Third Quarter 2013
Volume 96, Issue 3

BUSINESS REVIEW

Federal Reserve Bank of Philadelphia

INSIDE

ISSN: 0007-7011

The *Business Review* is published four times a year by the Research Department of the Federal Reserve Bank of Philadelphia. The views expressed by the authors are not necessarily those of the Federal Reserve. We welcome your comments at PHIL.BRComments@phil.frb.org.

For a free subscription, go to www.philadelphiafed.org/research-and-data/publications. Archived articles may be downloaded at www.philadelphiafed.org/research-and-data/publications/business-review. To request permission to reprint articles in whole or in part, click on Permission to Reprint at www.philadelphiafed.org/research-and-data/publications. Articles may be photocopied without permission. Microform copies may be purchased from ProQuest Information and Learning, 300 N. Zeeb Road, Ann Arbor, MI 48106.

The Federal Reserve Bank of Philadelphia helps formulate and implement monetary policy, supervises banks and bank and savings and loan holding companies, and provides financial services to depository institutions and the federal government. It is one of 12 regional Reserve Banks that, together with the U.S. Federal Reserve Board of Governors, make up the Federal Reserve System. The Philadelphia Fed serves eastern Pennsylvania, southern New Jersey, and Delaware.

THIRD QUARTER 2013

Media reports and policymakers' concerns about rising student loan balances and defaults have greatly intensified in recent years. **Wenli Li** sheds light on the economics behind these trends and discusses the implications for the broader economy.

Innovation, a key to economic growth, does not happen in a vacuum. Economists have studied the knowledge spillovers that occur when firms locate near one another. Yet, these spillovers have proved hard to empirically verify. **Gerald A. Carlino** and **Jake K. Carr** explain what they found through a more accurate way to measure the geographic concentration of research and development labs.

The failure and bailout of prominent financial institutions amid the crisis of 2007-09, and the effect these events had on the economy as a whole, have led policymakers to rethink how the global financial system is regulated. **Ronel Elul** explains the history behind bank capital regulation and how ongoing regulatory changes might help prevent future crises.

Abstracts of the latest working papers produced by the Research Department of the Federal Reserve Bank of Philadelphia.

The Economics of Student Loan Borrowing and Repayment

BY WENLI LI

Reports in the popular press and policymakers' concerns about student loans have greatly intensified in recent years because of rising student loan balances and defaults. Even greater cause for concern arose as student loans outstanding passed credit card debt to become the single largest nonmortgage household debt in 2012. Worries about the risk of massive default have even prompted a comparison with the subprime mortgage crisis.[1]

Existing theoretical and empirical work by economists on student loans can shed light on the economics behind this trend and, therefore, help provide answers to a number of important questions: What determines whether and how much a household borrows for student loans, and what determines whether and when a household repays these loans? What factors account for the widely noted increase in student loans outstanding and defaults? What are the implications of the trend for households' consumption and for the broader economy?

A SIMPLE THEORY OF STUDENT BORROWING AND REPAYMENT

What Makes Student Loans Different? Student loans are made solely for the purpose of financing higher education; that is, they are designed to help students pay for college tuition, books, and living expenses. They are different from other consumer loans, including credit card debt, auto loans, or mortgages; for those types of loans, households borrow to purchase goods they consume immediately, such as clothes, a car, or a house. Economists often view student loans as a means of financing investment in human capital. In other words, student loans help borrowers, through their college experience, to acquire knowledge as well as social and personal attributes that may enhance their ability to later perform in the economy and, thus, gain higher earnings.[2] It is in this sense that student loans are analogous to investment in physical capital such as an MRI machine purchased by a clinic. Unlike a pill given to a patient, the machine is not consumed immediately; rather it is used for future production (scanning patients), and with each use, it generates income from the fee a patient pays for each test.

Both Supply and Demand Factors Affect Student Borrowing. A household's decision to take out a student loan — the demand side — is obviously tied to its decision about whether to attend college. The majority of people in the U.S. go to college shortly, if not immediately, after high school. These people are often in their late teens or early 20s and lack the financial resources to pay for college, even with the help of their parents. Therefore, they need to borrow to cover the cost. Put simply, for a large fraction of the U.S. population, the decision about whether and when to take out a student loan is closely tied to the decision of whether, when, and

[1] For example, Steven Eisman titled his presentation on student loans at the Ira Sohn Conference "Subprime Goes to College."

[2] Of course, education serves other important purposes that are not captured by a narrow look at graduates' earning power, but in this article I focus solely on the economics of student loans.

Wenli Li is a senior economic advisor and economist at the Federal Reserve Bank of Philadelphia. The views expressed in this article are not necessarily those of the Federal Reserve. This article and other Philadelphia Fed research and reports are available at www.philadelphiafed.org/research-and-data/publications.

where to attend college. As a matter of fact, according to the *Chronicle of Higher Education*, about 60 percent of Americans who attend college borrow annually to cover costs.

As with any other economic decision, the decision of whether, when, and where to attend college depends on the difference between the benefits and the costs. The economic benefits of going to college are captured by the gain in future earnings, and the costs include the earnings a student forgoes while in school, in addition to tuition, books, and living expenses. Described this way, the prospective student's decision sounds very simple. But even if we imagine, as most economic analyses do, that the student has the ability to rationally calculate costs and benefits, the decision is actually fraught with uncertainty.

First, think about costs. While some of the costs — tuition, books, and living expenses — are immediately observable and are relatively easier to calculate and predict over, say, a two- or four-year period, real borrowing costs may fluctuate as interest rates and inflation rates fluctuate. In addition, students' forgone earnings may be very difficult to measure with any precision. The income gains from a college education are entirely in the future and need to be estimated and, thus, can be very imprecise. For example, a computer science major not only needs to figure out job prospects and prevailing salaries in four years' time, but he must also project job prospects and wages over the rest of his working life. To complicate the matter further, he also needs to factor in the possibility that he may end up disliking the field and taking up a different career with lower potential earnings.

The lender's decision — the supply side — would be relatively simple if students borrowed in a perfect capital market. The concept of a perfect capital market is an ideal benchmark used by economists, in which many real-world difficulties are assumed away. The concept is useful because it forces us to think carefully about the factors that may limit a student's capacity to borrow. In a perfect capital market, lenders can sign a contract that makes the payments conditional on borrowers' future earnings and can at no cost to themselves compel borrowers to work and earn enough to repay the loan. The factors that affect a lender's decision about whether to extend a student loan will thus be the opportunity cost of the funding (the interest the lender could have earned on other loans) and the riskiness of the gains (mainly due to the uncertainty about the borrower's income).

Two factors complicate our ideal world. First, human beings, not machines, are the ones producing earnings. In a civilized society, humans cannot serve as collateral because lenders cannot enslave borrowers, nor can they buy and sell them.[3] Second, although lenders can garnish borrowers' earnings when borrowers do not make payments, borrowers' earnings also depend on their effort. This is very different from machines, whose value depends mainly on their resale value, which is largely outside the control of the owners who use it as collateral. For example, a computer software engineer living in New Jersey can go to work for an investment house in New York City and make $60,000 a year with a commuting cost of $8,000 a year, or she can work for $50,000 for a local firm that has better work schedules and does not require any commute. Suppose the engineer has to give half of her income to the lender to service student loans. In the first case, it means that the engineer pays $30,000 to the lender and has $22,000 for herself after taking out commuting costs. In the second case, it means that the engineer pays $25,000 to the lender and the same amount to herself. The engineer will choose to work locally, since she makes the same amount of money in either case, but the lender will lose $5,000 if the engineer chooses to work in New Jersey rather than in New York City.

Over the years, the federal government has become the dominant supplier of student loans, first through its loan guarantee programs and more recently through direct loans.[4] *The Structure of the Student Loan Market* provides a brief discussion of the role of government in the student loan market. Therefore, a full account of the supply side of the market would require us to discuss the underlying political forces, since the total loan amount and interest rates are set by Congress. That is beyond the scope of this article.

The Repayment Decision. The student loan payment decision, like all other consumer loan payment decisions, depends on the borrower's ability to pay and the costs and benefits associated with default. The ability to pay depends on the borrower's income and assets. If a borrower loses his job or suffers a big loss in the stock market or a decline in the value of his primary residence, he may not be able to service his debt. The benefits of not paying one's student loans are the resources that are freed and that can be used for consumption purposes or to service other debt. Felicia Ionescu and Marius Ionescu show that households

[3] Prior to the mid-19th century, debtors' prisons were a common way to deal with unpaid debt. The father of the British writer Charles Dickens was sent to Marshalsea debtors' prison. As a result, Dickens used Marshalsea as the model for debtors' prison in his novels.

[4] Prominent arguments for government involvement are that social returns to education are greater than private returns. Furthermore, employers tend to underinvest in generalized training, since they do not fully capture the returns in the event the trained employees leave the firm.

The Structure of the Student Loan Market

There are three types of student loans: federally guaranteed loans made by banks and other lenders; federal loans made directly by the government; and private loans, which are essentially the same as other consumer loans from banks and companies. In the case of guaranteed loans, the government pays a subsidy to lenders that make the loans and also guarantees the amount loaned.*

Effective July 2010, in response to the changing market and the debate about the federal government's role in supporting student financing, Congress expanded federal aid to college students while ending federal subsidies to private lenders through loan guarantees.

The interest rate paid by students on both guaranteed loans and direct loans is fixed and set by Congress. The government pays the interest that accrues while the borrower is in school. Congress in 2007 temporarily reduced interest rates for low- and middle-income undergraduate borrowers to 3.4 percent from 6.8 percent until July 1, 2012. Congress then extended the freeze in interest rates until July 2013, at which time it pegged rates to the 10-year Treasury yield.

Private loans usually have worse terms than either type of federal loan, and interest rates on private loans can change over time. Because most students have limited credit histories, private lenders often require cosigners. The borrower is responsible for paying the interest that accrues.

* The top 10 holders of government guaranteed loans (FFELP loans) in the third quarter of 2010 were SLM Corporation, Nelnet, Wells Fargo, Brazos Group, JPMorgan Chase Bank, the Pennsylvania Higher Education Assistance Agency, College Loan Corporation, CIT, PNC, and Goal Financial. SLM Corporation had the largest market share (close to 60 percent), and each of the other institutions had under 10 percent of the market share.

have incentives to default on student loans first, before defaulting on credit card debt. By keeping their credit card account current, they can continue to use it as a transaction account or for borrowing purposes. Economists call this phenomenon "preserving liquidity."

The benefits from defaulting on student loans are, by contrast, limited. Unlike credit card debt, car loans, and other consumer loans, student loans cannot be discharged or reduced by a judge (known as "cramming down") under personal bankruptcy. Instead, borrowers who are late with their federal student loan payments have to enter into a repayment plan that can last 10 to 15 years, and during that time, a fraction of their earnings will be garnished, similar to what occurs in a Chapter 13 repayment plan under personal bankruptcy. The government can also garnish the borrower's tax returns and benefits. Other costs of defaulting on one's student loans include limited future access to the credit market, since the borrower's decision to default will affect his credit score from the credit bureau. Evidence from bankruptcy filers may give some sense of the order of magnitude of these costs. For instance, using data from the Federal Reserve's triennial Survey of Consumer Finances, Song Han and Geng Li find that bankruptcy filers are more than 40 percent less likely to have credit cards than comparable households that did not file for bankruptcy. If they do have cards, their lines of credit have far lower limits (by $12,000) compared with those who did not file for bankruptcy. Moreover, bankruptcy filers pay higher interest rates (1.2 percentage points higher) than people who did not file.[5]

With this theory in mind, we can now turn to the empirical evidence and discuss how and why student loans outstanding and defaults have increased sharply and the implications for the broader economy.

MORE TREND THAN CYCLE

Rising Student Loan Balances. The analysis here draws on the Federal Reserve Bank of New York (FRBNY)/Equifax Consumer Credit Panel dataset, a nationally representative random sample of anonymized credit reports from Equifax, one of three major consumer credit reporting agencies in the U.S., containing borrowers' ages, amounts borrowed, and repayment histories for bank and department store credit cards, car loans, mortgages, home equity loans, etc.[6]

Figure 1 shows the outstanding balances for various consumer loans, credit card debt, auto loans, home equity loans, and student loans. Note that I omit first mortgages because, unlike the other loans discussed here, first mortgages are of much larger value and collateralized.[7] Two observations are worth noting. Student loans have been trending up since the beginning of our sample period (the first quarter of 2003), and they did not come down until very recently. By comparison, credit card debt and auto loans did not

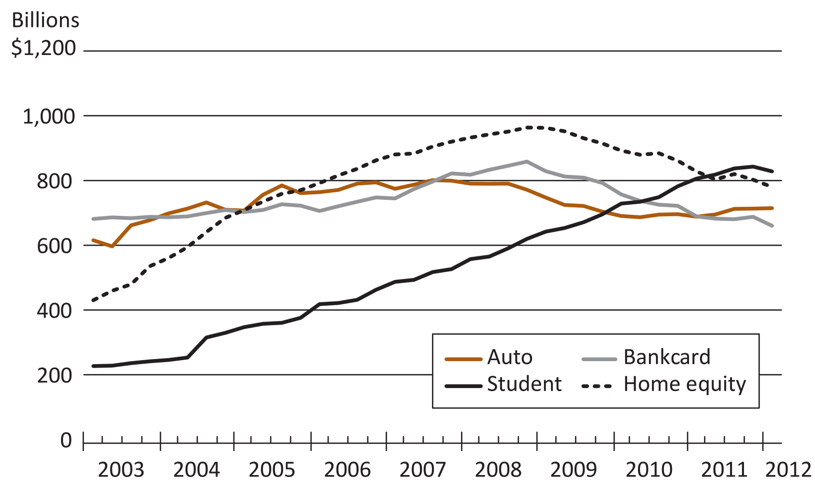
[5] It is likely that those who default on student loans will suffer a larger effect related to access to credit than bankruptcy filers. Bankruptcy wipes out some or all of a borrower's existing debts, a situation that is attractive to new lenders, who will not have to compete with old lenders to be repaid. But default does not wipe out student loans.

[6] The calculation is based on a 1 percent random sample of the FRBNY Consumer Credit Panel, while the panel accounts for about 5 percent of all households that have files with the credit bureau.

[7] Although car loans are also collateralized, cars depreciate much faster than houses. For most car loans, the resale value of the car is not the primary determinant of the loan terms.

FIGURE 1

Trend of Student Loan Balances vs. Other Loan Balances



Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

exhibit a comparable long-run trend, and their acceleration and deceleration coincided with the crisis. Home equity loans also experienced a long boom prior to the crisis. But balances came down immediately after the crisis, an immediate effect of the significant decline in house prices and the decline in households' equity in their homes.

The rise in student loan balances comes from the rise in both the number of people who borrowed and the amount each person borrowed. In contrast to other loans, the fraction of people with student loans has been increasing steadily over time and is now about 15 percent of the total population (Figure 2). The average student loan balance has also been moving up over the years for all age groups (Figure 3). In the first quarter of 2012, the average student loan balance for a 40-year-old was $30,000!

The Effects of Supply and Demand Factors. Although we cannot completely separate the effects of demand-side factors from supply-side factors, there are reasons to believe that both have contributed to the phenomenal rise in total student loans outstanding. On the demand side, estimates of the difference in lifetime earnings for those with college degrees versus only high school diplomas range from $650,000 to $1 million.[8] This is because a shift in the production technology over the past decade or two has favored skilled labor over unskilled labor by increasing skilled labor's relative productivity and hence its relative demand. For instance, the adoption of computers in the workplace has posed challenges for many workers. However, it is less costly for more educated, able, or experienced workers to learn to use computers and thus adapt to the new technology. The wage differential for educated workers has certainly not gone unnoticed by high school students deciding whether to enter the labor force. Indeed, more students are now accessing higher education than before. According to the Census Bureau, college enrollment as a fraction of the population between ages 16 and 25 rose from 34 percent in 1990 to 51 percent in 2010.

The rise in student loan borrowing per person reflects to a large extent the rising cost of higher education that has been going on for over a decade. According to the College Board, over the period 1997-98 to 2007-08, published tuition and fees for full-time in-state students at public four-year colleges and universities rose 54 percent in inflation-adjusted dollars — an average of 4.4 percent per year;[9] those for full-time students at two-year colleges and universities rose 17 percent in real terms — 1.5 percent annually; published tuition and fees for full-time students at public two-year colleges and universities rose 33 percent in real terms, 2.9 percent annually. Reduced funding from government is partially responsible for the rise in tuition and fees. According to the annual Grapevine Study, conducted by Illinois State University's Center for the Study of Education Policy with the cooperation of the State Higher Education Executive Officers, state appropriations for colleges and students sank by 7.6 percent in 2011-12, the largest such decline in at least half a century.

Finally, declines in family resources following the recent financial crisis have also driven up demand for student loans in the past five years. According to the Survey of Consumer Finances, between 1998 and 2007,

[8] See the paper by Anthony P. Carnevale, Stephen J. Rose, and Ban Cheah, Keith Sill's *Business Review* article on the skill premium, and http://www.pewsocialtrends.org/2011/05/16/lifetime-earnings-of-college-graduates/.

[9] In economics, the nominal value of something is its money value in different years. By contrast, real values adjust for differences in price levels of those years. As a result of the adjustment, any differences in real values are then attributed to differences in the amount of goods that money income could buy in each year.

while real median household income fell 3.9 percent, real median household net worth went up by 10 percent. Between 2007 and 2010, however, real median household income fell 11 percent, and median household net worth fell 39 percent over that same period.

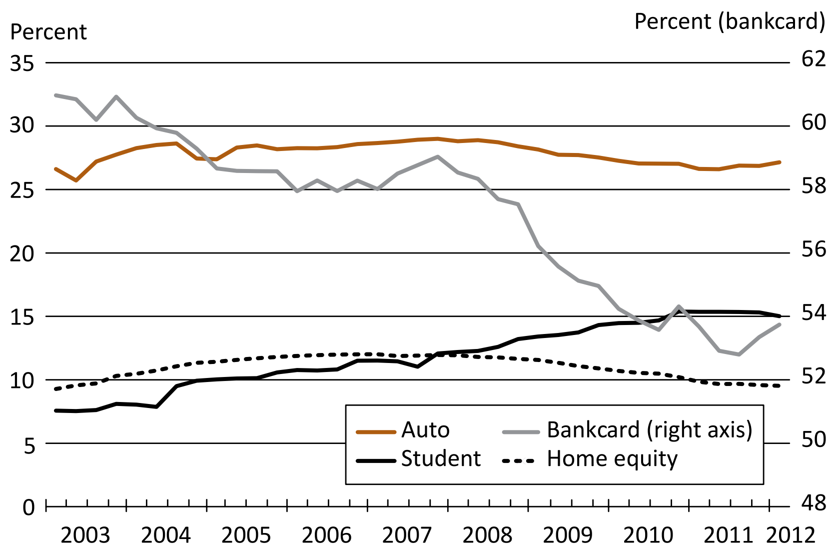
On the supply side, the U.S. government has played an increasingly important role in extending student loans that are cheaper than those the private market would offer, thus crowding out banks from the lending market (Figure 4). Furthermore, starting in July 2010, the government replaced loan guarantees with direct loans and effectively ended all subsidies to private lenders. According to the Department of Education, Federal Student Aid, an office of the department, managed or oversaw $713 billion in student loans in 2011, which accounts for close to 90 percent of the market. Most college students qualify for federal student loans. Students can borrow the same amount of money, at the same loan rate, regardless of their own income or their parents', regardless of their expected future income, and regardless of their credit history. Only students who have defaulted on federal student loans or have been convicted of drug offenses are excluded.

Trends in Past Due and Delinquent Loans. The trend in loans past due closely mirrors the rise in loans outstanding (Figure 5).[10] The total amount of past dues has been trending up since the beginning of our sample period, although the increase in past dues accelerated after 2007. This is again in contrast to the total amount of past dues of other consumer loans, which exhibit more of a cyclical pattern; that is, the amount of past dues for all other consumer loans was more or less flat until right around the crisis. Moreover, after 2009, the past due amount came down for all consumer debt except student loans.

The movement of delinquency rates tells a similar story (Figure 6). In terms of population, the delinquency rate on student loans has exceeded the delinquency rates on all three other types of consumer loans. My

FIGURE 2

Percent of Indebted Households by Loan Type

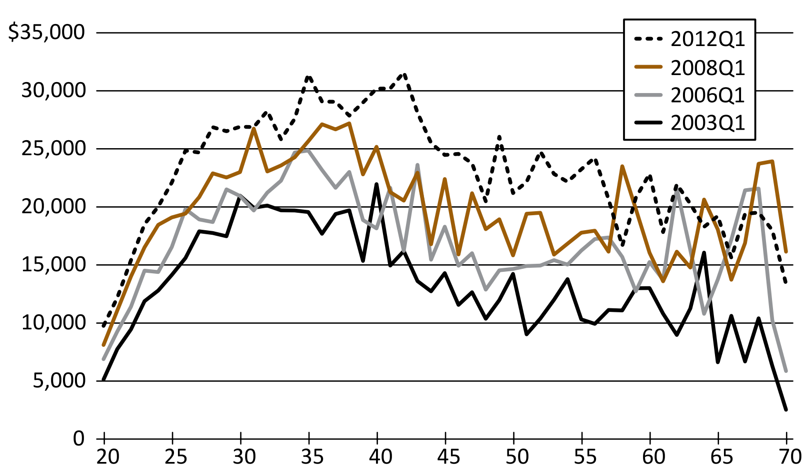


Note: Households includes those with credit histories on file.

Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

FIGURE 3

Average Student Loan Balance by Age

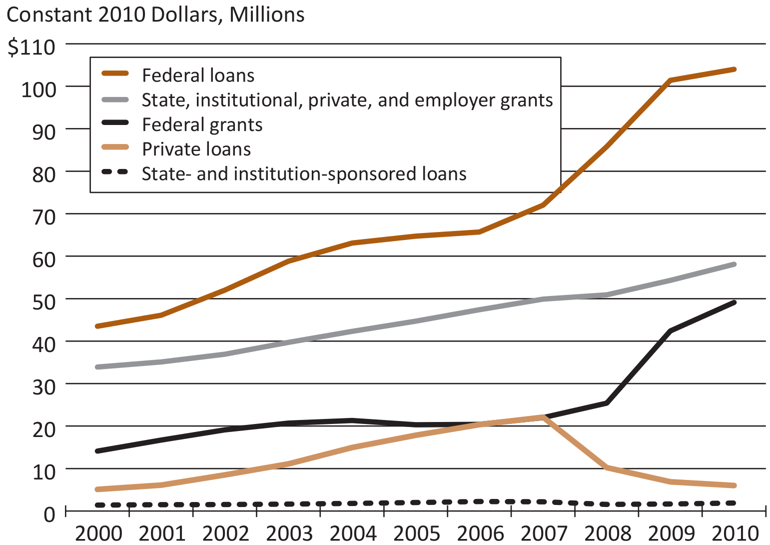


Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

[10] For private student loans, past dues are those with one missed payment. For government loans, past dues may include those with multiple missed payments because of their 270-day grace period.

FIGURE 4

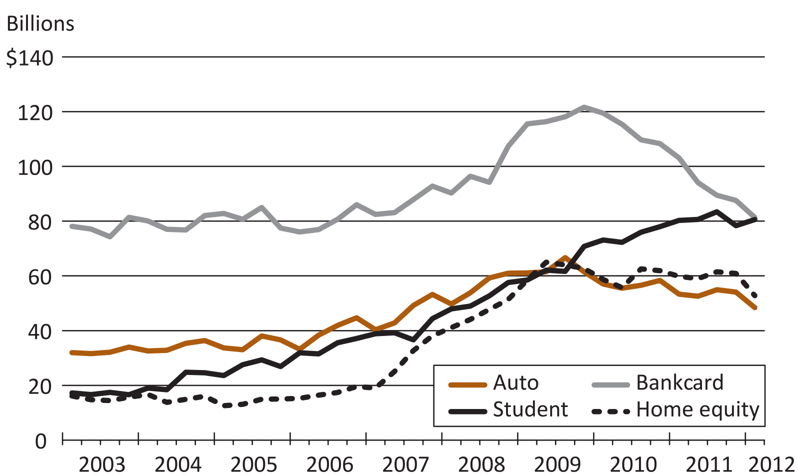
Federal and Nonfederal Student Loans and Grants



Source: The College Board

FIGURE 5

Past Due Balances on Consumer Loans



Note: Includes loans 30 days or more delinquent or charged off.

Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

estimate of a 14 percent to 15 percent student loan delinquency rate that we observed in 2012 is probably a lower bound for the actual delinquency rates for student loans. Other estimates by economists at the New York Fed put the delinquency rate as high as 26 percent.[11] Data limitations require the analyst to make assumptions, which I discuss further in the adjacent explanation, *Calculating Student Loan Delinquency Rates*.

Given the long-run factors that have increased the demand for higher education and the factors driving up college costs, in tandem with the slower rise in household incomes, it is not surprising that we saw a rise in student loan defaults long before the start of the crisis. The ensuing economic recession, in particular the weak labor market, nevertheless further drove up the defaults in student loans, as it did with most other consumer loans. For younger adults, particularly those in their 20s, who often hold student loans, the unemployment rates have been especially high (about 16 percent). Finally, part of the rise in student loan delinquency rates may also stem from portfolio adjustments as borrowers stop their student loan payments in order to keep their credit card payments current to preserve liquidity, as I discussed earlier.

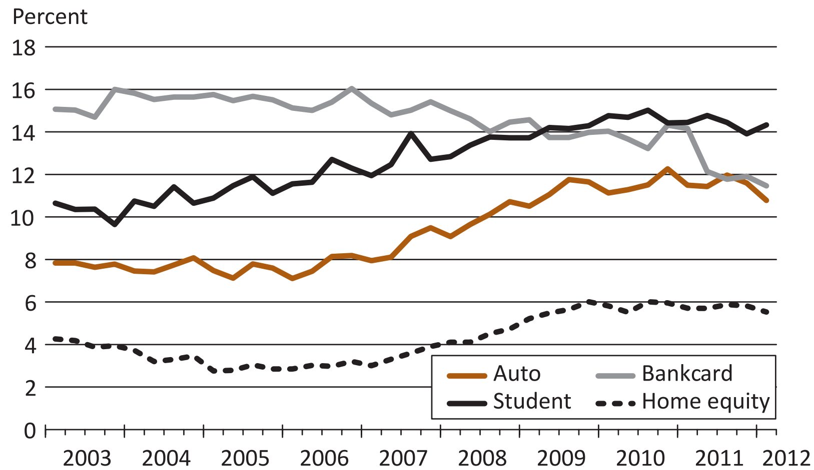
BROAD ECONOMIC IMPACT

Aggregate statistics and averages often mask substantial differences at the individual level. To gain further insight, it is often necessary to examine the differences among individuals in a more disaggregated way. These individual differences can lead to very different policy prescriptions. For example, suppose we find that very young people owe all of the loans and

[11] See the article by Meta Brown, Andrew Haughwout, Donghoon Lee, Maricar Mabutas, and Wilbert van der Klaauw.

FIGURE 6

Consumer Loan Delinquency Rates



Note: Includes charged-off loans.

Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

Calculating Student Loan Delinquency Rates

The calculation of student loan delinquency rates is somewhat involved due to the unique market structure of student loans. The key difficulty lies in the fact that the credit bureau data do not have information on whether a household needs to make student loan payments in the current quarter. The reason is that with federal loans, there is typically a six- or nine-month grace period, depending on the type of loan, after a borrower leaves school during which the borrower does not have to make payments on his loans. We do not want to count these borrowers in the denominator when calculating the default rate, which is defined as the ratio of the number of borrowers who are behind on their student loan payments over the number of borrowers who need to make student loan payments.

One way to get around this issue is to follow the New York Fed's approach* and exclude individuals who owed as much as or more than they did in the previous quarter while maintaining a zero past-due balance. The rationale behind this approach is that presumably those whose balance did not change across two quarters and who did not have student loan past dues do not need to make payments on their student loans yet. If I use this strategy, then the delinquency rates are much higher. For instance, 26 percent of borrowers would have past-due balances in the first quarter of 2012 by this calculation as opposed to 14 percent. However, this method is not perfect. For example, it might miss borrowers who negotiated smaller payments with their lenders through an income-based repayment plan. If their new payments are too low to cover accruing interest, their balances would be higher rather than lower. We wouldn't count these borrowers as being in delinquency using the proposed method even though they clearly need to be there.

* See the article by Brown and coauthors.

that they are the ones defaulting. In this case, we might argue that there is less cause for concern because young people have a long horizon over which to work out their situation. And the policy prescription may be to design programs to help these people find jobs or find better jobs. On the other hand, suppose a large fraction of loans are held by 50-year-olds and that these older households are defaulting in significant numbers. In this case, we might be much more concerned, since these people have much shorter horizons over which to recover from their financial difficulty. The corresponding policy prescription may require some degree of loan forgiveness.

To address questions like these, I reexamine student loan balances, past dues, and default rates by borrowers' age using the FRBNY/Equifax Consumer Credit Panel. Two main observations emerge from the analysis.

First, over time, average student loan balances have increased for all age groups, but more for those between ages 30 and 55. Furthermore, it appears to take longer to pay off loans than in the past. For example, in the first quarter of 2012 the decline in average balances really started after age 32, as opposed to the late 20s in the first quarter of 2003 (Figure 7). Balances didn't stabilize until age 45 in the first quarter of 2012, as opposed to the late 30s in the first quarter of 2003 (Figure 7).[12] Second, the trend toward older households with significant amounts of student debt is confirmed if we look at the fraction of people who have student loans by age. Those between ages 25 and 45 had the larg-

[12] A small part of the balance is accounted for by cosigned loans, and, as expected, cosigned student loans have two peaks: at age 25 (less than 10 percent of the total balance at that age) and at age 55 (less than 20 percent of the total balance). At age 25, borrowers have their parents as cosigners. At age 55, they most likely act as cosigners for their children.

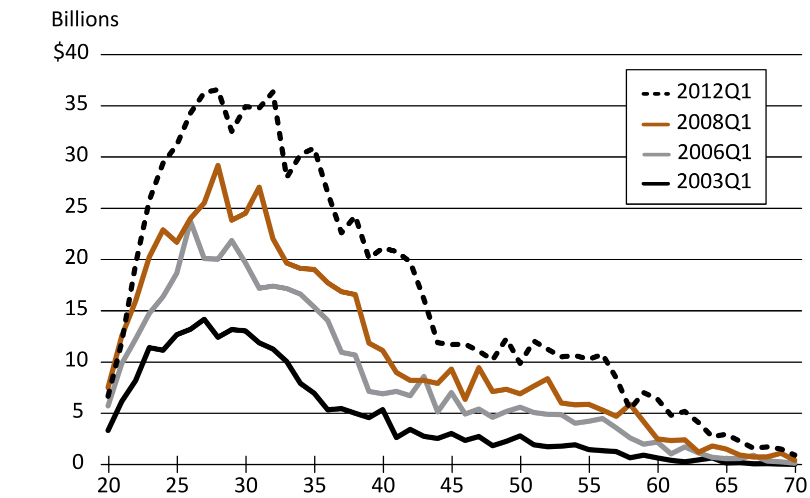
est increase. These two observations are striking, since they indicate that student loans are not just an issue for young borrowers as conventional wisdom perceives, but that the middle-aged (those 40 and above) actually shoulder a lot of the burden.[13]

An examination of the total amount of past dues by age confirms that it is indeed the middle-aged who are struggling with their student loan repayments (Figure 8).[14] To some extent, this trend is not surprising, since the growth in student loans has outstripped the growth in income for some time, as discussed earlier. The housing crisis obviously exacerbated the situation by further reducing households' net worth.[15]

Looking just at average borrowings obscures the fact that there are also substantial differences in the amount they borrowed. A high average balance might mean that the typical individual's balance is high. At the same time, it could mean that most individuals have very low balances, while a relatively small number of individuals have very burdensome debt levels. One way to think about this is to consider the difference between the mean and the median. The mean is simply the average: the total amount divided by the number of people. The median is the amount at which half of the population has more and half has less. A classic example to illustrate the difference between mean and median is that after Bill Gates walks into a bar that already has four unemployed people whose income is zero, everyone in the bar is, on average, a millionaire, since the mean income is over $1 million but the median is still zero (since half of them are still unemployed).[16]

FIGURE 7

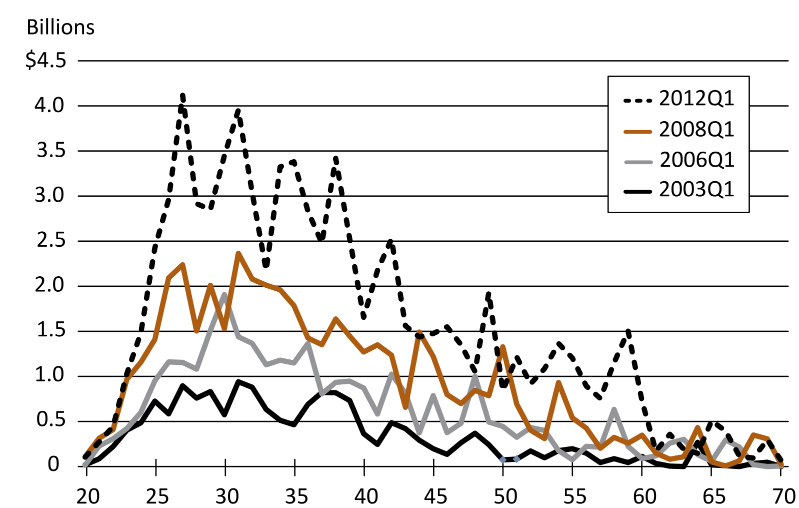
Student Loan Balances by Age of Borrowers



Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

FIGURE 8

Past Due Student Loan Balances by Age of Borrowers



Source: Federal Reserve Bank of New York/Equifax Consumer Credit Panel

[13] This may be due to a trend in the proportion of parents cosigning on loans while they are still paying down their own. Identifying this would require analyzing the individual trade lines, which appears to be out of scope for this paper.

[14] Brown and coauthors have also documented similar findings in their 2012 article.

[15] The harder question that we cannot pinpoint with the data is why so many people are still borrowing so much to finance their education. It could be that individuals are slowly learning about the change (lower) in expected income. Or it could simply be that receiving an education is a decision that involves a lot more than just having a higher income in the future.

[16] See a different version of the story at http://introductorystats.wordpress.com/2011/09/04/when-bill-gates-walks-into-a-bar/.

Although not as extreme, in our data, in the first quarter of 2012, the median balance at age 35 is $14,000, while the mean is close to $25,000. About 10 percent of borrowers have balances over $56,000, and 5 percent of the households have student loan balances over $81,000, suggesting that a relatively small number of households are seriously burdened by their debt level.

The Broad Economic Implications. One of the major concerns about ballooning student loans and student loan defaults is that these loans will have a negative impact on borrowers' consumption, since the borrowers need to devote a large fraction of their income to making loan payments. Furthermore, those who default on student loans will have more restricted access to credit because their credit scores will be lower. For credit-constrained families, such as those who need to borrow to buy a car, repair a roof, etc., this drop in credit scores may make all of this additional consumption infeasible. Indeed, the credit card utilization rate (credit card balance divided by credit limit) for those with student loan balances over $56,000 is 55 percent, compared with 39 percent for the general population in the first quarter of 2012. Economists have found that high credit card utilization rates are indicators of liquidity or income shocks.[17]

Andrew Glover, Jonathan Heathcote, Dirk Krueger, and Jose-Victor Rios-Rull show that older people will fare worse than the young after the recent financial crisis, since they do not have as long a horizon as the young to recover from the losses they have suffered: loss in income, loss in stock market investment and, more important, loss in their housing asset. My finding that middle-aged and older households are much more indebted by student loans than they used to be (the mean age of those with student loan balances over $56,000 is 38 years old, and the median age is 36 years old) and to a surprising extent before the crisis suggests that if we take student loan borrowing into consideration, middle-aged and older people may be even worse off.

Aside from these immediate economic concerns, researchers have found some longer-term social concerns. For example, researchers have found evidence that high debt burdens make students less likely to choose lower-paying careers such as teaching. Jesse Rothstein and Cecilia Rouse study a "natural experiment" generated by a change in financial aid policy by a highly selective university. The university introduced a "no loans" policy, in which it replaced the loan component of financial aid awards with grants. Interestingly, they find that debt causes graduates to choose jobs with substantially higher salaries, such as those in finance and consulting, and reduces the probability that students choose low-paid "public interest" jobs such as grade-school teacher or social worker.[18]

Additionally, Dora Gicheva suggests that each $10,000 in additional student debt decreases the borrower's long-term probability of marriage by 7 percentage points.[19] A 2010 poll found that 85 percent of college graduates were planning to move back home after graduation (Dickler 2010). The high unemployment rates and low income of new graduates are the leading causes behind these survey results. But having large student loans can certainly make things worse. Although currently there are more open questions than settled answers regarding the extent to which student loans hurt the formation of households, there is no doubt that reduced household formation has obviously hurt the recovery of the nation's housing market. According to the Census Bureau, the homeownership rate of those under age 35 declined from its 2006 peak of 42.6 percent to 36.8 percent in the first quarter of 2012. By comparison, the overall homeownership rate came down only 3.4 percentage points, from 68.8 percent to 65.4 percent. Of course, the reduced homeownership rates for the young also reflect their increased credit constraints that are not related to household formation. Further research is called for.

CONCLUSION

The substantial increase in student loans in recent years is a continuation of a trend that started a decade ago due to technological innovation. But the trend was exacerbated by the Great Recession. As households experienced significant contractions in income and wealth, housing wealth in particular, and as jobs became scarce, more students had to borrow increasingly large amounts to fund their educations. Moreover, student loans became delinquent as borrowers' payment ability declined. This article suggests that any policy to address student loans needs to target both secular and cyclical factors.

---

[17] See the article by Ronel Elul, Nicholas Souleles, Souphala Chomsisengphet, Dennis Glennon, and Bob Hunt.

[18] Two features of the policy change make this a natural experiment. First, the change was unexpected. This means that any change in students' employment choices was not affected by some expected change in financing policies. In addition, the change in a student's debt load was caused by a decision by the university, rather than a decision by the student. This means that it was the change in debt load that induced the change in students' employment choices, rather than the other way around. As with most natural experiments, though, the precise answers come at some cost to generality. Among other questions, it is natural to ask whether the behavior of students at a highly selective university is indicative of the behavior of students more generally.

[19] To deal with the issue that those with high student loan balances may be those who have less intention of forming a household in the first place, Gicheva uses exogenous changes in limits and eligibility of federal loans as instruments.

REFERENCES

Brown, Meta, Andrew Haughwout, Donghoon Lee, Maricar Mabutas, and Wilbert van der Klaauw. "Grading Student Loans," Federal Reserve Bank of New York, *Liberty Street Economics* (March 2012).

Carnevale, Anthony P., Stephen J. Rose, and Ban Cheah. "The College Payoff: Education, Occupations, and Lifetime Earnings," Georgetown University Center on Education and the Workforce, Washington, D.C. (2011).

Dickler, Jessica. "Boomerang Kids: 85% of College Grads Move Home," CNNMoney (November 15, 2010); http://money.cnn.com/2010/10/14/pf/boomerang_kids_move_home/index.htm.

Eisman, Steven. "Subprime Goes to College," (2010); http://www.scribd.com/doc/32066986/Steve-Eisman-Ira-Sohn-Conference-May-2010.

Elul, Ronel, Nicholas Souleles, Souphala Chomsisengphet, Dennis Glennon, and Robert M. Hunt. "What Triggers Mortgage Default?" *American Economic Review* 100:2 (2010), pp. 490-494.

Gicheva, Dora. "In Debt and Alone? Examining the Causal Link Between Student Loans and Marriage" (2013), manuscript, University of North Carolina at Greensboro.

Glover, Andrew, Jonathan Heathcote, Dirk Krueger, and Jose-Victor Rios-Rull. "Intergenerational Redistribution in the Great Recession," University of Pennsylvania Working Paper (2012).

Han Song, and Geng Li. "Household Borrowing After Personal Bankruptcy," *Journal of Money, Credit and Banking*, 43 (2011), pp. 491-517.

Illinois State University. "Fiscal Year 2011-2012," *Grapevine*, ISU Center for the Study of Education Policy; http://grapevine.illinoisstate.edu/.

Ionescu, Felicia, and Marius Ionescu. "The Interplay between Student Loans and Credit Cards and Amplification of Consumer Default," Colgate University Working Paper (2011).

Rothstein, Jesse, and Cecilia Rouse. "Constrained after College: Student Loans and Early Career Occupational Choices," *Journal of Public Economics*, 95:1-2 (2011), pp. 149-63.

Sill, Keith. "Widening the Wage Gap: The Skill Premium and Technology," Federal Reserve Bank of Philadelphia *Business Review* (Fourth Quarter 2002), pp. 25-32.

Clusters of Knowledge: R&D Proximity and the Spillover Effect

BY GERALD A. CARLINO AND JAKE K. CARR

The United States is home to some of the most innovative companies in the world, such as Apple, Facebook, and Google, to name a few. Inventive activity depends on research and development, and R&D depends on, among other things, the exchange of ideas among individuals. People's physical proximity is a key ingredient in the innovation process. Steve Jobs understood this when he helped to design the layout of Pixar Animation Studios. The original plan called for three buildings, with separate offices for animators, scientists, and executives. Jobs instead opted for a single building with a vast atrium at its core. To ensure that animators, scientists, and executives frequently interacted and exchanged ideas, Jobs moved the mailboxes, the cafeteria, and the meeting rooms to the center of the building.

There is nothing really new in the recognition that face-to-face contact among individuals is one key to innovation. Mervin Kelly, who for a time ran AT&T's legendary Bell Labs, was, according to a *New York Times* article, "convinced that physical proximity was everything."[1] According to the article, Kelly personally helped to design a building that opened in 1941 "where everyone would interact with one another." Hallways were designed to be so long that when walking a hall's length one would encounter "a number of acquaintances, problems, diversions and ideas. A physicist on his way to lunch in the cafeteria was like a magnet rolling past iron filings." Within this unique culture, Bell Labs' employees developed some of the most important inventions of the 20th century, including the transistor, the laser, and the solar cell.

Most American companies are small in size, and they obviously lack the resources of companies such as Apple, Facebook, and Google. Does their small size deprive these firms of the benefits of knowledge spillovers — the continuing exchange of ideas among individuals and firms — that physical proximity provides? The answer appears to be no. There is an exceptionally high spatial concentration of individual R&D labs in the Northeast corridor, around the Great Lakes, in Southern California, and in California's Bay Area. The high geographic concentration of R&D labs creates an environment similar to that found at Bell Labs, in which ideas move quickly from person to person and from lab to lab.[2] This exchange of ideas underlies the creation of new goods and new ways of producing existing goods.

In this article, we will discuss a recent study that we coauthored with Robert Hunt and Tony Smith. That

Gerald A. Carlino is a senior economic advisor and economist specializing in regional analysis at the Federal Reserve Bank of Philadelphia. **Jake K. Carr** is a former economic analyst in the Research Department of the Federal Reserve Bank of Philadelphia. The views expressed in this article are not necessarily those of the Federal Reserve. This article and other Philadelphia Fed research and reports are available at www.philadelphiafed.org/research-and-data/publications.

[1] Jon Gertner, "True Innovation," *New York Times*, February 25, 2012.

[2] Knowledge spillovers are the *unintended* transmission of knowledge that occurs among individuals and organizations. For example, as pointed out by AnnaLee Saxenian, although there is intense competition in California's Silicon Valley, a remarkable level of knowledge spillovers occurs.

study has two main goals. First, our study introduces a more accurate way to measure the extent of the spatial concentration of R&D activity. This new approach allows us to document the spatial concentration of more than 1,000 R&D labs in the Northeast corridor of the U.S. An important finding that emerged from this approach is that the clustering of labs is by far most significantly concentrated at very small spatial scales, such as distances of about one-quarter of a mile, with significant clustering attenuating rapidly during the first half-mile. The rapid attenuation of significant clustering is consistent with the view that knowledge spillovers are highly localized.

We also observe a secondary node of significant clustering at a scale of about 40 miles. This secondary node of clustering is interesting because its spatial scale is roughly the same as that of the local labor market. That is, firms will draw most of their workers and most residents will commute to jobs within 40 miles. Hence, this scale is consistent with the view that the efficiency gains and cost savings at the labor market level (e.g., better matching of workers' skills to the needs of labs) are important for innovative activity.

A second goal of our study is to provide evidence on the extent to which knowledge spillovers are geographically localized within the R&D clusters we identify. Data on patent citations have been used to track knowledge spillovers. Patents contain detailed geographic information about the inventors as well as citations to prior patents on which the inventions were built. If knowledge spillovers are localized within the clusters that we identify, then citations of patents generated within a cluster should come disproportionately from within the same cluster as previous patents. We find that citations are a little over four times more likely to come from the same cluster as earlier patents than one would expect based on the preexisting concentration of technologically related activities.

LEARNING IN CLUSTERS

An enormous increase in the material well-being of individuals has been achieved over the past 200 to 300 years. We not only have more of the same goods and services but also a variety of new goods and services — such as the personal computer, the Internet, and cellular phones — whose specific characteristics could not have been imagined just 50 years ago. It took an accumulation of knowledge to design and build these goods and services and bring them to market. Inventions or innovations do not happen in a vacuum but instead are created by individuals working together to solve common problems. Often, new knowledge is tacit knowledge, that is, knowledge that is highly contextual and difficult or even impossible to codify or electronically transmit.

Beginning with Alfred Marshall, economists have studied the benefits that individuals and firms gain from locating near one another, in what are referred to as *agglomeration economies*. Knowledge spillovers, an important aspect of agglomeration economies, have proved hard to empirically verify. The empirical evidence on knowledge spillovers is rather sparse. What the limited research suggests is that the transmission of knowledge rapidly diminishes the farther one gets from the source of that knowledge. Looking at innovative activity, Adam Jaffe, Manuel Trajtenberg, and Rebecca Henderson and, more recently, Ajay Agrawal, Devesh Kapur, and John McHale find that nearby inventors are much more likely to cite each other's inventions in their patents, suggesting that knowledge spillovers are indeed localized. Mohammad Arzaghi and Vernon Henderson look at the location pattern of firms in the advertising industry in Manhattan. They show that for an ad agency, knowledge spillovers and the benefits of networking with nearby agencies are extensive, but the benefits dissipate quickly with distance from other ad agencies and are gone after roughly one-half mile.

An important finding that emerged from our new approach is that the clustering of labs is by far most significantly concentrated at very small spatial scales, such as about one-quarter of a mile.

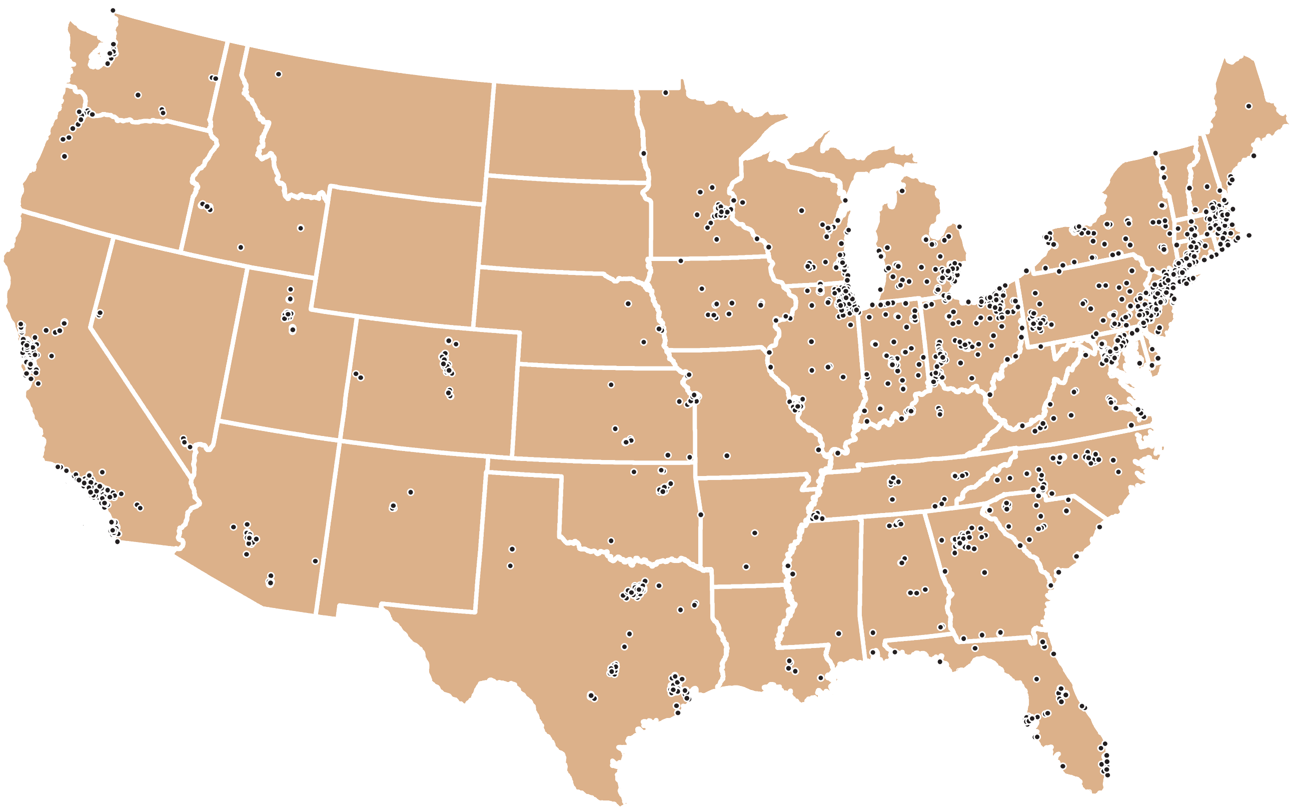
More than most economic activity, innovative activity such as R&D depends on knowledge spillovers. R&D labs will have an incentive to locate near one another if knowledge spillovers tend to dissipate rapidly with increasing distance from the source of that knowledge.

A map of the spatial distribution of R&D labs reveals a striking clustering of R&D activity (Figure 1). In places that have little R&D activity, each dot on the map represents the location of a single R&D lab. For example, there is only one lab in Montana, represented by the single dot. In counties with a dense clustering of labs, the dots tend to sit on top of one another, representing a concentration of labs.

FIGURE 1

Location of R&D Labs

Each dot on the map represents the location of a single R&D lab in 1998. In areas with dense clusters of labs, the dots tend to sit on top of one another.



Sources: Directory of American Research and Technology and authors' calculations

A prominent feature of the map is the high concentration of R&D activity in the Northeast corridor, stretching from northern Virginia to Massachusetts. There are other concentrations, such as the cluster around the Great Lakes and the concentration of labs in California's Bay Area and in Southern California.

The high geographic concentration of R&D labs creates an environment in which ideas move quickly from person to person and from lab to lab. Locations that are dense in R&D activity encourage knowledge spillovers, thus facilitating the exchange of ideas that underlie the creation of new goods and new ways of producing existing goods. The tendency for innovative activity to cluster raises a number of interesting and important questions. How strong is the tendency for R&D labs to cluster? Where in space do these labs cluster, and what are the geographic sizes of these clusters? How rapidly does the mutual attraction among labs attenuate with distance? Providing answers to these questions is an important objective of our study with Hunt and Smith.

MEASURING CLUSTERING OF ECONOMIC ACTIVITY

Although R&D labs tend to be spatially concentrated, a similar pattern of geographic concentration would be found for either population or employment. Thus, studies that look at the concentration of R&D labs need to control for the general tendency for economic activity and population to cluster spatially. In a 1996 study, David Audretsch and Maryann Feldman introduced the "locational Gini coefficient" to show that innovative activity at the state level tends to be considerably more concentrated than is manufacturing employment and that industries that stress R&D activity also tend to be more spatially concentrated.[3]

Glenn Ellison and Edward Glaeser have identified a potential problem with the Audretsch and Feldman study. They argue that an industry may appear to be spatially concentrated if that industry consists of a few large firms. In this instance, the industry would be classified as industrially concentrated but not necessarily spatially concentrated. Ellison and Glaeser developed an alternative measure of spatial concentration — called the EG index — that controls both for the overall concentration of economic activity and for the industrial organization of the industry. Typically, the EG index has been used to gauge the geographic concentration of various manufacturing industries with fixed spatial boundaries, such as states, metropolitan areas, and counties.[4]

The EG index suffers from a number of important aggregation issues that result from using fixed spatial boundaries. For example, when calculating EG indexes at the county level, researchers will not take into account any activity that crosses county borders. As a result, measures of spatial concentration will be underestimated for counties. For example, Philadelphia County shares a border with Montgomery County. One stretch of City Avenue divides these two counties. Economic activity on the Philadelphia side of City Avenue is allocated to Philadelphia County, while activity on the Montgomery County side is assigned to that county. But this partition of economic activity is artificial, since this activity is really part of the same cluster. As a result, concentration will be underestimated for both counties. To avoid problems associated with fixed spatial boundaries, authors of several recent studies have used geocoded data to identify the exact location of establishments. These studies base their approach on the actual distance between establishments and are, therefore, not bound by a fixed geographical classification.[5]

MEASURING THE CLUSTERING OF R&D LABS

In our study, we used 1998 data from the Directory of American Research and Technology to electronically code the R&D labs' addresses and other information. Since the directory lists the complete address for each establishment, we were able to assign a geographic identifier (using geocoding techniques) to more than 3,100 R&D labs in the U.S. in 1998. We limited our analysis to 1,035 R&D labs in the 10 states (Connecticut, Delaware, Maryland, Massachusetts, New Hampshire, New York, New Jersey, Pennsylvania, Rhode Island, and Virginia) and the District of Columbia that make up the Northeast corridor of the United States.

A key question we need to determine is whether an observed spatial collection of labs in this corridor is somehow unusual; that is, is it different from what we would expect based on the spatial concentration of manufacturing employment? We used manufacturing employment instead of manufacturing firms as our benchmark.[6] In our study, we start with a "global" measure of concentration that is based on the observed concentration of R&D labs at various distances, ranging from a quarter-mile to 100 miles. For example, suppose we want to calculate the average number of labs that are located within a quarter-mile radius of

[3] A locational Gini coefficient shows how similar (or dissimilar) the location pattern of employment (or innovative activity, in Audretsch and Feldman's case) in a particular manufacturing industry is to the location pattern of overall manufacturing employment. The larger the value found for the locational Gini, the more concentrated is employment (or innovative activity) in a particular industry relative to overall manufacturing employment. See the *Business Review* article by Kristy Buzard and Gerald Carlino for a discussion of the construction of the locational Gini coefficient. The study by Audretsch and Feldman looked at the spatial concentration of innovative activity by industry. Their analysis, which is at the state level, uses 1982 census data provided by the United States Small Business Administration. They construct a data set on innovations by state and industry that is culled from information on new product announcements in over 100 scientific and trade journals.

[4] For examples of studies that use the EG index, see the studies by Ellison and Glaeser; Stuart Rosenthal and William Strange; and Glenn Ellison, Edward Glaeser, and William Kerr. See the *Business Review* article by Buzard and Carlino for a discussion of the EG index.

[5] Another problem is that authors of studies based on the EG index often provide only indexes of localization, without any indication of the statistical significance of their results. Without such statistical analyses, it is unclear whether the concentrations found differ from concentrations that would have been found if the locations of economic activity were randomly chosen. See the article by Gilles Duranton and Henry Overman for a discussion of statistical issues with the EG index.

[6] The concentration of R&D establishments is measured relative to a baseline of economic activity as reflected by the amount of manufacturing employment in the Zip code, as reported in the 1998 vintage of Zip Code Business Patterns. Since one of our objectives is to describe the localization of total R&D labs, manufacturing employment represents a good benchmark because most R&D labs are owned by manufacturing firms. We elected to use manufacturing employment as our benchmark rather than the number of manufacturing establishments in a Zip code, since past studies (such as the study by Audretsch and Feldman) use manufacturing employment as their benchmark. When we look at the clustering of R&D labs in specific industries relative to the location of all R&D labs in our data set, we find that the patterns of clustering in specific industries are highly similar to the overall clustering of labs that we found when we used manufacturing employment as the benchmark.

one another. We start by choosing one of the labs and drawing a ring with a quarter-mile radius around that lab. We then count the number of *other* labs in that quarter-mile ring and enter that number in a spreadsheet. Next, we move to another lab and draw a quarter-mile ring around it; then we count the number of other labs in its quarter-mile ring and enter that number in the spreadsheet. We repeat this procedure for all of the 1,035 labs in the corridor. Finally, we can compute the global measure of concentration at the level of a quarter-mile by averaging the 1,035 entries in the spreadsheet. This gives us the *average* number of labs that are located within a quarter-mile of one another.

We computed the global measures of the concentration of R&D labs for distances ranging from a quarter-mile to 100 miles. Finally, R&D clusters for a given distance, such as a quarter-mile, are identified as "significant" only when they contain more R&D labs than would be expected at that distance based on manufacturing employment (see Appendix: *Measuring Concentration Based on K-Functions*). We show that for every distance we considered, the spatial concentration of R&D labs is much more pronounced than it is for manufacturing employment. As we have noted, physical proximity is a key ingredient in order for firms and individuals to maximize the benefits from knowledge spillovers. This suggests that we should expect to see evidence that the benefits from such spillovers decline rapidly with increasing distance among the labs. More important, we find that the concentration of labs is most significant when labs are located within a quarter-mile radius of one another and that the significance of clustering of labs relative to manufacturing falls off rapidly as the distance among labs increases. The rapid attenuation of significant clustering at small spatial scales is consistent with the view that knowledge spillovers are highly localized.

We also found evidence of a secondary node of statistically significant clustering at a distance of about 40 miles. This scale is roughly comparable to that of a local labor market, suggesting that such markets may provide additional spillovers that improve the efficiency of labs. One way dense locations improve efficiency is through the better quality of matches among labs and workers that occurs in large and dense labor markets. Workers and labs in larger, denser labor markets can be much more selective in their matches because the opportunity costs (the lost wages or profits when the worker or firm has not made a successful match) of waiting for a prospective partner are lower. That is because even though workers and labs are more selective, on average they form better matches and tend to match more quickly. As a result, the average output from matches (such as new ideas that lead to innovation) is higher, and a higher share of the workforce and labs is engaged in productive matches. Another possibility is that labs in larger and denser locations may share critical inputs into the production process. For example, Robert Helsley and William Strange argue that the necessary inputs into the process of innovation are more plentiful and more readily available in an area with a dense network of input suppliers. The dense network of input suppliers facilitates innovation by making it cheaper to bring new ideas to fruition.

PLOTTING THE CLUSTERING OF R&D LABS

The discussion to this point has revealed at what distances the clustering of labs is most significant, but it does not tell us where this clustering takes place. Therefore, we use a second approach, referred to as a "local" measure of clustering, to identify specific geographic areas within the corridor with high concentrations of R&D labs. Thus, a novel feature of our study is the use of a local measure of clustering to identify specific R&D clusters as well as the labs that belong to them.

R&D clusters for a given distance, such as a quarter-mile, are identified as "significant" only when they contain more R&D labs than would be expected at that distance based on manufacturing employment.

This approach allows us to show on a map the exact locations where the clustering of labs is occurring. For example, suppose we want to know how many other labs are located within a half-mile radius of a given lab. To find this, as we did for the global measure of clustering, we draw a circle with a radius of a half-mile around a particular lab and count the number of other labs that fall within that half-mile circle. Before, to get the global measure of clustering, we computed the average number of other labs across all 1,035 labs at a half-mile distance. To get the local measure of clustering, we are interested in the number of other labs in the individual clusters themselves. The local measures of clustering focus on the size and locations of specific R&D clusters.

Once again, we are confronted with the issue of whether the count of the labs in each of these half-mile circles is greater than would be expected based on the spatial concentration of manufacturing employment. Figure 2 shows the strength of the clustering of labs relative to manufacturing employment for labs located south of Central Park in New York City. The 11 black dots indicate that the data strongly support the concentration of labs relative to the concentration of manufacturing employment, while the grey dots indicate somewhat less support.

To identify a half-mile cluster in New York City, we start by drawing rings with a half-mile radius around each of the 11 black dots shown in Figure 2. Figure 3 shows the pattern resulting from the construction of these half-mile rings. Notice that these rings tend to overlap one another, indicating a mutual influence among these labs. Next, we take the union of these rings to form the "half-mile" cluster in New York City (Figure 4). An important thing to note about this half-mile cluster is that its actual geographic distance is greater than a half-mile.

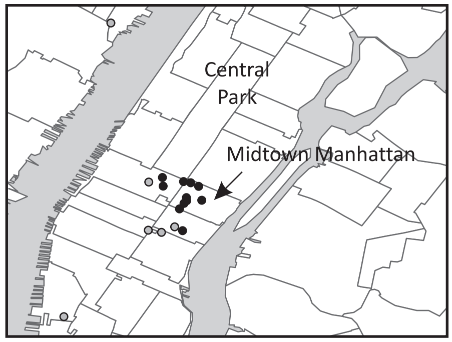
Figure 5 shows the locations of the four half-mile clusters we identified in the Boston area. The largest (both spatially and by number of labs) is found in Cambridge, MA, shown roughly at the center of the map. We also found two half-mile buffer clusters located along Route 128 and one such cluster located along Route 495.

We repeated the procedure used to create half-mile clusters, but this time we constructed one-mile rings around each of the 1,035 labs. We identified eight one-mile clusters in the Boston area, which are shown in Figure 6. Notice that all four half-mile clusters are each contained within a unique one-mile cluster. Next, we followed the same procedure to first create a five-mile cluster (of which there are two in Boston) and then a 10-mile cluster (of which there is one in Boston). Figure 7 shows the two five-mile clusters (solid black line) and the 10-mile cluster (dotted black line).

There are 187 R&D labs within

FIGURE 2

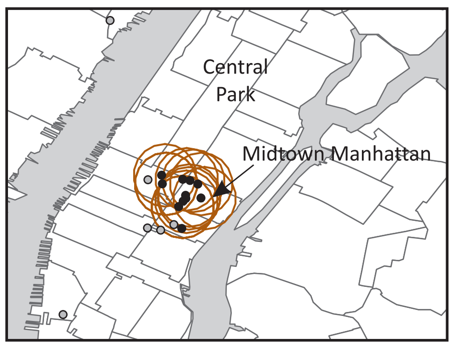
R&D Labs in New York City



Each dot represents the location of a single R&D lab. The black dots strongly indicate a local cluster of labs relative to manufacturing employment. The grey dots indicate a less significant concentration of labs relative to manufacturing employment.

FIGURE 3

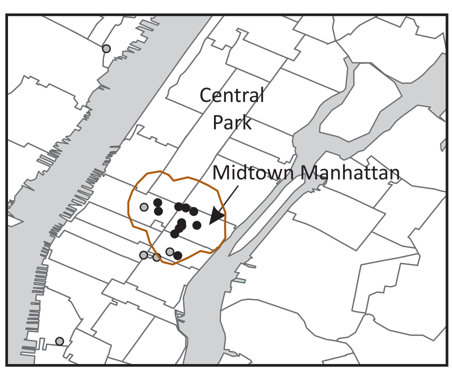
Constructing Half-Mile Buffer Rings



This half-mile cluster in New York City was created by constructing rings with a half-mile radius around each black dot. These rings tend to overlap one another, indicating a mutual influence among these labs.

FIGURE 4

Half-Mile Cluster in New York City

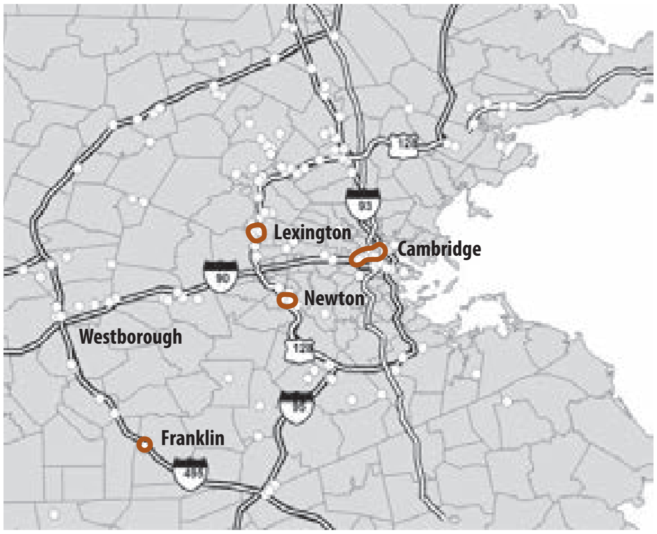


To identify New York City's half-mile cluster, we drew a line around the perimeter of the rings in Figure 3. It is important to note, however, that the actual geographic distance of this cluster is greater than a half-mile.

Sources: Directory of American Research and Technology and authors' calculations

FIGURE 5

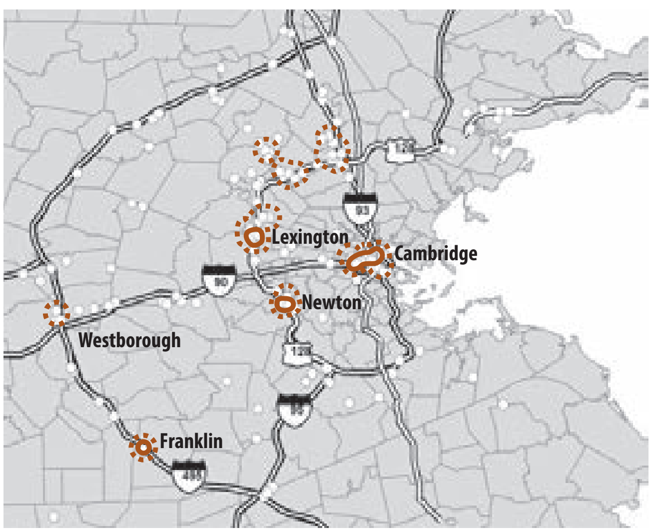
Half-Mile Clusters in Boston



This figure shows four half-mile clusters of labs in Boston, the largest of which is in Cambridge at the junction of Route 90 and Route 93.

FIGURE 6

One-Mile Clusters in Boston



Eight one-mile clusters of labs in Boston are indicated by dotted brown rings. Notice that all four half-mile clusters, which are indicated by solid brown rings, are situated within one-mile clusters.

Boston's single 10-mile cluster. Most of these labs conduct R&D in five industries: computer programming and data processing, drugs, lab apparatus and analytical equipment, communications equipment, and electronic equipment. The largest five-mile cluster, which is shown in Figure 7, contains 108 labs, which account for 58 percent of all labs in the larger 10-mile cluster. At the one-mile scale, Boston has eight clusters, six of which are centered in the largest five-mile cluster. The largest of these one-mile clusters contains 30 labs, half of which conduct research on drugs.

Figure 8 shows the clusters of R&D labs we identified in the Philadelphia region, where there are a total of 49 labs. The city of Philadelphia is shown by the darker grey area east of the center of the figure. The dotted black ring depicts Philadelphia's 10-mile cluster. Of the 49 labs in this broad cluster, 16 conduct research on drugs, and another 16 perform research in the plastics materials and synthetic resins industry. The Philadelphia region contains two five-mile clusters, shown by the solid black boundaries in Figure 8. The most prominent subcluster is centered in the King of Prussia area, directly west of the city of Philadelphia, and contains 30 labs, of which 40 percent conduct research on drugs. Within this subcluster, there is a much tighter concentration of labs (indicated by the dotted brown ring in Figure 8) located near Routes 76 and 276.

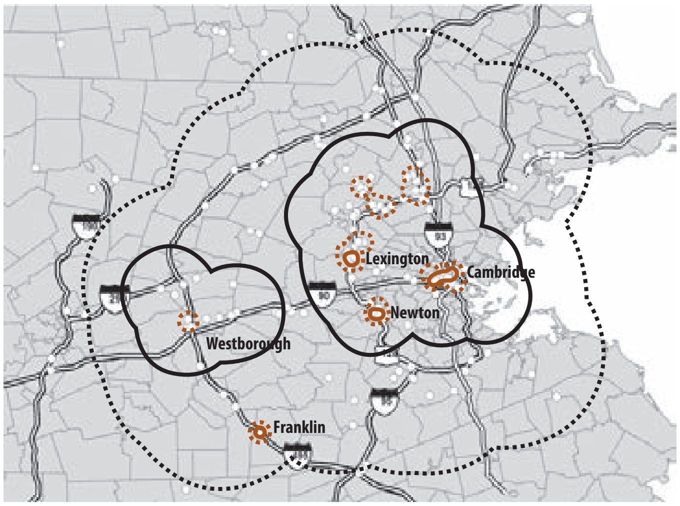
The second subcluster is centered in the city of Wilmington, DE, where about 25 percent of the labs are also engaged in research on drugs, but most (almost 60 percent) are conducting research on plastics materials and synthetic resins.

THE EFFECTS OF KNOWLEDGE SPILLOVERS

Innovation is important because it can directly affect a nation's productivity growth and the economic welfare of society through the introduction of new or improved goods and lower prices. In addition to these direct benefits, as we have argued in this article, the innovative activity of one person can also influence the innovative activity of others through knowledge spillovers. Paul Krugman has argued, however, that knowledge spillovers are impossible to measure empirically because they "are invisible; they leave no paper trail by which they may be measured and tracked." However, as Jaffe and coauthors have noted, "Knowledge flows

FIGURE 7

Ten-Mile Cluster in Boston



This figure shows the two five-mile clusters of labs in Boston (solid black lines) and the single 10-mile cluster (dotted black line). Notice that all four half-mile clusters (solid brown) identified for Boston are situated within one-mile clusters (dotted brown). Similarly, most of the one-mile clusters lay within the two five-mile clusters, and the two five-mile clusters are contained within the 10-mile cluster.

FIGURE 8

Ten-Mile Cluster in Philadelphia

In the Philadelphia region, we identified a single one-mile cluster that is located west of the city (the city is shown in dark grey) approximately in the King of Prussia, PA, area. The Philadelphia region has two five-mile clusters (solid black lines) and one 10-mile cluster (dotted black line). The second five-mile cluster is centered in the city of Wilmington, DE.

do sometimes leave a paper trail in the form of patent citations to prior art."

Jaffe and coauthors pioneered a method for studying the geographic extent of knowledge spillovers using patent citations. Every patent contains the names, hometowns, and Zip codes of the inventors named in the patent. A patent can be assigned to a location by using the Zip code of one of its inventors (usually the first person named). Patent citations are similar to citations received by academic articles in that patent citations reference prior technology or prior art on which the citing patent builds. Therefore, Jaffe and coauthors hold that patent citations are a useful proxy for measuring knowledge flows among inventors. If knowledge spillovers are localized within a given metropolitan area, then citations to patents within a given metropolitan area should disproportionately come from other inventors who are located within that metropolitan area.

However, Jaffe and coauthors point out that just because we observe a geographic clustering of technologically related activities, such as the clustering of the semiconductor industry in Silicon Valley, this clustering is not necessarily evidence of knowledge spillovers among these related activities. There are other sources of agglomeration economies in metropolitan areas, such as better matching and sharing, that could explain the spatial clustering of activities in the semiconductor industry. Jaffe and coauthors deal with the spatial clustering of related activities by constructing a set of control patents designed to match the existing geographic concentration of technologically related activities. To test for localized knowledge spillovers, Jaffe and coauthors construct three patent samples. The first sample consists of a set of originating patents. The second sample consists of a set of patents that cite one of the originating patents (referred to as citing patents). The final sample consists of a control patent chosen to match each of the citing patents. To qualify as a control patent, the patent must be as similar as possible (in terms of being in the same technology class and having an application date as close as possible) to the matched citing patent, but the control patent must not cite the matched originating patent. Jaffe and coauthors compute two geographic matching frequencies: one between the citing patents and the originating patents and one between the control patents and the originating patents. Their test for the localization of knowledge spillovers is whether the citation matching frequency for a given geographic definition (states and metropolitan areas) is significantly greater than that associated with the control matching frequency. Jaffe and coauthors find that patent citations are two times more likely to come from the same state and about six times more likely to come from the same metropolitan area as earlier patents than one would expect based on the control patents.

In our study, we adopt Jaffe and coauthors' methodology to look for evidence of localized knowledge spillovers, except that we use the boundaries determined by the nine five-mile clusters identified in our research instead of using state and metropolitan area boundaries.[7] State boundaries are politically determined, rather than economically justified, and states are too big to adequately capture knowledge spillovers, which are highly localized. In addition, the boundaries of metropolitan areas are determined by labor market flows; therefore, they are not well suited for analysis of spillovers among individuals engaged in innovative activity. Instead, we use the boundaries determined by our nine five-mile clusters as our basic geography, since these boundaries are determined by interrelationships among the R&D labs and more accurately reflect the appropriate boundaries in which knowledge spillovers are most likely to occur.

The patent citation counts that we use are constructed from the NBER Patent Citations Database. Patents are assigned to locations according to the Zip code of the first inventor named on the patent.[8] There were 9,105 patents applied for in the nine five-mile buffer clusters we identified in our study during the period 1996–1997. After removing self-citations, these originating patents received 90,159 forward citations during the period 1996–2006.[9] But we were able to find control patents for only about 55,000 of the citing patents. This limits our analysis to those citing patents for which we have controls.[10] We find that, on average, a patent that falls within one of our five-mile clusters is 4.3 times more likely to cite an earlier patent in the same five-mile cluster compared with a control patent (a finding that is highly statistically significant). Despite the fact that knowledge spillovers are not directly observable, they do leave a paper trail in the form of patent citations. We find that these paper trails provide evidence consistent with the geographic concentration of knowledge spillovers.

CONCLUSION

In this article, we summarize the findings from our study that uses distance-based measures to analyze the spatial concentration of over 1,000 R&D labs in the Northeast corridor of the United States. Rather than using a fixed spatial scale, such as counties and metropolitan areas, we attempt to describe the spatial concentration of R&D labs more precisely by considering the spatial structure at different scales. We find that the clustering of labs is by far most significant at very small spatial scales, such as distances of about one-quarter of a mile, with significance attenuating rapidly during the first half-mile. The rapid attenuation of significant clustering at small spatial scales is consistent with the view that knowledge spillovers are highly localized.

We introduce a novel way to identify the location of clusters and number of labs in these clusters. For example, this approach identified a number of clusters of R&D labs in the Boston, New York–Northern New Jersey, Philadelphia–Wilmington, and Washington, D.C., areas. We also found that each of these clusters has distinct characteristics, especially in terms of the mix of industries the R&D labs serve.

Using patent data, we are able to provide evidence that knowledge spillovers are highly localized within the clusters of R&D labs that we identify. We find that patent citations are a little over four times more likely to come from the same cluster as earlier patents than one would expect based on the preexisting geographic concentration of technologically related activities.

---

[7] We identified two five-mile clusters in Boston (Figure 7), three such clusters in New York, two in Philadelphia (Figure 8), and two in Washington, D.C. In this article, we present only the findings averaged across the nine clusters. See our working paper for details on the individual clusters.

[8] The patent and citation data we use from the National Bureau of Economic Research (NBER) Patent Data Project provide the name, town, and Zip code of the principal (or first named) inventor on each patent. As is standard when assigning patents to areas, we assign patents to our clusters using the Zip code of the first inventor named on the patent. Knowledge spillovers can occur among individuals who meet because they are part of either local technical or social networks. For example, AnnaLee Saxenian describes how Walker's Wagon Wheel bar in Mountain View, CA, became a popular place for engineers who lived in Silicon Valley to exchange ideas.

[9] Since self-citations may not result from knowledge spillovers, we excluded not only inventor self-citations but also citing patents owned by the same organizations as the originating patent.

[10] There was an insufficient number of control patents to confidently conduct the analysis for the one-mile or half-mile clusters.

Appendix: Measuring Concentration Based on K-Functions

The Global K-Function

A popular measure of concentration is Ripley's *K*-function, which we use to test for clustering at differing distances:

$$\hat{K}^{O}(d) = \frac{1}{n}\sum_{i=1}^{n} C_i(d)$$

where $C_i(d)$ is the count of additional labs within distance *d* from lab (location) *i* and *n* is the total number of locations in the study ($n = 1{,}035$ in our study). To see how this works, set *d* equal to one mile. Take the first lab and draw a one-mile circle around that lab. Count the number of other labs in that one-mile circle and enter the resulting count of other labs into a spreadsheet. Go to the next lab and construct a one-mile circle around that lab. Count the number of other labs in that one-mile circle and enter the resulting number into the spreadsheet. Repeat these steps for all 1,035 labs. Sum over the 1,035 observations and divide by 1,035 labs. This is the average value of concentration of labs at a distance of one mile, denoted by $\hat{K}^{O}(1)$. We calculate the average observed value of concentration, beginning at a quarter-mile and increasing at quarter-mile increments below one mile and at one-mile increments from one mile to 100 miles.

The key question of interest is whether the overall pattern of R&D locations in the 10 states and the District of Columbia exhibits more clustering than would be expected from the spatial concentration of manufacturing in those areas. To address this question statistically, our null hypothesis is that R&D locations are determined entirely by the distribution of manufacturing employment.

We use a two-step procedure for generating counterfactual observations that are used to test the null hypothesis. In the simulations, we randomly allocated labs to Zip codes based on a probability proportional to manufacturing employment in that Zip code so that Zip codes containing a large share of employment are more likely to be assigned labs. For each distance, we compute a simulated distribution of labs. We compared the observed value for their *K*-functions (the $\hat{K}^{O}(d)$) with values obtained from a simulated distribution of R&D labs. If the observed value for the *K*-function for a given distance is large relative to the simulated distribution, this is taken as evidence of significant clustering of labs relative to manufacturing employment. *P*-values can be computed as:

$$P(d) = \left\{ \frac{\text{The number of simulated values at distance } d \text{ that are at least as large as the observed value}}{\text{Number of simulation performed}} \right\}$$

For example, if we performed 1,000 simulations and there are 10 simulated values at least as large as $\hat{K}^{O}(d)$, then there is only a one-in-a-hundred chance of observing a value at least as large as $\hat{K}^{O}(d)$. In this example, there is significant clustering of R&D locations at the 0.01 level of statistical significance at spatial scale *d*. However, we found that the clustering of labs is so strong relative to manufacturing employment that the estimated *p*-values were uniformly 0.001 for all the distances we considered. We obtained sharper discrimination by calculating the *z*-scores for each observed estimate, $\hat{K}^{O}(d)$, as given by

$$z(d) \;=\; \frac{\hat{K}^{0}(d) - \bar{K}_d}{s_d} \quad , \quad d = \{0.25, 0.5, 0.75, 1, 2, \ldots, 99, 100\}$$

where $\bar{K}_d$ and s_d are the corresponding sample means and standard deviations for the $N+1$ sample *K*-values. These *z*-scores are shown along the vertical axis in the figure, while the horizontal axis shows distances among R&D labs. The higher the *z*-score for a given distance, the more spatially concentrated the R&D labs are at that distance relative to manufacturing employment. Notice that the highest *z*-score we found, which is more than 30 standard deviations away from the mean, occurs at the shortest distance among labs we considered (one-quarter of a mile) and declines rapidly with distance up to a distance of about five miles. The rapid decline in *z*-scores (significance of clustering of R&D labs) at short distances is consistent with the view that knowledge spillovers are highly geographically localized. Notice that the lowest *z*-score obtained, which occurs at a distance of about five miles, is still more than 7 standard deviations away from the mean, indicating that R&D labs are significantly more concentrated than manufacturing employment over all the distances we considered. We also observe a secondary mode of significance at a scale of about 40 miles, which is roughly associated with metropolitan areas.

The Local *K*-Function

Basically, the local version of Ripley's *K*-function for a lab at a given location is simply the count of all additional labs within distance *d* of the given lab. In terms of the notation, the local *K*-function, $\hat{K}_i$, at location i is given for each distance, *d*, by,

$$\hat{K}_i(d) = C_i(d)$$

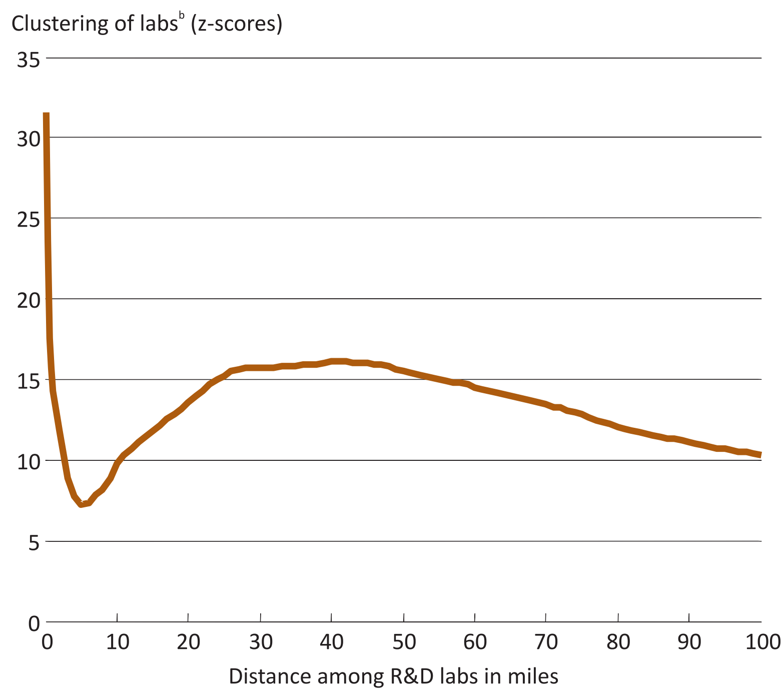
We use the same null hypothesis employed in the global *K*-function analysis that R&D labs are distributed in a manner proportional to the distribution of manufacturing employment. The only substantive difference from the procedure used in global *K*-function analysis is that the actual point associated with location *i* is held fixed when computing the simulated values for the local *K*-function. That is, for a given distance, holding the location of the lab fixed, we compute a simulated distribution of labs at that point. We compared the observed value for their *K*-functions (the $\hat{K}_i(d)$) with values obtained from a simulated distribution of R&D labs. If the observed value for the *K*-function at a given point is large relative to the simulated distribution, this is taken as evidence of significant clustering of labs relative to manufacturing employment at that location. The set of radial distances (in miles) used for the local tests was $D = \{0.5, 0.75, 1, 2, 5, 10, 11, 12.., 100\}$.

In our global analysis, the *p*-values were essentially the same for nearly all spatial scales. That is not the case for the local analysis. It is not surprising to find that many isolated R&D locations exhibit no local clustering whatsoever; therefore, wide variations in significance levels are possible at any given spatial scale. Thus, *p*-values are used in the local *K*-function analysis.

An attractive feature of these local tests is that the resulting *p*-values for each point i in the observed pattern can be *mapped*. This allows us to check visually for *regions* of significant clustering. In particular, groupings of very low *p*-values serve to indicate not only the location but also the *approximate size* of possible clusters.

Because we conduct tests for local clustering over many locations and spatial scales, we need to address two aspects of the "multiple testing" problem. First, suppose that there is, in fact, no local clustering of labs. In our simulations, we would nonetheless expect to find that 5 percent of the observed values for the local *K*-functions for a given distance are statistically significant at the 5 percent level of significance. Therefore, when many such tests are conducted (1,035 tests for each distance considered), we are likely to find some degree of significant clustering using standard testing procedures. The incidence of this type of "false positive" findings is mitigated by reducing the threshold level of significance (the *p*-value) deemed to be "significant." That is, we can minimize the incidence of false positives due to the multiple testing problem by focusing on labs with very high levels of statistical significance (*p*-values of 0.001 or lower). We refer to these as core points — the black dots in Figure 2 in the article.[a] A second condition of a core point is that there must be at least four other labs at a given distance. This condition is imposed to exclude isolated labs that happen to be in areas with little or no manufacturing employment.

Clustering of Labs Attenuates Rapidly with Distance



[a] The grey dots in Figure 2 are associated with p-values no greater than 0.005.

[b] *Z-scores are shown along the vertical axis, while the horizontal axis shows distances among R&D labs. The higher the *z*-score for a given distance, the more spatially concentrated the R&D labs are at that distance relative to manufacturing employment. For example, a *z*-score of 10, occurring at a distance of about two miles, indicates that the concentration of labs at that distance is 10 standard deviations away from the mean at that distance, indicating that labs are significantly more concentrated at that distance relative to manufacturing employment.

REFERENCES

Agrawal, Ajay, Devesh Kapur, and John McHale. "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data," *Journal of Urban Economics*, 64:2 (2008), pp. 258-69.

Arzaghi, Mohammad, and J. Vernon Henderson. "Networking Off Madison Avenue," unpublished manuscript (2005).

Audretsch, David B., and Maryann P. Feldman. "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86 (1996), pp. 630-40.

Buzard, Kristy, and Gerald A. Carlino. "The Geography of Research and Development Activity in the U.S.," Federal Reserve Bank of Philadelphia *Business Review* (Third Quarter 2008), pp. 1-10.

Carlino, Gerald A., Jake K. Carr, Robert M. Hunt, and Tony E. Smith. "The Agglomeration of R&D Labs," Federal Reserve Bank of Philadelphia Working Paper 12-22 (September 2012).

Directory of American Research and Technology, 23rd ed. New York: R.R. Bowker, 1999.

Duranton, Gilles, and Henry G. Overman. "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies*, 72 (2005), pp. 1077-1106.

Ellison, Glenn, and Edward L. Glaeser. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105 (1997), pp. 889-927.

Ellison, Glenn, Edward L. Glaeser, and William Kerr. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," Discussion Paper 2133, Harvard Institute of Economic Research (April 2007).

Helsley, Robert W., and William C. Strange. "Innovation and Input Sharing," *Journal of Urban Economics*, 51:1 (2002), pp. 25-45.

Jaffe, Adam, M. M. Trajtenberg, and R. Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108 (1993), pp. 577-98.

Krugman, Paul R. *Geography and Trade*. Cambridge, MA: MIT Press, 1991.

Rosenthal, Stuart, and William C. Strange. "The Determinants of Agglomeration," *Journal of Urban Economics*, 50 (2001), pp. 191-229.

Saxenian, AnnaLee. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*, 2nd ed. Cambridge, MA: Harvard University Press, 1996.

The Promise and Challenges of Bank Capital Reform

BY RONEL ELUL

The failure and bailout of some prominent financial institutions amid the crisis of 2007-09, and the effect these events had on the economy as a whole, have led policymakers to rethink how the global financial system is regulated.[1] These changes, commonly known as the Basel III Accords, will require banks to maintain more capital in reserve, hold higher-quality capital, and assign greater risk weights to certain types of assets.[2]

Why were these changes considered necessary? And how might the new standards help prevent future crises? To understand the rationale behind the changes, it is helpful to examine the history of bank capital regulation and explore some reasons why previous regulatory frameworks may have proved inadequate during the crisis.

HOW AND WHY WE REGULATE BANKS

Why We Need to Regulate Banks. Society may have a particular interest in financial stability — and in particular regulating financial institutions so as to reduce the incidence of their failure — for several reasons. One reason is the key role that banks play in channeling funds to firms throughout the economy. This means that the impact of a bank failure, or of a weak bank, can be greater than that of other kinds of businesses. Victoria Ivashina and David Scharfstein give an example of how a shock to banks can affect other parts of the economy. They show that banks that were members of lending syndicates with Lehman Brothers reduced their lending to a greater extent than other banks following the Lehman bankruptcy in September 2008.[3] Ivashina and Scharfstein reason that these banks expected to shoulder the commitments that Lehman could no longer honor, so they cut back on making other loans. Similarly, Manju Puri, Jorg Rocholl, and Sascha Steffen show that German savings banks that had significant exposure to U.S. subprime mortgages were more likely to reject loan applications.

Another reason why society is concerned with regulating banks is the interconnection among financial institutions; the failure of one can bring down others. This was cited, for example, in the bailout of AIG, whose failure would have led to significant losses at Goldman Sachs and the large French bank Société Générale, among others. Yet another reason that bank failures may be of social concern is that because U.S. bank deposits are guaranteed (through the FDIC), taxpayers may end up bearing the costs of bank failures.[4]

Finally, the regulation of banks may be important simply because they are particularly fragile, as compared with nonfinancial firms. Many financial firms are fragile because they tend

[1] The Federal Reserve Bank of St. Louis has compiled a timeline of the financial crisis at http://timeline.stlouisfed.org/index.cfm?p=timeline.

[2] The Basel Committee on Banking Supervision provides an overview and details on Basel III at http://www.bis.org/bcbs/index.htm.

[3] In a lending syndicate, a group of banks makes a shared commitment to make loans to a particular borrower at the customer's demand for some fixed period of time.

[4] Although the guarantee fund is paid for by an assessment on banks, taxpayers are on the hook to the extent that the funds needed to pay off depositors turn out to be greater than the funds available.

Ronel Elul is an economic advisor and economist specializing in microeconomics and financial market research at the Federal Reserve Bank of Philadelphia. The views expressed in this article are not necessarily those of the Federal Reserve. This article and other Philadelphia Fed research and reports are available at www.philadelphiafed.org/research-and-data/publications.

to fund their assets with debt. Furthermore, this debt often has much shorter maturity than the assets (for example, using demand deposits to fund mortgage lending). Thus, they are subject to the risk of bank runs in which lenders (including depositors) refuse to continue financing the bank. At the same time it may be difficult for the bank to raise funds by selling its assets, and so it is at risk of failure.

Capital Requirements Are an Important Regulatory Tool. One of the most important ways in which banks are regulated is through capital requirements. A financial institution's capital is its net worth: the difference between the values of its assets and liabilities. A bank's typical assets would include loans to businesses and households, and securities such as municipal bonds or mortgage-backed securities, while its liabilities would include deposits, loans from other banks or the central bank, and other types of debt.

But what's the best way to measure net worth? One way would be to consistently use market values for assets and liabilities, a measure that economists call "economic capital." But the capital measure used by regulators departs from this by relying more on accounting book values. One reason for this is that it may be hard to determine market values for assets, a particular problem during financial crises, when markets shut down and the number of trades falls to a trickle. Thus, for regulatory purposes, loans the bank made might be carried at historical cost until they reach a certain level of delinquency, for example 90 days delinquent, at which point they are written off.

A further reason book values are used is that market values fluctuate more often; this might create more uncertainty about when regulators would intervene. This uncertainty might make it more difficult for the bank to raise financing. The drawback of relying on book values, however, is that these tend to be backward-looking and, thus, generally represent a less up-to-date measure of the firm's worth.

Capital regulation usually takes the form of requiring the bank to hold a minimum level of capital, relative to the bank's assets. A typical capital ratio requirement would require the bank's equity financing to be at least a certain fraction of the value of some measure of its assets.[5] Requiring banks to hold capital has several benefits. One is that holding capital helps to absorb unanticipated losses, thereby inspiring confidence that the bank can continue as a going concern. In addition, it protects nonequity liability holders, especially depositors, and deposit insurers (and thus, the taxpaying public) against losses. Finally, it limits risk by restraining asset growth; to lend more, banks need to raise more capital.

For several reasons many economists feel that banks would not hold enough capital were they left to their own devices, and thus they must be regulated. One reason is that equity financing tends to be more expensive than debt financing because debt interest payments are tax deductible.[6] Another important reason is that the management team of a bank does not bear the full cost of the bank's failure; there can be spillovers to other financial institutions and to society more generally.

INTERNATIONAL CAPITAL REGULATION

Why Might We Want Regulatory Harmony? Since the 1970s, there has also been an effort to harmonize international capital regulations through the Basel Committee on Banking Supervision (BCBS).[7] Why would we need international harmonization of capital regulations? One reason is that bank failures in one country can spill over to other countries. One early example is the failure of the German Herstatt Bank in 1974. Herstatt had agreed to exchange Deutsche marks it received from its customers for U.S. dollars, which were to be delivered in New York, but the bank was shut down by German regulators before it could deliver the dollars (since New York markets opened later in the day). This led to turmoil in the interbank markets that banks use to borrow from each other. Another example is Lehman Brothers; one of the biggest creditors in its bankruptcy was the German Deposit Insurance Fund.

Another reason given for why we need international harmonization is the potential for a race-to-the bottom

> The first international agreement on capital regulation was the 1988 Basel Accord, commonly known as Basel I.

[5] I will discuss the various ways in which regulators measure assets for capital regulation below. The most commonly used measure is *risk-weighted assets*, in which the amount of capital required per dollar of an asset depends on the risk of the asset. As discussed below, the Dodd-Frank Act would require banks to maintain a 7 percent equity capital ratio by 2019.

[6] Another reason equity financing is more expensive than debt is that the value of equity is more sensitive to private information that insiders might have about the value of the bank, as discussed by Stewart Myers and Nicholas Majluf.

[7] The BCBS provides a forum for international cooperation on banking supervisory matters, including the harmonization of regulations.

in bank regulation.[8] That is, each national regulator will lower its standards in order to lure business to its jurisdiction. But are there any drawbacks to harmonization?

Giovanni Dell'Ariccia and Robert Marquez develop a model that analyzes the tradeoff between the benefits and costs of international harmonization of regulations. In their model, regulators are interested not only in the profitability of their home banks but also in financial stability. Competition among regulators leads to standards that are too lax because national regulators want to benefit home bank shareholders and don't fully take into account the benefits to other countries' banks of imposing tighter standards on their own banks. Specifically, tighter standards set by regulators on banks domiciled in that country lead to fewer bank failures in other countries in which the bank also does business. On the other hand, there is a cost to coordinating regulation: uniform standards may not fit each country. In Dell'Ariccia and Marquez's model, this is because the public in each country places different weights on financial stability versus the profitability of their home banks. But one can also imagine other salient differences, such as differences across countries in the concentration of the banking sector or in the relative sophistication of nonbank financial markets. So when is it good to harmonize regulations? In their model, a regulatory union is beneficial when countries are not too dissimilar, so that the benefits outweigh the costs.

The First Basel Accord. The first international agreement on capital regulation was the 1988 Basel Accord, commonly known as Basel I. Basel I required banks to hold at least 8 percent capital relative to risk-weighted assets. Asset classes perceived as less risky received lower risk weights. For example, sovereign debt was assigned a zero risk weight (so no capital was required), mortgages were given a 50 percent risk weight, and corporate bonds a 100 percent risk weight. This meant, for example, that the capital a bank was required to hold per dollar of mortgage loans made was only half that for corporate loans. Each country that was a party to Basel I agreed to write its own regulations that implemented these principles, although, in practice, the national authorities had considerable discretion in how to interpret them.

What was the effect of the first Basel Accord? Patricia Jackson and her coauthors survey the literature and find that this accord generally represented a tightening of regulations, since it led banks in the G-10 countries to raise their capital ratios, on average.[9] There may have been some negative consequences to this, however. First, some economists, such as Ben Bernanke (who later became Chairman of the Federal Reserve Board) and Cara Lown, have argued that this led to a credit crunch, or a decline in lending, during the 1990-91 recession in the U.S.

In addition, Basel I may also have encouraged *regulatory arbitrage*, that is, a shift toward risky activities that are not fully captured by the regulations. The reason is that with higher capital requirements, banks may have had an increased motivation to evade regulations in order to conserve capital. Furthermore, setting uniform international standards required more formal rules than had existed in the past, which could make it easier for banks to structure their activities in such a way so as to evade these regulations.

In his study, David Jones gives several examples of how banks could use securitization to reduce their regulatory capital requirements while still effectively retaining all of the risk of the loans. One way they can do this is by selling the most senior, safest parts of the assets to investors (thereby removing them from their balance sheets) while retaining the junior, riskier portions. Basel I's emphasis on credit risk alone may also have encouraged banks to increase their profits by taking on other risks. For example, Linda Allen, Julapa Jagtiani, and Yoram Landskroner find that, after the introduction of the first Basel Accord, some banks took on additional interest rate risk without increasing their capital.[10] In addition, Basel I did not distinguish between different risks *within* categories. Since all corporate loans received a 100 percent risk weight, for example, banks might lend to riskier customers, thereby increasing the risk of distress — a risk partially borne by other banks and taxpayers — without being required to hold more capital of its own. Finally, Basel I considered the credit risk of assets individually, rather than the riskiness of the bank's whole portfolio; thus, a well-diversified portfolio could have the same required capital as a poorly diversified portfolio. Notwithstanding these specific examples, a survey of the literature by

[8] The risk of a "race to the bottom" in banking regulation was cited as a reason that "standards be implemented uniformly and in a timely fashion" by Stephen Cecchetti, head of the monetary and economic department at the Bank for International Settlements, in an interview with the *Wall Street Journal* on October 30, 2012.

[9] The Group of Ten, or G-10, is composed of 11 nations that are members of the International Monetary Fund: Belgium, France, Germany, Italy, Japan, the Netherlands, Sweden, the United Kingdom, the United States, Canada, and Switzerland.

[10] By interest rate risk we mean holding assets whose values fluctuate more in response to variations in interest rates than do the values of the liabilities used to fund the assets. In particular, a rise in interest rates can lead to a large fall in assets with long maturities. While these assets yield high returns because they are riskier, they would not require more capital.

Linda Allen finds no consensus that banks increased their overall risk in response to Basel I.

Basel II Made Capital Requirements More Sensitive to Risk. The second Basel Accord (Basel II), published in 2004, was designed to address some of the shortcomings of Basel I, and its provisions remain in force in some countries. Basel II makes the standard framework more risk-sensitive than Basel I, especially within asset categories. It does this primarily by relying on credit ratings to calibrate risks. Thus, assets with a BBB rating from Standard & Poor's require less capital than those with a BB rating. Basel II also allows large banks to use their own internally developed risk models, the presumption being that these models more accurately reflect risk, particularly at the portfolio level. Note, however, that countries differed in how they implemented the accord. For example, while European regulators allow banks to estimate their own required capital using internal models, U.S. regulators permit U.S. banks to use their own internal models only for assets held in their trading book, and even then, they are more restricted than banks in other countries.

Shortcomings of Basel II. There are some shortcomings with the Basel II framework, however, some of which became apparent during the financial crisis.

First, the heavy reliance on credit ratings may have created problems. For instance, Basel II treats ratings inconsistently, with sovereign debt often receiving lower capital charges than corporate bonds with the same ratings. For example, a corporate bond with a rating between A– and A+ receives a 50 percent risk weighting, whereas a sovereign bond with the same rating (such as Greek bonds in 2009) would get only a 20 percent risk weighting. This inconsistency may help to explain the heavy holdings of risky sovereign debt by some European banks.

Another shortcoming of the Basel II capital accord is that it underweights "tail risk." That is, it arguably does not assign sufficient capital to protect against extreme events such as a nationwide collapse of the housing market or a financial crisis. Viral Acharya, Thomas Cooley, Matthew Richardson, and Ingo Walter have argued that in the run-up to the financial crisis, this aspect of the Basel II framework encouraged the biggest financial institutions to accumulate large amounts of tail risk without holding a commensurate amount of capital. One example is the most senior tranches of mortgage-backed securities (MBS), which had AAA ratings (and thus very low capital charges) and were often retained by large banks.[11] Such securities were considered safe, except in what was then considered the unlikely event of a large and widespread collapse in the housing market.

Another instance of Basel II underemphasizing tail risk is that, in some circumstances, it allows banks to use their own internal models and, in particular, encourages the use of value-at-risk (VaR), an approach to measuring the risk of loss in a given portfolio of assets.[12] However, in most common implementations of value-at-risk, the behavior in the tails, that is, in the case of extreme events, is not fully considered. That is, value-at-risk measures losses that occur with a large enough probability (for example, 99 percent of the time) but does not consider the potential severity of losses in the other 1 percent. Basel II may encourage tail risk in another way. The regulations have a similar impact across many banks, and thus, they may all align their portfolios in similar ways, thereby further heightening systemic risk.

Another potential problem with Basel II is that it tends to have a procyclical effect on capital charges. That is, capital requirements can go down in booms and rise following a period of financial instability. One reason for this procyclical effect is that the regulations rely on credit ratings, which generally go up in good times and down in bad times. Another factor contributing to procyclicality arises from the use of value-at-risk for setting capital requirements. Asset price volatility is an important input into value-at-risk calculations. Because data from the recent past are generally used to estimate volatility, following a period of financial stability in which asset volatilities are relatively low such as 2001-06, a bank's portfolio is likely to appear less risky and thus require less capital. Conversely, as can be seen from Figure 1, (which plots the level of the S&P 500 and stock market volatility as measured by the VIX index), during bad times prices tend to be more volatile, and so capital requirements increase.[13] As the joint report from the Financial Stability Forum[14] and the BCBS points out, one potentially undesirable consequence of this

---

[11] A tranche is a slice of a mortgage-backed security that is sold as a separate bond. The senior tranches of private MBS are those that have first claim on cash flows in the case of default and are thus less risky (and so obtain a higher rating). However, as became apparent during the financial crisis, they are by no means risk-free.

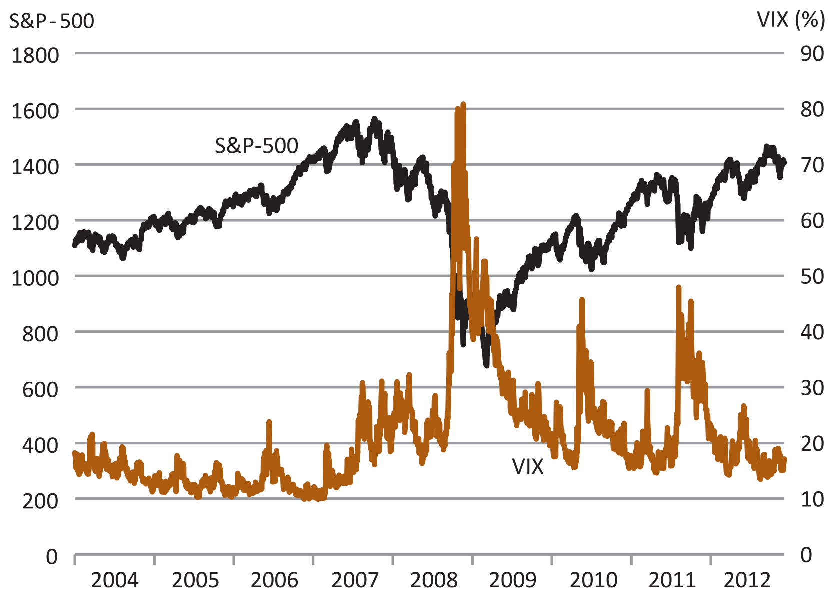
[12] For more on the use of value-at-risk by banks in meeting capital requirements, see the article by Mitchell Berlin and the book by Anthony Saunders.

[13] The VIX is an index disseminated by the Chicago Board Options Exchange that uses information from S&P 500 index options to infer the market's expectation of volatility over the next 30 days.

[14] The Financial Stability Forum was established in 1999 to promote international financial stability through enhanced information exchange and international cooperation in financial market supervision and surveillance. In 2009, it was replaced by the Financial Stability Board, which has a broader membership.

FIGURE 1

The S&P 500 and the VIX Volatility Index



Sources: Standard & Poor's, Chicago Board Options Exchange.

procyclicality is that it tends to encourage more lending during booms and, conversely, requires banks to sell assets when their prices have fallen, thus potentially amplifying these cycles.

Finally, although Basel II expands the range of risks that are considered in determining regulatory capital, some, such as liquidity risk, are still neglected.[15] One example of this risk is highlighted by the collapse of the British lender Northern Rock in September 2007. Hyun Song Shin shows that Northern Rock had obtained an unusually small share of its funding from traditional branch-based retail deposits. On the other hand, it relied heavily on deposits from offshore and Internet-based bank accounts and on "wholesale funding," in which short-term securities are sold to investors. And while traditional retail depositors tend to be slow to withdraw their funds from a bank, this was not the case for the other investors upon whom Northern Rock relied too heavily, and the lender was hurt when these investors fled risky investments at the start of the financial crisis in the summer of 2007 and refused to roll over their deposits at institutions such as Northern Rock.

Similarly, a paper by Viral Acharya, Philipp Schnabl, and Gustavo Suarez shows that Basel II was also subject to regulatory arbitrage in the run-up to the financial crisis because of its inconsistent treatment of credit and liquidity risk. Banks set up asset-backed commercial paper conduits that were "off balance sheet" for regulatory purposes. These conduits purchased medium- to long-term assets (often mortgage-backed securities) and held them until maturity. They were financed by issuing a type of short-term debt called asset-backed commercial paper (ABCP), with maturities of 30 days or less. Even though the assets were formally off the banks' balance sheets, in reality, the banks were exposed to the risk that they would be forced to take over the assets if investors stopped purchasing the ABCP. Banks were exposed to risk because they typically offered "liquidity guarantees" — promises to pay off maturing commercial paper as long as assets were not actually in default — to persuade investors to buy it. From the bank's perspective, this was an attractive deal because these liquidity guarantees carried lower capital charges than would have been the case had the assets been formally held on the bank's balance sheet. However, this structure really left the risk with the issuing bank because the short maturity of the ABCP meant that it would need to be paid off well before the assets were formally in default. Once investors, concerned about the risk of the underlying assets, stopped buying new commercial paper, the banks were forced to take these assets back onto their balance sheets, degrading their capital ratios.

REFORM OF BASEL II

Basel II.5. Recent revisions to the Basel Accords have addressed these concerns. Some of these revisions were proposed in 2009 and are colloquially known as Basel II.5. One area involves increasing capital requirements for certain assets, particularly for "resecuritizations" such as collateralized debt obligations (CDOs).[16]

[15] Liquidity risk refers to the problems of having assets that are difficult to sell and liabilities that have short maturities — for example, deposits. With this asset-liability structure, banks can be caught in a situation in which they must sell assets at fire-sale prices if liability holders such as depositors refuse to roll over their claims.

[16] A CDO is an asset-backed security in which the underlying collateral is itself composed of other debt securities. For example, during the subprime bubble, low-rated, junior mortgage-backed security tranches were sometimes packed into new securities. For more on CDOs and the risk they can carry, see the paper by Joshua Coval, Jakub Jurek, and Erik Stafford.

These were often created from risky tranches of mortgage-backed securities and performed particularly badly once mortgage defaults began to rise. In addition, liquidity guarantees offered by banks as part of securitizations (such as the ABCP discussed by Acharya and his coauthors) now receive higher risk weights and thus require more capital.

These revisions to Basel also introduced a "stressed VaR" calculation, in which banks would need to calculate their potential losses under a "period of significant financial stress."[17] This would address two issues raised above: the procyclicality of capital requirements based on VaR, and the fact that standard VaR implementations tend to underemphasize tail risk. One limitation of stress testing is that it is tempting to use past crises to inform the construction of the stress scenarios (indeed, the Bank for International Settlements explicitly refers to the period of 2007-08), but future crises are likely to be quite different from past ones. This is an intrinsic issue in all systemic risk regulation; while markets continue to evolve, regulators can be trapped in fighting the last crisis.

Basel III. More extensive revisions, known as Basel III, have also been adopted in principle, and individual countries are supposed to adopt rules that would phase them in by the beginning of 2019. In addition to the reforms of international capital regulations undertaken by the Basel committee, there is also a parallel effort under way in the United States. For more details, see *Dodd-Frank and Basel III.*

Strengthened capital requirements. First, capital requirements have been increased in several respects. There is a greater reliance on common equity capital, since equity is a more stable buffer against losses. By contrast, other forms of regulatory capital, which proved to be poor buffers during the financial crisis, now play a more limited role in meeting regulatory capital requirements. For example, two forms of capital used in the past — deferred tax losses and mortgage servicing rights — did not prove to be very good buffers during the financial crisis and are now more restricted.[18] An example of a security that previously was considered as capital but must be phased out under Basel III is trust preferred securities (TruPS). These are hybrid instruments having characteristics of both debt and equity. In particular, like equity, they could count toward capital, but like debt, their dividend payments were tax-deductible for the issuer, which made them attractive to issuing banks. Unfortunately, during the financial crisis it became clear that the debt-like element of these securities meant that they were not able to fully meet their role in stabilizing the bank. For example, TruPS have a fixed term and need to be replaced at maturity (unlike equity). Also, many of these securities had dividends that accumulated if they were not paid; this limited their ability to absorb losses.[19]

In addition, Basel III will also require a *capital conservation buffer.* This buffer consists of an additional 2.5 percent of risk-weighted assets that banks can draw on during times of stress, but doing so will place limits on earnings

Dodd-Frank and Basel III

The Basel framework envisions that each country will adopt the capital regulations at the national level. In the United States, the three large regulators — the Office of the Comptroller of the Currency, the Federal Reserve, and the Federal Deposit Insurance Corporation — adopted rules in July 2013 that detail how many of the revisions to Basel will be implemented.*

In addition, the Dodd-Frank Wall Street Reform and Consumer Protection Act, signed into law on July 21, 2010, also dramatically changes how financial institutions are regulated in the United States. Many of these provisions are quite similar to those formalized in Basel II.5 and Basel III (for example, stress-testing of bank portfolios), and thus little conflict should arise as Basel III is implemented. However, in some cases, Dodd-Frank envisions a very different regulatory approach. One notable example is the use of credit ratings for regulatory purposes: The Basel Accords continue to give these considerable weight, while under Dodd-Frank, regulatory agencies' reliance on credit ratings is drastically curtailed. And indeed, the recently released rules do not incorporate credit ratings. However, some aspects of Basel III are not covered by these rules, and considerable thought will have to be given to their implementation in the U.S.

* For further detail on these rules, see the Federal Reserve Bank of Philadelphia's *Banking Legislation and Policy*, 32:2 (Second Quarter 2012). For an overview of the Dodd-Frank Act, see *Banking Legislation and Policy*, 29:2 (Second Quarter 2010).

[17] The Basel committee gave the period from 2007 to 2008 as one example.

[18] Deferred tax losses were not very valuable when banks were suffering losses. And servicing rights declined in value when the securitized mortgage market shrank dramatically during the crisis.

[19] For further detail on trust preferred securities, see the article by Jennifer Salutric and Joseph Wilcox.

distributions. That is, if losses are large enough that a bank needs to use the buffer to meet its capital requirements, the bank will be restricted in its dividend distributions, stock repurchases, and discretionary executive compensation such as bonuses.[20] Rafael Repullo and Javier Suarez develop a model in which they show that this type of buffer can help mitigate the negative effects resulting from the procyclicality of the Basel II capital requirements.

Basel III will also introduce two capital ratios to supplement the existing one based on risk-weighted assets. The first is a *leverage ratio*, in this case a minimum 3 percent of capital against all assets, without any risk-weighting; the other is the liquidity coverage ratio, which is discussed below.[21] In addition to the leverage ratio adopted in Basel III, in July 2013 U.S. regulators proposed that large institutions be subject to stricter requirements, in particular 5 percent for the largest bank holding companies and 6 percent for their insured depository institutions.

Regulating leverage ratios has several benefits. First, as Tobias Adrian and Hyun Song Shin show, financial institution leverage tends to be very procyclical (rising during booms and falling during busts) and so imposing a maximum leverage ratio can help moderate these cycles. In addition, a simple rule like a leverage ratio is harder to manipulate by shifting portfolios away from activities with high risk weights toward risky activities with low risk weights. That is, the leverage ratio reduces the incentive for regulatory arbitrage. Finally, because it does not rely on complex models to determine the proper risk weight for assets, the leverage ratio may provide better protection against loss even when modelers — at both banks and regulatory agencies — have relatively imprecise knowledge about the true risks, as they inevitably do.[22] However, as Katia D'Hulster points out, the fact that it ignores the risk of assets can also be a weakness; thus, its proper place has typically been viewed as part of a broader framework for capital regulation, rather than as a substitute for risk-sensitive capital requirements.

Systemically important financial institutions (SIFIs). Finally, because of the transmission of shocks from one bank to another during the crisis, capital reform has also focused on increasing capital and supervisory measures for institutions deemed to be "systemically important." Under the Dodd-Frank Wall Street Reform and Consumer Protection Act, U.S. bank holding companies with assets of $50 billion or more will be designated as systemically important. These institutions will be subject to additional regulation; for example, they will be required to develop a "living will" to facilitate their orderly liquidation.[23] In addition, the act tasks the newly established Financial Stability Oversight Council with determining whether nonbanks should be designated as systemically important and subject to Federal Reserve oversight. For example, in June 2013, AIG and GE Capital disclosed that they had been designated as systemically important. The broadening of the SIFI category to include nonbanks is natural, given the key role that nonbank financial institutions — AIG in particular — played in the crisis. In addition to the SIFIs designated by U.S. regulators under the Dodd-Frank Act, the Financial Stability Board has published a list of 29 global systemically important financial institutions (G-SIFIs). Under Basel III, these institutions will be subject to additional capital requirements.

Finally, while I have focused on reforms to international capital regulations, Basel III also adds measures to reduce liquidity risk. See *New Liquidity Requirements Under Basel III*.

[20] Another proposed approach to providing additional capital during times of stress is *contingent capital*. This is debt that automatically converts into equity under certain conditions. For further discussion of contingent capital, see the article by Yaron Leitner.

[21] Some countries, such as the United States and Canada, already use leverage ratios for regulatory capital purposes.

[22] However, the leverage ratio is also subject to manipulation. As documented in the report of the examiner for the Lehman bankruptcy, Lehman Brothers used various accounting maneuvers (such as Repo 105) to reduce the level of debt on its balance sheet.

[23] For further details on how Dodd-Frank changes the regulation of institutions deemed to be systemically important, see the Federal Reserve Bank of Philadelphia's *Banking Legislation and Policy*, 30:4 (Fourth Quarter 2011).

New Liquidity Requirements Under Basel III

We have seen that Northern Rock failed in part because of illiquidity. Basel III adds liquidity requirements. One is the *liquidity coverage ratio*: the requirement that a bank have enough liquid assets to withstand outflows under a 30-day stress scenario. One example would be a significant runoff of wholesale deposits. Wholesale deposits are those obtained through nontraditional demand deposit accounts, such as from Internet accounts. Wholesale deposits tend to be much more mobile and typically evaporate when a bank gets into trouble. Another liquidity requirement added by Basel III is the *net stable funding ratio*, which requires that at least some fraction of long-term assets (such as loans with maturities greater than one year) be funded with long-term financing sources.

CONCLUSION

Capital requirements play an important role in regulating banks' risk-taking and mitigating the consequences of bank failures. Since the 1970s, there has been an effort to harmonize international regulation of banks through the Basel Accords. The financial crisis showed, however, that these regulations still have room for improvement, for example, in how they treat liquidity risk, underweight extreme or "tail" events, and continue to allow scope for regulatory arbitrage. The recent revisions to the Basel Accords are designed to address some of these concerns. Integrating all of these revisions with the Dodd-Frank Act will be another challenge.

REFERENCES

Acharya, Viral V., Thomas Cooley, Matthew Richardson, and Ingo Walter. "Manufacturing Tail Risk: A Perspective on the Financial Crisis of 2007–2009," *Foundations and Trends in Finance*, 4:4 (2009), pp. 247-325

Acharya, Viral, Philipp Schnabl, and Gustavo Suarez. "Securitization Without Risk Transfer," *Journal of Financial Economics*, 107(3) (2013), pp. 515-536.

Adrian, Tobias, and Hyun Song Shin. "Liquidity and Leverage," *Journal of Financial Intermediation*, 19:3 (July 2010), pp. 418-437.

Allen, Linda, "The Basel Capital Accords and International Mortgage Markets: A Survey of the Literature," *Financial Markets, Institutions & Instruments*, 13:2 (May 2004) pp. 41-108.

Allen, Linda, Julapa Jagtiani, and Yoram Landskroner. "Interest Rate Risk Subsidization in International Capital Standards," *Journal of Economics and Business*, 48:3 (August 1996), pp. 251-267.

Berlin, Mitchell. "Can We Explain Banks' Capital Structures?" Federal Reserve Bank of Philadelphia *Business Review* (Second Quarter 2011).

Bernanke, Ben S., and Cara S. Lown. "The Credit Crunch," Brookings Papers on Economic Activity, 2 (1992), pp. 205-239.

Coval, Joshua, Jakub Jurek, and Erik Stafford. "The Economics of Structured Finance," *Journal of Economic Perspectives*, 23:1 (Winter 2009), pp. 3-25.

Dell'Ariccia, Giovanni, and Robert Marquez. "Competition Among Regulators and Credit Market Integration," *Journal of Financial Economics*, 79:2 (2006), pp. 401-430.

D'Hulster, Katia. *The Leverage Ratio*. The World Bank, December 2009.

Financial Stability Forum and the Basel Committee on Bank Supervision Working Group on Bank Capital Issues. "Reducing Procyclicality Arising from the Bank Capital Framework," *FSF-BCBS Joint Report* (April 2009).

Ivashina, Victoria, and David Scharfstein. "Bank Lending During the Financial Crisis of 2008," *Journal of Financial Economics*, 97 (2010), pp. 319-338.

Jackson, Patricia. "Capital Requirements and Bank Behaviour: The Impact of the Basel Accord," Basel Committee on Banking Supervision Working Papers, 1 (April 1999).

Jones, David. "Emerging Problems with the Basel Capital Accord: Regulatory Capital Arbitrage and Related Issues," *Journal of Banking and Finance*, 24 (2000), pp. 35-58.

Leitner, Yaron. "Contingent Capital," Federal Reserve Bank of Philadelphia *Business Review* (Second Quarter 2012).

Myers, Stewart C., and Nicholas S. Majluf. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics,* 13:2, (1984), pp. 187-221.

Puri, Manju, Jorg Rocholl, and Sascha Steffen. "Global Retail Lending in the Aftermath of the U.S. Financial Crisis: Distinguishing Between Supply and Demand Effects." *Journal of Financial Economics*, 100.3 (2011), pp. 556-578.

Repullo, Rafael, and Javier Suarez. "The Procyclical Effects of Bank Capital Regulation," *Review of Financial Studies*, 26:2 (2013), pp. 452-490.

Salutric, Jennifer, and Joseph Wilcox. "Emerging Issues Regarding Trust Preferred Securities," Federal Reserve Bank of Philadelphia, *SRC Insights* (First Quarter 2009), available at http://www.philadelphiafed.org/bank-resources/publications/src-insights/2009/first-quarter/q1si4_09.cfm.

Saunders, Anthony. *Financial Institutions Management: A Modern Perspective*. Boston: Irwin McGraw Hill, 2000.

Shin, Hyun Song. "Reflections on Northern Rock: The Bank Run That Heralded the Global Financial Crisis," *Journal of Economic Perspectives*, 23:1 (2009), pp. 101-119.

Valukas, Anton R. "Report of Examiner to the United States Bankruptcy Court, Southern District of New York," Chapter 11 Case No 08-13555 (March 2010).

Research Rap

Abstracts of research papers produced by the economists at the Philadelphia Fed

Economists and visiting scholars at the Philadelphia Fed produce papers of interest to the professional researcher on banking, financial markets, economic forecasting, the housing market, consumer finance, the regional economy, and more. More abstracts may be found at www.philadelphiafed.org/research-and-data/publications/research-rap/. You can find their full working papers at http://www.philadelphiafed.org/research-and-data/publications/working-papers/.

Modeling the Credit Card Revolution: The Role of Debt Collection and Informal Bankruptcy

In the data, most consumer defaults on unsecured credit are informal, and the lending industry devotes significant resources to debt collection. The authors develop a new theory of credit card lending that takes these two features into account. The two key elements of their model are moral hazard and costly state verification that relies on the use of information technology. They show that the model gives rise to a novel channel through which IT progress can affect outcomes in the credit markets, and argue that this channel can be critical to understand the trends associated with the rapid expansion of credit card borrowing in the 1980s and over the 1990s. Independently, the mechanism of the model helps reconcile high levels of defaults and indebtedness observed in the U.S. data.

Working Paper 13-12. Lukasz A. Drozd, The Wharton School, University of Pennsylvania, Federal Reserve Bank of Philadelphia Visiting Scholar; Ricardo Serrano-Padial, University of Wisconsin.

Who Said Large Banks Don't Experience Scale Economies? Evidence from a Risk-Return-Driven Cost Function

The Great Recession focused attention on large financial institutions and systemic risk. The authors investigate whether large size provides any cost advantages to the economy and, if so, whether these cost advantages are due to technological scale economies or too-big-to-fail subsidies. Estimating scale economies is made more complex by risk-taking. Better diversification resulting from larger scale generates scale economies but also incentives to take more risk. When this additional risk-taking adds to cost, it can obscure the underlying scale economies and engender misleading econometric estimates of them. Using data pre- and post-crisis, they estimate scale economies using two production models. The standard model ignores endogenous risk-taking and finds little evidence of scale economies. The model accounting for managerial risk preferences and endogenous risk-taking finds large scale economies, which are not driven by too-big-to-fail considerations. The authors evaluate the costs and competitive implications of breaking up the largest banks into smaller banks.

Working Paper 13-13/R. Joseph P. Hughes, Rutgers University; Loretta J. Mester, Federal Reserve Bank of Philadelphia and The Wharton School, University of Pennsylvania.

Market Run-Ups, Market Freezes, Inventories, and Leverage

The authors study trade between an informed seller and an uninformed buyer who have existing inventories of assets similar to those being traded. They show that these inventories may lead to prices that increase even absent changes in fundamentals (a "run-up"), but may also make trade impossible (a "freeze") and hamper information dissemination. Competition may amplify the run-up by inducing buyers to enter loss-making trades at high prices to prevent a competitor from purchasing at a lower price and releasing bad news about inventory values. Inventories also prevent seller competition from delivering the Bertrand outcome, in which prices match sellers' valuations. The authors discuss both empirical implications and implications for regulatory intervention in illiquid markets.

Working Paper 13-14. Supersedes Working Paper 12-8. Philip Bond, University of Minnesota; Yaron Leitner, Federal Reserve Bank of Philadelphia.

The Cost of Delay

In this study, the authors make use of a massive database of mortgage defaults to estimate REO liquidation timelines and time-related costs resulting from the recent post-crisis interventions in the mortgage market and the freezing of foreclosures due to "robo-signing" revelations. The cost of delay, estimated by comparing today's time-related costs to those before the start of the financial crisis, is eight percentage points, with enormous variation among states. While costs are estimated to be four percentage points higher in statutory foreclosure states, they are estimated to be 13 percentage points higher in judicial foreclosure states and 19 percentage points higher in the highest-cost state, New York. They discuss the policy implications of these extraordinary increases in time-related costs, including recent actions by the GSEs to raise their guarantee fees 15-30 basis points in five high-cost judicial states. Combined with evidence that foreclosure delays do not improve outcomes for borrowers and that increased delays can have large negative externalities in neighborhoods, the weight of the evidence is that current foreclosure practices merit the urgent attention of policymakers.

Working Paper 13-15. Larry Cordell, Federal Reserve Bank of Philadelphia; Liang Geng, Federal Reserve Bank of Philadelphia; Laurie Goodman, Lidan Yang, Amherst Securities Group, LP.

Improving GDP Measurement: A Measurement-Error Perspective

The authors provide a new and superior measure of U.S. GDP, obtained by applying optimal signal-extraction techniques to the (noisy) expenditure-side and income-side estimates. Its properties — particularly as regards serial correlation — differ markedly from those of the standard expenditure-side measure and lead to substantially revised views regarding the properties of GDP.

Working Paper 13-16. S. Boragan Aruoba, University of Maryland, Federal Reserve Bank of Philadelphia Visiting Scholar; Francis X. Diebold, University of Pennsylvania, Federal Reserve Bank of Philadelphia Visiting Scholar; Jeremy Nalewaik, Federal Reserve Board; Frank Schorfheide, University of Pennsylvania, Federal Reserve Bank of Philadelphia Visiting Scholar; Dongho Song, University of Pennsylvania.

Competition in Bank-Provided Payment Services

Banks supply payment services that underpin the smooth operation of the economy. To ensure an efficient payment system, it is important to maintain competition among payment service providers, but data available to gauge the degree of competition are quite limited. The authors propose and implement a frontier-based method to assess relative competition in bank-provided payment services. Billion dollar banks account for around 90 percent of assets in the U.S., and those with around $4 to $7 billion in assets turn out to be both the most and the least competitive in payment services, not the very largest banks.

Working Paper 13-17. Wilko Bolt, De Nederlandsche Bank; David Humphrey, Florida State University, Federal Reserve Bank of Philadelphia Visiting Scholar.

Dynamics of Investment, Debt, and Default

How does physical capital accumulation affect the decision to default in developing small open economies? The authors find that, conditional on a level of foreign indebtedness, more capital improves the sovereign's ability to meet its obligations, reducing the likelihood of default and the risk premium. This effect, however, is diminishing in the stock of capital because capital also tames the severity of the contraction following default, making autarky more appealing. Access to long-term debt and costly capital adjustment are crucial for matching business cycles. Their quantitative model delivers default episodes that mimic those observed in the data.

Working Paper 13-18. Grey Gordon, University of Indiana; Pablo Guerrón-Quintana, Federal Reserve Bank of Philadelphia.

Estimating Dynamic Equilibrium Models with Stochastic Volatility

The authors propose a novel method to estimate dynamic equilibrium models with stochastic volatility. First, they characterize the properties of the solution to this class of models. Second, the authors take advantage of the results about the structure of the solution to build a sequential Monte Carlo algorithm to evaluate the likelihood function of the model. The approach, which exploits the profusion of shocks in stochastic volatility models, is versatile and computationally tractable even in large-scale models, such as those often employed by policy-making institutions. As an application, the authors use their algorithm and Bayesian methods to estimate a business cycle model of the U.S. economy with both stochastic volatility and parameter drifting in monetary policy. Their application shows the importance of stochastic volatility in accounting for the dynamics of the data.

Working Paper 13-19. Jesús Fernandez-Villaverde, University of Pennsylvania, Federal Reserve Bank of Philadelphia Visiting Scholar; Pablo Guerrón-Quintana, Federal Reserve Bank of Philadelphia; Juan F. Rubio-Ramírez, Duke University, Federal Reserve Bank of Philadelphia Visiting Scholar.

Subsidizing Price Discovery

When markets freeze, not only are gains from trade left unrealized, but the process of information production through prices, or price discovery, is disrupted as well. Though this latter effect has received much less attention than the former, it constitutes an important source of inefficiency during times of crisis. The authors provide a formal model of price discovery and use it to study a government program designed explicitly to restore the process of information production in frozen markets. This program, which provided buyers with partial insurance against acquiring low-quality assets, reveals a fundamental trade-off for policymakers: while some insurance encourages buyers to bid for assets when they otherwise would not, thus promoting price discovery, too much insurance erodes the informational content of these bids, which hurts price discovery.

Working Paper 13-20. Braz Camargo, Sao Paulo School of Economics – FGV; Kyungmin (Teddy) Kim, University of Iowa; Benjamin Lester, Federal Reserve Bank of Philadelphia.

Credit Ratings and Bank Monitoring Ability

In this paper the authors use credit rating data from two large Swedish banks to elicit evidence on banks' loan monitoring ability. For these banks, their tests reveal that banks' credit ratings indeed include valuable private information from monitoring, as theory suggests. However, their tests also reveal that publicly available information from a credit bureau is not efficiently impounded in the bank ratings: The credit bureau ratings not only predict future movements in the bank ratings but also improve forecasts of bankruptcy and loan default. The authors investigate possible explanations for these findings. Their results are consistent with bank loan officers placing too much weight on their private information, a form of overconfidence. To the extent that overconfidence results in placing too much weight on private information, risk analyses of the bank loan portfolios in the authors' data could be improved by combining the bank credit ratings and public credit bureau ratings. The methods the authors use represent a new basket of straightforward techniques that enable both financial institutions and regulators to assess the performance of credit rating systems.

Working Paper 13-21. Supersedes Working Paper 10-21. Leonard I. Nakamura, Federal Reserve Bank of Philadelphia; Kasper Roszbach, Sveriges Riksbank, University of Gronigen.

Trend-Cycle Decomposition: Implications from an Exact Structural Identification

A well-documented property of the Beveridge-Nelson trend-cycle decomposition is the perfect negative correlation between trend and cycle innovations. The authors show how this may be consistent with a structural model where trend shocks enter the cycle, or cyclic shocks enter the trend and that identification restrictions are necessary to make this structural distinction. A reduced-form unrestricted version such as that of Morley, Nelson and Zivot (2003) is compatible with either option, but cannot distinguish which is relevant. They discuss economic interpretations and implications using U.S. real GDP data.

Working Paper 13-22. Mardi Dungey, University of Tasmania, CFAP, University of Cambridge, CAMA; Jan P.A.M. Jacobs, University of Groningen, University of Tasmania, CAMA, CIRANO; Jing Tian, University of Tasmania; Simon van Norden, HEC Montréal, CAMA, CIRANO, CIREQ, Federal Reserve Bank of Philadelphia Visiting Scholar.

Large Capital Infusions, Investor Reactions, and the Return and Risk-Performance of Financial Institutions over the Business Cycle

The authors examine investors' reactions to announcements of large capital infusions by U.S. financial institutions (FIs) from 2000 to 2009. These infusions include private market infusions (seasoned equity offerings (SEOs)) as well as injections of government capital under the Troubled Asset Relief Program (TARP). The sample period covers both business cycle expansions and contractions, and the recent financial crisis. They present evidence on the factors affecting FIs' decisions to raise capital, the determinants of investor reactions, and post-infusion risk-taking of the recipients, as well as a sample of matching FIs. Investors reacted negatively to the news of private market SEOs by FIs, both in the immediate term (e.g., the two days surrounding the announcement) and over the subsequent year, but positively to TARP injections. Reactions differed depending on the characteristics of the FIs, and the stage of the business cycle. More financially constrained institutions were more likely to have raised capital through private market offerings during the period prior to TARP, and firms receiving a TARP injection tended to be riskier and more levered. In the case of TARP recipients, they appeared to finance an increase in lending (as a share of assets) with more stable financing sources such as core deposits, which lowered their liquidity risk. However, the authors find no evidence that banks' capital adequacy increased after the capital injections.

Working Paper 13-23. Supersedes Working Paper 11-46. Elyas Elyasiani, Fox School of Business and Management, Temple University, and Fellow, Wharton Financial Institution Center; Loretta J. Mester, Federal Reserve Bank of Philadelphia, The Wharton School; Michael S. Pagano, Villanova School of Business, Villanova University.

Credit Access and Credit Performance After Consumer Bankruptcy Filing: New Evidence

This paper uses a unique data set to shed new light on the credit availability and credit performance of consumer bankruptcy filers. In particular, the authors' data allow them to distinguish between Chapter 7 and Chapter 13 bankruptcy filings, to observe changes in credit demand and supply explicitly, to differentiate existing and new credit accounts, and to observe the performance of each credit account directly. The paper has four main findings. First, despite speedy recovery in their risk scores after bankruptcy filing, most filers have much reduced access to credit in terms of credit limits, and the impact seems to be long lasting. Second, the reduction in credit access stems mainly from the supply side as consumer inquiries recover significantly after the filing, while credit limits remain low. Third, lenders do not treat Chapter 13 filers more favorably than Chapter 7 filers. In fact, Chapter 13 filers are much less likely to receive new credit cards than Chapter 7 filers even after controlling for borrower characteristics and local economic environment. Finally, the authors find that Chapter 13 filers perform more poorly than Chapter 7 filers (after the filing) on all credit products (credit card debt, auto loans, and first mortgages). Their results, in contrast to prior studies, thus suggest that the current bankruptcy system does not appear to provide much relief to bankruptcy filers.

Working Paper 13-24. Julapa Jagtiani, Federal Reserve Bank of Philadelphia; Wenli Li, Federal Reserve Bank of Philadelphia.

Congestion, Agglomeration, and the Structure of Cities

Congestion pricing has long been held up by economists as a panacea for the problems associated with ever increasing traffic congestion in urban areas. In addition, the concept has gained traction as a viable solution among planners, policymakers, and the general public. While congestion costs in urban areas are significant and clearly represent a negative externality, economists also recognize the advantages of density in the form of positive agglomeration externalities. The long-run equilibrium outcomes in economies with multiple correlated, but offsetting, externalities have yet to be fully explored in the literature. To this end, the author develops a spatial equilibrium model of urban structure that includes both congestion costs and agglomeration externalities. The author then estimates the structural parameters of the model by using a computational solution algorithm and matches the spatial distribution of employment, population, land use, land rents, and commute times in the data. Policy simulations based on the estimates suggest that naive optimal congestion pricing can lead to net negative economic outcomes.

Working Paper 13-25. Jeffrey C. Brinkman, Federal Reserve Bank of Philadelphia.

Stress Tests and Information Disclosure

The authors study an optimal disclosure policy of a regulator who has information about banks' ability to overcome future liquidity shocks. They focus on the following trade-off: Disclosing some information may be necessary to prevent a market breakdown, but disclosing too much information destroys risk-sharing opportunities (Hirshleifer effect). The authors find that during normal times, no disclosure is optimal, but during bad times, partial disclosure is optimal. They characterize the optimal form of this partial disclosure. The authors also relate their results to the debate on the disclosure of stress test results.

Working Paper 13-26. Itay Goldstein, University of Pennsylvania; Yaron Leitner, Federal Reserve Bank of Philadelphia.

Reverse Mortgage Loans: A Quantitative Analysis

Reverse mortgage loans (RMLs) allow older homeowners to borrow against housing wealth without moving. In spite of growth in this market, only 2.1 percent of eligible homeowners had RMLs in 2011. In this paper, we analyze reverse mortgages in a life-cycle model of retirement, calibrated to age-asset profiles. The ex-ante welfare gain from RMLs is sizable at $1,000 per household; ex-post, low-income, low-wealth and poor-health households use them. Bequest motives, nursing-home moving risk, house price risk, and interest and insurance costs all contribute to the low take-up rate. The model predicts market potential for RMLs to be 5.5 percent of households.

Working Paper 13-27. Makoto Nakajima, Federal Reserve Bank of Philadelphia; Irina A. Telyukova, University of California, San Diego.

Banking Crises and the Role of Bank Coalitions

The goal of this paper is to provide a framework to analyze the effectiveness of bank coalition formation in response to an external aggregate shock that may cause disruption to the payment mechanism and real economic activity. The author shows that the kind of insurance mechanism provided by a specific type of bank coalition allows society to completely prevent any disruption to real activity that can be caused by a temporary drop in the value of banking assets, at least in the case of a shock that is not too big. If the shock is relatively large, then a private bank coalition will be unable to completely prevent a disruption in real activity even though it will be able to substantially mitigate the effects on equilibrium quantities and prices. Thus, the existence of a private bank coalition of the kind described in this paper can be an effective means of preventing significant disruptions in trading activity.

Working Paper 13-28. Daniel Sanches, Federal Reserve Bank of Philadelphia.

Macroeconomic Dynamics Near the ZLB: A Tale of Two Equilibria

This paper studies the dynamics of a New Keynesian dynamic stochastic general equilibrium (DSGE) model near the zero lower bound (ZLB) on nominal interest rates. In addition to the standard targeted-inflation equilibrium, the authors consider a deflation equilibrium as well as a Markov sunspot equilibrium that switches between a targeted-inflation and a deflation regime. The authors use the particle filter to estimate the state of the U.S. economy during and after the 2008–09 recession under the assumptions that the U.S. economy has been in either the targeted-inflation or the sunspot equilibrium. The authors consider a combination of fiscal policy (calibrated to the American Recovery and Reinvestment Act) and monetary policy (that tries to keep interest rates near zero) and compute government spending multipliers. Ex-ante multipliers (cumulative over one year) under the targeted-inflation regime are around 0.9. A monetary policy that keeps interest rates at zero can raise the multiplier to 1.7. The ex-post (conditioning on the realized shocks in 2009–11) multiplier is estimated to be 1.3. Conditional on the sunspot equilibrium, the multipliers are generally smaller and the scope for conventional expansionary monetary policy is severely limited.

Working Paper 13-29. S. Borağan Aruoba, University of Maryland, Federal Reserve Bank of Philadelphia Visiting Scholar; Frank Schorfheide, University of Pennsylvania, NBER, Federal Reserve Bank of Philadelphia Visiting Scholar.

100 YEARS™
FEDERAL RESERVE SYSTEM

To Mark Our Centennial

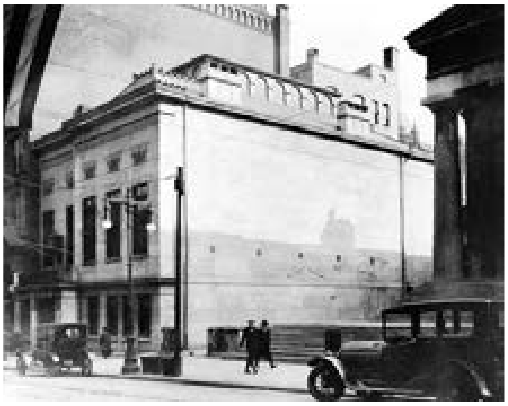
To mark the 100th anniversaries of the signing of the Federal Reserve Act in 1913 and the opening of the Federal Reserve Banks in 1914, the Fed is asking scholars, historians, and other members of the public to help compile an inventory of records, collections, and artifacts related to the history of the nation's central bank. Do you know of materials that should be included? Information may be submitted at http://www.federalreserve.gov/apps/contactus/feedback.aspx.

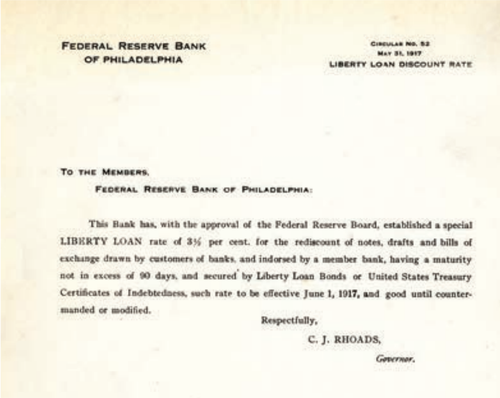
The inventory will give researchers, academics, and others interested in studying U.S. central banking a single point of electronic access to documents, photographs, and audio and video recordings from sources across the Federal Reserve System, universities, and private collections. Information is also being included about material not yet available online

On December 23, 1913, President Woodrow Wilson signed the Federal Reserve Act, establishing the Federal Reserve System as the U.S. central bank. Its mission is to conduct the nation's monetary policy; supervise and regulate banks; maintain the stability of the financial system; and provide financial services to depository institutions, the U.S. government, and foreign official institutions.

Congress designed the Fed with a decentralized structure. The Federal Reserve Bank of Philadelphia — serving eastern Pennsylvania, southern New Jersey, and Delaware — is one of 12 regional Reserve Banks that, together with the seven-member Board of Governors in Washington, D.C., make up the Federal Reserve System. The Board, appointed by the President of the United States and confirmed by the Senate, represents the public sector, while the Reserve Banks and the local citizens on their boards of directors represent the private sector.

The Research Department of the Philadelphia Fed supports the Fed's mission through its research; surveys of firms and forecasters; reports on banking, markets, and the regional and U.S. economies; and publications such as the *Business Review*.




FEDERAL RESERVE BANK
OF PHILADELPHIA

CIRCULAR NO. 52
MAY 31, 1917
LIBERTY LOAN DISCOUNT RATE

TO THE MEMBERS,
FEDERAL RESERVE BANK OF PHILADELPHIA:

This Bank has, with the approval of the Federal Reserve Board, established a special LIBERTY LOAN rate of 3½ per cent. for the rediscount of notes, drafts and bills of exchange drawn by customers of banks, and indorsed by a member bank, having a maturity not in excess of 90 days, and secured by Liberty Loan Bonds or United States Treasury Certificates of Indebtedness, such rate to be effective June 1, 1917, and good until countermanded or modified.

Respectfully,
C. J. RHOADS,
Governor.

Federal Reserve Bank of Philadelphia

Ten Independence Mall
Philadelphia, Pennsylvania 19106-1574

ADDRESS SERVICE REQUESTED

www.philadelphiafed.org

Past and current *Business Review* articles can be downloaded for free from the Federal Reserve Bank of Philadelphia website. There you will also find data and reports on the regional and national economy, the latest research by our economists, information on consumer finance and community development, resources for teachers, and more.

EXPLORE AND LEARN