Economic Insights

Exploring the Economic Effects of the Opioid Epidemic

Regional Spotlight

Implementing Monetary Policy in a Changing Federal Funds Market
Exploring the Economic Effects of the Opioid Epidemic
Adam Scavette examines how the opioid crisis affects the labor market, and why it’s been particularly hard on the Third District.

Regional Spotlight: Smart Growth for Regions of All Sizes
To build a healthy economy, a region must grow—or so says the conventional wisdom. Paul R. Flora examines the data and finds that growth isn’t always the best way forward.

Implementing Monetary Policy in a Changing Federal Funds Market
Benjamin Lester explores how the Fed set interest rates before, during, and after the Great Recession, and how the distribution of reserves might help it understand when those reserves are no longer ‘ample.’

Research Update
Abstracts of the latest working papers produced by the Philadelphia Fed.

About the Cover
Philadelphia’s most famous citizen, Benjamin Franklin, has graced the $100 bill since the newly created Federal Reserve began issuing “Federal Reserve Notes” in 1914. This particular image is taken from H.B. Hall’s engraving of Joseph-Siffred Duplessis’s 1785 portrait of Franklin, which is currently on view at the National Portrait Gallery in Washington, D.C. In the background are details from the 2009 redesign of the $100 bill, including a reproduction of the Declaration of Independence. Franklin served on the “Committee of Five” that drafted the Declaration and presented it to the Second Continental Congress, then meeting at the Pennsylvania State House, on July 4, 1776. The State House still stands today, just two blocks from the Federal Reserve Bank of Philadelphia, and is now known as Independence Hall.

Photo by Rich Wood.
Exploring the Economic Effects of the Opioid Epidemic

Hundreds of thousands of Americans have died from opioid overdoses in recent years. What has this epidemic done to the economy? And why is the crisis so much worse right here in the Third District?

BY ADAM SCAVETTE

In a single weekend in July 2018, more than 170 people in Philadelphia overdosed from what investigators said was a single “bad batch” of heroin. Ten died. Bags containing this particularly harmful compound were ominously stamped Santa Muerte (Spanish for “Holy Death”). In the preceding year, there were more than 70,200 drug overdose deaths in the United States, over four times the number of homicide deaths.1 The age-adjusted rate of drug overdose deaths tripled between 1999 and 2016 and jumped an additional 10 percent in 2017.2 Nearly 70 percent of those 2017 drug overdose deaths can be attributed to opioids.3

All three states in the Federal Reserve’s Third District have been struck particularly hard by this surge in drug overdose deaths. The age-adjusted drug overdose rate for the nation was 21.7 deaths per 100,000 people in 2017, while it was 30.0 for New Jersey, 37.0 for Delaware, and 44.3 for Pennsylvania. Only West Virginia and Ohio had drug overdose death rates higher than Pennsylvania’s in 2017, at 57.8 and 46.3, respectively. However, in terms of the absolute number of drug overdose deaths, Pennsylvania was first in the country at 5,388. That was 8 percent of the nation’s drug overdose deaths, even though Pennsylvania had less than 4 percent of its population. And as is true in the rest of the United States, the majority of these Pennsylvania deaths were opioid related.

This article examines the origins of the crisis, the nature of the crisis in the Third District, the relationship between the crisis and the labor market, the costs of the epidemic, and some policy countermeasures designed to alleviate the crisis.

What Is an Opioid?
An opioid is a substance that acts on the opioid receptors in the nervous system. Among other effects, opioids relieve pain and, when abused, produce euphoria. The Centers for Disease Control and Prevention (CDC) groups opioids into three primary categories: natural and semisynthetic opioid analgesics that are typically available by prescription (such as morphine, codeine, oxycodone, and hydrocodone); synthetic opioid analgesics (such as tramadol and fentanyl); and heroin.

The Origins of the Opioid Crisis in the United States
Studies suggest that a large percentage of abusers began their journey with prescription opioids. In one study, 80 percent of heroin users admitted to misusing prescription opioids before turning to heroin. Because users must obtain these prescription drugs either illicitly through diversion (the illegal transfer of opioids from the prescribed individual to others) or legally from a legitimate prescription, it is helpful to examine the rise in legitimate prescriptions for opioids.

Prior to the 1990s, opioids were prescribed mainly for cancer patients (or to treat chronic malignant pain). Beginning in the early 1990s, pharmaceutical companies encouraged physicians to prescribe opioids to treat noncancer pain.4 Noncancer patients tend to require longer-term administration of the drug than do cancer patients. As a result, practitioners of pain medicine as well as other medical specialties were taught to rely more on opioids for general pain treatment.5 At the time, the public didn’t realize how risky this would be. Later studies would show that noncancer pain patients are more likely than cancer

---

Adam Scavette is a senior economic analyst at the Federal Reserve Bank of Philadelphia. The views expressed in this article are not necessarily those of the Federal Reserve.
patients to become dependent on and eventually abuse opioids.

By decade’s end, 86 percent of patients using opioids were using them for non-cancer pain. The national prescribing rate of opioid medications climbed throughout the 2000s, peaking in 2012 (see Figure 2). While the prescribing rate in Pennsylvania roughly mirrored the nation’s, Delaware’s rate was typically 20 percent higher than the nation’s, and New Jersey’s prescribing rate was 25 percent lower. Thanks to increasing awareness and regulatory countermeasures to be discussed later in this article, by 2017 the rate of opioid prescriptions had fallen more than 25 percent from its peak in 2012.

As they become more dependent, abusers of prescription opioids often turn to heroin, which has experienced a large supply increase and thus a price decrease over the past 20 years due to major changes in its production and supply chain. According to recent Drug Enforcement Agency (DEA) estimates, from 1980 to 2016 the average retail price per gram of pure heroin decreased by more than 70 percent, from $3,260 (in 2012 dollars) to between $465 and $1,020. An even more recent trend has been a spike in the availability of fentanyl in the United States. Fentanyl is a synthetic painkiller about 50-100 times more potent than morphine. The drug is commonly prescribed in the form of transdermal patches or lozenges. In the United States it is available as a Schedule II drug, meaning it is legally available only through a non-refillable prescription. However, drug cartels began purchasing cheaply produced fentanyl from Chinese pharmaceutical labs and shipping it to Mexico to mix with the heroin supply before it enters the United States for sale. Because small doses of fentanyl are more likely to be fatal due to its potency, users are at a much higher risk of overdosing when abusing fentanyl or fentanyl-laced heroin. This has escalated the number of overdose deaths in the United States for which fentanyl is responsible (see Figure 4).

Natural-opioid overdose deaths, which can be attributed primarily to prescription pills, have increased at a steady rate since 1999. But starting in 2010, heroin and synthetic opioid overdose deaths have increased much more rapidly. As seen in the synthetic opioid series in Figure 4, fentanyl has become responsible for almost as many overdose deaths as natural opioids and heroin combined and has increased nearly tenfold since 2013.

The ‘Synthetic Problem’ in the Third District
An October 2018 New York Times feature article about the local and social effects of the crisis focused on the Philadelphia neighborhood of Kensington, which has long been recognized as the highly visible regional epicenter of the opioid epidemic due to its open-air drug markets, encampments of drug-addicted homeless users, and hyper-localized poverty. To address the crisis, Philadelphia Mayor Jim Kenney established an opioid task force upon his inauguration in 2016, and Governor Tom Wolf signed a statewide disaster declaration in 2018, an unprecedented public-health emergency measure in
Pennsylvania. We will take a closer look at public policy responses in a later section, but it is helpful to examine the data on our District to see how the local crisis became so deadly in the past few years.

Looking at opioid overdose death rates specifically, we see how quickly the opioid crisis worsened in our region. Figure 5 depicts opioid overdose death rates in the United States and the Third District states. Although Pennsylvania’s and New Jersey’s opioid overdose death rates were similar to the nation’s in 2015, by 2017 their rates had more than doubled, rising twice as fast as the national rate in two years. Meanwhile, in Delaware the opioid overdose death rate stood at nearly double the national rate in 2017.

Figure 6 breaks down these 2017 overdose death rates by type of opioid. Nationally, synthetic opioids (e.g., fentanyl) caused overdose deaths at a higher rate than other types of opioids. In the Third District, the overdose death rate from synthetic opioids was even higher.

Although the overdose death rates from natural opioids (e.g., prescription pills) in the Third District states are roughly in line with the national rate, the rate of overdose deaths for synthetic opioids is roughly double or more in each of our three states. Fentanyl has been a recent problem in the heroin supply of East Coast cities, which could contribute to the higher overdose death rates in the Third District—perhaps because East Coast heroin is sold as a powder and is thus easier to cut with an adulterant like fentanyl, whereas West Coast heroin is often sold as a dark brown paste.
The Opioid Epidemic’s Relationship with the Labor Market

On July 13, 2017, in her semiannual testimony before the Senate Banking Committee, then–Federal Reserve Chair Janet Yellen noted the intertwined but complex nature of the opioid crisis’s relationship with the labor market and the broader state of the economy: “I don’t know if [the crisis is] causal or symptomatic of long-running economic maladies that have affected these communities and particularly affected workers who have seen their job opportunities decline.”

Theoretically, the opioid crisis makes some workers less likely to search for and find suitable positions, causing problems with the labor supply. However, these same individuals could have been driven to use drugs as a result of poor health (e.g., chronic pain due to osteoarthritis or diabetes) or because they were discouraged due to a long-term decline in the U.S. demand for low-skilled workers (e.g., manufacturing jobs), a trend that is particularly noteworthy for males. (See Figure 7.) These labor market symptoms are difficult to discern in large urban areas such as Philadelphia or New York, but they become more apparent in less heavily urbanized regions affected by the epidemic.

A 2017 paper by Alan Krueger explores the relationship between a declining labor force participation rate and the opioid crisis. Krueger notes that “labor force participation is lower and fell more in the 2000s in areas of the United States that have a higher volume of opioid medication prescribed per capita than in other areas.” He goes on to suggest that 43 percent of the observed decline in the male labor force participation rate between 1999 and 2015 could be attributed to the increase in opioid prescriptions during that time. However, Krueger notes that it is unclear whether other factors that result in low labor force participation (e.g., poor health, discouraged workers) could have also resulted in high prescription rates of opioids in certain counties.

Exploring the relationship further, a 2018 NBER paper investigated the effect of opioid prescription rates on employment-to-population ratios at the county level. Its authors found that the effect is positive but small for women (suggesting that higher opioid use in specific counties allows more women to enter the labor force), and that there is no relationship for men. The case for causality is not strong enough to suggest that opioid prescriptions lead directly to poor employment outcomes in the above studies.

However, there have been reports of channels in which this link occurs. One of those channels is drug testing. A May 2018 Federal Reserve Bank of Cleveland report noted that, after soliciting input on how the opioid epidemic was affecting the

### Figure 7

**Non-Hispanic White Men of Working Age More Likely to Die from Drugs**

Drug poisoning mortality in the U.S., 1999–2016, by sex, age, and race/ethnicity

<table>
<thead>
<tr>
<th>Total Deaths, 2017</th>
<th>70,237</th>
</tr>
</thead>
<tbody>
<tr>
<td>By sex.</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>0%</td>
<td>20%</td>
</tr>
<tr>
<td>By age.</td>
<td></td>
</tr>
<tr>
<td>&lt;15</td>
<td>15–24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age-Adjusted Death Rate</th>
<th>By race.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>2005</td>
</tr>
</tbody>
</table>

business community, several contacts cited candidates’ inability to pass drug tests as being a hindrance to finding qualified employees.\textsuperscript{22} Thus, while there is not a clear enough causal link between opioid abuse and poor employment outcomes, the correlation is sufficient to warrant further study of the link and optimal policy for assuaging both problems.

**Accounting for the Costs of the Opioid Epidemic**

The exact costs of this ongoing problem are difficult to measure precisely, but a number of studies have made an attempt to reflect the societal costs across a variety of categories. A cost estimate of the opioid epidemic by the Council of Economic Advisors (CEA) placed the total cost in 2015 at $504 billion for the nation, with an expectation for that figure to grow if the crisis worsened.

In an attempt to quantify the costs of the opioid epidemic at a more local level, Alex Brill and Scott Ganz used the CEA estimates along with local variations of health-care costs, criminal justice services, and worker productivity to arrive at county- and state-level estimates.\textsuperscript{22} They found that the total per capita costs of the opioid epidemic varied from a high of $4,378 in West Virginia to a low of $394 in Nebraska in 2015, with a median of $1,672. For the Third District, the per capita costs of the opioid epidemic in 2015 were $1,907 in New Jersey, $1,945 in Pennsylvania, and $2,530 in Delaware, all above the median. The total cost of the opioid epidemic across the Third District states in 2015 was nearly $45 billion.

Who bears these costs? How are these costs, fatal and nonfatal, distributed throughout society? Studies disagree, so rather than try to estimate the specific numerical costs, we will examine the channels through which these costs flow.

For fatalities, there are the health-care costs of treating overdosed patients and the losses in future productivity. Much of the lost productivity is borne by the deceased’s family and the private sector. Given that the average age for an overdose fatality is 41, which is considered prime working age,\textsuperscript{23} the losses in future productivity are quite high. However, these losses also show up in federal, state, and local tax receipts.

Aside from these monetized costs, there are numerous unquantifiable effects on families and communities as a result of the opioid crisis. Opioid-dependent individuals suffer a substantial decrease in their quality of life. Their families experience pain and suffering as a result of this dependence and certainly as a result of any overdoses or overdose fatalities that may occur. Local communities may suffer any number of problems, including decreased property values and a loss of community well-being and safety.

**Policy Countermeasures**

Since 2010, when opioid prescriptions peaked in the United States, there have been a number of coordinated policy responses at the federal, state, and local levels aiming to counteract the opioid crisis. In 2011, the U.S. Office of National Drug Control Policy recommended that states have active prescription drug monitoring programs (PDMPs) to counteract overprescribing. A PDMP is an electronic database that tracks controlled-substance prescriptions in a state.\textsuperscript{24} These state-level databases track individuals who obtain prescriptions, informing doctors of patients’ histories in order to mitigate overprescribing. However, PDMPs are most effective in preventing overprescribing if the state requires clinicians to check the state’s PDMP before prescribing a controlled substance. States that have implemented such laws since 2011 have reduced oxycodone prescriptions and oxycodone deaths.\textsuperscript{25} Within a year of passing a 2012 law requiring prescribers to check the state’s PDMP before prescribing opioids, New York State saw a 75 percent drop in patients seeking multiple prescribers for the same drugs.

Another major public health response to the proliferation in opioid overdoses over the past 10 years has been the growing administration and supply of naloxone. Sold under the brand name Narcan, naloxone is an opioid antagonist used to temporarily reverse the effects of an opioid overdose.\textsuperscript{26} State and local governments have combated overdoses by increasing the supply of naloxone and training first responders on how to use it. In 2014 the New York Office of the Attorney General provided $1.2 million to supply 20,000 kits to police officers in the state.\textsuperscript{27} And in December 2018 the Pennsylvania Department of Health instituted a program for any Pennsylvania resident to receive naloxone free from any of 80 locations across the state.\textsuperscript{28} Although it is difficult to measure the exact number of overdose deaths that have been prevented with the drug, in 2017 alone more than 4,000 individuals were administered naloxone by Philadelphia Emergency Medical Services.\textsuperscript{29}

Improving access to addiction treatment is one of the most powerful tools for fighting the opioid epidemic, in that it offers a way out of the cycle of dependence. A number of public programs treat opioid addiction, including medication-assisted treatment (MAT), which is a combination of medication, counseling, and behavioral therapy. State and local governments can set up point-of-contact centers to counsel those seeking recovery on how to access various treatment options. (Pennsylvania refers to these as Centers of Excellence.)\textsuperscript{30} States fund these centers and treatments with Medicaid, a state government insurance program for limited-income individuals and families, but their availability varies by state.\textsuperscript{22}

**Final Thoughts**

As the opioid crisis intensifies in the nation and particularly in our region, it becomes ever more important to understand its impact on society. With the rise in the supply and abuse of highly potent synthetic opioids such as fentanyl, the crisis has entered its deadliest stage yet. Its costs to society are measured not only in terms of a diminished labor force and lost productivity but also in its impact on the health-care sector, the criminal justice system, and families and communities. Much of the research on the economic effects of the opioid crisis is still preliminary and has not captured the last two years’ worth of data, so there is still much to learn about its continuing effects on society and the economy.\textsuperscript{8}
Notes
1 See Federal Bureau of Investigation (2018).
2 The age-adjusted mortality rate standardizes the data by adjusting for age groups in the population. The age-adjusted rate of drug overdose deaths rose from 19.8 per 100,000 in 2016 to 21.7 per 100,000 in 2017.
3 Many drug overdose deaths occur from mixing nonopioid drugs with opioids. For example, the National Institute on Drug Abuse breaks down overdoses from cocaine and benzodiazepines with and without opioid involvement. The overdose death rates of these two drugs without opioid involvement are virtually flat, meaning that opioids, particularly synthetic opioids, are driving this uptick in overall drug overdose deaths. See Figures 7 and 8 at https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates.
4 See Lin et al. (2017).
5 See Jones et al. (2018).
8 Prior to the mid-1990s, the heroin market in the United States was mainly supplied by Asia, via production and transport through countries such as Afghanistan, Myanmar, and Thailand. In the mid-1990s, traffickers from Colombia and Mexico flooded the U.S. market with large amounts of cheap, pure heroin. The supply of heroin from this new source was large enough to reduce the price per gram of pure heroin throughout the country.
10 See O’Connor (2017).
12 See Mars et al. (2017).
13 See Vestal (2019).
15 See Lovelace (2017).
17 See Michaels (2017).
18 Because the Centers for Disease Control and Prevention does not break out opioid data by sex or race, this relationship must remain theoretical for now.
19 See Krueger (2018).
21 See Fee (2018).
22 See Brill and Ganz (2018).
23 See Rhyan (2017).
24 See Florence et al. (2016).
26 See Centers for Disease Control and Prevention (2017a).
28 See Durando (2014).
31 See Pennsylvania Department of Human Services (2019).
32 See Grogan et al. (2016).

References


Regional Spotlight

Smart Growth for Regions of All Sizes

Pittsburgh’s population has shrunk by 400,000 since 1969, making it the poster child for urban shrinkage. So why is it doing so well? Sometimes, smaller really is better.

BY PAUL R. FLORA

Among the truisms to which regional policymakers frequently adhere, the most pervasive may be that a region must grow to be successful. However, population growth and job growth are not preconditions for a region to become economically healthy. Rather, the composition and characteristics of jobs required to meet the demands of a region’s growing (or changing) industrial structure typically determine the health of a region’s economy. High-productivity industries and progressively managed firms generate high-skill, high-wage jobs that raise a region’s per capita income.

Indeed, for a region that has lost its prior locational advantage or is unable to attract high-productivity industries, growth may falter or reverse, but its economic health need not decline if policies recognize the transition; assist those most impacted; and generally lower the cost of living, including adjustments to the scope and cost of public infrastructure. Too often, shrinking regions refuse to accept the reality that their population will not resume a growth path.

Aside from the apparent fallacy of the growth paradigm, the policies that emanate from such a belief typically maintain unrealistic expectations given a region’s economic structure, often ignore stronger countervailing market forces, and routinely waste resources in pursuit of misplaced goals. Pittsburgh’s regional economy offers a persuasive counterexample to the growth paradigm even though its policymakers also often chased economic growth rather than economic health.

Pittsburgh has arguably held the crown for population decline among the nation’s largest metropolitan regions for more than half a century. The region’s population began falling in the 1960s. From 1969 to 2017, the Pittsburgh region’s population fell further, and more consistently year to year, than any other major metropolitan region in the country, including Detroit. Yet no sooner had most of Pittsburgh’s steel mills closed for good in the early 1980s than Pittsburgh began garnering accolades as the “most livable city” and generating a stream of positive media coverage describing its economic comeback.

Moreover, consider that in 1969 the Pittsburgh metro area was the ninth largest in the nation. Pittsburgh has since lost population in 41 of the 48 years through 2017, amounting to a net loss of more than 400,000 people. As Pittsburgh shrank, 17 other metro areas surpassed it in population (Figure 1).

While Pittsburgh’s net population was falling at a pace of 0.3 percent per year, Orlando, FL, added nearly 2 million people (3.4 percent annualized growth). Las Vegas, which will likely surpass Pittsburgh within the next five years, has grown at a 4.5 percent annualized rate since 1969. Has rapid growth in Orlando and Las Vegas made their economies healthier than Pittsburgh’s? In 2016, Pittsburgh’s real per capita income was $60,797—nearly $12,000 more than in Las Vegas and over $15,000 more than in Orlando. In 2017, Pittsburgh’s poverty rate was 11.8 percent (a five-year-average estimate); the poverty rate was 14.6 percent in Las Vegas and 15.4 percent in Orlando. In Las Vegas and Orlando, population and job growth have produced neither greater overall incomes per capita nor a more equitable distribution of income.

Of the 17 metro areas that have surpassed Pittsburgh in population since 1969, only three (Minneapolis, St. Louis, and Seattle) had higher real per capita incomes than Pittsburgh as of 2016. And only four (Baltimore, Denver, Minneapolis, and Seattle) had lower poverty rates as of 2017. Despite over five decades of population decline, Pittsburgh remains an economically better place to live than most of its now larger peers.

How does a region’s economy improve while declining in population? What does regional economic growth mean, and how should we measure it? Most important, how can we ensure that policymakers are governed by a realistic appraisal of a region’s prospects and develop strategies to grow better, not just bigger?

These questions are especially important in Pennsylvania. In seven of the state’s 15 more mature metropolitan statistical areas, population has been shrinking for most of the past 50 years.
In 1969, Pittsburgh was the ninth largest metropolitan region in the country, but its population had been falling since 1960, in step with the decline of the nation’s integrated steel industry. Other than auto manufacturing in Detroit, arguably no other industry had dominated a region more thoroughly than steel had dominated Pittsburgh. During the double-dip recession of the early 1980s, most of Pittsburgh’s large integrated steel mills closed their doors for good.

Meanwhile, the advent of air conditioning ushered in the migration from the Rust Belt to the Sun Belt and propelled numerous southern regions into periods of rapid growth. As of 2017, 17 metropolitan statistical areas (MSAs) had surpassed Pittsburgh in population. Miami, Dallas, Phoenix, and Orlando shot past on their own rapid trajectories. Others, such as St. Louis and Baltimore, edged past by rising slowly.

The seeds for Pittsburgh’s revival had been sown over many prior decades, beginning perhaps with the writings of the 1930s business economist Glenn McLaughlin, who warned city leaders that steel was a mature industry and the region should diversify to prepare for steel’s gradual decline. Pittsburgh engineered numerous critical changes, including strict county air pollution regulations, flood control projects, and downtown renewals such as Renaissance One, Renaissance Two, and several successors.

While later researchers have pointed to myriad current explanations for Pittsburgh’s economic revival, the decades of prior city renewal efforts enabled Pittsburgh to reinvent itself after the decline of steel. From the 1990s on, firms like Google and RAND have been attracted to Pittsburgh as Carnegie Mellon University and the University of Pittsburgh led advances in higher education, the medical sector, life sciences, computer science, and robotics. Meanwhile, PNC Bank supplanted Mellon Bank as the region’s major financial employer and has been a significant downtown developer since the 1990s. Thus the stability of Pittsburgh’s economic health during the past 35 years has resulted from the emergence of high-skill, high-paying professional and technical jobs, which eased the sting of losing thousands of lower-skill, well-paying steelworker jobs.

Yet the region’s population continued to shrink. Following the demise of steel, economic policymakers persisted with the assumption that the region would begin to grow next year and they planned accordingly—against the advice of regional economists that ongoing decline was more realistic. Pittsburgh likely would have fared better (with greater gains or a smoother transition) had policymakers continued to heed the advice of regional economists, as in prior decades.
Moreover, 12 of 20 smaller micropolitan statistical areas, such as Oil City, PA, have seen declining populations since at least 2000. Only seven micro areas have grown during that period, and only four grew sufficiently to earn a “promotion” to metro area status.

The Case for Regions to Pursue Per Capita Income Growth Rather Than Jobs

Regions are too often compared on size-based measures such as population growth, employment growth, GDP, and total personal income. Less often, or as a secondary measure, regions are compared on a per capita basis, such as per capita income. Let’s call these economic-health-based measures.7

Whereas GDP measures a region’s total economy, GDP per capita represents a region’s average potential spending power, which falls when population rises faster than GDP. Thus, growth in per capita measures is more vital to the well-being of a region’s people. Individuals and families are better off in regions with higher per capita income, or where per capita incomes are rising faster.8

In fact, a region’s population growth has a clear, strong correlation with the growth of its total real personal income. Since 1969, Austin, TX, Las Vegas, Orlando, Phoenix, and Raleigh, NC, have been the five fastest-growing regions for both population and total real personal income. By contrast, Cleveland, Detroit, Pittsburgh, and Buffalo, NY—the only four regions in which population has fallen since 1969—bring up the rear with the slowest growth of total real personal income (Figure 2).9

However, the strong positive correlation dissipates when real per capita income is examined in place of total income. Instead, a weak negative correlation appears. Austin and Raleigh still show strong growth rates for real per capita income, but Las Vegas, Orlando, and Phoenix show relatively weak growth rates. Meanwhile, Buffalo, Detroit, and Cleveland show below-average growth in real per capita income, while Pittsburgh’s growth is above average (Figure 3).

Why do size comparisons persist if per capita comparisons matter more for individual welfare? For one, size comparisons were once much simpler to make, and their use continues through inertia.10

Moreover, aggregate economic growth is generally perceived as a desirable outcome that benefits firms as well as state and local governments. Because local businesses and local politicians typically frame the economic development conversation, size comparisons tend to dominate the analysis. To be sure, population growth leads to new home construction, rising retail sales, and a larger tax base.

Thus, firms seeking to increase revenues or market share will expand into growing suburban areas and in the Sun Belt, where population and consumer demand are growing most rapidly. At least three unfortunate events await many of these firms.

First, retailers have often extended their suburban reach too far, only to be reminded that the urban center retains a locational advantage for well-heeled customers. Second, over time in regions with a rising percentage of lower-income households, local firms will find their profit margins squeezed by price-sensitive consumers. Finally, when growth does stop, local businesses will be left servicing less-profitable customers than their counterparts in high per capita regions.

**FIGURE 2**

Total Personal Income Growth Is Highly Correlated with Population Growth

Austin, TX, and Las Vegas are the latest U.S. boom towns.

Annualized change in aggregate real personal income vs. annualized percent change in pop., 1969 to 2017

Source: Bureau of Economic Analysis.
Taxing entities also prefer population growth, which often masks an unsustainable fiscal structure. Florida is able to provide state and local government services without an income tax by relying on impact fees for new development, a high sales tax that generates a large proportion of its revenues from tourists, and a “welcome stranger” property tax that caps increases for homestead property owners and shifts a disproportionate burden to new homebuyers.\textsuperscript{11}

However, when growth stops, the cost of providing and maintaining infrastructure and delivering services grows faster than tax revenues. In Florida, the Great Recession revealed an unpleasant, surprise feature of its tax cap for existing homeowners: As their property values plummeted, their assessed values continued to rise, which resulted in higher property taxes.\textsuperscript{12}

If a region’s population declines permanently, so too does its tax base, often forcing state and local governments to raise taxes on a population with lower incomes and less wealth. The resulting fiscal stress often prompts local officials to pursue growth strategies—unsuccessfully in the face of much stronger countervailing economic forces. Within a region, this problem may be further magnified by local government fragmentation, which can accelerate a migration of households with means from fiscally distressed cities to surrounding suburban and rural jurisdictions.

Understandably, firms and regions are inclined to try to stay on the easier path by pursuing continuous growth, but it is unrealistic to expect to be forever immune from economic shocks that cause growth to slow, stop, or reverse.

\textbf{Industrial Structure Remains Key to Assessing a Region’s Prospects}

Since François Quesnay published his \textit{Tableau économique} in 1758, economists have studied nations, and then regions, in terms of the industries that compose the economy to better understand what drives economic growth. Quesnay believed that agricultural surpluses were the prime mover.

Today, economists speak of a region’s export base (or economic base) as those sectors associated with the region’s production of goods or services in excess of local demand. The auto industry remains a significant part of Detroit’s economic base, film studios of Los Angeles’s economic base, and finance of New York’s economic base.

A region’s export base affects its economy in two key ways. First, employment typically rises (or falls) as industries present within a region’s export base grow (or decline). Second, per capita income is greater in regions whose export-base sectors utilize highly skilled, high-paid employees.

In turn, the usual multiplier effects that generate local jobs (e.g., carpenters, teachers, clerks, and wait staff) will be stronger in regions in which higher-paid export-base workers will consume high-value goods and services.

To better understand how a region attains high real per capita incomes with or without population growth, I compare four regional economies that represent four extremes of the distribution and examine their industry mix: Austin, with high population growth and high per capita income growth; Las Vegas, with high population growth and low per capita income growth; Cleveland, with

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{But Per Capita Income Growth Is Not Correlated with Population Growth}
\end{figure}

\textbf{Austin, TX, and Las Vegas have widely disparate rates of per capita income.}

\textbf{Annualized change in real personal income per capita vs. annualized percent change in pop., 1969 to 2017}

\textit{Source: Bureau of Economic Analysis.}
low population growth and low per capita income growth; and Pittsburgh, with low population growth and high per capita income growth.

Not surprisingly, the industry mix differs substantially among these four regions. A good sense of the differences can be gained by identifying within each region the top five industrial sectors (by location quotient, a measure of an industry’s concentration in an area) for which the sector employs at least 5,000 workers (Figure 4).13

Austin’s top five sectors are representative of an economy with a large concentration of high-paying, high-tech jobs at firms that design and produce computer hardware and software. Pittsburgh is still characterized by its steel industry legacy plus its education and health sectors. Cleveland retains concentrations in many small, diversified manufacturing industries (nine of its 10 largest location quotients were manufacturing sectors) but has also experienced a shift to health care. In contrast, the Las Vegas economy is heavily concentrated in tourism sectors that do not pay very high wages.

To provide a more comprehensive comparison, I computed a weighted average wage for employees who represent the export base of each region.14 The export-base employment of Austin generated an average weekly wage of $1,841 in 2016—significantly higher than Pittsburgh’s $1,346 and Cleveland’s $1,245. Export-base workers in Las Vegas averaged only $793 per week.

How a region grows, whether in high-skill, high-wage sectors or in low-skill, low-wage sectors, has important implications for the overall long-term health of the region’s households and of the region itself. Were growth alone responsible for lifting a regional economy and all of its participants, then 46 years of rapid growth should have turned Las Vegas into one of the healthiest economies in the nation. If instead population growth is driven primarily by low-wage jobs in sectors such as call centers, construction, tourism, and warehousing, then a region may grow poorer while its population is growing larger.

The distribution of 2016 real per capita incomes adjusted by regional price parities further demonstrates a lack of correlation with population growth and reflects instead the industrial structure (Figure 5). In every income bracket, one can find fast-growth and slow-growth regions. Despite very slow growth (or no growth), Cleveland, Philadelphia, Pittsburgh, and St. Louis enjoy per capita income levels equal to Denver, Houston, and Nashville, TN, which have all grown at twice the national rate. Of note, high-income outliers (San Jose, CA, San Francisco, and Boston) have significantly lower population growth rates than the low-income outliers (Orlando, Tucson, AZ, Las Vegas, Phoenix, and Tampa, FL).

Finally, 2017 poverty rates show a distribution that is likewise uncorrelated with population growth. Austin, Cleveland, Las Vegas, and Pittsburgh are once again located in separate quadrants of the scatter plot (Figure 6).15

Despite enduring population losses for over 50 years, the Pittsburgh region maintains higher real per capita income and a lower poverty rate than most of its peers. This outcome suggests that regional policymakers should not simply seek job growth but should pursue development strategies that emphasize the quality of jobs and the needs of the resident population.

**How to Grow a Healthier Economy—Without (Necessarily) Growing More Populous**

Regions face the same basic challenges, whether they are anchored by a large, mature, slow-growing city; a midsize, youthful, rapidly growing city; or a small, declining city contemplating the

---

**Figure 4**

*Industry Matters More Than Population Growth*

2017 annual average salary; 2017 average weekly wages; sectors employing at least 5,000 workers

*Four Examples of Extreme Distribution*

<table>
<thead>
<tr>
<th>Real per capita income growth</th>
<th>Population change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Pittsburgh</td>
<td>Austin</td>
</tr>
<tr>
<td>U.S.</td>
<td>Cleveland</td>
</tr>
<tr>
<td>Las Vegas</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Location quotient quantifies how concentrated an industry is compared to the U.S. average. A location quotient of 10 indicates jobs in a specified industry are 10 times more concentrated in a specified region as in the nation as a whole.


---

**Salaries for Sectors with Five Highest Location Quotients in Each Region**

<table>
<thead>
<tr>
<th>Annual average salary</th>
<th>Average Weekly Wage of Export-Related Employment</th>
</tr>
</thead>
<tbody>
<tr>
<td>$180,000</td>
<td>$2,000 - - -</td>
</tr>
<tr>
<td>$160,000</td>
<td>$1,750 - - -</td>
</tr>
<tr>
<td>$140,000</td>
<td>$1,500 - - -</td>
</tr>
<tr>
<td>$120,000</td>
<td>$1,250 - - -</td>
</tr>
<tr>
<td>$100,000</td>
<td>$1,000 - - -</td>
</tr>
<tr>
<td>$80,000</td>
<td>$750 - - -</td>
</tr>
<tr>
<td>$60,000</td>
<td>$500 - - -</td>
</tr>
<tr>
<td>$40,000</td>
<td>$250 - - -</td>
</tr>
<tr>
<td>$20,000</td>
<td>$0 - - -</td>
</tr>
</tbody>
</table>

**Source:** Bureau of Labor Statistics.
promise from the latest resource boom. Regions typically strive to deliver public services and to enable the provision of amenities to meet the needs of residents and firms within a fiscally sound, long-range budget constraint. A region’s industrial structure is a major determinant of the budget constraint. First, policymakers should approach economic development with a realistic understanding of their region’s place and prospects in the world economy. A comprehensive economic base analysis provides a good start to avoid setting unattainable goals and wasting resources on empty strategies. This analysis should undergird any objective assumptions about future population or employment trends.

Next, regions should develop an infrastructure plan and pragmatic policy solutions for addressing the economic needs of existing and future residents, including a sustainable fiscal path for the region’s local governments. Ideally, each region would produce a multi-jurisdictional fiscal impact analysis of its long-range comprehensive plan to ensure its efficiency and feasibility.

In addition, all regions, but especially those with a high proportion of low-wage jobs, may find that they need to budget for strategies that reduce the cost of living for households living on those minimum-wage jobs. Policymakers may need to reduce barriers to education and labor force participation by, for example, assisting with day care, health care, job training, and transportation needs for marginally attached workers, and by encouraging the provision of affordable housing near jobs and in transit-oriented locations.

Finally, and especially for regions that are stagnant or declining in population, policymakers should consider developing people-oriented policies that help persons relocate to regions with greater job opportunities and consider rationalizing public infrastructure with incentives for the region’s residents to consolidate into a more compact urban form.

These recommendations do not include policies to target and attract particular businesses. Better perhaps is a policy to not target or subsidize any business that does not utilize high-value occupations, or that does not make immediate use of a region’s existing underutilized labor force.

A final key point is that all of the above policy prescriptions are (optimally) regional. Cities, their suburbs, and their hinterlands will realize their greatest economic success by working as one. Ideally, states would encourage municipal consolidations that expand the political/fiscal base to match a region’s economic footprint. Since municipalities are creatures of the state, the burden of failing to relieve the inefficiencies of local government fragmentation will fall to the state to address.
Notes
1 Cleveland’s population fell during the same period at a slower rate than did Pittsburgh’s. Detroit’s population fell at a much slower rate and rose in more years than it fell. Buffalo’s population declined at a similar rate, but less consistently, and Buffalo was, and is, half the size of Pittsburgh.

2 Unless otherwise noted, region and metro area refer to official metropolitan statistical area (MSA). Analysis in this article is based on data for each MSA as delineated in the Office of Management and Budget Bulletin 18-03 issued April 10, 2018. This article truncates these official names to the names of their largest principal cities.

3 The brief interludes of population growth occurred in years that followed economic recessions, suggesting that some of the Pittsburgh diaspora returned home after losing jobs in other regions. (They may have felt that it is better to be unemployed near family and friends than in a relatively strange place.)

4 Recently released census estimates of 2018 population indicate that Sacramento, CA, became the 18th metro area to surpass Pittsburgh’s population.

5 Per capita income estimates have been adjusted for cost-of-living differences and are expressed in 2018 dollars.


7 Other economic-health-based measures include poverty rates, unemployment rates, and comparative cost-of-living measures.

8 This statement assumes that other variables are the same, including potential income inequality.

9 The article analyzes the 53 metro areas in the United States with populations greater than 1 million in 2017. The United States as a whole is also represented. The variables are based on data from the U.S. Census Bureau, the Bureau of Labor Statistics, and the Bureau of Economic Analysis.

10 Adjusting economic data to eliminate potentially distorting underlying factors—such as population growth or the presence of an unusual number of college students, retirees, migrants, or prisoners—was difficult before computers and remains complicated today.

11 Florida’s constitution limits the annual increase in assessed value of properties with a homestead exemption to 3 percent or the change in the Consumer Price Index, whichever is lower. New homebuyers can face tax bills that are several times greater than their long-tenured neighbors in comparable properties.

12 The recapture rule of the Florida law requires that homestead properties with an assessed value below market value must be assessed the legislated increase even when the market value has fallen, as long as the assessed value is below market value.

13 This selection was necessarily an arbitrary one that misses sectors with smaller location quotients, which may employ significantly more workers or pay significantly higher wages. Moreover, key sectors with higher location quotients may have been suppressed in the Quarterly Census of Employment and Wages data set based on nondisclosure rules of the Bureau of Labor Statistics.

14 The average weekly wage was calculated on a weighted basis across all sectors within each region with a location quotient of 1.2 or higher. The average weekly wage in each of these sectors was multiplied by the number of workers in each sector in excess of the number required to reach a location quotient of 1.2. The employment and wage data are 2017 annual averages for all sectors in a region except those sectors for which the BLS suppressed data because of nondisclosure rules.

15 Some shifting and some compression would occur in this scatter plot if poverty rates could be adjusted for regional price parities, as was done with per capita income.


18 See Giarratani and Houston (1989).

19 See Giarratani and Houston (1989).

References


Implementing Monetary Policy in a Changing Federal Funds Market

As the Fed normalizes its balance sheet, it helps to understand how the federal funds market used to operate, how it changed in the wake of the crisis, and what comes next.

BY BENJAMIN LESTER

Every six weeks or so, the financial world watches as the Federal Open Market Committee (FOMC) decides on a target interest rate in the federal funds market. But what happens next? How do policymakers make sure that interest rates in the fed funds market trade within the target range?

Though not widely discussed, the framework that the FOMC uses to implement monetary policy has changed over the last decade and continues to evolve today. Before the financial crisis—when reserves were scarce—policymakers used one set of instruments to achieve the target rate. However, several important policy interventions introduced soon after the crisis drastically altered the landscape of the fed funds market. This new environment—with ample reserves—necessitated a new set of instruments for monetary policy implementation. Now, as the FOMC begins to unwind the effects of these policy interventions, the question arises: What happens next as the fed funds market converges to a “new normal”?

Implementing Monetary Policy Before the Crisis

Banks hold reserves in an account at the Fed and are required to maintain a balance above a certain fraction of their deposits—so-called required reserves. Prior to the onset of the Great Recession in December 2007, a defining feature of the fed funds market was that reserves were scarce. As a result, throughout the day a bank’s reserves would fluctuate as payments were made or received, and some banks would find themselves short of their reserve requirements at the end of the day. In order to avoid borrowing at the Fed’s discount window, these banks would look to borrow from other banks in the fed funds market.

At the same time, some other banks would find themselves holding excess reserves at the end of the day. Since the Fed didn’t pay interest on excess reserves deposited overnight, these banks would look to lend in the federal funds market to earn a positive rate of return. As there were a significant number of banks on both sides of the market—some looking to borrow and others looking to lend—trading volume in the fed funds market was substantial, and interbank trades dominated market activity. For instance, Afonso, Entz, and LeSueur estimate an average daily trading volume of approximately $200 billion in the fourth quarter of 2006, of which approximately 60 percent was accounted for by bank-to-bank lending.

In this environment of scarce reserves, monetary policy implementation was fairly straightforward. The Open Market Trading Desk (the Desk) at the Federal Reserve Bank of New York would implement the desired target for the effective federal funds rate (EFFR) by adjusting the supply of reserves via open market operations. For example, if the Desk wanted to increase market rates, it would sell securities (such as Treasury bills) in the market, thereby decreasing the supply of cash held by banks. As banks’ supply of cash became scarcer, the rate at which they would be willing to lend would rise. Hence, as in the usual model of supply and demand, a reduction in the supply of reserves in the market would lead to an increase in the fed funds rate. (See Figures 1 and 2.) As the fed funds rate rose, market rates would rise as well.

Three Important Changes

The landscape of the fed funds market was altered dramatically following the financial crisis. First, and most important, the Fed’s large-scale asset purchase programs left depository institutions awash with reserves. Over three rounds of “quantitative easing” in 2008, 2010, and 2012, the Fed purchased assets such as U.S. Treasury debt and agency mortgage-backed securities. As the Fed bought these assets, the banks that sold them saw their reserve balances soar. As a result, excess reserves held by depository institutions reached nearly $2.7 trillion by August 2014.

To put that in perspective, in the precrisis years, excess reserves typically hovered between just $1 and $2 billion. Second, changes in the assessment of FDIC fees made borrowing in the interbank market more expensive for domestic banks.
Insurance Corporation (FDIC) changed the basis for its fees from a bank’s deposits to its assets. Since a bank’s reserves are included in the calculation of its assets, this policy change increased FDIC fees and, hence, the cost of borrowing reserves on the interbank market. Economists estimate that these policy changes implied an additional cost between 4 and 7 basis points for each extra dollar of cash on a bank’s balance sheet.

However, FDIC fees are imposed only on banks with U.S. deposits, and branches of foreign banks typically don’t hold U.S. deposits, so this policy change raised the cost of borrowing for domestic banks while leaving foreign banks with U.S. subsidiaries largely unaffected.

Third, in October 2008, in the hope of putting a “floor” beneath market rates, the Fed started paying an interest rate of 25 basis points on overnight reserves deposited by banks. However, this overnight rate was not made available to other financial institutions, including government-sponsored entities like the Federal Home Loan Banks (FHLBs) as well as money market funds. As a result, the introduction of interest on reserves (IOR), with eligibility restrictions, created a gap between the interest rates available to different types of financial institutions.

**Postcrisis Implications**

These changes altered the fed funds market in a number of important ways, including the types of financial institutions that were trading, the rates at which they were borrowing and lending, and the tools available to the FOMC that could effectively influence these market rates.

Because banks were awash with reserves, their desire to borrow effectively vanished, and bank-to-bank lending largely disappeared. However, once the Fed started paying interest on reserves to some (but not all) financial institutions, a new lending opportunity emerged.

To understand this opportunity better, consider a financial institution ineligible to receive interest on reserves at the Fed, such as an FHLB. At the end of the day, it likely holds some amount of cash, but the highest overnight interest rate it could receive—what economists call its “outside option”—was a zero percent net return.
Eligible financial institutions, however, had a better outside option, since they could deposit money at the Fed and earn the IOR rate (initially set at 25 basis points), less any costs associated with expanding their balance sheet. Because only domestic banks incurred FDIC fees from increasing their asset position, foreign banks faced smaller costs and thus had an advantage in borrowing.

Hence, an opportunity for arbitrage emerged: The FHLB could lend to an eligible bank at a rate above zero (its outside option) but less than the IOR rate, and the eligible bank could lend those reserves to the Fed at the IOR rate (its outside option). “Arbitrage in the Fed Funds Market” describes in greater detail the arbitrage opportunity that emerged because of differing outside options, the effects of borrowing costs like FDIC fees, and the determination of a mutually agreeable interest rate.

As a result of the many changes in the immediate aftermath of the crisis, the majority of trading in the fed funds market was occurring between ineligible financial institutions, like FHLBs, and eligible financial institutions with low costs of borrowing, like U.S. branches of foreign banks, at rates below the IOR rate being offered at the Fed. Moreover, with no bank-to-bank lending, the overall market volume dropped precipitously, to $80 billion or less per day. (See Figure 3.)

**Implementing Monetary Policy After the Crisis**

These changes to the fed funds market required policymakers to devise a new system for implementing monetary policy. Since the market rate was no longer primarily determined by banks’ supply and demand for reserves, typical open market operations would have essentially no effect on market rates. Instead, when the FOMC decided to raise interest rates after a long period at zero, it did so by adjusting the outside options of the lenders and the borrowers in this market via administered rates.

The Fed had been controlling the outside option of eligible banks via the IOR rate since October 2008. However, if the Fed adjusted this rate alone, the gap between the two outside options would widen as the IOR increased and, as a result, market rates might not rise in sync with the IOR. So in September 2013 the FOMC introduced an instrument to adjust the outside option of ineligible institutions, too, via the overnight reverse repurchase agreement facility, or ON RRP.

In a reverse repurchase, the Desk sells a security to an eligible counterparty with an agreement to buy the security back at a specified date and price, with the interest rate computed from the difference between the original purchase price and the (higher) repurchase price. Importantly, the FOMC included a wide range of market participants as eligible counterparties at the ON RRP facility, including FHLBs and key money market funds. By adjusting the rate being offered at the ON RRP facility, the FOMC was thus adjusting the outside option of essentially all major financial institutions ineligible to earn IOR at the Fed.

---

**Arbitrage in the Fed Funds Market**

Between October 2011 and September 2013, an FHLB could earn a zero net return on any cash it held at the end of the day. However, it could lend that money to a bank eligible to earn the IOR rate, 25 basis points, less any costs associated with expanding its balance sheet. Suppose these costs were 5 basis points, so there were “gains from trade” between the FHLB and the bank of 25−5=20 basis points. This means the two parties would agree to trade at any interest rate between 0 and 20 basis points.

What determines the interest rate at which they actually trade? In bilateral transactions like this, we often assume that the two parties negotiate or “bargain.” Moreover, we assume that the interest rate at which they agree to trade depends on each party’s relative negotiating skill or “bargaining power.” If the bank has more bargaining power, it negotiates an interest rate r closer to zero so that its profit, 20−r, is relatively large. If the FHLB has more bargaining power, it negotiates an interest rate closer to 20 so that it earns more profit on its overnight loan.

A number of factors could determine the bargaining power of a bank or an FHLB. For example, an FHLB that can quickly and easily find an alternative bank to trade with would be in a relatively strong bargaining position. However, a bank that was desperate to borrow to avoid violating reserve requirements would be in a relatively weak bargaining position.
Since the FOMC began raising the target rate in December 2015, it has used these two instruments—the IOR and ON RRP rates—to raise and control the fed funds rate in a market characterized by ample excess reserves. In particular, as Armenter and Lester (2017) describe, the FOMC has raised rates by increasing both the ON RRP and IOR rates at the same time, while it has adjusted where the fed funds rate falls within the target range by adjusting the IOR rate.

The top panel of Figure 4 illustrates the relationship between the ON RRP rate, the IOR rate, and the fed funds rate between December 2015 and September 2018. The bottom panel of Figure 4 plots the spread between the IOR and ON RRP rates between June 2017 and September 2018, and it plots where the EFFR rate falls within this spread (the red line).

From the time it “lifted off” from zero until 2018, the FOMC raised the IOR and ON RRP rates in tandem, with a 25 basis point spread between the two. The EFFR followed suit, staying safely within the target range until the second quarter of 2018. At that time, however, the outside option of ineligible financial institutions began rising, putting upward pressure on the EFFR. In response, when the FOMC raised the target range in June 2018, it increased the ON RRP rate by 25 basis points but the IOR rate by only 20 basis points. Decreasing the spread between the IOR and ON RRP rates puts downward pressure on the fed funds rate, helping to keep it within the target range.

**Normalization**

In the summer of 2017 the FOMC announced its intention to stop reinvesting the proceeds from maturing assets (such as mortgage-backed securities) on its balance sheet. This decision marked the beginning of the Fed unwinding or “normalizing” its balance sheet. As the Fed’s balance sheet shrinks, excess reserves in the banking sector decline. However, at the time, the FOMC did not provide an explicit endpoint for this process.\(^{10}\)

More recently, in January 2019 the FOMC announced how it planned to hold “no more securities than necessary to implement monetary policy efficiently and effectively”: by using a “regime in which an ample supply of reserves ensures that control over the level of the federal funds rate and other short-term interest rates is exercised primarily through the setting of the Federal Reserve’s administered rates, and in which active management of the supply of reserves is not required.”\(^{11}\) In other words, the FOMC decided to shrink the balance sheet until reaching the minimal size still consistent with “ample” excess reserves, and to use the ON RRP and IOR rates to achieve the target fed funds rate.

This decline in aggregate excess reserves changes the individual behavior of market participants, and this in turn affects overall market conditions in the fed funds market, including interest rates and trading volume. In particular, if total excess reserves decline enough, the market will transition from the ample-reserve regime—in which open market operations have little effect—to the precrisis scarce-reserve regime. However, it is difficult to forecast when this transition will occur because it depends not only on the level of excess reserves in the market but also on the distribution of these reserves across banks, which is hard to predict.

---

**FIGURE 4**

**How the Fed Changes Rates Post-Great Recession**
The Fed uses the IOR and ON RRP rates to adjust the EFFR.

Source: FRED, Federal Reserve Bank of St. Louis.
Who Trades with Whom, and at What Price?
In the fed funds market, a bank can try to find a counterparty to borrow from (either an ineligible financial institution, like an FHLB, or another bank), it can try to find a counterparty to lend to (another bank), or it can remain idle. When all banks are awash with reserves, there is no motive to lend, since nobody in the market is willing to pay more than the IOR rate. Hence, when reserves are ample, banks with sufficiently low balance-sheet costs (such as banks not subject to FDIC fees) will borrow from institutions such as FHLBs at a rate between the ON RRP rate and the IOR rate, and the remainder of banks (with higher costs from expanding their balance sheets) will remain idle.

However, as total reserves decline, some banks will find themselves lending in the fed funds market, and they will do so, as total reserves decline further, there will be more and more desperate banks looking to borrow.

When this occurs, banks that are far from their reserve requirements will face a choice. These “nondesperate” banks can continue looking to borrow from an FHLB at a rate below the IOR rate, pocketing the difference (less any balance-sheet costs), or they can try to lend to desperate banks at a rate above the IOR rate. As the Fed’s balance sheet shrinks and reserves become increasingly scarce, the demand for reserves from desperate banks will grow, the supply of reserves from nondesperate banks will shrink, and lending to desperate banks will become more attractive. At some point, nondesperate banks will once again find themselves lending in the fed funds market, and they will do so at rates above the IOR rate.

This shift in the behavior of individual market participants has several important implications for the fed funds market as a whole. First, the fed funds rate, which is an average of all rates in the fed funds market, will no longer reside within the corridor formed by the ON RRP and IOR rates. It will instead lie within the corridor formed by the IOR and discount-window rates. Second, as bank-to-bank lending resumes alongside trades between FHLBs and banks, trading volume should also increase. Lastly, since the market rate will be determined by supply and demand once again, the fed funds rate will be sensitive to relatively small changes in the supply of reserves.

When Are Reserves No Longer ‘Ample’?
How much must total reserves shrink before we see these changes? Because the logic above suggests that the fed funds rate should move from one corridor to another when enough banks find themselves with scarce reserves, it is not sufficient to know the total level of reserves. In addition, we need to know the distribution of those reserves across banks! To see why, consider what would happen if the total amount of excess reserves declined by $100 billion and the entirety of this decline came off the balance sheets of banks already close to their reserve requirements. This would immediately force a number of banks to enter the fed funds market as borrowers, prompting other banks to lend above the IOR, thus raising rates. However, if this decline in reserves came off the balance sheets of banks far from their reserve requirements, it would have little effect; all banks would continue to borrow from FHLBs at rates below the IOR.

Hence, to forecast the level of reserves at which the market transitions from ample to scarce reserves, we need to predict the distribution of reserves across banks as the Fed’s balance sheet shrinks. Several factors determine this distribution, including each bank’s size and the regulatory costs they face. In a recent paper with Afonso and Armenter, we estimate the total quantity of reserves consistent with the fed funds rate returning to a corridor between the IOR and discount-window rates. Our benchmark model suggests an answer of approximately $900 billion. However, we find that our estimates are quite sensitive to what we assume about the evolution of the distribution of reserves. In particular, assuming that the majority of the decline in aggregate reserves is absorbed by the smallest or largest banks, respectively, produces estimates as large as $1.1 trillion and as small as $500 billion.

Conclusion
In response to the financial crisis, the Federal Reserve introduced new programs and policies to stabilize markets, restore liquidity, and spur economic activity. However, a byproduct of these changes was that the fed funds market was dramatically altered, necessitating a new framework for monetary policy implementation. More recently, as the Fed began to unwind some of these programs, it was forced to reassess the long-run size of its balance sheet—and the tools it intended to use for monetary policy implementation—given the current economic and regulatory environment. It has chosen to maintain a balance sheet that is sufficiently large to support a market with ample reserves, and to use the administered (IOR and ON RRP) rates to achieve the target range. A lingering challenge is identifying the minimum balance-sheet size consistent with these goals, as this requires forecasting the evolution of the distribution of reserves across banks.
Notes

1 Although not all banks are depository institutions, and not all depository institutions are banks, we will use "bank" to refer to depository institutions trading in the fed funds market, including bank holding companies, standalone commercial banks, and thrifts. However, institutions other than banks also trade in the federal funds market. Under current regulation, once deposits exceed a minimal threshold, these banks are required to hold at least 10 percent of any additional deposits as reserves at the Fed.

2 Banks would try to avoid borrowing at the discount window because the rate was higher than the typical rate being offered in the fed funds market, and because there was a stigma associated with borrowing at the discount window. See Ennis and Weinberg (2013).


4 For more details on quantitative easing, see Yu (2018).

5 A basis point equals one hundredth of 1 percent. McCauley and McGuire (2014) estimate a cost of 4 basis points, while Banegas and Tase (2016) find a cost of 7 basis points.

6 This policy change was made possible when Congress passed the Financial Services Regulatory Relief Act in 2006, clearing the way for the Federal Reserve to start paying interest on reserves to eligible depository institutions effective October 1, 2011. This date was later moved up to October 1, 2008, as a result of the Emergency Economic Stabilization Act of 2008.

7 The Federal Home Loan Banks provide funds to depository institutions in the form of loans collateralized by real estate. They were initially set up to provide liquidity to savings and loans but are now a source of funds for all banks.

8 If the Fed tried to conduct policy on precrisis terms, it would have had to execute very large open market operations to drain reserves in relatively short order. Selling large quantities of certain assets in a very short period would have negative side effects, as prices in these markets would likely experience sudden declines.

9 For more information about eligible counterparties at the ON RRP facility, see https://www.newyorkfed.org/markets/rrp_counterparties.

10 In its June 14, 2017, statement, the FOMC announced only that “the Federal Reserve’s securities holdings will continue to decline in a gradual and predictable manner until the Committee judges that the Federal Reserve is holding no more securities than necessary to implement monetary policy efficiently and effectively.”


References


The views expressed in these papers are solely those of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of Philadelphia or Federal Reserve System.

Beautiful City: Leisure Amenities and Urban Growth

Modern urban economic theory and policymakers are coming to see the provision of consumer-leisure amenities as a way to attract population, especially the highly skilled and their employers. However, past studies have arguably only provided indirect evidence of the importance of leisure amenities for urban development. In this paper, we propose and validate the number of tourist trips and the number of crowdsourced picturesque locations as measures of consumer revealed preferences for local lifestyle amenities. Urban population growth in the 1990–2010 period was about 10 percentage points (about one standard deviation) higher in a metro area that was perceived as twice more picturesque. This measure ties with low taxes as the most important predictor of urban population growth. “Beautiful cities” disproportionally attracted highly educated individuals and experienced faster housing price appreciation, especially in supply-inelastic markets. In contrast to the generally declining trend of the American central city, neighborhoods that were close to central recreational districts have experienced economic growth, albeit at the cost of minority displacement.


Building Credit History with Heterogeneously Informed Lenders

This paper examines a novel mechanism of credit-history building as a way of aggregating information across multiple lenders. We build a dynamic model with multiple competing lenders, who have heterogeneous private information about a consumer’s creditworthiness, and extend credit over multiple stages. Acquiring a loan at an early stage serves as a positive signal—it allows the borrower to convey to other lenders the existence of a positively informed lender (advancing that early loan)—thereby convincing other lenders to extend further credit in future stages. This signaling may be costly to the least risky borrowers for two reasons. First, taking on an early loan may involve cross-subsidization from the least risky borrowers to more risky borrowers. Second, the least risky borrowers may take inefficiently large loans relative to the symmetric-information benchmark. We demonstrate that, despite these two possible costs, the least risky borrowers often prefer these equilibria to those without information aggregation. Our analysis offers an interesting and novel insight into debt dilution. Contrary to the conventional wisdom, repayment of the early loan is more likely when a borrower subsequently takes on a larger rather than a smaller additional loan. This result hinges on a selection effect: Larger subsequent loans are only given to the least risky borrowers.

The Firm Size and Leverage Relationship and Its Implications for Entry and Concentration in a Low Interest Rate World

Larger firms (by sales or employment) have higher leverage. This pattern is explained using a model in which firms produce multiple varieties and borrow with the option to default against their future cash flow. A variety can die with a constant probability, implying that bigger firms (those with more varieties) have lower coefficient of variation of sales and higher leverage. A lower risk-free rate benefits bigger firms more as they are able to lever more and existing firms buy more of the new varieties arriving into the economy. This leads to lower startup rates and greater concentration of sales.


Capitalization as a Two-Part Tariff: The Role of Zoning

This paper shows that the capitalization of local amenities is effectively priced into land via a two-part pricing formula: a “ticket” price paid regardless of the amount of housing service consumed and a “slope” price paid per unit of services. We first show theoretically how tickets arise as an extensive margin price when there are binding constraints on the number of households admitted to a neighborhood. We use a large national dataset of housing transactions, property characteristics, and neighborhood attributes to measure the extent to which local amenities are capitalized in ticket prices vis-à-vis slopes. We find that in most U.S. cities, the majority of neighborhood variation in pricing occurs via tickets, although the importance of tickets rises sharply in the stringency of land development regulations, as predicted by theory. We discuss implications of two-part pricing for efficiency and equity in neighborhood sorting equilibria and for empirical estimates of willingness to pay for nonmarketed amenities, which generally assume proportional pricing only.


Mortgage Loss Severities: What Keeps Them So High?

Mortgage loss-given-default (LGD) increased significantly when house prices plummeted and delinquencies rose during the financial crisis, but it has remained over 40 percent in recent years despite a strong housing recovery. Our results indicate that the sustained high LGDs postcrisis are due to a combination of an overhang of crisis-era foreclosures and prolonged foreclosure timelines, which have offset higher sales recoveries. Simulations show that cutting foreclosure timelines by one year would cause LGD to decrease by 5–8 percentage points, depending on the trade-off between lower liquidation expenses and lower sales recoveries. Using difference-in-differences tests, we also find that recent consumer protection programs have extended foreclosure timelines and increased loss severities in spite of their benefits of increasing loan modifications and enhancing consumer protections. Supersedes Working Paper 17-08.


Demographic Aging, Industrial Policy, and Chinese Economic Growth

We examine the role of demographics and changing industrial policies in accounting for the rapid rise in household savings and in per capita output growth in China since the mid-1970s. The demographic changes come from reductions in the fertility rate and increases in the life expectancy, while the industrial policies take many forms. These policies cause important structural changes; first benefiting private labor-intensive firms by incentivizing them to increase their share of employment, and later on benefiting capital-intensive firms resulting in an increasing share of capital devoted to heavy industries. We conduct our analysis in a general equilibrium economy that also features endogenous human capital investment. We calibrate the model to match key economic variables of the Chinese economy and show that demographic changes and industrial policies both contributed to increases in savings and output growth but with differing intensities and at different horizons. We further demonstrate the importance of endogenous human capital investment in accounting for the economic growth in China.

Commuting, Labor, and Housing Market Effects of Mass Transportation: Welfare and Identification

Using a panel of tract-level bilateral commuting flows, I estimate the causal effect of Los Angeles Metro Rail on commuting between connected locations. Unique data, in conjunction with a spatial general equilibrium model, isolate commuting benefits from other channels. A novel strategy interacts local innovations with intraurban geography to identify all model parameters (local housing and labor elasticities). Metro Rail connections increase commuting between locations containing (adjacent to) stations by 15 percent (10 percent), relative to control routes selected using proposed and historical rail networks. Other margins are not affected. Elasticity estimates suggest relatively inelastic mobility and housing supply. Metro Rail increases welfare $146 million annually by 2000, less than both operational subsidies and the annual cost of capital. More recent data show some additional commuting growth.


Consumer Lending Efficiency: Commercial Banks Versus a Fintech Lender

We compare the performance of unsecured personal installment loans made by traditional bank lenders with that of LendingClub, using a stochastic frontier estimation technique to decompose the observed nonperforming loans into three components. The first is the best-practice minimum ratio that a lender could achieve if it were fully efficient at credit-risk evaluation and loan management. The second is a ratio that reflects the difference between the observed ratio (adjusted for noise) and the minimum ratio that gauges the lender’s relative proficiency at credit analysis and loan monitoring. The third is statistical noise. In 2013 and 2016, the largest bank lenders experienced the highest ratio of nonperformance, the highest inherent credit risk, and the highest lending efficiency, indicating that their high ratio of nonperformance is driven by inherent credit risk, rather than by lending inefficiency. LendingClub’s performance was similar to small bank lenders as of 2013. As of 2016, LendingClub’s performance resembled the largest bank lenders—the highest ratio of nonperforming loans, inherent credit risk, and lending efficiency—although its loan volume was smaller. Our findings are consistent with a previous study that suggests LendingClub became more effective in risk identification and pricing starting in 2015. Caveat: We note that this conclusion may not be applicable to fintech lenders in general, and the results may not hold under different economic conditions such as a downturn.


Elasticities of Labor Supply and Labor Force Participation Flows

Using a representative-household search and matching model with endogenous labor force participation, we study the interactions between extensive-margin labor supply elasticities and the cyclicity of labor force participation flows. Our model successfully replicates salient business-cycle features of all transition rates between three labor market states, the unemployment rate, and the labor force participation rate, while using values of elasticities consistent with micro evidence. Our results underscore the importance of the procyclical opportunity cost of employment, together with wage rigidity, in understanding the cyclicity of labor market flows and stocks.

A Generalized Factor Model with Local Factors

I extend the theory on factor models by incorporating local factors into the model. Local factors only affect an unknown subset of the observed variables. This implies a continuum of eigenvalues of the covariance matrix, as is commonly observed in applications. I derive which factors are pervasive enough to be economically important and which factors are pervasive enough to be estimable using the common principal component estimator. I then introduce a new class of estimators to determine the number of those relevant factors. Unlike existing estimators, my estimators use not only the eigenvalues of the covariance matrix, but also its eigenvectors. I find strong evidence of local factors in a large panel of U.S. macroeconomic indicators.


Institution, Major, and Firm-Specific Premia: Evidence from Administrative Data

We examine how a student’s major and the institution attended contribute to the labor market outcomes of young graduates. Administrative panel data that combine student transcripts with matched employer-employee records allow us to provide the first decomposition of premia into individual and firm-specific components. We find that both major and institutional premia are more strongly related to the firm-specific component of wages than the individual-specific component of wages. On average, a student’s major is a more important predictor of future wages than the selectivity of the institution attended, but major premia (and their relative ranking) can differ substantially across institutions, suggesting the importance of program-level data for prospective students and their parents.

Forthcoming

Is the Phillips Curve Dead?

Collateral Damage: House Prices and Consumption During the Great Recession

Banking Trends: Have Foreign Banks Changed How They Operate?
You can find Economic Insights via the Research Publications part of our website.