

Essays from our Annual Reports, Part II

John A. Weinberg

This issue represents a second collection of essays originally published in the Richmond Fed's *Annual Report* since 2007. The first issue collected essays relating to the financial crisis and its aftermath. But this was not all that was going on in the last eight years. In other annual reports we have tried to address broad issues affecting the economic well-being of U.S. households. The topics addressed in these essays largely deal with longer-run trends in the economy—phenomena that play out over longer intervals than the typical business cycle. As such, these issues shape the backdrop to the Federal Reserve's monetary policymaking.

The first of these was our 2007 essay on household savings, with a particular focus on life-cycle aspects of savings—saving for retirement. One frequently hears concerns that many households approach retirement age with savings inadequate to support a continuation of their pre-retirement standard of living. These concerns are often motivated by the shift away over time from defined benefit pensions in the work place and by a declining personal savings rate in the U.S. aggregate data. Looking carefully at the breadth of available empirical and theoretical research, our 2007 essay found a more nuanced picture. Looking at household behavior through the lens of the life-cycle theory of saving and consumption suggested that, while there was variation, most households were approaching retirement reasonably well-prepared. Two caveats to this finding are important to note. First, the essay looked entirely at evidence from before the deep recession that began in late 2007 and severely disrupted household finances. Second, for a large portion of the population, government transfers from the Medicare and Social Security programs made up—and continue to make up—a significant portion of their retirement resources. And our aging population means that these transfers will become an increasing source of fiscal strain for the federal government.

Budgetary issues more generally have been a topic of broad concern in the United States in recent years. The recession of 2007–09 brought with it a large increase in federal deficits, and the polarized political environment has led to near shut downs of the government on a number of occasions, as the Treasury has run up against its legislated debt ceiling. In 2011, a year in which fiscal policy and government borrowing was the subject of intense political debate (and gamesmanship), we examined the notion of sustainability in government financial plans. In our essay for that year, we explained that economic forces would put an upper limit on how much debt a government could issue, although identifying that limit, or how close we are to it at any point in time, is difficult. But financial market investors simply wouldn't be willing to hold—not without sharply higher yields—a level of debt so high that future taxes would not be capable of paying it off. As a government neared such a level, the public would come to expect some adjustment to planned spending and taxes. But uncertainty about the shape that such adjustments might take could itself have the effect of dampening investment spending and thereby weaken growth. Further, fiscal policy stretching the limits of sustainability raises the risks that monetary policy will lose its ability to control the price level. Hence, the possible consequences of continued political gridlock for the long-term performance of the U.S. economy were (and are) significant.

One of the defining characteristics of the recession of 2007–09 was the dramatic increase in the share of the unemployed who were unemployed longer than 26 weeks. This episode was the worst economic contraction in the United States since the Great Depression, with the unemployment rate topping out at 10 percent. This peak was a little below the peak in the 1982 recession. In that earlier period, however, the long-term unemployed never accounted for more than about a quarter of all unemployed. By contrast, in the recent recession the share of the long-term unemployed soared to more than 40 percent of overall unemployment. Our 2010 *Annual Report* essay examined these facts, together with the related phenomenon of “duration dependence”—the tendency for unemployed with longer unemployment spells to have lower probabilities of finding a job. The essay argues that the distinct patterns in the data for the Great Recession suggest that an unusually large number of workers with relatively low employability entered unemployment in the contraction. This finding is consistent with a major source of the economic dislocation of the period coming from structural changes in the economy that reduced the relative demand for lower-skilled workers. While this is a topic that has continued to attract considerable research attention, our 2010 essay's early look at

the period suggested that much of the labor market damage in the Great Recession was beyond the ability of monetary policy to correct.

Perhaps related to structural changes that altered the relative demands for workers of varying skill levels, the period since the recession has seen varying rates of improvement for workers and households across the income distribution. Indeed, inequality has increasingly become a topic of economic and political commentary. Our 2012 *Annual Report* essay touched on this topic, but with a focus more on the dynamics of household economic standing, as opposed to the distribution at a point in time. In particular, the essay surveys the evidence on intergenerational mobility—the frequency with which children born into the lower parts of the income distribution are able to climb the economic ladder in their lifetimes. The fact that there is a fair amount of persistence across generations is taken as evidence that people likely face different levels of challenges in developing the human capital necessary for advancing one’s standard of living. And these disparities in access to human capital investment likely begin very early in life—pointing to a targeted approach to public early childhood education investments as an avenue for improving the economic mobility of those at the bottom of the distribution.

The broad trends discussed in these essays, as well as the public policy implications, fall mostly well outside the scope of monetary policy. But by identifying forces shaping the real economy, they help us understand the context in which the Federal Reserve pursues its mandates for price stability and sustainable employment growth.

Are We Saving Enough? Households and Retirement

Doug Campbell and John A. Weinberg

On October 15 of last year, a retired school teacher from Earleville, Maryland, sat down at a computer terminal and typed responses to four “yes” or “no” questions, beginning with, “Are you at least 61 years and 9 months old?” In answering affirmatively, Kathleen Casey-Kirschling made history. Born one second after midnight New Year’s Day, 1946, she was the first of the 78-million-member baby boom generation to apply for Social Security benefits. She became eligible to collect with the turn of 2008.

The reason Casey-Kirschling’s otherwise everyday act made news is no mystery. In part because there are so many baby boomers relative to the overall population, Social Security payments to retirees are projected to exceed payroll tax revenues in less than 10 years. By 2041, benefits will have to decline, or taxes or government borrowing will have to increase. In the case of Medicare, the health care insurance system for the American elderly, similar changes are expected to be necessary as early as 2019.

The baby boom generation’s retirement brings into focus perhaps the most significant demographic shift in United States history. Baby boomers, the moniker for the generation born between 1946 and 1964, comprise about 26 percent of the overall U.S. population. Their sheer numbers assure that future growth in the labor force will slow by comparison to recent decades. The birth rate seems unlikely to ever spike up to that experienced in the 1950s, and life expectancy continues to increase.

In 1940, people who had already reached the age of 65 were expected to live to be 77.7 years. By 2030, life expectancy for 65-year-olds

■ The views expressed are the authors’ and not necessarily those of the Federal Reserve System.

is projected to reach 83.7 years. At the same time, birth rates are falling: In 1955—the core of the baby boom—the average woman had 3.5 children in her lifetime; by 2005, the birth rate had leveled off to about 2 children per woman, a trend that is projected to hold steady for the next 25 years.

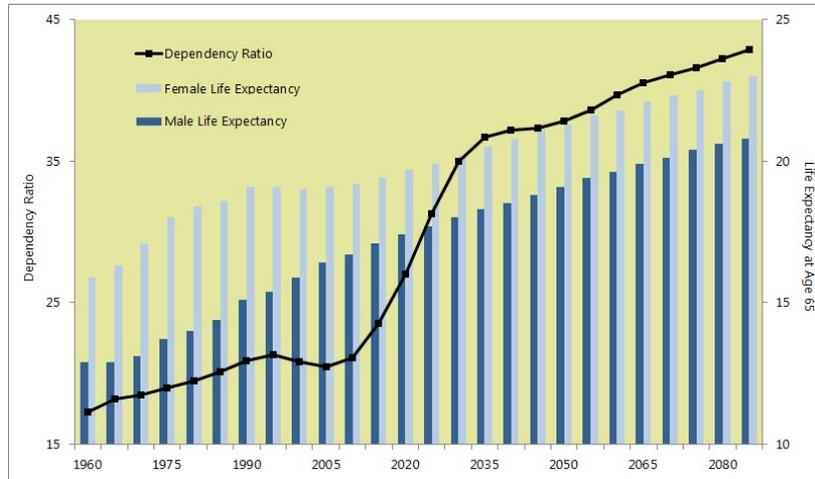
These trends signify long-term ramifications for the economic well-being of American households. First, the big picture: Population aging presents a problem of consumption maintenance. If a growing number of older people move into retirement, then there are fewer people working as a share of the population, increasing the so-called dependency ratio shown in the first figure in this article. So on a per-person basis, there would be relatively fewer goods and services being produced. The upshot is that people could have less to consume than in the absence of population aging. This statement requires the economist's usual "other-things-being-equal" qualification, which means that other factors in the economy affect economic output per person. Most importantly, productivity growth resulting from technical change or improved work force skills increases output per worker. But regardless of the status of such other factors, an aging population probably means lower average consumption-per-person than would otherwise be possible.

Second, beyond the sustainability of national consumption, population aging threatens the sustainability of the nation's entitlement programs. Social Security and Medicare are pay-as-you-go programs, meaning younger generations of workers finance the retirements of older generations. For Social Security, the present value of benefits promised to older cohorts is \$13.6 trillion greater than the present value of scheduled tax contributions to the system, according to the Treasury Department. As large numbers of baby boomers retire, they will drain those promised benefits to the point where incoming tax revenue will no longer be sufficient to keep the programs solvent. A similar issue exists at the state and local levels, where many public-employee pension and retiree health care plans are less than fully funded. The growth of public-sector obligations to retiring boomers could strain government budgets at many levels.

Strains on entitlements are only one type of challenge facing today's workers. Employees also are adjusting to the predominance of a relatively new form of retirement saving—defined-contribution pensions, most commonly in the form of employer-sponsored 401(k) plans, in which workers are the main suppliers to their retirement plans. Under such a plan, a household could potentially outlive its savings, unless it was effectively annuitized.

Fading away are defined-benefit pensions, which provide guaranteed income streams for retirees that they can't outlive. At precisely the

Figure 1 Longer Lives, Greater Dependency



Source: *Annual Report of the Social Security and Medicare Boards of Trustees for 2007*.

Notes: *Dependency Ratio per 100 workers; Population aged 65 and over, divided by population aged 20–64; **Life expectancy at age 65.

time at which public transfer backstops that mitigate the problem of longer lifespans are in trouble, responsibility for saving is being placed upon the shoulders of individuals, as is the investment risk.

Despite the heightened importance of individual preparedness for retirement, a puzzling observation is that Americans seem to be saving less than ever. The personal saving rate, for example, has been declining. And there is no shortage of anecdotes about overboard consumer spending and people entering their 60s with no nest eggs.

All of this adds up to quite a laundry list of concerns. To review, we have an aging population, which means problems in maintaining national consumption as well as maintaining entitlement programs like Social Security and Medicare. Then we have the tricky transition from guaranteed defined-benefit pensions to employee-driven, defined-contribution 401(k) plans. And finally we have economic statistics that appear to show that Americans are saving at historically low rates. We are left with a big question: Are U.S. households going to be financially prepared for retirement?

In this essay, we initially aim to clear up some misconceptions about Americans’ saving habits. We look at the data on demographics,

pensions, and wealth, seeking to identify which trends merit concern and action, and which may not.

Our emphasis is on households. Why households? In the popular media, the cited statistics are almost always aggregate—they consider the state of things across the board rather than by household. The popular press reports endless stories about perilously low saving rates; the implication is that “the economy” is in trouble. But our interest isn’t in the aggregate economy but in the economic well-being of individual households—people, couples, and families. In fact, when you look at the data on individual households—that is, disaggregated data—a surprisingly different picture emerges. Most households near retirement are saving adequately. Crucially, insofar as future policies are concerned, their saving is as modern economic theory predicts: They are mostly doing the best they can given their incomes.

Then we consider the future. The finding that households are now saving optimally assumes that the government will deliver on promised Social Security and Medicare benefits. But the demographic shift will stress the federal budget, imperiling those benefits. In addition, we face the related problem that the demographic shift may reduce the size of the overall pie that households can consume (relative to a world in which no demographic change occurs). It might seem wise to find a way to spread these burdens across generations so that future generations don’t take the biggest hit. Those ways might include saving more, taxing more, or borrowing more. We will explore the effects of these different approaches, with particular attention to their unintended effects. Understanding the economic tradeoffs inherent in each of these strategies may help us choose well.

1. MEASURING SAVINGS

It goes without saying that saving is important. Taking income from present consumption and moving it to savings allows us to finance spending on both physical and human capital to increase the future standard of living. The growth of future living standards depends on how much income is set aside for savings, as well as growth in productivity.

Concern about Americans’ readiness for retirement generally can be traced to a single source—the personal saving rate. The most widely cited measure of personal saving comes from the U.S. government’s National Income and Products Accounts (NIPA). Boiled down, the NIPA measure is disposable—or after-tax—income minus spending.

This measure held mostly steady between 7 percent and 10 percent of disposable income from the 1950s through the early 1980s. It then

Figure 2 Personal Saving Rate

Source: U.S. Department of Commerce.

began to fall, going south of 7 percent in 1990, to 4 percent in 1996, and 2.3 percent in 2001. In 2005, it went into negative territory. In 2006, Americans saved an average of 0.4 percent of their disposable income, and the saving rate has hovered around zero since then. It is impossible to ignore the sharp downward movement that this rate has displayed over the past two decades, and it has fallen more sharply than in most other developed countries.

Why have savings trended so far south? Many assume the main problem is self-control, or lack thereof. People may spend to satisfy immediate needs or cravings, ignoring reality or hoping against all evidence that the future will bring more wealth. A related story is that credit has become easy to obtain, leading households to take on more debt—or at least saving less because they know they can borrow in an emergency.

The components of the NIPA saving rate are worth a closer look. C. Alan Garner, an economist with the Federal Reserve Bank of Kansas City, points out several potential shortcomings. The NIPA rate computes how much household income is put aside for other uses, such as investments in homes or businesses. But it excludes capital gains and losses on existing assets. Therefore, it doesn't include potential changes in wealth from assets ranging from stocks to home equity.

The 1990s and early 2000s saw significant increases in both stocks and housing values. Perhaps households, feeling wealthier, were

motivated to spend more. Indeed, some economists believe that there is a “wealth effect” on consumption; when household wealth rises or falls, consumption will go in the same direction.

Measured savings is a consequence of households’ consumption decisions and shows the difference between measured income and the resulting consumption. Consumption can grow with no corresponding increase in measured income, which drives the saving rate lower but this could be because actual income increased more rapidly than measured income. Meanwhile, those examining the NIPA rate don’t have the same perspective as consumers, whose confidence in their future earnings or wealth isn’t directly observed. To observers, it may look like some households are saving too little; for some of those households, it may just be a case of spending now in anticipation of higher income later.

For data on household wealth, the Federal Reserve’s Flow of Funds Accounts provides some aggregate figures. Overall, wealth has gone up almost every year (it dropped in the 2001 recession), though the growth has slowed in recent decades. It may seem surprising that the saving rate has gone down while net household wealth has gone up. But the two are not historically connected, as wealth changes are a product mostly of changes in stock and real estate asset prices, which are not taken into account by the standard measures of saving. By itself, the NIPA rate doesn’t tell us whether Americans are likely to reach retirement with sufficient wealth.

As with all national economic indicators, later revisions can change initial results. Historically, the NIPA saving rate has mostly been revised upward, and sometimes by large amounts. Leonard Nakamura and Tom Stark, economists with the Federal Reserve Bank of Philadelphia, find that initial estimates of personal savings from 1965 to 1999 on average were revised upward by 2.8 percentage points. For the fourth quarter of 1981, for example, the revision was up 7.3 percentage points. Nakamura and Stark attribute the differences to new methodologies that take into account new sources of household income. New data from Census revisions also may play a role in adjusting estimated business sales, which in turn affect personal consumption expenditures captured in NIPA.

Finally, the saving rate is an aggregate measure. It gives no sense of savings across the population’s distribution. How much are low-income households saving compared with high-income households? The NIPA saving rate, as generally cited, does not address this question.

2. A CLOSER LOOK AT WEALTH

Many studies have looked at more robust measures of household wealth. Alicia Munnell and Mauricio Soto, economists at Boston College, analyzed the Health and Retirement Study (HRS), which provides panel data from an initial 1992 sample of 7,600 households aged 51 to 61. It provides a close-up look at where household savings are located at the cusp of retirement.

Financial planners often rely on replacement rates to gauge whether their clients are saving as much as they should. A replacement rate assesses the amount of spending a retired household's savings can sustain relative to its pre-retirement income. A typical rule of thumb is that a retired household should plan to spend between 75 percent and 85 percent of annual income before retirement, because even though expenditures on things like health care might increase, living expenses generally are lower for old people. Munnell and Soto calculate average income replacement rates for households of adult couples with pensions at 79 percent and those without pensions at 62 percent.

Clearly, households with replacement rates of 62 percent can expect to experience declines in their living standards upon retirement. On the other hand, these couples who lack pensions make up just 25 percent of the sampled population.

Sizing up these figures, Munnell and Soto conclude that: "The majority of households retiring today are in pretty good shape. Regardless of how retirement income and pre-retirement income are defined, households with pensions appear to meet the threshold of adequacy." Importantly, Munnell and Soto found that for the mean of the middle 20 percent of soon-to- retire U.S. households, expected payments from Social Security represent an average of 48 percent of their wealth. Their prediction about the adequacy of household wealth assumes that entitlement programs like Social Security will remain solvent.

Economists at Williams College and the Federal Reserve Board of Governors take the next step by analyzing the HRS data for insights into the distribution of savings across the population. David Love, Paul Smith, and Lucy McNair develop a new measure they term "comprehensive wealth," asking whether U.S. households are "adequately" saving for retirement. The authors take one of the first looks at the 2004 wave of the HRS, which captures the "early baby boomers" born between 1948 and 1953. They begin with financial net worth, which they define as the sum of stocks, checking accounts, and CDs, minus non-vehicle and non-housing debts. Also added are balances from defined-contribution pension plans, typically 401(k)s, and IRA balances. Moreover they added present values of defined-benefit pensions,

Social Security, and welfare, plus expected future labor income. And they added employer matches to defined-contribution plans.

Their findings show that, “overall, households hold comprehensive wealth that is several multiples” of the wealth level necessary to sustain consumption at the official poverty line. The median ratio of wealth to the present value of future poverty lines is 3.56; the median annuity value of wealth is \$32,000. (These are, admittedly, not large nest eggs, but since old people consume less than young people, they may well be sufficient. A retirement annuity of \$32,000 per person represents a 75 percent replacement rate for a worker earning \$42,700 a year.) Still, about 12 percent of households lack enough comprehensive wealth to bring them over the poverty line, and 9 percent (with ratios between 1.0 and 1.5) are “near” the line. “Not surprisingly,” they write, “there is a close correlation between lifetime earnings and the share of households below or near the poverty line.”

Put another way, the working poor often don’t have enough savings when older to lift them out of poverty in retirement. Poor households in their working years remain poor in their retirement. “Overall, our findings show a generally optimistic view of retirement savings adequacy among current older cohorts, though with a notable pocket of inadequacy concentrated among those with the lowest lifetime earnings.” Like Munnell and Soto, these authors find that expected Social Security payments represent a large share of retirement wealth for those at or below the middle of the lifetime earnings distribution.

3. A THEORY OF SAVING

The main reason that looking at aggregate statistics on saving can be misleading is founded on two 50-year-old economic theories. In his 1957 book, *A Theory of the Consumption Function*, Milton Friedman found that current income matters less in consumption than “permanent” income, by which he meant a long-run average of anticipated income. People tended to smooth their consumption throughout their lifetimes based on how much they expected to earn.

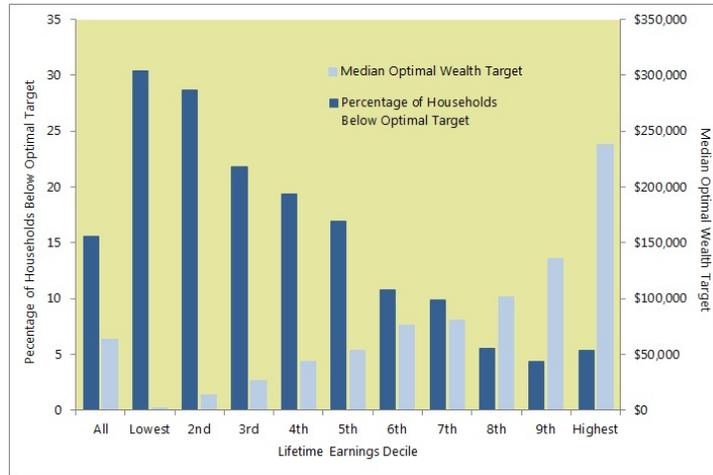
Also in 1957, Albert Ando and Franco Modigliani tested the prediction that people’s natural inclination is to smooth their consumption over their lifetimes. When younger and earning less income, people may borrow more and save less. During middle age, when labor income is typically at its peak, people will ratchet up their saving. In retirement, as income diminishes, people spend off their savings. Overall, households estimate the stream of resources over their lifetimes and use that as their benchmark in deciding how much to spend at any given period.

It turns out that these theories, known as the permanent income and life-cycle hypotheses, have matched up with the data fairly well over time. One of the more recent studies on this front comes from John Karl Scholz, Ananth Seshadri, and Surachai Khitatrakun. What was most unique about their study was that it gained access to previously unavailable Social Security earnings data, providing more precise measures of actual earnings and lifetime income than previously available. They developed optimal decision rules for consumption for each household in the sample, with rules that differed depending on household characteristics, and then plotted the distribution of optimal net worth across households in the HRS.

It should be noted here that the “optimality” of saving as examined by Scholz, Seshadri, and Khitatrakun is related to a specific theory of household behavior. While this theory is the standard approach of economists studying saving and consumption, it necessarily abstracts from many forces that might affect behavior. Still, this idea of optimality is a useful notion that builds on the idea of “adequacy” by taking into account the most important economic factors affecting household choices. Their two most important findings:

- More than 80 percent of households in the observed HRS sample have accumulated wealth above the targets implied by the model, while 15.6 percent of surveyed households with a member nearing retirement age fell short of wealth targets. But the authors note that most of the people who are undersaving aren’t undersaving by much.
- At the same time, they find that “undersavers are concentrated in the bottom half of the lifetime earnings distributions.” In the lowest earnings decile (basically, people whose incomes are at or below the poverty level), 30.4 percent of households are below the optimal target; in the highest decile, 5.4 percent are below. (The authors caution, however, that this result may be more strongly related to whether a person is in a single or married household.)

What’s important about the Scholz, Seshadri, and Khitatrakun model is that it confirms the theoretical notion that households tend to save the amount necessary to provide the maximum level of smoothed consumption over their expected lifetimes. The model takes into account that each household experiences different fluctuations in earnings and life expectancy. Though it may sound odd, viewed through this lens, seemingly paltry levels of wealth may actually be quite consistent with reasonably effective saving behavior, given a household’s income experience. At the aggregate level, it is impossible to identify this household-level activity.

Figure 3 Optimal Savings

Source: Scholz, Seshadri, and Khitatrakun.

According to these economists and their high-quality data, most people are doing precisely what economic theory says they should be doing. Most people are doing the best they can given their situations. (In fact, one of the authors' main findings is that many people seem to be oversaving.) Most households save enough to generate the highest level of smoothed consumption over their expected lifetimes. As with Love, Smith, and McNair, these authors find that undersavers are also the poorest, suggesting once again that America faces less of a retirement savings problem than a poverty problem.

A downside to the optimality approach, some economists counter, is that what's "optimal" may still make a household "wealth poor" at retirement. It might be the case that for one household, whose wage earners lose jobs or get sick, entering retirement with only Social Security as a backstop is "optimal," as it provides the smoothest possible consumption over their lifetime. But some may consider relying on Social Security alone—with average monthly payments around \$1,000 a month at present—as simply inadequate. Of course, optimality should not be confused with desirability. Morally or ethically, we might be predisposed to wanting people to have more resources at retirement. Additionally, models that generate optimal consumption paths rely on assumptions that may not be correct, including mortality and risk preferences.

Having acknowledged these challenges, we can still agree that life-cycle theory seems to be generally squaring with the facts. Given the resources that people acquire throughout their lifetimes, most are arranging for their nonworking years in retirement as best as they can. Addressing poverty—where evidence of undersaving is greatest—is in many ways a different problem.

4. DEMOGRAPHIC CHANGE

The judgment that most Americans are saving reasonably well does not mean we should be sanguine about the future. As the disaggregated data show, Social Security accounts for a significant portion of expected retirement income for many households. But the aging of the U.S. population will put strains on the ability of the government to make good on its promised Social Security payments. On top of that, the demographic shift could mean lower economic output and consumption than in the absence of population aging.

We now face choices about how to prepare for these changes and to make good on our promises to workers. Will we have to raise taxes on current or future workers? Many analysts have postulated that higher household saving rates are desirable because they could help ease the burden of higher taxes or lower spending that might otherwise be passed to future generations. To properly evaluate the choices, let's first consider the size of the shift and what it might seem to imply about future consumption possibilities.

When we talk about population aging, it is important to take into account both the larger share of old people and the smaller share of children, because they can have opposite effects on overall consumption levels (with old people consuming more because of their medical needs, and children, less). The declining birth rate means a lowered dependency burden, which ordinarily would be a good thing with regards to per capita consumption. But in this case, it is swamped by the growing number of old people per worker.

At face value, what these trends mean is that younger generations of workers will support larger numbers of old people. Equally, it means there could be fewer goods and services to go around compared with a world in which there is no demographic change. This result is because, in general, consumption per person depends on output per person. So while productivity growth raises output per person, a growing share of retirees in the population holds down those gains on a per-person basis.

To get a clearer understanding of the implications of population aging on consumption, consider the ratio of working-age people (ages 20 to 64) to elderly people (older than 65). Currently, there are five

working-age adults for every person aged 65 and above. By 2030, there will be three working-age adults for every elderly person. Overall, annual growth in the size of the labor force is expected to slow from 1 percent at present to 0.2 percent after 2020. (Obviously, these figures could change if, for example, more people stay in the labor force past the usual retirement age of 65. Immigration of young workers could also pick up some of the slack.)

Louise Sheiner, Daniel Sichel, and Lawrence Slifman, economists with the Federal Reserve Board of Governors, argue that the best gauge of the macroeconomic effects of population aging is what they call a “weighted support ratio.” This takes into account both the heightened consumption needs of the elderly (primarily because of their greater demand for health care) and the lower needs of children. Their weighted support ratio is peaking now at about 0.64 (workers to old people and children, with these populations’ consumption needs weighted) as most baby boomers remain in the work force. But it is projected to drop sharply over the next decade, to 0.60 in 2020, then to 0.56 in 2040. That seemingly small decline actually represents major changes in growth of the U.S. labor force; it means the number of workers to dependent population will be much lower than we’ve recently experienced, as well as lower than the previous low point in the early 1960s. The weighted support ratio falls farther than the simple support ratio, implying a larger impact on the economy.

Now, it is a bit more complicated than that. A society’s potential level of consumption depends, among other things, on capital per worker, technical advancement, and the return to capital. Given current trends, Sheiner, Sichel, and Slifman conclude that we will experience a significant reduction in per capita consumption relative to a baseline in which there is no demographic change. (These trends include assumptions about labor force participation among the elderly and levels of immigration.) This is because the population bulge has made our production bulge as well. We have, in short, experienced a period of low dependency during which per capita output was high. With fertility low relative to that of the baby boom generation, we received a temporary benefit in the form of greater consumption available per person.

5. WHAT NOW?

The data presented earlier on household-level wealth holdings suggest that older baby boomers are reasonably well-prepared for retirement. On the other hand, dependency ratio calculations like those presented in the previous section imply a real economic cost of the demographic

bulge that will weigh on the consumption opportunities of future retirees, future workers, or both. How do we square these two facts? A key assumption in the calculations of household wealth is that future Social Security payments will be made according to current policy. This assumption is important, since for many low- and moderate-income households, expected Social Security payments represent a large fraction of retirement wealth. But as we discuss elsewhere, current Social Security payment policy, together with current taxation policy, creates large fiscal deficits. These will ultimately require changes either in payments or in taxes (or both) and will ultimately affect some people's consumption patterns.

People's responses to the aggregate economic changes brought on by the demographic trends will depend on the prices households face in making consumption decisions and the returns households receive on their labor time and savings. By prices, we mostly refer to wages and interest rates. Does population aging somehow affect prices in such a way that individual households are hindered in their ability to prepare for retirement? We now explain both how population aging could affect prices and then how it doesn't have to.

Intuitively, the most obvious and simple plan might seem to be to save our way out of demographic change—to put more money aside now while we've got more people working. This would require people to consume less, of course, but it would also help lower the burden on future generations. With extra savings, we could add to the capital stock and thus make future workers more productive.

If only it were as simple as that. The effect of increasing the capital stock may actually discourage saving. Federal Reserve Board economists Douglas W. Elmendorf (now with the Brookings Institution) and Sheiner assume that current consumption and saving rates are close to optimal (an assumption supported by other research cited in this essay) to isolate the impact of population aging. They point out that forcing greater saving on current workers is not an obviously beneficial approach to the looming demographic trends.

Here is why: Recall that the U.S. workforce is growing more slowly now with the aging of the baby boomers. With fewer workers, we require less in the way of investment to provide new workers with capital. So if we are trying to save our way out of uneven consumption, we increase the future capital-to-labor ratio (because we have less labor and more capital than before). This means returns on capital are smaller than before, and investment payoffs are lower.

This is not to argue that we should simply kick the burden of demographic change and supporting entitlement programs to future

generations. Rather, it is to explain the possible complications of that approach. In fact, it's fair to say that all approaches are imperfect.

A study by economists Laurence Kotlikoff, Kent Smetters, and Jan Walliser considered how the combination of demographic change and the burden of Social Security might play out. They conclude that payroll taxes would have to jump by 77 percent, and that this increased tax burden would swamp the extra capital to workers that ordinarily would accompany an aging society. Alternatively, there is the research of Nobel Prize-winning economist Edward C. Prescott and Arizona State University's Kathryn Birkeland: They argue that addressing the solvency of entitlement programs while maintaining the overall welfare of the U.S. population is as simple as having the government issue more debt. Prescott and Birkeland's point is that in the existing tax-and-transfer system, households may pare back their labor in the face of high taxes. Despite the risks, issuing more government debt along with a mandatory worker "saving-for-retirement system" would mean that workers' productive time is rewarded with a larger savings nest egg. This results in a larger capital stock awaiting future generations.

By no means is this an endorsement of any of these approaches. Our aim is to briefly point out what sort of consequences we can expect with each one. You can ask households to save more, but doing so would tend to lower everybody's rates of return. While there are many other ways that economists approach the retirement/entitlement problem amid demographic change, the most useful are those that model households as rational, forward-looking units that respond to incentives. If households face a pricing environment where saving makes sense, they will do so.

While understanding the tradeoffs involved with preparing for demographic change is important, it is also important to take action as soon as possible. In their study, Sheiner, Sichel, and Slifman conclude that if we made no changes to our saving habits, future generations would see their per capita consumption fall 14 percent compared with what it would have been without demographic change. By contrast, if we alter saving rates now as a means to spread the burden equally across generations, the relative decline in per capita consumption is reduced to just 4 percent. While there is always uncertainty around such projections, the desirability of a timely response to demographic change is clear enough.

6. CHANGES AHEAD

As the first retiring baby boomer, Kathleen Casey-Kirschling became a symbol for America's demographic transition. Her arrival on Social

Security's doorstep made long lingering questions more urgent: Do retirees have enough savings? Will her cohorts bankrupt our entitlement programs? Will the sheer size of her generation cause living standards to decline in the future?

We have shown that, contrary to popular opinion, most Americans near retirement are saving largely as economic theory predicts they should. Most of the nation's undersavers are also the poorest. While lack of savings isn't exclusively a problem of the poor, that's where the problem is largest. Our chief concern should be for those who are poor even before retirement.

The aging of the U.S. population is not a surprise. It is a predictable event that we can plan for. Research on household saving behavior shows that most households plan reasonably well. But the important caveat in this conclusion is that household planning appears to be predicated on the assumption that Social Security and other retirement benefits will be paid according to current policy. The fiscal stresses that these policies face imply difficult choices. Increasing taxes to prop up entitlement programs would create additional problems. Should the government take on more debt? Some economists believe that approach is not as unwise as it first sounds—it could ease the burden on current workers while allowing interest rates to remain high enough to encourage household saving.

As we have described, an older society portends a time when the growth of consumption per person might be held down, and saving might become harder. The somewhat natural lengthening of time that older workers stay in the labor force may cushion the demographic blow, as might increased immigration. But in general, whichever approach we take, our focus should be on making sure that households both today and tomorrow are not impaired in their ability to save. If there is no consensus about what to do next, there is agreement that to delay action will exacerbate the problem for future generations. The earlier we embark on this effort, the more likely we are to achieve a desirable outcome.

APPENDIX

The following sections appeared as sidebars in the original text.

1. WHY AREN'T (SOME) PEOPLE SAVING MORE?

Though careful studies show that many people are saving enough, it's also clear that some people aren't. Why not?

The first possible explanation is that figuring out how much to save is complicated. Economic theory holds that people seek to smooth consumption over their lifetimes. But in the 21st century, this is not such an easy calculation. Annamaria Lusardi and Olivia Mitchell explain the difficulty this way: "The consumer must understand present discounted values, the difference between nominal and real amounts, and be able to project expected future labor income, pensions and Social Security benefits, retirement ages, and survival probabilities, among many other factors. These requirements are inherently complex and demanding."

Within the "it's complicated" explanation fall several subcategories. Some people, for example, may use simple rules of thumb in planning for retirement. These rules might include aiming for certain replacement rates of income upon reaching retirement. But because of the complex nature of investment decisions, this sort of planning may still fall short, and retirees may have to shift down their consumption to adjust. (Some economists dispute the notion that consumers have to understand every detail to properly save for retirement. These economists argue that most people make estimates that turn out to be accurate.)

A related explanation is that people may believe themselves to be financially sophisticated when they really aren't. However, some recent research by Lusardi and Mitchell discounts this notion, finding that most people who classify themselves as financially literate indeed score well on related testing.

A branch of economics is interested in the idea that undersaving reflects a lack of self-control. Some surveys have shown that households themselves cite "lack of willpower" for their low savings, while others admit to procrastination. Behavioral economists use these examples in support of their theories of why people deviate from standard economic rules.

Though much of this kind of work is open to question, recent behavioral research on participation in 401(k) plans is striking. The research has shown that if employers make "opt in" the default choice for such plans, more people automatically end up saving than if "opt out" is the default. This evidence runs contrary to traditional theory, which holds that people ought to be making the same decision whether it is the default or not.

Many studies point to a graver problem than the misperception that most people aren't saving enough—it's that undersaving is most widespread among the poor. A possible explanation is that because

they have less to gain, poor people invest less in financial planning that would help them save more. They may also face disincentives to saving because of financial backstops like Social Security and welfare transfers. Like a lot of research on savings, this finding points to the need for raising wealth for those with low incomes as much as for increasing their savings.

2. THE EFFECTIVENESS OF FINANCIAL EDUCATION PROGRAMS

Of all the ways to encourage higher saving rates, perhaps none is more popular than financial education. If only Americans were made aware of the importance of retirement planning—and given some pointers on how to get started—then changes in savings behavior would surely follow.

That’s the conventional wisdom, at least. But despite the seemingly obvious link between knowledge and behavior, economists have struggled to measure the degree to which financial literacy efforts actually work. It is well documented that some people have a poor grasp of basic economic concepts, and that shortfalls of knowledge are particularly evident about Social Security and pensions. But the connection between the effect of being exposed to financial education and subsequent improvements in saving habits is tenuous.

The trick is distinguishing between causation and correlation. There are definite correlations between wealth and retirement planning. Among baby boomers who reported that they undertook even “a little” retirement planning, wealth holdings were twice as large as non-planners, according to economists Annamaria Lusardi and Olivia Mitchell. Meanwhile, many studies have documented that households that do little financial planning tend to be the less educated and minorities. But does that mean that planning can lift these households into more secure retirements?

Lusardi and Mitchell, who are two of the world’s leading researchers on the topic, created a “financial literacy index” based on a survey of Americans in their prime working years, with most respondents between 40 and 60, as well as the Health and Retirement Study. With the index, the economists identify which traits and concepts are predictive of retirement planning. In general, they conclude that “financial literacy is a key determinant of retirement planning” and that literacy is highest among those exposed to economics in school and to those who attended company-sponsored programs.

This supports some of their earlier research, which considered the possibility that wealthier households planned more because they had

more to gain. They couldn't find any effect of wealth on planning, however, and concluded that planning is more likely to cause wealth, rather than vice-versa.

"Saving for retirement is becoming a more and more challenging and a more important objective requiring ever-greater levels of financial sophistication," Lusardi and Mitchell wrote. "Clearly it is urgent to target effective programs to those who can put this necessary financial knowledge to work."

As it happens, the most effective programs do not come cheap. In a survey of the literature on financial education, Richmond Fed economist Matthew Martin concludes that there are returns from such programs, especially to low-income and lesser-educated households. However, Martin finds that one-size-fits-all efforts may not succeed: "Financial education programs are most effective when they are tailored to the needs of the recipient and include face-to-face time, either with a counselor or in a classroom setting." As a result, the most effective programs also tend to be the most costly.

3. DOES THE DECLINE OF DEFINED-BENEFIT PENSIONS SIGNAL TROUBLE FOR AMERICANS' RETIREMENT YEARS?

Retirement, as we know it today, is a relatively new concept. Back in 1880, eight in 10 men aged 65 and up still worked. When they stopped, it usually was because they were physically unable to carry on. They relied on family for financial support until their deaths. Self-financed retirement was a luxury affordable mainly to the rich.

Over time, workers came to rely on employer-sponsored pensions (plus payments from Social Security, which launched in 1937). The Pennsylvania Railroad Pension is touted as having kicked off the private pension era with its creation in 1900. Its "defined-benefit" formula generally has been followed ever since.

Defined-benefit pensions provide an annuity at retirement that workers can't outlive. Benefits are a function of years of service and highest salary. The assets of defined-benefit pensions are professionally managed and the employer bears most of the investment risk. Employers first started offering defined-benefit pensions in part to help with worker loyalty and to ward off strikes.

Today, defined-contribution plans, predominantly 401(k)s, have replaced defined-benefit plans as the leading form of employer-provided pension. This transition has raised concerns among some observers, in part because defined-contribution pensions place more of the burden of saving, not to mention the portfolio risk, on individuals. Participation

in such plans is voluntary, meaning some will opt out of them, even if it would not seem to be in their best interest to do so. And smaller firms don't yet en masse offer 401(k) plans, whose big appeal is the matching contributions that employers make.

Given current trends, what will household portfolios look like as they reach retirement? In one study, economists with Williams College and the Federal Reserve Board of Governors point out that though personal retirement accounts (with defined-contribution plans being the leading contributors) are small in size among people nearing retirement, this doesn't necessarily suggest that Americans have inadequate savings. Instead, it is mostly evidence that they are relatively new vehicles for savings.

In the 2004 Health and Retirement Study, less than a third of households aged 75 and older had personal retirement accounts, compared with about half of households aged 62 to 75 and 61 per cent of those between 51 and 61. Despite this transition to defined-contribution coverage, "we do not find evidence of a steep deterioration in retirement adequacy among the younger households in our sample."

Growth in defined-contribution plans is widely evident. In 1985, assets in private defined-benefit pensions almost doubled those in defined-contribution plans—\$814 billion to \$417 billion. In 2005, assets in defined-contribution plans were on top, \$3 trillion versus \$2.2 trillion. By 2040, 401(k) assets are projected to grow eightfold from their 2000 level.

By one study, the number of people covered by defined-benefit pensions over the past 20 years fell by about 30 percent, while the number covered only by a 401(k) plan grew 300 percent. The number of participants in defined-contribution plans grew from about 19 million in 1980 to more than 52 million in 2004.

Meanwhile, even though growth in 401(k) coverage has slowed in recent years, participation rates are expected to climb well into the future. Among those 60-year-olds in the 2nd earnings decile (i.e., people whose earnings put them between the lowest 10th percentile and 20th percentile of the total population), 401(k) participation in 2000 was 23 percent. By 2040, it's expected to increase to 53 percent. It would seem that even for the relatively poor, pension participation will rise. Overall, participation rates at age 60 are expected to be much higher, topping 80 percent, from the 70th earnings percentile on up.

Economists James Poterba, Steven Venti, and David Wise find that the average 65-year-old in 2040 will have more than \$450,000 in personal retirement accounts (in 2000 dollars). Of course, there is wide variance in accumulations. Those in the 2nd earnings decile are

Figure 4 Mean Projected 401(k) Assets for Cohorts Retiring in 2000, 2020, and 2040



Source: Poterba, Venti, and Wise.

expected to have about \$51,000 in mean projected 401(k) assets; those in the 9th decile (90th percentile and up) about \$1.1 million.

A lingering concern about 401(k) pensions is that so much of their assets are in equities, which tend to be volatile. According to one study, 61 percent of 401(k) assets in 2001 were in stocks. But this extra risk has been shown to be offset by the portability of such plans. Employees who take new jobs can take their 401(k) assets with them, but defined-benefit plans effectively penalize workers who leave.

In general, these projections point to future retirement security for most Americans, not the opposite. While the assets of low-income households remain low in retirement, many economists are optimistic that the transition away from defined-benefit pensions is one that ultimately will lead to more wealth for U.S. households: “The advent of personal account saving is projected to yield very large increases in the financial assets of future retirees across the lifetime earnings spectrum,” wrote Poterba, Venti, and Wise.

4. THE FED’S ROLE

The Federal Reserve’s role in the coming demographic transition is several-fold. First, the Fed can encourage households to make sound financial decisions, supporting financial education efforts that inform people about their choices and the importance of saving. In its

regulation of financial institutions, the Fed ensures that consumers receive adequate disclosures. These roles will be increasingly important as the United States begins its demographic shift.

Most importantly, the Fed abides by its two-part mission—to keep prices stable and promote maximum sustainable economic growth. People decide whether and how much to save based principally on their current and expected lifetime income and interest rates. By keeping inflation low, the Federal Reserve helps keep a stable economic environment. In fighting inflation, the Fed makes it easier for people to save.

5. NO EASY FIX FOR ENTITLEMENT PROGRAMS

The financial burden of paying for Social Security and Medicare is growing as the U.S. population ages. That’s hardly a revelatory statement, but it bears repeating as the first baby boomers enter retirement and begin to draw benefits from entitlement programs.

Both Social Security and Medicare are essentially “pay-as-you-go” programs, with retiree benefits funded by current payroll taxes levied on employers and employees. The 2007 Treasury Department report calculates that, thanks to population aging, the present value of Social Security’s scheduled benefits surpasses the present value of scheduled tax receipts by \$13.6 trillion—that’s the difference between the amount older cohorts put in to the program and the amount they plan to withdraw from it.

Meanwhile, Medicare expenses are expected to overtake income as soon as 2010, with trust fund reserves depleted by 2019. The present value of the unfunded liability for Medicare is close to \$70 trillion over an infinite horizon. Federal spending to support the two programs is expected to rise from 6 percent of GDP in 2005 to 20 percent in 2080. Another way to look at it is to focus on the program revenues and outlays as percentages of taxable payroll—income stays relatively flat into the future while expenditures continue to climb.

How do we close these unsustainable financing gaps facing Social Security and Medicare? There is no shortage of proposed reforms. Broadly, they fall into four categories:

- Keep more workers in the labor force, thereby reducing the growth in the number of retired Americans receiving benefits to workers paying taxes that fund those benefits.
- Raise taxes on workers.
- Reduce benefits.

- Allow greater numbers of young immigrants into the U.S. workforce.

Additionally, there are proposals to phase out the system in favor of private accounts, such as the program that President Bush promoted unsuccessfully in 2005. And there is a school of thought that argues that Social Security should be abolished because its existence has a number of undesirable effects, including that it discourages private savings that might otherwise supplement the program, and that it encourages early exits from the labor force.

So what should be done? One thing that most economists agree upon is that whatever reform is adopted, it will be easier to swallow—as well as more evenly spread across generations—if it is taken sooner rather than later. By government estimates, closing the 75-year unfunded liability of Social Security would require an immediate increase in the payroll tax of about 2 percentage points; waiting until 2041 would require approximately a 4-percentage-point boost. Medicare faces a similar scenario—it needs an immediate 3-percentage-point hike to fix the liability, or waiting until 2020 would require a gradual 10-percentage-point increase over the following 55 years. But this would still not make the systems permanently solvent; it would merely put them into balance for 75 years.

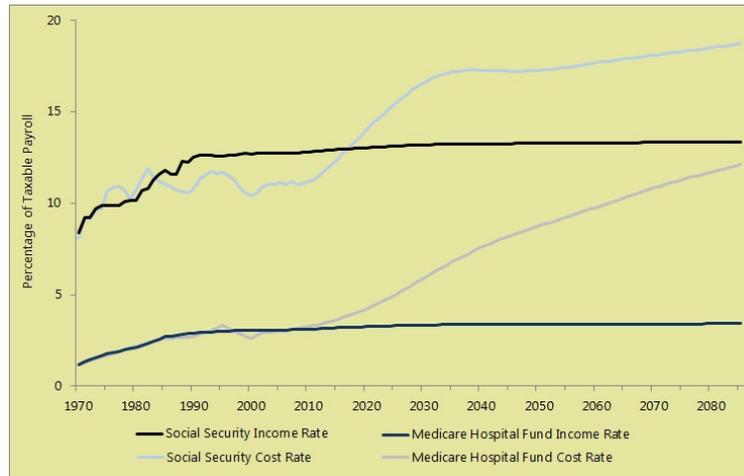
From a fairness perspective, this observation from the Treasury Department is worth considering: “Each time new legislation has ratcheted up taxes and real benefits, substantial windfalls have been conveyed to individuals in mid-to-late working life at the time of the change, as these individuals face increased taxes for only a relatively few years but are entitled to receive the full advantage of the benefit increases.”

The late Edward Gramlich, a former Federal Reserve Board governor, was one of the nation’s leading thinkers on the topic of reforming Social Security and Medicare. His proposal consisted of two main parts: First, he would eliminate the now \$102,000 (but rising slowly each year) cap on wages that are taxable for Social Security, thus bringing in more revenue.

On the benefit side, Gramlich wanted to raise both the early eligibility age and the normal retirement age for Social

Security, and then keep qualifying ages common across both Social Security and Medicare. On balance, the changes would be sufficient to permanently fund Social Security, but would still leave large holes in parts of the Medicare system. And politically, Gramlich conceded, it might be a tough sell. “The package of taxing all payrolls for Social Security and advancing the normal retirement age is indeed strong medicine,” he said in a 2005 speech.

Figure 5 Social Security and Medicare Income and Cost Rates



Source: *Annual Report of the Social Security and Medicare Boards of Trustees for 2007*.

REFERENCES

- Aguiar, Mark, and Erik Hurst. 2005. "Consumption vs. Expenditure." *Journal of Political Economy* 113 (5): 919–48.
- Bernheim, Douglas, Jonathan Skinner, and Steven Weinberg. 2001. "What Accounts for the Variation in Retirement Wealth Among U.S. Households?" *American Economic Review* 91 (4): 832–57.
- Birkeland, Kathryn, and Edward C. Prescott. 2007. "On the Needed Quantity of Government Debt." Federal Reserve Bank of Minneapolis *Quarterly Review* 31 (1): 2–15.
- Campbell, John. 2006. "Household Finance." *Journal of Finance* 61 (4): 1,553–604.
- Elmendorf, Douglas W., and Louise M. Sheiner. 2000. "Should America Save for its Old Age? Fiscal Policy, Population Aging, and National Saving." *Journal of Economic Perspectives* 14 (3): 57–74.

- Friedman, Milton. 1957. *A Theory of the Consumption Function*. (1st ed.) Princeton, N.J.: Princeton University Press.
- Garner, C. Alan. 2006. "Should the Decline in the Personal Saving Rate Be a Cause for Concern?" Federal Reserve Bank of Kansas City *Economic Review*, Second Quarter: 5–27.
- Gramlich, Edward. 2005. "A First Step in Dealing with Growing Retirement Costs." Remarks at the Spring 2005 Banking and Finance Lecture, Widener University, Chester, Penn., April 21, 2005.
- Gramlich, Edward. 2006. "Another Attempt at Social Security Reform." University of Michigan Retirement Research Center Policy Brief (February).
- Kotlikoff, Laurence, Kent Smetters, and Jan Walliser. 2002. "Finding a Way Out of America's Demographic Dilemma." National Bureau of Economic Research Working Paper 8258.
- Lettau, Martin, and Sydney Ludvigson. 2004. "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption." *American Economic Review* 94 (1): 276–300.
- Love, David, Paul Smith, and Lucy McNair. 2007. "Do Households Have Enough Wealth for Retirement?" Federal Reserve Board of Governors *Finance and Economics Discussion Series* 2007-17.
- Lusardi, Annamaria, and Olivia S. Mitchell. 2007. "Baby Boomers Retirement Security: The Role of Planning, Financial Literacy, and Housing Wealth." *Journal of Monetary Economics* 54 (1): 205–24.
- Lusardi, Annamaria, and Olivia S. Mitchell. 2007. "Financial Literacy and Retirement Planning: New Evidence from the Rand American Life Panel." University of Michigan Retirement Research Center Working Paper 2007-157.
- Martin, Matthew. 2007. "A Literature Review on the Effectiveness of Financial Education." Federal Reserve Bank of Richmond Working Paper 07-03.
- Munnell, Alicia H., and Mauricio Soto. 2005. "What Replacement Rates Do Households Actually Experience in Retirement?" Center for Retirement Research at Boston College Working Paper 2005-10.
- Nakamura, Leonard, and Tom Stark. 2005. "Benchmark Revisions and the U.S. Personal Saving Rate." Federal Reserve Bank of Philadelphia Working Paper 05-6.

- Poterba, James, Steven Venti, and David A. Wise. 2007. "Rise of 401(k) Plans, Lifetime Earnings, and Wealth Retirement." National Bureau of Economic Research Working Paper 13091.
- Poterba, James, Steven Venti, and David A. Wise. 2007. "The Changing Landscape of Pensions in the U.S." National Bureau of Economic Research Working Paper 13381.
- Samwick, Andrew, and Jonathan Skinner. 2004. "How Will 401(k) Plans Affect Retirement Income?" *American Economic Review* 94 (1): 329–43.
- Scholz, John Karl, Ananth Seshadri, and Surachai Khitatrakun. 2006. "Are Americans Saving 'Optimally' for Retirement?" *Journal of Political Economy* 114 (41): 607–43.
- Sheiner, Louise, Daniel Sichel, and Lawrence Slifman. 2007. "A Primer on the Macroeconomic Implications of Population Aging." Federal Reserve Board of Governors *Finance and Economics Discussion Series* 2007-01.
- "Social Security Reform: The Nature of the Problem." U.S. Department of the Treasury, Issue Brief No. 1 (undated).
- Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112 (1): S164–87.
- "The 2007 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Trust Funds." 2007. Social Security Administration, Washington, D.C.

The Rise in Long-Term Unemployment: Potential Causes and Implications

Andreas Hornstein and Thomas A. Lubik

The dramatic rise in long-term unemployment has been one of the most striking features of the Great Recession of 2007–09. The number of unemployed workers who have been out of a job for more than half a year has reached heights that the U.S. economy has not witnessed since the Great Depression.

The overall unemployment rate has been elevated for a substantial period of time, although it has not reached its post–World War II peak of 10.8 percent. Underlying this dire unemployment picture is the rise in long-term unemployment and an overall lengthening of the duration of unemployment spells, which are now far above their levels in previous recessions.

The U.S. labor market historically has been characterized by relatively short unemployment durations for an average worker. The high level of long-term unemployment we are currently seeing represents a sharp break with previous experiences. In the past, most job losses led to only short unemployment spells, as the labor market was able to quickly absorb newly unemployed workers into employment relationships. Although some workers, in particular older ones, experienced long periods of unemployment, the incidence of long-term unemployment in the United States was far less than in other OECD countries. Moreover, although recessions have always been characterized by lengthening unemployment spells, a quick increase in hiring when

■ The authors would like to thank Marianna Kudlyak, Jessica Romero, Aaron Steelman, John Weinberg, and Alexander Wolman for discussions and detailed comments and Nika Lazaryan for excellent research assistance. The views expressed are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

coming out of a recession kept the incidence of long-term unemployment low. The Great Recession seems to be different in that respect.

The high level of unemployment, in combination with a high fraction of long-term unemployment, presents challenges for both monetary and fiscal policymakers. Many of the efforts of the Federal Reserve were aimed at halting the decline in output and employment in the wake of waves of adverse shocks. Arguably, the Fed's policies were successful in that respect. However, the U.S. economy has been operating under extremely low nominal interest rates for such an extended period that additional expansionary monetary actions, such as quantitative easing, are possibly only marginally effective.

In this article we discuss how long-term unemployment has become such a dominant feature of the labor market during the Great Recession. We first summarize the data on aggregate unemployment and the duration distribution of unemployment for the United States since 1960. We then show that, in terms of pure accounting, the composition of the unemployment pool is determined by the inflow and outflow of workers, that is, by the rates at which workers lose and find jobs. We begin from an aggregate perspective and argue that the increase in long-term unemployment can largely be explained by a decline in the exit rate from unemployment. The severity of the recession led to high initial job losses, but the persistent and substantial increase in unemployment and unemployment duration is mainly due to a decline in job finding rates. In response to the increase in long-term unemployment, Congress extended the maximum duration of unemployment benefits from six months to close to two years. We discuss the effects of this extension on unemployment duration and argue that the effects have been limited.

We then proceed to a more disaggregate analysis and study how unemployment of different demographic groups was affected by the Great Recession. We show that unemployment rates and duration differ substantially across demographic groups, but that almost all groups were equally affected by the increase in unemployment rates and duration. We then discuss how negative duration dependence, that is, the apparent decline in job finding rates with the length of time unemployed, affects long-term unemployment. We find that accounting for duration dependence allows us to better model long-term unemployment in the U.S. labor market. This accounting framework also suggests that a significant part of the increase in long-term unemployment is indeed due to the inflow into unemployment of workers with relatively low job finding rates. We conclude by arguing that given the increased contribution to overall unemployment of unemployed workers with inherently low job

finding rates, monetary policymakers may want to exercise caution in the use of policy to respond to the level of unemployment.

1. A LOOK AT THE DATA

The standard measure of unemployment comes from the monthly Current Population Survey (CPS) conducted by the Census Bureau for the Bureau of Labor Statistics. This survey, commonly known as the household survey, is a randomly selected sample of about 60,000 households that report on their employment status and other characteristics.¹

A respondent is classified as employed, unemployed, or out of the labor force. A respondent is classified as unemployed if he or she reports not being employed but is actively searching for a job. The labor force is defined as those respondents who are either employed or unemployed, and respondents who are neither employed nor actively searching for a job are classified as being out of the labor force. The unemployment rate is the ratio of the number of unemployed respondents to the number of workers in the labor force. Conditional on the employment state, there are follow-up questions that further characterize the employment state. Employed respondents are asked about the type of employment (part-time or full-time), their occupation, and the industry of employment, among other questions. Unemployed respondents are asked about the length of the ongoing unemployment spell and their previous occupation and industry. Basic demographic information is also collected, such as the sex, age, race, and education level of the respondent.²

There are two notable features to the pattern of the rise and fall of unemployment over the business cycle. First, unemployment rises

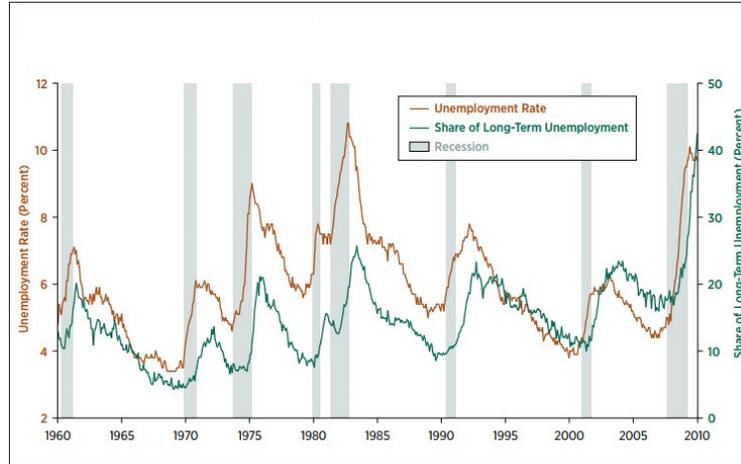
¹ The Bureau of Labor Statistics provides a detailed description of the CPS at <http://www.bls.gov/cps/>.

² We can look at the household survey as providing information on the supply of labor. There are two other surveys that report on the state of the U.S. labor market from the demand side for labor.

The Current Employment Statistics (CES) program, commonly known as the establishment survey, reports on the number of jobs from a sample of about 440,000 establishments in the U.S. nonfarm sector. By construction, the establishment survey provides information on employment only, not unemployment. Furthermore, the establishment survey provides information on jobs and not on household employment. For example, a household survey respondent who works two jobs is counted as employed once, but the establishment survey would count two jobs. Finally, the establishment survey does not cover unincorporated self-employment.

Another recently introduced survey, the Job Openings and Labor Turnover Survey (JOLTS), tries to capture how establishments change their employment. JOLTS provides monthly data on job openings, hires, quits, layoffs, etc., for a sample of about 16,000 establishments.

More detailed descriptions of the CES and JOLTS are provided at <http://www.bls.gov/ces/home.htm> and <http://www.bls.gov/jlt/home.htm>, respectively.

Figure 1 Long-Term Unemployment

Sources: Bureau of Labor Statistics, Haver Analytics, authors' calculations.

Notes: The share of long-term unemployment (more than 26 weeks) as a percent of total unemployment typically increases during recessions (the shaded areas). But following the most recent recession, the share is nearly double the previous peak after the 1981–82 recession.

rapidly at the onset of a recession, but it comes down only slowly over the course of the recovery. Second, long-term unemployment increases sharply with overall unemployment.

Figure 1 depicts the unemployment rate (dark orange line, left axis) and the share of total unemployment that is long-term unemployment (green line, right axis) for the U.S. economy from 1960 through 2010, with recessions highlighted in grey.³ The average unemployment rate for this period is about 6 percent, but unemployment increases substantially in recessions. For example, in the 1981–82 recession the unemployment rate increased by about three percentage points within

³ The recessions are dated using the business cycle peaks and troughs as announced by the National Bureau of Economic Research (NBER). NBER business cycle dates are a widely accepted definition of recessions in the United States. The NBER procedure to date the beginning and end of a recession is supposed to reflect a widespread and significant decline in economic activity. As such, the NBER procedure incorporates a large number of measures of economic activity, including production, sales, income, and employment. Unemployment tends to lag the NBER recession dates, in the sense that the unemployment rate peaks after the end of the recession. For more detailed information on the NBER business cycle program, see <http://www.nber.org/cycles/main.html>.

one and a half years to reach a peak of 10.8 percent in October 1982. In the expansion phase, the unemployment rate then usually declines slowly from its peak. This pattern is especially noticeable for the 1990–91 recession and the 2001 recession, and has given rise to the idea of a “jobless recovery,” in which economic growth picks up, but employment gains are small and unemployment declines only slowly. This pattern seems to be repeating itself in the current recovery.

Long-term unemployment is defined as being unemployed for more than 26 weeks. This is the conventional measure of long-term unemployment since unemployment benefits typically last for about half a year. In Figure 1 we see that the average share of long-term unemployment is quite low, about 15 percent from 1960 to 2010, but in every recession the share of long-term unemployment increases sharply with the unemployment rate. A similar observation applies to the mean duration of unemployment for all those who report job search durations in any month. From 1960 to 2010 the average mean duration of unemployment is about 14 weeks, but mean duration increases significantly in recessions.

The increase in unemployment during the Great Recession stands out for its severity, especially the substantial increase in long-term unemployment. Unemployment peaked at 10.1 percent in October 2009, about one quarter after the official end of the recession according to the National Bureau of Economic Research (NBER) dating scheme, and stayed close to this level for almost one year. For the postwar period, this peak unemployment rate is second only to the 10.8 percent unemployment rate after the 1981–82 recession. The share of long-term unemployment peaked at 46 percent in the second quarter of 2010, and averaged a bit more than 43 percent for all of 2010. This peak value for the share of long-term unemployment is significantly higher than the previous peak of 26 percent that was attained following the 1981–82 recession. Finally, mean duration of unemployment had increased to about 35 weeks by the middle of 2010, again a substantial increase over the previous peak for mean unemployment duration of 21 weeks after the 1981–82 recession. Never before in the postwar period have unemployed workers been unemployed for such a long time.

2. ACCOUNTING FOR UNEMPLOYMENT

We now take a more systematic look at how total unemployment is related to unemployment duration. For this purpose we study how the inflows into unemployment and the outflows from unemployment determine total unemployment. One can think of total unemployment as the water level in a bathtub, which is determined by the inflow of

new water and the rate at which the water drains. The total number of unemployed workers is determined by the rate at which workers become newly unemployed and start looking for work (the entry rate) and the rate at which current unemployed workers find work (the exit rate).⁴ Other things being equal, the more workers who become newly unemployed, that is, the higher the entry rate, then the higher the total number of unemployed workers. Similarly, at a given inflow rate of newly unemployed workers, the less likely it is that an unemployed worker finds a new job, then the higher the total number of unemployed workers will be eventually. For a slightly more formal representation of this model see Box 1.

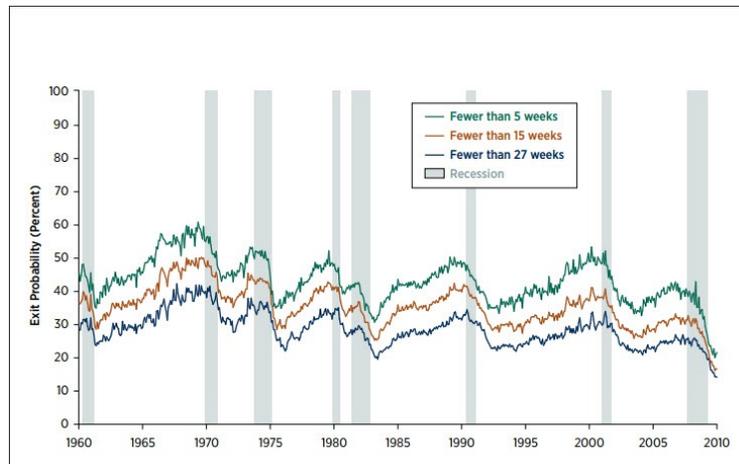
While total unemployment depends on both the entry and exit rates, the average duration of unemployment depends mainly on the behavior of the exit rate. The lower the exit rate, that is, the lower the chance that an unemployed worker becomes employed, then the longer the average unemployment duration and the larger the share of workers who have been unemployed for a long time.

Robert Shimer (2007) shows how one can recover measures of the entry and exit rate from data on total unemployment and data on short-term unemployment, that is, workers who have been unemployed for fewer than five weeks.⁵ The maintained assumption of his accounting exercise is that all unemployed workers are homogeneous in the sense that they all have the same exit rate. This is a simplifying assumption that provides some valuable first insight into the dynamics of unemployment and the interpretation of long-term unemployment. We will relax that assumption below.

Since the increase in the unemployment rates during recessions is usually accompanied by a substantial lengthening of unemployment duration, a declining exit rate must be an important source of high unemployment. In other words, it is hard to find a job during recessions. This observation also suggests that more long-term unemployment does not necessarily mean that the long-term unemployed are in any way different from the short-term unemployed. Even if all unemployed workers face the same exit rate, a decline in the exit rate will

⁴ Given the definition of unemployment, workers may exit unemployment not only because they find work, but also because they stop searching, that is, the workers drop out of the labor force. Alternatively, they may enter unemployment not only because they lose a job, but also because they decide to (re)enter the labor force and search for a job. In the analysis we disregard the flows in and out of the labor force. For most purposes this is not a restrictive assumption (Shimer 2007).

⁵ Similar exercises have been performed by Elsby, Michaels, and Solon (2009) and Fujita and Ramey (2009). An important debate in this literature concerns the relative importance of variations in the job finding rates and the job separation rates in accounting for variations of the unemployment rate.

Figure 2 Duration Dependence in Exit Rates

Sources: Bureau of Labor Statistics, Haver Analytics, authors' calculations.

Notes: Workers who have been unemployed for fewer than five weeks have the highest probability of exiting unemployment within the next month. The likelihood of exit typically declines for all workers during recessions (the shaded areas), but it continued to decline well after the troughs of the last three recessions.

yield higher average unemployment duration and an increased share of long-term unemployment.

We follow Shimer's (2007) simple accounting framework and recover entry and exit rates of homogeneous unemployed workers. In Figure 2 we display the implied exit rates of workers who have been unemployed for fewer than 5 weeks, fewer than 15 weeks, and fewer than 27 weeks. The green line displays the exit rate from unemployment implied by data on short-term unemployment, that is, those workers who have been unemployed for fewer than 5 weeks.⁶ Most of the time unemployed workers find a job quite quickly: the average probability that an unemployed worker finds work within a month is about 40 percent, and at the peak of an expansion this job finding probability can be as high as 60 percent. As we also can see, the exit rate from

⁶ The exit rates displayed in Figure 2 are actually derived from the steady state relationship between unemployment duration shares and the exit rate as described in Box 1. We also calculate exit and entry rates using data on short-term unemployment while not imposing the steady state condition as in Shimer (2007). With a few exceptions the two procedures essentially yield the same series for the exit rate.

unemployment drops sharply in a recession, falling to about 35 percent in previous recessions. Furthermore, in the jobless recoveries after the 1990–91 recession and the 2001 recession, the exit rate from unemployment declined significantly even two years after the recessions' troughs. The 2007–09 recession again stands out in terms of the speed and magnitude of the decline in the exit rate from unemployment. One year after the trough, the probability of finding a job within a month declined to about 20 percent, about half the average exit rate from unemployment and substantially less than in previous recessions.

A model with homogeneous unemployment is consistent with the qualitative features of long-term unemployment in recessions, but it cannot account for the magnitude of long-term unemployment in recessions. Using the entry and exit rates from our unemployment accounting exercise, we can construct counterfactual duration distributions for unemployment. By construction, the parameters of the simple model exactly match total unemployment and the number of workers unemployed for fewer than 5 weeks. A model with homogeneous unemployed workers is not a good match for medium- and long-term unemployment, however. Consistent with the data on previous recessions, the model does predict a sharp rise in medium- to long-term unemployment in recessions. But the model significantly understates the magnitude of long-term unemployment: for almost all recessions the model predicts only one-third of those workers who are unemployed for more than 26 weeks.

3. THE EFFECTS OF UNEMPLOYMENT INSURANCE ON UNEMPLOYMENT

We use the share of unemployed workers who have been unemployed for more than 26 weeks as a measure of long-term unemployment. As noted above, the particular cutoff duration for this conventional definition of long-term unemployment is related to the maximum duration of unemployment benefits, usually 26 weeks. Unemployment compensation programs are administered at the state level, and the amount and duration of benefits may vary across states. The duration of unemployment compensation tends to increase in response to increased unemployment following a cyclical downturn. These changes occur at the state and federal level. In particular, in response to the increase in long-term unemployment in 2008, in June of that year Congress authorized an Emergency Unemployment Compensation (EUC) program that provided an additional 13 weeks of benefits for unemployed workers who were eligible under state programs. After various additional authorizations, by early 2010 the maximum duration of unemployment benefits

was 99 weeks, with some variation across states. For a description of the different programs, see Daniel Aaronson, Bhashkar Mazumder, and Shani Schechter (2010).

The duration of unemployment benefits is extended in order to lessen the negative impact of unemployment on long-term unemployed workers. A side effect of extended benefits can be to lengthen the average duration of unemployment. If we assume that unemployed workers make choices about whether to accept or reject job offers, then increasing or extending unemployment benefits will affect how these choices are made. On the one hand, unemployed workers who are currently eligible for unemployment benefits may be willing to hold out for longer until they receive what they think is a more acceptable offer. This will reduce the exit rate from unemployment for these workers and thereby increase the average duration of unemployment. On the other hand, not every unemployed worker qualifies for unemployment benefits. In order to qualify, a worker must have had a job and must have been laid off. If a worker does not qualify for unemployment benefits, lengthening the duration of unemployment benefits does not mean much for the worker now, but it does make taking a job much more attractive since the worker then qualifies for the extended unemployment benefits should the worker become unemployed again. Thus one might expect that unemployed workers who are ineligible for unemployment benefits become more willing to accept job offers if unemployment benefits are extended.

There is a considerable amount of empirical work on the possible effects of extended unemployment benefits on unemployment duration. Aaronson et al. (2010) and Michael Elsby, Bart Hobijn, and AyŞegül Şahin (2010) survey that work and provide some estimates on how the EUC program may have affected the duration of unemployment. They start with estimates that an extension of unemployment payments by one week tends to increase unemployment duration by 0.1 to 0.2 weeks. Considering that the duration of unemployment benefits has been increased by up to 73 weeks, and that only about 50 percent of unemployed workers are eligible for unemployment benefits, they estimate that the EUC program may have lengthened the average duration of unemployment by between 2 and 6 weeks. This is a significant lengthening, but much less than the actual 18-week increase of average unemployment duration from about 17 weeks in 2008 to 35 weeks in the middle of 2010. Furthermore, as Aaronson et al. (2010) and Elsby et al. (2010) point out, the lower bounds of their estimates are likely to be more relevant than the upper bounds. Finally, using the simple bathtub model of unemployment, one can translate the estimated

increase in unemployment duration to an increase in the unemployment rate of between one and three percentage points.⁷

4. DIFFERENT UNEMPLOYMENT EXPERIENCES

The previous discussion considers only the behavior of total unemployment in the economy. But the labor market experience in the United States is not the same for all members of the labor force. Unemployment rates differ vastly across demographic groups. For instance, as of December 2010 the unemployment rate among individuals who have not completed high school was more than three times the unemployment rate of those with a college degree. It is therefore tempting to hypothesize that some of the higher unemployment and the longer unemployment duration might be due to composition effects. This term describes the idea that during a recession the composition of newly unemployed workers shifts toward demographic groups characterized by lower exit rates and longer durations. In other words, the overall unemployment picture hides deeper dynamics in the labor market that affect specific groups, occupations, or industries in markedly different ways.

In Table 1 we show the average unemployment rate, mean duration of unemployment, and share of long-term unemployment for several demographic groups for the available sample after 1960 and for the year 2010. Three things are apparent: First, unemployment rates and unemployment durations differ significantly across demographic groups. Second, during the 2007–09 recession, unemployment tended to increase more in some groups that in the past were less susceptible to job loss in recessions. Third, for all demographic groups, unemployment rates, mean durations of unemployment, and long-term unemployment shares are significantly higher than their sample averages prior to the recession. While the first two observations point to the possibility that changes in the composition of unemployment inflows might contribute to the overall increase in unemployment, the third observation suggests that changes in composition cannot be a complete explanation for the overall increase in unemployment.

⁷ This estimate is based on the relationship between mean unemployment duration and the exit rate from unemployment described in Box 1. The calculations are conditional on an average monthly job separation rate of 3 percent per month (Shimer 2007). A similar exercise is performed by Mazumder (2011).

Table 1 Demographic Difference in Unemployment

	Unemployment Rate	Unemployment Duration	Share of Long-Term Unemployment
	Sample Avg.* 2010 Avg.	Sample Avg. 2010 Avg.	Sample Avg. 2010 Avg.
Total	6.1 9.6	14.9 33.1	15.0 43.3
Gender			
Female	6.3 8.6	14.8 32.0	14.9 41.5
Male	5.9 10.5	17.6 33.9	18.9 44.7
Age			
24 and younger	12.4 18.4	11.8 23.3	10.6 29.7
25-54	4.8 8.6	18.7 35.5	19.9 46.9
55 and older	3.6 7.0	22.1 40.6	40.1 79.3
Race			
White	5.3 8.7	15.5 32.1	15.9 42.0
Black (1972-2010)	12.2 16.0	18.3 36.7	20.0 48.4
Asian (2000-2010)	4.9 7.5	22.0 36.9	26.1 48.5
Hispanic (1973-2010)	9.1 12.5	17.1 30.7	15.5 39.4
Occupation (2000-2010)			
Management, business, and financial operations	2.9 5.1	22.4 38.4	26.8 52.0
Professional and related occupations	2.9 4.5	19.9 32.6	22.6 43.3
Services	6.9 10.3	17.7 29.9	19.4 38.9
Production, transportation, and material moving	7.9 12.8	20.3 36.5	25.9 61.8

Sources: Bureau of Labor Statistics, Haver Analytics, authors' calculations.

*Monthly observations from 1960 to 2010 unless otherwise noted. ** Data for unemployment duration and share of long-term unemployment are not readily available. Unemployment rates and durations differ across demographic groups. In the 2007-09 recession, unemployment nearly doubled in groups that traditionally have lower unemployment rates, such as college-educated workers and those in management, business, and financial operations.

Table 1 (Continued) Demographic Difference in Unemployment

Industry (2000-2010)	Unemployment Rate	Unemployment Duration	Share of Long-Term Unemployment
	Sample Avg.* 2010 Avg.	Sample Avg. 2010 Avg.	Sample Avg. 2010 Avg.
Construction	10.3	16.9	33.3
Manufacturing	6.3	22.3	39.8
Wholesale and retail trade	6.1	19.2	34.6
Transportation and utilities	5.2	20.4	37.9
Information	5.8	22.8	37.9
Financial activities	3.8	21.0	38.0
Professional and business services	7.2	19.5	33.7
Education and health services	3.6	18.3	30.1
Leisure and hospitality	8.6	16.6	29.5
Education	9.0	14.9	
(1992-2010)**	5.4	10.3	
Less than high school High school completed	4.4	8.4	
Some college	2.6	4.7	
College completed			

Sources: Bureau of Labor Statistics, Haver Analytics, authors' calculations.

*Monthly observations from 1960 to 2010 unless otherwise noted. ** Data for unemployment duration and share of long-term unemployment are not readily available. Unemployment rates and durations differ across demographic groups. In the 2007-09 recession, unemployment nearly doubled in groups that traditionally have lower unemployment rates, such as college-educated workers and those in management, business, and financial operations.

The average unemployment rate in 2010 was more than 50 percent higher than the average unemployment rate from 1960 through 2010. Although the unemployment rate of males tends to be somewhat lower than the unemployment rate of females, in 2010 the unemployment rate of males increased relative to that of females. Across age groups, the unemployment rate of younger workers (under 25) tends to be higher than that of older workers (over 55), but in 2010 the unemployment rate of older workers increased relative to that of younger workers.⁸ Overall, the unemployment rate of workers who have not completed high school is about three times as high as that of workers with a college degree, yet in 2010 the unemployment rate of college-educated workers increased slightly relative to that of workers without a high school degree. Workers in managerial occupations related to business and financial operations have some of the lowest unemployment rates among all occupations, certainly compared to service-oriented occupations and occupations in the production sector, but in 2010 the unemployment rate for managerial occupations increased somewhat relative to these other occupations.⁹ Across industries, the average unemployment rate in construction in 2010 was nearly three times the rate in financial activities, but in both industries the rate almost doubled over the sample average. Since construction and financial services were at the heart of the 2007–09 recession, it should not be too surprising that workers affiliated with both industries experienced some of the biggest increases in unemployment rates.

The differences in unemployment rates across demographic groups are accompanied by similar differences in mean unemployment duration and long-term unemployment shares, although the relationship between these variables is not particularly tight. For example, on average the mean unemployment duration of older workers is more than twice the mean duration of younger workers, yet the unemployment rate of older workers is less than half that of younger workers. This observation highlights that there are two determinants of unemployment, inflows and outflows, as discussed previously.¹⁰ The low unemployment

⁸ Across race groups, unemployment rates tend to be lower among whites, but relative to the whole sample unemployment rates among whites increased in 2010. One should be careful when comparing the unemployment rate changes of different race groups since the sample periods do not coincide.

⁹ The occupation and industry affiliation of an unemployed worker refer to the last job held by that worker. It is not uncommon for workers to change occupations or industries, even without an intervening unemployment spell. The classification of an unemployed worker by last known employment can be useful if it reflects on the human capital that an unemployed worker has acquired and that affects the job search decisions of that worker.

¹⁰ See also the discussion of unemployment in OECD countries in Box 2.

rate for older workers, then, is mainly due to a very low inflow rate into unemployment—in other words, a very low probability of losing a job. On the other hand, once an older worker loses a job and becomes unemployed, the probability of finding a new job is very low compared to a younger worker. Nevertheless, it appears as if a general decline in job finding rates was an important driver of the increased unemployment rate in the 2007–09 recession. Across all demographic groups, there are comparable increases in mean unemployment duration and long-term unemployment shares in 2010.

We have documented differences in the way unemployment rates, mean duration, and long-term unemployment changed for different demographic groups, but we do not want to overemphasize these differences since essentially all groups experienced significant increases in unemployment. A more thorough analysis of the role of demographic changes and their contributions to the average duration of unemployment is provided by Aaronson et al. (2010). They compare average unemployment duration in (1) the expansion phases following the 1981–82 and the 2001 recessions, and in (2) the first six months following the 1981–82 and the 2007–09 recessions. To a first approximation, Aaronson et al. (2010) calculate the change in total unemployment that is attributable to two different factors. First, they calculate the change in unemployment duration that would have occurred given the change in the demographic composition of the labor force, but assuming that the unemployment durations within demographic groups do not change. Second, they calculate the change in unemployment duration that would have occurred assuming no change in the demographic composition of the labor force, but allowing for the observed change of unemployment durations within demographic groups.¹¹ They find that comparing the expansion phases after the 1981–82 and 2007–09 recessions, changes in the labor force composition account for less than half of the trend change in unemployment duration. Furthermore, comparing the periods immediately after the 1981–82 and 2007–09 recessions, changes in the labor force account for only one-fifth of the difference in unemployment duration.

¹¹ There is also a third effect, which captures any interactions between changes in the relative size of demographic groups and changes in durations within demographic groups.

5. DURATION DEPENDENCE OF UNEMPLOYMENT

People are different, and survey measures do not capture all the characteristics that are relevant to unemployment duration. Some characteristics that are relevant to the chances of an unemployed worker finding work can be quite persistent yet unobservable, and these characteristics might actually be related to the unemployment experience itself. For example, consider two equal pools of unemployed workers who at the beginning of the month share the same observable characteristics, except for the time that they have been unemployed already. On average, at the end of the month relatively more workers from the pool with the shorter unemployment duration will have found work. In other words, the longer a worker has been unemployed already, the less likely it is that he or she will find a job. This apparent decline in exit rates with the length of time unemployed is called “negative duration dependence.”¹²

Negative duration dependence is clearly inconsistent with the simple model of homogeneous unemployment that we discussed previously, since that model assumes that in any given month all unemployed workers have the same chance of finding work, independent of how long they already have been unemployed. This observation may account for the fact that the simple model understates the prevalence of long-term unemployment.

As noted previously, we can use the entry and exit rates from our unemployment accounting exercise to construct counterfactual duration distributions for unemployment. When we account for unemployment in the previous section, we use the share of workers who were unemployed for fewer than 5 weeks to calculate estimates of the exit rate from unemployment. Suppose we were instead to use the share of workers who were unemployed for fewer than 15 weeks to calculate the exit rate from unemployment. The simple model imposes the same exit rate on workers who were unemployed for fewer than 5 weeks and on workers who were unemployed for between 5 and 15 weeks.

But if there is negative duration dependence, the exit rate we calculate when we use the share of workers who have been unemployed for fewer than 15 weeks should be less than the exit rate we calculate when we use the share of workers who have been unemployed for fewer than 5 weeks. In Figure 2 we display the exit rates from unemployment based on different segments of the duration distribution of

¹² See Machin and Manning (1999) for a survey on the role of duration dependence in the determination of long-term unemployment in Europe.

unemployment: the share of workers who have been unemployed for fewer than 5 weeks, fewer than 15 weeks, and fewer than 27 weeks. In fact, consistent with negative duration dependence, the implied exit rates decline monotonically as unemployment duration increases.

Two explanations have been proposed for the observed negative duration dependence of exit rates from unemployment. The first explanation simply assumes that for each unemployed worker, the exit rate is a declining function of elapsed unemployment duration. Then the exit rate from the unemployment pool declines with the duration that the pool's members have been unemployed. This approach is called "true duration dependence." An alternative explanation is to assume that newly unemployed workers already differ according to their exit rates from unemployment. Even if the exit rate for an individual worker does not change over time, the composition of the pool will change over time, which implies a change in the average exit rate from the pool. In particular, over time workers with a high exit rate will make up a smaller and smaller share of the remaining pool of workers who have not yet found work, which implies a declining average exit rate from the pool. This approach is called "unobserved heterogeneity."

Various reasons can account for true duration dependence in exit rates. For one, over time unemployed workers tend to lose skills associated with actual work experience and work-related training. This decline in human capital implies that the average wage offer an unemployed worker could obtain probably also would decline over time.

If the benefits of staying unemployed remain constant over time, for example via constant unemployment insurance payments, while the average wage offer is declining, then the likelihood that an unemployed worker accepts an offer probably also declines over time, and so would the exit rate. Additionally, over time unemployed workers lose attachment to networks that may aid in finding new jobs. Finally, potential employers might interpret a prolonged unemployment spell as a signal of ability, irrespective of the true, underlying characteristics of the unemployed worker. All of this means that exit rates from unemployment would decline over time.¹³

Unobserved heterogeneity does not need any particular story. Clearly surveys do not capture all the information that is relevant to the determination of exit rates from unemployment. For example, a worker who loses a job for reasons that are idiosyncratic to the previous employer

¹³ See Ljungqvist and Sargent (1998) for an example that studies the implications of duration dependence due to human capital depreciation. See Blanchard and Diamond (1994) for an example of duration dependence due to employer screening of long-term unemployed.

may have skills that are valued by a wide range of employers, and may find work relatively quickly. On the other hand, if a worker loses a job in an industry or occupation that is in secular decline, the skills of that worker may not be easily transferable to a wide range of employers, and this worker may stay unemployed for a long time.

The two explanations of duration dependence potentially have different policy implications. If true duration dependence is widespread among unemployed workers, then current high levels of unemployment might imply high future unemployment because more unemployed workers make the transition to long-term unemployment. In this case, a reduction in current unemployment, if possible, would reduce future unemployment. On the other hand, if unobserved heterogeneity accounts for duration dependence and the increase in unemployment, and the duration of unemployment is mainly due to an influx of workers with low exit rates, it might be a signal that unemployment is due to a “mismatch” of skills. In this case, high unemployment may reflect structural change and may not be amenable to monetary policy actions.

Preliminary work by one of the authors of this essay indicates that a simple framework with two types of unemployed workers—short-term unemployed with a relatively high exit rate from unemployment and long-term unemployed with a relatively low exit rate—can account quite well for the observed variation in the duration distribution of unemployment (Hornstein 2011). This framework allows for two special cases. The first case consists only of true duration dependence: all unemployed workers are initially short-term unemployed with a high exit rate from unemployment, but during unemployment workers make a random transition from short-term to long-term unemployment. The second case consists only of unobserved heterogeneity: workers are from the beginning either short-term or long-term unemployed, and they never change types.

The two special cases of the framework provide different accounts of unemployment volatility. In the true duration dependence case, almost all of the unemployment rate fluctuations are attributed to exit rate fluctuations of the two types. Changes in entry rates of short-term unemployed workers and transition rates from short-term to long-term unemployment—that is, the true duration mechanism itself—have only a limited effect. In the unobserved heterogeneity case, on the other hand, a substantial portion of the unemployment rate fluctuation is attributable to changes in the entry rate of long-term unemployed workers. In this case, fluctuations in entry and exit rates of long-term unemployed workers account for about two-thirds of unemployment rate

volatility. Of the two cases, the unobserved heterogeneity approach provides a better match to the unemployment duration distribution.

The more general framework that allows for both true duration dependence and unobserved heterogeneity yields results that are closer to the special case of only unobserved heterogeneity. In the general framework, unemployment volatility is about equally accounted for by changes in the entry rate of long-term unemployed workers and the exit rates of both types. Furthermore, it appears as if the observed increase in unemployment in the 2007–09 recession is mainly driven by the increased entry rate and reduced exit rate of long-term unemployed workers.¹⁴ Given the above interpretation of long-term unemployment due to unobserved heterogeneity, one could then argue that most of the increase in unemployment in the 2007–09 recession represents an increase in structural unemployment.

6. LONG-TERM UNEMPLOYMENT AND MONETARY POLICY

A simple view of the statistical relationship between unemployment and inflation—the Phillips curve—suggests that the choices for monetary policymakers who want to promote employment and price stability are clear. Unemployment is high and inflation is low, therefore monetary policy can and should be expansionary. Yet many modern macroeconomists argue that movements in real quantities matter for inflation dynamics only to the extent that they depart from their natural level. Furthermore, the best way to attain low unemployment volatility in the long run is to follow policy rules that promote price stability.¹⁵ In this view, the labor market situation in the wake of the Great Recession still presents a challenge to monetary policymakers.

Macroeconomic theory defines the natural rate of unemployment as the hypothetical level of unemployment that would obtain in the absence of any distortions, such as impediments to free adjustment of nominal prices and wages. The difference between actual and natural unemployment is often referred to as the “unemployment gap.” It constitutes a measure of the degree of slack, or under-utilization of resources, in the economy; a large and positive unemployment gap may constrain inflationary pressures. With a large pool of unemployed workers to hire from, wages are unlikely to increase, which therefore

¹⁴ According to the more general framework, the behavior of unemployment in the 2007–09 recession is similar to its behavior in the 1981–82 recession. It is unlike the 1973–75 and 1990–91 recessions, where the increase in the unemployment rate was mainly driven by a general decline in exit rates.

¹⁵ For one exposition of this view, see Lacker and Weinberg (2006).

limits pricing pressures stemming from rising input costs. This scenario seemingly describes the recent economic climate, with unemployment persistently high and inflation trending gradually lower over the course of the recession. How useful this argument is for policy decisions depends on how easy it is for policymakers to discern the level of the natural rate. A main point of contention in the current policy debate is whether the natural rate has substantially shifted upward over the course of the Great Recession. The existence of very high long-term unemployment has implications for this debate.

We have shown that the prevalence of long-term unemployment is related to the fact that the exit rate out of unemployment declines with the duration of unemployment. Since a high fraction of long-term unemployed workers find it difficult to transition to employment, the pool of workers who can reasonably expect to be hired may be effectively smaller than it appears from the raw unemployment numbers alone. Thus the natural rate of unemployment would be higher, and the unemployment gap smaller, than what one might infer from the measured level of unemployment. Furthermore, a preliminary account of unemployment during the Great Recession seems to indicate that overall unemployment increased because of increased entry rates and reduced exit rates of long-term unemployed workers, suggesting that structural unemployment, and therefore the natural rate of unemployment, has increased.

How, then, should monetary policymakers respond to the increase in long-term unemployment? To the extent that the exceptionally large share of long-term unemployment reflects structural change and a higher natural rate of unemployment, policymakers should seriously consider the possibility that a high unemployment rate does not necessarily equate to a large unemployment gap. Furthermore, if higher long-term unemployment were to become a permanent feature of the U.S. labor market, then the level of unemployment would be even less likely to respond to short-term monetary stimulus. Any policy options to deal with permanent long-term unemployment would likely have to take the form of structural labor and product market reforms that increase the ability and willingness of the unemployed to find work, and reduce the costs of generating and maintaining employment relationships. Reforms of this kind arguably reduced the incidence of long-term unemployment in, for instance, the United Kingdom in the 1980s and Germany during the past decade.

APPENDIX

The following sections appeared as Box 1 and Box 2 in the original article.

1. A SIMPLE FRAMEWORK FOR UNEMPLOYMENT ACCOUNTING

We can formalize the bathtub model of unemployment described in the text as follows. The labor force consists of a fixed number of L workers who are either employed, E , or unemployed, U , and $L = E + U$. For simplicity assume that employed workers become unemployed at a constant rate σ and that unemployed workers become employed at a constant rate λ . Then the rate of change of unemployment, \dot{U} , is simply the difference between inflows and outflows,

$$\dot{U} = \sigma E - \lambda U.$$

Inflows and outflows and the change in unemployment are to be interpreted as occurring instantaneously. The unemployment rate is $u = \frac{U}{L}$ and the employment rate is $e = \frac{E}{L} = 1 - u$. The rate of change of the unemployment rate is

$$\dot{u} = \sigma(1 - u) - \lambda u.$$

Inflow and outflow rates may change over time, but if these rates remain constant, unemployment will converge to a rest point or steady state, u^* . If inflows exceed (fall short of) outflows, the unemployment rate will increase (decline), $\dot{u} > 0$ ($\dot{u} < 0$), toward the steady state. In the steady state, inflows and outflows just balance such that the unemployment rate remains constant, $\dot{u} = 0$,

$$u^* = \frac{\sigma}{\sigma + \lambda}.$$

If the exit rate from unemployment is large relative to the inflow into unemployment, convergence to the steady state will be fast. In this case, not much is lost in thinking about unemployment in any given month as steady state unemployment corresponding to the inflow and outflow rates for that month, and ignoring the convergence to the steady state.

This simple model assumes that every unemployed worker faces the same chance of exiting the unemployment pool. In particular, this

exit rate is independent of the time the worker has been unemployed. Again, assuming that the inflow and outflow rates remain unchanged, we can calculate the implied duration distribution of unemployment in the steady state. The share of unemployed workers who have been unemployed for no more than duration T is then given by

$$\omega_T^* = 1 - e^{-\lambda T}.$$

Entry and exit rates in U.S. unemployment are indeed quite high, and we can interpret unemployment and the duration distribution of unemployment as being close to their steady states. In the text we use data on the duration distribution to recover estimates of the exit rate from unemployment. Given an estimate of the exit rate, we then use data on unemployment to obtain estimates of the entry rate into unemployment.

Another way to relate the exit rate from unemployment to observables is to consider its implications for the average duration of unemployment. Our description of the outflows from unemployment—that an unemployed worker becomes employed at the instantaneous rate λ independent of how long that worker has been unemployed—corresponds to a particular stochastic process, namely a Poisson process. For such a process the average duration that a worker is unemployed is simply the inverse of the exit rate,

$$\bar{T} = \frac{1}{\lambda}.$$

The relationship between the exit rate from unemployment and the average duration of unemployment allows us to obtain a back of the envelope calculation of the effect of extended unemployment benefits on the unemployment rate. Suppose that an extension of the length of unemployment benefits increases the average duration of unemployment from T_0 to T_1 . That implies a reduction in the exit rate from unemployment from $\lambda_0 = \frac{1}{T_0}$ to $\lambda_1 = \frac{1}{T_1}$. Everything else the same, that is, with no change in the separation rate, the steady state unemployment rate increases from $u_0 = \frac{\sigma}{(\sigma+1/T_0)}$ to $u_1 = \frac{\sigma}{(\sigma+1/T_1)}$.

2. THE PERSISTENCE OF LONG-TERM UNEMPLOYMENT: AN INTERNATIONAL COMPARISON

Many European economies experienced high rates of unemployment associated with significant long-term unemployment throughout the

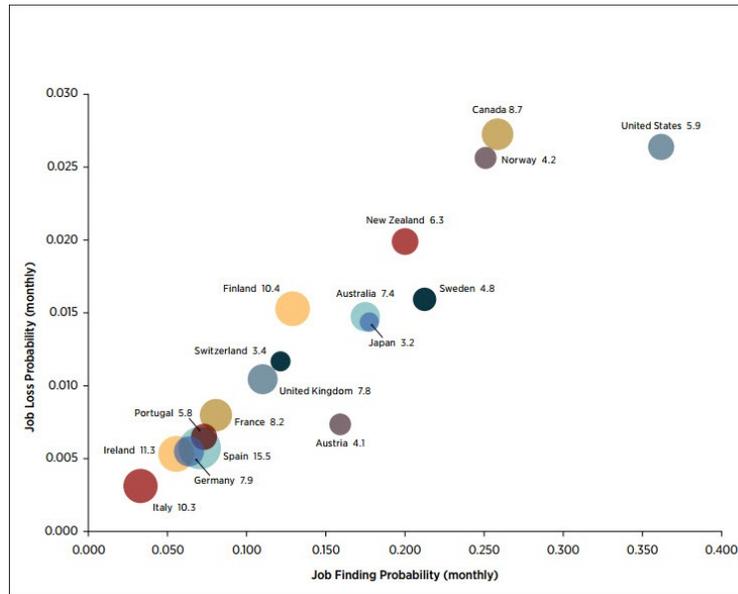
1980s and 1990s.¹⁶ This high European unemployment has been attributed to the interaction of labor market institutions with structural and monetary shocks. For example, one of the authors of this essay (Hornstein), together with Per Krusell and Giovanni Violante (2007), argues that in response to a common acceleration of embodied technological change, different labor market institutions in continental Europe and the United States led to a differential response of unemployment and wage inequality in these countries. On the other hand, Laurence Ball (1997) sees the original common shock in a series of disinflations induced by monetary policy around 1980, but also argues that the impact on unemployment differed depending on the countries' labor market institutions. European unemployment rates eventually declined, in some countries arguably due to structural reforms in product and labor markets in the wake of the European Monetary Union.

In the following discussion we provide a short summary of the determinants of unemployment in OECD countries based on the data set provided by Michael Elsby, Bart Hobijn, and AyŞegül Şahin (2011). In our analysis of cyclical long-term unemployment in the United States, we have suggested that a general decline in exit rates from unemployment is an important source of increased long-term unemployment following a recession. The cross-sectional data for the OECD countries, on the other hand, suggest that both entry rates and exit rates are important drivers of unemployment in the long run. Based on the analysis of the simple model in Box 1, we construct job finding rates using the fraction of workers who have been unemployed for fewer than three months.¹⁷ Using this exit rate from unemployment and the unemployment rate, we construct job separation rates. We construct job finding and job separation rates for each country for each available year. In Figure 3 we display the average job finding and job separation rates for each country. We express these rates as the probability that in any month an employed (unemployed) worker will become unemployed (employed).

¹⁶ For a survey, see Machin and Manning (1999).

¹⁷ Our procedure assumes (1) that the country data for unemployment rates and duration distributions reflect steady states, and (2) that there is no duration dependence in exit rates from unemployment. The steady state assumption is a good approximation for labor markets with relatively high job finding rates, such as the U.S. labor market. Since the job finding rates are much smaller for almost all other OECD countries, our procedure is potentially less reliable for these countries. It turns out that our estimates from the simple steady-state-based procedure are not that different from the estimates one obtains if transition dynamics are taken into account. Elsby et al. (2011) argue that for most of the continental European countries, there is no significant evidence for duration dependence in exit rates, but that there is evidence for duration dependence in most of the other countries.

Figure 3 Labor Market Turnover in Selected OECD Countries



Sources: Elsby, Hobijn, and Şahin (2011), OECD, authors' calculations.

Notes: The number to the right of each country name is the average unemployment rate from 1968–2006. U.S. workers on average are more likely to become unemployed than in other OECD countries, but they also find new work more quickly. In countries with low job turnover, such as Italy and Spain, the average unemployment rate is much higher than in the United States. The number to the right of each country is the country's average unemployment rate from 1968–2006.

The fluidity of the U.S. labor market stands out when compared to the labor markets of almost all other countries. This is especially true when compared to several continental European countries that have high unemployment rates. In the United States, the average unemployed worker has a more than 35 percent chance of finding work within a month, while at the same time there is a less than 3 percent chance that a worker becomes unemployed within a month. The high job finding rate more than counteracts the high job separation rate, such that at 6 percent the average U.S. unemployment rate is quite low compared to most other countries. The Italian labor market, on the other hand, displays very low turnover. Workers rarely become unemployed and unemployed workers take a very long time to find work; job

separation and job finding probabilities are one-tenth of those in the United States. In Italy this extremely sclerotic labor market results in an unemployment rate that is almost twice that of the United States.

Looking at the cross-section of countries in Figure 3, it appears that the less turnover there is in a country's labor market, the higher the unemployment rate. Most of the continental European countries with high average unemployment rates are concentrated in the lower left hand corner of Figure 3, with low job finding and job separation probabilities. One should be careful not to draw too many conclusions from this observation about the causality between transition rates and the level of unemployment, but there is evidence that low separation rates due to rigid labor market laws can lead to low job finding rates. (For example, see Giuseppe Bertola and Andrea Ichino [1995].) Finally, comparing Figures 2 and 3 we can see that the U.S. job finding rate, even at an extreme cyclical trough like the one after the 2007–09 recession, is still higher than the average job finding rate in most other OECD countries.

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 (2): 28–51.
- Ball, Laurence M. 1997. "Disinflation and the NAIRU." In *Reducing Inflation: Motivation and Strategy*, edited by Christina Romer and David Romer, 167–94. Chicago: University of Chicago Press.
- Bertola, Giuseppe, and Andrea Ichino. 1995. "Wage Inequality and Unemployment: United States versus Europe." In *NBER Macroeconomics Annual*, edited by Ben S. Bernanke and Julio J. Rotemberg, 13–66. Boston: MIT Press.
- Blanchard, Olivier Jean, and Peter A. Diamond. 1994. "Ranking, Unemployment Duration, and Wages." *Review of Economic Studies* 61 (3): 417–34.
- Elsby, Michael W. L., Bart Hobijn, and AyŞegül Şahin. 2010. "The Labor Market in the Great Recession." *Brookings Papers on Economic Activity* Spring: 1–48.

- Elsby, Michael W. L., Bart Hobijn, and AyŞegül Şahin. 2011. "Unemployment Dynamics in the OECD." Federal Reserve Bank of San Francisco Working Paper 2009–04.
- Elsby, Michael W. L., Ryan Michaels, and Gary Solon. 2009. "The Ins and Outs of Cyclical Unemployment." *American Economic Journal: Macroeconomics* 1 (1): 84–110.
- Fujita, Shigeru, and Garey Ramey. 2009. "The Cyclicalities of Separation and Job Finding Rates." *International Economic Review* 50 (2): 415–30.
- Hornstein, Andreas. 2011. "Accounting for Unemployment: The Short and Long of It." Mimeo, Federal Reserve Bank of Richmond.
- Hornstein, Andreas, Per Krusell, and Giovanni L. Violante. 2007. "Technology-Policy Interaction in Frictional Labour-Markets." *Review of Economic Studies* 74 (4): 1,089–124.
- Lacker, Jeffrey M., and John A. Weinberg. 2006. "Inflation and Unemployment: A Layperson's Guide to the Phillips Curve." Federal Reserve Bank of Richmond *2006 Annual Report*, 4–26.
- Ljungqvist, Lars, and Thomas J. Sargent. 1998. "The European Unemployment Dilemma." *Journal of Political Economy* 106 (3): 514–50.
- Machin, Stephen, and Alan Manning. 1999. "The Causes and Consequences of Longterm Unemployment in Europe." In *Handbook of Labor Economics Vol. 3*, edited by Orley Ashenfelter and David Card, 3,085–139. Amsterdam: Elsevier.
- Mazumder, Bhashkar. 2011. "How Did Unemployment Insurance Extensions Affect the Unemployment Rate in 2008–10?" Federal Reserve Bank of Chicago *Chicago Fed Letter* no. 285.
- Shimer, Robert. 2007. "Reassessing the Ins and Outs of Unemployment." Mimeo, University of Chicago.

Unsustainable Fiscal Policy: Implications for Monetary Policy

Renee Haltom and John A. Weinberg

The debt of the United States government that is held by the public reached its highest point since World War II in 2011, at 67.7 percent of gross domestic product (GDP).¹ Annual deficits surpassed 10 percent of GDP in 2009, the highest level since 1945, dipping to 8.7 percent of GDP in 2011. The early-to-mid 1980s was the only other point in the postwar period in which deficits exceeded 5 percent of GDP.

Recent numbers are high by historical comparison, but more important than the current size of the deficit and debt is the path they are likely to follow in the future. Federal debt held by the public was actually higher after World War II than it is today—109 percent of GDP in 1946, the highest level on record—but a key difference was that large deficits then were almost entirely associated with the temporary war effort. The same cannot be said today; several factors point to large demands on fiscal resources for most of the foreseeable future. Most

■ The authors would like to thank Andreas Hornstein, Thomas A. Lubik, Aaron Steelman, and Alexander Wolman for discussions and detailed comments. The views expressed are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

¹ There are two common ways to measure the government's debt burden. Debt held by the public, used in this essay, reflects government borrowing from private financial markets. Total federal debt, the second common measure, comprises debt held by the public (private investors, including the Federal Reserve) and debt held by government accounts. The two measures have different implications. Debt held by the public can affect the current economy by crowding out private borrowing. In contrast, debt held by government accounts reflects internal transactions that are not traded in capital markets. However, that debt is nonetheless a legal liability of the federal government and a burden on taxpayers, which is why total debt is also used as a measure of the government's overall debt burden. We focus on debt held by the public because that is the measure for which long-term projections are readily available.

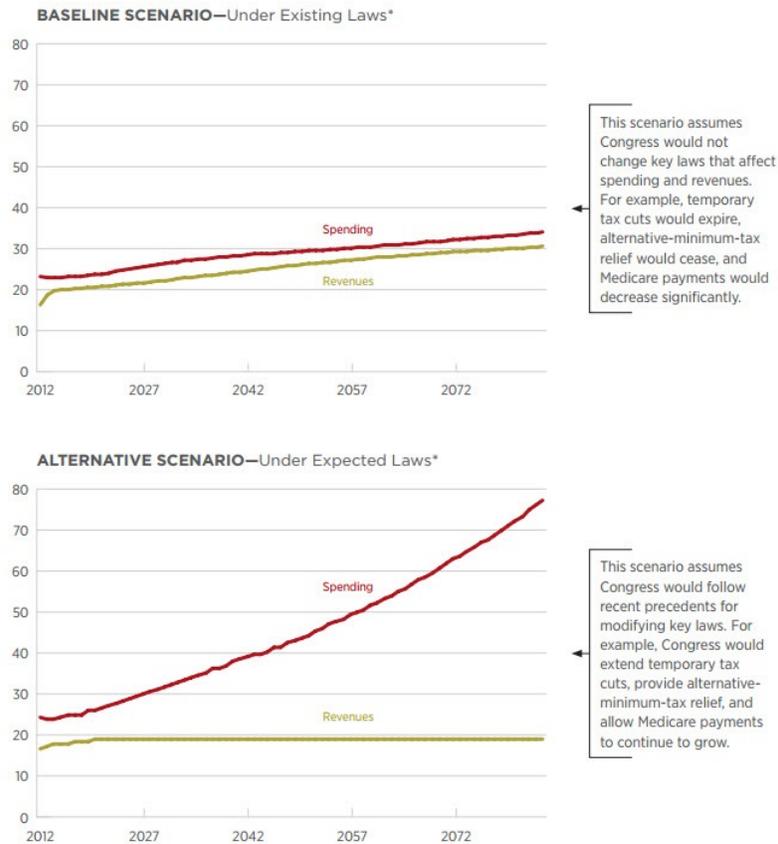
prevalent is the aging population. The first baby boomers reached retirement age in 2011, and the fraction of the population aged 65 or older will surpass 20 percent by 2035, compared to 13 percent today. For the past 30 years, there have been roughly five working people in the United States for every person of retirement age; that number will drop to 2.8 after 2035. This “dependency ratio” is a rough approximation of the number of working individuals in the economy that support, through taxes and Social Security contributions, the people drawing age-related benefits from the government. The aging population will impose significant demands on federal resources through Social Security, Medicare, and Medicaid. These programs are written into law, which means their spending is not determined annually by the federal budgets created by the U.S. president and Congress, but instead can only be reduced through major overhauls to law.²

The nonpartisan Congressional Budget Office (CBO) projects the federal government’s long-term budget outlook under two scenarios: a “baseline” scenario that holds current laws constant and an “alternative” scenario that incorporates the effects of laws the CBO deems likely to pass. (The budget outlooks under both scenarios are displayed in Figure 1.) The baseline scenario reflecting current laws presents the more optimistic view of the future path of fiscal policy. Tax revenues are projected to reach much higher levels than in recent history, while each category of spending except that on Social Security, health care entitlements, and interest payments on debt is projected to fall to its lowest level since World War II. Still, the increase in revenues and decline in other spending would be slightly more than offset by increased spending on Social Security, Medicaid, and Medicare as the population ages. Therefore, deficits would remain positive, causing debt levels to grow slowly over time. Under the baseline scenario, debt held by the public would rise to 84 percent of GDP by 2035, staying in that ballpark for the remaining decades of the forecast. (See Figure 2.)

The alternative scenario—the one the CBO considers more likely—presents a more alarming picture of the growth in federal debt. In

² The aging population may not be the only source of coming strains on government budgets. Additional, though less certain, liabilities stem from the government’s implicit support of other sectors of the economy. This is the support that market participants may assume the federal government will provide to certain markets in the event of trouble, including contingent support to the housing agencies Fannie Mae and Freddie Mac, as well as private pension funds. Whether the government ever will provide this implicit support is highly uncertain, but John Walter and Nadezhda Malysheva (2010) estimated that more than half the private financial sector—potentially \$25 trillion in liabilities, far greater than the size of the economy—was likely to enjoy some explicit or implicit federal backing at the end of 2009. Not included in their analysis were public sector pensions, which are underfunded by more than \$3 trillion, more than triple states’ outstanding debts, according to the most pessimistic estimates.

Figure 1 Projected Budget Gaps (as a Percent of GDP)

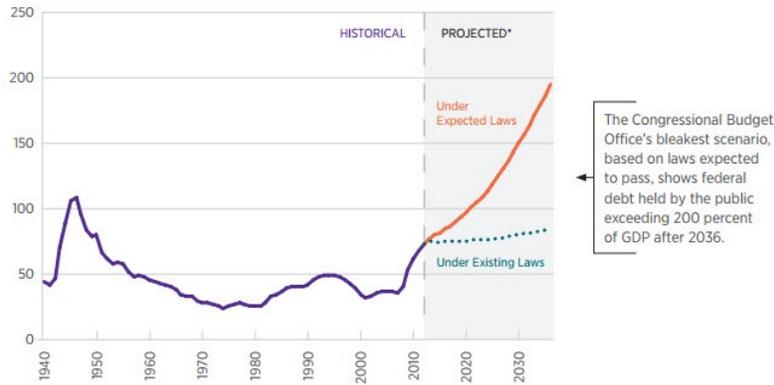


Source: Congressional Budget Office’s 2011 Long-Term Budget Outlook.

Notes: The Congressional Budget Office produces two long-term budget projections: the “baseline” scenario, based on current laws, and the “alternative” scenario, based on laws expected to pass. *Projections begin with the 2012 budget.

that scenario, revenues do not rise much from where they are today, yet spending grows rapidly. This is because of law changes the CBO deems likely to take place, including an extension of the tax cuts that were enacted since 2001 and extended in 2010. The CBO also assumes that tax laws will be changed to keep tax revenues close to their long-run average of 18.4 percent of GDP, rather than rising to historically large levels as they do in the baseline scenario. In addition,

Figure 2 Federal Debt Held by the Public (as a Percent of GDP)



Source: Congressional Budget Office's 2011 Long-Term Budget Outlook.

Federal debt held by the public consists primarily of U.S. Treasury securities, including those held by the Federal Reserve. It does not include debt held in federal government accounts or securities issued by Fannie Mae or Freddie Mac. *Projections begin with the 2012 debt level.

Medicare payments are not assumed to decrease as current law dictates, health care spending under the major reform bill passed in 2010 is not assumed to decrease after 2021 as current law prescribes, and spending on non-entitlement programs is not assumed to fall as rapidly as in the baseline scenario. Under these conditions, federal debt held by the public would rise sharply after 2011, exceeding its historical record of 109 percent of GDP as early as 2023. It would surpass 200 percent of GDP—far more than double today's share of GDP—by the late 2030s.

The two scenarios represent optimistic and pessimistic alternatives from a range of possible outcomes. The exercise shows that the evolution of the federal government's fiscal position depends largely on policy decisions that have yet to be made. Given the demands on fiscal resources coming from the aging population under existing laws, achieving a path toward fiscal balance will involve very difficult tradeoffs for fiscal policymakers.

1. UNSUSTAINABLE FISCAL POLICY

Economists use the word “unsustainable” to describe debt levels projected by the CBO’s alternative scenario, a characterization reflecting the likelihood that financial markets would force a painful adjustment in fiscal policy before such debt levels could be reached. That notion is based on a simple framework called the government’s intertemporal budget constraint. “Intertemporal” simply means “over time,” while a budget constraint is a basic accounting identity that says an entity must pay for everything that it purchases. The government’s intertemporal budget constraint says that the value of the government’s outstanding debt must equal the present value of its expected future surpluses—that is, what financial markets believe surpluses will be, calculated in today’s dollars.

The intertemporal budget constraint suggests that any time the real debt increases by even a small amount—a budget deficit is run in a single year—the expectation of future taxes or spending must adjust to put the equation in balance. However, the equation says only that surpluses must eventually rise; it provides no guidance on when that must occur. Historical experience doesn’t provide a great deal more insight. For example, the U.S. government ran moderate deficits, averaging roughly 3 percent of GDP every year, from 1970 to 1997, with no obvious concern from financial market participants about the sources of future surpluses. That experience would imply that governments can sustain moderate deficits seemingly indefinitely.

That is less likely to be true when the imbalance between outstanding debt and future surpluses is very large. The larger the debt grows, the larger future surpluses—revenues in excess of spending—must be to satisfy the equation. However, there are limits to future surpluses. Spending cannot drop to zero; to the contrary, spending is expected to rise to historically high levels as a percent of GDP even under the CBO’s most optimistic scenario, and tax revenues have an upper limit. As tax rates grow higher, they distort incentives to work and produce, and at very high rates would shrink the revenue collected by the government. There are likely to be political limits to tax revenues even before that point is reached, a reality reflected in the CBO’s alternative scenario assumption that tax revenues will revert to their historical average of 18.4 percent of GDP within a decade. With debt levels predicted to grow much larger than GDP within two decades, it is clear that many years of higher taxes would be required to produce enough surpluses to resolve the resulting imbalance. There is some level of debt that is high enough—although how high is difficult to predict—that generating the amount of future surpluses required would simply be infeasible.

That point is what economists have called the “fiscal limit.” At the fiscal limit, the government cannot borrow further, and the government’s existing spending promises therefore cannot be funded. At least one of two events must occur at the fiscal limit: the government would reduce its debt levels by defaulting, or real debt levels would be reduced through actions taken by the central bank.

There are two main ways in which central banks can improve governments’ fiscal positions. The first is through “seigniorage,” the revenue that governments effectively receive when central banks create money. In the United States, it comes from the interest the Fed earns on the Treasury securities it purchases to expand the money supply. The Fed retains only the interest revenue that it requires to fund operations, and turns the rest over to the Treasury each fiscal year.³ The level of seigniorage remitted annually does not significantly affect debt: it amounts to slightly more than 1 percent of revenues in most years.⁴ The governments of most developed nations do not regularly rely on seigniorage as a funding strategy because overreliance on seigniorage—that is, on money creation—will inevitably lead to rising inflation. Perhaps the most famous example of printing money to fund government operations is Germany in the early 1920s, when the price level doubled every two days. This action is sometimes called “monetizing” government debt: if the market grows unwilling to purchase government debt at low rates, the central bank can step in to purchase that debt directly from the government. Stanley Fischer, Ratna Sahay, and Carlos Vegh (2002) estimate how much government revenue can be created through seigniorage from a sample of 24 countries in the post-World War II period. Those nations created enough money to push annual inflation above 100 percent. During those episodes, seigniorage amounted to just 4 percent of GDP on average—not enough to cover their average deficits of just below 5 percent of GDP. By comparison, deficits under

³ This revenue for the Treasury effectively is a tax on the public’s holdings of non-interest-bearing money—the currency and bank reserves issued by the Fed—since the public would have otherwise earned interest from holding those treasuries.

⁴ Since 2009, the Fed has produced a larger than average amount of seigniorage because the Fed has earned greater interest revenue due to the large expansion of the Fed’s balance sheet to treat the financial crisis. From 2001 through 2008, the Fed turned an average of \$26 billion over to the Treasury each fiscal year, averaging 1.1 percent of gross fiscal receipts. From 2009 through 2011, the Fed turned an average of \$67.9 billion over to the Treasury each year, or roughly 2.7 percent of gross fiscal receipts. Data for the Fed’s annual remissions to the Treasury can be found in the annual reports of the Federal Reserve Board of Governors, available on its website. Though the seigniorage revenue remitted to the Treasury has been larger in recent years due to the Fed’s increased interest income, partially offsetting that increased income is the fact that the Fed, as of 2008, pays banks interest for the reserves they hold. The Federal Reserve System paid \$3.8 billion to banks in 2011 in interest on reserves and term deposits.

the CBO's alternative scenario are projected to grow from a low of 5.6 percent of GDP in 2014 to more than 57 percent of GDP by 2085.

Aside from seigniorage, a central bank can reduce the government's debt burden by creating inflation that was not anticipated by financial markets. Inflation allows all borrowers, the government included, to repay loans issued in nominal terms with cheaper dollars than the ones they borrowed. In the United States, inflation tends to be low and predictable from year to year. Inflation that is higher than expected, and therefore not priced into the contract interest rate, tends to produce only a small transfer of wealth from lenders to borrowers. (Indeed, this is one strong rationale behind the Fed's price stability objective for monetary policy.) However, roughly 90 percent of the federal government's debt is issued in nominal terms at prices that reflect the market's expectations for inflation over the life of the loan. A significant deviation from those expectations would produce a larger transfer of wealth from lenders to borrowers. Historically, some central banks—though never the Federal Reserve—have even produced inflation for the sole purpose of eroding the value of the government's debt.

Today, the central banks of most developed nations operate independently of fiscal policy considerations, and none that the authors are aware of produce inflation for the explicit purpose of reducing government debt levels. Between low, stable inflation and minimal seigniorage revenue, the Federal Reserve's policies generally have little direct impact on the government's debt burden. (See Box 1 for an overview of other ways in which fiscal and monetary policies interact.) This could change, however, if financial markets began to view hitting the fiscal limit as a possibility. That situation would inevitably invite monetary policymakers to intervene since inflation presents one possible source of revenue. (See the Appendix for a discussion of ways in which this pressure could arise in a crisis.)

In fact, economic research suggests that high debt levels ultimately could overwhelm a central bank's efforts to keep prices stable. The remainder of this essay will argue that these outcomes should be avoided in the United States by putting fiscal policy on a sustainable path.

2. SOURCES OF FISCAL INFLATION

Even without direct political pressures on the central bank to create inflation, unsustainable fiscal policy may be able to force that outcome. Inflation is commonly argued to be “always and everywhere a monetary phenomenon,” a statement reflecting the monetarist notion that in the long run, inflation can be created only by the central bank's actions to increase the money supply. However, economists Thomas Sargent and

Neil Wallace (1981) show that the central bank may not have control over inflation in times of fiscal crisis. This stems from the idea that the public has a limited demand, based on its private portfolio preferences, to hold government debt as a percent of GDP. Sargent and Wallace model a scenario in which the government has reached that limit on debt, yet continues to run budget deficits. If the government is to avoid default, the central bank has no choice but to produce inflation to reduce debt levels and satisfy the intertemporal budget constraint. In this scenario, monetary policymakers uncharacteristically focus on stabilizing debt, while inflation is determined by deficit policy.⁵

Does this scenario resemble the way monetary and fiscal policies are conducted in the United States? In the Sargent and Wallace framework, fiscal authorities “move first” by choosing levels of debt and surpluses, leaving monetary policymakers to make up for any imbalance. However, the central bank may be able to constrain the actions of fiscal authorities by making the first move; that is, by firmly establishing the expectation among both fiscal authorities and market participants that it will not step in to reduce debt levels with inflation.⁶ One could argue that this is the way monetary policy is conducted in the United States, such that the inflationary outcome that Sargent and Wallace describe need not be a concern. Since the early 1980s, American monetary policy has tended to adjust interest rates fairly predictably in response to the performance of inflation and unemployment.

As a result of this consistent stance in opposition to inflation, financial markets view the Fed’s inflation objectives as highly credible, as evidenced by anchored inflation expectations. The same is true for the central banks of many other developed nations. Some central banks even face legally binding price stability mandates, such as the Bank of England, which must explain its failures to the Chancellor of the Exchequer, as well as the actions that are being taken to correct them. The credibility that these central banks have earned is bolstered by the operational independence most of them have been granted by their

⁵ Sargent and Wallace label this outcome the “unpleasant monetarist arithmetic” of chronic fiscal deficits. Variations of this model are presented by Eric Leeper (1991), Christopher Sims (1994), John Cochrane (1999), and Michael Woodford (2001), among others.

⁶ Eric Leeper (1991) describes this as an “active monetary policy/passive fiscal policy” framework. An active policy is one that chooses its objectives—surplus or deficit levels for fiscal policy, or money supply growth for monetary policy—as it sees fit, leaving the “passive” entity to stabilize debt. If monetary policy is “active,” it generally follows a policy that adjusts interest rates in response to inflation. When fiscal policy is active, it pursues the spending and tax policies it desires without necessarily stabilizing debt. If it chooses large debt levels, it will ultimately determine inflation as a result of Sargent and Wallace’s “unpleasant arithmetic.”

governments, which insulates monetary policy from pressure to set aside price stability to temporarily boost the economy.

In practice, however, a central bank's credibility cannot constrain fiscal policy in any meaningful sense: it cannot stop fiscal policymakers from running budget deficits that continually expand the debt. As a result, whether high debt levels would lead to inflation depends critically on whether the public believes fiscal authorities will balance the intertemporal budget constraint, or instead leave fiscal imbalances to be addressed by inflation. Unfortunately, neither theory nor experience provides a good rule of thumb for when those expectations might begin to change, potentially unleashing a fiscal crisis, though it is reasonable to expect that such a shift becomes more likely as projected debt levels grow ever larger. For example, Eric Leeper (2010) imagines a scenario in which the federal government is almost at its fiscal limit, but fiscal authorities still have some ability to adjust fiscal policy to stabilize debt levels. Being near the fiscal limit is enough to enable an equilibrium in which markets expect the central bank to accommodate the debt with inflation in the future. The public's expectation of higher inflation can push actual inflation higher before the central bank decides to create a single dollar.⁷

To emphasize the power of expectations in creating inflation, it is worth noting that a change in expectations also could bring an inflationary episode to a quick end. Sargent (1981) looked at the hyperinflations experienced by Austria, Hungary, Germany, and Poland after World War I. Each country financed massive government deficits and war reparations through sales of government debt to the central bank, resulting in hyperinflation. In each case, hyperinflation was brought to a sudden end through drastic regime changes in both fiscal and monetary policies: each nation established an independent central bank that was legally prohibited from extending credit to the government and established rules that limited fiscal policy to financing debt through private markets. In each case, the regime change credibly convinced market participants that the central bank would no longer finance fiscal policy.

The lesson from this literature is that when the public expects fiscal authorities to take action to satisfy the budget constraint while they still can, inflation need not rise. This is perhaps the situation in the United States today: debt projections under the CBO's more

⁷ This effect presents an outcome similar to the "unpleasant monetarist arithmetic"—that chronic fiscal deficits can lead to inflation—except that here inflation can arise even without monetary accommodation provided by the central bank. Accordingly, this branch of literature is called "the fiscal theory of the price level." Several of the references provided in footnote five follow this line of thinking.

likely scenario exceed historical records for most developed countries, yet markets appear perfectly willing to purchase government debt at low interest rates, indicating that inflation expectations remain low. Apparently markets believe fiscal imbalances will be resolved through fiscal policy rather than through inflation. However, as long as there is uncertainty over the feasibility of generating sufficient future surpluses, policymakers cannot be sure that market expectations will not shift unexpectedly and produce inflation. Leeper (2010) argues that a way to reduce that uncertainty would be to establish clear rules that govern fiscal policy in times of fiscal strain to avoid long-term imbalances, a topic discussed at the end of this essay. In the meantime, since uncertainty remains over how current fiscal imbalances would be resolved, it is useful to consider the options facing the central bank in an environment of fiscal crisis.

3. ENCOURAGING SUSTAINABLE POLICY

Credible monetary policy may help postpone the spike in inflation expectations that the above literature describes by convincing the public that the central bank will not quickly or easily agree to erode the debt through inflation. In many developing countries, central banks have a history of creating large amounts of inflation to help governments finance spending. For countries with that history, fiscal imbalances may more easily lead to a spike in inflation. Fortunately, the United States has no such history. The Fed can preserve its credibility by continuing to meet its price stability objectives, a task made more complicated in times of economic turbulence. In the past few years, weak economic conditions have greatly influenced the policies of the Fed and many other central banks, while inflation has perhaps been less of an immediate concern. It is useful to remember that the Fed's credibility helps make policies aimed at supporting real economic growth more effective. For example, markets remained confident in 2008 that the Fed would act to constrain any inflation pressures that emerged, even as the Fed added extraordinary liquidity to the banking system.

There are additional steps that can be taken to bolster the Fed's credibility. Elected leaders could reaffirm the central bank's independence to reassure markets that the Fed will not face political pressure to erode the debt through inflation, similar in spirit to the formal accord struck between the Fed and the Treasury Department in 1951. (See Appendix.) A formal target for inflation, like the one adopted by the Fed in early 2012, may strengthen the central bank's perceived commitment to avoiding inflation.

However, these steps may not be sufficient. As research by Sargent and Wallace and others describes, fiscal policy that does not contain the debt may lead to inflation even if monetary policymakers have the best intentions. This is due to the incontrovertible nature of the government's intertemporal budget constraint. When the expected path for fiscal policy does not by itself achieve balance in the constraint over time, the price level is the only other factor that can adjust to provide it.

It is useful to consider how much inflation would be required to adequately reduce current debt levels. The opening paragraphs of this essay noted that the historical peak of the U.S. debt-to-GDP ratio was reached after World War II. Counting only the portion of that debt that could easily be bought and sold in public markets, George Hall and Sargent (2011) estimate that it took 30 years for debt to fall from 97.2 to 16.9 as a percent of GDP. They estimate that about 20 percent of that debt reduction came from inflation. (Annual inflation, measured by the Personal Consumption Expenditures Price Index, averaged 3.2 percent over that time period.) To consider how much inflation would be required today to address current debt imbalances, Michael Krause and Stéphane Moyon (2011) estimate that a moderate rise in inflation to 4 percent annually sustained for at least 10 years—in effect a permanent doubling of the Fed's inflation objective—would reduce the value of the additional debt that accrued during the 2008–09 financial crisis, not the total debt, by just 25 percent. If the rise in inflation lasted only two or three years, a 16 percentage point increase—from roughly 2 percent inflation today to 18 percent—would be required to reduce that additional debt by just 3 percent to 8 percent. Such inflation rates were not reached even in the worst days of the inflationary 1970s. The reason inflation has such a minimal impact on debt in Krause and Moyon's estimates is that while inflation erodes the value of existing nominal debt, it increases the financing costs for newly issued debt because investors must be compensated to be willing to hold bonds that will be subject to higher inflation. This effect would be greater for governments such as the United States that have a short average maturity of government debt and therefore need to reissue it often.

With these estimates in mind, it is worth recalling the CBO's projection that debt held by the public may triple as a percent of GDP within 25 years. The estimates cited above suggest that inflation is simply not a viable strategy for reducing such debt levels. In addition, it is important to remember that inflation is costly on many levels. Inflation high enough to significantly erode the debt would inflict considerable damage on the economy and would require costly policies for the Fed to regain its credibility after the fact. Inflation that was

engineered specifically to erode debt would provide a significant source of fiscal revenue without approval via the democratic process, and so would raise questions about the role of the central bank as opposed to the roles of Congress and the executive branch in raising fiscal revenues.

Ultimately, the solution to high debt levels must come from fiscal authorities. Decades of monetary policy research suggests that rules and institutions can help ensure that central bankers take a long-run view of their policy objectives, even when doing so entails difficult or unpopular policy choices in the short term. Monetary policymakers have increasingly adopted transparent and consistent practices that make their policy rules credible and reduce uncertainty over their priorities.

The same rules-based institutions do not currently exist for fiscal policy. To a degree, this is a matter of necessity: the distributional nature of fiscal policy ought to be subject to the approval of the general public via the political process. However, it may be possible to create better rules for the more objective aspects of fiscal policy, a point argued by Leeper (2010). Just as Congress has agreed to set long-run objectives for the Fed while leaving day-to-day policy choices to independent monetary policymakers, fiscal policymakers could adopt objective long-run goals for fiscal policy—such as appropriate long-run targets for the ratio of debt to economic growth, guidelines for when unusual circumstances justify a large increase in debt, and how quickly fiscal imbalances should be resolved in that situation—while leaving the distributional details to the democratic process.

With that said, guaranteeing that policymakers will remain committed to those rules is difficult in practice.

The recent fiscal crisis in Europe provides telling proof. As a precondition to joining the European monetary union, 17 nations agreed to the Stability and Growth Pact, an agreement obligating each nation to maintain annual deficits of less than 3 percent of GDP and overall debt levels of less than 60 percent of GDP. Even the threat of sanctions for breaching this agreement was not enough to bind the fiscal policies of many European nations, including ones that have been the focus of the recent debt crisis and ones currently in relative fiscal health. If everyone knows that there are circumstances under which the rules will be violated—such as a demographic shift or an unprecedented financial crisis that calls upon national resources—then those rules will fail to anchor expectations. Though rules may be helpful, they may not be enough without some mechanism for enforcing them.

Despite the difficulties of establishing fiscal rules to reduce uncertainty over how fiscal imbalances would be resolved, there are encouraging examples from within the United States of fiscal policymakers

adopting a longer-term perspective. Before the Constitution was created, the federal government had no power to levy taxes without unanimous approval from the states. After a period in which both federal and state debt became significantly devalued, the fiscal regime was changed in 1790 by creating new powers for federal taxation and, as a quid pro quo, nationalizing state debt. This policy established an unfortunate precedent for relieving local governments of their debt burdens. Nearly 50 years later, the states again had incurred heavy debts and defaulted after the recession of the late 1830s. Creditors again looked to the federal government, but Congress rejected proposals to take on state debt, arguing that states had entered into debt of their own accord to finance local projects. The decision was costly to the federal government. Its reputation suffered because international creditors did not distinguish between state and federal debt, yet the decision forced states to rewrite their treatment of debt in their constitutions. Many adopted the balanced-budget amendments they retain today. Sargent (2011) describes this episode as an example of how fiscal crises can lead to positive institutional changes.

Ultimately, the solution to current fiscal imbalances will require our elected authorities to make difficult decisions. The Fed's best contribution to this process is to maintain its commitment to monetary policy objectives, including low and stable inflation. For the time being, markets appear to believe that fiscal policymakers will put future debt, spending, and tax levels on a more sustainable path. If they are correct, our nation will not have to experience the significant economic challenges of a world in which those expectations have changed.

APPENDIX

The following sections appeared as sidebars in the original article.

1. THE INTERACTION BETWEEN FISCAL POLICY AND MONETARY POLICY

Several of the everyday interactions between fiscal policy and monetary policy do not have a large effect on their respective goals to support a strong economy.

The most direct interaction in the United States is that monetary policy is conducted in the secondary market for U.S. Treasury securities. The Fed buys treasuries to put money into the banking system

when it wants to accommodate economic growth, and sells them to remove money and suppress inflation. The Fed does not exchange securities directly with the U.S. Treasury, but instead conducts transactions with private financial market participants, which avoids conflicts of interest that could otherwise arise from this relationship. The Fed also affects the government's borrowing costs when it raises interest rates in times of strong economic growth. Today the Fed's independence avoids pressure to make borrowing cheaper for the government, but this was not always the case.

More fundamentally, both fiscal policy and monetary policy affect the broader economy through the spending and investment decisions of households and businesses—though neither has a perfect ability to manage the economy in this way—and as a result their policies can affect each other's goals. (This, too, has led to political pressures throughout the Fed's history, as discussed in the sidebar.) So the Fed must consider the effects of current fiscal policy when it sets monetary policy to pursue its goals of price stability and healthy employment. For example, the Fed must consider how fiscal actions are likely to affect private demand based on how and when people expect those actions to be paid for by increased taxes or future expenditure reductions. Another possible effect of debt-financed fiscal stimulus—and another way in which fiscal and monetary policy interact—is that it could put upward pressure on interest rates in the economy as government borrowing rises.

Finally, as the main essay discusses, fiscal policy can have costly implications for monetary policy in times of fiscal crisis.

2. COULD THE FED'S MONETARY POLICY INDEPENDENCE WITHSTAND A FISCAL CRISIS?

On March 4, 1951, the Federal Reserve and the Treasury Department publicly agreed that the Fed would end its nine-year program in support of fiscal policy. Soon after the United States entered World War II, the Fed had committed to regularly purchasing enough Treasury debt to keep the government's financing costs low. The agreement to end that program became known as the Fed-Treasury accord, and it marked the end of an era of strong Treasury influence over monetary policy decisions, helping to usher in a new era of Fed independence. The accord asserted the Fed's authority to independently determine the size of the money supply to reach its congressionally established goals, which today include stable prices and healthy employment. This separation

of authority has been essential to keeping the Fed accountable while shielding monetary policy from short-term political influence.

The 1951 accord has not completely insulated the Fed from political intervention, however. Pressures on the Fed often have been motivated by a short-term interest in economic stimulus, but the Fed also has experienced pressures to place greater weight on price stability, including recently. Since the 1980s, despite occasional pressures, appreciation has grown both inside and outside of central banks for monetary policy independence as the best way to achieve both objectives.

The main essay points to research suggesting that fiscal imbalances can lead to inflation. This could occur most directly through explicit pressure from elected leaders to create inflation, but it also could stem from the central bank's desire to soothe an economy suffering from fiscal crisis.

It is useful to consider the conditions that likely would arise in fiscal crisis. The federal government would face two extreme choices: defaulting on its debt or enacting some combination of painful spending cuts and tax increases. The prospect of the first option would wreak havoc in financial markets as investors become concerned about the growing risk associated with U.S. Treasury securities. This effect has been demonstrated by the unfolding sovereign debt crisis in Europe. In early 2010, markets began to demand higher yields to hold debt issued by European governments that sustained large projected debt levels. The debt of some nations was downgraded by credit rating agencies, damaging the financial position of the many

European banks that hold large amounts of sovereign debt because the banks were then forced to raise more capital. A similar effect would arise in a U.S. fiscal crisis since Treasury securities are widely held by financial institutions and play an important role in many private market transactions as well. The European Central Bank responded by purchasing sovereign debt and also accepting that debt as collateral in loan agreements to banks. (The ECB's purchases were "sterilized," meaning that an equal amount in liquidity was removed from the financial system so that the purchases would not add to the overall money supply.)

The second option facing governments, a combination of sudden tax increases and broad cuts to services, could cause economic weakness in the short run. Independent of the possible short-run effects of fiscal "austerity," rational households and businesses are likely to hold back spending in anticipation of fiscal retrenchments even before such decisions are announced, particularly if there is uncertainty over the specific forms those adjustments would take. Without knowing whether payroll taxes will be higher in five years, a planned government

investment project will come to fruition, or employer health care costs will change abruptly, firms may delay a broad spectrum of spending, hiring, and investment decisions until those various sources of uncertainty have been resolved. In Europe, too, the uncertain resolution of fiscal imbalances has dampened spending and economic activity. Though monetary policy cannot resolve this type of uncertainty, it is clear that both default and extreme fiscal retrenchment may threaten the central bank's economic objectives.

That is why the dynamics of fiscal crisis can create difficult short-term tradeoffs for the central bank: the economic pain associated with fiscal crisis versus the longer-term costs of central bank intervention to reduce debt levels—including the risk of inflation, damaged central bank credibility, and a precedent for rescuing the government from its debt. At the same time, even the most conservative central banker might feel compelled to intervene in hopes of limiting a panic before it could grow more severe, despite the known costs of doing so. (A related discussion is presented by Jeffrey Lacker, 2011.)

Averting fiscal crisis entails making people believe that difficult fiscal policy choices will be made before they are forced by financial markets. Thus, creating that expectation may require fiscal constraint before it seems strictly necessary. Yet because of the difficult and unpopular tradeoffs required to achieve fiscal balance, it may be tempting for elected officials to delay action in hopes that monetary policy will relieve imbalances.

Experience since the 1951 accord and the prospects for how a fiscal crisis could unfold make clear the conditional nature of monetary policy independence. Extreme conditions could stress both the consensus in support of independence and the central bank's ability to act independently. While formal agreements like the accord can make overt political intervention in monetary policy more difficult, such "rules" cannot ensure that the central bank would escape difficult choices in times of crisis.

REFERENCES

- Cochrane, John H. 1999. "A Frictionless View of U.S. Inflation." In *NBER Macroeconomics Annual 1998*, edited by Ben S. Bernanke and Julio J. Rotemberg. Cambridge, Mass.: MIT Press.

- Congressional Budget Office. 2011. "2011 Long-Term Budget Outlook."
- Fischer, Stanley, Ratna Sahay, and Carlos A. Vegh. 2002. "Modern Hyper- and High Inflations." *Journal of Economic Literature* 40 (3): 837–80.
- Hall, George J., and Thomas J. Sargent. 2011. "Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics." *American Economic Journal: Macroeconomics* 3 (July): 192–214.
- Krause, Michael U., and Stéphane Moyen. 2011. "Public Debt and Changing Inflation Targets." Deutsche Bundesbank (April 29).
- Lacker, Jeffrey M. 2011. "Understanding the Interventionist Impulse of the Modern Central Bank." Speech to the CATO Institute 29th Annual Monetary Conference, Washington, D.C. (November 16).
- Leeper, Eric M. 1991. "Equilibria Under 'Active' and 'Passive' Monetary and Fiscal Policies." *Journal of Monetary Economics* 27: 129–47.
- Leeper, Eric M. 2010. "Monetary Science, Fiscal Alchemy." Paper presented at the Kansas City Fed Economic Policy Symposium at Jackson Hole (August).
- Malysheva, Nadezhda, and John R. Walter. 2010. "How Large Has the Federal Financial Safety Net Become?" Federal Reserve Bank of Richmond *Economic Quarterly* 96 (Third Quarter): 273–90.
- Sargent, Thomas J. 1981. "The Ends of Four Big Inflations." Federal Reserve Bank of Minneapolis Working Paper No. 158 (May).
- Sargent, Thomas J. 2011. "United States Then, Europe Now." Nobel Prize Lecture (December 8).
- Sargent, Thomas J., and Neil Wallace. 1981. "Some Unpleasant Monetarist Arithmetic." Federal Reserve Bank of Minneapolis *Quarterly Review* 5 (Fall).
- Sims, Christopher A. 1994. "A Simple Model for Study of the Determination of the Price Level and the Interaction of Monetary and Fiscal Policy." *Economic Theory* 4 (3): 381–399.
- Woodford, Michael. January 2001. "Fiscal Requirements for Price Stability." NBER Working Paper No. 8072.

Land of Opportunity: Economic Mobility in the United States

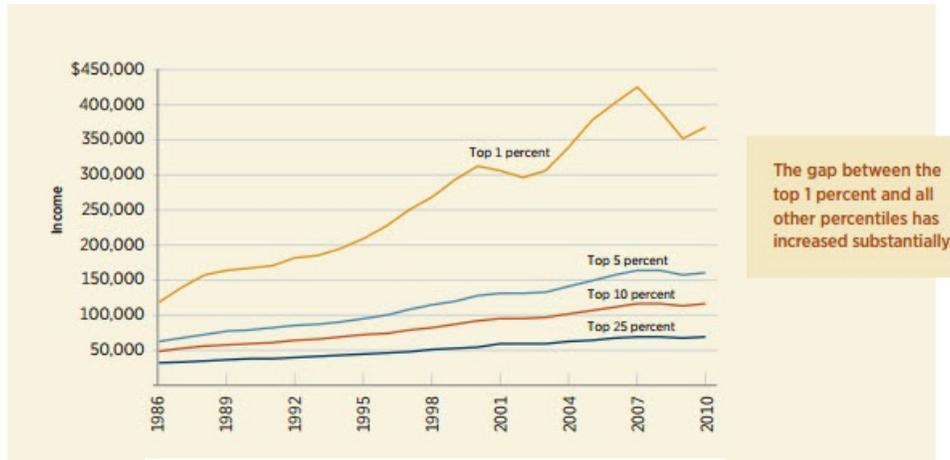
Kartik Athreya and Jessie Romero

The gap between people in the highest percentiles of earnings and wealth distributions and the rest of society has grown significantly during the past several decades, a fact that has led to considerable public discussion about the nature of opportunities available in the United States. Often overlooked in this debate, however, is the importance of economic mobility—the extent to which people are able to move up and down the income ladder—in determining what inequality implies for opportunity. If mobility is high, for example, the level of inequality at any point in time is not necessarily cause for concern, since it’s possible that today’s poor will be tomorrow’s rich. The potential for such upward mobility is the foundation of the American dream that has lured generations of immigrants to the United States.

The dream endures today. Nearly half of Americans aged 18–29 believe they will become rich at some point in their lifetimes, according to a 2012 Gallup Poll. But the odds are against them: In 2010 (the most recent year for which the Internal Revenue Service has published data), only about 5 percent of U.S. households earned more than \$150,000 per year, and about 1 percent earned more than \$350,000 per year. (See Figure 1). Most of those people, moreover, were not born to poor parents—especially not in recent years.

Understanding economic mobility is essential to understanding how observed levels and patterns of economic inequality relate to the implicit promise of American life. But this is complicated. Mobility and

■ The authors would like to thank Huberto Ennis, Arantxa Jarque, Marianna Kudlyak, Karl Rhodes, and Aaron Steelman for valuable discussions and insights. The views expressed are those of the authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

Figure 1 Thresholds for Selected Income Percentiles

Source: Internal Revenue Service.

Notes: An “income percentile threshold” is the lowest amount earned by a household in that percentile. Income is “adjusted gross income,” which is income minus certain deductions. Amounts are in current dollars.

inequality are determined jointly by random chance, by policy, and—most confounding of all for social scientists—by the deliberate actions of individuals or their parents. Regarding the latter determinant, it is clear that people differ according to their aptitude for various tasks, their appetite for risk, and their preferences for work versus leisure, among other characteristics. Both mobility and inequality thus will arise at least in part because different people make different choices. (See Appendix.)

This reality creates a challenge for economists seeking to understand the sources of observed levels of mobility and inequality, and for policymakers who hope to influence those levels. If everyone has the same opportunities for movement, then differences in income, wealth, or education must at least partially reflect deliberate choices and not market structure. This is not a setting in which many people would find efforts to alter outcomes via policy compelling. In contrast, to the extent that inequality continues across generations because people do not have the same chances, then inequality and immobility can be partially chalked up to market structure. From a normative standpoint, there

thus might be support for policy interventions that seek to equalize opportunities, rather than those that would equalize outcomes.

One such intervention is greater investment in early education. High-quality early-childhood education equips children with the skills they need to succeed at each subsequent stage of life, yet in the United States, access to such education appears to strongly depend on parents' income. Children of poor parents are thus at a disadvantage from the very beginning—a disadvantage from which it is very difficult to recover. But these children are not the only ones who are affected; all else equal, a more skilled workforce increases the productivity of society as a whole. Enhancing early education opportunities for the initially disadvantaged could therefore lead to better economic outcomes for everyone.

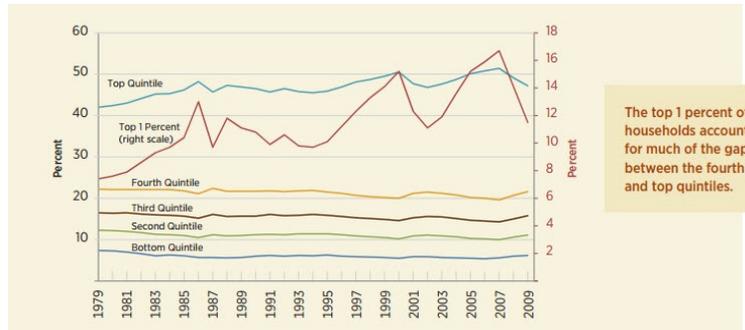
This essay will review both recent and longer-run features of U.S. economic mobility, with a focus on how those trends affect the interpretation of data on income inequality. It then will discuss some of the challenges and choices facing policymakers seeking to alter observed outcomes.

1. INEQUALITY IN THE UNITED STATES

By nearly any measure, income inequality in the United States is increasing.¹ In particular, today's rich are both richer than their counterparts in the past and richer relative to those around them. In 1979, the top 1 percent of households took home 7.4 percent of total after-tax income in the United States. By 2007, the share had more than doubled to 16.7 percent (Congressional Budget Office 2011).² At the same time, the share of income earned by households at all levels of the remaining distribution stayed flat or declined. Those in the middle three quintiles (fifths), for example, saw their share decrease from 51 percent to 43.9 percent. The picture looks the same for pretax income; the share accruing to the top 1 percent rose from 8.9 percent to 18.7

¹ Economists also study consumption inequality, or differences in the amounts of goods and services that households purchase. Consumption inequality might differ from income inequality because of savings, taxes, or in-kind benefits such as food stamps. Some recent research suggests consumption inequality is much less pronounced than income inequality (e.g., Meyer and Sullivan [2013]), although other research finds that the trends in income and consumption inequality are very similar (e.g., Aguiar and Bils [2011]).

² The CBO defines after-tax income as market income (labor income, business income, capital gains, capital income, and other income) plus government transfers (such as Social Security payments, unemployment benefits, or in-kind transfers such as food stamps) minus taxes paid.

Figure 2 Income Distribution by Quintiles

Source: Congressional Budget Office.

Notes: Quintiles are displayed on the left scale; the top 1 percent is displayed on the right scale. After-tax income is defined as market income (labor income, business income, capital gains, capital income, and other income) net of transfer payments and taxes.

percent (Congressional Budget Office 2011).³ These changes are a result both of increasing concentration of all types of income at the top of the distribution and a shift in the composition of income toward business income and capital gains (Congressional Budget Office 2011). This compositional change also makes incomes at the top of the distribution more volatile, but the trend is clearly one of growing inequality. (See Figure 2.)

Other research shows similar trends. Thomas Piketty and Emmanuel Saez (2003) find that after remaining flat throughout the 1950s and 1960s, the share of pretax income earned by the top 10 percent of households increased from 31.5 percent in 1970 to 41.4 percent in 1998.⁴ As in the CBO's analysis, this increase was largely driven by those at the very top of the distribution. While the income share for those in the 90th through 99th percentiles increased from 23.7 percent to 26.9 percent, the share for those in the very top percentile nearly doubled, from 7.8 percent to 14.6 percent.⁵

³ Data are from the supplemental data tables posted at www.cbo.gov/publication/43373.

⁴ In Piketty and Saez (2003), the unit of analysis is a tax unit, defined as two married people living together (with or without dependents) or a single adult (with or without dependents). Their income measure excludes capital gains.

⁵ Updated data are available at elsa.berkeley.edu/~saez/TabFig2011prel.xls.

The trend continued after the 2007–09 recession. Although average real income for the top 1 percent fell about three times more than for the remaining 99 percent, the decline was almost entirely due to the stock market crash. As markets recovered in 2010, incomes for the top 1 percent increased 11.6 percent, compared to only 0.2 percent for all other households (Saez 2013).

Income shares for the 90th–99th percentiles and the top 1 percent continued to increase, to 29.1 percent and 17.4 percent, respectively, in 2011 (Piketty and Saez 2003, updated data).

These data have garnered a great deal of attention from economists, policymakers, and the public, but do they shed light on what is actually happening to individuals or households?

2. MOBILITY: A CENTRAL FORCE BEHIND INEQUALITY

An observation of inequality at any point in time is only a snapshot; it does not shed light on how that snapshot developed. For example, imagine three different worlds: In the first world, the first inhabitants flip coins to determine not only their income, but also the income of all future generations; each descendant earns either \$1,000 or \$100,000 per year, depending on his ancestor's original coin toss. In the second world, the members of each new generation flip coins, but they do so just once at birth to determine whether they will earn \$1,000 or \$100,000 per year during their lifetimes. In the third world, individuals get to flip a coin each year to determine their income for that year.

The people in these worlds face very different lifetime risks. The first world, which is akin to a caste system, is very risky from the perspective of the first ancestor, who is determining outcomes for an entire dynasty. The second world also is risky since the die is cast for one person's entire life, but each of her descendants gets a chance to flip the coin, making it unlikely that bad luck will persist across many generations. The third environment is the least risky since it is very unlikely that an individual's average annual income over his lifetime would be significantly different than \$50,500, the average annual income he can expect over many years.

Despite these differences, snapshots of these economies in any given year look the same. In each, about half the population earns \$1,000 per year, while the other half earns \$100,000. Clearly, then, inequality data alone do not reveal the underlying prospects of individuals. For this, one must study economic mobility.

3. TRENDS IN ECONOMIC MOBILITY

Economists and policymakers generally are interested in two types of mobility: intragenerational and intergenerational. Intragenerational mobility describes how a given person's economic status changes over the course of his lifetime. Intergenerational mobility reflects the degree to which a person's economic status as an adult differs from that of her parents or ancestors. Status is usually measured by earnings (wage income), income (all sources of income, including wages), or less frequently wealth (the value of assets minus liabilities). Most research focuses on relative intra- and intergenerational mobility, or how a person's status changes in comparison to others. But it is also important to recognize that a person might experience absolute mobility even in the absence of relative mobility. She might occupy the same place in the earnings distribution as her parents, remaining in the same position relative to the rest of society, but still have a higher standard of living than her parents did, depending on the rate of economic growth.⁶

Intragenerational Earnings Mobility

Does the top of the income distribution comprise the same people year in and year out, or do individuals flow in and out of the highest percentiles over their lifetimes? If intragenerational mobility is high, then any snapshot of inequality will overstate the actual long-term inequality among individuals. For example, it is possible that the large gap in recent years between those in the top percentile and the rest of the distribution reflects an increase in the variation of annual earnings due to stock options and large bonuses. If that were the case, short-term inequality might be high, but long-term inequality could be much lower, reflecting high mobility.

In addition, in most modern societies, there is a clear life-cycle pattern to earnings and income. Imagine an extreme case where half the population earns \$1,000 during the first half of their lives and \$100,000 during the second half, while the other half of the population earns \$100,000 early in life and \$1,000 later. Income inequality would be high at a point in time, but everybody has the same lifetime income. Assuming that individuals could save and borrow to smooth their consumption over time, the snapshot of income inequality might not accurately reflect people's well-being since consumption inequality—a truer, and harder to measure, barometer—would be relatively low.

⁶ For example, see Easterlin (2000).

Anthony Shorrocks (1978) formalized these ideas by developing an index in which mobility is defined as the extent to which income inequality decreases over a given timeframe. Wojciech Kopczuk, Emmanuel Saez, and Jae Song (2010) calculate Shorrocks indices comparing inequality in annual earnings and in earnings averaged over five years for workers between 1937 and 2004. They find that short-term (five-year) mobility has not changed over the period, which implies that greater volatility of short-term earnings is not the source of observed higher inequality. Instead, higher inequality is likely the result of increased variation in lifetime earnings, including higher earnings at the top of the distribution. The authors conclude that mobility has not been sufficient to offset the rise in inequality, and thus that short-term inequality likely reflects lifetime inequality.

Kopczuk, Saez, and Song (2010) also find that long-term income mobility, from the beginning to the end of working life, actually increased significantly for all workers between 1942 and 1999. There is significant heterogeneity among groups of workers, however. Although on average men are more upwardly mobile than women, men's mobility was stable or declining during the sample period. Women's mobility, however, has increased greatly since the 1960s, as more women have moved into higher-paying professions. Thus, the increase in mobility for all workers has been driven by the labor market experiences of women.

Heterogeneity in intragenerational mobility also is apparent across the income distribution. Gerald Auten, Geoffrey Gee, and Nicholas Turner (2013) find that about 75 percent of taxpayers aged 35–40 who were in the second, third, or fourth quintile in 1987 were in a different quintile in 2007. (About 60 percent of those who changed position moved up or down a single quintile.) But they find greater persistence at the top and bottom of the distribution: 43 percent of taxpayers in the bottom quintile were still there 20 years later, and 46 percent of taxpayers in the top quintile maintained their positions. The authors also find that the very top earners tended to remain top earners: From 1992 through 2006, between 60 percent and 70 percent of the top 1 percent in a given year were in the top 1 percent in the following year.

Intergenerational Mobility

A commonly used measure of intergenerational mobility is the intergenerational elasticity of earnings (IGE). The IGE describes in percentage terms how much of the difference between the earnings of families in one generation persists into the next generation, typically by comparing the correlation of the earnings of fathers and sons. For example, an IGE of 0.5 means that a 10 percent difference between the income of

two fathers translates into a 5 percent difference in the income of their sons. The smaller the IGE, the greater the amount of mobility.

Important early studies of the United States and other developed countries found a high degree of mobility, with an IGE of 0.2 or less (Becker and Tomes 1986). Later research, however, found that data used in this work featured biases that would lead to artificially low measurements of the true level of earnings persistence. (See Stokey [1996] for a review of this research.)

New and better data suggest that mobility in the United States has been historically lower than initial estimates implied, and that it has declined even further in recent decades. Daniel Aaronson and Bhashkar Mazumder (2008) construct a time series of intergenerational elasticity from 1950 to 2000. They find that mobility increased between 1950 and 1980—the IGE decreased from 0.40 to 0.32—but decreased significantly during the 1980s and 1990s, with the IGE reaching 0.58 by 2000.

Although exact international comparisons are not possible, most research suggests that people in the United States are somewhat less mobile than people in Canada, Denmark, Finland, and Norway, where the IGE is about 0.15 to 0.2. In Germany and Switzerland, the IGE is about 0.3, and people in the United Kingdom and France also are relatively immobile, with IGEs of about 0.4 to 0.5 (Corak 2006).

While the IGE is a widely used statistic in work on intergenerational mobility, it only reflects average mobility across the entire distribution of individuals; it does not reveal anything about the direction of mobility or how it varies across different groups. To learn more about such mobility, Mazumder (2008) calculates transition rates, the likelihood of moving from one point in the distribution to another, across generations. He finds that, as with intragenerational measures, the amount of mobility varies significantly according to income. For example, there is a great deal of “stickiness” at the top and bottom of the distribution; people whose parents are in the bottom quintile of income are more likely to be in the bottom quintile themselves, and those whose parents are in the top quintile are likely to remain there. More than 60 percent of children whose parents are in the bottom quintile will end up in the bottom or second quintile, compared to 23.3 percent of those whose parents are in the top quintile. Only 7.4 percent of people who reach the top quintile are from families in the bottom quintile. (See Figure 3.) There also are stark differences between black people and white people and between men and women. Whites appear to be more upwardly mobile and less downwardly mobile than blacks. Mazumder (2008) finds that about 24.9 percent of whites remain in the bottom quintile, compared to 43.7 percent of blacks. And 38.9 percent of whites remain in the top quintile, compared to 21.3 percent of blacks. In addition, more

Figure 3 Intergenerational Income Quintile Transition Rates

Source: Mazumder (2008).

Notes: The figure shows what percentages of adolescents from families in a given income quintile remained in that quintile or transitioned to a different quintile as adults. For example, 33.5 percent of adolescents from families in the bottom quintile remained in the bottom quintile, while 26.9 percent moved to the second quintile. Income data were gathered from 1979 through 1980 and again from 1997 through 2003.

than twice as many whites as blacks experience the “rags-to-riches” scenario of moving from the bottom quintile to the top quintile, 10.6 percent compared to 4.1 percent. Mazumder also finds a large gender gap. While 40.5 percent of women from families in the lowest quintile remain there, only 27.2 percent of men do. Conversely, 43.0 percent of men from families in the top quintile remain in that quintile, compared to 31.9 percent of women. Men are thus more upwardly mobile and less downwardly mobile than women. The gender gap is trumped by the race gap, however: Both black men and black women tend to be the most likely to remain in the bottom quintile and the most likely to fall out of the top quintile.⁷

⁷ Isaacs (2008) finds similar differences in black and white mobility.

Mobility of Immigrants

For centuries, the American dream has drawn immigrants to the United States, from the waves of German and Irish immigrants in the late 1800s to the nearly 12 million Mexican immigrants who arrived during the past four decades.⁸ But how likely is it that the dream becomes a reality?

Decennial census data indicate that immigrants' earnings increase rapidly after they arrive in the United States; the earnings gap between them and their native-born peers appears to shrink substantially over time. Comparing natives and immigrants with similar work experience, Darren Lubotsky (2007) finds that the positive earnings gap between natives and the cohort of immigrants who came to the United States between 1965 and 1969 fell from 38 percent in the 1970 Census to 16 percent in the 1980 Census, and vanished by the 1990 Census. The gap between natives and immigrants who arrived in the late 1980s fell from 55 percent to 36 percent between the 1990 and 2000 censuses. This mobility might be spurious, however. Up to one-third of immigrants eventually return to their home countries; if these immigrants tend to be those with lower earnings, then the apparent earnings growth actually reflects fewer low earners in the data pool. Lubotsky (2007) corrects for this "selective out-migration" by studying longitudinal rather than cross-sectional data, and finds that earnings growth is significantly lower. In the cross-sectional data, immigrants' relative earnings increase 20 percent during their first decade in the United States and an additional 10 percent to 20 percent in each following decade. In the longitudinal data, however, immigrants' earnings grow between 12 percent and 15 percent during their first 15 years in the country and then stagnate.

The mobility of the second generation also appears to be decreasing. Throughout the 20th century, the children of immigrants not only earned more than their parents, but they also earned more on average than the rest of the non-immigrant population, perhaps reflecting some of the selection effects Lubotsky (2007) observed. But that advantage is shrinking. In 1940, the second generation earned 17.8 percent more than non-immigrants on average. In 1970, the difference was 14.6 percent, and by 2000, the difference had fallen to 6.3 percent (Borjas 2006). The reason might be a shift in the composition of immigrants. There has long been significant heterogeneity in earnings among immigrant

⁸ The number includes undocumented immigrants. Since the 2007–09 recession, net migration from Mexico has fallen to virtually zero. Between 2007 and 2011, the number of undocumented Mexican immigrants in the United States declined by about 1 million (Passel, Cohn, and Gonzalez-Barrera 2012).

groups, and in recent times, immigrants from developed countries tend to earn more than those from developing countries. Immigrants from Germany earned 24.9 percent more than non-immigrants in 1970 and their children earned 19.5 percent more in 2000, for example, while those from Mexico earned 31.6 percent less in 1970 and their children earned 14.6 percent less in 2000 (Borjas 2006).⁹ While wages in the second generation tend to regress toward the mean, overall earnings show significant persistence into the second generation. Borjas (2006) finds that across all immigrant groups, the intergenerational elasticity over the period 1970 to 2000 is 0.43. As the composition of immigrants increasingly shifts toward people from less-developed countries, who tend to have lower skills and levels of education, the wage gap is likely to persist through successive generations of immigrants (Haskins 2008).¹⁰ Irrespective of how quickly immigrants' earnings approach the earnings of natives, many immigrants still improve their economic status significantly by immigrating to the United States. In this sense, the move to the United States is a powerful form of economic mobility, and the United States' absorption of both legal and illegal immigrants makes it an engine of global mobility.

This last point must be part of any meaningful assessment of the mobility offered by a society. Even a calcified society, in which intergenerational or intragenerational mobility of natives is low, may be a source of mobility for the world's residents via its openness to immigrants. Conversely, societies that promote intergenerational mobility of natives through intensive early intervention and generous social safety nets but limit entry of immigrants—perhaps out of fear that they will exploit the generous safety nets—might hinder equality of opportunity in a global sense.¹¹

4. WHAT GENERATES PERSISTENCE?

The preceding discussion has highlighted empirical findings on the persistence of economic outcomes both within and across generations. But these findings do not explain why persistence across generations

⁹ Because the flow of immigrants from Mexico has been substantially greater than the flow from developed countries, the average wage of first-generation immigrants is still lower than the average wage of their native-born peers.

¹⁰ Immigrant mobility matters not only for the prospects of the immigrants themselves, but also for measured inequality in society as a whole. Imagine a room in which everyone is six feet tall. If a group of shorter people enter the room, measured inequality in height will increase. In the context of immigration, the arrival of a group with wealth, skills, or education significantly different from those of natives can mechanically increase inequality at a point in time.

¹¹ See, for example, Pritchett (2006).

exists in the first place or why it might have increased. As Aaronson and Mazumder (2008) note, intergenerational elasticities do not reflect causality. Instead, measures like the IGE are simply omnibus measures of everything correlated with parents' income and children's future earnings—factors ranging from the neighborhood where a child grew up to the availability of health care, among many others.

Intuitively, parents' decisions to invest in developing their children's skills, or "human capital," are important. Their willingness to make such investments stems in large part from altruistic concern for their children.¹² One model that incorporates this dynamic was created by Gary Solon (2004). He relates this investment decision to the rate of return to human capital and to the progressivity of public investment in children's human capital, such as government provision of education and health care. Solon's model suggests several things: that higher-income parents invest more in their children's human capital, that more progressive public investment in children's human capital partially crowds out parents' investment, and that parents are likely to invest more when the returns to human capital increase. The model predicts that intergenerational mobility will decrease during a period of increasing returns to human capital because rich parents are able to invest more than poor parents, and that mobility will increase during a period of more progressive public investment.

Recent trends in intergenerational mobility do correspond to Solon's predictions (Mazumder 2012). The returns to college education dropped during the 1940s, remained steady for several decades, and then began rising around 1980. These turning points in the returns to college education match the turning points in intergenerational elasticity observed in Aaronson and Mazumder (2008), as well as in other studies of mobility trends.

In Solon's (2004) model, the degree of progressivity of public education is exogenous—that is, determined outside the model. Andrea Ichino, Loukas Karabarbounis, and Enrico Moretti (2011) develop a model in which the degree of progressivity is the outcome of sociopolitical forces. In their model, public education is an insurance system that increases the future income of children without much innate talent at the expense of the future income of children with high innate talent. Public education thus increases mobility. But currently rich dynasties prefer low mobility for their descendants (as will be discussed in more detail in the following section), so in countries where rich dynasties are more politically active, spending on public education will be lower.

¹² For a thorough treatment, see Mulligan (1997).

In the United States, spending on public education mostly begins with kindergarten. But children face differences even before they begin school that may determine their future success. Mazumder (2008) finds that educational attainment alone is not enough to explain different mobility rates among black and white children. Black and white people who have completed the same number of years of school still have different intergenerational mobility rates, particularly at the level of high school completion and below. Other research also has found that educational attainment can explain less than half of the intergenerational transmission of earnings (Bowles, Gintis, and Groves 2008).

What this research implies is that human capital embodies more than the number of years spent in school. For example, adolescents who score higher on the Armed Forces Qualifying Test (AFQT) are more likely to move out of the bottom income quintile, and differences in AFQT scores can explain nearly all of the black/white mobility gap (Mazumder 2008).¹³ These test scores, however, capture much more than innate intelligence or academic achievement; non-cognitive skills such as work ethic, the ability to follow instructions, motivation, and patience also are essential to success on such standardized tests (Bowles, Gintis, and Groves 2008; Heckman 2008). In fact, these non-cognitive skills may be just as important as cognitive skills in determining future success in the labor market. For example, the General Educational Development (GED) credential is supposed to demonstrate cognitive equivalence between people who have graduated from high school and people who have dropped out and taken the GED exam instead. But GED holders have much poorer labor market outcomes than high school graduates despite obtaining equivalent knowledge. The reason, James Heckman and other economists have concluded, is that many students who earn a GED lack the non-cognitive skills that would have enabled them to complete high school—the same skills that would help them succeed in the labor market (Heckman, Humphries, and Mader 2010).

Recognizing the importance of non-cognitive skills begs an important question: How do children acquire these skills? A consensus now exists that the foundation is laid very early in life, even from infancy. Skill development is hierarchical; the early mastery of basic emotional, social, and other non-cognitive skills makes it easier to learn more complex cognitive skills throughout life. And children who fall behind early have difficulty catching up. Gaps in cognitive skills that are important

¹³ The AFQT is administered by the military to determine qualification for enlistment. AFQT scores have been widely used by economists as a measure of pre-labor market skills.

for adult outcomes are present as early as age 5 and tend to persist into adulthood (Heckman 2008).

The data suggest that poor and minority children are much more likely to fall behind. A recent report from the Brookings Institution (Sawhill, Winship, and Grannis 2012) examines the likelihood of achieving certain social and economic milestones on the path to the middle class, defined in the report as having a family income at least 300 percent of the poverty level, or about \$70,000 for a married couple with two children. Only 48 percent of children from families in the bottom income quintile are ready for school at age 5, compared to 78 percent of children from families in the top quintile.¹⁴ There also is a large disparity in early childhood outcomes according to race. Sixty-eight percent of white children are ready for school at age 5, versus only 56 percent of black children and 61 percent of Hispanic children. The gap between white and black widens throughout the lifespan. By age 11, 73 percent of white children versus 52 percent of black children have basic reading and math skills. By age 29, only 33 percent of black people have successfully transitioned to adulthood (defined by the authors as living independently and having either a college degree or a family income at least 250 percent of the poverty level), while 68 percent of white people reach this milestone. Hispanic people fare somewhat better; 66 percent achieve the age-11 milestone, and 47 percent reach the age-29 milestone.

5. CHALLENGES FOR POLICYMAKERS

What is the role for public policy, if any, in addressing economic inequality and mobility? Answering this question requires asking several others: What would policy try to achieve, and in particular, whose well-being would it attempt to enhance? Would the goal be to improve opportunities for current cohorts or for future generations? Would policy treat individuals at different moments in time as discrete units, irrespective of their ancestors, or would it emphasize dynasties by taking into account how family members invest in descendants?

From a policymaker's point of view, mobility might be inadequate as a measure of what a good society should provide its members. First of all, there is a tradeoff between mobility and predictability. Recall the imaginary world resembling a caste system described earlier. This setting is utterly immobile and risky for each dynasty's first member. But it is perfectly safe for the members of each successive generation

¹⁴The authors define "school-ready" as having acceptable pre-reading and math skills and behavior that is generally school-appropriate.

since income is completely stable. In fact, for a person whose ancestor flipped the \$100,000 coin, this world is not only safe, but also quite comfortable. On the macro level, it is possible that the costs of large fluctuations and risky income patterns outweigh the benefits of high mobility and reduced inequality. Peter Gottschalk and Enrico Spolaore (2002) study a model in which there are large welfare gains from greater mobility if aversion to inequality is the only consideration. But if aversion to income fluctuations is considered, those gains disappear. Of course, this might not be of great consolation to a person whose ancestor flipped the \$1,000 coin.

In addition, a world in which mobility is high is one where parents are of little consequence, despite their desire or ability to position their children and grandchildren for future success. Few parents would want to live in a world where their investments in their children have no influence beyond their lifetimes. The flip side is that descendants of people who were not altruistic or who made poor decisions would not be as constrained by their ancestors' actions.

Viewed in this light, what most people might agree on is trying to promote individual productivity while limiting downward mobility. Broadly speaking, the former goal involves ensuring preparedness at labor market entry, while the latter involves insuring households against low innate abilities, poor health, or job loss. Knowing the extent to which these forces matter is crucial for policy interventions to be effective. For example, if workers were similarly prepared at the time of entry into the labor market, and shocks in working life were important, the question would be how, if at all, to better insure workers, and not how to alter educational investment decisions. Conversely, if preparedness differed and shocks during working life were unimportant, further insuring workers would yield little benefit. Instead, changes to the educational system would be more effective.

Both factors are important, according to a recent line of work exemplified by Mark Huggett, Gustavo Ventura, and Amir Yaron (2011). They find that about 60 percent of the observed disparity in lifetime earnings is due to individual differences that exist before people enter the labor market, and the remainder is due to shocks that buffet them as they work, such as job losses. Their research stresses that the observed evolution of earnings inequality over lifetimes is consistent with a simple setting in which all workers accumulate skills through experience and effort, but do so at substantially different rates that reflect their initial "learning" ability. At the same time, their estimates clearly indicate that a substantial portion of inequality is generated during working life. This suggests that shocks to earnings are essential to a successful theory of earnings dispersion in the economy.

A critical point here is that the disparity in learning ability likely arises not only from differences in innate ability, but also from forces such as the quality of K-12 education and parental and cultural influences. These forces are very different for children from poor versus rich families—a dynamic that is magnified by a labor market that demands increasing levels of skill.

6. INVESTING IN HUMAN CAPITAL

For most people—all but a lucky few—labor is what they can sell to generate income. They can increase the value of their labor by acquiring greater skills, but the value of their labor is only partially under their control. It also depends on the supply and demand for their skills in the marketplace.

The industrial revolution, for example, created factories that made workers more productive and more valuable without substantially increasing their skills. But the information revolution has created a marketplace that rewards personally acquired skills, such as computer programming or mathematical analysis. In this new environment, an individual's innate ability and early life education become critical because they largely determine the levels of skills each person can develop to “rent” to the marketplace.

Given the large earnings gap between workers with and without college degrees, many policies aim to increase college access, for example by increasing federal subsidies for student loans. But it's not clear that college is the best focus for policymakers. The observed disparity between high school and college graduates applies to students who have graduated from college already; those students who have not yet enrolled might not necessarily receive the same benefit, perhaps because they are not as well prepared. For example, Lutz Hendricks and Oksana Leukhina (2012) find in preliminary work that about 70 percent of the lifetime earnings gap between high school and college graduates results from ability selection rather than from attaining the college degree per se. In other words, the college graduates were likely to be better earners even before entering college.

Intervening well before college could yield much higher returns. As noted above, the skills learned early in life prepare children to obtain more complex skills later in life. Heckman and many other researchers have found that the return on a dollar invested in human capital is highest when the investment occurs at age 3, and that children who receive high quality early education fare much better on a variety of socioeconomic measures (Heckman 2008).

The most cost-effective policy for increasing equality of opportunity is thus likely to be one that shifts funding away from universal college subsidies and toward early childhood interventions. Elizabeth Caucutt and Krishna Kumar (2003) find that a large increase in college subsidies with the goal of reducing the “enrollment gap” leads to very inefficient use of education resources, with little or no welfare gain, because more poorly prepared students enroll and the dropout rate increases. In a model of human capital transmission in which parents invest in their children, Diego Restuccia and Carlos Urrutia (2004) find that subsidies for investment in early education are much more effective at mitigating persistence in earnings than subsidies for college.

Investments in early childhood education can be viewed as a form of insurance against the risk of being born to poor parents, among other things. And while the public provision of such insurance could yield a big “bang for the buck” by enabling current generations to invest more in the education of future generations, one must also acknowledge the potential for moral hazard. A public system that equalizes the educational opportunities (or far more ambitiously, the home environments) of poor and rich children could reduce the incentives of all parents to invest in children.¹⁵

Greater public investment in early childhood education cannot replace the advantages that some parents are able to bestow upon their children, nor can it guarantee that all children will grow up to be prosperous. But such investments could give more children the necessary foundation for future acquisition of skills, and ensure that large amounts of human capital are not foregone simply because many children are born to poor families. This foregone human capital is a loss not only for the child, but also for society as a whole. According to an influential line of research, long-run economic growth depends on the amount of human capital in a society.¹⁶ Unlike physical capital, which exhibits decreasing returns to scale, human capital might well exhibit increasing returns. Knowledge leads to new ideas and new technologies, which lead to higher productivity, thus raising per capita income and living standards for society as a whole.

As this essay has discussed, economic inequality has increased significantly in the United States in recent years. At the same time, data suggest that economic mobility also has decreased, particularly for those born at the top and the bottom of the income distribution.

¹⁵ See Chang and Kim (2012) and Seshadri and Yuki (2004) for more on the “price of egalitarianism.”

¹⁶ Influential papers on “endogenous growth theory” include Romer (1986) and Lucas (1988).

Many factors contribute to the attainment and persistence of economic status, including innate ability, preferences for present versus future rewards, aversion to risk, and quite a bit of luck. But for nearly all people, advancement depends critically on opportunities to obtain human capital—and those opportunities are not the same for children born to poor versus rich families. Policies that aim to equalize these opportunities, particularly very early in life, appear to yield a very high return on investment, although much remains to be learned about the feasibility of implementing such interventions on a large scale. Nonetheless, such efforts have the potential to help the United States achieve a more inclusive prosperity.

APPENDIX

The following section appeared as a sidebar in the original article.

1. THE ROLE OF CHOICE

Inequality and immobility partially reflect deliberate choices related to the fact that people differ in their tolerance for risk or in their willingness to defer gratification (what economists call “time discounting”). But these differences cannot be directly observed. Instead, economists must make inferences based on actual outcomes, such as occupational choice, savings, and consumption.

Risk tolerance has a large impact on occupational choice, and thus on income and wealth. Beginning with Frank Knight’s *Risk, Uncertainty, and Profit* (1921) and continuing in modern work since Richard Kihlstrom and Jean-Jacques Laffont (1979), economists have modeled entrepreneurs as less risk averse than other people and therefore more likely to undertake high-risk/high-return enterprises. To the extent that people genuinely vary in risk aversion, this model suggests that the rich and the poor disproportionately will be those with high risk tolerance, while those in the middle will be more risk averse. This is consistent with data that show a disproportionate number of self-employed people at both ends of the earnings and wealth spectrums. They also figure more prominently among households in financial distress (Sullivan, Warren, and Westbrook 2000).

Additional evidence for the role of risk tolerance in personal economic outcomes comes from Sam Schulhofer-Wohl (2011), who finds that risk-tolerant workers tend to have jobs more exposed to

economy-wide or “aggregate” risk. Movements in these workers’ incomes thus tend to be more volatile even when they have insured themselves against individual-level, or “idiosyncratic,” risks, such as job loss or illness. As a result, volatility in their consumption of goods and services is not necessarily evidence of poor insurance possibilities in the marketplace. Indeed, Schulhofer-Wohl (2012) finds that after correcting for this bias, U.S. households do not appear to be bearing any significant uninsurable risk. (A variety of other research, however, has found that certain types of shocks, such as a long-term disability, are clearly not fully insured.)

Observed inequality also might reflect different preferences for consumption in the present versus the future. Per Krusell and Anthony Smith (1998) show, for example, that a model that includes variation in “impatience,” or the willingness of households to borrow against future earnings, successfully matches observed wealth inequality in the U.S. population. Emily Lawrance (1991) and Marco Cagetti (2003) also find that data on consumption and wealth suggest the presence of significant differences in preferences, especially in risk-aversion and time discounting. They find that less-skilled and less-wealthy individuals generally are less patient—meaning they place a higher value on current versus future consumption—than their more-skilled and wealthier counterparts. More recently, Lutz Hendricks (2007) has measured the extent of differences in households’ discount factor by noticing that households vary a great deal in their wealth even though they have and can expect to have very similar lifetime incomes.

Taken as a whole, economists’ work suggests that many of the observed differences in the way households make decisions can be understood as arising from differences in risk tolerance or time discounting. A caveat, however, is that a variety of difficult-to-model environmental forces might play a large role in generating these differences. In a society with low life expectancy or a high violent crime rate, for example, individuals might not be “choosing” to be impatient so much as making a rational decision to value current over future consumption. Likewise, not attending college might indicate an individual with a high discount factor who chose not to invest in K-12 education—or it might indicate a person facing strong institutional barriers to attending college. It is important to keep such environmental factors in mind when interpreting any model that includes heterogeneity in preferences.

REFERENCES

- Aaronson, Daniel, and Bhashkar Mazumder. Winter 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources* 43 (1): 139–72.
- Aguiar, Mark A., and Mark Bils. February 2011. "Has Consumption Inequality Mirrored Income Inequality?" National Bureau of Economic Research Working Paper No. 16807.
- Auten, Gerald, Geoffrey Gee, and Nicholas Turner. January 4, 2013. "Income Inequality, Mobility and Turnover at the Top in the United States, 1987–2010." Paper presented at the Allied Social Science Associations Annual Meeting, San Diego.
- Becker, Gary S., and Nigel Tomes. July 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4 (3) Part 2: S1–39.
- Borjas, George J. Fall 2006. "Making It in America: Social Mobility in the Immigrant Population." *Future of Children* 16 (2): 55–71.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne Groves. 2008. "Intergenerational Inequality Matters." In *Unequal Chances*, edited by Samuel Bowles, Herbert Gintis and Melissa Osborne Groves, 1–22. Princeton, N.J.: Princeton University Press.
- Cagetti, Marco. July 2003. "Wealth Accumulation Over the Life Cycle and Precautionary Savings." *Journal of Business and Economic Statistics* 21 (3): 339–53.
- Caucutt, Elizabeth M., and Krishna B. Kumar. 2003. "Higher Education Subsidies and Heterogeneity: A Dynamic Analysis." *Journal of Economic Dynamics and Control* 27: 1,459–502.
- Chang, Yongsung, and Sun-Bin Kim. January 2012. "The Price of Egalitarianism." *B.E. Journal of Macroeconomics* 11 (1).
- Congressional Budget Office. July 2012. "The Distribution of Household Income and Federal Taxes, 2008 and 2009."
- Congressional Budget Office. October 2011. "Trends in the Distribution of Household Income Between 1979 and 2007."
- Corak, Miles. 2006. "Do Poor Children Become Poor Adults? Lessons from a Cross-Country Comparison of Generational Earnings Mobility." In *Research on Economic Inequality*, Vol. 13, edited by John Creedy and Guyonne Kalb, 143–88. Bingley, U.K.: Emerald Group Publishing.

- Easterlin, Richard A. Winter 2000. "The Worldwide Standard of Living since 1800." *Journal of Economic Perspectives* 14 (1): 7–26.
- Gottschalk, Peter, and Enrico Spolaore. 2002. "On the Evaluation of Economic Mobility." *Review of Economic Studies* 69 (1): 191–208.
- Haskins, Ron. 2008. "Immigration: Wages, Education, and Mobility." In *Getting Ahead or Losing Ground: Economic Mobility in America*, edited by Julia B. Isaacs, Isabel V. Sawhill, and Ron Haskins, 81–90. Washington, D.C.: The Pew Charitable Trusts.
- Heckman, James J. July 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46 (3): 289–324.
- Heckman, James J., Lance J. Lochner, and Petra E. Todd. February 2008. "Earnings Functions and Rates of Return." National Bureau of Economic Research Working Paper No. 13780.
- Heckman, James J., John Eric Humphries, and Nicholas S. Mader. June 2010. "The GED." National Bureau of Economic Research Working Paper No. 16064.
- Hendricks, Lutz. September 2007. "How Important Is Discount Rate Heterogeneity for Wealth Inequality?" *Journal of Economic Dynamics & Control* 31(9): 3,042–68.
- Hendricks, Lutz, and Oksana Leukhina. May 2012. "The Return to College: Selection Bias and Dropout Risk." Manuscript, University of North Carolina, Chapel Hill, and University of Washington.
- Huggett, Mark, Gustavo Ventura, and Amir Yaron. December 2011. "Sources of Lifetime Inequality." *American Economic Review* 101 (7): 2,923–54.
- Ichino, Andrea, Loukas Karabarbounis, and Enrico Moretti. January 2011. "The Political Economy of Intergenerational Income Mobility." *Economic Inquiry* 49 (1): 47–69.
- Isaacs, Julia B. 2008. "Economic Mobility of Black and White Families." In *Getting Ahead or Losing Ground: Economic Mobility in America*, edited by Julia B. Isaacs, Isabel V. Sawhill and Ron Haskins, 71–80. Washington, D.C.: The Pew Charitable Trusts.
- Kihlstrom, Richard E., and Jean-Jacques Laffont. August 1979. "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion." *Journal of Political Economy* 87 (4): 719–48.
- Knight, Frank H. 1921. *Risk, Uncertainty, and Profit*. Hart, Schaffner, and Marx Prize Essays, no. 31. Boston: Houghton Mifflin.

- Kopczuk, Wojciech, Emmanuel Saez, and Jae Song. February 2010. "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data since 1937." *Quarterly Journal of Economics* 125 (1): 91–128.
- Krusell, Per, and Anthony A. Smith, Jr. October 1998. "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy* 106 (5): 867–96.
- Lawrance, Emily C. February 1991. "Poverty and the Rate of Time Preference: Evidence from Panel Data." *Journal of Political Economy* 99 (1): 54–77.
- Lubotsky, Darren H. October 2007. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy* 115 (5): 820–67.
- Lucas, Robert E. Jr. July 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (1): 3–42.
- Mazumder, Bhashkar. May 2008. "Upward Intergenerational Economic Mobility in the United States." Economic Mobility Project, The Pew Charitable Trusts.
- Mazumder, Bhashkar. April 2012. "Is Intergenerational Mobility Lower Now than in the Past?" Federal Reserve Bank of Chicago *Chicago Fed Letter*, no. 297.
- Meyer, Bruce D., and James X. Sullivan. January 4, 2013. "Consumption and Income Inequality in the U.S. since the 1960s." Paper presented at the Allied Social Science Associations Annual Meeting, San Diego.
- Mulligan, Casey B. 1997. *Parental Priorities and Economic Inequality*. Chicago: University of Chicago Press.
- Passel, Jeffrey, D'Vera Cohn, and Ana Gonzalez-Barrera. April 2012. "Net Migration from Mexico Falls to Zero—and Perhaps Less." Pew Research Hispanic Center.
- Piketty, Thomas, and Emmanuel Saez. February 2003. "Income Inequality in the United States, 1913–1998." *Quarterly Journal of Economics* 118 (1): 1–39.
- Pritchett, Lant. 2006. *Let Their People Come: Breaking the Gridlock on Global Labor Mobility*. Washington, D.C.: Center for Global Development.
- Restuccia, Diego, and Carlos Urrutia. December 2004. "Intergenerational Persistence of Earnings: The Role of Early and College Education." *American Economic Review* 94 (5): 1,354–78.

- Romer, Paul M. October 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94 (5): 1,002–37.
- Saez, Emmanuel. January 2013. "Striking It Richer: The Evolution of Top Incomes in the United States (Updated with 2011 Estimates)." Manuscript.
- Sawhill, Isabel V., Scott Winship, and Kerry Searle Grannis. September 2012. "Pathways to the Middle Class: Balancing Personal and Public Responsibilities." Brookings Institution Center on Children and Families.
- Schulhofer-Wohl, Sam. October 2011. "Heterogeneity and Tests of Risk Sharing." *Journal of Political Economy* 119 (5): 925–58.
- Seshadri, Ananth, and Kazuhiro Yuki. October 2004. "Equity and Efficiency Effects of Redistributive Policies." *Journal of Monetary Economics* 51 (7): 1,415–47.
- Shorrocks, Anthony. December 1978. "Income Inequality and Income Mobility." *Journal of Economic Theory* 19 (2): 376–93.
- Solon, Gary. 2004. "A Model of Intergenerational Mobility Variation over Time and Place." In *Generational Income Mobility in North American and Europe*, edited by Miles Corak, 38–47. Cambridge, England: Cambridge University Press.
- Stokey, Nancy L. 1996. "Shirtsleeves to Shirtsleeves: The Economics of Social Mobility." In *Frontiers of Research in Economic Theory: The Nancy L. Schwartz Memorial Lectures, 1983-1997*, edited by Donald P. Jacobs, Ehud Kalai and Morton I. Kamien, 210–41. Cambridge, England: Cambridge University Press.
- Sullivan, Teresa A., Elizabeth Warren, and Jay Lawrence Westbrook. 2000. *The Fragile Middle Class: Americans in Debt*. New Haven and London: Yale University Press.