with a group of bank chief executive officers, who rejected the plan (Nishikawa 2011, pp. 137–38).

As a result, Japan's banks found it difficult to recognize and dispose of bad loans in large amounts. Instead, the disposition of bad loans became a long and protracted process, which in turn made the declines in property prices protracted as well. As property prices kept falling, giving no indication of bottoming out, nonfinancial firms holding property increasingly felt that their balance sheet deterioration was permanent. They started to repay excessive debt by cutting spending, especially investment in structures, which is the component of aggregate demand most sensitive to property prices.[3] This component of aggregate demand alone subtracted about 0.4 percent per year from Japan's GDP growth during the 1990s. In turn, property prices declined more.

The sharp fall in banks' capital buffer led to another characteristic of Japan's deleveraging process—credit misallocation. Banks were obliged to lend to "zombie" companies in order to avoid recognition of losses on their balance sheets. As a result, banks found it difficult to increase lending to more-promising firms (given the capital constraint on the expansion of total assets). New entry into banking was strictly controlled by regulators. Banks became an obstacle to creative destruction and sowed the seeds for a long period of stagnation.

There is a large literature on such misallocation of credit. Using firm-level data, Fukao and Kwon (2006) present a striking result that the productivity level of exiting firms was higher than that of staying firms in many industries in Japan during

---

[3] For details of how Japan's nonfinancial corporations deleveraged, see Figure A1 in the online Appendix available with this paper at ⟨http://e-jep.org⟩.

1994–2001. Peek and Rosengren (2005) go further by showing that Japanese banks allocated credit to severely impaired borrowers. Caballero, Hoshi, and Kashyap (2008) provide evidence of the negative effects of zombie survival on other more efficient firms in the same industry. This literature also provides a bridge between the literature emphasizing reduced productivity growth (for example, Hayashi and Prescott 2002) and the works that focus on financial factors in the analysis of Japan's stagnation.

Thus, Japan offers a typical example of the pro-cyclicality of simple capital adequacy rules as discussed by Kashyap and Stein (2004). It was, however, compounded by the absence of prompt recapitalization of banks. The pro-cyclicality took the form of a negative feedback loop among asset prices, financial instability, and the economy. The failure to react promptly was a serious mistake on the part of regulators given that they could have learned from the U.S. savings and loan crisis in the late 1980s (for an example in this journal, see Kane 1989).

In passing, I would add that through the postwar period until about the mid 1990s, Japan's regulatory authorities used the so-called "convoy approach" to the resolution of troubled financial institutions. That is, they let healthier financial institutions take over troubled ones, thus protecting depositors and other debt holders. The public's belief that the authority would continue to honor this approach was an important factor behind the absence of serious liquidity problems for Japanese banks and thus the absence of sharp fire sale pressures in the early 1990s, until circumstances changed dramatically for the worse in the late 1990s (as discussed in the next subsection).

How does the deleveraging of the U.S. banking sector since the Great Recession compare with that of Japan in the 1990s? In Figure 2, on the horizontal axis, time T = 0 is taken to be the year that was the peak of the stock market: that is, 1990 for Japan and 2007 for the United States. The solid and dotted lines show bank loan growth rates before and after the collapse in asset prices. Bank loan growth in Japan fell sharply in the early 1990s, but stayed in positive territory until the late 1990s. At this time, many of Japan's banks were still supporting zombie companies by rolling over loans. Japan's banks became earnest about bad loan disposal in year 1995 (year T + 5) onwards, as is shown by the bar graph superimposed on that figure. However, by this time bad loans had become much larger than would have been the case had they been addressed in the early 1990s. The growth rate of U.S. bank loans became sharply negative in year T + 2, but has stabilized around zero since then. So far, the U.S. pattern is consistent with much more swift adjustment of the banking sector than in Japan. But much depends on what will happen from here.[4]

The International Monetary Fund (2012) presents estimates of the leverage of U.S. large banks, which rose to 28 in December 2008 from 20 in March 2006, but

---

[4] U.S. household leverage calculated from the flow of funds statistics increased sharply in 2008 as property prices declined, but has returned to levels in the early 2000s due mainly to declines in household debt. Part of this seems to be a result of the nonrecourse nature of U.S. mortgage loans, which shifts the burden to the lenders. However, the level of debt relative to GDP still remains very high compared with the late 1990s for both households and financial institutions.

Figure 2

Bank Loan Growth Rate in Japan and the United States, and Bad Loan Disposals in Japan
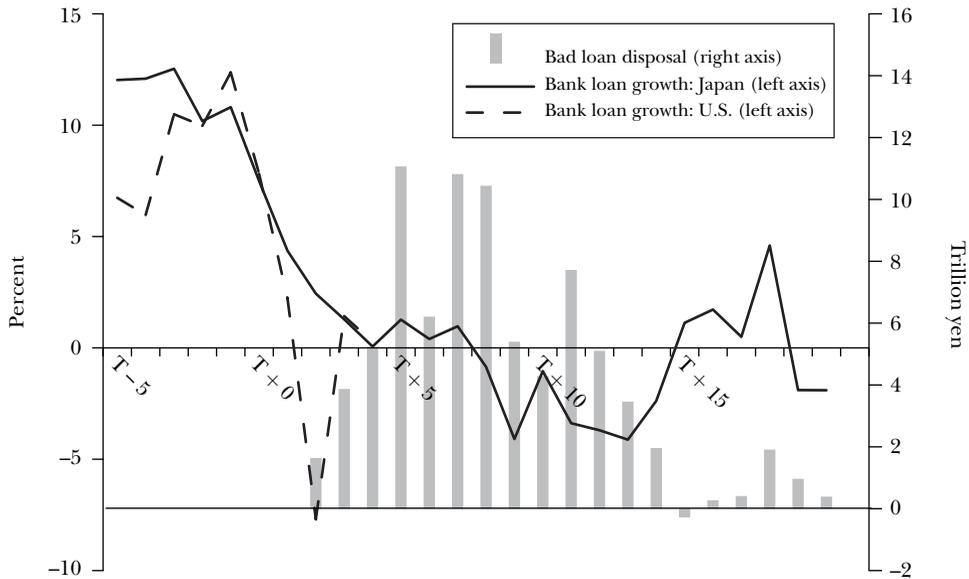
Source: Datastream, and Japan's Financial Services Agency (FSA). FSA data available at ⟨http://www.fsa .go.jp/en/regulated/npl/index.html⟩.
Note: T = 0 corresponds to 1990 for Japan and 2007 for the United States.

was down to 15 by December 2011. The sharp decline in leverage between 2008 and 2011 reflects the recapitalization by public money and contributions from retained earnings. The spreads between lending and borrowing rates have been much larger for U.S. banks than Japanese banks, and thus have helped U.S. banks more. On these measures, it appears that the deleveraging pressure remaining in the U.S. economy five years after 2007 is less serious than it was in Japan in the late 1990s and early 2000s, ten years and more after Japan's crisis erupted.

Political Dynamics of Crisis and Response

The East Asian economic crisis of 1997–98 and a hike in Japan's consumption tax rate in 1997 became a trigger for a serious financial crisis in Japan in 1997–98. A medium-sized securities company, Sanyo Securities, went under in November 1997 and defaulted on call market loans (remember, these are the overnight loans between financial institutions similar to those governed by the federal funds interest rate in the United States); it was the first time such a thing had happened in Japan during the post–World War II period. A financial panic ensued, which led to a series of bankruptcies for Japanese banks and securities firms. This event was Japan's equivalent of the Lehman bankruptcy that rocked the U.S. financial system

Figure 3
Money Market Risk Premium
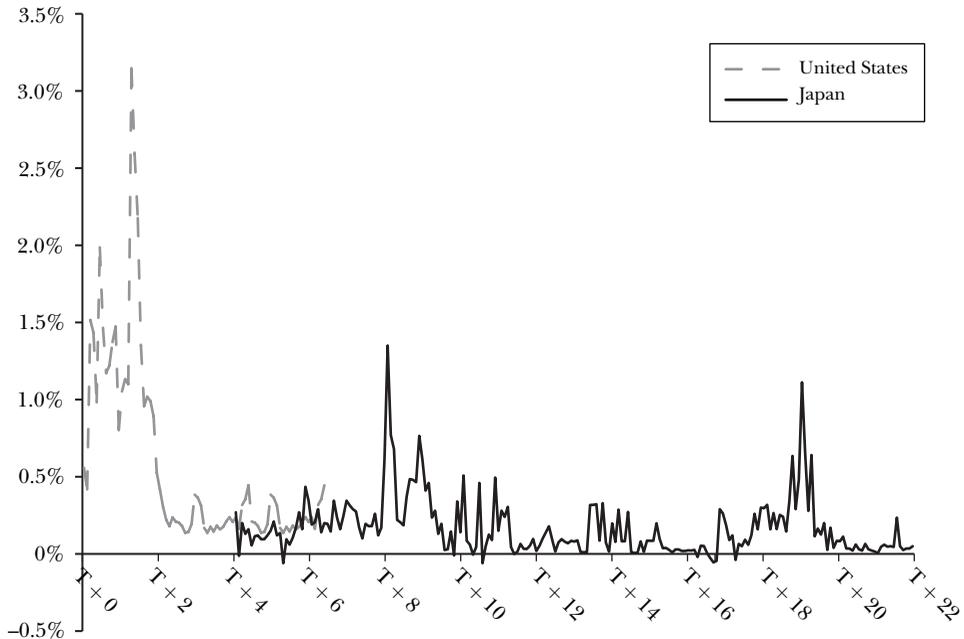
Source: Bloomberg.
Notes: The risk premium is the 3-month LIBOR (London Inter-Bank Offered Rate) minus the 3-month Treasury bill rate for the U.S., and the 3-month uncollateralized call market rate minus 3-month treasury bill rate for Japan. T = 0 stands for June 2007 for the U.S. and January 1990 for Japan. The line for Japan starts at T + 4 because the three-month call market rate is available only since 1994 (see footnote 5).

in September 2008. Japan's financial system could no longer stand the weight of mushrooming bad loans. It is also noteworthy that the Japanese authorities underestimated the consequences of a failure of a broker/dealer.

Figure 3 presents the behavior of the money market risk premium for the United States and Japan, which is a useful measure of financial instability. In the figure, T + 0 is set to June 2007 for the United States and January 1990 for Japan. A risk premium for the banking system can be measured by the difference between an interest rate that fluctuates with the perceived risk of the banking sector minus a risk-free interest rate. For the United States, the risk premium is measured by the three-month LIBOR (London Inter-Bank Offered Rate), which is the rate at which banks lend money to each other, minus the three-month Treasury bill rate. This measure of the risk premium for U.S. banks increased immediately in 2007, after property and stock prices started to fall in the United States, and then spiked higher in late 2008 and early 2009. For Japan, the risk premium is measured by the three-month uncollateralized call market rate minus Japan's three-month treasury

bill rate.[5] Despite Japan's meltdown of asset prices in 1990 and 1991, it was not until late 1997 and early 1998 that this risk premium rose sharply in Japan. As pointed out above, the market believed that the authorities would in the end act to avoid an outright default in financial obligations. Hence, the call loan defaults by Sanyo securities took the rest of the market by surprise.

Japan's government finally responded by recapitalizing large banks in two stages in 1998–99 and by committing to use public money to protect all deposits and other bank debt. In March 1999, a more significant capital injection to 15 large banks took place in the amount of 7.5 trillion yen.[6] By 2001, 17 other banks received capital injections. Figure 3 indicates that these measures succeeded in containing the money market risk premium.

The years 1997–98 marked a departure of Japan's banking policy from what was earlier called the "convoy approach," in which weak banks and financial institutions would be absorbed by stronger ones. The public's anger toward "finance" turned towards the cozy relationship between bureaucrats and bankers, and resulted in the separation of the Financial Service Agency from the Ministry of Finance and a revision of the Bank of Japan law to make the bank more independent. Government and central bank officials were prohibited from having meals with bankers. As an undesired by-product of these rules, smooth communication between officials and bankers was impaired. The Financial Service Agency focused on prompt resolution of the bad loan problem and did not seem to consider its macroeconomic implications.

These events worsened Japan's credit crunch. Banks intensified their efforts at deleveraging; they now had the capital to recognize bad loans. Bank loan growth turned negative and did not come back to positive territory until the mid 2000s (as shown in Figure 3). Many researchers have found significant negative effects of the deterioration of bank balance sheets on business fixed investment, especially for this period (for example, Sekine 1999; Kasahara, Sawada, and Suzuki 2011). At the same time, the efforts of Japan's nonfinancial corporations to use savings to repay existing debt also intensified after 1998. Both the demand and supply sides of the bank loan market were shrinking. This pattern explains why Japan's "credit crunch" was not accompanied by very high borrowing costs except for the brief period in 1997–98.

The negative feedback loop became even more serious after the credit crunch of 1997–98. Events during these years led to declines in expectations about

---

[5] A better measure of the bank borrowing rate is the Tokyo Interbank Offered Rate. But it is available only back to late 1995. Also, the Japanese three-month treasury bill rate is available only since 1992. The Japanese government started to issue short-term debt in the market only in the mid 1980s, and its market was not well developed until sometime in the 1990s.
[6] All large banks except the Bank of Tokyo Mitsubishi were pressured into receiving capital. Capital was injected in the form of preferred shares, subordinated loans, and bonds. The government put governance pressure on banks by threatening that it would turn preferred shares into common stocks if banks do not perform well. On the other hand, the government continued to encourage misallocation by asking banks to lend to small and medium-sized companies.

Table 1

Financial Assets/Liabilities of Financial Intermediaries in 2001 as Percent of Total Financial Assets/Liabilities

	Japan	*United States*	*Euro area*
Depository corporations	59	25	60
Insurance and pension funds	18	28	13
Others	23	47	27

Source: Flow of Funds, Bank of Japan.

inflation and growth.[7] Moreover, Japan's property prices had largely returned to pre-bubble levels by the late 1990s or earlier 2000s, but have continued to decline thereafter, which also suggests the possibility of negative interaction between growth and asset prices.

Japan's government and the Bank of Japan were slow to perceive this emerging dynamic. For example, the Bank of Japan's official economic report did not recognize the negative interaction between financial factors and the real economy until the fourth quarter of 1993. It took the financial crisis in 1997–98 to persuade the public and the government of the need for recapitalization. To put it another way, one important reason for Japan's forbearance approach through most of the 1990s was that Japan didn't actually experience a serious financial panic until late 1997.

In contrast, the U.S. financial system exhibited serious instability starting in 2007, almost immediately after the collapse of the property and credit market bubble, and the U.S. economy experienced a severe financial crisis from September 2008 into early 2009. Thus, the U.S. government and the Federal Reserve found it easier to gather sufficient support for addressing problems in the banking system, including the bank recapitalization that occurred in 2008, than Japan did in the 1990s; so the U.S. government was able to move more quickly.

The acuteness of the financial crisis in the United States can be explained in part by the fact that the U.S. financial system is more market-oriented and less bank-centered than in many other countries. Table 1 compares Japan, the United States, and the euro area in terms of the share of financial assets/liabilities of different types of financial intermediaries in the economy in 2001. (The choice of the year of comparison does not matter much.) The U.S. economy is clearly an outlier with a large weight for the "others" component, which consists of investment

[7] In the Appendix available with this paper at ⟨http://e-jep.org⟩, Figure A2 shows expected inflation calculated by implied forward rates from the SWAP curve and growth expectations compiled by the Cabinet Office. Although Figure A2 suggests that medium- to long-term inflation expectations fell in advance of deflation, it does not seem to support the Benhabib, Schmitt-Grohe, and Uribe (2001) story that an exogenous emergence of deflationary expectations was a cause of a zero interest rate and deflation. Inflation expectations in the figure are mostly positive and their declines in 1996–1998, as argued in the text, seem to have been a result of developments in the economy.

banks, hedge funds, dealer/brokers, various special purpose vehicles, and so on. In a market-oriented financial system, stresses spread across the system swiftly and the financial authorities are obliged to respond. This may be the major reason for the differences in the speed of authorities' response to the financial crisis between the three areas.

In addition, the U.S. authorities may have learned from Japan's mistake. Remember that Ben Bernanke, chairman of the U.S. Federal Reserve after 2007, was giving speeches back in 2000 about how Japan's central bank should react. However, Bernanke's earlier writings often focused on the threat of deflation. It seems the U.S. authorities failed to learn from the earlier Japanese experience that the bankruptcy of an investment bank can have implications for systemic risk across an economy. For example, in Japan's case the failure of securities companies led to a systemic crisis in 1997. The risk of a similar outcome could have been foreseen when Lehman Brothers was allowed to fail in September 2008.

Monetary Policy for Addressing Financial Instability and Deflation

Just a few years ago, discussions of monetary policy would have largely begun and ended with changes in the policy rate, normally a short-term interest rate, and perhaps a mention of other conventional methods of conducting monetary policy, like adjustments to reserve requirements or discount rates. But starting in the late 1990s, the Bank of Japan began implementing various nonconventional monetary policies, and in the aftermath of the Great Recession, the Federal Reserve and other central banks around the world have followed suit.

The experience of the central banks during this period reveals some close relationships between the deleveraging process ongoing in these economies and monetary policy. On the one hand, monetary policy has been an indispensable tool to mitigate the pain of deleveraging. On the other hand, beyond restoring a degree of stability in the financial system, central banks have had a difficult time actually stimulating the economy.

Differences in individual central bank experiences, however, do exist. We will mainly discuss those between the Bank of Japan and the Federal Reserve. Most notably, the United States, unlike Japan, has essentially escaped deflation—at least so far. Although asset prices declined sharply in the United States as in Japan, at least U.S. stock prices have rebounded more briskly. With these differences in mind, let us now compare the Federal Reserve with the Bank of Japan in their use of monetary policy, starting with conventional monetary policy and then proceeding to unconventional approaches.

Conventional Monetary Policy

The major difference between the Federal Reserve and the Bank of Japan's use of the conventional monetary policy tool is the speed with which they responded to the financial and economic crises. Figure 4 presents the movements in the overnight

Figure 4
Real and Nominal Overnight Rates in Japan and the United States
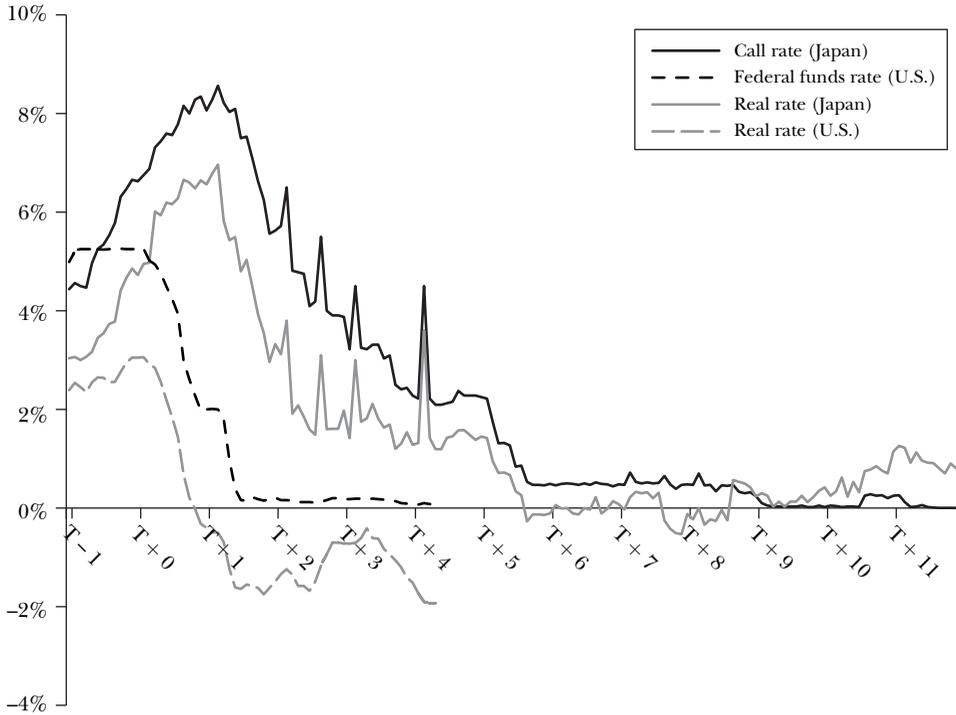
Source: Bloomberg.
Notes: Figure 4 presents the movements in the overnight call rate, which is targeted by the Bank of Japan, and the federal funds rate, which is targeted by the U.S. Federal Reserve. The figure shows both nominal overnight rates as well as the real rates calculated using the core inflation rate that was later announced. T = 0 corresponds to June 2007 for the United States and January 1990 for Japan.

call rate that is targeted by the Bank of Japan and the federal funds rate that is targeted by the U.S. Federal Reserve. The figure shows both nominal overnight rates as well as the real rates calculated using the core inflation rate that was later announced—that is, the Consumer Price Index excepting food and energy prices. Again, the horizontal axis is the time elapsed after the start of the collapse of the stock market bubble, with T = 0 as 1990 for Japan and 2007 for the United States.

The Bank of Japan started to lower the call market rate in 1991. By the second half of 1995 the rate had been brought down to less than 0.5 percent. However, even the rate cuts amounting to more than 800 basis points over four years did not turn Japan's economy around. The behavior of the real interest rate suggests that the Bank of Japan was cutting the nominal rate faster than the speed with which inflation fell, and thus was providing stimulus to the economy. (Otherwise, the real interest rate would rise and exert negative effects on the economy.) The decline in the real interest rate, however, came to a halt and started to move upward in the

late 1990s as the deflationary trend set in. We see here the severe constraint on monetary policy that results from the inability of a central bank to cut interest rates below zero percent—the so-called "zero lower bound." Since the late 1990s, the Bank of Japan's nominal policy rate has been in the zero to 0.5 percent range for more than a decade and a half.

Figure 4 shows that the Federal Reserve reacted to the financial crises and recession of 2007–2008 much more rapidly than did the Bank of Japan in the 1990s. The policy rate was brought down to near-zero within about 18 months of the start of the crisis. Given that U.S. core inflation has remained positive (that is, the inflation rate after stripping out the more volatile energy and food prices), the real interest rate has been clearly negative. In contrast, the real policy rate was never below −0.5 percent in Japan in the 1990s. Faced with the severity of the financial crisis, the Federal Reserve moved quickly to the zero percent lower bound within a short period, which surely is one of the reasons why the U.S. economy has avoided deflation so far.

In both Japan's economy of the 1990s and in the U.S. economy during and after the Great Recession, however, near-zero policy interest rates failed to stimulate the economy adequately. Japan has not escaped from its deflation scare yet, and the U.S. has suffered from an extremely weak labor market.

A Typology of Nonconventional Monetary Policy Measures

Given that the policy rate reached the zero lower bound, the Bank of Japan has adopted many nonconventional monetary policy measures starting in the late 1990s, and the Federal Reserve has done so starting in 2007. Indeed, central banks all around the world began to use such policies in the aftermath of the Great Recession. They can be classified into "quantitative easing" and "forward guidance of interest rates." Quantitative easing, in turn, consists of large-scale asset purchases in distressed markets and in more normal markets, and "pure quantitative easing" (defined below). The term "large-scale asset purchases" is usually used when the central bank is concerned with what type of assets are purchased, while "pure quantitative easing" is used when the bank is only concerned with the size of its balance sheet. In Ueda (2012a), I offer more details on this classification.

Large-scale asset purchases have occurred in many forms. The theoretical rationale for such actions seems to rest on the existence of market imperfections. During a financial crisis, a sharp decline in investors' ability to take risks reduces market liquidity in certain segments of the financial system. In such markets, central bank purchases of assets can lower liquidity/risk premiums and in this way support the economy. Allen and Gale (2007), Curdia and Woodford (2010), Gertler and Karadi (2012) discuss the usefulness of such operations, which are sometimes called "credit easing." In addition to security markets, interbank markets can become dysfunctional due to heightened counterparty risks, especially for horizons longer than a few days. Central banks can advance term loans against some collateral in order to contain risk premiums. Such operations may also be regarded as credit easing.

Other types of large-scale asset purchases by central banks are purchases of Treasury bonds or private financial instruments in more normal market conditions.

For example, many central banks have purchased long-term government bonds and expanded their balance sheets. Such an operation can be decomposed into pure quantitative easing (to be discussed below) and a so-called "operation twist," involving the central bank purchases of long-term Treasury bonds while at the same time selling short-term Treasury bills. The operation twist part of the measure affects the yield curve if investors in such securities are segmented or have "preferred habitats." The effects could spill over into other markets such as the corporate bond market through portfolio rebalancing effects.

Some have argued that irrespective of what a central bank buys, an expansion of the central bank balance sheet generates an easing effect by itself. An example would be central bank purchases of Treasury bills in order to supply liquidity beyond the level required for a zero percent policy rate. Such a policy may be called "pure quantitative easing." At a zero interest rate, however, the economy is largely satiated with liquidity. Hence, it is not clear why attempts to add still more liquidity will produce any significant results. Of course, it would be a different story if the central bank was financing government purchases of goods and services—a helicopter drop of money. Consequently, many researchers now consider it more important what types of assets central banks purchase in their pursuit of nonconventional policies, rather than the size of their balance sheet increases per se. Let us say, however, that the effectiveness of pure quantitative easing remains an open question.

An entirely different form of unconventional monetary easing is *forward guidance*—providing assurance to the market that the key policy interest rate, like the federal funds interest rate, will be lower in the future than currently expected. To affect market expectations of future short rates, the central bank needs to commit to monetary easing even after the economy no longer requires it. This promise of unnecessary easing in the future creates an expectation of rising inflation. As a result, the current market interest rates will be lowered up to a certain maturity, but raised beyond that maturity if inflation expectations rise. Bauer (2012) argues that large-scale asset purchases, by sending the signal that the central bank will continue to be aggressive in monetary easing in the future, also entail an element of forward guidance—a signaling effect.

The underlying logic of how nonconventional monetary policy measures work suggests limits on what they can be expected to achieve. Credit easing—that is, operations in temporarily dysfunctional markets—should come to an end once the markets have adjusted. Forward guidance is an attempt to narrow long–short interest rate spreads up to a certain maturity. Asset purchases in more normal markets may reduce risk premiums. But there are likely to be limits to the extent of the fall in interest rate spreads or risk premiums. Also, as the size of such operations becomes very large, one has to start worrying about distortions generated by direct central bank involvement in financial intermediation.

Nonconventional Monetary Policy at the Bank of Japan and the Federal Reserve

Table 2 illustrates some of the typical nonconventional measures adopted by the Bank of Japan and the Federal Reserve. Although details are different, the

Table 2
Examples of Unconventional Monetary Policy Measures

Operations in dysfunctional markets or sectors (credit easing)	Japan	Funds-supplying operations in term markets (1–12 months)
		Purchases of commercial paper, equities, asset-backed securities, and corporate bonds
	United States	Term Auction Facility
		Lending against asset-backed securities
		Lending to money market mutual funds and broker-dealers
		Purchases of mortgage-backed securities, agency bonds
Asset purchases in more normal markets	Both countries	Purchases of long-term government bonds
	United States	Operation Twist (2011–12)
Pure quantitative easing	Japan	Setting target on the current account balances at the Bank of Japan (2001–2006)
Forward guidance	Japan	"maintain the current zero interest rate until deflationary concerns [are] dispelled" (1999–2000)[a]
	United States	"[T]he Committee . . . anticipates that economic conditions . . . are likely to warrant exceptionally low levels for the federal funds rate at least through late 2014." (2012)[b]

[a] Bank of Japan (1999).
[b] Federal Reserve (2012a).

two central banks have adopted many similar measures, except that the Federal Reserve has not resorted to what I call "pure" quantitative easing nor to purchases of stocks.

In early years of financial stress—that is, the late 1990s to early 2000s for Japan and 2007–2010 for both the United States and Japan—both central banks employed credit-easing measures extensively. The specific measures used reflect the characteristics of the financial system in the two countries. In Japan, efforts focused on channeling funds into the banking sector, the major players in the financial system. Even the Bank of Japan's early purchases of equities (2002–2004) were from banks; that is, they were designed to contain a negative systemic externality stemming from banks' forced sale of equities during their deleveraging. Also the Bank of Japan's purchases of asset-backed securities were aimed at substituting for the impaired ability of banks to make loans, but this move was unsuccessful given the underdeveloped nature of the market in Japan. The Federal Reserve channeled funds into a variety of agents, reflecting the more market-based nature of the U.S. financial system, including lending to money market mutual funds and broker-dealers. It also lent against asset-backed securities to respond to the stress in the housing finance market. In addition, the Fed bought huge amounts of mortgage-backed securities and debt from the quasi-government agencies of Fannie Mae and Freddie Mac in 2008–2009.

As the acute phase of the financial crisis passed, both the Bank of Japan and the Federal Reserve shifted to purchasing mostly government bonds. In both countries, the size of the operations was unprecedented. The Bank of Japan during 2001–2006 proceeded to target on the "current account balances" held at the Bank of Japan—essentially bank reserves—and raised its target from 5 trillion yen initially to 30–35 trillion yen in 2004. This was by itself pure quantitative easing, but it was also accompanied by government bond purchases and forward guidance. Forward guidance has been used extensively by the Bank of Japan. It was first introduced in April 1999 when the bank committed to maintaining a near-zero interest rate "until deflationary concerns were dispelled." The commitment was lifted in August 2000, but was reintroduced in March 2001 with quantitative easing, which was continued until March 2006. The Bank has made a similar commitment since 2009.

The Federal Reserve has used the forward guidance approach since December 2008 when it announced "the [Federal Open Market] Committee anticipates that weak economic conditions are likely to warrant exceptionally low levels of the federal funds rate for some time" (Federal Reserve 2008).[8] It was strengthened in January 2012 to the maintenance of a near-zero rate "at least through late 2014" (Federal Reserve 2012a).

Evidence on the Effectiveness of Nonconventional Monetary Policy Measures

Turning to the effectiveness of the measures adopted, let us first examine the extent and effectiveness of the expansions of central bank balance sheets. Figure 5 shows the monetary base (that is, currency outstanding plus bank reserves) relative to GDP for four countries including the United States and Japan. The monetary base has behaved in almost the same way in the four countries since 2007. It has more than doubled, except in Japan where it started from a higher level. The Bank of Japan carried out its massive balance sheet expansion in the early to mid 2000s. Despite such injections of money, none of the economies have expanded by as much as a simple monetarist calculation would suggest (for example, the doubling of money has not led to a parallel doubling of nominal GDP, nor anything close to it).

The reason the expansion of base money failed to stimulate the economy can be inferred from Figure 6, where the so-called money multiplier is shown: that is, the ratio between a broader measure of the money supply, M2, and the monetary base. In general, to varying degrees, the period of sharp expansions in the monetary base saw corresponding declines in the multiplier. That is, central bank money supplied has been largely held by financial institutions, rather than used for credit creation. Otherwise, M2 would have increased more sharply. This pattern accords with the discussion of Japan's deleveraging earlier in this paper. Japan's banks did not trust each other and were not able to borrow in the interbank market. Even those who had plenty of liquidity did not lend, either because lending was constrained by absence of capital buffer or nonfinancial firms were deleveraging. The central bank

---

[8] A similar statement was actually first used in 2003 when the Federal Reserve lowered the target rate to 1 percent (Federal Reserve 2003).

Figure 5
Monetary Base/GDP
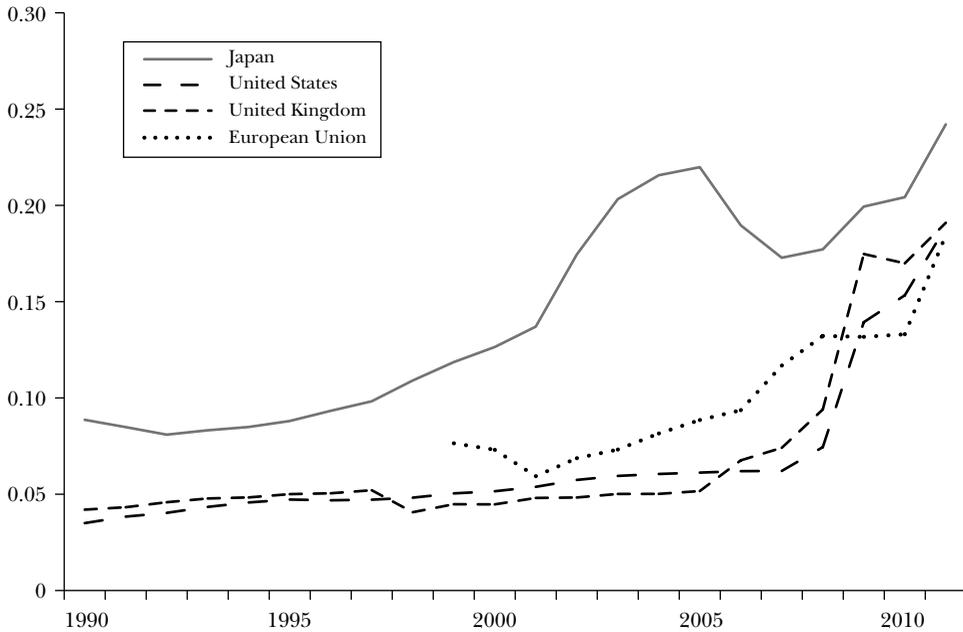
Source: Datastream.
Note: Figure 5 shows the monetary base (that is, currency outstanding plus bank reserves) relative to GDP for four countries.

could provide liquidity by stepping in, but this pure quantitative easing was not very effective in stimulating the real economy. Such stories seem to apply to other countries as well.

However, there is a literature that has found that nonconventional monetary policies do have some effect on interest rates and other asset prices. Among other things, various credit easing measures have contained risk premiums that might otherwise have created considerable instability. For example, the Bank of Japan's fund-supplying operations reduced money market risk premiums almost to zero (Baba, Nakashima, Shigemi, and Ueda 2006; Bank of Japan 2009). This pattern can be seen informally in Figure 3, where the money market risk premium has been kept at very low levels with only two exceptions, one, during Japan's credit crunch of 1997–1998 (T = 7, 8) and, the other, right after the September 2008 shock in the United States (T = 18, 19). The same figure suggests that the Federal Reserve's credit easing measures also succeeded in containing the money market risk premium after 2009 (T = 2).

Event study approaches have been used to analyze the effects of the Federal Reserve's large-scale asset purchases on asset prices. (The common use of this approach reflects the limited availability of time-series data for carrying out more

Figure 6
Money Multiplier (M2/Monetary Base)
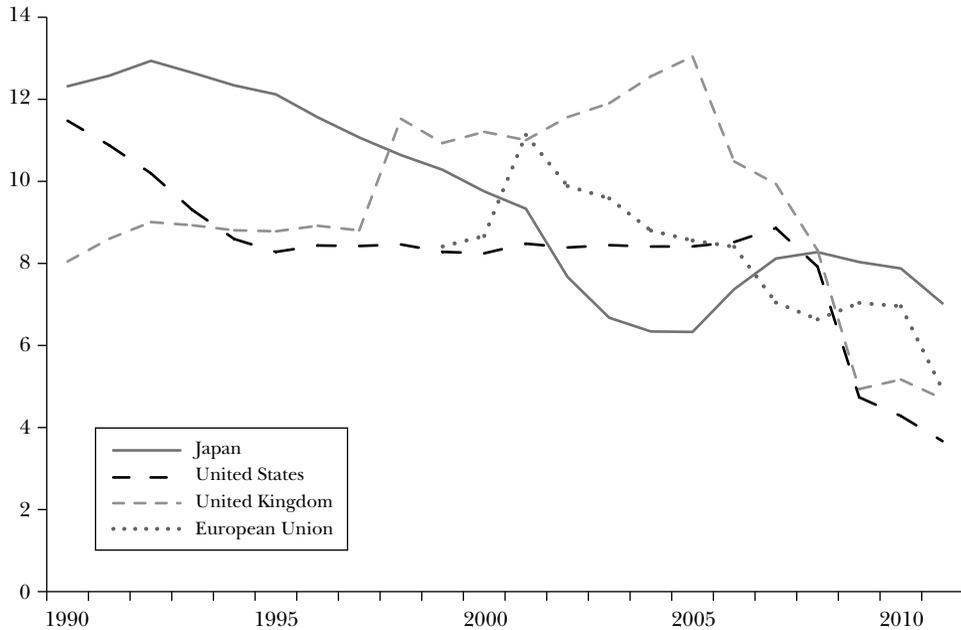
Notes: M2 equals currency outstanding plus bank deposits.

detailed analyses.) This literature has several findings. First, asset purchases in general have affected interest rates (Williams 2011). Second, the transmission channel of the purchases remains unclear. Some emphasize portfolio rebalancing effects, that is, the effects depend on what assets are bought (Krishnamurthy and Vissing-Jorgensen 2011), while others consider the signaling channel of asset purchases more important (Christensen and Rudebusch 2012). Third, the effects on interest rates decline as we move from the Fed's earlier purchases in 2008 and 2009 to those in 2010 and 2011. The decline in effectiveness seems to be due to the disappearance of the credit easing aspect of the purchases as markets returned to normalcy and also due to the repeated use of other measures such as forward guidance to lower interest rates (Bauer 2012).

With regard to Japan's experience, Lam (2011) and I (Ueda 2012b) apply a similar approach, with some similar findings. First, the Bank of Japan's asset purchases have affected government bond and corporate bond yields as well as stock prices. Thus, there is some evidence of portfolio rebalancing effects. However, the effects on the yen exchange rate are not discernible in most cases. Second, the first introduction of forward guidance in 1999 and quantitative easing in 2001 generated larger effects than other measures. Third, pure quantitative easing, that is, increases

in the target amount of the current account balances without increases in government bond purchases did not generate a significant response of asset prices.

More formal analyses of the Bank of Japan's forward guidance also find significant effects on the term structure of interest rates (for example, Okina and Shiratsuka 2004; Baba, Nakashima, Shigemi, and Ueda 2006; Oda and Ueda 2007). However, few find significant effects of the pure quantitative easing.[9] In short, a body of evidence suggests that the two central banks' nonconventional monetary policy measures, with the exception of pure quantitative easing, have had nonnegligible effects on asset prices. The effects of the measures on the real economy, however, have been less well analyzed.

Failure to Stop Deflation/Stagnation

The above summary of findings on the effectiveness of the two central banks' monetary policy raises the obvious question of why they have so far failed to stop deflation in Japan and economic stagnation in the United States. If we assume that the avoidance of deflation in the United States so far has been partly due to monetary policy, there is also the question of why the Federal Reserve was more successful in avoiding deflation than the Bank of Japan.

Surely the first and foremost reason why monetary policy in Japan and the United States seems to have been ineffective is the sheer size of the negative shocks to the economy. As we have discussed, deleveraging forces in the aftermath of the burst of the bubbles generated tremendous difficulties in the two economies. As of June 2012, real interest rates in the United States are at two-decade lows for Treasuries and high-rated corporate bonds. Still, growth is sluggish and the labor market is extremely weak. Real interest rates in Japan are not as low given the deflationary trend, but nominal interest rates have been at record low levels for more than a decade. As Eggertsson and Krugman (2011) suggest, the natural rate of interest may well have been negative in the two countries.

In Japan's case, a second reason for the ineffectiveness of monetary policy lies in the suboptimal policy making that in some ways intensified deleveraging forces. As with bank recapitalization and conventional monetary easing, nonconventional monetary policy measures should have been adopted earlier in Japan. The effectiveness of the measures to stimulate Japan's economy may have been undermined by the emergence of deflationary expectations. For example, the forward guidance strategy of the Bank of Japan, if it had been successful, would have lowered expected future short-term rates up to that point in the future where the forward guidance extends, but then raise expected interest rates beyond that (because forward guidance should bring an expectation of increased inflation in the future). However, in

---

[9] One exception is Honda, Kuroki, and Tachibana (2007), who find, using vector autoregression analysis, that an expansion of bank reserves (as measured by the Bank of Japan's current account balance) exerted significant effects on stock prices and in turn on output. Given the methodology, however, it is unclear which aspect of the unconventional monetary policies generated such effects. The analysis also does not include a variable representing changes in perceptions about the stability of the financial system and hence runs the risk of picking up spurious correlation between money and output.

Japan, that expectation of future inflation—and hence, a steepening of the yield curve beyond some maturity—has not materialized.

Moreover, a significant portion of the favorable response of asset prices to nonconventional measures as discussed above did not last long. For example, the introduction of quantitative easing by the Bank of Japan in March 2001 led to a significant increase in stock prices, which faded away in a few months as the economy remained weak. On this point, it is interesting to note that the response of stock prices in the United States to the Fed's Treasury purchases during 2010–2011 was very similar to the pattern in Japan, although the period of favorable response was longer.

A third aspect of the difficulty the Bank of Japan had in stimulating the economy with nonconventional monetary policy relates to the lack of well-developed capital markets. In the U.S. economy, the markets for mortgage-backed securities and for corporate bonds are comparable in size to the government bond market. The Federal Reserve has been able to carry out operations in these capital markets to affect private financial intermediation. In particular, mortgage-backed securities and debt from Fannie Mae and Freddie Mac exceeded 50 percent of the Federal Reserve's total assets in early 2010. The Fed also lent against asset-backed securities as collateral. These operations reduced risk premiums in the capital markets and contained the deleveraging attempts by holders of the instruments.

In contrast to the case in the United States, the size of Japan's market for nonfinancial corporate bonds is only 10 percent of GDP while Japan's market for long-term government bonds exceeds 130 percent of GDP. Japan's market for asset-backed securities is even smaller than its corporate bond market. As of the end of 2011, capital market debt and stocks made up only 4.1 percent of the assets of the Bank of Japan. The rest was long-term government bonds and short-term lending to financial institutions backed mostly by government debt. The Bank of Japan has been able to ease the funding pressure of financial institutions by lending to them, but it has less ability to affect private capital markets through asset purchases, due to the fact that such markets are smaller in Japan. Purchases of stocks have been an attempt to alleviate this problem. Given the obvious risks involved, however, the Bank of Japan tiptoed into this measure in 2002 after a 70 percent fall in stock prices from the peak.

Finally, as nonconventional measures are used over time within a stagnant economy, risk premiums/interest rate spreads are lowered to extreme levels, and they could be approaching their limits, especially in Japan. For example, back in January 1999, the spread between ten-year and two-year Japanese government bonds was 1.45 percent. Just a few months later in April 1999, after the announcement of the zero interest rate policy—the first wave of nonconventional policy using forward guidance—the spread had fallen to 1.29 percent. By May 2003, in the quantitative easing period, the spread was down to 0.5 percent. However, by June 2012, despite massive bond purchases by the Bank of Japan, the same spread was up to 0.67 percent. Perhaps these spreads can decline again, but one cannot escape the impression that with the protracted stagnation in Japan's economy and

the repeated application of nonconventional monetary policy measures, such policies are close to a lower bound in reducing the spread between long-term and short-term interest rates.

Something similar may be taking place in the United States, as the result by Bauer (2012)referred to above points out. However, with the spread between ten-year and two-year Treasury bonds at 1.39 percent in June 2012, the traction left for this aspect of the Fed's nonconventional monetary policy seems roughly equal to that of the Bank of Japan in the late 1990s as far as the government bond market is concerned. The Fed also probably has room for more action elsewhere, like in the markets for corporate bonds and mortgage-backed securities.

But overall, nonconventional monetary policies in both economies have experienced strong headwinds from deleveraging pressures. Their repeated application gradually lowered effectiveness. In addition, in Japan the same factors that made the deleveraging process severe—namely, the slow response of policymakers and the absence of well-developed markets for private debt—have limited the scope of nonconventional monetary policies.

Concluding Remarks

The parallel between Japan's economy since the 1990s and the U.S. economy since the Great Recession is far from a perfect correspondence. First, the growth rate of Japan's economy was probably due for a slowdown. Japan's economy grew at an annual per capita rate of 3.3 percent per year from 1974 to 1990. As Japan's economy approached the technological frontier, this growth rate was unlikely to be sustained. Second, Japan faces a severe demographic adjustment. The size of Japan's workforce started shrinking in the late 1990s, and the size of Japan's population began declining in 2005. Third, Japan's economy has long been powered by firms focused on international markets, but has yet to encourage innovation and greater competition in many service industries in the domestic market.

But while Japan's economy was unlikely to sustain its boom period forever, Japan's economic stagnation since the bursting of Japan's asset price bubble in 1990 is at least partly due to mismanaged macroeconomic policy. Some of the Bank of Japan's interest rate increases in the early 1990s may have been unnecessary or even counterproductive, and its interest rate cuts in the early to mid 1990s could have been more aggressive. Japan's government should have acted to recapitalize banks in the early to mid 1990s. As a result, Japan's economy was more vulnerable to a severe financial crisis in 1997–1998 and a negative feedback loop developed among asset prices, financial stability, and growth. Deflation, if mild, ensued in response, and along with strong expectations of a low-growth, low-inflation future, the Bank of Japan's ability to stimulate the economy through conventional and nonconventional monetary policy measures became highly constrained.

Although the U.S. Federal Reserve and the Treasury have sometimes been criticized for not acting aggressively enough to stimulate the U.S. economy in

recent years, U.S. policymakers have been far more aggressive since 2007 than were their Japanese counterparts back in the 1990s. The Federal Reserve lowered the federal funds interest rate virtually to zero percent within about 18 months of the onset of the financial crisis and has adopted various nonconventional monetary policy measures as well. U.S. banks were recapitalized in late 2008. Thanks to these measures, the U.S. financial system has resumed stability to a certain extent and core inflation has stayed in positive territory. Moreover, the Federal Reserve strengthened forward guidance in January 2012 by stating that a near-zero policy rate would continue until late 2014, even though core inflation was already around its target of 2 percent.

In contrast, financial markets have questioned the Bank of Japan's resolve to fight deflation. Its initial use of forward guidance was discontinued in August 2000 when Japan's core inflation was still −0.5 percent. Similarly, the Bank of Japan exited from its second wave of nonconventional monetary policy in March 2006 with a core consumer price inflation rate of −0.5 percent, despite an earlier promise of the continuation of the policy until "inflation is stably positive." Such a seeming difference between the two central banks' resolve to fight deflation may have contributed to the differences in the effects of policy measures.

The U.S. economy in mid 2012 remains far from full recovery from the financial and economic crisis of 2007–2009. Households are still deleveraging. U.S. property prices relative to the previous peak are now roughly at levels where Japanese land prices were in the mid to late 1990s, which is when the negative feedback loop between falling asset values and the real economy in Japan became more significant. While the Federal Reserve does have some additional room to ease credit with nonconventional monetary policies, such choices are unlikely to be as effective as they were several years ago. The power of the Fed's tools may or may not be enough to counteract possible negative forces coming from further deleveraging in the U.S. economy or from other external shocks such as instability in the euro area.

Both the U.S. government and the Federal Reserve have been quick to respond to financial stresses in the economy. There is a deeper question here, however, as to whether the prompt responses of U.S. policymakers during the crisis have sowed the seeds of future crises by generating moral hazard on the part of private investors, who will expect such actions to continue in the future. This possibility will need to be addressed in future studies and taken into account in future policy making.

References

Allen, Franklin, and Douglas Gale. 2007. *Understanding Financial Crises.* Clarendon Lectures in Finance. Oxford: Oxford University Press.

Baba, Naohiko, Motoharu Nakashima, Yosuke Shigemi, and Kazuo Ueda. 2006. "The Bank of Japan's Monetary Policy and Bank Risk Premiums in the Money Market." *International Journal of Central Banking* 2(1): 105–35.

Ball, Laurence M. 2012. "Ben Bernanke and the Zero Bound." NBER Working Paper 17836.

Bank of Japan. 1999. "Minutes of the April 9, 1999 Meeting." (English translation prepared by the Bank staff based on the Japanese original.) http://www.boj.or.jp/en/mopo/mpmsche_minu/minu_1999/g990409.htm/.

Bank of Japan. 2001. "Developments in Profits and Balance Sheets of Japanese Banks in Fiscal 2000 and Banks' Management Tasks." *Quarterly Bulletin* 9(4): 73–130.

Bank of Japan. 2009. *Financial Markets Report.* Financial Markets Department, August 31.

Bauer, Michel D. 2012. "Fed Asset Buying and Private Borrowing Rates." *Economic Letters*, Federal Reserve Bank of San Francisco, May 21.

Benhabib, Jess, Stephanie Schmitt-Grohe, and Martin Uribe. 2001. "The Perils of Taylor Rules." *Journal of Economic Theory* 96(1–2): 40–69.

Bernanke, Ben S. 2000. "Japan's Slump: A Case of Self-Induced Paralysis?" Chap. 7 in *Japan's Crisis and Its Parallels to the U.S. Experience,* Special Report 13, edited by Adam S. Posen and Ryoichi Mikitani. Institute for International Economics.

Bernanke, Ben S. 2003. "Some Thoughts on Monetary Policy in Japan: Remarks before the Japan Society of Monetary Economics, Tokyo, Japan." May 31, http://www.federalreserve.gov/boarddocs/speeches/2003/20030531/default.htm.

Caballero, J. Ricardo, Takeo Hoshi, and Anil K. Kashyap. 2008. "Zombie Lending and Depressed Restructuring in Japan." *American Economic Review* 98(5): 1943–77.

Christensen, Jens H. E., and Glen D. Rudebusch. 2012. "The Response of Interest Rates to U.S. and U.K. Quantitative Easing." Working Paper no. 2012-06, Federal Reserve Bank of San Francisco.

Curdia, Vasco, and Michael Woodford. 2010. "The Central-Bank Balance Sheet as an Instrument of Monetary Policy." NBER Working Paper 16208.

Eggertsson Gauti B., and Paul Krugman. 2011. "Debt, Deleveraging, and the Liquidity Trap." Paper presented at the Japan Project Meeting of NBER, June 24–25, Tokyo.

Federal Reserve. 2003. "Press Release." Federal Open Market Committee Statement, August 12. http://www.federalreserve.gov/boarddocs/press/monetary/2003/20030812/default.htm.

Federal Reserve. 2008. "Press Release." Federal Open Market Committee Statement, December 16. http://www.federalreserve.gov/newsevents/press/monetary/20081216b.htm.

Federal Reserve. 2012a. "Press Release." Federal Open Market Committee Statement, January 25. http://www.federalreserve.gov/newsevents/press/monetary/20120125a.htm.

Federal Reserve. 2012b. "Transcript of Chairman Bernanke's Press Conference." April 25. http://www.federalreserve.gov/monetarypolicy/fomcpresconf20120425.htm.

Fukao, Kyoji, and Hyeog Ug Kwon. 2006. "Why Did Japan's TFP Growth Slow Down in the Lost Decade? An Empirical Analysis Based on Firm-Level Data of Manufacturing Firms." *Japanese Economic Review* 57(2): 195–228.

Gertler, Mark, and Peter Karadi. 2012. "QE1 vs. 2 vs. 3 . . . A Framework for Analyzing Large Scale Asset Purchases as a Monetary Policy Tool." Paper presented at the Federal Reserve Board conference on "Central Banking: Before, During, and After the Crisis," March 23–24, 2012, Washington D.C. http://www.federalreserve.gov/newsevents/conferences/GertlerKaradi.pdf.

Hanson, Samuel G., Anil K. Kashyap, and Jeremy C. Stein. 2011. "A Macroprudential Approach to Financial Regulation." *Journal of Economic Perspectives* 25(1): 3–28.

Hayashi, Fumio, and Edward C. Prescott. 2002. "The 1990s in Japan: A Lost Decade." *Review of Economic Dynamics* 5(1): 206–35.

Honda, Yuzo, Yoshihiro Kuroki, and Minoru Tachibana. 2007. "An Injection of Base Money at Zero Interest Rates: Empirical Evidence from the Japanese Experience 2001–2006." Osaka University, Discussion Papers in Economics and Business, no. 07-08.

Hoshi, Takeo, and Anil Kashyap. 1999. "The Japanese Banking Crisis: Where Did It Come From and How Will It End?" *NBER Macroeconomics Annual*, vol. 14, pp. 129–212.

International Monetary Fund. 2012. *Global Financial Stability Report.* April.

Kane, Edward J. 1989. "The High Cost of Incompletely Funding the FSLIC Shortage of Explicit Capital." *Journal of Economic Perspectives* 3(4): 31–47.

Kasahara, Hiroyuki, Yasuyuki Sawada, and Michio Suzuki. 2011. "Investment and Borrowing Constraints: Evidence from Japanese Firms." Paper

presented at the Japan Project Meeting of the NBER, June 24–25, Tokyo.

Kashyap, Anil K, and Jeremy C. Stein. 2004. "Cyclical Implications of the Basel II Capital Standards." Federal Reserve Bank of Chicago *Economic Perspectives* 28(Q1): 18–31.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity*, Fall, pp. 215–287.

Krugman, Paul. 2012. "Earth to Ben Bernanke: Chairman Bernanke Should Listen to Professor Bernanke." *New York Times Magazine*, April 24.

Lam, W. Raphael. 2011. "Bank of Japan's Monetary Easing Measures: Are They Powerful and Comprehensive?" IMF Working Paper, WP/11/264. http://www.imf.org/external/pubs/ft/wp/2011/wp11264.pdf.

Nishikawa, Yoshifumi. 2011. *Nishikawa Yoshifumi Kaikoroku* [*Autobiography of Yoshifumi Nishikawa*]. Tokyo: Kodansha.

Oda, Nobuyuki, and Kazuo Ueda. 2007. "The Effects of the Bank of Japan's Zero Interest Rate Commitment and Quantitative Monetary Easing on the Yield Curve: A Macro-Finance Approach." *Japanese Economic Review* 58(3): 302–28.

Ogawa, Kazuo, and Kazuyuki Suzuki. 1998. "Land Value and Corporate Investment: Evidence from Japanese Panel Data." *Journal of the Japanese and International Economies* 12(3): 232–49.

Okina, Kunio, and Shigenori Shiratsuka. 2004. "Policy Commitment and Expectation Formation: Japan's Experience under Zero Interest Rates." *North American Journal of Economics and Finance* 15(1): 75–100.

Peek, Joe, and Eric S. Rosengren. 2005. "Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan." *American Economic Review* 95(4): 1144–66.

Sekine, Toshitaka. 1999. "Firm Investment and Balance-Sheet Problems in Japan." IMF Working Paper, WP/99/111.

Ueda, Kazuo. 2000. "Causes of Japan's Banking Problems in the 1990s." Chap. 3 in *Crisis and Change in the Japanese Financial System*, edited by T. Hoshi and H. T. Patrick, 59–81. Kluwer Academic Publishers.

Ueda, Kazuo. 2012a. "Japan's Deflation and the Bank of Japan's Experience with Non-traditional Monetary Policy." *Journal of Money, Credit and Banking* 44(2): 175–190.

Ueda, Kazuo. 2012b. "The Effectiveness of Non-traditional Monetary Policy Measures: The Case of the Bank of Japan." *Japanese Economic Review* 63(1):1–22.

Williams, John, C. 2011. "Unconventional Monetary Policy: Lessons from the Past Three Years." *FRBSF Economic Letters* 2011-31, Federal Reserve Bank of San Francisco, October 3.

The Relationship between Unit Cost and Cumulative Quantity and the Evidence for Organizational Learning-by-Doing

Peter Thompson

The concept of a learning curve for individuals has been in widespread use in the psychology literature since the beginning of the twentieth century. Ebbinghaus (1885) famously demonstrated the first learning curve by memorizing ever-longer strings of nonsense syllables; Bryan and Harter (1899) studied learning curves exhibited by telegraph operators sending and receiving Morse code; and Book (1908) studied a learning curve in typing skills. The idea that a phenomenon analogous to the learning curve might also apply at the level of the organization took longer to emerge, but it had begun to figure prominently in military procurement and scheduling at least a decade before Wright's (1936) classic paper providing evidence that the cost of producing an airframe declined as cumulative output increased.

Wright (1936), who was careful *not* to describe his empirical results as a learning curve, proposed three explanations for the relationships between cost and cumulative quantity produced that he observed. The first was the "improvement in proficiency of a workman with practice" (p. 124), characterized by the individual learning curve. The others were "the ability to use less skilled labor as more and more tooling and standardization of procedure is introduced," and "the greater spread of machinery and fixture set up time in large quantity production" (p. 124). Only the first of these is what we would unambiguously identify as a source of organizational learning; the others are consistent with organizational learning but also with standard static economies of scale.

Kenneth Arrow's (1962) seminal paper, "The Economic Implications of Learning by Doing," marked an important new direction for research on the relationship

■ *Peter Thompson is Professor of Organization & Management, Goizueta Business School, Emory University, Atlanta, Georgia. His email address is ⟨peter.thompson@emory.edu⟩.*

between cost and cumulative output. Arrow was among the first to apply the concept to a time horizon far beyond those typical of applications to procurement and scheduling of narrowly-defined products, thus demonstrating that organizational learning could have important economic implications. Arrow was the first to propose an explicit interpretation of the relationship between cost and cumulative quantity produced as an organizational learning curve, and he also seems to have been the first to apply the term learning-by-doing at the level of the organization rather than the individual. In particular, Arrow (p. 156) argued that the decline in production costs was a *consequence* of the accumulated experience, because "it is the very activity of production which gives rise to problems for which favorable responses are selected over time."

It quickly became apparent that the notion of organizational learning as a by-product of accumulated experience has important consequences for firm strategy. This realization was made first, not in the economics literature, but in the emerging industry of business strategy consulting. In 1966, the Boston Consulting Group (BCG) built its fledgling consulting business around the concept of what it branded the *experience curve*. Billed as a radical departure from traditional learning curves "which applied only to direct labor," according to Henderson (1973), the BCG asserted that cost reductions associated with cumulative output applied to all costs, were "consistently around 20–30% each time accumulated production is doubled, [and] this decline goes on in time without limit" (Henderson 1968). The BCG soon found experience curves everywhere it looked. John Clarkeson, at that time a partner in BCG, later recalled (as quoted in Kiechel, 2010, pp. 37–8): "For the next five years, maybe more, we applied experience curves to anything that moved, and a lot of things that didn't." The experience curve led the BCG to advise firms that new products "should be priced as low as necessary to dominate their market segments or probably not be sold at all," and that "market share should be maintained at all costs" (Henderson 1968).

There is, perhaps, some hubris in the BCG's telling of the history of its experience curve. After all, Wright (1936) had 30 years earlier documented cost–quantity relationships not only in the direct labor component of total costs, but also in materials use, intermediate inputs, and indirect costs.[1] Nonetheless, outside observers share BCG's view of the profound influence of its formulation of organizational learning. For example, the editors of the *Harvard Business Review* recently selected the experience curve as one of five charts that "changed the world" (Ovans 2011). In a recent history of the strategy consulting business, Kiechel (2010, p. 31) concluded: "The experience curve was, simply, the most important concept in launching the

---

[1] Twenty years later, Asher (1956, p. 16) observed that Wright was "still the most frequently quoted authority on the subject, and his early formulation of the principle is accepted by most airframe producers and by a large number of Air Force Personnel." However, in defense of BCG's version of events, Asher (p. 20) also noted that "for some inexplicable reason, Wright's analysis of materials cost has never been so important as his analysis of labor cost."

[business] strategy revolution . . . no other idea was to set in motion such an alteration in corporate consciousness."

In the world of economic research, formal explorations of the strategic implications of organization learning began to appear as part of the expanding application of dynamic methods to industrial organization. For example, organizational learning was shown to affect dynamic pricing strategy because production costs are expected to fall as cumulative production increases (Rosen 1972; Spence 1981; Clarke, Darrough, and Heinecke 1982). Organizational learning can also promote industry concentration and imperfect competition (Fudenberg and Tirole 1983; Dasgupta and Stiglitz 1988); indeed, it can create concentration out of initially identical firms by inducing them to be indifferent between dramatically different pricing and survival strategies (Petrakis, Rasmusen, and Roy 1997). Organizational learning can also create incentives for incumbent firms to engage in predatory behavior that deters entry and promotes exit (Cabral and Riordan 1994, 1997; Hollis 2002). As I demonstrate in Thompson (2010), the specific strategic consequence of learning is, in each case, dependent on the auxiliary assumptions made in any given model, but the sense that organizational learning has potentially profound consequences is pervasive in the literature.

In the 50 years since Arrow's (1962) paper appeared, empirical research on organizational learning curves has continued unabated. Numerous studies have appeared each year documenting that the relationship between cost and cumulative quantity produced, that was first described by Wright, can be found in enormously varied settings (for reviews of this literature, see Yelle 1979; Argote and Epple 1990; Dutton and Thomas 1994; Dar-El 2000, chap. 8). As a result, the negative relationship between unit production costs and cumulative output is one of the best-documented empirical regularities in economics.

Nonetheless, the thesis of this paper is that the conceptual transformation of the relationship between cost and cumulative production into an organizational learning curve with profound strategic implications has not been sufficiently supported with direct empirical evidence. In the next subsection, we describe the standard formulation of the empirical relationship between cost and cumulative output and highlight particular features of the standard model that give rise to important strategic implications of organizational learning: in particular, that the learning must continue for an extended period of time and the amount learned in any interval of time is causally influenced by the rate of output. The remainder of the paper attempts to demonstrate that the evidence supporting these features is surprisingly tenuous. The paper concludes with discussion of the sort of research that is needed to close this gap in the empirical evidence.

The Standard Formulation of the Organizational Learning Curve

The standard empirical formulation of the organizational learning curve assumes that the current unit cost of a firm of age t, $c(t)$ is a decreasing function

of its cumulative prior output, $y(t)$. The form of the relationship in the standard model is the power rule, $c(t) = c(0) y(t)^{-\beta}$. In this specification, unit cost declines by a constant proportion with each doubling of cumulative output so, although learning is unbounded, costs decline at a rate that falls asymptotically to zero when the current output rate is held constant. The rate of learning, as captured by β, is frequently summarized by the *progress ratio*, $r = 2^{-\beta}$. (A lower progress ratio is associated with faster progress.) A progress ratio of 80 percent, for example, is obtained when $\beta = 0.32$, and implies a unit cost reduction of 20 percent with each doubling of cumulative output. The standard model follows neither Wright's (1936) empirical formulation, which relates *cumulative* unit cost to cumulative production, nor to Arrow's (1962) theoretical model, which relates current unit cost to cumulative *investment*. The first exposition of the standard formulation appears to be in an unpublished empirical paper by Crawford (c. 1942). The standard theoretical formulation of the learning curve is much the same: cost is a deterministic declining function of cumulative output and, when parameterized, the power rule is the preferred functional form.

When interpreted as a statistical association, the standard formulation is not especially contentious. It has been compared with other specifications, and has usually been found to fit the data better than, or as well as, the alternatives. As a caveat, however, note that learning curves relate two trending variables—rising cumulative output and declining costs—so the explanatory power of a variety of specifications is inevitably high: the resulting horse race (between typically non-nested specifications) is consequently reduced to a comparison of high coefficients of determination that differ by margins of no real economic or statistical significance. While there is evidence that the standard formulation performs worse over longer-time horizons and out-of-sample predictions are often poor (Alchian 1963; Hirsch 1952, 1956; Conway and Schultz 1959), it has repeatedly proved useful as a planning tool over the time horizons typical for operations research problems.

However, when the standard model is used as the basis for theorizing about strategy, its particular formulation takes on considerable importance. Two features in particular seem fundamental. First, the interval of time over which cost reductions are secured needs to be sufficiently extended so that dynamic considerations matter. Second, there must be a causal effect of cumulative production on current cost. This causal effect is of course, likely to be transmitted through some mediating variable, but the distinctive implications of learning derive from the premise that increasing output today will secure greater cost reductions tomorrow. Wherever these two features are absent, familiar strategic and economic implications of learning curves may be either ameliorated or overturned. This point is perhaps best illustrated with three examples.

First, while unbounded learning by doing has long been associated with an incentive on the part of a first-mover to engage in practices that deter subsequent entry (for example, Cabral and Riordan 1994), bounded learning by doing may have the opposite effect. Hollis (2002) has shown that, if the first mover has attained, or nearly attained, a terminal productivity at which learning by doing effects have

attenuated, that firm will likely prefer *more* entry rather than less. The reason is that more entry will divide the market among more followers and slow the rate at which their costs catch up with those of the first-mover. The desire to face more competitors can even induce the leading firm to subsidize entry—for example by offering to license technology at a low cost.

Second, if learning quickly arrives at a terminal productivity, then long-run improvements in performance cannot be the result of learning. This realization led to the development of hybrid models of endogenous (macroeconomic) growth that combine passive learning with the development of new, superior, vintages of technology from research and development activities (for example, Young 1993; Stein 1997). In these models, almost any relation is possible between the rates of learning and economic growth: an increase in the rate of learning may have no effect on long-run growth, it may increase it, or it may decrease it. For example, too much learning may lead to stagnation, and stagnation may arise independently of the rate of learning; learning may also induce clustering of innovations and, more generally, cyclical patterns of growth. In Thompson (2010), I provide a formal analysis of these claims.

Third, the standard formulation of the organizational learning curve induces strategic behavior in which firms set prices lower and output higher than would be warranted by static profit maximization. Increasing output above the level that equates marginal cost and marginal revenue yields a benefit in the form of lower future costs. The optimal strategy increases output until the gap between marginal cost and revenue (that is, the foregone profit resulting from a marginal increase in the distance of output from its static optimum) is equal to the discounted present value of all future cost savings obtained from an increment to output today (Rosen 1972); the effect can be strong enough to induce optimal paths for price that are constant, that rise, or that change non-monotonically even as cost falls (for examples, see Spence 1981; Clarke, Darrough, and Heinecke 1982; Petrakis, Rasmusen, and Roy 1997).

These dynamic pricing results are built on the assumption that the association between output and the rate of decline of cost is causal. As a result, the firm has only one instrument, price, to attain divergent goals of static profit maximization and dynamic cost reductions. Suppose, for example, that organizational learning is a function of time rather than output. In this case, cumulative output and cost are correlated, but changes to output have no influence on the path of cost. The firm sets static marginal cost to marginal revenue at every point in time, because only the goal of static profit maximization is susceptible to manipulation by the firm's policy choices. Suppose, instead, that cost reductions are secured not by passive learning from experience but by purposive investments in research and development. In this case, current output and the rate of cost reduction are correlated—thereby giving the appearance of a learning curve—but the underlying reason is that the optimal rate of investment is generally increasing in firm size (for example, Klepper 1996). In this setting the optimal static price is set in every period because the firm now has two instruments with which to pursue its two objectives.

To a somewhat lesser extent, it also matters that the learning rate is reasonably predictable. If the learning rate is uncertain, the value of strategically increasing output now depends on the *expected* cost savings that will be obtained in the future as a result of having greater cumulative output. Under the standard formulation, however, cost savings are a concave function of the learning parameter, so any mean-preserving increase in the subjective variance of the rate of learning reduces the expected value of strategic behavior today. In fact, unless the occurrence of negative learning rates has zero probability (and we will see later that it does not), the expected cost, $E[c(y)]$ need not be a monotonic function of y: for example, if in the standard formulation β is believed to be a draw from the normal distribution with mean $\bar{b}$ and variance σ^2, $E[c(y)]$ is strictly increasing in cumulative output for any $y > e^{\bar{\beta}/\sigma}$.

Empirical Learning Curves

In this section we review some evidence on the empirical performance of the standard formulation. We will first provide evidence suggesting that learning rates are indeed very unpredictable except in unusual cases, and that terminal productivities are often quickly attained. We will then show how, even in cases where the standard formulation appears to work well, it may do so even though cost reductions are not in fact caused by increases in cumulative output.

Unpredictable Learning Curves

It is well-known that younger firms grow faster than older firms (Dunne, Roberts, and Samuelson 1989), and there is evidence that at least part of this is explained by more rapid productivity growth among young firms (de Kok, Brouwer, and Fris 2006). At the same time, old plants using dated technology on average have similar productivity to new plants using the latest equipment (Jensen, McGuckin, and Stiroh 2001). These phenomena are all consistent with the standard formulation of the organizational learning curve. However, there remain enormous cross-sectional differences in firm productivity even after controlling for age and output, the causes of which remain poorly understood (Bartelsman and Doms 2000). One possible explanation, of course, is that there are large variations in rates of organizational learning.

Recall that the Boston Consulting Group based its advice back in the 1960s on the claim that there was a "consistent" decline of 20–30 percent in costs each time cumulative output doubled; the empirical evidence does not support this claim. Figure 1 plots the distributions of 271 progress ratios collected for two different review articles (Ghemawat 1985; Dutton and Thomas 1994), along with my own calculations for 11 shipyards engaged in building Liberty ships in World War II (broken out with dots below the horizontal axis). The central tendency of these estimates is clearly around the 70–80 percent progress ratio that BCG claimed as the standard. But the variation around that level is in fact very large. Indeed, the figure

Figure 1
271 Estimated Progress Ratios
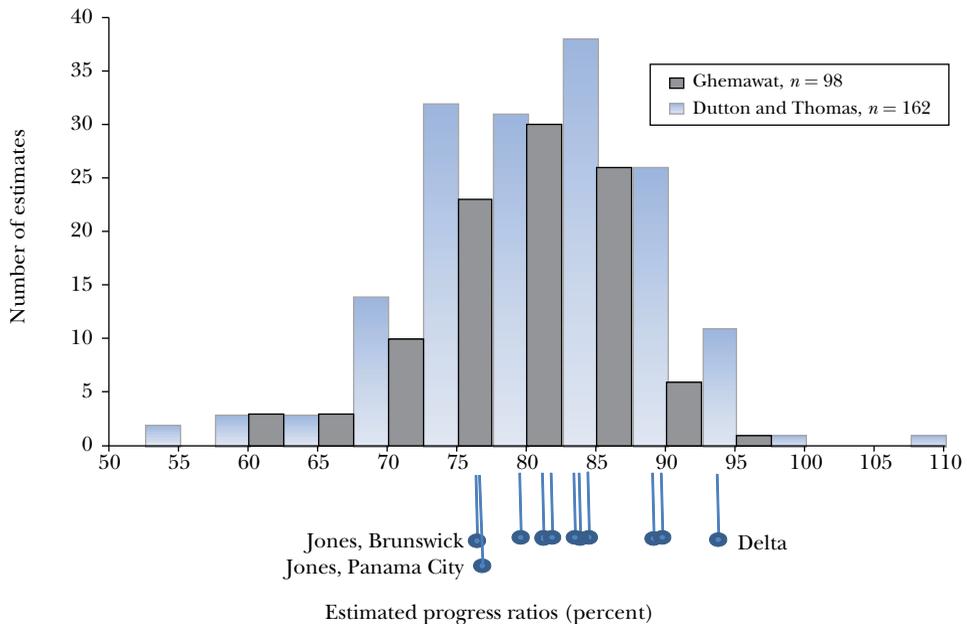
Source: Ghemawat (1985), Dutton and Thomas (1994), and author's calculations for Liberty shipyards.
Notes: Figure 1 plots the distributions of 271 progress ratios collected for two different review articles, along with my own calculations for 11 shipyards engaged in building Liberty ships in World War II (broken out with dots below the horizontal axis). An unknown fraction of the observations from the two papers Ghemawat (1985) and Dutton and Thomas (1994) are common to the two frequency distributions. "Jones" is the J.A. Jones Construction Company. "Delta" is the Delta Shipbuilding Corporation.

almost certainly underestimates the end of the distribution with near-zero and even negative learning rates that would be produced by studying a random sampling of firms. After all, there is little value in estimating and attempting to publish learning rates for industries where organizational learning is not expected to be present.

Why do progress ratios vary so much? One explanation is that technologies differ across industries and products, and some are more amenable to learning than others. Jordan (1958), for example, concluded that in the airframe industry the proportion of labor whose pace of work was determined by machines had a strong influence on the progress ratio. Similarly, Hirsch (1956) found that progress ratios in assembly jobs were about twice the size of machine-paced ratios.

However, the extent to which technology can explain this variation is unclear. Wide variation in rates of learning have also been documented for different plants operating the same technology, and even for different runs on the same product within plants (Yelle 1979). Even within technologies and plants, differences in learning rates may reflect the varied outcomes of "bets" by management and workers as to which approach will be most readily mastered (Nelson 1981).

Figure 2
Unit Labor Requirements on Liberty Ships, Three Shipyards
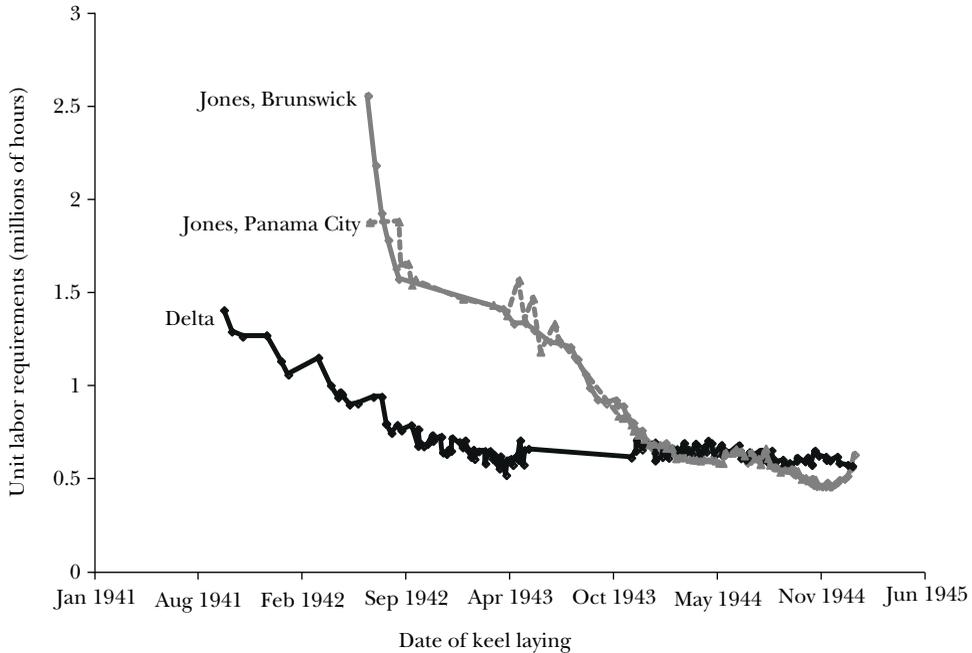
Source: Author's calculations. Data available at ⟨http://pthompsonecon.com/⟩.
Note: Observations on modified designs, tankers, and colliers are omitted.

However, look at the variation in Figure 1 of the progress ratios for 11 yards engaged in the production of Liberty ships during World War II. Two yards, both operated by J.A. Jones Construction Company, learned much more rapidly than other yards, attaining ratios of about 75 percent. At the other end of the scale, the Delta Ship-building Corporation was able to attain a progress ratio of only 93 percent. What is particularly puzzling about the Liberty ship experience is that these yards had limited ability to make different bets: they were engaged in the production of an identical, standardized ship, using the same methods in purpose-built yards with similar layouts. Moreover, these yards were not in competition with each other, and considerable effort was made to share lessons learned in one yard with the others.

Figure 2, which plots time-series of unit labor requirements for the Jones and Delta yards, suggests one reason for the difference in learning rates. The Jones yards appeared to learn so much not because they outperformed Delta, but because they performed so poorly until they caught up with Delta in December 1943. Unsur-prisingly, then, cost forecasts obtained by looking at the learning rates observed in other firms are likely to be highly unreliable. Of course, contemporaneous planners probably often have more information at their disposal than just the prior empirical record, so that they can adjust their observations of learning at other firms based

on their own situation. The managers at the Jones yards, for example, were well aware that their startup was unusually poor. However, the extensive historical record (Lane 1951) offers no explanation for the poor start in the Jones yards, nor any evidence to suggest that managers had anticipated it.

Formal evidence on the accuracy of contemporaneous expectations about future learning is hard to come by. There is, however, some indirect evidence contained in the contracts under which Liberty ships were built. In contract negotiations, each yard and the government agreed on the average production speeds and unit labor requirements for all ships to be delivered under the contract. The base payment for each ship was set as cost plus fixed fee. However, the government attempted to build incentives into each contract by paying bonuses for early completion, charging fees for late delivery, allowing the yard to retain part of any labor cost savings, and charging the yard for excess labor costs. To protect both yards and the government from excessive fees and penalties, the contracts placed upper and lower bounds on the fees that could be earned. The negotiated averages were adjusted as the war progressed, to reflect past experience and expectations of future efficiency gains. Of course, if projections about future efficiency were off the mark, so that actual and agreed hours differed greatly, the fees would hit their bounds. More often than not, this is exactly what happened. In my own study of 36 contracts awarded by the government (Thompson 2001), I found that in fully two-thirds of the contracts signed either the minimum or maximum fees were earned. In this case, contemporaneous expectations seem askew with outcomes, at least in the sense that contractual protections were written that view these outcomes as extreme cases.

Learning curves contained in tender documents also provide information about expectations of learning, although few have found their way into the academic literature. Figure 3 provides one example. In 2001, the German firm Babcock Noell Nuclear submitted a tender to produce superconducting dipole magnets, the largest and most complex equipment that was installed in CERN's Large Hadron Collider. Figure 3 shows the expected learning rate for the collared coils (the largest part of the magnets), as evidenced by the tender document, along with the realized unit labor hours. The agreement between expectations and outcomes is remarkable, and stands in stark contrast to the frequent forecast errors evidenced in the Liberty ship contracts. Indeed, three firms were involved in production of the coils, and the forecast accuracy illustrated in Figure 3 was typical (Rossi 2004, 2007).

However, the contracts for the magnets were unusual in many respects. CERN had already conducted ten years of research and development on the production process before Babcock Noell became involved, during which time it had built prototypes and made numerous design changes. CERN then shared this work with the firm as part of an extensive program of technology transfer. CERN also developed and tested specialized tools, which were provided to Babcock Noell, along with all the main components and ongoing engineering and technical support. As a result, "the dipoles were 'build to print' and 'build to process', with only minor degrees of freedom left to the companies in certain areas, notably in the coil winding and

Figure 3

Actual and Expected Learning Curves for Dipole Collared Coils at Babcock Noell Nuclear
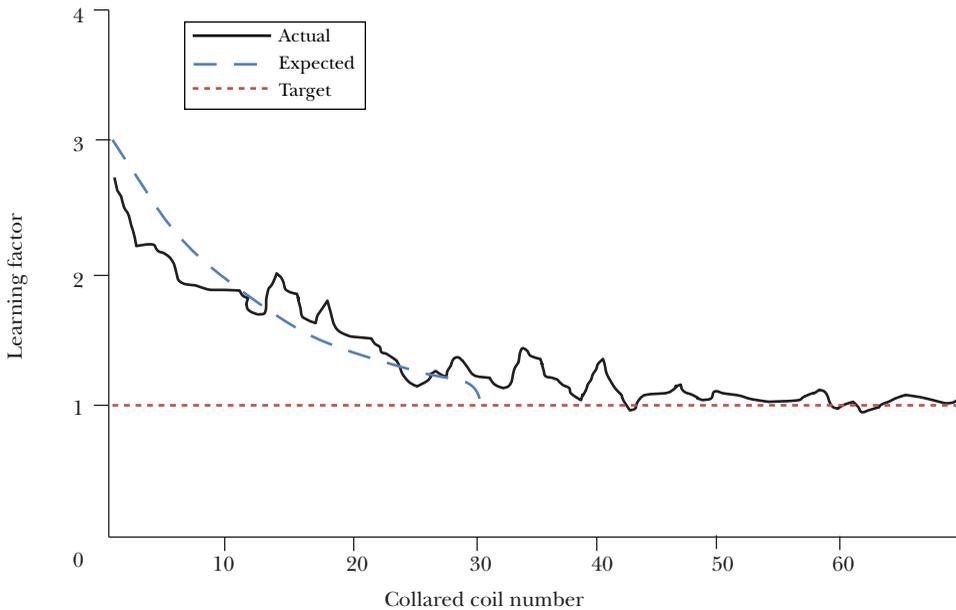
Source: Rossi (2004).
Notes: Not including manufacturing hours for breakdowns, repairs, and interruptions. Includes estimates for magnets started but not yet completed.

pole assembly" (Rossi 2007). The firm itself also undertook extensive preparatory research, in conjunction with the University of Hanover, on the sequence of operations necessary to manufacture the magnets. Clearly, many uncertainties that would otherwise have made learning curve forecasting a formidable task had already been resolved. The unusual circumstances surrounding the contracts for the magnets perhaps serves mainly to reveal, by their absence from the situation surrounding the Babcock Noell tender, the driving forces behind the uncertainty over future learning curves that usually prevails in more typical settings.

Terminal Productivities

Learning curves often drop toward a terminal productivity: for example, this flattening out of the cost reductions appears in the Liberty ship example in Figure 2 and the expected and actual Babcock Noell learning curves in Figure 3. This characteristic of learning is often overlooked. Empirical learning curve studies in other settings, especially in the earlier literature, contain production runs that are often too short to ascertain whether a terminal plateau has been reached or is close at hand. But the evidence from longer production runs suggests terminal productivities can be attained rather quickly.

Figure 4
A Learning Curve for Final Assembly of a Large Electromechanical Product
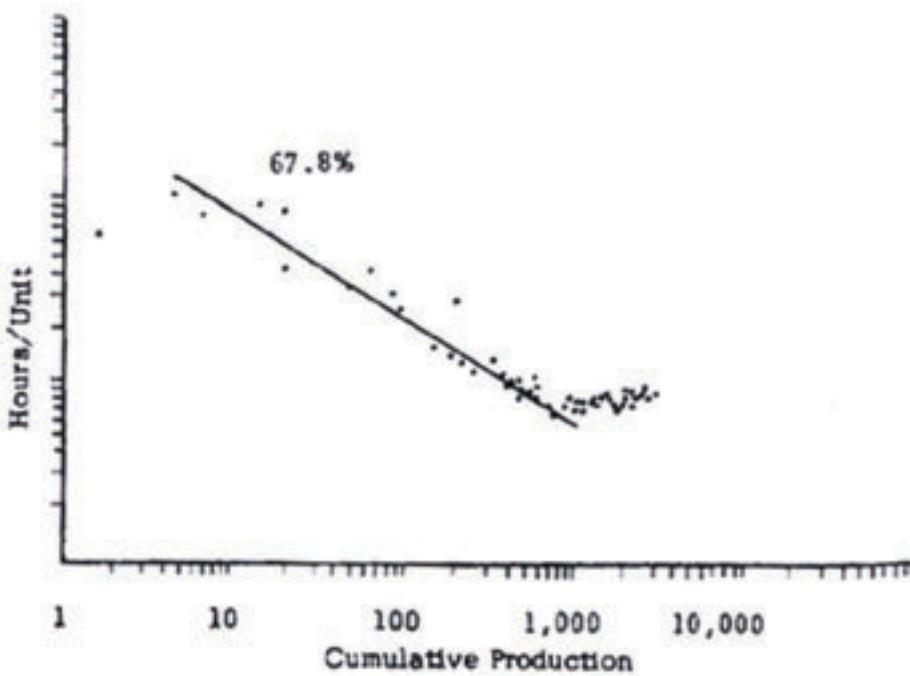
Source: Conway and Schultz (1959, figure 9).
Notes: Both axes are log scales. The regression line fits a 67.8 percent progress ratio indicating that prior to attaining terminal productivity, unit costs fell by 32.2 percent with each doubling of output. The source did not provide a scale for the vertical axis.

To provide just one additional example, Figure 4, taken from Conway and Schultz (1959), shows labor hours per unit required in the final assembly of "a large electromechanical product" during a long production run in a single plant; evidently, the attainment of a terminal productivity can at times be extremely abrupt. In Conway and Schultz's paper, in fact, six out of ten plots revealed compelling evidence that a terminal productivity had been attained and progress had stopped altogether. Their study was one among several that led Baloff (1966) to assert that although the power rule curve may describe the start-up phase in manufacturing, it does not describe behavior in the long term.

Cumulative Output versus Time

The standard formulation of the learning curve supposes a *causal* effect of cumulative production on current unit cost. But suppose that learning proceeds as a function of time rather than accumulated output (as in, for example, Jovanovic and Nyarko 1995). In this case, increasing output would not lead to more rapid learning, undermining much of the strategic importance of the learning curve.

However, a naive regression with cumulative production as the only explanatory variable would, by virtue of the trend inherent in the regressor, produce spurious evidence in favor of the standard formulation.

There is, however, a substantial literature that has confirmed the robustness of the power rule against the alternative of a possible time trend by running regressions of the form

$$\ln(c_t) = a_0 + a_1 \ln(y_{t-1}) + a_2 t + u_t,$$

in which c_t is unit cost at time t, y_{t-1} is cumulative output during the previous time period, and the coefficient on t captures the time trend. Typical results (for example, Rapping 1965) are that separate regressions of $\ln(c_t)$ on $\ln(y_{t-1})$ and on t each fit about equally well. However, when the two regressors are included together, the coefficient on $\ln(y_t)$ is statistically significant and negative while the coefficient on calendar time is either insignificant or of the wrong sign.

Under conditions typical of learning curve studies, it is hard to determine which of these results is more reliable. High degrees of serial correlation in productivity regressions is a common phenomenon (for example, Olley and Pakes 1996), and they appear to be especially problematic in learning curve estimation. For example, Benkard (2000, table 1) reports first-order serial correlation coefficients of 0.73 and 0.97 in his basic learning curve estimates for Lockheed aircraft, while Argote, Beckman, and Epple (1990) report significant third-degree serial correlation in their shipbuilding study. The problem for ordinary least squares estimation of a regression equation, of course, is that serial correlation induces a correlation between lagged cumulative output and the disturbance: a negative shock to cost generally induces a firm to hire more inputs and increase output, so that next period's shock is correlated not only with the current shock but also current cumulative output. This correlation leads to an overestimation of the magnitude of the coefficient on cumulative output and, because time and cumulative output are positively correlated, to an attenuation of the coefficient on time.

To show that serial correlation may induce serious problems in learning curve regressions, we constructed artificial data sets under the assumption that costs decline only as a function of time, such that the true coefficients in the earlier equation are $a_1 = 0$ and $a_2 = -.03$. Samples of 100 observations were then constructed, where the firm's output level is chosen optimally given its current cost and the elasticity of demand. We then used the regression above, which include both a term for cumulative output and a time trend, to estimate the coefficients for this data. The results are summarized in Figure 5. When serial correlation is modest with a correlation coefficient of $\rho = 0.2$, the regressions return accurate results: the coefficient on time is around $-.03$, while the coefficient on cumulative output is centered around zero. In contrast, when serial correlation becomes more extreme with $\rho = 0.8$, the coefficient on time is strongly attenuated, sometimes the wrong sign, and frequently statistically insignificant, while the coefficient on

Regression Coefficients on Artificial Learning Curve Data

Estimated coefficient for log of cumulative output

cumulative output is large in magnitude (and always significant). These effects are most pronounced when the elasticity of demand is high, so that output responds strongly to cost shocks. As a result, when serial correlation is sufficiently strong, an ordinary least squares approach incorrectly supports the standard formulation.

The solution, of course, is to find settings in which cost shocks do not induce changes in input use, or to find valid instruments for cumulative output. In Thompson (2001), I made the case that the Liberty ship program is an instance of the former solution, but such settings are almost vanishingly rare. Unsurprisingly, since the much greater part of the empirical learning curve literature predates the wide use of instrumental variable techniques (Angrist and Krueger 2001), the body of literature offering reliable support for the standard formulation is smaller than is generally supposed. One exception worth mentioning is Benkard's (2000)

study of learning in the production of Lockheed L-1011 Tri-Star passenger jets, one of the few to have deployed a plausible instrumental variable approach.[2]

Sources of Learning

How does a rise in cumulative output reduce costs? Remarkably little is understood about what this mechanism might be. After reviewing the evidence on learning in Liberty shipyards, Lucas (1993, p. 262) concluded:

> There is . . . considerable ambiguity about what this evidence means. Is it the individual worker who is doing the learning? The managers? The organization as a whole? Are the skills being learned specific to the production process on which the learning takes place, or more general? Does learning accrue solely to the individual worker, manager, or organization that does the producing, or is some of it readily appropriable by outside observers?

We do not have answers to these questions. However, it seems implausible that organizational learning can be explained as an aggregation of worker learning, and therefore the reduction in costs as cumulative output rises must be in some broad sense organizational in nature. In this section, after first making the case against aggregation of individual learning, we will review some examples in which the principle sources of sustained organizational learning seem to have been identified. In a number of examples where apparent learning has been sustained, the drivers of cost reduction within the organization are often not fully consistent with the standard formulation of the organizational learning curve.

Can Individual Learning Explain Organizational Learning?

The standard pattern in datasets of individual learning is that productivity quickly attains an upper bound. For example, Jovanovic and Nyarko (1995) collected a number of datasets on individual learning. As one example, they offer evidence that the productivity of new line-workers at a British munitions factory in World War I attained an upper bound within five weeks of initiating employment. Waldman, Yourstone, and Smith (2003) report similar results for surgeons, while Mazur and Hastie (1978) have found rapid exhaustion of learning in laboratory experiments. This evidence is consistent with popular theoretical models of individual learning, including the replacement and accumulation models popular among psychologists (Restle and Greeno 1970), as well as with the Bayesian "dial-setting model" of Jovanovic and Nyarko (1995).

---

[2] Benkard (2000) provides clear evidence that the standard formulation is misspecified: learning essentially stops halfway through the production run. His findings are consistent with the attainment of a terminal productivity. However, Benkard makes a strong case that learning appears to stop because organizations also forget.

One of the challenges in drawing inferences about organizational learning from individual learning is that workers come and go, so the translation from worker learning to organizational learning cannot be a one-to-one mapping. Is there some way in which the continual arrival of new workers might translate into an extended period of organizational learning? Probably not—but here is some of the underlying analysis.

First, imagine that labor turnover is independent of job tenure. In this case, with workers coming and going, individual learning does nothing to extend the duration of the organizational learning phase: mean output per worker simply attains, at the same speed, a lower terminal value than individual output. Next, imagine that the rate at which workers separate from jobs tends to decline with job tenure, which seems the common pattern (Jovanovic 1979). This pattern does serve to delay the time at which the organization attains its terminal productivity, but it seems doubtful that the effect can be sufficiently strong for our purposes. For example, I carried out simulation results in which workers attain their terminal productivity after six months employment, and the separation rate of workers declines in exaggerated fashion, from 20 percent in the first period to only 2 percent for workers with tenure of six months or more. The effect is only to increase the time taken to approach the firm's terminal productivity from 6 to 12 months, not nearly the length of the organizational learning curves that we see. (Conversely, a separation rate that increases with job tenure would accelerate the speed at which individual learning translates into the associated organizational learning.)

Further delays in the translation of individual learning into the attainment of the organization's terminal productivity could arise if the firm is able to learn and exploit variations in individual learning rates to reallocate workers to the most appropriate tasks (Prescott and Visscher 1980), and if this reallocation takes a sufficient amount of time. The process may also be drawn out beyond the period in which individuals learn as the firm dismisses workers that have failed to learn and hires replacements who have yet to demonstrate their ability to learn. However, Figure 6, which charts time-to-build for the French supplier of magnets to the Large Hadron Collider, offers some good evidence against these hypotheses. The observed learning curve shows three episodes in which there is a large decline in the speed of production. These are associated with intensive recruitment episodes, at times indicated by the bold arrows, involving increases in employment of direct labor of 300 percent, 50 percent, and 33 percent. Despite these very disruptive adjustments to the size of the workforce, and the specialized firm-specific human capital required for production, productivity returned to trend very quickly. The evidence strongly suggests that worker learning and organizational learning are driven by substantively different processes.

I know of only one case in which variations in labor turnover can explain a mapping from rapid individual learning to sustained and more gradual organizational learning. The story is worth telling, largely because it is exceptional. David (1973) documents an example of learning in the Lawrence Manufacturing Company Mill No. 2, a textile mill established in Lowell, Massachusetts, in 1834. There was

Figure 6

Actual and Expected Learning Curves for the Manufacturing of a Dipole Magnet at Jeumont
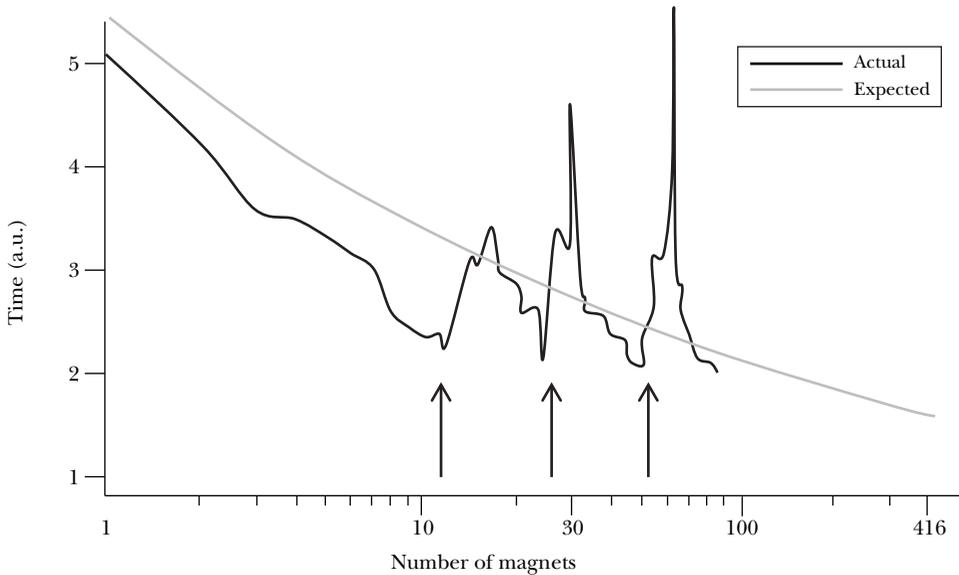
Source: Rossi (2004).
Note: Arrows indicate points in time when new workers were hired.

essentially no capital investment during 20 years that followed the founding of the mill—in particular, every loom that had been installed in 1834 was still in operation in 1856. Nonetheless, output per worker rose by an average 2 percent per year during this period.

Using detailed personnel records that survive in the Baker Library at Harvard, Bessen (2003) shows how the organizational learning at Lowell is explained by a fundamental demographic shift that took place in Lowell over the period 1834–56. In the 1830s, the labor force in the mill consisted primarily of "Yankee farm girls," who lived in boarding houses under paternalistic contractual arrangements with the mill. The farm girls were literate, but two characteristics limited their productivity. First, they tended to have little experience, it being the norm to abandon work in the mills upon marriage. Second, they frequently did not work in the mills during the summer, either returning to the farm to help during a busy time of year or taking summer teaching jobs. Both characteristics limited the extent to which the farm girls could learn from experience.

Beginning in the late 1830s the supply of farm girls began to fall behind demand: the number of mills in Lowell doubled between 1835 and 1847, at the same time that the New England farming population was declining. Offsetting these changes, the population of Lowell rose markedly, primarily due to an influx of Irish-born immigrants. These new arrivals had fewer outside opportunities and

consequently lower labor turnover rates. Coincident with these demographic shifts, managers began to increase the number of looms per worker first from two to three in the early 1840s, and then four a few years later. The capital deepening has been termed the "stretch-out."

Bessen argues that the increase in the number of looms per worker was a natural response to declines in the rate of labor turnover. To support this claim, Bessen shows that workers assigned to just two looms learned more quickly than those assigned to three or four, although the latter eventually became more productive. Initial productivity for those working on two looms was about 25 percent of terminal productivity, and it took about six months to attain terminal productivity. For those working on three or four looms, initial productivity was less than 20 percent of terminal productivity, which took a year to attain. The profitability of the stretch-out therefore depended upon the labor turnover rate: workers must have been expected to remain in the job long enough to recoup the greater initial investment in human capital associated with assignment to more than two looms. Bessen calculates the profitability of the stretch-out directly as a function of the turnover rate: it was profitable in the 1840s after sufficient numbers of immigrants were working the mills, he concludes, but not in 1834 when Yankee farm girls dominated the workforce.

While the Lowell mills illustrate a case in which bounded individual learning translated into much longer-term organizational productivity growth, it did so only as a result of profound changes in demography and the characteristics of labor. But this situation is obviously rare. When most people refer to organizational learning curves, they are usually referring to a roughly similar workforce and capital stock over time, which nonetheless experiences steady reductions in unit costs as cumulative output rises.

Organization-Level Drivers

The most obvious danger in estimating organizational learning curves is that the conventional measure of experience, cumulative output, is correlated with variables known to be associated with higher productivity but that are often not available to the researcher. For example, greater cumulative output is often correlated with a higher base of installed capital, and the omission of capital investment has proved particularly problematic in the interpretation of some classic case studies. For example, in Thompson (2001), I point out that early studies of the Liberty shipbuilding program, which did not have access to data on the capital stock, constructed a crude proxy for capital that was essentially constant over time. I recovered capital stock data from the National Archives for six of the 13 Liberty shipyards, showing dramatic increases in installed capital that coincided with reductions in the unit labor requirement. I concluded that for these yards, more than half of the increase in output per worker was accounted for by capital deepening. Similarly, Mishina (1999) undertook a closer look at Alchian's (1963) sample of aircraft factories, concluding that, inter alia, capital investments were a significant source of labor productivity growth in the production of the flying fortress bomber.

Table 1

Distribution of Ordinary Least Squares Estimates of Learning Parameter for 99 Specialty Chemicals

Range	Formal R&D group	Informal R&D group	
		Testing reduced	*No reduction in testing*
$\hat{\beta} < -1.4$	0	0	4
$-1.4 < \hat{\beta} < -1.0$	0	0	5
$-1.0 < \hat{\beta} < -0.6$	0	0	2
$-0.6 < \hat{\beta} < -0.2$	2	0	11
$-0.2 < \hat{\beta} < 0.2$	1	1	12
$0.2 < \hat{\beta} < 0.6$	5	4	12
$0.6 < \hat{\beta} < 1.0$	5	0	4
$1.0 < \hat{\beta} < 1.4$	3	4	1
$1.4 < \hat{\beta}$	9	4	3
Number	25	13	59
Means	1.20	1.04	−0.09
		0.11	

Source: Sinclair, Klepper, and Cohen (2000).

An especially interesting example of potential omitted variable bias is found in Sinclair, Klepper, and Cohen's (2000) look inside the specialty chemicals division of a Fortune 500 company. The authors had privileged access to a wealth of information, including batch-specific manufacturing costs and output for over 100 chemicals, and (most remarkably) *chemical-specific* expenditures on research and development. What makes this study especially interesting is that it documents, within a single case, instances of likely omitted variable bias because of the coexistence of different sources of cost reduction within a single firm, including instances where production experience clearly plays a role.

Sinclair, Klepper, and Cohen (2000) estimated chemical-specific learning curves for 120 chemicals produced in batches. These chemicals were divided into two groups. In one, chemicals were the subject of a formal research and development effort aimed at improving the production process. In the other group, which we shall call the informal research and development group, no attempt was made to modify the production process, but all of the chemicals were the subject of a project intended to reduce the amount of interim testing that took place during production. For each chemical in this latter group, a team was formed to study which stages of the production process always seemed to run smoothly and therefore did not need testing.

Table 1 reports the distribution of estimated learning rates, $\hat{\beta}$. There is a marked contrast between the informal and formal research and development groups. The $\hat{\beta}$s are positive for almost all the chemicals in the formal research and development groups, with an average of 1.20; the average for the informal R&D group is only one-tenth the size, with many estimates reporting rising costs over

time. The table further divides the informal research and development group into two subgroups: chemicals for which the study team was successful in reducing testing costs and those for which it was not. The former subgroup looks much like the research and development group: all learning rates are positive, with a mean of 1.04. Among the latter subgroup, in contrast, the learning rates are all over the place with on average a negative learning rate.

What can be learned from these results? Consider first the formal research and development group, which exhibited the highest learning rates. One possible conclusion is that the apparent learning curves reflect an omitted variable bias— that research and development, and not learning in the sense intended by the standard formulation, is the driver of cost reductions. An alternative conclusion is that that production experience revealed which chemicals had problems that might be addressed by research and development, so formal research and development just happens in this case to be the channel through which experience was mediated. Sinclair, Klepper, and Cohen (2000) argue that experience had little effect on the amount of formal research and development conducted, noting that requests for process research and development on a particular chemical most commonly came from marketing and sales personnel after they had identified a large potential demand if only production could be scaled up and units costs reduced. Unfortunately, we cannot know how much experience aided the marketing and sales team in their deliberations.

Ambiguities also exist in the interpretation of the results for the informal research and development group. The information used to decide whether eliminating a test was prudent clearly came from prior experience, indicating that in this group informal research and development was indeed a mediating channel for experience. However, we do not know whether testing reductions were more likely to be made among chemicals with greater cumulative output. Thus, even within the informal research and development group, the study is silent on the accuracy of the standard formulation of the experience curve, on which so much of the strategic importance of organizational learning depends.

Conclusion

In this paper, we have attempted to offer a cautionary note about the thesis of organizational learning-by-doing. In the standard formulation of organizational learning, cost reductions are obtained as a predictable by-product of accumulated production volume, through a process of learning that continues without limit. The reliability of these assumptions matters, because much of our understanding of the economic and strategic implications of organizational learning is built upon models that use the standard formulation. However, key components of the standard model may be less reliable than is typically supposed: not only are variations in the rate of learning difficult to predict, they are difficult to understand after the fact; learning often stops suddenly, with productivity frequently attaining a terminal

value after quite a short period of time; and evidence that cumulative output is the driver of cost reductions is contaminated by a variety of econometric difficulties.

A large part of the difficulty is that the standard formulation of the organizational learning curve is a reduced form for what is in reality a complex and varied set of processes by which firms secure increases in productivity. In some cases—for example where workers in a new firm learn on the job—production experience may have quite direct effects on performance. In such cases, however, the learning process appears to be bounded and quite short-lived. In other instances, experience may affect performance only when managers undertake costly actions, such as research and development. These mediated effects may well lead to learning gains that are more drawn out, but they are especially difficult to interpret. Indeed, in many cases the observed relationship between unit and cumulative quantity produced may be only coincidental, rather than evidence of a true mediated learning-by-doing effect.

Economists continue to have remarkably little understanding of the processes by which production experiences translate into organizational learning-by-doing and, until we do, we will continue to lack the sort of evidence necessary to justify our theorizing about its strategic importance. Developing an understanding requires that we continue to dig deep into the workings of individual firms. However, as Sinclair, Klepper, and Cohen's (2000) study of the specialty chemical firm illustrates, doing so depends on fortuitous access to internal information that only rarely becomes available. Even then, the complexity of real life within large firms is likely to raise more questions than the available information can answer.

One way around such complexity is to identify instances in which the effects of experience are expected to be relatively transparent. The construction of the dipole magnets for CERN discussed earlier is one such case, and there is surely much more we can learn from that historic construction project. The repeated improvements in yields observed in the production of successive generations of semiconductors (for example, Irwin and Klenow 1994) may be another such setting. In a notable recent example of the sort of research we have been lacking, Levitt, List, and Syverson (2012) study how a large automobile plant eliminated defects in a new car model. Exploiting what appears to be essentially unfettered access to the plant's internal records, they are able to show, inter alia, that organizational learning was not embodied in workers and that defects arose from problematic operations undertaken at individual stations along the production line, and in addition to document the speed with which these operations were improved.

However, limiting studies to cases where the effects of experience are expected to be transparent can only be the first step. This is because in such cases the learning curve will, almost by construction, be short-lived, and hence of no relevance to strategy. Indeed, in Levitt, List, and Syverson (2012), 90 percent of the total reduction in defect rates was attained in just six weeks. Developing analogous studies over the much longer time horizons relevant to strategy seems a much harder problem.

References

Alchian, Armen. 1963. "Reliability of Progress Curves in Airframe Production." *Econometrica* 31(4): 679–93.

Angrist, Joshua D., and Alan B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15(4): 69–85.

Argote, Linda, Sara L. Beckman, and Dennis Epple. 1990. "The Persistence and Transfer of Learning in Industrial Settings." *Management Science* 36(2): 140–54.

Argote, Linda, and Dennis Epple. 1990. "Learning Curves in Manufacturing." *Science* 247(4945): 920–24.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29(3): 155–73.

Asher, Harold. 1956. *Cost–Quantity Relationships in the Airframe Industry.* Santa Monica, CA: RAND Corporation.

Baloff, Nicholas. 1966. "Startups in Machine-Intensive Production Systems." *Journal of Industrial Engineering* 17(1): 25–32.

Bartelsman, Eric J., and Mark Doms. 2000. "Understanding Productivity: Lessons from Longitudinal Microdata." *Journal of Economic Literature* 38(3): 569–94.

Benkard, Lanier C. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90(4): 1034–54.

Bessen, James. 2003. "Technology and Learning by Factory Workers: The Stretch-Out at Lowell, 1842." *Journal of Economic History* 63(1): 33–64.

Book, William F. 1908. *The Psychology of Skill.* University of Montana Publications in Psychology, Bulletin No. 53.

Bryan, W. N., and N. Harter. 1899. "Studies in the Telegraphic Language: The Acquisition of a Hierarchy of Habits." *Psychological Review* vol. 6, pp. 345–75.

Cabral, Luis M. M., and Michael H. Riordan. 1994. "The Learning Curve, Market Dominance, and Predatory Pricing." *Econometrica* 62(5): 1115–40.

Cabral, Luis M. M., and Michael H. Riordan. 1997. "The Learning Curve, Predation, Antitrust, and Welfare." *Journal of Industrial Economics* 45(2): 155–69.

Clarke, Frank H., Masako N. Darrough, and John M. Heineke. 1982. "Optimal Pricing Policy in the Presence of Experience Effects." *Journal of Business* 55(4): 517–30.

Conway, R. W., and Andrew Schultz, Jr. 1959. "The Manufacturing Progress Function." *Journal of Industrial Engineering* 10(1): 39–54.

Crawford, J. R. c.1942. *Learning Curve, Ship Curve, Ratios, Related Data.* Burbank, CA: Lockheed Aircraft Corporation.

Dar-El, Ezey M. 2000. *Human Learning: From Learning Curves to Learning Organizations.* Dordrecht: Kluwer Academic Publishers.

Dasgupta, Partha, and Joseph E. Stiglitz. 1988. "Learning-by-Doing, Market Structure, and Industrial and Trade Policies." *Oxford Economic Papers* 40(2): 246–68.

David, Paul A. 1973. "The 'Horndal Effect' in Lowell, 1834–1856: A Short-Run Learning Curve for Integrated Cotton Textile Mills." *Explorations in Economic History* 10(1): 131–50.

de Kok, Jan, Peter Brouwer, and Pieter Fris. 2006. "On the Relationship between Firm Age and Productivity Growth." Scales Research Reports H200617, EIM Business and Policy Research, Netherlands.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1989. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics* 104(4): 671–98.

Dutton, John M., and Annie Thomas. 1994. "Treating Progress Functions as a Managerial Opportunity." *Academy of Management Review* 9(2): 235–47.

Ebbinghaus, Herman. 1885. *Memory: A Contribution to Experimental Psychology.* Translated by Henry A. Ruger & Clara E. Bussenius [1913]. New York: Columbia University.

Fudenberg, Drew, and Jean Tirole. 1983. "Learning-by-Doing and Market Performance." *Bell Journal of Economics* 14(2): 522–30.

Ghemawat, Pankaj. 1985. "Building Strategy on the Experience Curve." *Harvard Business Review* 63(2): 143–49.

Henderson, Bruce. 1968. "The Experience Curve." *BCG Perspectives* No. 87.

Henderson, Bruce. 1973. "The Experience Curve—Reviewed. II. History." *BCG Perspectives* No. 125.

Hirsch, Werner Z. 1952. "Manufacturing Progress Functions." *Review of Economics and Statistics* 34(2): 143–55.

Hirsch, Werner Z. 1956. "Firm Progress Ratios." *Econometrica* 24(2): 136–43.

Hollis, Aiden. 2002. "Strategic Implications of Learning by Doing." *International Journal of the Economics of Business* 9(2): 157–74.

Irwin, Douglas A., and Peter J. Klenow. 1994.

"Learning-by-Doing Spillovers in the Semiconductor Industry." *Journal of Political Economy* 102(6): 1200–27.

Jensen, J. Bradford, Robert H. McGuckin, and Kevin J. Stiroh. 2001. "The Impact of Vintage and Survival on Productivity: Evidence from Cohorts of U.S. Manufacturing Plants." *Review of Economics and Statistics* 83(2): 323–32.

Jordan, R. B. 1958. "Learning How to Use the Learning Curve." *N.A.A. Bulletin* 39(5): 27–39.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87(5): 972–90.

Jovanovic, Boyan, and Yaw Nyarko. 1995. "A Bayesian Learning Model Fitted to a Variety of Empirical Learning Curves." *Brookings Papers on Economic Activity. Microeconomics* vol. 1995, pp. 247–305.

Kiechel, Walter. 2010. *The Lords of Strategy.* Cambridge, MA: Harvard Business Press.

Klepper, Steven. 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle." *American Economic Review* 86(3): 562–83.

Lane, Frederic C. 1951. *Ships for Victory: A History of Shipbuilding under the U.S. Maritime Commission in World War II.* Baltimore: Johns Hopkins Press.

Levitt, Steven D., John A. List, and Chad Syverson. 2012. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." Working paper, University of Chicago.

Lucas, Robert E. Jr. 1993. "Making a Miracle." *Econometrica* 61(2): 251–72.

Mazur, James E., and Reid Hastie. 1978. "Learning as Accumulation: A Reexamination of the Learning Curve." *Psychological Bulletin* 85(6): 1256–74.

Mishina, Kazuhiro. 1999. "Learning by New Experiences: Revisiting the Flying Fortress Learning Curve." In *Learning by Doing in Markets, Firms, and Countries,* edited by Naomi Lamoreaux, Daniel M. Raff, and Peter Temin, 145–84. Chicago: University of Chicago Press.

Nelson, Richard R. 1981. "Research on Productivity Growth and Productivity Differences: Dead Ends and New Departures." *Journal of Economic Literature* 19(3): 1029–64.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64(6): 1263–98.

Ovans, Andrea. 2011. "The Charts that Changed the World." *Harvard Business Review,* December.

Petrakis, Emmanuel, Eric Rasmusen, and Santanu Roy. 1997. "The Learning Curve in a Competitive Industry." *RAND Journal of Economics* 28(2): 248–68.

Prescott, Edward C., and Michael Visscher. 1980. "Organization Capital." *Journal of Political Economy* 88(3): 446–61.

Rapping, Leonard. 1965. "Learning and World War II Production Functions." *Review of Economics and Statistics* 47(1): 81–86.

Restle, Frank, and James G. Greeno. 1970. *Introduction to Mathematical Psychology.* Reading, MA: Addison-Wesley.

Rosen, Sherwin. 1972. "Learning by Experience as Joint Production." *Quarterly Journal of Economics* 86(3): 366–82.

Rossi, Lucio. 2004. "LHC Dipole Production Begins to Take Off." *CERN Courier,* January 27.

Rossi, Lucio. 2007. "The Longest Journey: The LHC Dipoles Arrive on Time." *CERN Courier,* October 5.

Sinclair, Gavin, Steven Klepper, and Wesley Cohen. 2000. "What's Experience Got to Do With it? Sources of Cost Reduction in a Large Specialty Chemicals Producer." *Management Science* 46(1): 28–45.

Spence, A. Michael. 1981. "The Learning Curve and Competition." *Bell Journal of Economics* 12(1): 49–70.

Stein, Jeremy C. 1997. "Waves of Creative Destruction: Firm-Specific Learning-by-Doing and the Dynamics of Innovation." *Review of Economic Studies* 62(2): 265–88.

Thompson, Peter. 2001. "How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study." *Journal of Political Economy* 109(1): 103–37.

Thompson, Peter. 2010. "Learning by Doing." In *Handbook of Economics of Innovation,* edited by Bronwyn Hall and Nathan Rosenberg, 429–76. Amsterdam: Elsevier/North-Holland.

Waldman, J. Deane, Steven A. Yourstone, and Howard L. Smith. 2003. "Learning Curves in Health Care." *Health Care Management Review* 28(1): 41–54.

Wright, Theodore P. 1936. "Factors Affecting the Cost of Airplanes." *Journal of the Aeronautical Sciences* 3(4): 122–28.

Yelle, Louis E. 1979. "The Learning Curve: Historical Review and Comprehensive Survey." *Decision Sciences* 10(2): 302–28.

Young, Alwyn. 1993. "Invention and Bounded Learning by Doing." *Journal of Political Economy* 101(3): 443–72.

Recommendations for Further Reading

Timothy Taylor

    This section will list readings that may be especially useful to teachers of under-graduate 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.

Potpourri

    How much is household production worth in the U.S. economy and how has it changed over time? Benjamin Bridgman, Andrew Dugan, Mikhael Lal, Matthew Osborne, and Shaunda Villones tackle this question in "Accounting for Household Production in the National Accounts, 1965–2010." "We find that incorporating home production in GDP raises the level of GDP 39 percent in 1965 and 25.7 percent in 2010." "The impact of home production has dropped over time because women have been entering the workforce. This trend is driven by an increasing trend in the wage disparity between household workers and employees (that is, the opportunity cost of household labor)." "During 1965 to 2010, the annual growth rate of nominal GDP was 6.9 percent. When household production is included, this growth rate

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives*, based at Macalester College, Saint Paul, Minnesota. He blogs at* ⟨*http://conversableeconomist.blogspot .com/*⟩*.*

drops to 6.7 percent." *Survey of Current Business,* May 2012, pp. 23–36. At ⟨http://www.bea.gov/scb/pdf/2012/05%20May/0512_household.pdf⟩.

The International Monetary Fund, in its April 2012 *World Economic Outlook,* uses the U.S. Home Owners' Loan Corporation, established in 1933 and eventually ended in 1951, as its primary example of how best to restructure large numbers of household mortgages to "help avert self-reinforcing cycles of household defaults, further house price declines, and additional contractions in output." "To prevent mortgage foreclosures, HOLC bought distressed mortgages from banks in exchange for bonds with federal guarantees on interest and principal. It then restructured these mortgages to make them more affordable to borrowers and developed methods of working with borrowers who became delinquent or unemployed, including job searches. HOLC bought about 1 million distressed mortgages that were at risk of foreclosure, or about one in five of all mortgages. Of these million mortgages, about 200,000 ended up foreclosing when the borrowers defaulted on their renegotiated mortgages. The HOLC program helped protect the remaining 800,000 mortgages from foreclosure, corresponding to 16 percent of all mortgages. HOLC mortgage purchases amounted to $4.75 billion (8.4 percent of 1933 GDP), and the mortgages were sold over time, yielding a nominal profit by the time of the HOLC program's liquidation in 1951. . . . A key feature of HOLC was the effective transfer of funds to credit-constrained households with distressed balance sheets and a high marginal propensity to consume, which mitigated the negative effects on aggregate demand discussed above. . . . Accordingly, HOLC extended mortgage terms from a typical length of 5 to 10 years, often at variable rates, to fixed-rate 15-year terms, which were sometimes extended to 20 years. . . . In a number of cases, HOLC also wrote off part of the principal to ensure that no loans exceeded 80 percent of the appraised value of the house, thus mitigating the negative effects of debt overhang discussed above." April 2012. At ⟨http://www.imf.org/external/pubs/ft/weo/2012/01/pdf/text.pdf⟩.

A panel of the National Research Council headed by Daniel S. Nagin and John V. Pepper has published *Deterrence and the Death Penalty.* The report refers back to a 1978 NRC report which concluded that "available studies provide no useful evidence on the deterrent effect of capital punishment." The latest study reaches the same conclusion: "The committee concludes that research to date on the effect of capital punishment on homicide is not informative about whether capital punishment decreases, increases, or has no effect on homicide rates." What are some of the problems of such studies? "Properly understood, the relevant question about the deterrent effect of capital punishment is the differential or marginal deterrent effect of execution over the deterrent effect of other available or commonly used penalties, specifically, a lengthy prison sentence or one of life without the possibility of parole. One major deficiency in all the existing studies is that none specify the noncapital sanction components of the sanction regime for the punishment of homicide. Another major deficiency is the use of incomplete or implausible models of potential murderers' perceptions of and response to the capital punishment component of a sanction regime. Without this basic information, it is impossible

to draw credible findings about the effect of the death penalty on homicide." The report can be ordered or a free PDF can be downloaded at ⟨http://www.nap.edu /catalog.php?record_id=13363⟩.

The "Hart-Scott-Rodino Annual Report: Fiscal Year 2011" from the Federal Trade Commission summarizes the number of mergers and acquisitions, and their size, and provides some discussion of antitrust enforcement last year. "The total dollar value of reported transactions rose dramatically from fiscal years 1996 to 2000, from about $677.4 billion to about $3 trillion. . . . [T]he dollar value declined to about $565.4 billion in fiscal year 2002, and $406.8 billion in fiscal year 2003. This was followed by an increase in the dollar value of reported transactions over the next four years: about $630 billion in fiscal year 2004, $1.1 trillion in fiscal year 2005, $1.3 trillion in fiscal year 2006, and almost $2 trillion in 2007. The total dollar value of reported transactions declined to just over $1.3 trillion in fiscal year 2008, and to $533 billion in fiscal year 2009, increased to $780 billion in fiscal year 2010, and $979 billion in fiscal year 2011." Available at ⟨http://www .ftc.gov/os/2012/06/2011hsrreport.pdf⟩.

Michael Lipton was one of the winners of the 2012 Leontief Prize for Advancing the Frontiers of Economic Thought awarded by the Global Development and Environment Institute at Tufts University. Lipton gave his acceptance talk on "Income from Work: The Food-Population-Resource Crisis in the 'Short Africa.': "'Africa' in this talk is 'the short Africa': excluding N Africa, Madagascar, Mauritius and South Africa. All these are sharply distinct from the rest of Africa environmentally, agriculturally and economically, and generally well ahead in mean income; poverty reduction; growth; farming (irrigation, fertilizer, seeds); and demographic transition. The short Africa is itself highly diverse, but no more so than is India or China." "'Scientific smallholder intensification' in Africa is no easy path to development. From global evidence, we know it's possible. Is it necessary? Initially, yes. Farm development is only the start of modernization away from agriculture; I'm no agricultural or smallholder fundamentalist. But I'm an income-from-work fundamentalist. 'The short Africa' by 2050 will have 2.3 times today's population—but 3.7 times today's 15–64-year-olds. They need an affordable initial path to workplaces giving income and respect. Otherwise, potential demographic dividend will become demographic disaster. But, with half the people still in severe poverty and States cash-strapped too, what initial path is 'affordable'? One, trodden elsewhere, is scientific intensification of smallholder farms. If there's an alternative, what is it?" At ⟨http://www.ase.tufts.edu/gdae /about_us/leontief/LiptonLeontiefPrizeComments.pdf⟩.

Mark Lemley discusses "Fixing the Patent Office." "Most patents don't matter. They claim technologies that ultimately failed in the marketplace. They protect a firm from competitors who for other reasons failed to materialize. They were acquired merely to signal investors that the relevant firm has intellectual assets. Or they were lottery tickets filed on the speculation that a given industry or invention would take off. Those patents will never be licensed, never be asserted in negotiation or litigation, and thus spending additional resources to examine them would

yield few benefits. . . . Some bad patents, however, are more pernicious. They award legal rights that are far broader than what their relevant inventors actually invented, and they do so with respect to technologies that turn out to be economically significant. . . . Compounding the problem, bad patents are too hard to overturn. Courts require a defendant to provide 'clear and convincing evidence' to invalidate an issued patent. In essence, courts presume that the Patent Office has already done a good job of screening out bad patents. Given what we know about patents in force today, that is almost certainly a bad assumption. . . . "The problem, then, is not that the Patent Office issues a large number of bad patents. Rather, it is that the Patent Office issues a small but worrisome number of economically significant bad patents and those patents enjoy a strong, but undeserved, presumption of validity." Lemley discusses proposals like allowing patent applicants to pay extra for a patent that would be considered more carefully, and would have a stronger legal presumption of validity. Long-time devotees of the *JEP* may recognize this argument, because it follows up on some arguments that Lemley made with co-author Carl Shapiro in "Probabilistic Patents" in the Spring 2005 issue. SIEPR (Stanford Institute for Economic Policy Research) Discussion Paper No. 11-014, May 21, 2012). At ⟨http://siepr.stanford.edu/?q=/system/files/shared/pubs/11-014.pdf⟩.

Why does the U.S. spend so much more on health care than other countries? David Squires assembles some of the evidence in "Explaining High Health Care Spending in the United States: An International Comparison of Supply, Utilization, Prices, and Quality." "This analysis uses data from the Organization for Economic Cooperation and Development and other sources to compare health care spending, supply, utilization, prices, and quality in 13 industrialized countries: Australia, Canada, Denmark, France, Germany, Japan, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The U.S. spends far more on health care than any other country. However this high spending cannot be attributed to higher income, an older population, or greater supply or utilization of hospitals and doctors. Instead, the findings suggest the higher spending is more likely due to higher prices and perhaps more readily accessible technology and greater obesity. Health care quality in the U.S. varies and is not notably superior to the far less expensive systems in the other study countries. Of the countries studied, Japan has the lowest health spending, which it achieves primarily through aggressive price regulation." Commonwealth Fund, May 2012. At ⟨http://www.commonwealthfund.org/~/media/Files/Publications/Issue%20Brief/2012/May/1595_Squires_explaining_high_hlt_care_spending_intl_brief.pdf⟩.

Energy and the Environment

The International Energy Agency discusses *Golden Rules for a Golden Age of Natural Gas.* (The IEA is an autonomous international agency with 28 member countries and a staff of 260. It was founded back in 1973–74 as part of the response to the oil price shock at that time.) "The United States and Canada still account

for virtually all the shale gas produced commercially in the world . . . There are large resources of all three types of unconventional gas across the United States. Of the 74 trillion cubic metres (tcm) of remaining recoverable resources of natural gas at end-2011, half are unconventional; in total, gas resources represent around 110 years of production at 2011 rates." The "Golden Rules" mentioned in the title are detailed guidelines for exploiting unconventional gas resource in an environmentally sensitive manner. "We estimate that applying the Golden Rules could increase the overall financial cost of development [of] a typical shale-gas well by an estimated 7%." 2012. At ⟨http://www.worldenergyoutlook.org/media/weowebsite/2012/goldenrules/WEO2012_GoldenRulesReport.pdf⟩.

Michael Greenstone and Adam Looney compare the full social cost of various ways of producing electricity in "Paying Too Much for Energy? The True Costs of Our Energy Choices." "Our primary sources of energy impose significant health costs—particularly on infants and the elderly, our most vulnerable. For instance, even though many air pollutants are regulated under the Clean Air Act, fine particle pollution, or soot, still is estimated to contribute to roughly one out of every twenty premature deaths in the United States. Indeed, soot from coal power plants alone is estimated to cause thousands of premature deaths and hundreds of thousands of cases of illness each year. The resulting damages include costs from days missed at work and school due to illness, increases in emergency room and hospital visits, and other losses associated with premature deaths. . . . The National Academy of Sciences recently estimated total non-climate change-related damages associated with energy consumption and use to be more than $120 billion in the United States in 2005." Their estimates suggest that existing coal resources are the lowest method of generating electricity in terms of private costs, but when social costs are taken into account, new natural gas resources are the least expensive method. *Daedalus*, Spring 2012, pp. 10–30. At http://www.amacad.org/publications/daedalus/12_spring_greenstone_looney.pdf.

Drew Shindell discussed "Beyond CO2: The Other Agents of Influence." "Black carbon ("soot") and ozone, of which methane is a primary precursor, are the only two agents known to cause both global warming and degraded air quality. Black carbon is emitted during incomplete combustion of fossil fuels and biomass, and methane is a gas emitted from many sources, including landfills and coal mines." Shindell reports on the results of an international team of researchers who started off by looking at 400 possible ways of reducing soot and methane emissions, and settled on 14 as having especially beneficial cost–benefit calculations: "Of the 14 measures selected, 7 target methane emissions (from coal mining, oil and gas production, long-distance gas transmission, municipal waste and landfills, wastewater, livestock manure, and rice paddies). The other 7 controls target black carbon emissions from incomplete combustion and include both technical measures (for diesel vehicles, biomass stoves, brick kilns, and coke ovens) and regulatory measures (for agricultural waste burning, high-emitting vehicles, and domestic cooking and heating)." *Resources*, 2012, no. 180, pp. 20–23. At ⟨http://www.rff.org/Publications/Resources/Pages/180-Beyond-CO2.aspx⟩.

From the Federal Reserve Banks

The Federal Reserve has published results from the most recent Survey of Consumer Finance, the triennial survey that is the canonical source for data on the wealth of households, in a bulletin called *Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances,* by a team of authors led by Jesse Bricker, Arthur B. Kennickell, Kevin B. Moore, and John Sabelhaus. "[O]verall, median net worth fell 38.8 percent, and the mean fell 14.7 percent . . . Median net worth fell for most groups between 2007 and 2010, and the decline in the median was almost always larger than the decline in the mean. The exceptions to this pattern in the medians and means are seen in the highest 10 percent of the distributions of income and net worth, where changes in the median were relatively muted. Although declines in the values of financial assets or business were important factors for some families, the decreases in median net worth appear to have been driven most strongly by a broad collapse in house prices . . ." *Federal Reserve Bulletin,* June 2012, vol. 98, no. 2. At ⟨http://www.federalreserve.gov/pubs/bulletin/2012/PDF/scf12.pdf⟩.

Willem Van Zandweghe has a more recent take, with updated evidence, in "Interpreting the Recent Decline in Labor Force Participation." "At the turn of the 21st century, labor force participation in the United States reversed its decades-long increase and started trending lower. A more startling development has been the recent sharp decline in the labor force participation rate—from 66.0 percent in 2007 to 64.1 percent in 2011—a far bigger drop than in any previous four-year period. . . . This article presents a variety of evidence—including data on demographic shifts, labor market flows, gender differences, and the effects of long-term unemployment—to disentangle the roles of the business cycle and trend factors in the recent drop in participation. Taken together, the evidence indicates that long-term trend factors account for about half of the decline in labor force participation from 2007 to 2011, with cyclical factors accounting for the other half." This article is a useful follow-up to the article "Changes in Labor Force Participation in the United States," by Chinhui Juhn and Simon Potter in the Summer 2006 issue of this journal. Federal Reserve Bank of Kansas City *Economic Review*: First Quarter 2012, pp. 5–34 At ⟨http://www.kc.frb.org/Publicat/EconRev/PDF/12q1VanZandweghe.pdf⟩.

David B. Humphrey and Robert Hunt tell the story of "Getting Rid of Paper: Savings from Check 21." "On September 11, 2001, planes were grounded and check float—the value of checks in the process of transportation and collection—rose to $47 billion (about eight times the normal daily level), while electronic payments were unaffected. Although the technology has been available for almost two decades to digitize check images and collect checks electronically on a same-day basis, the legal requirement of physical presentment inhibited its adoption. . . . The September 2001 disruption spurred the Federal Reserve to ask Congress to allow a paper representation of the digital image of the front and back of a check (called a substitute check) to be legally the same as the original physical item for purposes of collection and presentment. This legislation, adopted in 2003 and known as Check 21 (Check

Clearing for the 21st Century Act), along with other initiatives, currently permits almost all of the 24.5 billion checks paid annually in the U.S. (worth $32 trillion) to be collected electronically on a same-day (or next-day) basis once they are deposited at a bank. The original check is imaged and transported electronically, and a substitute check is printed close to where the paying bank is located. The substitute check is then physically presented for payment. Since accepting billions of substitute checks is more costly than accepting and paying the electronic image itself, almost all paying banks now receive and pay the image." The authors estimate that the change to electronic checks saves about $3 billion per year. Federal Reserve Bank of Philadelphia, Working Paper 12-12, May 2012. At ⟨http://www.philadelphiafed.org/research-and -data/publications/working-papers/2012/wp12-12.pdf⟩.

About Economists

The July 2012 issue of *Public Choice* is a special issue on "The intellectual legacy of Gordon Tullock," guest edited by Charles K. Rowley and Daniel Houser, with 16 articles mostly focused on an array of topics in political economy. The opening essay, by Rowley and Houser, tells of "The Life and Times of Gordon Tullock." Tullock had an unconventional career: for example, several years before Tullock held his first position as an economics professor, he had taken one total course in economics, served as an infantry rifleman during World War II, finished law school, been a Foreign Service Officer, and written articles on the economics of China and Korea that would be published in the *American Economic Review,* the *Journal of Political Economy*, and the *Economic History Review.*

Aaron Steelman has an "Interview with John B. Taylor." What is the Taylor rule for monetary policy? "The rule is quite simple. It says that the federal funds rate should be 1.5 times the inflation rate plus .5 times the GDP gap plus one. The reason that the response of the fed funds rate to inflation is greater than one is that you want to get the real interest rate to go up to take some of the inflation pressure out of the system. To some extent, it just has to be greater than one—we really don't know the number precisely. One and a half is what I originally chose because I thought it was a reasonably good benchmark. . . . In the late 1990s Chairman Greenspan told me that it explained about 80 percent of what they were doing during his tenure, but that doesn't mean that he was looking at it explicitly." Federal Reserve Bank of Richmond *Region Focus,* First Quarter 2012, pp. 29–33. At ⟨http://www.richmondfed.org/publications/research/region_focus/2012/q1 /pdf/interview.pdf⟩.

Discussion Starters

Jeffrey Passel, D'Vera Cohn, Ana Gonzalez-Barrera estimate that "Net Migration from Mexico Falls to Zero—and Perhaps Less." "The largest wave of immigration in

history from a single country to the United States has come to a standstill. . . . The U.S. today has more immigrants from Mexico alone—12.0 million—than any other country in the world has from all countries of the world. Some 30% of all current U.S. immigrants were born in Mexico. The next largest sending country—China (including Hong Kong and Taiwan)—accounts for just 5% of the nation's current stock of about 40 million immigrants. . . . Beyond its size, the most distinctive feature of the modern Mexican wave has been the unprecedented share of immigrants who have come to the U.S. illegally. Just over half (51%) of all current Mexican immigrants are unauthorized, and some 58% of the estimated 11.2 million unauthorized immigrants in the U.S. are Mexican." They discuss possible causes of a leveling off or drop in Mexico–U.S. immigration: on the U.S. side, a combination of heightened enforcement along with the recession and sluggish recovery; on the Mexican side, changes in birthrates and improved economic prospects. 2012. At ⟨http://www.pewhispanic.org/files/2012/04/PHC-04-23a-Mexican-Migration.pdf⟩.

Gary Clyde Hufbauer and Sean Lowry evaluate "US Tire Tariffs: Saving Few Jobs at High Cost." "Starting on September 26, 2009, Chinese tires were subjected to an additional 35 percent ad valorem tariff duty in the first year, 30 percent ad valorem in the second year, and 25 percent ad valorem in the third year." The higher tariffs did reduce tire imports from China. For example, "radial car tires imported from China fell from a high of approximately 13.0 million tires in 2009Q3 to 5.6 million tires during 2009Q4—a 67 percent decrease." "[T]he total cost to American consumers from higher prices resulting from safeguard tariffs on Chinese tires was around $1.1 billion in 2011. The cost per job manufacturing saved (a maximum of 1,200 jobs by our calculations) was at least $900,000 in that year. Only a very small fraction of this bloated figure reached the pockets of tire workers. Instead, most of the money landed in the coffers of tire companies, mainly abroad but also at home." Peterson Institute for International Economics, Policy Brief Number PB12-9, April 2012, At ⟨http://www.iie.com/publications/pb/pb12-9.pdf⟩.

Mark Lino of the U.S. Department of Agriculture has authored this year's version of *Expenditures on Children by Families, 2011.* "In 1960, average expenditures on a child in a middle-income, husband-wife family amounted to $25,229, or $191,723 in 2011 dollars. By 2011, these estimated expenditures climbed 23 percent in real terms to $234,900 . . . Housing was the largest expense on a child in both time periods and increased in real terms over this time. Food was also one of the largest expenses in both time periods, but decreased in real terms. . . . Perhaps the most striking change in child-rearing expenses over time relates to child care and education expenses. It should be noted that in 1960, child care/education expenses included families with and without the expense. Even so, these expenses grew from 2 percent of total child-rearing expenditures in 1960 (for families with and without the expense) to 18 percent (for families with the expense) in 2011. Much of this growth is likely related to child care." Available at ⟨http://www.cnpp.usda.gov/Publications/CRC/crc2011.pdf⟩.

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, we will also print a shorter, one-paragraph version of your notice in the "Notes" section of the* Journal of Economic Perspectives. *It is best to* **send announcements for JEP "Notes" before** *March 20 for the JEP Spring issue, which mails the end of May; before June 20 for the JEP Summer issue, which mails the end of August; before September 20 for the JEP Fall issue, which mails the end of November; and before December 10 for the JEP Winter issue, which mails the end of February. We reserve the right to edit material received.*

The Annual Meeting of the American Economic Association will be held in San Diego, California, January 4–6, 2013. The headquarters will be the Manchester Grand Hyatt. Registration and exhibits will also be in the Manchester Grand Hyatt. 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 San Diego. The location of the interview tables will be the San Diego Marriott Marquis & Marina. Registration will open mid-September, for additional information or to register for the meeting, please go to ⟨www .vanderbilt.edu/AEA⟩.

The Association's 2013 Continuing Education Program will feature three concurrent programs on January 6, 7, and 8, 2013, in San Diego, California. Participants can choose from among: Time Series Econometrics; Labor Economics; and Public Finance.

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. This is applicable even when the authors have no relevant interests to disclose. Authors will need to provide a separate Disclosure Statement for each coauthor at the time of submission. Submissions that do not include the statements will be considered incomplete and will not be reviewed. 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 Amy Finkelstein was awarded the John Bates Clark Medal for 2012.

2012 Distinguished Fellows. The Association is pleased to announce that the Distinguished Fellows for 2012 are Truman F. Bewley, Marc L. Nerlove, Neil Wallace, and Janet L. Yellen.

Foreign Honorary Members. The Association is pleased to announce Aloisio Araujo, Kenneth George Binmore, and Lars E. O. Svensson have been recognized as Foreign Honorary Members.

2013 Nominating Committee of the AEA. In accordance with Article IV, Section 2 of the Bylaws of the American Economic Association, President-elect Claudia Goldin appointed a Nominating Committee for 2013 consisting of Robert Hall (Chair), Marianne Bertrand, Jeff Kling, Robert Margo, Ted O'Donoghue, Valerie Ramey, Nancy Rose, and Andrei Shleifer. Attention of members is called to the part of the by-law reading: "In addition to appointees chosen by the President-elect,

the Committee shall include any other member of the Association nominated by petition including signatures and addresses of not less than two percent of the members of the Association, delivered to the Secretary before December 1. No member of the Association may validly petition for more than one nominee for the Committee. The names of the Committee shall be announced to the Membership immediately following its appointment and the membership invited to suggest nominees for various offices to the Committee."

Nominations of AEA Officers for 2013. 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 web sites 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 14, 2012.**

The Committee on the Status of Women in the Economics Profession (CSWEP) seeks nominations for the 2013 Elaine Bennett Research Prize, awarded every other year to recognize, support, and encourage outstanding contributions by young women in the economics profession. The next award will be presented in January 2013. Nominees should be at the beginning of their career and be within seven years of completing their dissertation but have demonstrated exemplary research

contributions in their field. 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 14, 2012.**

Call for abstracts. The AEA **Committee for the Status of Women in the Economics Profession (CSWEP)** will sponsor sessions at the January 2014 American Economic Association meetings in Philadelphia, PA. We will be organizing three sessions on gender-related topics and three sessions on econometrics topics. Accepted papers will be considered for publication in the Papers and Proceedings issue of the *American Economic Review*. Abstracts of individual papers and complete session proposals will be considered. Email a cover letter (specifying to which set of sessions the paper is being submitted) and a copy of a one- to two-page abstract (250–1000 words) clearly labeled with the paper title, authors' names, affiliation, and contact information for all the authors **by March 1, 2013**, to ⟨cswep@econ.duke.edu⟩.

Submissions invited for Warren Samuels Prize. This prize is awarded to a paper being presented at the January, 2013, ASSA meetings that best exemplifies scholarly work that: 1) is of high quality, 2) is important to the project of social economics, and 3) has broad appeal across disciplines. It is preferable, but not required, that the paper is presented at one of the ASSA sessions sponsored by the Association for Social Economics (ASE). The winning paper may, subject to peer review, be published in the *Review of Social Economy*, and the winner receives a $500 stipend. **Send papers before December 5, 2012,** to Wilfred Dolfsma, Corresponding Editor, *Review of Social Economy*, at ⟨w.a.dolfsma@rug.nl⟩. See the ASE website for details.

The 2012 Julius Shiskin Memorial Award for Economic Statistics will go to William D. Nordhaus, Sterling Professor of Economics at Yale University. This award recognizes unusually original and important contributions in the development of economic statistics or in the use of statistics interpreting the economy. Professor Nordhaus is recognized for his contributions to the measurement of environmental-economic accounts and economic welfare and his active participation with the U.S. statistical system.

Call for papers. A special issue of the *Journal of Family and Economic Issues* invites original empirical or theoretical research papers relevant to the **Health**

Systems Strengthening Experience, Economics, and the Role of Individuals and the Families. Health strengthening policies and activities mainly focus on the governance, financing, resources, organization and management of healthcare systems; they may miss the intricate relationship between the individuals and families and their economic environments. When the individuals' and families' choices and constraints, as well as, the self-interests of regulators and implementers are not taken into consideration, outcomes may be suboptimal. Policymakers, regulators, and the bureaucrats should incorporate individual and family dynamics in generating, maintaining, and consuming health, such that, the nation's welfare is maximized. Papers on this theme may be submitted at ⟨http://www.springer.com/social+sciences/journal/10834⟩ **before October 3, 2012.** For further information contact the Guest Editor Manouchehr (Mitch) Mokhtari, ⟨mokhtari@umd.edu⟩, at the School of Public Health, University of Maryland.

Call for papers. The 4th International Conference on Institutional and Technological Environment for Microfinance (ITEM4) will be organized by the Banque Populaire Chair in Microfinance of Burgundy School of Business on April 11–12, 2013 in Paris, France, and will focus on financial inclusion in developed and developing countries, including lessons the North can learn from the South, and lessons the South can learn from the North. Debates will also be focused on financial inclusion as an economic instrument for social solidarity. Other microfinance related topics will be discussed. Concurrent tracks will be held in French and in English. **Paper submission deadline is October 1st, 2012.** Papers must be sent to ⟨microfinancechair@escdijon.eu⟩. Indicate whether your paper is targeted for publication in *Cost Management*, *La Revue des Sciences de Gestion*, or *Strategic Change*. For further information, go to ⟨http://item4.weebly.com/english-version.html⟩.

2012 Application/Renewal for Membership

AMERICAN ECONOMIC ASSOCIATION
2014 Broadway, Suite 305
Nashville, TN 37203
Ph. 615-322-2595 fax: 615-343-7590
Federal ID No. 36-2166945
www.vanderbilt.edu/AEA

RENEWING MEMBERS, ENTER ACCT. NUMBER & EXP. DATE	IF PAYING BY CREDIT CARD, PLEASE FILL OUT BELOW
ACCOUNT NUMBER:	CARD NUMBER:
EXPIRATION DATE:	EXP DATE: CSC CODE:

FIRST NAME:	MI:	LAST NAME:
ADDRESS:		
CITY:	STATE/PROVINCE:	ZIP:
COUNTRY:	☐ Check here if non-US	
PHONE:	FAX:	
PRIMARY FIELD OF SPECIALIZATION:		
SECONDARY FIELD OF SPECIALIZATION:		
EMAIL:	☐ Check here to exclude your email address from the public directory	

Please include my email address to receive:
☐ Announcements about public policy affecting economists or the economics profession
☐ Surveys of economists for research purposes
☐ Commercial advertising

MEMBERSHIP DUES — Based on annual income. Please select one below.

☐ Annual income of $70,000 or less	$20	$
☐ Annual income of $70,000 to $105,000	$30	$
☐ Annual income over $105,000	$40	$

The AEA dues above include online access to all seven AEA journals.
For print or CD subscription(s) indicate preference below and add appropriate charge(s).

Journal		Print	Int'l Postage*	CD*	
AER	(7 issues, incl. P&P)	☐ Add $20	☐ Add $25	☐ Add $15	$
AER Papers & Proceedings Only*		☐ Add $10	n/a	n/a	$
JEL	(4 quarterly issues)	☐ Add $15	☐ Add $15	☐ Add $15	$
JEP	(4 quarterly issues)	☐ Add $15	☐ Add $15	☐ Add $15	$
AEJ: Applied	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$
AEJ: Policy	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$
AEJ: Macro	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$
AEJ: Micro	(4 quarterly issues)	☐ Add $15	☐ Add $15	n/a	$

* Int'l postage applies only to print journals mailed outside of the U.S. No additional postage is required for CDs or the AER Papers and Proceedings.

AEA Journals via JSTOR online

JSTOR	☐ Add $16	$

	Sub Total	$

Check One:	☐ 1 Year	☐ 2 Years	☐ 3 Years	TOTAL AMOUNT	$

Make checks payable to: American Economic Association.
Must be drawn on a US bank.
Apply online at http://www.aeaweb.org/membership.php

Payments must be made in advance. We accept checks (in US dollars only, with correct coding for processing in US banks) and credit cards; online or by faxing or mailing the application. Please choose one method; it is the Association's policy NOT TO REFUND dues.

American Economic Association

Membership Information

Our Publications include:
AMERICAN ECONOMIC REVIEW (AER) & the AER PAPERS AND PROCEEDINGS (MAY ISSUE)
JOURNAL OF ECONOMIC LITERATURE (JEL)
JOURNAL OF ECONOMIC PERSPECTIVES (JEP)
AMERICAN ECONOMIC JOURNAL: APPLIED ECONOMICS (AEJ: AE)
AMERICAN ECONOMIC JOURNAL: ECONOMIC POLICY (AEJ: EP)
AMERICAN ECONOMIC JOURNAL: MACROECONOMICS (AEJ: MAC)
AMERICAN ECONOMIC JOURNAL: MICROECONOMICS (AEJ: MIC)

Membership – In addition to having access to all seven academic journals of the Association, your benefits include:

- Receive the ***AER, JEL***, and ***JEP*** online and in print or CD.
- Receive online and in print the ***American Economic Journals (Applied Economics, Economic Policy, Macroeconomics, and Microeconomics).***
- Access to pre-publication accepted articles for **AEA journals**
- Discounts on submission fees for the ***AER*** and the ***AEJs.***
- Quarterly **AEA Virtual Field Journals:** Notification of articles in all of the AEA journals in the subject classifications of your choice.
- Access to **AEA journals in JSTOR** for an additional $16 annually.
- **EconLit for Members:** The EconLit online bibliography (without all the search features and full-text links provided in institutional settings).
- EconLit for Members update **alerts by JEL code(s)** of your choice.

- **Continuing Education Program** discounts.
- Listing in the **AEA Directory of Members**.
- **Group Term Life Insurance & Short Term Recovery Health Care** through Marsh US Consumer

    ### Only AEA members may:
- Vote in the annual election of officers and at the **Annual Business Meeting**.
- Contribute to AEJ and JEP online **"Discussion Forums"**
- Submit papers to be considered for presentation at the **AEA Annual Meeting Program.**
- View **webcasts of certain AEA Annual Meeting and Continuing Education sessions** online.

Regular Membership dues are based on annual income.

Student memberships are $35 and are for one year only. A multiple-year membership option is not available for students. Student status must be certified by a faculty adviser or school registrar with written verification.

Family Membership, for a person living at the same address as a regular member, is $14. This is an additional membership without publications in print or on CD. A family member must obtain access to AEA journals online.

Online Member benefits begin immediately. Requested journals in print or CD begin with the issue following posting of your payment. Membership may not be back-started. Journals are mailed second class; please allow 6 to 8 weeks for arrival of journals shipped outside the U.S. CDs are mailed First Class.

Contact Information

American Economic Association   **Phone: (615) 322-2595**
2014 Broadway, Suite 305   **Fax: (615) 343-7590**
Nashville, TN 37203   **www.vanderbilt.edu/AEA/**
aeainfo@vanderbilt.edu

It is important to include your e-mail address and to keep it up to date. It often is used for verification of services. In addition, we notify members of important dates and new services by e-mail.

The American Economic Association

Symposia

Labor Markets and Unemployment

Mary C. Daly, Bart Hobijn, Ayşegül Şahin, and Robert G. Valletta,
"A Search and Matching Approach to Labor Markets:
Did the Natural Rate of Unemployment Rise?"
Hilary Hoynes, Douglas L. Miller, and Jessamyn Schaller,
"Who Suffers During Recessions?"

Government Debt

Philip R. Lane, "The European Sovereign Debt Crisis"
Carmen M. Reinhart, Vincent R. Reinhart, and Kenneth S. Rogoff,
"Public Debt Overhangs: Advanced-Economy Episodes Since 1800"

Articles

Justin M. Rao and David H. Reiley, "The Economics of Spam"
Bruce D. Meyer and James X. Sullivan, "Identifying the Disadvantaged: Official
Poverty, Consumption Poverty, and the New Supplemental Poverty Measure"
Karen N. Eggleston and Victor R. Fuchs, "The New Demographic Transition:
Most Gains in Life Expectancy Now Realized Late in Life"
Gary Charness and Matthias Sutter,
"Groups Make Better Self-Interested Decisions"
Kazuo Ueda, "Deleveraging and Monetary Policy:
Japan Since the 1990s and the United States Since 2007"
Peter Thompson, "The Relationship between Unit Cost and Cumulative
Quantity and the Evidence for Organizational Learning-by-Doing"

Recommendations for Further Reading • Notes

AMERICAN
ECONOMIC
ASSOCIATION