



2017

# Richmond Fed Research Digest

Summaries of work by economists in the Bank's Research Department  
published externally from June 1, 2016, through May 31, 2017

Welcome to the sixth 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 29, 2018, when the Bank will publish the next issue of the *Richmond Fed Research Digest*.

---

## Indeterminacy and Learning: An Analysis of Monetary Policy in the Great Inflation

By Thomas A. Lubik and Christian Matthes

*Journal of Monetary Economics*, September 2016, vol. 82, pp. 85–106

There are three basic narratives about the Great Inflation and the Great Moderation in the academic literature. One narrative offers a good luck/bad luck explanation. According to this view, the 1970s was a decade with frequent and strong economic shocks. It was simply bad luck to be a central banker at that time because it proved difficult to stabilize the economy. In the 1980s, however, fewer and less persistent shocks led to the Great Moderation. An almost opposite viewpoint is based on the quality of policy. According to this view, the Federal Reserve conducted bad policy in the 1970s under Chairman Arthur Burns, who was not aggressive enough in fighting inflation. Chairman Paul Volcker adopted good policy in the 1980s by raising interest rates enough to vanquish the Great Inflation and usher in the Great Moderation. A third narrative takes an intermediate view that the Federal Reserve did not perceive the economic situation of the 1970s correctly. Substantial data errors and misperceptions about the state of the economy led the Fed to implement policies that delivered bad outcomes. This problem abated in the 1980s as an improved understanding of the state of the economy led to better monetary policy.

In an article in the *Journal of Monetary Economics*, Richmond Fed economists Thomas Lubik and Christian Matthes integrate the bad/good policy narrative with the misperception narrative. The authors argue that the Great Inflation can be understood as the result of an equilibrium in which loose monetary policy engendered excess volatility in macroeconomic aggregates and prices. The Fed inadvertently pursued policies that were not sufficiently anti-inflationary because the Fed did not fully understand the economic environment in which it was operating. Specifically, it had imperfect knowledge about the structure of the U.S. economy, and it was subject to data misperceptions. This combination of imperfect knowledge and data-measurement errors resulted in policies that seemed optimal at the time but led to an indeterminate equilibrium in the economy.

Lubik and Matthes refine the interpretation of the Great Inflation and the Great Moderation in two ways. First, they note that the recession of 1974–75 coincided with a switch to an aggressively anti-inflationary monetary policy stance that led to a determinate equilibrium. The Fed leaned against a perceived strong increase in inflation, but it erroneously reversed course in late 1975 when the data indicated a relative decline in inflation. The loosening of policy, as described by an optimal policy rule, was then enough to induce indeterminacy. Second, the authors find that the so-called Volcker disinflation appears not as an abrupt shift in policy but rather as the culmination of a gradual adjustment process that made policy more inertial after 1975.

Interestingly, the authors' results also provide some support to the good luck/bad luck way of thinking because in a model where the central bank has access to the true, final data, a stable determinate equilibrium does not occur until 1993. In other words, measurement errors in the 1970s represented bad luck because they led to bad policy, but measurement errors in the 1980s represented good luck because they magnified the perception of persistent inflation and output dynamics, which strengthened the Fed's resolve to maintain an aggressive policy stance.

<http://dx.doi.org/10.1016/j.jmoneco.2016.07.006>

---

## Payment Choice and Currency Use: Insights from Two Billion Retail Transactions

By Zhu Wang and Alexander L. Wolman

*Journal of Monetary Economics*, December 2016, vol. 84, pp. 94–115

For several decades, the U.S. payments system has been migrating from cash and check toward debit and credit cards. As economists study this trend, one major impediment is the difficulty of tracking consumers' use of cash. Most studies rely on data from consumer surveys with relatively small samples and relatively narrow coverage of locations and time.

An article in the *Journal of Monetary Economics* by Zhu Wang and Alexander Wolman helps to bridge that gap. The authors report new evidence on cash use in retail transactions—as well as on credit, debit, and check use—based on data from a discount retail chain with thousands of stores across the country. The data cover nearly two billion transactions involving millions of consumers from April 2010 through March 2013. The authors use the data to explore three themes: payment variation across transaction sizes and locations, payment variation by day of week and day of month, and payment variation over the longer term. They link their empirical findings to theoretical work on money demand and payment choice, paying special attention to models that offer a “threshold framework,” in which each consumer has a threshold transaction size for using cash. According to this framework, the share of cash transactions is the fraction of shoppers whose transaction sizes are below their thresholds, so the cash share is lower for larger transactions.

**The authors report new evidence on cash use in retail transactions—as well as on credit, debit, and check use—based on data from a discount retail chain with thousands of stores across the country.**

Regarding payment variation across transaction sizes and locations, the authors confirm that the use of cash decreases with transaction size in a given zip code. They also find that the cross-location dispersion of the payment mix increases with transaction size, which indicates that the threshold distributions across locations exhibit more variation for larger transaction sizes. A quantitative decomposition reveals that the estimated threshold distributions cannot be explained by transaction-size fixed effects. Instead, they are determined primarily by location-specific characteristics, which in part proxy for access to noncash payment options.

Regarding day-of-week and day-of-month patterns, the authors find that the fraction of cash payments declines over the month and also has a regular weekly pattern. Over the course of a week, the cash and debit fractions are nearly mirror images of each other, but credit comes closer to mirroring cash over the course of a month. In addition, the number of transactions shows time patterns similar to those for the cash share. The authors interpret the high correlation between the number of transactions and the cash share of transactions as indicating that consumers are subject to time-varying financial or cash constraints that likely are related to when they receive wages or benefits.

Regarding longer-run trends, the authors estimate that the overall cash fraction declined by 2.46 percentage points per year in their three-year sample period, largely replaced by debit. They argue that the decline in cash at this particular retailer was likely not driven by transitory factors, and if the decline were to continue, only a relatively small fraction could be explained by forecasted changes in the zip-code-level variables, including age-cohort composition. This leaves a large fraction of the time trend to be explained, with prime candidates being technological progress in debit and changing consumer perceptions of debit relative to cash.

<http://dx.doi.org/10.1016/j.jmoneco.2016.10.005>

---

## Ad Valorem Platform Fees, Indirect Taxes, and Efficient Price Discrimination

By Zhu Wang and Julian Wright

RAND Journal of Economics, Summer 2017, vol. 48, no. 2, pp. 467–484

Platforms that facilitate transactions between buyers and sellers typically charge sellers ad valorem fees. In other words, they capture a percentage of the price set by the seller. Well-known examples include online marketplaces, such as Amazon and eBay, and payment card networks, such as Visa and MasterCard. Since these platforms do not incur significant costs that vary with transaction prices, sellers and regulators have questioned their use of ad valorem fees. Ad valorem sales taxes are also widely used in practice, with their desirability vis-à-vis per-unit taxes the subject of debates.

In an article in the *RAND Journal of Economics*, Zhu Wang of the Richmond Fed and Julian Wright of the National University of Singapore offer a new rationale for the use of ad valorem fees and taxes. The key idea of their theory is that when a market involves many different goods that vary widely in their costs and values that may not be directly observable, ad valorem fees and taxes represent an efficient form of price discrimination because the value of a transaction is plausibly proportional to the cost of the good traded. In such a setting, per-unit fees and taxes distort the price elasticity of demand across goods, as they add proportionally more to the price of a low-cost, low-value item compared with a high-cost, high-value item, thus reducing the efficiency of revenue extraction.

Ad valorem fees and taxes do not lead to such distortions, so they can ensure optimal Ramsey pricing. The authors show that in their model if there are no variable costs of enabling trades or collecting taxes, charging ad valorem fees or taxes allows the platform, regulator, or tax authority to achieve the same level of profit or welfare that could be obtained under third-degree price discrimination. It's as if the relevant authority could perfectly observe the cost and valuation for each good traded and set a different optimal fee for each transaction.

Wang and Wright then extend their theory to accommodate the fact that many platforms charge sellers a small fixed fee for each transaction in addition to the main ad valorem fee. To make this extension, the authors allow for the fact that platforms typically incur a small marginal cost per

transaction. They show that an affine fee schedule (a fixed fee per transaction plus a fee proportional to the transaction price) is optimal if and only if demand for the sellers belongs to the class with constant curvature of inverse demand (which includes linear demand, constant-elasticity demand, and exponential demand).

This result allows the authors to explore policy questions surrounding the use of ad valorem fees by platforms. The issue as it pertains to the payment card networks is currently being debated in many countries, including Australia, Canada, and the United States. The authors address this policy question by considering a regulated setting in which the regulator seeks to maximize social welfare while also allowing platforms to recover their costs, including their fixed costs. In contrast to policymakers' concerns regarding the use of ad valorem fees, Wang and Wright show that for the class of demand functions that rationalizes a platform's use of affine fee schedules, welfare in this constrained case is always higher when a proportional fee is used—in addition to a fixed per-transaction fee—to recover costs. This suggests that a policymaker who wishes to regulate platform fees to cover costs should also use ad valorem fees.

<http://dx.doi.org/10.1111/1756-2171.12183>

---

## Unauthorized Immigration and Fiscal Competition

By Subhayu Bandyopadhyay and Santiago M. Pinto  
*European Economic Review, February 2017, vol. 92, pp. 283–305*

Enforcement of immigration policy has traditionally been the exclusive responsibility of federal or supranational governments. In the United States, a few states have promoted state-level initiatives to deter unauthorized immigrants. In a paper published in the *European Economic Review*, Subhayu Bandyopadhyay of the St. Louis Fed and Santiago Pinto of the Richmond Fed examine the economic impacts of these initiatives using a spatial model of immigration policy enforcement.

**The authors consider this theoretical framework under four institutional arrangements, ranging from a completely centralized case to a completely decentralized case.**

Their model consists of two countries—a “source” and a “host”—with legal restrictions on the movement of labor between the two. The host country can enforce these restrictions at the border, by preventing immigrants from entering the country, and internally, by monitoring and penalizing firms that employ unauthorized immigrants. In the first part of the paper, the host consists of two regions that share a border with the source country and independently provide regional public goods, such as schools and hospitals. Citizens of the source country face two decisions: whether to move illegally to the host country and, once there, whether to move between regions. In the model, unauthorized immigration has two effects on the host country. First, it reduces wages for labor and increases rents on capital, leading to redistribution of income among domestic owners of factors of production. Second, it increases costs for deportation and for providing regional public goods, leading to higher taxes for domestic citizens.

The authors consider this theoretical framework under four institutional arrangements, ranging from a completely centralized case to a completely decentralized case. In decentralized environments, the authors observe several externalities between the regions. Provision of regional goods creates a positive externality by drawing unauthorized immigrants away from the other region. Border enforcement also has a positive external effect by reducing the total number of unauthorized immigrants. Internal enforcement has a negative external effect, as it diverts unauthorized immigrants to the other region.

When unauthorized immigrants can move freely within the host country, the centralized case leads to lower levels of unauthorized immigration. Decentralized policymaking tends to underprovide border enforcement and regional goods while overproviding internal enforcement.

As the cost of moving between host regions rises, decentralized internal enforcement declines while border enforcement increases. When moving between regions is impossible, internal enforcement becomes redundant, and decentralization could actually lead to greater provision of regional goods than in the centralized case.

Bandyopadhyay and Pinto go on to examine several extensions of their model. First, they assume that governments are able to restrict the access of unauthorized immigrants to regional goods. This causes internal enforcement to become relatively more effective in the decentralized case because regional governments can increase their provision of regional goods without attracting unauthorized immigrants. Second, they assume that the supply of unauthorized immigrants is endogenous—that wages in the source country adjust upward as laborers leave. Third, they consider a host country with both “border” and “interior” regions, a situation that more closely resembles the European Union. The main conclusions of the basic model hold up under each of these extensions, though each additional assumption offers new insight and makes the model more realistic.

The authors conclude by suggesting other extensions to their model: different mobility costs for workers of different skill levels, international mobility of capital, and firms’ uncertainty as to the immigration status of the people they are hiring.

<http://dx.doi.org/10.1016/j.eurocorev.2016.12.010>

---

## Innovation, Deregulation, and the Life Cycle of a Financial Service Industry

By Fumiko Hayashi, Bin Grace Li, and Zhu Wang  
Review of Economic Dynamics, forthcoming

As an industry evolves from birth to maturity, typically the number of firms in the industry initially rises and later falls. Quite often, only a small number of firms survive. Most theories that attempt to explain this “shakeout” have focused on the role of technological innovation in manufacturing industries. Only a few studies have looked at service industries, some of which also have experienced shakeouts due to either technological innovation or government deregulation.

In an article forthcoming in the *Review of Economic Dynamics*, Fumiko Hayashi of the Kansas City Fed, Bin Grace Li of the International Monetary Fund, and Zhu Wang of the Richmond Fed study the life cycle of the U.S. automated teller machine (ATM) and debit card industry, where both technological innovation and deregulation have played important roles. Specifically, they study networks that deploy ATMs and provide ATM services to cardholders from multiple financial institutions. These networks started to emerge in the early 1970s, and their numbers increased rapidly to a peak in the mid-1980s but then declined sharply in spite of continuing growth in ATM transaction volumes.

The authors identify two major shocks at the outset of this shakeout. One was a product innovation, the debut of debit cards in the mid-1980s. The debit card innovation allowed bank customers to use their ATM cards to make retail purchases in addition to making cash withdrawals, a new synergy that greatly increased the optimal size of ATM networks. Another major shock was banking deregulation, which started taking effect around the same time. The U.S. banking industry had maintained an almost constant number of commercial banks until the mid-1980s. Then the number of banks started to decline as deregulation relaxed size constraints on banks, and the resulting larger banks facilitated larger ATM networks. In addition, as a part of the banking deregulation, the U.S. Supreme Court ruled in 1985 to uphold a lower court’s decision that national banks’ use of shared ATM networks did not violate federal branching restrictions.

To explain how these two major shocks interacted with each other and drove the shakeout of ATM networks, Hayashi, Li, and Wang construct a dynamic equilibrium model of the industry's evolution. Their model distinguishes a major product innovation (the debit card) from subsequent process innovations. More importantly, the authors incorporate deregulation in the model—both general banking deregulation (the advent of interstate banking) and ATM-specific deregulation (the Supreme Court ruling). The authors show that their model aligns well with the rich data they collected on network entries, exits, sizes, and product offerings. The model also allows them to conduct counterfactual analyses to evaluate the roles that innovation and deregulation each played in the industry's evolution.

On one hand, the authors find that debit innovation was the main reason for the decline in the number of networks. On the other hand, while deregulation explained little of that decline, they find that it generated substantial welfare gains through cost and price reductions. They further find that the welfare effects of deregulation and innovation were magnified by how they interacted with each other. The authors' finding that deregulation could have a substantial welfare impact is consistent with observations from studies of some other service industries. Moreover, with a structural model, the authors are able to disentangle the effects of deregulation and innovation, which is a major step forward from those earlier studies.

<http://dx.doi.org/10.1016/j.red.2017.04.001>

---

## Market-Based Incentives

By Borys Grochulski and Yuzhe Zhang

*International Economic Review*, May 2017, vol. 58, no. 2, pp. 331–382

Misaligned incentives can induce wasteful individual behavior and lead to disastrous aggregate outcomes, so it's important to understand how incentives work. Economists have identified two main types of incentives: those in contracts between counterparties of an economic relationship and those that come from outside the relationship. These two types are almost always studied separately, but in reality both types are present in many economic relationships.

In an *International Economic Review* article, Borys Grochulski of the Richmond Fed and Yuzhe Zhang of Texas A&M University build a model that shows how external, market-based incentives can change the structure of direct, performance-based incentives in fully flexible long-term contracts. At equilibrium in their model, stronger job performance increases a worker's market value—the value she could obtain from a new contract with a different employer. When this market value becomes sufficiently high, the threat of her quitting forces the employer to increase her compensation. The prospect of obtaining this raise gives the worker a reason—a market-based incentive—to exert greater effort, thereby reducing the need for direct, performance-based incentives.

**The prospect of obtaining a market-based pay raise reduces the need for direct, performance-based incentives.**

The market-based incentive is strongest at the worker's quitting boundary, where the market value of her productivity is equal to the continuation value she receives from the contract with her current employer. When the worker is nowhere near the quitting boundary, the long-term contract helps insure against her idiosyncratic productivity shocks. But as her performance improves, the continuation value rises more slowly than the outside market value, moving her closer to the quitting boundary. When she reaches the boundary, the continuation value of the long-term contract no longer can rise more slowly than the outside market value. If it did, she would quit. Since performance is the sum of effort and productivity shocks, the continuation value at this point is highly sensitive to the worker's level of effort, which means her incentive to exert greater effort is



maximized. At the boundary, therefore, the contract no longer insures the worker—her incentive becomes entirely market based.

Connecting pay to performance still would be necessary if a worker's performance is persistently weak, but when Grochulski and Zhang relax the assumption of full commitment on the part of the firm and allow it to fire workers, performance pay becomes completely unnecessary. In other words, when firms can fire workers without incurring substantial costs, market-based incentives are stronger because workers are motivated not only by the prospect of pay raises, but also by the risk of losing their jobs.

The authors' analysis suggests that market-based incentives exist in principal-agent relationships beyond the particular setting of their model as long as the agent's efforts (or other actions benefiting the principal) improve the agent's standing in the market outside the present principal-agent relationship. For this reason, they expect that market-based incentives play an important role in many firm-employee and, perhaps particularly so, firm-executive relationships. Market-based incentives also may be important in lender-borrower relationships, where the borrower's outside option (for example, refinancing terms) can depend on the performance of her outstanding debt.

<http://dx.doi.org/10.1111/iere.12220>

---

## Does Redistribution Increase Output? The Centrality of Labor Supply

By Kartik Athreya, Andrew Owens, and Felipe Schwartzman  
*Quantitative Economics, forthcoming*

In an article forthcoming in *Quantitative Economics*, Richmond Fed researchers Kartik Athreya, Andrew Owens, and Felipe Schwartzman take a step, both analytically and quantitatively, toward evaluating the short-run effects of wealth redistribution on consumption and output.

The recession of 2007–09 and the slow recovery that followed have drawn renewed attention to the possibility that transferring wealth from rich households to poor households can stimulate output. The conventional argument that such transfers have a stimulative effect centers on the notion that poor households have a relatively high marginal propensity to consume out of wealth. Under standard “Keynesian” intuition, where individual components of aggregate expenditure decisions can be an important determinant of output, the resulting boom in consumption leads to an increase in output. Such a view puts household heterogeneity front and center in determining the aggregate short-run response to a change in wealth-transfer policy. In particular, it implies that any quantitatively plausible evaluation of the aggregate impact requires a model that accurately captures observed wealth heterogeneity.

The authors' qualitative analysis of such a model complements the conventional intuition by demonstrating how the direction and size of the output effects of wealth redistributions depend on labor supply decisions. On one hand, the authors show that the conventional intuition indeed holds for aggregate consumption. But on the other hand, they note that the same intertemporal considerations that induce this nonlinearity in consumption also push low-wealth households toward increasing their leisure by relatively more so that, barring countervailing forces, the same incentives that lead to a consumption boom also can induce a decrease in labor supply that leads to a reduction in output. In an important class of settings, the authors 1) show analytically that wealth redistributions are output-neutral unless they alter aggregate labor supply and 2) tease out conditions under which redistribution would lead to either a boom or a bust in output. They dem-

onstrate that this centrality of labor supply can hold even in the presence of aggregate demand externalities and sticky prices.

The authors' quantitative benchmark is a standard incomplete-markets model of consumption and labor supply that incorporates nominal rigidities and, in its quantitative version, accurately captures the U.S. wealth distribution. Within this model, they highlight the role of wealth effects on labor that are strong enough to largely negate the output effects of a consumption boom triggered by wealth redistribution. The authors' findings are robust to numerous modifications. In particular, they are virtually indistinguishable from those arising in a model that allows for nominal rigidities and a zero-lower-bound constraint on monetary policy. This combination has been invoked in other contexts as creating an environment in which fiscal stimulus can operate through aggregate labor demand. The authors' results therefore highlight that, as a quantitative matter, labor supply behavior remains dominant even in this leading case. While their results do not rule out the possibility of being overturned by an alternative calibration, they do suggest that under a reasonable set of assumptions, labor supply plays a central role in determining the stimulative impact of wealth transfers.

Athreya, Owens, and Schwartzman conclude that research aimed at measuring the impact of wealth redistribution on output would benefit substantially from further empirical research on how marginal propensities to work vary with wealth. They point to a narrow literature aimed at measuring the likely response of labor supply to plausibly exogenous changes in wealth, but they could find essentially no work tracking how this sensitivity varies with initial household wealth.

<http://qeconomics.org/ojs/forth/578/578-3.pdf>

---

## Preventing Bank Runs

By David Andolfatto, Ed Nosal, and Bruno Sultanum  
*Theoretical Economics, forthcoming*

**B**anking is the business of transforming long-maturity illiquid assets into short-maturity liquid liabilities, such as demand deposits. But economists and regulators view this maturity transformation as an inherently fragile financial structure that is susceptible to runs. When large numbers of depositors withdraw funds based on fear rather than real needs, banks cannot meet those obligations because a fire sale of their long-maturity assets will not generate sufficient funds to cover all of those liabilities. So the fear of a run can become a self-fulfilling prophecy.

**The idea is to resurrect Diamond and Dybvig's suspension scheme with suspension being triggered not by redemption activity but by "credible rumors of an impending run."**

A 1983 paper by Douglas Diamond of the University of Chicago and Philip Dybvig of Yale University is most often viewed as the seminal theory of bank runs, but what is often overlooked is their prescription for preventing bank runs. Diamond and Dybvig propose a suspension clause in short-term bank deposit contracts that would activate when redemption activity exceeds a certain threshold. This idea works well when depositors' liquidity needs are known in advance by the banks, but in the presence of uncertainty—when the fraction of depositors' liquidity needs is not known—the authors acknowledge that suspending redemptions is not optimal.

In an article forthcoming in *Theoretical Economics*, David Andolfatto of the St. Louis Fed, Ed Nosal of the Chicago Fed, and Bruno Sultanum of the Richmond Fed propose a solution to this problem. Any model with a bank-run equilibrium assumes that in equilibrium depositors know that a bank run is occurring. This assumption is necessary to generate the phenomenon of self-fulfilling prophecy. But if a model assumes that depositors know what is going on—in terms of their own collective beliefs—they should be able to communicate those beliefs to a mechanism. And if depositors



could be incentivized to report their beliefs, then the threat of suspension based on such information might prevent a bank run from occurring. The idea is to resurrect Diamond and Dybvig's suspension scheme with suspension being triggered not by redemption activity but by "credible rumors of an impending run."

Andolfatto, Nosal, and Sultanum depart from the direct mechanism approach, typically employed in the bank-run literature, and expand the set of messages that depositors can send to banks (in a direct mechanism, a depositor simply requests to withdraw or not). The authors' indirect mechanism expands this communication to permit the depositor to also express his belief that a run is occurring. This mechanism gives the depositor a higher payoff for sharing his belief than the payoff associated with concealing his belief and making an early withdraw. Upon receiving such a message, the mechanism fully suspends all further redemptions. The authors demonstrate that if commitment is possible, then under some weak parameter restrictions, their indirect mechanism uniquely implements an allocation that can be made arbitrarily close to the constrained efficient allocation as an equilibrium. In other words, their mechanism could prevent bank runs with approximately zero welfare cost to investors.

<https://econtheory.org/ojs/index.php/te/article/viewForthcomingFile/1970/16067/1>

---

## A Proposal to Clarify the Objectives and Strategy of Monetary Policy

By Robert L. Hetzel

*Journal of Macroeconomics, forthcoming*

The academic argument for a rule to discipline the formulation of monetary policy has generated enormous amounts of discussion over the years with no clear influence on the actual conduct of monetary policy. In 2012, the FOMC did adopt an inflation target but with no articulation of a strategy for achieving it.

In an article forthcoming in the *Journal of Macroeconomics*, Robert Hetzel of the Richmond Fed puts forth a proposal to bridge the gap between academic economists who want an arithmetic rule that would prescribe individual changes in the federal funds rate and policymakers who say they do not want and cannot follow such a rule. Hetzel accepts the latter view while maintaining that the FOMC still should clarify the objectives and strategy of monetary policy. Instead of an abrupt shift to an arithmetic rule, he builds on the FOMC's existing procedures and its stated desire for greater transparency.

Under existing procedures, FOMC participants make individual forecasts of the economy in the Summary of Economic Projections (SEP). These forecasts, which are based on unarticulated assumptions about "appropriate" monetary policy, are summarized statistically in various ways. In a separate report—the so-called "dot plot"—FOMC participants make unattributed forecasts of the federal funds rate target.

Hetzel wants to reorganize these procedures around three simple graphs that would reveal a consensus FOMC forecast that would replace the SEP and clarify the FOMC's strategy for achieving monetary policy objectives. The first graph would show two hypothetical paths for inflation—a price level target (currently 2 percent per year) and the FOMC's projected path for the actual price level. The second graph would show two hypothetical paths for real GDP—a benchmark path for the estimated level of potential real GDP and the FOMC's projected path for actual real GDP. The third graph would show two potential paths for interest rates—the federal funds rate projected by the futures market and the federal funds rate projected by the FOMC. The FOMC could use these

three graphs to communicate to markets the strategy for monetary policy that underlies its pursuit of the dual mandate.

In practice, the FOMC would routinely discuss monetary policy strategy to arrive at a consensus over the qualitative look of the price level graph and the real GDP graph—particularly the gaps between the lines in those figures. At least four times per year, the FOMC would be compelled to ask: What is the current rate of underlying inflation and how does that rate compare to the inflation target? What is the path of potential output—its level and growth rate? What then are the inflation gap and the output gap? Given these forecasts, how does the FOMC want to handle the relationship between the inflation gap and the output gap? Ultimately, the third graph would reveal the path for the federal funds rate that the FOMC deems consistent with its forecasts of inflation and output.

“The complexity of monetary policy and of the monetary standard renders accountability difficult,” Hetzel concludes. But “by providing a framework that makes explicit FOMC objectives and its forecasts, both of the economy and of the path of the funds rate, the proposal here would facilitate the communication with the public required for accountability.”

<https://doi.org/10.1016/j.jmacro.2017.03.004>

---

## Social Interactions and the Effectiveness of Urban Policies

By Santiago M. Pinto

Review of *Regional Studies*, Summer 2016, vol. 46, no. 2, pp. 117–126

In the presidential address before the Southern Regional Science Association, later published in the *Review of Regional Studies*, Santiago Pinto of the Richmond Fed argued that social interactions are crucial to consider when evaluating the effectiveness of urban policies. The social context may explain, for instance, why policies that share similar characteristics lead to different social and economic outcomes. Policy interventions may modify how people interact, creating unexpected social costs and endogenous social effects. These complications make predicting the effects of urban policies extremely challenging.

**Pinto cited an experimental U.S. program called Moving to Opportunity (MTO), which gave families in high-poverty neighborhoods housing vouchers to move to low-poverty neighborhoods.**

Pinto described two types of social interactions: neighborhood effects, occurring between individuals who live near each other, and network effects, which are not conditioned necessarily by physical proximity. The two concepts often interact and overlap in complex ways. Pinto indicated that newly available data could help disentangle them, which, though challenging, is important from a policy standpoint. Another complication is that models incorporating social network effects tend to have multiple equilibria. While it can be argued that models with multiple solutions may not provide new insights, these kinds of models are capable of explaining a variety of policy outcomes, such as those observed in reality.

Pinto cited an experimental U.S. program called Moving to Opportunity (MTO), which gave families in high-poverty neighborhoods housing vouchers to move to low-poverty neighborhoods. In the medium term, families that moved experienced no improvements to their labor market outcomes, welfare use, or academic performance. Adults and young females in families that moved enjoyed improved mental and physical health, while young males showed increased propensity for risky behaviors. Long-term effects on children depended on how old they were when their families moved: those younger than thirteen experienced strong positive socioeconomic benefits later in life, while older children were negatively impacted.

Pinto compared MTO to a similar experiment implemented in the Indian city of Ahmedabad. Residents of the city's slums were given the opportunity to relocate to better neighborhoods. Those who moved experienced no significant socioeconomic benefits, and about half returned to the slums within ten years. Additional evidence obtained through surveys suggests that moving made it harder for participants to stay connected with their families and friends. This program had effects different from MTO partly because it was implemented in a different social context, but both programs demonstrate that urban policies can have unexpected social costs and varied economic consequences.

Policy interventions also can cause the endogenous formation of networks. For example, a 2013 experiment split an incoming U.S. Air Force Academy freshman class into a group of middle-ability students, a mixed group of high- and low-ability students, and a randomly assigned control group. Even though pretreatment data suggested that low-ability students would benefit from spending time with high-ability students, low-ability students in the mixed group actually performed worse posttreatment. The reason was that during the experiment, students in the mixed group ended up segregating themselves by ability—low-ability students interacted mostly with other low-ability students. This study illustrates, among other things, how policies can create endogenous social dynamics leading to unanticipated consequences.

Pinto emphasized some other considerations to keep in mind when examining the effectiveness of urban policies. First, the quantity and quality of information may differ between types of social connections. Second, many studies claiming to find neighborhood effects are likely observing reverse causality. Researchers should keep in mind that endogenous sorting into neighborhoods can lead to selection bias.

<http://journal.srsa.org/ojs/index.php/RRS/article/view/46.2.1>

---

## Relationship Lending and the Great Depression: Measurement and New Implications

*By Jon Cohen, Kinda Cheryl Hachem, and Gary Richardson*  
NBER Working Paper No. 22891, December 2016

The Great Depression remains ground zero for studying the nonmonetary effects of financial crises. Despite abundant scholarship on the period, the lack of disaggregated data on lending activities has limited economists' ability to determine how much the collapse in long-term lending relationships affected the real economy. It stands to reason that failures of lenders who had accumulated substantial knowledge of their borrowers were more likely to disrupt the flow of funds than failures of lenders who were just starting to build relationships or those who only provided credit against easy-to-evaluate collateral. If that assumption is true, then the key to unlocking the nonmonetary effects of bank failures on real economic activity resides in the proper measurement of long-term lending relationships on the eve of the Great Depression.

In a National Bureau of Economic Research working paper, Jon Cohen of the University of Toronto, Kinda Cheryl Hachem of the University of Chicago, and Gary Richardson of the University of California, Irvine, and the Richmond Fed propose a novel way to extract cross-sectional differences in relationship lending from geographically aggregated financial statements. They unbundle the financial statements of banks, make a sharp distinction across jurisdictions between continuing-relationship loans and other types of credit activity, and, as a result, demonstrate that (1) the rise and fall of long-term lending relationships contributed significantly to the intensity and duration of the Great Depression and (2) the revival of those relationships varied across locales and played an important role in determining the pattern and pace of economic recovery.

**On net, the authors estimate that failures of all national banks could have accounted for one-eighth of the economic contraction observed during the Depression.**

First, the authors show that the marginal impact of bank suspensions on economic activity in the early 1930s was much higher in areas with more long-term lending relationships. They then show, through counterfactual analysis, that small bank failures on their own could have generated roughly one-third of the economic contraction observed during the Great Depression. Running the counterfactual analysis with both small and large banks, they find that distress at large banks actually played a mitigating role. In particular, the reallocation of deposits toward surviving relationship lenders appears to have generated economic gains. On net, Cohen, Hachem, and Richardson estimate

that failures of all national banks could have accounted for one-eighth of the economic contraction observed during the Depression. Second, the authors study the role of relationship lending in recovery. They show that cross-sectional differences in rebuilding these relationships are important for understanding cross-sectional differences in economic performance during the 1937–38 recession. This result suggests that the effects of the Depression persisted long after recovery had begun. The authors find that areas that rebuilt the long-term relationships observed in the 1920s fared

better during the 1937–38 recession than otherwise similar areas that failed to rebuild. Their findings seem to suggest that ongoing lending relationships played a significant role in recovery, much as they did during the contraction.

In sum, the authors contend that banking distress, if measured properly, has substantial negative consequences for real economic activity and that the effects of severe financial crises are likely to linger long after the events themselves have passed. “While it may appear that modern, highly sophisticated financial systems are immune to such stresses, the Great Recession and its aftermath would seem to provide a salutary corrective to such overconfidence,” they conclude.

<http://www.nber.org/papers/w22891>

---

## Rethinking Detroit

*By Raymond Owens III, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte*  
NBER Working Paper No. 23146, February 2017

The city of Detroit, an iconic example of urban development associated with modern industrialization, is in disarray. Large declines in population over several decades have resulted in an urban structure that severely impairs economic outcomes. Hundreds of thousands of people work in Detroit’s central business district, but the city’s core is surrounded by residential areas that have been substantially abandoned. This structure violates one of the most basic economic principles of urban design—that people prefer to live close to their workplaces to minimize commuting costs.

Detroit’s problems are not unique. New production technologies and trading possibilities have led to declines in many once-thriving cities. And as the comparative advantages of countries and regions continue to evolve, economists expect similar secular changes to continue. So dealing effectively with declining cities could generate more efficient use of resources and better spatial allocation of economic activity in cities throughout the world.

In a National Bureau of Economic Research working paper, Ray Owens and Pierre Sarte of the Richmond Fed and Esteban Rossi-Hansberg of Princeton University develop a quantitative spatial model to explain the plight of Detroit and evaluate various policy responses. The big question for Detroit is: Why don’t investors redevelop abandoned neighborhoods that are close to downtown workplaces? The authors argue that potential new residents simply do not want to be isolated in a redeveloped neighborhood surrounded by abandoned ones. Hence, neighborhoods have two

---

equilibria, one that features residential areas with enough residents to make them sustainable and one that features residential areas that remain mostly abandoned.

Previous studies of Detroit's ailments have acknowledged the problems that can arise from a lack of coordination among developers, residents, and local governments. To address this issue, Owens, Rossi-Hansberg, and Sarte add housing externalities to a trade-based general-equilibrium quantitative model of spatial urban activity. In this model, when residential neighborhoods with desirable economic fundamentals develop to a sufficient scale, residents' aversion to isolation is overcome. The problem then is how to coordinate developers to achieve the necessary scale when no single developer can afford to take the initiative. To solve the coordination problem, the authors propose using development guarantees by a government or other credible institution to attract developers. The authors use their quantitative spatial framework to estimate the magnitude of the guarantees that would be necessary to achieve coordination in equilibrium with a positive number of residents and residential investment. They note that if the policy succeeded, externalities from the new development would generate private demand for housing in these largely abandoned areas. In that case, the guarantees would not be called upon, making them costless to the guarantor.

The authors collect data from a variety of sources that include bilateral census commuting data, local development and price data from assessors, local Detroit organizations focused on measuring urban blight, and Google Analytics data on actual commuting times and distances between census tracts. The resulting dataset is quite rich and allows them to quantify the model at a high level of detail. Based on this model, the authors conclude that newly designed policies, such as development guarantees, could yield hundreds of millions of dollars in additional land rents per year and could perform substantially better than other types of proposals examined by the authors.

<http://www.nber.org/papers/w23146>

---

## **Arthur Burns and Milton Friedman: Why Did the Master (Burns) and the Disciple (Friedman) Understand Inflation in a Diametrically Opposed Way?**

*By Robert L. Hetzel*

*In Milton Friedman: Contributions to Economics and Public Policy, edited by Robert A. Cord and J. Daniel Hammond, New York: Oxford University Press, 2016, pp. 292–312*

Milton Friedman considered Arthur Burns a mentor. These two eminent economists had summer homes near each other in Vermont. Burns reviewed drafts of every important paper that Friedman wrote. So why did they clash in the 1970s over inflation and the Federal Reserve's role in controlling it?

Robert Hetzel of the Richmond Fed poses that question in chapter 16 of *Milton Friedman: Contributions to Economics and Public Policy*. One common conjecture is that Burns understood that inflation was a monetary phenomenon, but he pursued expansionary policy in the face of rising inflation to promote a partisan political agenda. But over the years, Hetzel has argued that while Burns "often made tactical decisions based on political considerations," his policy positions were "a reflection of the Keynesian temper of the times."

Burns, who chaired the Federal Open Market Committee (FOMC) from February 1970 until March 1978, was not a Keynesian economist per se, but “his understanding of the world shared much in common with the Keynesian consensus,” Hetzel writes. Like the Keynesians, Burns believed that market power pervaded U.S. markets. According to this view, large corporations were using their market power to raise prices, and labor unions were using their market power to raise wages. This explanation fit the Keynesian consensus that inflation was a real phenomenon—not a monetary manifestation. Also, like the Keynesians, Burns relied on the past relationship between the unemployment rate and inflation to predict inflation. When nominal wages and prices continued to rise despite the 1970 recession, Burns did not discard this analytical apparatus even after the data clearly contradicted it.

**“The Great Inflation of the 1970s turned out to be one of the great experiments in monetary history, and, professionally, it was the high point of Friedman’s career.”**

Friedman, on the other hand, believed that in the absence of government restrictions on entry, markets would be competitive. As a result, he did not think that market power was driving inflation. According to this view, market power could affect relative prices, but it could not influence the overall price level. Instead, Friedman believed that inflation was a monetary phenomenon that hinged on how much money the Fed created.

Friedman didn’t just challenge Burns; he challenged Keynesian economics and its policy prescription of managing aggregate demand to achieve a low unemployment rate. And his challenge was not just theoretical. Friedman also provided a methodological critique based on his 1953 essay, “The Methodology of Positive Economics.” He viewed historical combinations of monetary policy and fiscal policy as semicontrolled experiments that could be used to identify which factors drive the business cycle and inflation. By advocating this approach, Friedman transformed economics “into a discipline capable of testing hypotheses,” Hetzel writes.

But Burns never did embrace Friedman’s methodology for testing hypotheses or his monetary view of inflation. So despite some similarities between the two economists, their differences were fundamental and long-standing. These differences led to their dramatic confrontation over how the Fed should respond to the high inflation and assumed high unemployment that emerged during Burns’s tenure as chairman of the FOMC. Ultimately, Friedman’s views were vindicated, Hetzel concludes. “The Great Inflation of the 1970s turned out to be one of the great experiments in monetary history, and, professionally, it was the high point of Friedman’s career.”

<http://dx.doi.org/10.1093/acprof:oso/9780198704324.003.0017>

---

## Regional Policy and Fiscal Competition

*By Santiago M. Pinto*

*In Regional Research Frontiers – Vol. 1: Innovations, Regional Growth and Migration, edited by Randall Jackson and Peter Schaeffer, Cham, Switzerland: Springer International, 2017, pp. 199–215*

The absence of centralized or coordinated regional policymaking may lead to games of “fiscal competition” in which regions compete to attract businesses by lowering taxes, initiating a “race to the bottom” that results in an inefficiently low provision of local public goods. In chapter 13 of *Regional Research Frontiers – Vol. 1*, Santiago Pinto of the Richmond Fed examines the difficulties of modeling and estimating the economic consequences of fiscal competition and identifies future research focus areas.

A 1956 model of fiscal competition from economist Charles Tiebout portrayed fiscal competition in a positive way. If households and resources are perfectly mobile, if policymaking is uncoordinated between regions, and if regional policies do not have external effects, then decentralized policymaking allows consumers of local public goods to “vote with their feet,” Tiebout reasoned. In



response, local governments would compete to provide services that most closely match consumers' preferences at the lowest cost. More recent models of fiscal competition, however, focus on some of the negative aspects of decentralization, such as the race to the bottom example.

Pinto writes that the race to the bottom is "one of the most compelling results in the traditional tax competition literature," but he also points out that "more recent work reaches different conclusions." By modifying some of the underlying assumptions of the basic model, results may change and competition could become welfare enhancing. The emergence of agglomeration economies, for example, might allow regions to sustain higher tax rates and lead to a "race to the top" of sorts. Another strand of the literature argues that policymaking in one region may affect other regions through informational externalities. This "yardstick competition" model suggests that voters use information about policies in other regions to evaluate the performance of their local government, putting pressure on policymakers to mimic behavior from other regions.

While it is clear that regional policies generate many different external effects, Pinto notes that no single theoretical framework can encompass all of these effects in a unified way or account for the rich policymaking environment in which fiscal competition occurs. Existing models tend to focus on one or only a few aspects of fiscal competition, making simplifying assumptions such as symmetric Nash equilibria, uniformly distributed ownership of capital, and/or a simplified policy space. Pinto emphasizes theory's lack of consensus and the ambiguity of predictions from more sophisticated models.

The empirical literature on fiscal competition is also inconclusive. Empirical work based on spatial econometric models has provided a sense of the degree to which regions are interdependent, but this work is faced with an "extremely challenging" identification problem and is unable to really quantify the effects or welfare implications of fiscal competition.

Pinto suggests that further advancement of the literature on fiscal competition will require methodological innovations. He focuses on newer developments in quantitative models that "combine calibration and structural estimation techniques with data," an increasingly popular approach in urban and regional economics. This approach has advantages and drawbacks. Structural methods help overcome the identification challenge, can quantify consequences for welfare, and are more useful for policy evaluation. But they also are harder to estimate empirically, may have less generalizable conclusions, and are data and computationally demanding. Ultimately, Pinto concludes that these methods should be considered complements to, rather than substitutes for, current research efforts.

[http://dx.doi.org/10.1007/978-3-319-50547-3\\_13](http://dx.doi.org/10.1007/978-3-319-50547-3_13)

---

This issue of *Richmond Fed Research Digest* was written and edited by Karl Rhodes, Lisa Kenney, and Kody Carmody. It was published by the Research Department at the Federal Reserve Bank of Richmond and may be photocopied or reprinted in its entirety. Each individual summary also may be photocopied or reprinted in its entirety. Please credit the Federal Reserve Bank of Richmond and include the statement below.



The *Richmond Fed Research Digest* summarizes articles and other works that reflect the views of their respective authors and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.