2016
Richmond Fed Research Digest

Summaries of work by economists in the Bank's Research Department
published externally from June 1, 2015, through May 31, 2016

Welcome to the fifth annual issue of the *Richmond Fed Research Digest*. The Federal Reserve Bank of Richmond produces several publications that feature the work of economists in its Research Department, but those economists also publish extensively in other venues. The *Richmond Fed Research Digest*, a mid-year annual, brings this externally published research together in one place with brief summaries, full citations, and links to the original work. (Please note that access to articles may require registration or payment.) So bookmark this spot on the Richmond Fed website and mark your calendar for June 30, 2017, when the Bank will publish the next issue of the *Richmond Fed Research Digest*.

A Road Map for Efficiently Taxing Heterogeneous Agents

By Marios Karabarbounis
American Economic Journal: Macroeconomics, *April 2016, vol. 8, no. 2, pp. 182–214*

Marios Karabarbounis of the Richmond Fed analyzes how labor-income taxes should vary by age, household assets, and filing status within a life-cycle model with heterogeneous, two-member households and endogenous human capital. The paper's key innovation is a labor supply elasticity that varies endogenously among households.

Karabarbounis finds that tax distortions should be hump-shaped according to age, decrease according to household assets, and be lower for joint filers relative to single filers.

Standard public finance principles imply that the government should decrease tax distortions on workers with a larger value of labor supply elasticity. To distinguish between workers of low and high labor supply elasticity, the government can use information on their characteristics. For example, a worker closer to retirement is more likely to quit her job if her taxes increase, as is a worker who is not the main breadwinner.

Karabarbounis finds that tax distortions should be hump-shaped according to age, decrease according to household assets, and be lower for joint filers relative to single filers. Age and assets act as complements within this optimal tax policy. In contrast, filing status neither complements nor crowds out the age and asset tags.

The key modeling assumption that generates differences in labor supply elasticity is an extensive margin of labor supply coupled with rich heterogeneity. Heterogeneity is introduced through i) a life-cycle dimension, ii) permanent and temporary uninsurable labor productivity shocks, iii) endogenous human capital accumulation, and iv) households with two (potential) earners, a male and a female. Changes in tax rates will affect only those workers whose reservation wage is sufficiently close to the market wage.

The first finding is an optimal tax schedule hump-shaped in age: young households receive tax cuts since they tend to earn less. By contrast, middle-aged households are strongly attached to their jobs, so large tax distortions tend to have small efficiency loss. In contrast, older households are very sensitive to tax changes, because they are more likely to retire early if their taxes increase. By decreasing their tax rates, accordingly, the alternative tax code encourages these households to delay retirement. The second finding is an optimal labor-income tax that decreases (linearly) in household assets. A novel feature of the optimal tax code is the strong complementarity between age- and asset-dependent taxes.

The third finding makes the case for a large tax subsidy toward households with two earners. Most single-earner households would switch to a two-earner household in this scenario, while only a small fraction would switch to non-employment. This reflects the large labor supply elasticity for the secondary earner and the relatively lower elasticity for the primary earner. When age (or household assets) is used together with filing status, the gains are almost equal to the additive sum of the separate policies. Hence the policies do not interact, but they do not crowd out each other, either.

Selection and Monetary Non-Neutrality in Time-Dependent Pricing Models

Infrequent price changes at the microeconomic level do not necessarily imply that monetary disturbances have large real macroeconomic effects. For the same frequency of price changes, the real effects of a monetary shock are small if the firms adjusting their prices are also the ones most likely to change prices by a large amount. The importance of this selection effect has been well-understood at least since 1987, when Andrew S. Caplin and Daniel F. Spulber published "Menu Costs and the Neutrality of Money" in the *Quarterly Journal of Economics*. In their model, large price adjustments by a small fraction of firms completely offset monetary shocks and induce money neutrality because there is self-selection. In other words, firms always have the option of incurring a menu cost to adjust their prices, so adjusting firms are also the ones that would like to adjust their prices by the greatest amount.

In an article in the *Journal of Monetary Economics*, Carlos Carvalho of Pontifical Catholic University of Rio de Janeiro and Felipe Schwartzman of the Richmond Fed argue that selection effects do not necessarily hinge on self-selection. They investigate different ways in which the shape of the distribution of duration of price spells influences the real effects of nominal aggregate demand shocks. They highlight a mechanism that has received very little attention in the literature: a selection for the time since prices were last adjusted. In fact, they show that selection provides a complete characterization of the distribution of price durations in time-dependent sticky-price models. This is because the real effects of a monetary shock differ depending on whether adjusting firms are more or less likely to have prices that pre-date the shock. More fundamentally, the authors show that in such an economy, for a given average frequency of price changes, the real effects of a monetary shock depend solely on this type of selection. In particular, the real effects of nominal shocks are larger if older prices are relatively less likely to be adjusted.

Carvalho and Schwartzman also analyze the implications of selection for private and social efficiency in the presence of monetary and other kinds of shocks. The fact that firms and society

face different objective functions, together with the fact that aggregate and idiosyncratic shocks may have different dynamic properties, can lead to a tension between private and social efficiency. Therefore, they find that firms may choose a pattern of selection that is suboptimal from a social perspective.

The authors conclude that a careful characterization of the distribution of price durations is of crucial importance for the proper evaluation of the aggregate implications of nominal price stickiness. While their results are derived for the case of time-dependent pricing, there is no reason why the selection effect they identify should not hold some relevance more broadly whenever the timing of price changes is not entirely up to firms. This suggests that further research on price setting would do well to focus on models that are able to fully account for the distribution of price spells and investigate the extent to which the mechanisms emphasized by Carvalho and Schwartzman continue to matter.

What Inventory Behavior Tells Us about How Business Cycles Have Changed

By Pierre-Daniel Sarte, Felipe Schwartzman, and Thomas A. Lubik
Journal of Monetary Economics, *November 2015, vol. 76, pp. 264–283*

Real business cycle (RBC) models have been highly successful at explaining business cycles that occurred before 1984, but since then, shifts in comovements and relative volatilities of key economic aggregates have challenged their preeminence. Prior to 1984, for example, labor productivity was strongly procyclical, but after 1984, it became nearly acyclical. At the same time, labor productivity switched from comoving positively with hours before 1984 to comoving negatively with hours after 1984. Both of these shifts challenge the notion that the productivity shocks used in the standard RBC model drive business cycles all by themselves.

By extending the standard RBC model to incorporate financial frictions, the authors are better able to explain the comovement and relative volatilities of key economic aggregates in post-1984 business cycles.

In a *Journal of Monetary Economics* article, Pierre-Daniel Sarte, Felipe Schwartzman, and Thomas A. Lubik of the Richmond Fed attempt to explain post-1984 business cycles by extending the neoclassical RBC model to include multiple stages of production, a process that starts with planning and design and progresses through coordination of suppliers, manufacturing of products, distribution to retailers, and ultimately, sales to consumers. This feature combines detailed production structures from the RBC literature with a mechanism that provides an incentive for firms to redirect some of their currently available labor and capital toward the production of goods in the future. This redirection of resources is what gives rise to inventories.

Analyzing inventory behavior in this way enables the authors to distinguish variations in the physical return to investment from variations in the discount rate. Variations in the discount rate become distinguishable because they affect inventories and fixed investment in similar ways, whereas variations in the physical return on investment affect fixed investment disproportionately. The discount rate influences how firms allocate resources over time, and these decisions are reflected in levels of inventories. So a substantial increase in fluctuations of inventories relative to output (captured by a less-countercyclical inventory/sales ratio) indicates that variations in the discount rate have become key drivers of post-1984 recessions.

The authors test their results by comparing their estimates of variations in the discount rate to other independent measures of credit-market frictions. They find that the measure that emerges from their model correlates well with a wide array of measures of credit-market frictions, including the lagged spreads between Treasury bonds and Baa-rated bonds, dividend payouts to business owners, and the fraction of U.S. banks that tighten lending standards.

In the spirit of the RBC literature, Sarte, Schwartzman, and Lubik attempt to gauge how well variations in technology alone account for the changing nature of business cycles. They find that the effects of fluctuations in technology are qualitatively in line with various comovement properties and relative volatilities of economic aggregates prior to 1984, but they are unable to account for the bulk of the variation in hours after that date and, therefore, changes in the behavior of labor productivity. By extending the standard RBC model to incorporate financial frictions, however, they are better able to explain the comovement and relative volatilities of key economic aggregates in post-1984 business cycles.

Optimal Monetary Policy with Heterogeneous Money Holdings

In a variety of models, monetary policy has persistent effects on the distribution of wealth, including money and other assets. These distributional effects are often muted or kept to a minimum in theoretical models to preserve tractability.

In an article in the *Journal of Economic Theory*, Francesco Lippi of the Einaudi Institute for Economics and Finance, Stefania Ragni of the University of Sassari, and Nicholas Trachter of the Richmond Fed explore how these distributional concerns affect the optimal monetary policy in a monetary model that allows for a full dynamic analysis of two-way interactions between monetary policy and wealth distribution.

Two forces shape the design of optimal policy in their setup. As in other monetary models, distortions are minimized when the return on money equals the rate of time discount (i.e., Friedman's rule). However, with uncertainty and incomplete markets, an expansionary policy can be desirable due to insurance needs. A trade-off arises as expanding the liquidity base dampens the return on the asset, therefore reducing the production incentives. The optimal regulation of the money supply strikes a balance between these two forces. Since the relevance of these forces depends on the wealth distribution, which evolves through the business cycle, the optimal (anticipated) monetary policy is state dependent.

The novelty of their research is that they acknowledge this dependence and explore how the state-dependent policy balances the costs of anticipated inflation with the needs for insurance along the business cycle. This policy allows for a dramatic improvement in welfare compared with a policy that does not respond to the state. The optimal policy expands the supply of liquidity when the unproductive agents are poor (that is, when the insurance needs are large), whereas it contracts the liquidity base otherwise to maximize production incentives. The principle underlying this prescription is due to the state-dependent redistributive role of monetary policy, and it differs from the one arising in sticky-price models. Because aggregate production is low when the unpro-

ductive group is poor and high when they are rich, the best policy can be interpreted as counter-cyclical. Interestingly, in spite of the policy being far away from contracting the money supply at the rate of preference, the optimal policy "echoes" Friedman's rule as the expected real return of money approaches the rate of preference.

While the specific predictions of the analysis likely depend on the details of the model, Lippi, Ragni, and Trachter highlight the potential relevance of a transmission channel of monetary policy that seems reasonable and relatively unexplored.

Equilibrium Price Dispersion with Sequential Search

By Guido Menzio and Nicholas Trachter

In this paper, Guido Menzio of the University of Pennsylvania and Nicholas Trachter of the Richmond Fed study equilibrium pricing in a product market with search frictions. As in Peter A. Diamond's 1971 article, "A Model of Price Adjustment," buyers search sequentially for sellers. But in contrast to Diamond's model, buyers do not meet all sellers with the same probability. In particular, some of the buyers' searches lead to one large seller, while their remaining searches lead to small sellers.

In this environment, the small sellers would like to set a price that makes the buyers indifferent between purchasing the good and searching for another seller. The large seller would like to price the small sellers out of the market. These incentives give rise to a game of cat-and-mouse, whose only equilibrium involves mixed strategies for both large and small sellers that generate price dispersion. The fact that the small sellers play mixed strategies implies that there is price dispersion in equilibrium. The fact that the large seller plays mixed strategies implies that prices and allocations vary from one day to the next.

The authors' theory of price dispersion follows directly from the assumption that there are some sellers whose pricing decisions affect the search strategy of the buyers and, hence, the number of buyers who end up visiting their stores.

Menzio and Trachter show that buyers capture a positive fraction of the gains from trade as long as the large seller has some, but not total, market power. When this is the case, the fraction of gains from trade accruing to the buyers depends in a natural way on the extent of search frictions. Specifically, buyers capture all of the gains from trade when search frictions vanish, and they capture none of the gains from trade when search frictions become infinitely large. The only difference between the environment considered in this paper and the one analyzed by Diamond is the presence of a large seller. The large seller's price affects the buyers' decision of whether to purchase from the small sellers and, hence, the number of buyers who meet and purchase from the large seller. This mechanism is sufficient to break the Diamond Paradox.

The authors' theory of price dispersion follows directly from the assumption that there are some sellers whose pricing decisions affect the search strategy of the buyers and, hence, the number of buyers who end up visiting their stores. First, for a seller to impact the buyers' search strategy, it has to be the case that (some) buyers meet that seller with positive probability. This condition seems reasonable. Second, for a seller to impact the buyers' search strategy, it has to be the case that (some) buyers know the seller's price before they start searching. This condition seems at least plausible. For example, it may be the case that some buyers learn about the seller's price in the newspaper and, afterward, meet a random sequence of

sellers on their way to and from work. Third, for a seller's price to impact the number of buyers who visit his store, it has to be the case that (some) competitors cannot respond to his pricing decision. This condition seems reasonable, as some competitors may incur physical costs and time delays in adjusting their prices.

http://dx.doi.org/10.1016/j.jet.2015.09.004

Drifts and Volatilities under Measurement Error: Assessing Monetary Policy Shocks over the Last Century

By Pooyan Amir-Ahmadi, Christian Matthes, and Mu-Chun Wang
Quantitative Economics*, forthcoming*

Pooyan Amir-Ahmadi of Goethe University Frankfurt, Christian Matthes of the Richmond Fed, and Mu-Chun Wang of the University of Hamburg study more than 100 years of U.S. economic data on inflation, real output, short-term and long-term nominal interest rates, as well as money growth to assess how the dynamics of the economy have changed and how the impact of monetary policy shocks has evolved over time.

Before tackling these questions, they confront the measurement issues inherent in historical macroeconomic time series, as discussed by Christina D. Romer in her 1989 article, "The Prewar Business Cycle Reconsidered: New Estimates of Gross National Product, 1869–1908." Romer concluded that the large volatility in U.S. GDP data before the end of World War II was substantially due to measurement error. With this in mind, Amir-Ahmadi, Matthes, and Wang combine a model of possibly mismeasured historical data with a time-varying parameter VAR model with stochastic volatility similar to those employed by Giorgio E. Primiceri in his 2005 article, "Time Varying Structural Vector Autoregressions and Monetary Policy," and by Timothy Cogley and Thomas J. Sargent in their 2005 article, "Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII US." The authors' model of measurement error builds on those used by Cogley, Sargent, and Paolo Surico in their 2015 article, "Price-Level Uncertainty and Instability in the United Kingdom," as well as Cogley and Sargent in their 2014 article, "Measuring Price-Level Uncertainty and Instability in the U.S., 1850–2012," and a 2014 paper by Frank Schorfheide, Dongho Song, and Amir Yaron, "Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach."

Amir-Ahmadi, Matthes, and Wang use this combined framework to assess the importance of measurement error and time variation in the dynamics of historical macroeconomic time series. Like Romer, they find substantial evidence of measurement error before and during World War II (particularly in GDP), time variation in volatilities of reduced form innovations, and substantial time variation in the correlations between the macroeconomic variables they consider. In particular, their model associates the early 1980s with substantial shifts in the structure of the economy.

The authors also use their model to study whether the evolution of the Federal Reserve as an institution over the past 100 years has altered the transmission of monetary policy shocks. They find wide variation in responses to monetary policy shocks, but they attribute most of this variation to changes in exogenous volatility. Once they condition on the size of the shock (i.e., the initial impact on short-term nominal interest rates), their most noteworthy finding is the surprising stability in the transmission of monetary policy shocks during the past century.

http://qeconomics.org/ojs/forth/475/475-4.pdf

Optimal Banking Contracts and Financial Fragility

Banks and other financial intermediaries engage in maturity transformation by issuing short-term liabilities while investing in long-term assets. In fact, many of an intermediary's liabilities are payable on demand. During normal times, these liabilities are like debt in the sense that the value of a claim typically does not depend on the precise timing of withdrawal. But during episodes of high withdrawal demand, claims that are repaid late in the process can be discounted considerably. Such intermediation arrangements often appear to be fragile in the sense of being susceptible to runs — events in which liability holders rush *en masse* to redeem their claims, causing an inefficient allocation of resources.

Ennis and Keister show how both the simple, debt-like features of banking contracts and financial fragility may have a common origin in the gradual revelation of information that is inherent in banking arrangements.

In an article in *Economic Theory*, Huberto M. Ennis of the Richmond Fed and Todd Keister of Rutgers University provide a model of financial intermediation that accounts for all of these features simultaneously. Their model is a finite-depositor version of the Diamond–Dybvig model of financial intermediation in which the bank and all depositors observe withdrawals sequentially, as they occur. The authors derive the constrained efficient allocation of resources in closed form and show that this allocation provides liquidity insurance to depositors. The contractual arrangement that decentralizes this allocation resembles a standard bank deposit with a demandable debt-like structure. When withdrawals are unusually high, however, depositors who withdraw relatively late experience significant losses. This contractual arrangement can be fragile, admitting an equilibrium in which some depositors run on the bank by withdrawing funds regardless of their liquidity needs.

In Ennis and Keister's model, a depositor's actions are observed by the bank and other depositors only when she withdraws. The authors show this specification of sequential service is tractable and generates results that are intuitive and perhaps more realistic than those in the existing literature. Depositors, for example, are able to learn before making a decision what payout they would receive by withdrawing. Moreover, even though the authors' approach allows for very complex patterns of payouts, the optimal arrangement resembles in many ways a traditional banking contract, particularly when the number of depositors is large. Under this arrangement, depositors who withdraw early receive (approximately) the face value of their deposits. If the number of withdrawals becomes unexpectedly high, however, depositors begin experiencing significant discounts in what they receive from the bank. Because of these discounts, a run in this model is necessarily partial, with only some depositors participating.

In much of the previous literature, the banking arrangement is fragile (in the sense of being susceptible to a self-fulfilling run) only if banking contracts are exogenously restricted to be of a simple form. Ennis and Keister impose no such restrictions in their model. Instead, they show how both the simple, debt-like features of banking contracts and financial fragility may have a common origin in the gradual revelation of information that is inherent in banking arrangements.

Learning about Fiscal Policy and the Effects of Policy Uncertainty

The financial crisis and accompanying recession of 2007–09 have motivated greater interest in the role that uncertainty about government policy plays in determining economic outcomes. In a paper in the *Journal of Economic Dynamics and Control*, Josef Hollmayr of the Deutsche Bundesbank and Christian Matthes of the Richmond Fed examine the impact of uncertainty about fiscal policy on the economic effect of fiscal policy changes when the government tries to counteract a deep recession.

Hollmayr and Matthes model fiscal policy in an environment in which agents adaptively learn about the economy. In contrast to papers such as Chryssi Giannitsarou's 2006 article, "Supply-side Reforms and Learning Dynamics," Stefano Eusepi and Bruce Preston's 2012 article, "Debt, Policy Uncertainty, and Expectations Stabilization," and Emanuel Gasteiger and Shoujian Zhang's 2014 article "Anticipation, Learning and Welfare: The Case of Distortionary Taxation," Hollmayr and Matthes construct a model in which agents know the behavior of all other agents except for the fiscal authority. Such agents are uncertain about current as well as future fiscal policy rules, but they are aware that the government budget constraint must hold.

Aside from the learning by agents, the model is a mostly standard real business cycle (RBC) model of a frictionless closed economy with distortionary taxation and government spending, based on that of Eric Leeper, Michael Plante, and Nora Traum's 2010 paper, "Dynamics of Fiscal Financing in the United States." Private firms and households share the same beliefs and act as Bayesian econometricians, using observed economic outcomes to estimate policy rule coefficients. The authors use a Monte Carlo simulation of 100 periods, beginning at the steady state associated with the original policy parameters. The simulation then experiences a negative technology shock in period nine, followed by a permanent average increase in government spending starting in period ten. First running this simulation under a full-information rational expectations model, the authors find that this fiscal policy change causes a small short-term increase in output, followed by a long-run return to previous steady-state output levels; however, the capital stock is permanently lowered due to increases in debt and the capital tax rate.

Next, the authors run this simulation under the learning model described above. They find that under a benchmark in which fiscal policy represents a two-standard-deviation shock under the beliefs of agents, uncertainty increases substantially during the policy change but rapidly decreases following it, implying that agents learn quickly. Nevertheless, in the long run consumption decreases below original levels, leading to a persistent negative difference in GDP and the capital stock compared to the rational expectations model. This outcome suggests that a short period of "confusion" can have persistent effects.

Tweaking the model to analyze scenarios in which agents have either more or less prior uncertainty than in the benchmark, Hollmayr and Matthes find a clear pattern: more uncertainty is associated with longer learning periods, higher short-run volatility, and bigger average outcome differences between learning and rational expectations models. The meaningful differences between the two models point toward a possible role for communicating policy changes. However, the authors caution that a number of issues can arise when a policymaker wants to announce a policy change but is unsure of the effects of such communication on the public's views.

http://dx.doi.org/10.1016/j.jedc.2015.08.002

Discussion of "Inflation Targeting: A Victim of Its Own Success" by Gillitzer and Simon

By Alexander L. Wolman

In the *International Journal of Central Banking*, Alexander Wolman of the Richmond Fed has written a discussion comment to accompany "Inflation Targeting: A Victim of Its Own Success," by Christian Gillitzer and John Simon of the Reserve Bank of Australia. Their paper joins a broader debate among economists about what inflation targeting (IT) can and cannot do, and why it was not able to prevent the Great Recession.

Focusing on Australia, the authors illustrate the success of IT in multiple dimensions, including a decreased sensitivity of inflation and inflation expectations to shocks, and a "de-linking" of traded-goods prices from inflation. They use the successes of inflation targeting to refute critics urging for wholesale changes, instead arguing for changes at the margin. Wolman largely agrees with the authors but provides a slightly different perspective than the authors on the decomposition of Australian traded and nontraded goods inflation. He also adds some comments on the theme of IT as a victim of its own success and suggests it may make sense to put more weight on some price changes than others in determining the optimal volatility of inflation.

> **While the paper addresses the theme of how much inflation variability should be tolerated, Wolman argues that another question deserves much more attention: What is the smallest feasible variability in inflation?**

Gillitzer and Simon compare the pre- and post-IT regimes in Australia, noting that the variance of nontradables price changes fell dramatically, while there was little change in the variance of tradables price changes; meanwhile, the covariance between tradable and nontradable price changes vanished. The authors view the prices of tradables as reflecting external influences to a large extent, whereas IT succeeded where they think it mattered, in stabilizing the prices of nontradables. Wolman agrees with this interpretation but argues that it is really the relative price change of tradables that should be viewed as reflecting external influences. He adds that the decline in variance may reflect a more stable overall inflationary environment, so that price changes became more effective.

Wolman adds an observation about IT regimes worldwide, noting that they were introduced in the hope that they would bring about or reinforce a secular decline in inflation. Foreseeing success (as many countries did), economists also could have foreseen the inevitable criticism: real fluctuations wouldn't disappear and would lead to a discussion of whether monetary policy should have done more to dampen them. Wolman concludes that economists cannot answer the question of what monetary policy should do without knowing what monetary policy can do. And economists still are not certain as to what monetary policy can do or how much inflation stability it can achieve. (See the summary of "What Monetary Policy Can Do" beginning on page 13.)

In sum, while the paper addresses the theme of how much inflation variability should be tolerated, Wolman argues that another question deserves much more attention: What is the smallest feasible variability in inflation? By looking at the distributions of realized inflation across countries, he says, economists can at least find upper bounds for the smallest feasible variation in inflation. From the U.S. perspective, this question is especially relevant right now. Inflation is widely perceived to be "low" over the last three years, but is it meaningfully low given the kind of variation a central bank must accept as inevitable?

http://www.ijcb.org/journal/ijcb15q4a9.pdf

Price Cap Regulation in a Two-Sided Market: Intended and Unintended Consequences

In this paper, Zhu Wang of the Richmond Fed studies the consequences of price cap regulation in the two-sided payment card market. Under the 2010 Dodd-Frank Act (DFA), a new debit card regulation was intended to lower merchants' card-acceptance costs by capping interchange fees at the issuer cost. However, for small-ticket transactions, the interchange fee instead actually rose after the regulation. To address the puzzle, Wang constructs a two-sided market model that shows how card demand externalities between large-ticket and small-ticket transactions could rationalize the pricing response of card networks. Based on the model, the paper also provides a welfare assessment of the regulation and discusses possible improvement.

Payment cards are considered two-sided markets because card networks serve two distinct end-user groups: cardholders and merchants. In practice, card networks and their issuers charge high interchange fees to merchants who accept their cards, but they also provide rewards to consumers for using those cards. There is concern that this highly skewed pricing structure may distort payments efficiency by inflating merchants' costs of accepting cards. As a result, more than twenty countries have regulated or started investigating interchange fees. In the United States, the DFA's Durbin Amendment required the Federal Reserve to regulate debit card interchange fees. The maximum fee for covered issuers was capped at half of its pre-regulation industry average level. This cap caused covered issuers to lose billions of dollars in annual interchange revenues. However, the regulation also generated unintended consequences. Prior to the regulation, merchants were charged differentiated interchange fees based on transaction sizes. Post-regulation, however, card networks set a uniform interchange fee at the maximum cap amount. As a result, small-ticket transactions that used to pay lower interchange fees now face an increased rate.

Why would card networks raise fees on small-ticket transactions in response to a fee cap? According to Wang's research, merchants engage in transactions of different sizes, and consumers' benefits from using cards at certain transaction sizes are positively affected by their card usage in others, or "ubiquity externalities." Prior to the regulation, card networks and issuers were willing to offer subsidized interchange fees for small-ticket transactions because their card acceptance boosted consumers' card usage for large-ticket purchases—and from the latter, card issuers could collect higher interchange fees. Once a cap on interchange fees was imposed, however, card issuers profited less from this externality, so they discontinued the subsidy.

The analysis shows that absent regulation, the market-determined interchange fees yield little total user surplus (the sum of consumer surplus and merchant profit). This explains why policymakers who care about end users (consumers and merchants) wanted the regulation in the first place. The analysis also shows in spite of hurting small-ticket transactions, the regulation may improve the total user surplus by capping down interchange fees. However, an issuer cost-based regulation lacks theoretical foundation and could result in unintended consequences. For one thing, such a regulation ignores the two-sided nature of the market and may run the risk of undershooting or overshooting. Especially in the latter case, the regulation could push the interchange fee too low so that a higher interchange fee may actually improve both the total user surplus and the issuer profit. For another, the regulation overlooks card demand externalities across different transactions, which may lead to the negative impact on small-ticket transactions that was seen in the market. In light of the model findings, some alternative regulatory approaches are discussed.

Mutual Assistance between Federal Reserve Banks: 1913–1960 as Prolegomena to the TARGET2 Debate

By Barry Eichengreen, Arnaud Mehl, Livia Chitu, and Gary Richardson

The trans-European automated real-time gross settlement express transfer system (TARGET2) has served as the European Monetary Union's real payment and settlement service. During the European debt crisis, TARGET2 imbalances emerged as euro area "periphery" countries experienced capital outflows and turned to euro area national central banks for funding against collateral. These imbalances sparked a heated debate between those who believed TARGET2 balances were integral to the functioning of the euro area and those who viewed TARGET2 imbalances as disguised bailouts of troubled euro area banks and sovereigns. Participants in this debate regularly invoked the Federal Reserve's structure as an inspiration for possible remedies to TARGET2's problems.

In a *Journal of Economic History* article, Barry Eichengreen of the University of California, Berkeley, Arnaud Mehl and Livia Chitu of the European Central Bank, and Gary Richardson of the University of California, Irvine and the Richmond Fed examine the history of mutual assistance between the twelve Federal Reserve Banks from 1913–60 in order to speak to the current debate over TARGET2 balances in Europe. Before 1975, gold settlements were on paper drawn from individual Reserve Bank reserve holdings daily, resembling a possible approach to reforming the TARGET2 system by capping balances at a specific level over a specified period. In practice, however, Reserve Banks assisted each other by mutualizing gold reserves in emergencies, such as liquidity crises and bank runs, in what were known as interdistrict accommodation operations, creating gold flows from districts with payment surpluses to those with deficits.

The authors assemble data on accommodation operations by the Federal Reserve districts during the practice of daily settlements, paying particular attention to severe asymmetric shocks such as the post-World War I slump, the Great Depression, and World War II, all of which bear strong similarities to the recent crisis in Europe. During the 1920–21 slump, coastal Reserve Banks took part in accommodation operations by lending to interior Reserve Banks, whose districts were hit by declines in the price of agricultural products. The authors find a significant negative impact of bank suspensions on accommodation during this period. During the Great Depression, the New York Fed was hit especially hard by gold losses, and Chicago and other Banks were hesitant to share gold reserves with New York due to their own difficulties. In March 1933, the Board of Governors instructed the other Reserve Banks to rediscount paper on behalf of the New York Fed; the authors argue that this accommodation between Reserve Banks was essential to the stability of the U.S. monetary union. Examining these shocks and others reveals that the Federal Reserve System did not have a well-defined "core" or "periphery"—districts frequently went from being emergency recipients of gold to providers. Furthermore, regional payment imbalances stabilized and reversed once panics subsided, with interdistrict accommodations reaching low levels, so mutual assistance did not create high regional tensions. Analyzing the dynamics in TARGET2 balances from 1999–2014, the authors find similar patterns to those of Federal Reserve System balances from 1913–60.

The history of mutual assistance among Reserve Banks, the authors conclude, illustrates the importance of maintaining a cooperative spirit for the smooth operation of the euro area. The tendency of Reserve Bank imbalances to narrow following crises suggests that concerns about TARGET2 imbalances boundlessly growing are unfounded and that the notions of the euro area's "core" and "periphery" are not set in stone.

Review of *Strained Relations: U.S. Foreign-Exchange Operations and Monetary Policy in the Twentieth Century* by Bordo, Humpage, and Schwartz

In *Strained Relations: U.S. Foreign-Exchange Operations and Monetary Policy in the Twentieth Century*, Michael D. Bordo, Owen F. Humpage, and Anna J. Schwartz write a history of intervention in foreign exchange markets by the U.S. Treasury and Federal Reserve System. Examining such events as the 1934 implementation of the Exchange Stabilization Fund and the Plaza and Louvre Accords, the authors conclude that frequent intervention in foreign exchange markets failed to solve the "trilemma." Instead, "the intervention and its associated institutions weakened the Federal Reserve's credibility for price stability."

Strained Relations documents a number of failures resulting from monetary authorities attempting to treat the foreign exchange value of the dollar as an independent instrument. For instance, U.S. efforts to use the dollar as a weapon against Germany and Japan following the 1987 Louvre agreement necessitated inflationary monetary policy, requiring monetary contraction starting in 1989. Such failures seem to confirm Milton Friedman's hypothesis that such attempts would destabilize the economy, and the authors' analysis shows interventions in exchange rates produced desired movements no more often than would have occurred by chance.

In a recent review of *Strained Relations* in the *Journal of Economic History*, Robert L. Hetzel of the Richmond Fed notes these findings and asks how policymakers can learn from their mistakes. His answer is that identifying such mistakes requires characterization of the evolution of monetary policy, but monetary policymakers' language of discretion makes such characterization difficult. By methodically recording the behavior of the Fed regarding foreign exchange markets, the authors of *Strained Relations* have, Hetzel argues, greatly aided efforts to characterize the evolution of monetary policy in order to prevent central banks from making future mistakes.

Comment on "The ECB as Lender of Last Resort: A Monetary Perspective" by Winkler

In a recent paper in the *Journal of Economics and Statistics*, Adalbert Winkler of the Frankfurt School of Finance and Management argues that the European Central Bank's (ECB) program of outright monetary transactions (OMT) was required in order to preserve the stability of the European Monetary Union's (eurozone's) financial system. In a controversial departure from existing policy, with the OMT program, the ECB proposed purchasing the sovereign debt of individual member countries of the eurozone. As a result, OMT effectively made the ECB into a lender of last resort for eurozone governments. Winkler concludes there is a proper role for OMT in the ECB,

although decisive steps toward fiscal and banking union are needed for it to act as a successful long-term lender of last resort for governments.

In a comment in the same journal, Robert L. Hetzel of the Richmond Fed highlights the issues of moral hazard and democratic accountability raised by using the ECB's balance sheet to make the kinds of fiscal transfers inherent in the OMT program. Hetzel observes that fiscal implications necessarily accompany the program because of the consequences for the creation and distribution of seigniorage revenues. Seigniorage concerns how money creation transfers resources to governments, thereby reducing their budget deficits.

Hetzel also places the debate in the context of desirable monetary policy in the eurozone. According to Hetzel, there are two competing views of the causes of disinflation and the back-to-back recessions in the eurozone: the forced-debt-deleveraging view and the contractionary-monetary-policy view. The former sees the recession as punishment for excessive assumption of debt while the latter emphasizes contractionary monetary policy and the incompatibility of a common foreign exchange value of the euro for countries with vastly different export sectors.

Both the Winkler and Hetzel views therefore highlight a number of critical fiscal issues involved in a policy of quantitative easing. Because long-term bonds issued by eurozone governments have extraordinarily low yields, Hetzel argues that a program of quantitative easing will require a massive replacement of bonds with bank reserves. Such a policy would succeed only if the ECB convinces the public that it is essential to the survival of the eurozone and that it will be carried out aggressively until successful.

http://www.degruyter.com/view/j/jbnst.2015.235.issue-3/issue-files/jbnst.2015.235.issue-3.xml

What Monetary Policy Can Do

By Jeffrey M. Lacker
Cato Journal*, Spring/Summer 2016, vol. 36, no. 2, pp. 261–268*

Reconciling the behavior of monetary policy with the behavior of inflation has been more difficult since the financial crisis of 2008. The dramatic increase in the Federal Reserve System's monetary liabilities—from just under $1 trillion to over $4 trillion—led to warnings from some critics that surging inflation was imminent. But that hasn't happened. To the contrary, inflation has remained persistently below the Federal Open Market Committee's 2 percent goal, prompting other critics to question whether monetary policy still has the ability to determine inflation and the price level over time.

In a speech before the Cato Institute's 33rd Annual Monetary Conference, later published in a special issue of the *Cato Journal*, Richmond Fed President Jeffrey M. Lacker expressed confidence that the Fed still does have the capacity to control inflation—even though the mechanism for doing so has changed. Before the crisis, if the Fed wanted to lower short-term interest rates, it conducted open market operations to increase bank reserves. Now reserve account balances earn interest, and bank reserves have increased dramatically. Some economists have argued that in this new regime, bank reserves are perfect substitutes for short-term Treasury securities. As a result, monetary policy may have become relatively impotent because institutions simply hold fewer Treasury securities and more bank reserves, leaving economic activity unaffected.

Lacker noted, however, that this theory overlooks a key characteristic of bank reserves: while any financial entity can hold Treasury securities, only banks can hold bank reserves. Banks could shed

other assets to accommodate larger reserve account balances, but there is an upper limit to that response. At some point, banks would have to raise more capital to accommodate larger reserve account balances. This action would force broader changes in banks' portfolios that would inevitably affect economic outcomes, including the price level.

In a July 2014 working paper at the Richmond Fed, economist Huberto M. Ennis provides an explicit model that captures this logic. The intuition is that when the quantity of bank reserves is small enough and interest rates are above what the Fed pays on excess reserves, price-level determination would work in the usual way. And when the quantity of bank reserves is large enough, banks would be forced to adjust their balance sheets, and again, the quantity of Fed liabilities would affect the price level. Between these two scenarios, however, there is a broad zone in which the quantity of bank reserves could vary without affecting the price level.

"This analysis bolsters my confidence that the intuition of the standard approach remains relevant and monetary policy still has the capacity to determine inflation and the price level over time," Lacker said. He also emphasized that the Fed's ability to do so is essentially independent of its capacity to affect real economic activity—a capacity he views as limited and temporary. Lacker concluded that the Fed should exercise caution in trying to use monetary policy to affect real economic activity. "Conducting monetary policy to achieve low and stable inflation over time, without doing damage to real activity, is hard enough," he said.

http://object.cato.org/sites/cato.org/files/serials/files/cato-journal/2016/5/cj-v36n2-4.pdf

Comprehending and Regulating Financial Crises: An Interdisciplinary Approach

By Nina Bandelj, Julia Elyachar, Gary Richardson, and James Owen Weatherall
Perspectives on Science, *forthcoming*

A broad perspective on the 2008 financial crisis does not yet exist, either in an understanding of its causes or in prescriptions for preventing future crises. In a forthcoming *Perspectives on Science* paper, Nina Bandelj, Julia Elyachar, and James Owen Weatherall of the University of California, Irvine, as well as Gary Richardson of Cal-Irvine and the Richmond Fed, argue that this lack of a consensus results from "silo thinking," wherein experts in one area ignore the methods and research of other areas. The authors contend that only the development of a critical interdisciplinary approach to financial markets and financial regulation can achieve a broad, macro-scale understanding of the 2008 financial crisis.

To make their argument, they describe four in-progress research projects attempting to answer a single question: With what kinds of models and financial crises do market actors and regulators operate, and with what consequences? These projects come from four disciplines—philosophy of science, sociology, economics, and anthropology—and each project rests on assumptions or raises issues to which only other disciplines can properly respond.

The first project studies the methodology of financial model construction. Traditionally, such models have assumed that assets have normally distributed returns; however, the 1987 Black Monday stock market crash and subsequent crises have led to a more widespread use of models with heavy-tailed distributions. An understanding of how methodologies of such models vary among communities, how they are understood by different groups, and how crises are defined in relation to these models would require the use of sociology and anthropology.

Recognizing the important role the news media plays in influencing financial markets, the second project relies on sociology to inquire into the kind of notions about financial models and crises that are proliferated in the media. Drawing upon the insights of behavioral economics and business psychology, this project examines how interpretations of crises and uses of models vary through time and place and by social position of interpreters.

Bandelj, Elyachar, Weatherall, and Richardson conclude that investors, regulators, and academics must move beyond silo thinking toward a critical interdisciplinary approach in order to avoid future crises.

The third project draws on experimental economics to randomly and anonymously assign forty individuals to ten four-person investment groups from which individuals can withdraw their investments at the beginning of each round. After the last round, groups in which no one withdrew receive a higher payoff than individuals who withdrew; individuals who stayed in a group from which others withdrew receive a payoff of zero. Preliminary results of this experiment suggest that investors pulling out can trigger panics. Informed by sociology and philosophy of science, this experiment incorporates unique design variations and questions assumptions usually unexamined in economic experiments.

The fourth project employs methods from anthropology, specifically the ethnography of financial crises, to focus on regulatory agencies and central banks. Gathering information from research assistants, economists, mid-level officials, and mainframe computers, the researchers will first conduct an ethnography of the New York Fed to study policy interventions not only as they are planned, but also as they work out on paper. Next, the researchers will interview employees of local investment and regulatory groups in Orange County, Calif., to investigate the mechanisms for transmission of knowledge and models among academics, the policy world, financial services, and consumers.

Each of these projects relies on considerations from other fields to identify hidden assumptions and fill gaps left by intradisciplinary approaches. This reliance demonstrates the broader and deeper insight that interdisciplinary studies can provide. Thus, Bandelj, Elyachar, Weatherall, and Richardson conclude that investors, regulators, and academics must move beyond silo thinking toward a critical interdisciplinary approach in order to avoid future crises.

http://dx.doi.org/10.1162/POSC_a_00216

Liquidity Risk, Bank Networks, and the Value of Joining the Federal Reserve System

By Charles W. Calomiris, Matthew Jaremski, Haelim Park, and Gary Richardson
NBER Working Paper No. 21684, October 2015

A large volume of research shows that the Federal Reserve System succeeded to an extent in its founding goal of stabilizing the American banking system by offering banks access to liquidity through its discount window. However, the Fed failed to achieve its goal of universal membership, as less than 8 percent of state-chartered banks joined the Federal Reserve in its first decade. The Fed would have more effectively limited systemic liquidity risk had it achieved universal membership, as discussed by Kris James Mitchener and Gary Richardson in their 2016 paper, "Network Contagion and Interbank Amplification during the Great Depression," which is summarized on page 18.

In a National Bureau of Economic Research working paper, Charles Calomiris of Columbia University, Matthew Jaremski of Colgate University, Haelim Park of the Department of the Treasury,

and Gary Richardson of the University of California, Irvine and the Richmond Fed study the slow response of state-chartered banks to the opportunity to join the Federal Reserve System. The authors first describe the regulatory environment at the time of the Fed's founding. From 1864 onward, the U.S. banking system was a dual system of state-chartered and national banks. This framework lacked an efficient means of routing payments, so the Federal Reserve was founded to operate an efficient nationwide payments system in addition to correcting regulatory flaws. Only national banks were required to join the Fed, with members being subjected to bank size and zero-interest reserve requirements, among other regulations. By June 1920, only 1,374 of 20,962 state-chartered banks had joined.

The authors propose several reasons for this low level of participation. First, in 1914 regulations in all but twenty states imposed heavy costs on membership. Then, World War I imposed wartime tasks upon the Federal Reserve, extending access to the discount window beyond Fed members. Even after the war, nonmember banks could indirectly access the discount window and access cash through correspondent member banks. Also, members faced high reserve requirements, narrow restrictions on assets that could be brought to the discount window, and regulatory red tape.

Next, the authors use a sample of 190 banks and seventy-seven trust companies from New York to provide the first detailed bank-level analysis of state banks' decision whether to join the Fed in its first decade. Using a log-logistic survival model, the authors find that a bank's size, capital sufficiency, and loan swing were all significant determinants of membership, with increases in all three leading to quicker adoption of membership. These findings support the view that banks large enough to absorb compliance costs of membership but too large to rely on local members for pass-through lending found greater net value in membership. Banks with greater pass-through access to the Federal Reserve were less likely to become members, meaning the moral hazard problem of shadow banking was already present during this era.

Using the same sample, the authors examine how membership changed banks' behavior from 1915–24. After joining the Fed, banks experienced an immediate increase in the ratio of loans to assets and a more gradual decrease in seasonal loan variation. These changes suggest that Fed membership reduced the liquidity risk of greater lending and that banks with a lower ability to mitigate seasonal loan demand variation had stronger incentives to join the Federal Reserve.

http://www.nber.org/papers/w21684

Relative Price Dispersion: Evidence and Theory

By Greg Kaplan, Guido Menzio, Leena Rudanko, and Nicholas Trachter
NBER Working Paper No. 21931, January 2016

In this paper, Greg Kaplan of Princeton University, Guido Menzio of the University of Pennsylvania, Leena Rudanko of the Philadelphia Fed, and Nicholas Trachter of the Richmond Fed use the Kilts-Nielsen Retail Scanner Dataset to measure the extent and sources of price dispersion. They focus on the dispersion of prices for the same good in different stores in the same market in the same week. They find that a significant fraction of price dispersion is due to the fact that stores that are equally expensive choose to set persistently different prices for the same good. They call this phenomenon "relative price dispersion."

The authors develop a theory of relative price dispersion in the context of the canonical model established by Kenneth Burdett and Kenneth L. Judd in their 1983 article, "Equilibrium Price Dispersion." Kaplan, Menzio, Rudanko, and Trachter's model is an equilibrium manifestation of the

sellers' attempts to discriminate between different buyers. In particular, an individual seller finds it optimal to charge asymmetric prices for different goods to discriminate between high-valuation buyers who need to purchase everything in the same location and low-valuation buyers who are willing to purchase different goods at different locations. In equilibrium, for every seller who charges a relatively high price for one good, there must be another seller who is equally expensive, on average, but charges a relatively low price for the same good. The authors show that their theory of relative price dispersion can account not only for the extent and sources of price dispersion, but also for the extent and sources of variation in the prices paid by different households for the same basket of goods.

The authors suggest several extensions of their theory. On the descriptive side, they propose combining their model with a model of temporary price reductions as in Joel Sobel's 1984 article, "The Timing of Sales," or in Victor Aguirregabiria's 1999 article, "The Dynamics of Markups and Inventories in Retailing Firms." The resulting framework would offer a comprehensive theory of price dispersion, in which price dispersion occurs because some stores are cheap and some are expensive, because equally expensive stores have different average prices for the same good, and because the same store posts different prices for the same good on different days. As in this paper, the resulting theory could be tested using data on the dispersion of prices paid by different households. On the normative side, the authors suggest using the calibrated model to measure the extent of inefficiency in the retail market, identifying which policies might be welfare-improving and which ones might exacerbate inefficiencies.

http://www.nber.org/papers/w21931

Equilibrium Price Dispersion across and within Stores

By Guido Menzio and Nicholas Trachter
NBER Working Paper No. 21493, August 2015

For centuries, economists have adhered to the law of one price, the theory that the same good should sell for the same price in all locations of a free and efficient market. But real-world observations of substantial and pervasive price dispersion have contradicted this theory too often to be explained away as brief deviations from equilibrium.

Over the years, models of price dispersion have tended to be spatial (different prices at different locations), while models of price discrimination have tended to be intertemporal (different prices at different times). But in a National Bureau of Economic Research working paper, Guido Menzio of the University of Pennsylvania and Nicholas Trachter of the Richmond Fed combine the insights from both theories into a unified search-theoretic model that generates price dispersion across and within stories. Their new framework attributes price dispersion primarily to differences among buyers' ability and willingness to shop around. The fact that some buyers shop at multiple stores drives spatial price dispersion, and the fact that some buyers shop at different times drives intertemporal price dispersion.

Menzio and Trachter model a market for an indivisible good. On the demand side, some buyers purchase the good from only one seller, while other buyers shop around. In addition, some buyers shop only during the day, while others shop during the day and during the night. On the supply side, there are identical sellers, and each seller can vary the daytime and nighttime price of the good. (In describing their model, Menzio and Trachter use "daytime" as shorthand for convenient times and "nighttime" as shorthand for less-convenient times.) Equilibrium in their model always

features price dispersion among stores because sellers encounter some buyers who shop at only one store and other buyers who shop at multiple stores. This difference prompts sellers to periodically post lower prices (hold sales) to attract some portion of the shop-around crowd. Moreover, if the buyers who shop day and night also shop from multiple stores, then equilibrium also features price variation within stores because sellers can charge lower prices at night—to compete for some portion of the shop-around crowd—without losing revenues from their daytime customers.

Menzio and Trachter suggest that it would be interesting to estimate their model using the econometric techniques developed by Han Hong and Matthew Shum in their 2006 article, "Using Price Distributions to Estimate Search Costs," and José Luis Moraga-Gonzáles and Matthijs R. Wildenbeest in their 2008 article, "Maximum Likelihood Estimation of Search Costs." Menzio and Trachter also suggest that it would be useful to integrate their model with the multiproduct model developed in their 2016 paper with Greg Kaplan and Leena Rudanko, "Relative Price Dispersion: Evidence and Theory," to build a unified framework for studying retail pricing. (A summary of that article begins on page 16.)

http://www.nber.org/papers/w21493

Network Contagion and Interbank Amplification during the Great Depression

By Kris James Mitchener and Gary Richardson
NBER Working Paper No. 22074, March 2016

For decades, scholars have questioned how financial networks propagate shocks, increase systemic risk, and magnify economic downturns, and economic theory suggests many channels through which financial networks may transmit shocks. In contrast to that theoretical inquiry, a National Bureau of Economic Research working paper by Kris James Mitchener of Santa Clara University and Gary Richardson of the University of California, Irvine and the Richmond Fed provides an empirical analysis of how interbank connections amplified the effects of the Great Depression.

The authors find that interbank networks did indeed worsen the contraction in lending during the Great Depression. Banking panics induced banks in the hinterland to withdraw interbank deposits from Federal Reserve member banks that were located in reserve and central reserve cities. These correspondent banks responded by curtailing lending to businesses.

In particular, this paper documents how interbank balances exposed correspondent banks to shocks afflicting banks in the hinterland. Interbank deposits were a liquid source of funds that could be employed to meet sudden demands by depositors to convert claims to cash, and the removal of those deposits from correspondent banks peaked during periods that the authors classify as banking panics. During these panics, withdrawals of interbank deposits forced correspondent banks to reduce lending to businesses. These interbank outflows led to a substantial decline in aggregate lending, equal to approximately 15 percent of the total decline in commercial bank lending from the peak in 1929 to the trough in 1933, according to the authors.

This paper illuminates both a mechanism for transmission (interbank deposits) as well as a source of amplification (balance-sheet effects). It also introduces an additional channel through which banking distress deepened the Great Depression and complements existing research on how bank distress during the Great Depression influenced the real economy.

These findings also may relate to the 2008 financial crisis. By the mid-2000s, shadow banks had accumulated assets that were at least as large as those in the commercial banking system. Some scholars have argued that the initial distress in financial markets originated from shadow bank transactions, such as those in the asset-backed commercial paper and repurchase agreement ("repo") markets. The run on the shadow banking system began in 2007, prior to the economic downturn in early 2008, and distress was spread from these markets to other parts of the financial system partly through shadow-bank linkages with traditionally regulated banking and insurance firms.

The authors are careful to note, however, that in the Great Depression, output fell prior to severe financial institution distress. Industrial production in the United States peaked in the summer of 1929, with the first banking panic coming more than a year later in October 1930. Further, the precise way in which network linkages led to disintermediation differs in the two crises. In the 1930s, the decline in interbank balances was sufficient to reduce aggregate lending, and the effect was sizeable. Some scholars of the recent crisis, on the other hand, argue that the destruction of shadow banks has been the source of overall disintermediation, suggesting that commercial banks have been unable to fill the void in credit resulting from their disappearance.

http://www.nber.org/papers/w22074

Bank Leverage and Regulatory Regimes: Evidence from the Great Depression and Great Recession

By Christoffer Koch, Gary Richardson, and Patrick Van Horn
American Economic Review: Papers and Proceedings, *May 2016, vol. 106, no. 5, pp. 538–542*

Policymakers and scholars have argued that leverage choices of the largest banks in the United States from 2001 through 2007 contributed to the financial crisis of 2008–09. And regulators and legislators have acted on this belief, strengthening capital requirements via Basel III and the Dodd-Frank Act. Economic theory indicates, however, that regulations that reduce the risk of runs, such as deposit insurance, distort incentives for investment, generating inefficient allocations and shifting risks from bankers to governments.

A key insight of this literature—that modern institutions distort banks' behavior toward risk—has been difficult to test directly. An ideal test would examine a panel of banks operating in an environment without these modern institutions and then randomly change the incentives of some banks relative to others. In an article in *American Economic Review: Papers and Proceedings*, Christoffer Koch of the Dallas Fed, Gary Richardson of the Richmond Fed, and Patrick Van Horn of Southwestern University examine historical events that approximate such an experiment. They compare the capital choices of large versus small banks in the United States before and after the creation of the modern institutions that distort commercial banks' capital choices. They focus on two episodes when individual banks' capital choices may have had aggregate consequences: the booms preceding the Great Depression and the Great Recession. In the 1920s, stockholders of commercial banks faced double liability. Directors and most senior executives typically held substantial equity in their institutions, and they often faced civil suits and criminal prosecution when courts liquidated their failed banks. Large firms received no special treatment. In sharp contrast, on the eve of the Great Recession, bankers' liabilities were limited, deposit insurance was ubiquitous, and governments had a history of bailing out banks believed to be too big to fail.

The authors' estimation strategy relies on the concept of treatment and control. From the 1920s to the 2000s came the rise of institutions deemed too big to fail. So the authors build a panel of banks in which the treated institutions are the Manhattan money center banks of the 1920s and their modern descendants. The untreated group contains many smaller institutions that existed in both periods, but some of these banks operated in only one period.

The authors find that during the boom before the Great Depression, large banks held capital substantially above regulatory minimums, had capital ratios much lower than smaller banks, and increased capital relative to assets (deleveraging) as the boom progressed. At the peak of the cycle, the capital ratios of the largest banks matched and in many cases exceeded those of the smaller banks. In contrast, before the Great Recession, the largest commercial banks in the United States typically held the minimum capital required by law until the peak of the expansion. The largest banks held much less capital relative to assets or to risk than smaller institutions, and this relationship between larger and smaller institutions remained constant over the cycle.

The authors conclude that the changing behavior of the money center (treated) banks in their study is consistent with the theoretical literature on the impact of too-big-to-fail policies and modern prudential policy initiatives. Their findings further suggest that scholars and policymakers need to be realistic about the potential benefits of institutional and regulatory reforms. While the institutional framework of the 1920s induced money center banks to hold excess capital relative to risk, U.S. and international financial systems still collapsed.

http://dx.doi.org/10.1257/aer.p20161045

FEDERAL RESERVE BANK OF RICHMOND®

Richmond ▪ Baltimore ▪ Charlotte