Research Review

Featured Paper

Wealth Shocks and Macroeconomic Dynamics
Daniel Cooper and Karen Dynan

FEDERAL RESERVE
BANK OF BOSTON™

Executive Summaries in This Issue

Federal Reserve Bank of Boston
Research Department Papers Series

Public Policy Discussion Papers present research bearing on policy issues. They are generally written for policymakers, informed business people, academics, and the informed public. Many of these papers present research intended for professional journals.

Working Papers present statistical or technical research. They are generally written for economists and others with strong technical backgrounds, and they are intended for publication in professional journals.

Public Policy Briefs present analysis on topics of current interest concerning the economy. These briefs are written by Boston Fed staff economists, based on briefing materials or presentations prepared by Boston Fed research staff for senior Bank executives or outside audiences.

Research Reports present research on economic and policy issues of concern to New England's state and local policy leaders and others engaged in developing and implementing public policy in the region. These reports are written by Boston Fed economists and policy analysts.

Research Data Reports present research that focuses primarily on methodologies and methods involved in sourcing, gathering, compiling, transforming, or managing data and datasets that shed light on economic topics.

Research department papers are available online.
http://www.bostonfed.org/economic/respubs.htm

p-13-1

When Does Delinquency Result in Neglect? Mortgage Distress and Property Maintenance

by Lauren Lambie-Hanson

abstract and complete text: http://www.bostonfed.org/economic/ppdp/2013/ppdp1301.htm
e-mail: lauren.lambie-hanson@phil.frb.org

Motivation for the Research

Mortgages are financial contracts between borrowers and lenders, but when a borrower loses a property through foreclosure, the process impacts parties external to the contract. If foreclosures result in vacancy, deferred maintenance, or vandalism and other crime, then tenants and neighboring owners may suffer. This paper examines the timing of one type of foreclosure externality, reduced property upkeep, which is measured using conditions reported by constituents in Boston, Massachusetts.

Most existing studies of foreclosure externalities use neighboring house prices as the metric for spillovers. While prices are easy to measure and may literally put a dollar value on foreclosure spillovers, these studies are typically unable to distinguish between whether foreclosures hurt neighbors' home values because of deferred maintenance and vacancy or because foreclosures add to the supply of low-priced properties on the market, pushing down prices. Moreover, these studies often find only negligible evidence of spillovers, perhaps because the valuation of a property, as a long-lived asset, may be based more on the expected future value than on the short-term use value of the home. Since neighboring foreclosures represent only a temporary nuisance, buyers may not adjust their willingness to pay for a home with distressed sales nearby, even though those properties may, at least in the short run, harm neighborhood quality of life. Finally, price spillover studies tell us little about how foreclosures impact neighboring owners who do not sell their properties.

The purpose of this paper is to fill these gaps in the existing literature by determining whether (and when) properties owned by delinquent borrowers and lenders become public nuisances in their neighborhoods.

Research Approach

Using a rich administrative dataset from Boston, the author captures information on when residents in a neighborhood report problems about particular properties to local government. She links this property-level dataset of constituent complaints and requests to three other datasets—a property-level dataset of sales transactions and mortgage originations, a loan-level dataset of mortgage performance for subprime and Alt-A mortgage borrowers, and real estate sale listings data from the area's multiple listing service. Using this four-part, master dataset, she estimates a set of multi-level longitudinal models to compare the incidence and timing of complaints, identifying when in the delinquency and foreclosure process a property becomes the subject of resident complaints. She also distinguishes between owners who attempt to sell their properties through short sales and those who do not try to sell short.

Key Findings

- There is no apparent relationship between property upkeep and short sale attempts. However, the level of property maintenance varies during different stages of the foreclosure process. Borrowers begin neglecting maintenance when they are seriously (90 days or more) delinquent, and property distress becomes more common once the owner has been in foreclosure for over a year.

- Properties are most likely to be the subject of constituent complaints when they are bank owned. This is particularly true for single-family properties, which are more than nine times as likely to be the subject of a constituent complaint when bank owned as before the borrowers become delinquent.

Implications

The findings suggest that distressed properties are most problematic when owned by banks, both before and during lenders' attempts to sell the properties. Lenders often work to bring properties up to code to enable sales to buyers who require FHA mortgage financing (Sinnock 2012), although perhaps greater bank accountability for properties is needed. Finding the parties responsible for the upkeep on bank-owned properties can be challenging, even when these properties have a designated real estate agent. Zillow, a self-described "home and real estate marketplace," recently began providing open access on its website to property records and valuation information for foreclosed properties that have not yet been listed—and in some cases, properties on which foreclosures have not even been completed. The introduction of this type of publicly accessible information may have the supplementary benefit of increasing public awareness about the ownership status of nearby properties and lessening banks' abilities to "hide in the shadows" while their properties become community nuisances.

The collateral harm caused by bank-owned properties suggests that more might still be done to hold banks accountable for property maintenance, including providing easier access to the contact information of property caretakers. Since mortgages terminated through short sales avoid bank ownership entirely, allowing short sales should impose less damage on the neighborhood quality of life. Finally, well-intentioned policy interventions that lengthen the foreclosure timeline while failing to prevent foreclosures may lead to longer periods in which foreclosure externalities are likely to plague neighborhoods.

In February 2008, the City of Boston passed a foreclosure registration ordinance, which requires that lenders holding foreclosed properties register them with the city each year and pay a $100 fee. The purpose of the ordinance is to help the city track contact information for the stewards of foreclosed properties, in case these properties become unsafe, unsecured, or poorly maintained. More city inspections and code enforcement in distressed neighborhoods may help, although according to the results in this paper, in order to be most effective, these efforts would need to begin before properties become bank owned—and so before they are registered under the ordinance, a daunting task. Having a large student population, Boston devotes a significant share of its inspectional services resources to routine inspections of rental housing following occupant turnover. This leaves limited resources for the city's inspectional services department to respond to foreclosure-related disinvestment in neighborhoods.

Longer periods in serious delinquency and foreclosure generate negative externalities for neighbors, as demonstrated by this study and by Ellen, Lacoe, and Sharygin (2012) and

The collateral harm caused by bank-owned properties suggests that more might be done to hold banks accountable for property maintenance.

Constituent Complaints by Type, June 2009–December 2011
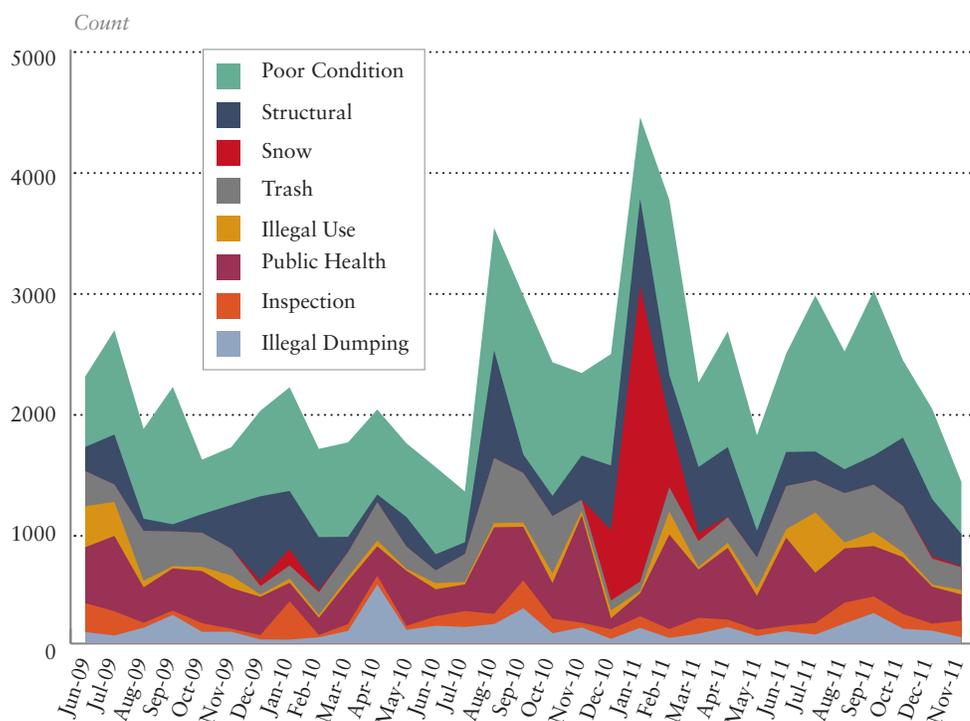
Source: City of Boston Constituent Response Management System, Inspectional Services Department. Cases tabulated by author.

Gerardi et al. (2012). All these studies offer evidence that properties are nearly twice as likely to be the subject of a constituent complaint once the owners are in foreclosure. Policymakers should consider this finding when designing well-intentioned policies that lengthen the foreclosure timeline. As discussed by Gerardi, Lambie-Hanson, and Willen (2013), judicial foreclosure proceedings and state-specific right-to-cure periods lengthen the average foreclosure timeline but do not improve the probability that borrowers self-cure their mortgage defaults or receive mortgage modifications. Policies that lengthen the foreclosure process extend the time properties are in ownership limbo, which could result in more problems from deferred maintenance.

Short sales, which are gaining steam nationally and are the most common form of "aid" lenders grant distressed borrowers (Berry 2012), are shown in this paper to result in shorter durations that properties spend in "ownership limbo" (owned by a bank or a borrower who is not making mortgage payments). Even though properties do not appear to receive better upkeep when owned by a borrower pursuing a short sale, the shorter duration spent in uncertain ownership should make properties sold through short sales less detrimental to their neighborhoods than foreclosures. Of course, short sales can pose problems of their own, particularly fraudulently low prices. A growing share of short sales has been followed by quick resales, at suspiciously high prices (CoreLogic 2011).

Contrary to expectations, the author's results also indicate that owners are not more susceptible to generating property complaints if they have less equity in the property. In order to verify the robustness of these results, a potential next step is to analyze code violations and building permit data from the City of Boston, and to devise ways of capturing more accurate measures of equity levels and borrowers' perceptions of their equity. Finally, this paper leaves open the question of how properties fare after being resold to third-party buyers. More information is

needed on this topic, particularly to improve our understanding of the role of investors and homeowners in purchasing foreclosed properties and stabilizing neighborhoods.

This paper is the first of its kind to use constituent requests for local government services as an indicator of foreclosure externalities. While other recent studies have found small or nonexistent spillovers of foreclosures on neighboring property values, this could reflect the fact that having a nearby property in foreclosure typically poses only a temporary threat to a neighborhood. In many cases, the types of complaints captured in the author's dataset reflect issues that certainly impact neighboring owners' and tenants' quality of life for a period of time, but may not have a material effect on the prices of housing (a long-lived asset). This could explain why Fisher, Lambie-Hanson, and Willen (2012) find no price spillover effects from single-family foreclosures, despite the fact that these properties are far more likely (over nine times as likely) to receive complaints while bank owned than before the borrowers defaulted.

About the Author
Lauren Lambie-Hanson is a senior specialist in the Risk Assessment, Data Analysis, and Research Group at the Federal Reserve Bank of Philadelphia. She wrote the paper while she was a research associate at the Federal Reserve Bank of Boston.

The Role of Proximity in Foreclosure Externalities: Evidence from Condominiums

by Lynn M. Fisher, Lauren Lambie-Hanson, and Paul S. Willen

abstract and complete text: http://www.bostonfed.org/economic/ppdp/2013/ppdp1302.htm
e-mail: lynn_fisher@unc.edu, lauren.lambie-hanson@phil.frb.org, paul.willen@bos.frb.org

Motivation for the Research
Starting with Immergluck and Smith (2006), researchers have documented that properties located near foreclosed properties sell at a discount relative to otherwise identical properties that have no foreclosures nearby. The authors extend this literature by focusing on a sample of Boston condominiums that allows them to identify the precise mechanism that generates these price effects. In particular, they aim to distinguish between two popular theories, the first being that foreclosures cause price declines through a "supply effect," resulting from the fact that a foreclosed property is a close substitute for nearby properties.

An alternative and not mutually exclusive explanation is that an owner has no incentive to invest in the property during the foreclosure process, and so the property deteriorates, generating a physical externality. The results have important implications for policy. If foreclosures affect prices merely by increasing the supply on the market, then the effect of foreclosures on nearby properties is a pecuniary externality, implying that the market outcome is not necessarily inefficient and that government intervention risks choosing winners and losers rather than increasing overall welfare. In contrast, a physical externality always allows for welfare-improving policy interventions.

Research Approach
In this paper the authors use regression analysis to examine these different explanations of the effect of foreclosures on neighboring properties using a dataset of condominium transactions

in Boston over the years 1987 to 2012, for which they have rich data on the size and location of condo associations. The principal data source for the analysis in this paper is a dataset of public records and assessors' files compiled for these Boston properties by the Warren Group, a New England firm that tracks real estate transactions.

The reason this dataset is particularly useful is that it includes information on the condo associations to which individual units belong, enabling the authors to distinguish between units within a foreclosed property's same association and those that are neighbors in other associations. Specifically, these data allow the authors to divide their sample of condo pairs into three groups: same-association, same-address (SASA); same-association, different-address; (SADA); and different-association, different-address (DADA) units.

To explain why these distinctions are useful, consider some alternative hypotheses. If foreclosures drive down prices because of the supply effect, we would expect this association to matter more than location and the effect of SASA and SADA foreclosures to be roughly equivalent, assuming that units within the same condo association are closer substitutes for one another than for units in neighboring associations. If the externality works through the association itself—for example, without the dues income, the association may have trouble maintaining the common spaces—we would also expect to see SASA and SADA having similar-sized effects. But if the externality is related purely to the physical condition of the distressed property, we would expect the effects of SASA foreclosures to matter the most, and we might expect SADA foreclosures to be comparable to DADA foreclosures in their effects of exerting downward pressure on house prices.

The authors pay special attention to the fact that the owner of a condo in mortgage foreclosure has little incentive to make association payments. Failure to pay these fees will result in the association's draining its reserves or deferring maintenance while attempting to recover the fees, either scenario potentially making the building and association less desirable to prospective buyers. High vacancy rates, nonowner occupancy, and unpaid condo association fees can trigger the loss of a property's eligibility for Federal Housing Administration (FHA) financing or securitization with Fannie Mae or Freddie Mac, potentially making it difficult for an owner to sell to a buyer who needs to use mortgage financing.

A major concern about regressing prices on foreclosures is that because falling house prices reduce borrowers' equity and lead to foreclosures, the estimated effects could unwittingly reflect the impacts of prices on foreclosures rather than the damaging effect of foreclosures on the sales prices of nearby properties. Following Gerardi et al. (2012) and others, the authors address this problem by adopting a repeat sales methodology and using census tract controls for neighborhoods and comparability controls for property characteristics, meaning that the authors' estimation strategy amounts to comparing two observably identical properties in the same census tract that were bought in the same year and sold in the same year and that differ only in the number of foreclosures nearby. Since a census tract is small—typically containing about 4,000 inhabitants—the authors can rule out explanations for any estimated effects that rely on differences across neighborhoods or markets. To offer an alternative explanation for why one observes a price discount near a foreclosed property, one must explain why properties in one part of a census tract appreciate at different rates than properties in another part of the tract. Given the small size of tracts, this is usually challenging. For example, buyers shopping for a house will typically not restrict their search to just one part of a census tract.

Key Findings

- The effect of SASA foreclosures is much stronger than the effect of either a SADA foreclosure or a DADA foreclosure, neither of which has a statistically or economically significant impact on neighboring house prices. All else being equal, an additional SASA foreclosure reduces the sale price of a nearby condominium by an average of 2.4 percent, whereas an additional SADA foreclosure reduces the price by 0.5 percent, and an additional DADA foreclosure reduces the price by 0.2 percent, with only the SASA effect being statistically significant. Almost all the SADA foreclosures took place in associations with more than 12 units, and the results hold even when the authors focus only on large associations.

- The effects of SASA foreclosure are much stronger in small condo associations than in large associations. The estimates show that an additional foreclosure in an association with 12 or fewer units lowers the price by 6.1 percent. The effect of SADA foreclosure in small associations is not statistically significant, because such foreclosures are rare, and so the results of those regressions lack statistical power.

Implications

The authors view their results as evidence that the main source of the effects of foreclosures on the prices of neighboring properties is the physical externality. They adopt this view because they reason that one would expect SADA properties to be very close substitutes, so if the supply effect were powerful, one would expect these foreclosures to depress prices—yet the effect they measure is neither economically nor statistically significant. In small associations, however, where same-association foreclosures are usually located at the same address, the physical externality cannot easily be disentangled from the association effects, so the authors believe that both effects may be depressing house prices.

About the Authors

Lynn M. Fisher is associate professor of real estate and the David D. and Carol Ann Flanagan Scholar at the University of North Carolina Kenan-Flagler Business School, Lauren Lambie-Hanson is a senior specialist in the Risk Assessment, Data Analysis, and Research Group at the Federal Reserve Bank of Philadelphia, and Paul S. Willen is a senior economist and policy advisor in the research department at the Federal Reserve Bank of Boston.

p-13-3

Economic Distress and Resurgence in U.S. Central Cities: Concepts, Causes, and Policy Levers

by Yolanda K. Kodrzycki and Ana Patricia Muñoz

Motivation for the Research

This paper provides a review of the literature on U.S. central city growth and distress during the second half of the twentieth century with the aim of gaining a better understanding of what factors can contribute to the resurgence of a distressed city.

Research Approach

The authors begin by reviewing the methodologies, findings, and implications of the literature on central city growth. They then consider the literature on distressed cities. Finally, the authors compare two different methods of identifying distressed and resurgent cities. One approach is to make comparisons across a comprehensive national cross-section of cities, while the other is to focus on a more limited sample of peer cities that share similar characteristics, challenges, or opportunities.

Key Findings

- Based on the literature on central city growth, metropolitan areas with favorable weather, higher growth, and greater human capital tended to experience higher growth, while distress was strongly correlated with a legacy of city-level manufacturing.

- Of all the forces that influenced population, employment, and economic well-being, the declines in manufacturing employment are arguably the most relevant for explaining city distress. Among all U.S. central cities, those that relied on a heavily industrial base experienced the greatest negative shocks in the second half of the twentieth century. These cities are also the most likely to have weak economies and low family incomes today. Distressed cities suffer from an erosion of physical and social capital, as well as deterioration in their civic infrastructure. Moreover, they have comparatively high shares of high school dropouts and low shares of residents with college degrees, making them unattractive locations for employers.

Some cities have managed to achieve resurgence through strong leadership, collaboration across sectors and institutions, clear and broad-based strategies, and significant infrastructure investments.

- Distress has been highly persistent, but some cities have managed to achieve resurgence through a combination of strong leadership, collaboration across sectors and institutions, clear and broad-based strategies, and significant infrastructure investments. In cities that have successfully revived their fortunes, public officials, private-sector employers, and nonprofit institutions have coalesced around a long-term vision and have collaborated for a sustained period of time in implementing broad-based revitalization strategies. Such strategies include attracting and retaining competitive businesses across a variety of industry sectors, fostering innovation and knowledge transfer, and improving both human and physical capital.

- The two methods for identifying distressed cities yield similar lists of economically troubled cities. They tend to disagree only in the few cases where cities have made uneven economic progress over time, or where their performance varies substantially across indicators. Further examination reveals that relatively small changes in time period or criteria would have resulted in a consistent categorization in at least one-half of these cases. It appears that reporting on alternative criteria and sample periods is particularly valuable for studying cities that place somewhere in the middle of the continuum of most distressed to least.

- Regardless of the sample and methodology used, most distressed older industrial cities in the United States have yet to revitalize their broader economic and civic underpinnings.

Implications

This review of the literature on distressed and resurgent cities reveals continued gaps in knowledge. Somewhat to the authors' surprise, researchers' understanding of the structural relationship between human capital and central city economic conditions remains far from

complete. Further work is needed to illuminate the contributions of city and metropolitan area educational institutions and of both intraregional and interregional migration to determining the distribution of skills within and across cities. The differing types of shocks to manufacturing employment and the varying economic consequences of apparently similar shocks across cities are additional topics that deserve further research. Finally, policymakers at the city, regional, state, and national levels would benefit from more analysis of the effectiveness of specific policy tools to improve the economics of distressed cities.

About the Authors

Yolanda K. Kodrzycki *is a vice president and the director of the New England Public Policy Center in the research department of the Federal Reserve Bank of Boston. The Policy Center conducts research on key economic and policy issues in New England and engages with regional partners in advancing identified policy options.* **Ana Patricia Muñoz** *is a senior policy analyst in the regional and community outreach department of the Federal Reserve Bank of Boston.*

Wealth Shocks and Macroeconomic Dynamics
by Daniel Cooper and Karen Dynan

Motivation for the Research

Economists have long been intrigued by how fluctuations in household wealth affect consumer spending, which in the United States accounts for approximately 70 percent of GDP. Since the mid-1990s the U.S. economy has experienced two major stock market booms and busts, as well as a dramatic rise and fall in house prices that precipitated a financial crisis and a very severe recession. Many observers believe that fluctuations in household wealth influence real economic activity. For example, by early 2009, the value of stocks held by U.S. households had plunged about 50 percent from peaks attained just a few years before, while the value of real estate owned by U.S. households had fallen roughly 25 percent, again within a two-to-three-year period. This sharp drop in household wealth is often cited as an important contributing factor to the slow recovery from the Great Recession.

Gaining a better understanding of wealth effects, meaning the impact that changes in household wealth may have on consumption and, in turn, on the overall macroeconomy, has become particularly important for policymakers. Over the last two decades a great deal of empirical research has focused on this relationship using many different data sources, such as macroeconomic time series, regional data, household survey results, and credit records. Yet significant questions about wealth effects remain unanswered.

Research Approach

The authors first review where the literature currently stands. In brief, standard macroeconomic theory offers the permanent income hypothesis (PIH) as a framework for thinking about wealth effects. The PIH posits that households consume a constant fraction of the present discounted value of their lifetime resources, and contends that changes in wealth that permanently alter a household's resources should therefore cause its consumption

patterns to change. While macroeconomic models are useful for depicting the average historical relationship between aggregate consumption and wealth, these models are limited because they cannot account for household heterogeneity—an important consideration regarding household wealth.

To address these limitations, much of the existing wealth effects literature has utilized microeconomic data. One source of potential heterogeneity that the literature focuses on is credit constraints, since different types of households may respond differently to changes in housing wealth. Since younger households and lower-income households are more likely to be credit constrained, these households are more apt to increase consumption in response to wealth gains, such as a rise in house prices. The spatial dispersion of wealth shocks can also vary over time, representing another possible source of heterogeneity in wealth effects. Within the United States over the last decade, the areas with the largest house price gains—such as Las Vegas—experienced the largest house price declines during the housing collapse. The various states also have different demographic characteristics. Another source of heterogeneity is the fact that the distribution of changes in aggregate wealth across households varies over time, due to differences in the composition of assets held. Moreover, the fraction of people having certain characteristics, such as being credit constrained, can shift over time. For instance, a household may be credit constrained when younger, but may escape this condition in later adulthood through gains in income and wealth.

In sum, there is a need to learn more about the underpinnings of wealth effects at the household level. The authors focus their attention on four key issues regarding wealth effects and macroeconomic dynamics that likely matter to policymakers. Three of these topics center on how housing wealth, stock market wealth, and household debt may influence consumer spending. The fourth topic considers why the wealth effect, at both the aggregate and household levels, may change over time.

The recent boom and bust in U.S. residential real estate prices has spurred much interest in how wealth effects from housing influence the macroeconomy, particularly whether gains in housing wealth have the same effect as gains in financial wealth on household spending decisions. More U.S. households own homes than own stocks and bonds. Housing is both an asset and a consumption good that provides shelter services, so to fully realize gains, homeowners have to sell this asset, something that in the short-to-medium term many will not be keen to do. However, housing wealth may have an indirect impact on consumption, for when home prices rise, homeowners can borrow more against their home equity and obtain funds to spend on goods and services, such as home improvements, college costs, and automobiles. Homeowners who previously were credit constrained are likely to increase their consumption spending when home prices rise. The share of housing wealth as a portion of total household wealth is much higher for lower-income homeowners than for higher-income homeowners, so gains in housing wealth among households in the lower parts of the income distribution, which tend to be more constrained, might have a greater impact on their marginal propensity to consume. Therefore, an increase in housing wealth has a greater effect on consumer spending than does a similar change in financial wealth. Yet the empirical evidence is mixed regarding the relative impact on consumption that accrues from housing wealth effects and financial wealth effects.

Typically, financial wealth is thought to affect consumption directly, since it tends to be held in fairly liquid forms like stocks and bonds. Still, some researchers argue that the positive

relationship between aggregate consumption and aggregate stock prices does not stem from a direct wealth effect. Rather, rising or falling stock prices could act as a signaling channel, whereby stock prices are seen as a proxy for household expectations about future wage growth. If stock prices are rising, consumers may increase their consumption based on optimistic expectations, while they may curtail consumption if stock prices fall. Older findings suggest that stockholders' consumption reacts more strongly than nonstockholders' consumption to stock price increases. Yet over the last decade this question has been under-researched, perhaps because of the greater attention paid to housing wealth effects. Dynan (2010) finds preliminary evidence that incorporating more recent household-level data in analyses of the differential behavior of stockholders versus nonstockholders weakens the earlier evidence considerably. Moreover, behavioral work by Choi, Laibson, and Metrick (2009) finds that individuals tend to raise their retirement plan contributions after experiencing good stock market returns—if these higher contributions were funded by reduced consumption, this would result in a wealth effect that is opposite to the standard positive relationship between higher consumption and higher stock market returns.

While household consumption is influenced by wealth effects stemming from the movements of asset prices, such as residential housing and stocks, households also may make spending decisions based on the amount of their outstanding debt. In the wake of the Great Recession, the sluggish growth of the U.S. economy has raised the issue of whether high levels of household mortgage debt and leverage resulting from the housing boom and bust have played a substantial role in dampening consumption and hence the recovery. While traditional economic models suggest that debt does not exert an independent effect on consumption beyond its indirect effect through household net worth, a case can be made that household debt does impede consumer spending to a certain extent. Some households might target holding a certain level of debt relative to income or assets, and cut back on spending to pay down debt. High-debt households might be concerned about future credit access, and cut their consumption in order to increase their savings. In separate studies using household-level data, Dynan (2012) and Cooper (2012) both find that levels of high household debt had a negative impact on consumption growth during the Great Recession—even after controlling for income, net worth, and other factors likely to affect spending. Household debt may have had a larger impact on consumption recently since more U.S. households are burdened by high debt and leverage than in earlier periods. Mian, Rao, and Sufi (2013) use regional data and conclude that leverage helped to amplify the negative wealth effect on consumption associated with falling house prices during the Great Recession. Still, Cooper (2012) shows, using aggregate data, that consumer spending has not behaved unusually in the aftermath of the Great Recession, given movements in income and net worth.

As suggested above, if household spending responds differently to changes in wealth according to the asset class affected, then the strength of the aggregate wealth effect should differ depending on the source of a given movement in aggregate household wealth. Yet there are other reasons why the wealth effect, at both the aggregate and individual household levels, might change over time. With shorter remaining lifespans, older households experiencing wealth shocks might consume more than younger households. With the aging of the baby boom generation, it is possible that the average marginal propensity to consume out of wealth across all households will rise. Starting in the early 1980s, institutional developments, including regulatory and tax codes changes, have reduced credit constraints, thereby increasing the availability of borrowing and lowering its cost. But the effect of these changes on the size of the wealth effect is unclear. Previous empirical work suggests that credit constraints tend to

Changes in Net Worth and Consumption

4-Quarter Percentage Change   *4-Quarter Percentage Change*

Net Worth
Real Consumption
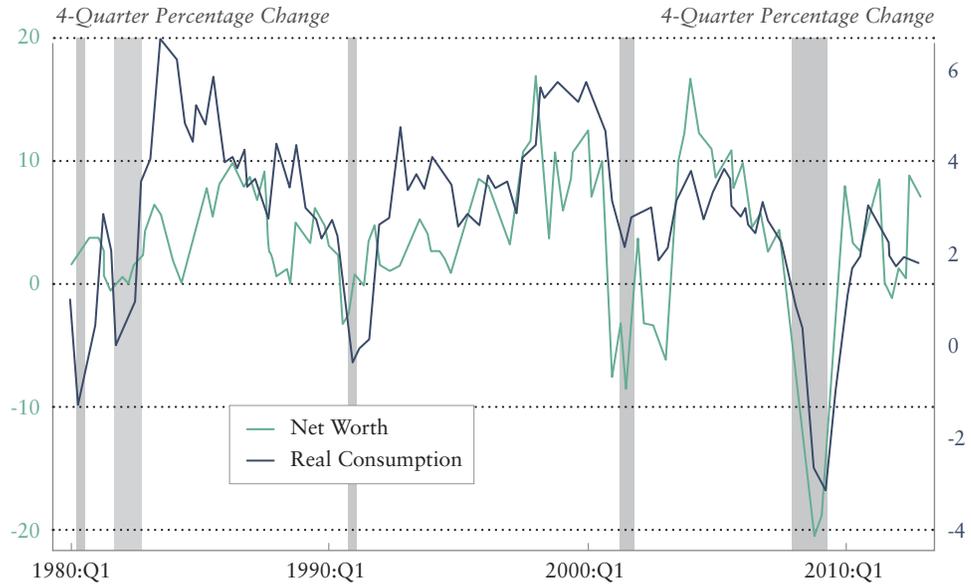
Source: Federal Reserve Board, BEA/Haver Analytics.
Note: Shaded areas indicate recession.

Net Worth to Income Ratio

Ratio
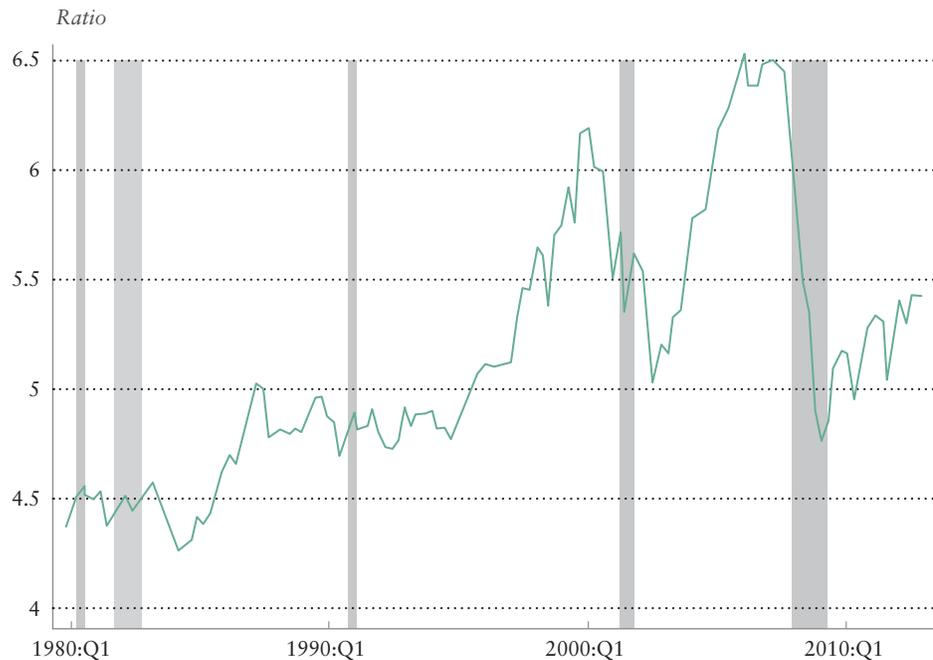
Source: Federal Reserve Board, BEA/Haver Analytics.
Note: Shaded areas indicate recession.

be associated with a stronger wealth effect. However, fairly recent financial innovation has made it easier and less costly to capture housing wealth gains through home equity loans and cash-out refinancing options; these developments may have increased the aggregate wealth effect. The growth of stock mutual funds and 401(k) accounts has enabled more households, especially lower-income households, to own stocks. Since lower-income households have a higher marginal propensity to consume based on positive wealth shocks, this development may also have raised the aggregate wealth effect.

Key Findings

- More empirical work is needed to improve our understanding of how gains in housing wealth affect the macroeconomy. Since current consumption models typically include just total household wealth, rather than breaking household wealth down according to the various asset classes that contribute to this total, the conventional coefficient on wealth simply captures the average experience over time. If households have different propensities to consume based on the type of wealth gain or loss, such as housing versus stocks, taking account of these differential movements should result in more accurate macroeconomic forecasts.

- Research on wealth effects based on stock prices has lagged over the last decade, despite dramatic swings in stock prices since the late 1990s. These recent episodes mean that more data are available to enrich our understanding of the underpinnings of the stock price wealth effect. Preliminary evidence indicates that the earlier findings establishing a strong positive relationship between aggregate consumption and aggregate stock prices might be weakened if more recent data are taken into account.

- The household-level empirical research on how debt relates to consumption is limited. At best, it considers only the period through the Great Recession and does not speak directly to the U.S. economy's weak performance during the recovery. A problem in many of these studies is that the standard errors are very large. Another drawback is that the emphasis has been on establishing the relationship rather than on discerning why a relationship exists, an issue that is highly relevant to the policy discussion pertaining to what, if anything, should be done to address the situation. All of these shortcomings suggest promising theoretical and empirical avenues for further research on the complex connection between macroeconomic activity and household leverage.

- Many factors may have changed the aggregate wealth effect over time, but because of small sample sizes, it is difficult to assess the direction in which it may have changed, using only aggregate data. Using household-level or regional data, where the price variation is much richer, should yield a better understanding of how the wealth effect has functioned in more recent periods.

- While macroeconomic datasets have limitations that prevent establishing the empirical relationship between wealth and consumption, household-level datasets also have many shortcomings. Many household-level datasets lack all the elements needed to estimate consumption functions—a panel dimension, complete balance sheet information, broad measures of consumption, good income measures, and demographic information, which can proxy for preferences, risk of job loss, and credit access, among other factors. Administrative records, such as those from credit bureaus or financial services companies, offer more detailed and accurate information. Recent research using regional data on consumption, income, and wealth, derived by aggregating data from records such as these, shows promise and should be pursued further. Since using regional data for identification may always be impaired because the set of covariates is not as rich as in the household data, an ideal solution might be to merge the less noisy administrative data with the available variables in household surveys in order to capture the strengths of each data source.

If households have different propensities to consume based on the type of wealth gain or loss, accounting for these differential movements should yield more accurate macroeconomic forecasts.

Implications

Developing a better understanding of how wealth shocks impact consumption should help to resolve some current conundrums regarding this relationship. More importantly, gaining a deeper sense of this connection should improve forecasts of consumer spending and overall economic growth. In addition, a better understanding of how wealth shocks are transmitted through the economy will improve assessments of risks to the economic outlook. These issues are particularly important during periods of substantially fluctuating asset prices.

About the Authors

Daniel Cooper *is a senior economist in the research department at the Federal Reserve Bank of Boston.* **Karen Dynan** *is vice president, co-director of economic studies, and the Robert S. Kerr Senior Fellow at the Brookings Institution.*

....
Working Papers
....

The Impact of Managed Care on the Gender Earnings Gap among Physicians

by Alicia Sasser Modestino

....

complete text: http://www.bostonfed.org/economic/wp/wp2013/wp1301.pdf
e-mail: alicia.sasser@bos.frb.org

Motivation for the Research

During the 1980s, market competition in the U.S. health care industry increased significantly with the advent and spread of managed care. All three main types of managed care providers—health maintenance organizations, preferred provider organizations, and independent practitioner associations—contract with a selected network of physicians and hospitals to limit their fees in exchange for inclusion in the network. Managed care organizations monitor physician practice patterns and discourage excessive medical services and unnecessary procedures. Because managed care organizations combine the functions of paying for and providing medical care, these organizations are able to deliver more cost-effective services than the traditional fee-for-service insurance system, which separated the provider from the payer and thus lacked incentives for cost containment. In 1980 over 90 percent of privately insured individuals had fee-for-service coverage, but by 1992 only 4 percent of them had a fee-for-service arrangement.

Besides this dramatic change in the U.S. health care market, the 1980s also marked a decrease in the gender earnings gap among all college-educated U.S. women and a rise in the number of female physicians practicing in the United States. Yet while the improvement in the gender earnings gap among all college-educated U.S. workers stalled after 1989, the gender earnings gap among U.S. physicians continued to improve rapidly through the mid-1990s. This paper explores whether the widespread adoption of managed care in the 1980s may have had a differential impact on the labor market outcomes of female versus male physicians in ways that unintentionally favored female physicians.

Research Approach

The author uses cross-state variation in health maintenance organization (HMO) enrollments to examine whether recent changes in the U.S. health care industry have helped close the gender earnings gap among physicians. She employs a differences-in-differences methodology that compares changes in the female-to-male earnings differential for physicians working in states with high growth in managed care to those for doctors practicing in states with low growth in managed care. A state is classified as high growth or low growth based on its HMO penetration relative to the nation as a whole. Between 1980 and 1990 there was considerable regional variation in HMO enrollments—states in the New England, Middle Atlantic, Mountain, and Pacific regions had large increases in managed care penetration, while states in the South and Central regions lagged behind. To control for changes in HMO penetration between 1986 and 1990, the author uses a state's average number of employees per firm and a Herfindahl index measuring the degree of concentration in each state's hospital market in 1985. The results from a triple differences exercise are translated into a regression framework to compare the change in wages of female physicians relative to the change in wages of their male counterparts in high managed care states versus low managed care states between 1986 and 1990. Robustness checks are performed to compare the labor market experiences of physicians with other groups of professionals with advanced degrees, such as lawyers.

The information on physician earnings and characteristics comes from Practice Patterns of Young Physicians, a survey jointly administered in 1987 and 1991 by the American Medical Association, Mathematica Policy Research, and the Robert Wood Johnson Foundation. Informally known as the Young Physicians Survey (YPS), it was designed to investigate the factors that influenced the career decisions of doctors under 40 years of age who had been practicing medicine continuously for two to five years. The YPS gathered information on specialties, practice settings, hours, annual income, and other professional and demographic characteristics. The sample means for male and female physicians display little variation in terms of demographic and labor market characteristics. The mean physician age is 35 years and the mean practice length is 3.4 years. In 1986 female physicians had a lower rate of board certification, but by 1990 equal proportions of men and women were board-certified physicians. The annual incomes of women physicians were about one-third lower than the incomes of their male counterparts and their hourly earnings were about 15 percent lower. The difference between the annual and hourly earnings gap is largely due to gender differences in the number of hours and weeks worked. On average, female physicians worked 7–10 hours less per week and one week less per year than did male physicians. Part of the remaining difference in income can be attributed to different specialties and practice settings. Women are more likely to be primary care physicians, while men are more apt to be medical and surgical subspecialists. Women are more likely to choose salaried positions in institutionalized settings such as HMOs, hospitals, universities, public health clinics, and government. These work environments offer more regular schedules, fewer hours, and an established patient base in exchange for less prestige and lower salaries. Men are more apt to work in traditional solo or group practice settings as full or partial owners. Hence male physicians are more likely to receive fee-for-service reimbursements and share in the income of group practices. Women physicians tend to have a higher percentage of patients who are African-American or Hispanic, covered by Medicaid, or entirely without insurance coverage.

The second source of physician data used in the paper is the decennial U.S. population census in 1980, 1990, and 2000. Together, the first two censuses cover a decade, thus capturing

Changes in Female/Male Earnings Ratio
and Managed Care Enrollments Over Time

Ratio: Female/Male Annual Earnings   *Percentage of Population*

Legend:
- Ratio for College Graduates (left scale)
- Ratio for Physicians (left scale)
- HMO Enrollments (right scale)
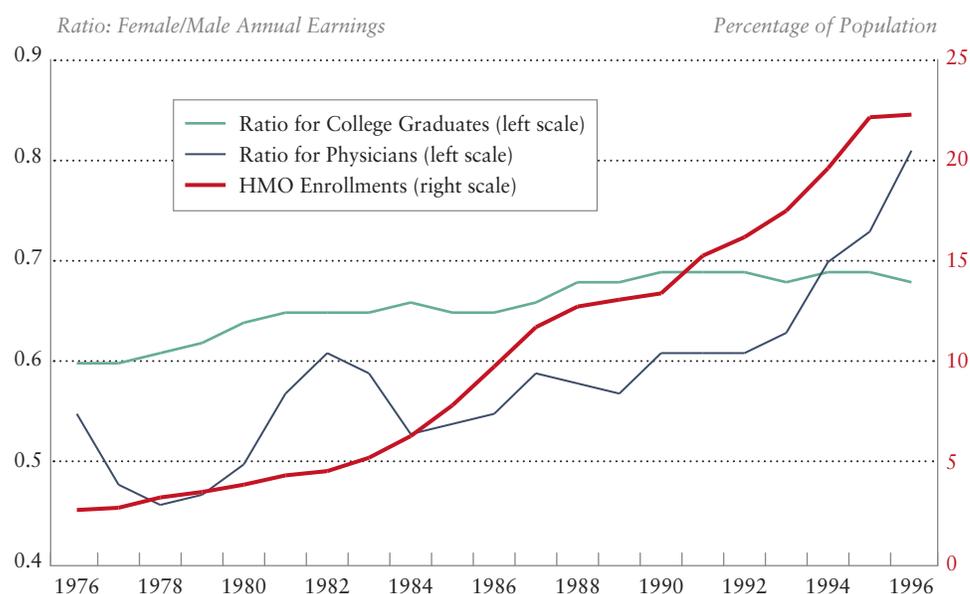
Source: Ratios for female physicians and college graduates are author's calculations using data from the U.S. Bureau of the Census, Current Population Survey. HMO enrollments from Gruber, Shadle, and Polich (1988), American Association of Health Plans (1995), and InterStudy (1996).
Note: Female/male earnings ratios are five-year moving averages as calculated by the author.

larger changes in managed care penetration than the four years separating the two YPS. By including physicians of all ages, the census data make it possible to test whether managed care has had a similar impact on all practicing physicians, not just those starting their professional careers. Most importantly, the census data contain information on individuals in other professional occupations, yielding placebo groups that can be used to control for factors associated with high-growth, managed care states that may affect the earnings of all professional women, not just physicians.

Regression estimates show that managed care reduces the differential in hourly earnings between male and female physicians by 10 percentage points.

Key Findings

- When controlling for demographic and professional characteristics as well as medical specialties and practice settings, regression estimates show that managed care reduces the differential in hourly earnings between male and female physicians by 10 percentage points. This results in a two-thirds reduction of the overall gender gap as measured in hourly earnings and improves women physicians' annual earnings by 7.3 percent, reducing the overall gender gap in annual income by about one-fifth.

- Managed care appears to have affected the overall distribution of physician earnings. Controlling for basic demographic characteristics shows that women physicians in states with high managed care growth saw their relative incomes improve by 17.8 percent between 1980 and 2000 when compared with women physicians in states with low managed care growth. Decomposing the gender earnings gap shows that the changes in the wage structure can account for about one-third of the improvement in the gender earnings gap among physicians in the high-growth states. The remaining two-thirds of the improvement are attributed to gender-specific factors that moved women up in the male earnings distribution. These gender-specific factors are partly related to the impact managed care has had on the relative demand for different medical specialties and practice settings.

Detailed interactions for primary care specialties reveals that managed care has had a positive impact on the earnings of pediatricians and general internists.

- A robustness check on the earnings of lawyers and other professionals with advanced degrees shows that the gender gap for women did not narrow more rapidly in states with high growth in managed care. This result indicates that managed care continues to have a positive and significant effect on the relative earnings of female physicians. Time-series evidence shows that while the improvement in the gender gap among all college-age workers has stalled since the early 1990s, the gender wage gap among physicians has continued to narrow as HMO penetration has continued to increase.

Implications

The paper's results suggest that market changes can have important consequences for the gender earnings gap when there are large pre-existing differences between men and women in a profession. In the case of the U.S. health care industry, the move to managed care has encouraged the greater use of less costly preventive care services, a shift that possibly increases the relative demand for primary care physicians such as family practitioners, general internists, and pediatricians, specialties chosen by a high fraction of female physicians.

About the Author

Alicia Sasser Modestino is a senior economist in the New England Public Policy Center, part of the research department of the Federal Reserve Bank of Boston.

The Power of Sunspots: An Experimental Analysis

by Dietmar Fehr, Frank Heinemann, and Aniol Llorente-Saguer

Motivation for the Research

The term "sunspots" refers to extrinsic random variables that may influence economic behavior but are not related to more fundamentally informative considerations such as payoffs, preferences, technologies, or endowments. In *The General Theory of Employment, Interest and Money* (1936), Keynes explained how price fluctuations may occur in equity markets. Using the example of a beauty contest in which people looked at six women's photographs and judged which one was the most beautiful, Keynes stated that a naïve strategy would have a player follow his or her own inclinations when choosing the most beautiful woman but that a more sophisticated strategy would choose based on which woman the majority of players would regard as the most beautiful. This second choice is based on some inference of public perceptions. In terms of asset prices, the lesson is that it may matter more how much other people value an asset than how much value an individual person assigns to it. The beauty contest example illustrates the idea that extrinsic information may affect agents' choices and behavior; for instance, in the real economy sudden swings in expectations based on extrinsic signals could trigger a financial crisis.

Azardias (1981) and Cass and Shell (1983) were the first researchers to theoretically explore the influence that extrinsic information may have on economic behavior. Both

studies showed that whenever there are multiple equilibria, there are also sunspot equilibria upon which agents condition their actions given publicly observable but intrinsically uninformative signals. Sunspots may serve as focal points for agents' beliefs, and the public nature of these sunspots may enable beliefs conditioned on those signals to become self-fulfilling. In the field, it is difficult to identify a particular extrinsic event that may affect an agent's choice, and even if such an event is identified, it is difficult to establish a causal relationship between an extrinsic event and an economic outcome. Laboratory experiments offer a controlled environment that permits a systematic empirical exploration of the impact that extrinsic information has on economic behavior.

To date, only a few studies have used laboratory experiments to investigate sunspot-driven behavior. Duffy and Fisher (2005) were the first to provide direct evidence for the occurrence of sunspots in a market with two distinct equilibrium prices. They found that the presence of sunspot equilibria depends on the market institution's particular information structure and a shared contextual understanding of these signals so that the information is interpreted in the same manner. In order to achieve a common understanding of sunspot variables, Duffy and Fisher included an initial training phase that alerted their subjects to the existence of high and low price equilibria in combination with the respective announcements. An important finding from related studies on third-party recommendations is that subjects only follow "credible" recommendations and tend to disregard advice to play an imperfect or less efficient equilibrium. Yet little is known about how the impact of extrinsic signals depends on their noise structure, meaning number of signals, the signal distribution, and whether these signals are publicly observable. An increasingly relevant question is whether sunspots emerge naturally and how the likelihood of observing sunspot equilibria depends on the nature of these signals. The authors designed an experiment that reliably produced sunspot-driven

Average Distance of Choices from the Risk-Dominant Equilibrium
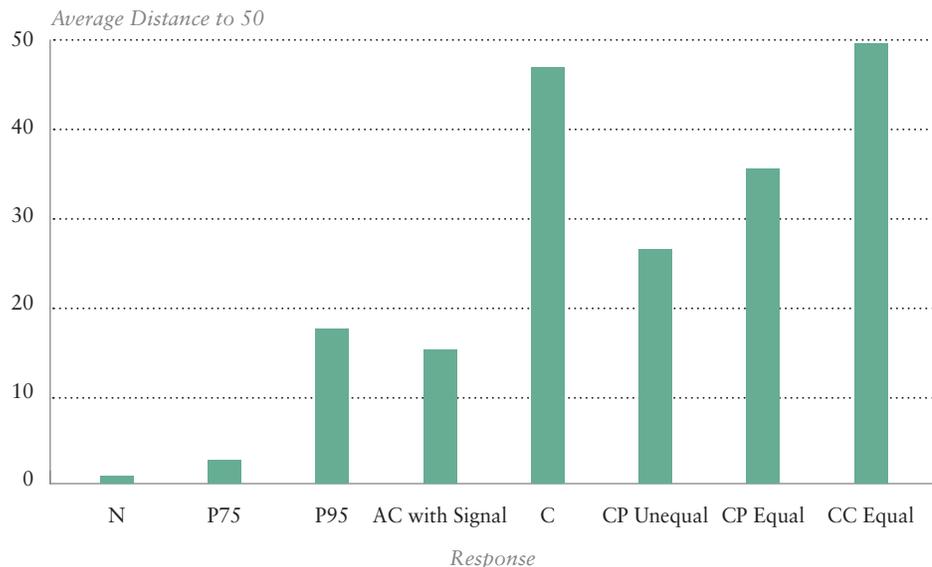
Source: Authors' calculations.
Note: In the benchmark treatment (Treatment N), subjects played the coordination game with payoff function (2) and received no extrinsic information. In all the other treatments, the subjects received extrinsic information (signals) and the authors varied their public nature and the number of signals.

behavior without explicitly priming or recommending that subjects follow extrinsic signals, in order to investigate how the influence of extrinsic random signals may depend on their noise structure. Since the payoffs were not contingent on the signals, the players were free to ignore the signals, in which case sunspot equilibria would not occur.

Research Approach

The authors designed a game that used extrinsic signals and systematically varied the information structure of these signals to control the available extrinsic information and its effect on the subjects' behavior. The authors used a simple two-player coordination game with random matching, a setup that can be considered a reduced form of a market setting. The players independently and simultaneously picked a number between zero and 100, maximizing their payoffs if they chose the same number. Deviations were punished with a quadratic loss function. Each coordinated number selection constituted a Nash equilibrium and payoffs did not depend on the number that players coordinated on. The game had a risk-dominant equilibrium, picking 50, that served as a natural focal point in the absence of a coordination device. In the experiment, the extrinsic signals (sunspots) were binary random variables, either zero or 100, that were unrelated to payoffs. These signals had two properties that the experiment exploited. First, these signals were semantically meaningful because they clearly map to the action space and can easily be used as coordination devices, providing a second focal point in addition to the risk-dominant equilibrium. Second, these semantically meaningful signals were extreme in the sense that they pointed to the lowest or highest possible action, thereby maximizing the tension between different focal points. Since the risk-dominant criterion allowed ordering the different equilibria by their distance from 50, the authors could measure the power of sunspots by how distant the players' actions were from the risk-dominant equilibrium. The sunspot equilibria arose endogenously without the need for an initial training period required in earlier studies. The authors varied the number of signals, their distribution, and their degree of public availability, measuring the average distance between the chosen actions and the risk-dominant strategy and the portion of groups that converge to sunspot equilibria. They then investigated to what extent publicly available information was necessary for sunspot-driven behavior to occur and how subjects aggregated the available information. The design allowed the authors to isolate the welfare effects of the miscoordination induced by extrinsic information.

In all the experimental treatments, the subjects repeatedly played the simple coordination game for 80 rounds. The subjects were randomly assigned to matching groups of six that were fixed for a given session. In each period the subjects were randomly matched into pairs within a matching group. Since there was no interaction between the subjects from different matching groups, the data from different matching groups were regarded as independent observations. The players were aware that in each round they were randomly paired with another subject from their matching group and that they would never face the same subject twice in a row. Public signals were revealed to both players in a pair and the subjects were aware that each of them received the same signals. In treatments with private signals, each subject received an independently drawn signal that was not revealed to the other player. The experiment varied the number of signals that subjects received. In some treatments players got either a private or a public signal, and in two treatments they each received two signals, either one private and one public signal, or two public signals. Different treatments varied the probability of both players receiving the same signal. After each round, the players learned their partner's choice, the distance between their own choice and their partner's choice, and the resulting payoff. In treatments with private signals, their partner's private

signal was never revealed. Subjects were seated at computer stations, informed of the payoff methods, and had to answer questions about the game's procedures and payoffs to ensure they understood the rules before the experiment began. A total of 288 students participated in the laboratory experiment held at Technical University Berlin. At the end of a session 10 out of the 80 periods were randomly selected for payment, with each point converted to 1 euro cent. The sessions lasted for about an hour, and subjects received a fee of 3 euros for showing up. On average, each player earned 21 euros.

The authors analyzed the data by checking whether subjects within a matching group converged to a common strategy and identified the strategies on which they converged. The distance of these choices from the risk-dominant equilibrium was used as a measure of the sunspots' power. These distance measures were then used to perform a detailed analysis of the differences in behavior across different treatments. Two convergence criteria were used. The strong convergence criterion required that all six subjects in a matching group played according to the same strategy, allowing a deviation of plus or minus 1 for periods 70–79. The weak convergence criterion required that at least four subjects in a matching group followed the same strategy, allowing for a deviation of plus or minus 3 for periods 70–79. If a group converged to a common strategy, the strategy to which it converged was determined by the choices of the majority of subjects who fulfilled the convergence criterion. For converging groups, four types of coordinating strategies were identified: 1) 50, the risk-dominant strategy, 2) intermediate sunspot strategies, such as 25/75 or 10/90, in which subjects chose the lower number when the signal was zero and the higher number when the signal was 100, 3) 0/100, following the signals, and 4) a mean that played the average of both signals.

Key Findings

- Extrinsic public signals that are easily aggregated lead to almost perfect coordination on the sunspot equilibrium implied by the semantics of the signals. This salient sunspot equilibrium reliably showed up when subjects just received two public signals, even when the sunspot equilibrium is associated with higher strategic risk than any other strategy.

> *The authors found, contrary to theoretical predictions, that the power of sunspots was significantly lower if private and public signals were combined.*

- Coordination on the salient sunspot equilibrium was less pronounced when public and private signals were both present, as some subjects then conditioned their actions on the private signal, which either prevented full coordination of actions or led to an intermediate sunspot equilibrium. While theory predicts the same set of equilibria as in a game with just one public signal, the authors found that the power of sunspots was significantly lower if private and public signals were combined. When subjects received both private and public signals, the different groups of subjects coordinated on different equilibria for the same external conditions.

- In the absence of public signals, the risk-dominant equilibrium dominated. However, sunspot-driven behavior can be observed for highly correlated private signals. This observation indicates that the likelihood of sunspot-driven actions may be a continuous function of the correlation of signals, while equilibrium theory predicts that sunspot-driven behavior can occur only if the signals from different agents are perfectly correlated.

- The occurrence of sunspot-driven behavior or sunspot equilibria largely depends on the distribution of strategies in the early periods of the game. In treatments where different groups coordinated on different equilibria, there was a significant correlation between

behavior that occurred in the first and the last periods of the game, in line with previous results on coordination games.

- Subjects' payoffs are U-shaped in the power of sunspots, measured by how distant the actions are from the risk-dominant equilibrium; hence, between treatments there are significant differences in average payoffs. Miscoordination arises from a slower convergence process toward a common strategy, or a lack of convergence and coordination on a non-equilibrium strategy, particularly if sunspot-driven behavior imposes negative externalities on agents who do not receive signals.

- The authors' results contrast with previous findings that subjects only follow credible recommendations. This experiment shows that subjects may follow a random coordination device, even if it is riskier to do so and even if such behavior has no equilibrium. Unlike the theoretical prediction, highly precise private signals may not only impede coordination but may also lead to coordination on nonequilibrium strategies.

- Different information structures induce very different behavior. Purely public information reliably generates sunspot equilibria but receiving no information or imprecise private information leads to the risk-dominant equilibrium. In terms of welfare, the chosen equilibrium does not matter, but it does matter whether and how fast subjects converge to an equilibrium. If a certain information structure results in a lower convergence process, there is frequent miscoordination in the early periods that yields welfare losses, judged by the group's average payoff.

Implications

The finding that the impact of sunspots is reduced by the presence of private signals, which impedes the ability of groups to coordinate on an action and leads to welfare losses, has an interesting implication: in economies where salient private signals exist, adding an extrinsic coordination device with similar semantics may make it more difficult to coordinate actions. Since the introduction of extrinsic information influences subjects' perceptions of focal points, considerable miscoordination may result. Hence the authors' results show that focal points can be quite fragile.

It remains an open question whether sunspots may be powerful enough to move agents away from a payoff-dominant equilibrium. Since risk-dominance seems to work well in measuring the power of extrinsic signals, the authors envision that their game form might be used for testing the salience of other messages or signal combinations. It may also be possible to use similar experiments to measure the common understanding of messages expressed in ordinary language.

About the Authors

Dietmar Fehr cowrote this paper while he was a visiting scholar in the research department of the Federal Reserve Bank of Boston. He is a research fellow at the Wissenschaftszentrum Berlin (WZB). Frank Heinemann, the corresponding author, is a professor of economics at Technische Universität Berlin, where he is the macroeconomics chair. Aniol Llorente-Saguer is a senior research fellow at the Max Planck Institute for Research on Collective Goods in Bonn.

Do Real-Time Okun's Law Errors Predict GDP Data Revisions?

by Michelle L. Barnes, Fabià Gumbau-Brisa, and Giovanni P. Olivei

Motivation for the Research

The empirical regularity of a negative relationship between movements in the unemployment rate and GDP, first established by Arthur Okun (1962) and known as Okun's Law, is an important tool for the conduct of monetary policy. Errors in Okun's Law, if persistent, tend to be heavily scrutinized, as they may convey information about changes in potential GDP and/or the natural rate of unemployment. Having a reliable estimate of the size of the output and unemployment rate gaps is of crucial importance for the proper conduct of monetary policy. So it is not very surprising to see policymakers continually reverse-engineering Okun's Law as one way of drawing inferences about these gaps. For example, the recent noticeable decline in the unemployment rate from a peak of 10 percent in October 2009 to 7.9 percent in October 2012, coupled with GDP growth averaging 2.2 percent over the period 2009:Q4 to 2012:Q2, could be taken as evidence of a decline in potential GDP growth from the 2 to 2.5 percent estimates that prevailed before the onset of the 2008–2009 recession.

Real-time errors in Okun's Law, however, convey other economic information that is not related to changes in potential GDP and/or the natural rate of unemployment. This paper examines and seeks to understand the information contained in Okun's Law errors.

Research Approach

The authors estimate errors in Okun's law using real-time data and a first-difference specification of the Okun's Law relationship, in which the change in the unemployment rate is a function of the change in the natural log of real GDP. Their methodology relies on estimating the relationship in real time and considering at each point in time the most recent error in the relationship. Using real-time data implies that in every quarter over the estimation period spanning 1965:Q4 to 2012:Q4 they take the latest vintage of GDP data available on the 15th day of that quarter's middle month. This latest vintage contains information up to the previous quarter. However, the authors consider as real-time the information encompassing at least the third GDP release (the "final" release) from the Bureau of Economic Analysis (BEA). In practice, this means that at any given quarter t they take the latest GDP vintage available at that time, up to quarter t–2. They follow the same timing approach for the real-time unemployment rate, although revisions to the unemployment rate are minor and related only to adjustments in the seasonal factors. This is an essential feature for the purpose of the exercise, the authors' premise being that the unemployment rate series is not subject to material revision and that therefore, as filtered through Okun's Law, it may feature information about the state of the real economy that the real-time GDP data do not fully capture.

The authors estimate both a baseline specification, using ordinary least squares, and a variant, using a maximum-likelihood-based Kalman filter. They introduce the variant to account for the possibility of changes in the equilibrium unemployment rate and in potential growth over a relatively short time frame. The real-time vintages of data are used in regressions with a rolling window of 60 quarters and are taken from the "Real-Time Dataset for Macro-

economists" maintained by the Federal Reserve Bank of Philadelphia. The object of interest is the last estimated Okun's Law error from each rolling regression. The procedure mimics a real-time exercise where, at each point in time, the econometrician estimates Okun's Law with the most up-to-date information available. The most recent error in the estimated relationship is taken as providing a real-time assessment of how closely the real-time measurement of GDP growth is reflected in movements in the unemployment rate. The authors then turn to evaluating whether the estimated series of Okun's Law errors predicts revisions to GDP, relate their findings to the existing literature, and then assess the stability of the Okun's Law relationship over time. Next, they assess the ability of Okun's Law errors in real time to predict revisions to GDP in the post-2007 period. They then perform robustness checks, including looking at the sensitivity to the revision date, the form of the specification, the estimation period, and the width of the rolling window. The authors explore whether, when projecting future economic activity, forecasters take into account discrepancies between the unemployment rate and real-time GDP readings as seen through lens of Okun's Law. They conclude by summarizing and interpreting the results.

Key Findings

- Real-time errors in Okun's Law contain information about future revisions to GDP. If the unemployment rate increases (decreases) by more than the amount that Okun's Law predicts on the basis of real-time GDP readings, then those GDP readings will later be revised to show less (more) growth than the statistical agency was first assessing.

- According to the authors' estimates, a change in the unemployment rate that is 1 percentage point greater than predicted by Okun's Law in real time is associated with a roughly 2 percent downward revision to GDP growth two years later. These predicted revisions are

Okun's Law Errors in Real Time
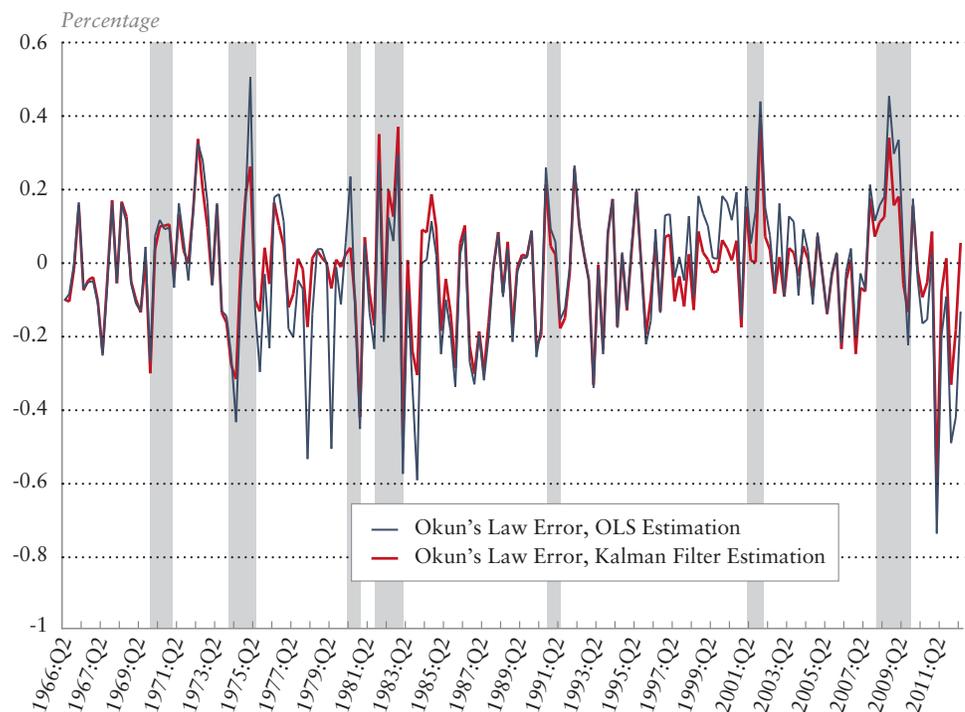
Source: Authors' calculations.
Note: Shaded areas indicate recession.

larger in the post-1983 period, where a 1 percentage point positive error in Okun's law is found to translate into a downward revision to GDP growth in that quarter at an annual rate of 4 percent.

- The information in Okun's Law errors about the true state of real economic activity also helps to improve GDP forecasts in the near term. For example, during the 2008–2009 recession the unemployment rate increased by more than the amount that Okun's Law would have predicted on the basis of real-time GDP data. Later, those GDP data were revised to show substantially less growth than initially thought. GDP growth forecasts from the Survey of Professional Forecasters tend to be overstated (understated) when real-time Okun's Law errors are suggesting weaker (stronger) growth than the real-time signal from GDP.

Implications

It is the portion of the change in the unemployment rate not explained by Okun's Law in real time that has predictive power for future GDP revisions.

This paper contributes to a vast literature on GDP data revisions. While the ability of the unemployment rate—which, unlike GDP, is not subject to revision aside from minor seasonal adjustments—to predict future GDP revisions has already been noted, this is the first paper, to the authors' knowledge, to show that it is the portion of the change in the unemployment rate not explained by Okun's Law in real time that has predictive power for future GDP revisions. Moreover, the authors show that this predictive power is significant from an economic standpoint. The paper also establishes that information contained in real-time Okun's Law errors can produce some improvement in the near-term forecasts of GDP growth in the Survey of Professional Forecasters. Thus, the paper provides additional evidence on the predictability of forecast errors in the Survey of Professional Forecasters.

The paper also relates to the literature on the relevance and stability of Okun's Law itself, as it provides a caveat against drawing overly strong conclusions about potential output or the natural rate of unemployment, when working with real-time GDP data. Ultimately, however, the authors view the paper's main contribution as adding another dimension to the policy debate on what signal to extract from Okun's Law errors in real time. While typically the policy discussion around these real-time errors is framed in terms of changes to potential GDP, the equilibrium unemployment rate, or transitory changes in labor's intensive margin (possibly via changes in effort), this paper stresses the information content of these errors for GDP data revisions and for assessing the true pace of output growth in real time.

About the Authors

Michelle L. Barnes is a senior economist and policy advisor in the research department at the Federal Reserve Bank of Boston. Fabià Gumbau-Brisa is a senior economist in the research department at the Federal Reserve Bank of Boston. Giovanni P. Olivei is a vice president and economist who oversees the macroeconomic/international group in the research department at the Federal Reserve Bank of Boston.

Window Shopping

by Oz Shy

....

Motivation for the Research

Consumer Reports magazine recently surveyed over 10,000 readers and found that 18 percent of them bought electronic products online after they had examined these same products in a brick-and-mortar store. More than half of this group eventually bought online from Amazon.com. The author refers to the practice of inspecting products at a walk-in retailer before buying them online as "window shopping."

Research Approach

The author constructs an analytical model of potential buyers who differ in their preference for after-sale services that are not offered by online sellers. Technically speaking, the walk-in retailer and the online seller are assumed to be vertically differentiated, implying that, in the absence of the transportation costs incurred from going to the store, all buyers would prefer to buy the product at the brick-and-mortar store if its price did not exceed the online price. While a trip to the brick-and-mortar store is costly for some buyers, it confers the informational advantage of mitigating the uncertainty as to whether the product will suit the buyer's needs.

The paper derives equilibrium prices, profits, consumer welfare, and social welfare in order to examine the relationship between the equilibrium number of window shoppers and the socially optimal number. The analysis first concentrates on a duopoly market structure where an online seller and a walk-in retailer compete to attract potential buyers. The same investigation is then conducted for single ownership of (or a merger between) the online and the brick-and-mortar outlets.

The model draws heavily on Shin (2007). Both papers model consumers who are uncertain as to whether the product suits their needs and therefore would benefit from expert, in-store advice on this matter. In both models, a retailer that does not provide pre-sale service may be able to free ride on pre-sale service provided by the rival vendor. However, there are some substantial differences between the two models. In Shin's model, the two retailers are identical in all respects (including buyers' transportation costs) and both are capable of providing identical pre-sale services. In the author's model, the online and the walk-in sellers differ in their ability to provide pre- and post-sale services, and buyers' costs for transportation and shopping time depend on whether they shop online or at the walk-in store. Consequently, in Shin's model, if the two retailers charge identical prices, all potential buyers (informed and uninformed) would patronize the retailer that offers the pre-sale service. In contrast, in the author's model, under equal prices each seller will face some positive demand.

Key Findings

• In equilibrium, some consumers will travel to the store, inspect the product, and then, if they find the product suitable, will leave the store and purchase the product online because the online price is cheaper. These consumers decide to buy the product online because once

a potential buyer travels to the store, the buyer views the transportation costs as sunk and therefore these costs do not influence the decision about where to purchase the product.

- From a social welfare perspective, assuming that the sellers have the same marginal costs, window shopping behavior is excessive; that is, the equilibrium number of window shoppers exceeds the optimal number. The reason, according to the model's assumptions, is that a potential buyer who has already traveled to the store and found the product to be suitable should buy it at the store rather than online because the store provides after-sale service and the online seller does not.

- If the walk-in store and the online seller merge and operate as a single profit-maximizing firm, and if the consumer's utility function is modified to include a reservation utility of zero, then joint ownership of the online and the walk-in store does not eliminate window shopping. This activity serves an important function for consumers who have low transportation costs or place a low value on their time, even for those who do not expect to benefit very much from after-sale service, because it allows these buyers to decide whether the product suits their needs before purchasing it.

- The gap between the equilibrium number of window shoppers and the optimal number becomes smaller with an increase in the walk-in store's marginal cost.

From a social welfare perspective, assuming that the sellers have the same marginal costs, the equilibrium number of window shoppers exceeds the optimal number.

Implications

At this stage of preliminary research it is difficult to draw definite policy or regulatory conclusions because in reality the cost of providing in-store services may affect fixed costs more than they affect marginal costs. If service costs affect marginal costs only, this paper shows that the gap between the optimum and the equilibrium number of window shoppers likely narrows to a degree that makes policy intervention unwarranted.

In reality, the competition between online and walk-in retailers is more complicated than the environment modeled in this paper. The following list suggests some possible extensions of the model. First, the Internet provides product reviews by other buyers, which can be used by online consumers as well as by walk-in store buyers to make buy/not buy decisions. Second, the retail industry environment is evolving in many ways. Fearing a loss of customers to online retailers, many large brick-and-mortar retailers now offer online shopping with either home delivery or store pickups. In addition, many online retailers offer easy returns, some offer "free returns," and some provide links to webpages where customers can find aftermarket service providers in their area. In addition, online sellers keep introducing more and more products, such as eyeglasses, that until recently were available only in walk-in stores. Online offers for these products feature significant price reductions that are possible because online merchants have learned how to bypass the middlemen and shorten the supply chain. Finally, online retailers are exempt from sales tax in most states where they do not have a physical presence. This cost advantage may disappear in the future if the legal environment changes with respect to sales taxation of online purchases.

About the Author

Oz Shy is a senior economist and a member of the Consumer Payments Research Center in the research department at the Federal Reserve Bank of Boston.

Cyclical Unemployment, Structural Unemployment

by Peter Diamond

Motivation for the Research

Between December 2007 and June 2009, the dates formally assigned by the National Bureau of Economic Research (NBER) to mark the technical beginning and end of what is widely known as the Great Recession, the United States experienced its worst economic contraction since the Great Depression. A housing market boom and bust sparked this most recent downturn, and it is well established in the macroeconomic literature that recessions stemming from financial crises tend to have very long recoveries. For the last four years, the U.S. economy has suffered from persistently high unemployment rates that have impeded a vigorous economic rebound.

The Beveridge curve compares the number of job vacancies (openings) to the unemployment rate and is regarded as a proxy for how well the labor market's matching function is working. During a recession it is expected that lower vacancy rates will be accompanied by higher unemployment rates, while during an expansion higher vacancy rates will be coupled with lower unemployment rates. Since the Great Recession's end in June 2009, the Beveridge curve pattern has been erratic—there have been two periods when the vacancy rates have risen with little impact on unemployment, and two periods when unemployment had fallen but job openings showed no steady rise. Since September 2009, all the Beveridge curve observations have been noticeably above a curve connecting the observations that took place before and during the recession. Thus, the present situation is that the U.S. unemployment rate is high, the number of job openings is low, and vacancies are higher than at the same unemployment rates during the Great Recession.

Whenever unemployment remains high for an extended period, it is very common for a debate to center on whether the high unemployment rate is due to structural or to cyclical reasons. Since movements along the Beveridge curve are typically assumed to reflect cyclical factors, while a shift in the curve itself is taken to indicate structural effects, these periods of rising vacancies unaccompanied by falling unemployment suggest that structural unemployment may have increased in the United States—in other words, the U.S. economy may now have a "new normal" in terms of a higher long-term level of unemployment. This structural interpretation of recent moves in the Beveridge curve can be taken to imply that policymakers should not be so concerned about stimulating aggregate demand through monetary and fiscal measures, as the structural shift indicates a more permanent change in the employment rate that will be unresponsive to cyclical stimulus. Yet despite a possible increase in structural unemployment, many recent analyses conclude that a sizable component of current U.S. unemployment is due to cyclical factors. In an effort to help resolve the current policy debate, the author offers a conceptual critique of issues that complicate how shifts in the Beveridge curve are interpreted.

Research Approach

The author surveys the existing literature's methodological assumptions underlying employment measures used to interpret the Beveridge curve. The paper is organized around the

central inquiry: to what extent do outcomes in the labor market as seen through the Beveridge curve imply that when the economy has recovered, the future target for unemployment should be different from the level of unemployment in the period before the onset of the Great Recession. Beveridge (1944) characterized full employment as "more vacancy jobs than unemployed men." Dow and Dicks-Mireaux (1958) used a definition of the equality between vacancies and the unemployed to separate times of high and low demand. The supply of vacancies helps to determine the full employment point. There can be different supply functions of vacancies and different ratios of unemployment to vacancies at different full-employment equilibrium points. A standard matching function approach, which relates the flow of hires to the stocks of unemployed workers and job vacancies, plays a central role in many interpretations of changes in the Beveridge curve. The efficiency parameter of the matching function can affect the speed with which jobs are filled.

The first main section follows Barlevy (2011) and examines the magnitudes of cyclical and structural unemployment under two assumptions: 1) that the shift in the Beveridge curve to fit the recent data would last through the recovery and 2) that a new full-employment equilibrium would lie on that curve at a point consistent with a higher unemployment-vacancy rate than at the previous full employment point. The next section explores what factors help to account for the decline in the efficiency parameter of the matching function and whether these causes can be expected to last through a full recovery, thus lending credence to a structural interpretation of the recent shift in the Beveridge curve. This analysis considers both the patterns of previous U.S. recoveries in the postwar period, and the possible implications posed by differences in the causes and magnitude of the Great Recession. Using firm-level data from December 2000 through 2006 and published data through December 2011, Davis, Faberman, and Haltiwanger (2012b) find that the speed of filling a vacancy varied by industry,

The Beveridge Curve (Job Openings vs. Unemployment Rate)
(Seasonally Adjusted)
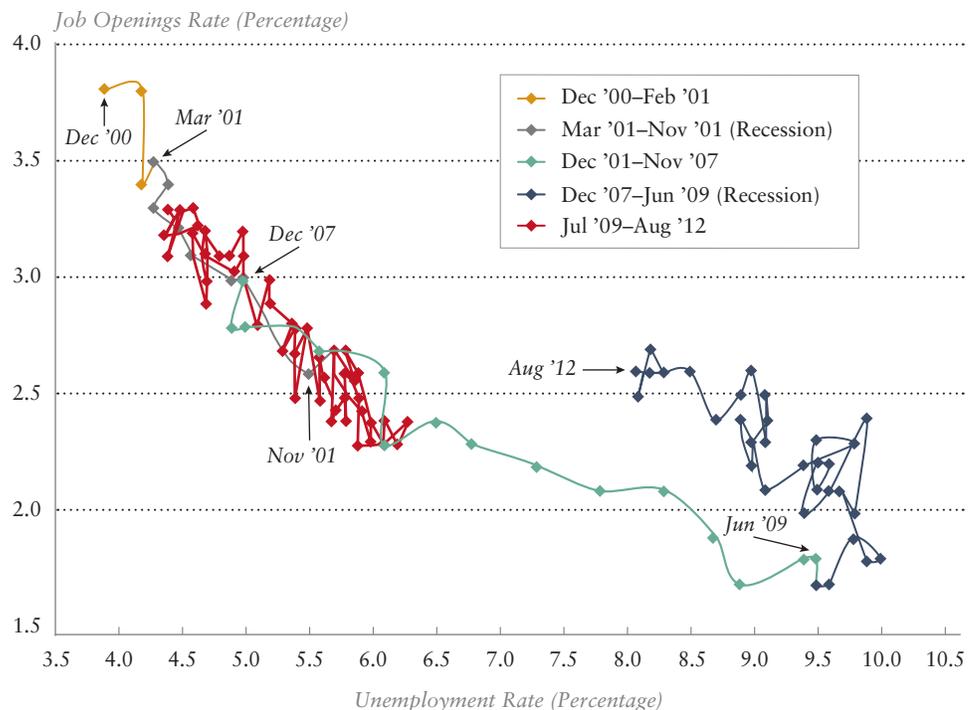
Source: Bureau of Labor Statistics, Current Population Survey, and Job Openings and Labor Turnover Survey, October 10, 2012.

by firm size, by turnover, and by firm growth, and that the proportion of hiring in different industries varies over the business cycle. Their simulation yields a better fit than one determined by the standard matching function. Looking at the Great Recession, they find that recruiting intensity per vacancy declined by over 21 percent between December 2007 and June 2009, and remained 11 percent below its pre-recession level as of September 2011. The next section considers how flows into unemployment, divided between workers who remain in the labor force and are counted as unemployed and workers who drop out of the labor force, affect the Beveridge curve. Kudlyak and Schwartzman (2012) analyze the direct effect of the four flows involving nonparticipation on the unemployment rate and conclude that, compared with the other postwar recessions, the 2007–2009 downturn produced a particularly large increase in the unemployment rate and a slow decline from this peak rate. The author then takes up the concept of "mismatch" in the labor market, which means that the existing job vacancies are not readily filled by the stock of unemployed workers for reasons related to skills, location, and so on. He finishes by discussing the limitations of a Beveridge curve analysis based on a strict technical interpretation.

Key Findings

- Concentrating on the steady-state relationship between unemployment and vacancies, an approach that does not assign a role to the dynamic pattern of movements around the steady-state curve, the results in Barlevy (2012) show that a decline in matching new hires began around December 2007. The author suggests that a decline in the matching function may be part of the normal Beveridge curve pattern in a recession.

- Additional factors influence the relationship between the Beveridge curve and the matching function, which is a relationship between hiring and two proxy variables for hiring, the stock of unemployed and job vacancies. But unemployed workers account for only a fraction of new hires, who are also drawn from the ranks of currently employed workers and labor market nonparticipants (those people not counted in the labor force as either employed or actively searching for work). Therefore, the matching function relationship is affected if these other variables change their patterns relative to unemployment, hiring, or vacancies or if a disaggregation of vacancies implies a changed relationship between the aggregates.

- A decline in the measured efficiency parameter of the standard matching function during the Great Recession, a very severe downturn, and a continued low level of the efficiency parameter during the recovery would contribute to a wider loop in the dynamics around the Beveridge curve. This might help to account for the unusual patterns observed since June 2009.

- Using the results in Davis, Faberman, and Haltiwanger (2012b), the author contends that the drop in recruiting intensity caused the measured efficiency parameter of the standard matching function to decline. He concludes that there is no reason to think that this decline in the matching function will be long-lasting once the economy fully recovers. So the additional unemployment from a decline in the matching function cannot be viewed as producing a structural, long-term change in the unemployment rate.

- Since differences in hiring across firms impact measurements of the matching function, and a decrease in the matching function appears to be a normal cyclical movement as the economy slows, ignoring such changes in the matching function may be problematic when

projecting the estimated Beveridge curve beyond the range of historical data, as in the case of the Great Recession.

- The author argues that several factors likely combined to generate the unemployment patterns found in by Kudlyak and Schwartzman (2012). The severity and length of the Great Recession and its aftermath means that the pool of long-term unemployed workers (those out of work for more than six months) increased. The extension of unemployment benefits authorized by Congress likely increased the time that an idle worker would remain in unemployment instead of exiting the labor force. If the stock of unemployed workers increases, this pattern will contribute to the appearance of a shift in the Beveridge curve. However, much of this effect is likely to go away when the economy rebounds and these extended benefits end.

Implications

While the Beveridge curve conveys important information about the state of the labor market, it should not be viewed as a tight technical relationship, and inferences made from it should be based on factors underlying the curve's unemployment and vacancy observations. Characterized by a financial crisis and the bursting of a housing bubble, the Great Recession was marked by a length and severity that have distinguished it from previous postwar recessions and recoveries. The author shows that one cannot necessarily interpret the recent outward shift in the Beveridge curve as indicating an increase in structural unemployment. Distinguishing between structural changes and cyclical changes requires more detailed analysis before concluding that policies aimed at stimulating the economy may be unwarranted. Since the efficiency parameter of the standard aggregate matching function should vary over the course of a business cycle, the author suggests that labor market analysis would benefit from a sharp distinction between trend and cycle issues, similar to the macroeconomic trend addressed by a Solow production function and the cyclical component treated in Okun's law. The author finds considerable evidence that cyclical unemployment accounts for much of the high unemployment currently present in the U.S. economy. He does not believe there is a single reason that fully explains the depth and severity of the current unemployment cycle, but suggests that inadequate aggregate demand is a major factor in explaining why the current U.S. unemployment rate is not falling faster.

About the Author

Peter Diamond wrote this paper while he was a visiting scholar in the research department of the Federal Reserve Bank of Boston. He is an Institute Professor and professor of economics, emeritus, at the Massachusetts Institute of Technology and was a co-recipient of the 2010 Nobel Memorial Prize in Economic Sciences.

b-13-1

A Decomposition of Shifts of the Beveridge Curve

by Rand Ghayad

Motivation for the Research

The apparent outward shift of the Beveridge curve—the empirical relationship between job openings and unemployment—has received much attention among economists and policymakers in the recent years. Many analyses point to the availability of extended unemployment benefits as a reason behind the shift. However, other explanations have also been proposed for this shift, including worsening structural unemployment. If the increased availability of unemployment insurance (UI) benefits to the long-term unemployed is responsible for the shift in the Beveridge curve, then allowing these benefits to expire should move many of the long-term unemployed back to work (or out of the labor force).

With the sharp increase in the unemployment rate during the recent recession, Congress enacted a series of UI extensions, allowing jobless individuals to collect up to 99 weeks of benefits in some states. Even though the U.S. labor market has been improving, there are still nearly three unemployed workers for each job opening, and the average duration of unemployment is currently 40 weeks—longer than the 26 weeks of benefits that an unemployed worker is normally eligible to collect. Under the Emergency Unemployment Compensation program, eligible workers could receive up to 53 weeks of coverage to regular and extended benefits for a combined total of 99 weeks in states with the highest unemployment rates.

This policy brief is an extension of recent work by Ghayad and Dickens (2012) on the Beveridge curve that intends to answer more succinctly the question economists have been asking: "Will the Beveridge curve move back when unemployment benefits expire?" Evidence in the earlier policy brief confirmed that the increase in job openings relative to unemployment—depicted by the outward shift of the Beveridge curve—has taken place only among the long-term unemployed, suggesting a possible role for extended UI benefits.

Research Approach

In this brief the author decomposes the aggregate Beveridge curve gap to estimate the contribution of the different unemployment categories to the deviation of the vacancy and unemployment rates from their historical empirical estimation. He uses a similar method to fit empirical Beveridge curves for job leavers, new entrants, and re-entrants, as well as job losers. In each category, he estimates the deviation in the unemployment rate of each group from its fitted curve for the period September 2009 onwards.

According to the classification scheme of the UI program, an unemployed worker's reason for being unemployed is a major factor in determining whether or not the worker is eligible to collect unemployment benefits. Job losers, who are often qualified to receive unemployment benefits, constitute only about half of the total unemployed (53 percent in January 2013), while the remaining portion is composed of job leavers, new entrants, and unemployed

Monthly Vacancy and Unemployment Rate Using Job Leavers, New Entrants, and Unemployed Reentrants
January 2001–January 2013
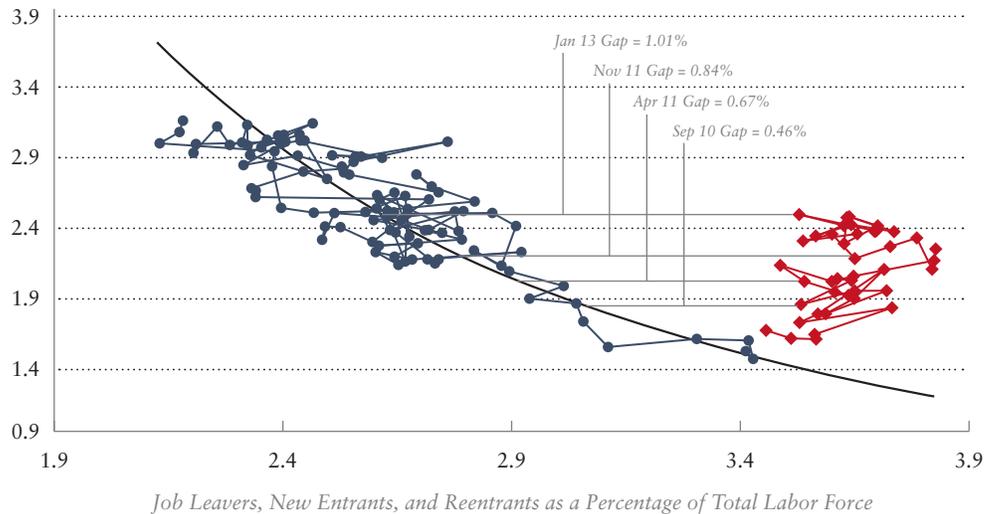
Job Openings as a Percentage of the Labor Force

Job Leavers, New Entrants, and Reentrants as a Percentage of Total Labor Force

Source: Bureau of Labor Statistics, Current Population Survey, and Job Openings and Labor Turnover Survey.
Note: The graph plots the job openings versus the unemployment rate using job leavers, new entrants, and re-entrants. The blue dots are the observations for 2001:m1–2009:m9. The red diamonds are the observations for 2009:m9–2013:m1. Data are seasonally adjusted monthly rates. The black curve is a fitted estimation using data prior to September 2009. For a given job opening rate, the gap is calculated by measuring the deviation of the actual unemployment rate from that implied by the fitted curve.

Monthly Vacancy and Unemployment Rates Using Job Losers
January 2001–January 2013
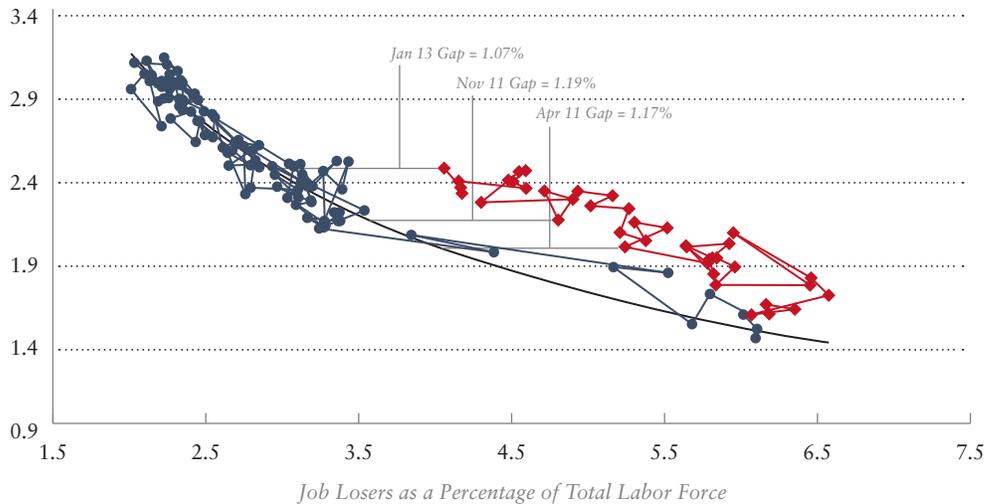
Job Openings as a Percentage of the Labor Force

Job Losers as a Percentage of Total Labor Force

Source: Bureau of Labor Statistics, Current Population Survey, and Job Openings and Labor Turnover Survey.
Note: The graph plots the vacancy rate versus job losers as a fraction of the entire labor force. The blue dots are the observations for 2001:m1–2009:m8. The red diamonds are the observations for 2009:m9–2013:m1. Data are seasonally adjusted monthly rates. The black curve is a fitted estimation using data prior to September 2009. For a given vacancy rate, the gap is calculated by measuring the deviation of the actual unemployment rate from that implied by the fitted curve.

Beveridge Curve Gap Decomposed by Reason for Unemployment

Percentage
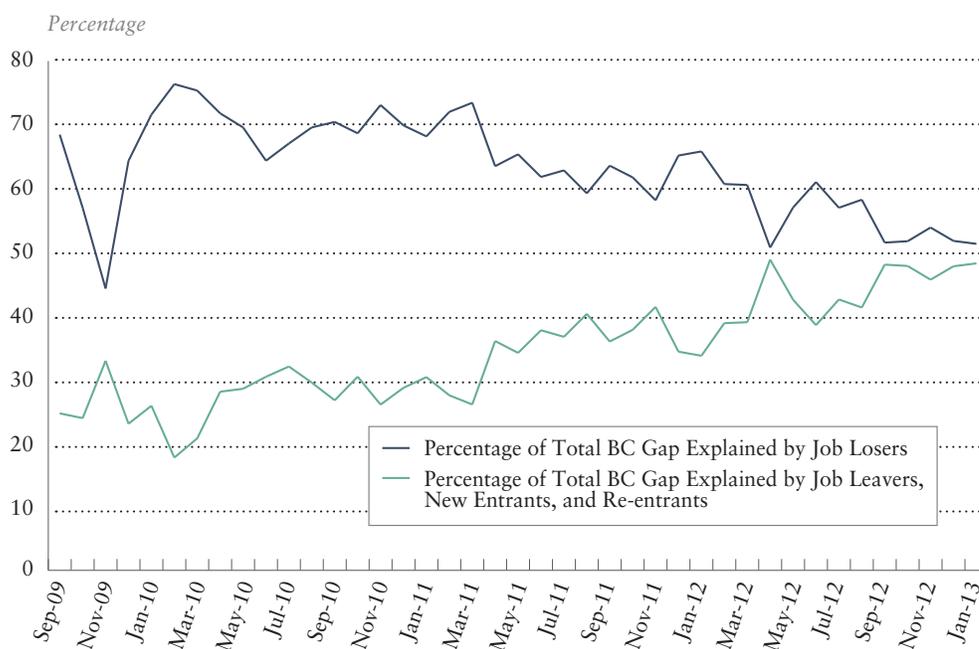
Source: Author's calculations.
Note: The residual unexplained gap is due to measurement error.

re-entrants to the labor market, groups that are generally not eligible to receive unemployment benefits. Thus, if part of the current shift in the Beveridge curve is explained by unemployed workers who are ineligible to collect benefits, then the Beveridge curve will not shift back to its pre-recession position when benefits for the long-term unemployed are discontinued.

To estimate which groups account for the breakdown in the job vacancy and unemployment relationship, the author decomposes the recent deviations from the Beveridge curve into different parts, using data on job openings from the Job Openings and Labor Turnover Survey (JOLTS) and unemployed persons by reason of unemployment obtained from the Current Population Survey (CPS). The findings will put an upward bound on the extent to which the increase in unemployment relative to job openings is due to reduced search effort caused by the extended availability of unemployment insurance.

While a plot of the Beveridge curve beginning in January 2001 clearly shows a stable, downward-sloping relationship between job openings and unemployment rates up to August 2009, the deviation of the points starting in September of 2009 from the stable Beveridge curve has been attributed by many economists to factors such as a rise in the mismatch between the skills of the unemployed and the skills desired by employers (it is standard in the literature to interpret movements along the Beveridge curve as cyclical movements in labor demand, and to interpret shifts in the Beveridge curve as indicative of shifts in the efficiency of job-worker matching) or to the supplemental and extended UI benefit programs that were designed to attenuate the hardships of involuntary job losses over the course of the Great Recession.

Key Findings

- A rough calculation suggests that job leavers, new entrants, and unemployed re-entrants— most of whom are not eligible for unemployment benefits—constituted approximately

48.5 percent of the aggregate gap in January 2013, while job losers accounted for the remaining part during the same month.

- While the vacancy and unemployment relationship appears to have shifted outward for job losers and unemployed entrants, exploring the relationship of each group across different age cohorts reveals that most of the shift among job losers is concentrated among persons over 44 years of age. When the job openings rate was plotted against job losers as a percentage of the total labor force for the age ranges of 16–19, 20–24, 25–34, and 35–44 years, there was little or no change in the historical Beveridge curve relationship. In contrast, exploring the relationship across different age groups using new labor market entrants and unemployed re-entrants reveals an outward shift among all categories.

- Job losers younger than 45 years of age appear to have benefitted more than the older cohorts from the increase in job openings over the recent period.

Implications

This brief uncovers new facts that emerge from disaggregating the unemployment rate into different categories based on the reason for unemployment. The findings suggest that up to half of the increase in the U.S. unemployment rate relative to the fitted Beveridge curve is explained by job leavers, new entrants, and re-entrants—those who are ineligible to collect unemployment benefits.

Up to half the increase in the U.S. unemployment rate relative to the fitted Beveridge curve is explained by job leavers, new entrants, and re-entrants—those who are ineligible to collect unemployment benefits.

Because unemployed job seekers who do not qualify to receive benefits compete for jobs with unemployed job losers who are eligible to collect UI, an unattractive vacancy that is refused by a job loser is likely be grabbed quickly by a new entrant or unemployed re-entrant who is not subject to any incentive effects. However, the evidence from the decompositions suggests that the increase in the unemployment rate relative to job openings will persist when unemployment benefit programs expire.

About the Author
Rand Ghayad is a visiting fellow at the Federal Reserve Bank of Boston and a Ph.D. candidate at Northeastern University.

....

Research Data Report

....

d-13-1

Merchant Steering of Consumer Payment Choice: Lessons Learned from Consumer Surveys

by Oz Shy and Joanna Stavins

abstract and complete text: http://www.bostonfed.org/economic/rdr/2013/rdr1301.htm
e-mail: oz.shy@bos.frb.org, joanna.stavins@bos.frb.org

Motivation for the Research

Until recently, credit card networks prohibited merchants in the United States from giving price discounts on debit card purchases or surcharging consumers on any card transactions.

Recent U.S. legislation and court settlements have allowed merchants to offer price discounts intended to steer customers toward payment methods that merchants prefer—typically payment instruments that are less costly for them to accept. Steering may also involve disclosing merchant fees, for example by posting notices at the store or online. Realizing that the new policies could change the way merchants interact with their customers, either by differentiating their prices based on payment method or by introducing other incentives, the authors attempted to discover the extent to which merchants implemented steering practices in the wake of the regulatory changes.

Anticipating these legislative and regulatory changes, the authors included three questions in the 2010 and 2011 pilots of the Diary of Consumer Payment Choice (DCPC), a collaborative effort of the Federal Reserve Banks of Boston, Richmond, and San Francisco that is administered by the RAND Corporation. The questions were intended to learn whether buyers were being steered by merchants toward using payment methods that are less costly to merchants but may not be otherwise selected by the customers. The initial attempts to use the diary survey to extract information from consumers about whether merchants were influencing their payment choices were not very successful. Some respondents were confused by some of the questions and provided inconsistent answers. In other cases, respondents interpreted the questions in ways that differed from what the authors had intended. In this data report the authors analyze some preliminary findings on the frequency and direction of merchants' attempts to influence consumer payment choices as reported by respondents to the pilot diary surveys. The authors then discuss other potential ways to measure the extent and effect of merchant steering and price discounting on the use of a particular payment instrument and propose alternative survey questions that might generate more consistent responses than the ones they obtained from the 2010 and 2011 DCPC pilots.

Research Approach

The authors used the DCPC as a vehicle to attempt to understand the extent of merchant steering of consumer payment choice. The DCPC provides more detailed, transaction-specific information than the broader Survey of Consumer Payment Choice (SCPC; see Foster, Meijer, Schuh, and Zabek 2011). Because the diary respondents also fill out the SCPC, the authors have access to a wide array of information about them. The DCPC is representative of U.S. consumers, but the sample, collected over three days, is small: 353 respondents in 2010 and 389 in 2011, and the standard errors are relatively large. (Subsequently, the 2012 DCPC, which includes revised questions about price discounts and surcharges, was administered to a full sample of over 2,500 respondents.)

The DCPC is a consumer survey, in which the aim was to solicit information from consumers about their payment choices, and not a merchant survey, in which the perspective of respondents might be different. Respondents to the DCPC were not expected to understand or even to be aware of the recent policy changes—they were simply asked to record their experiences while conducting transactions. The three questions the authors added to the DCPC asking about each transaction were as follows:

- Question 1: Did the merchant accept the payment method you most preferred to use for this purchase? (If yes, please leave blank. Otherwise, please indicate the payment method you most preferred to use.)

- Question 2: Did the merchant try to influence your choice of payment method by offering discounts or incentive programs, posting signs, or refusing to accept certain payment methods? (Please circle Y for yes or N for no.)

- Question 3: Did the merchant give you a discount on your purchase for the payment method you used? (Please circle Y for yes or N for no.)

The objectives of the three questions were as follows:

- Objective of Question 1: This question was intended to assess potential steering of consumer payment choices by merchant acceptance decisions. If the merchant accepts the buyer's most preferred payment instrument, but the buyer pays with a different instrument, one might infer that the merchant was able to influence the buyer's payment choice.

- Objective of Question 2: This question was intended to understand consumers' perspectives on whether merchants actively attempted to influence buyers by steering them toward less costly payment methods, and ultimately to measure directly which merchants try to steer their customers and how often. Note that a "Yes" answer implies only that the merchant tried to influence the customer's choice of payment, not necessarily that the attempt was successful.

- Objective of Question 3: This question asked about a very specific method of steering, namely about providing price discounts for using less-costly payment instruments.

In order to understand how the respondents had assessed the diary questions, the Federal Reserve Bank of San Francisco and the RAND Corporation commissioned cognitive interviews conducted with a subsample of the diary respondents. Cognitive interviews are sometimes used to improve survey instruments. In this case, the cognitive interviews were conducted by a Carnegie Mellon University professor who specializes in behavioral decision-making and survey methodology. The main goal of the cognitive interviews was to identify potential confusion about the payment diary and misinterpretation of the questions, and to solicit respondents' feedback about the clarity of instructions, questions, and categories of payment methods and merchants. Interviewees were also asked whether the diary booklet had provided a good memory aid to assist in accurately recording transactions. Because the 2010 and 2011 diaries were pilot surveys, the authors used the feedback received from these interviews to improve future designs of the diary.

Because the initial goal of the research discussed in this report was to gain an understanding of merchant steering practices of consumer payments, the authors made some adjustments to the survey data to attempt to reconcile contradictory results in some of the survey responses. These adjustments are described in the report.

Key Findings
- The interviews revealed that respondents had interpreted the three questions in a variety of ways. For example, some respondents had answered "Yes" to Question 1 if a merchant did not accept Discover cards but accepted other cards or "Yes" to Question 2 if a merchant offered a discount for using a store-branded credit card but not for using other cards. These examples indicate that at least some respondents had interpreted the questions as referring to specific types of payment cards, rather than to the entire category of credit cards, as the question had intended.

- Steering was not widely reported by the survey respondents and few price discounts were noted. Both steering and price discounts are costly to merchants (Briglevics and Shy 2012) and may be confusing to consumers. After adjusting the survey responses to the three questions about customers' perceptions of merchant steering, the authors found the following:

  – In over 96 percent of transactions, merchants accepted the payment method preferred by the buyer.

  – Merchants attempted to steer buyers in about 6–7 percent of the recorded transactions.

  – Merchants gave discounts on the payment instrument used in approximately 3–4 percent of the transactions, but the authors believe that discounts based on the choice of payment method were offered on only 2 percent of the transactions.

Implications

The authors found problems with the way the diary survey questions were formulated and evidence that the questions were interpreted differently by different people. While there are possible explanations for the responses received, it is impossible to confirm or reject the validity of these responses based on the available data. In order to assess the extent of merchant steering or price discounting based on payment method—and therefore to assess whether the policies that relaxed restrictions on merchants were implemented in practice—a better survey method must be applied. Survey methodology literature provides some help in how to ask survey questions (for example, Groves et al. 2009; Fowler 1995).

In addition to consumer surveys, other options for measuring the effects of policy changes should be evaluated. One option under consideration is conducting consumer focus groups or cognitive interviews, although the high cost of developing and administering such tools would likely result in small sample sizes. Because the DCPC was designed to be a representative survey of U.S. consumers but not necessarily a representative survey of U.S. merchants, developing surveys of merchants should also be considered. However, merchant surveys present challenges that must be carefully evaluated. For example, small merchants might be afraid to disclose their steering methods. In particular, small merchants who already impose surcharges on credit card transactions may not admit to that practice.

About the Authors

Oz Shy is a senior economist and Joanna Stavins is a senior economist and policy advisor; both are members of the Consumer Payments Research Center in the research department at the Federal Reserve Bank of Boston.

r-13-1

The Quest for Cost-Efficient Local Government in New England: What Role for Regionalization?

by Yolanda K. Kodrzycki

complete text: http://www.bostonfed.org/economic/neppc/researchreports/2013/rr1301.htm
e-mail: yolanda.kodrzycki@bos.frb.org

Motivation for the Research

Local governments constitute an important sector of the U.S. economy. Collectively, spending in 2007 by the nation's roughly 89,000 local governments (cities, towns, counties, independent school districts, and special districts) totaled $1.5 trillion, approximately 11 percent of U.S. GDP.

The Great Recession and its aftermath have made it more difficult for localities to maintain this level of spending. Budget shortfalls have led many states to cut aid to local governments, and falling property values have constrained local own-source revenues in many parts of the nation. As a result, local governments have been forced to enact a range of cost-cutting measures, including reductions in services, staffing, and employee compensation.

Revenues to fund local government operations are expected to remain constrained for the foreseeable future. As the federal government takes steps to bring its budget closer to balance, it is likely to pare back discretionary grants to state and local governments. In turn, state and local governments are likely to face continued pressures to pre-fund employee retirement benefits, possibly at the expense of other budget items that are arguably more discretionary. Thus, policymakers at all levels may find themselves re-examining cost-cutting options that once seemed unpalatable, including reorganizing service responsibilities across geographic or political boundaries.

Motivated by the prospect of continuing strain on local government finances, this study examines the extent to which a move to provide local government services at the regional rather than the local level could potentially reduce costs. It focuses especially on the expected long-term savings in the New England states, providing specific numerical estimates for Massachusetts and Connecticut. Where possible, the study also addresses the effects of regionalization on service quality, and indicates whether the available evidence on quality reinforces or mitigates the results based on costs alone.

Research Approach

Recognizing that local control has deep historical roots in New England, this study focuses on mechanisms that allow localities to continue to exist as distinct units but that take advantage of economies of scale by transferring responsibilities for specific municipal services to a consolidated government organization or a consortium of local governments. One such mechanism is the intermunicipal partnership, sometimes referred to as intergovernmental (or interlocal) cooperation. Under this form of regional consolidation, a locality enters into a formal agreement to provide certain public services jointly with one or more other localities.

Projected Potential Savings from Public Safety Answering Point (PSAP)
Consolidation in Massachusetts and Connecticut
### Scenarios Based on 2010 Michigan Data

Panel A. Michigan

Expenditure Per Call at Given Call Volumes (Dollars)

Panel B. Massachusetts

Expenditure Per Call at Given Call Volumes (Dollars)

Panel C. Connecticut

Expenditure Per Call at Given Call Volumes (Dollars)
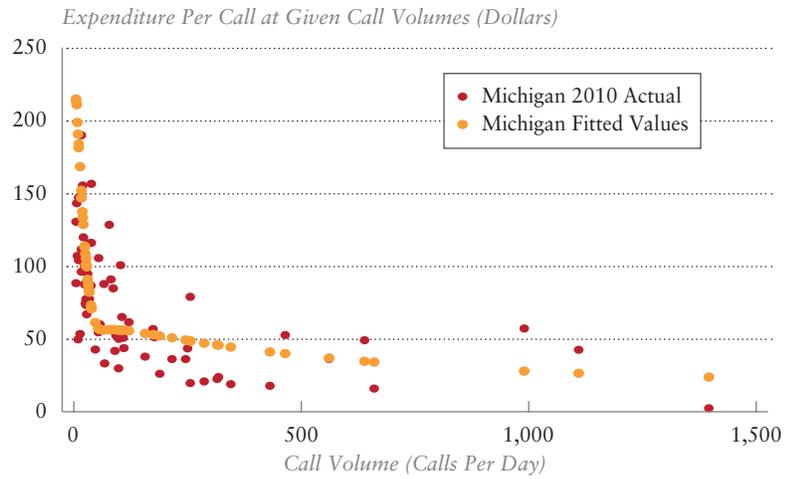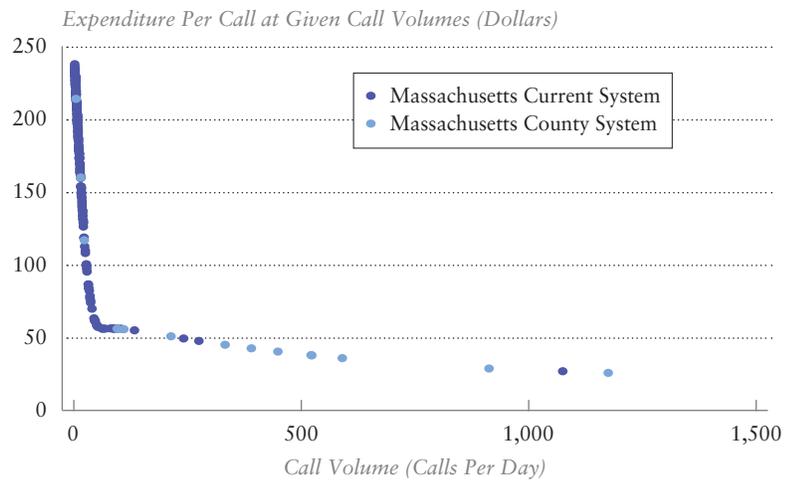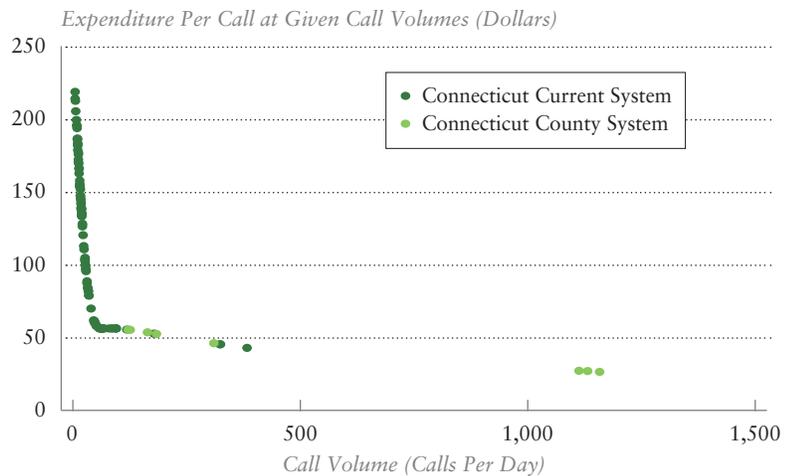
Source: Author's calculations based on data from 2011 Annual Report to the
State Legislature produced by the Michigan State 911 Committee.

Another mechanism is to centralize responsibility for designated municipal services in an existing regional (or state) authority or government. While full-scale mergers of local governments have remained extremely rare, intergovernmental cooperation and service sharing appear to be on the rise.

Local governments perform many functions, including ensuring public safety, maintaining roads, collecting trash, and educating children. As policymakers consider regionalizing the services currently provided by cities and towns (but not combining cities and towns into larger units), they need empirical evidence on the merits of consolidation at the service category level. This includes information on the scale at which government services are currently provided, which services could be provided more effectively at a larger scale, and how large the associated cost savings or quality improvements are likely to be. Finally, once policymakers have formulated their regionalization priorities, they will likely want to consider alternative mechanisms by which to achieve their objectives.

The study begins by summarizing the evidence on regional consolidation of public services from individual case studies and broader research. Based on these summarized findings, the remainder of the study focuses on three services for which the arguments favoring regionalization are particularly strong and the available data allow analysis of the likely savings associated with specific consolidation scenarios. These services are emergency call handling and dispatch, public health services, and public pension plan administration.

To estimate the potential long-term savings from regional consolidation for these services, the author first compares the degree to which the provision of these services is fragmented in each of the New England states as compared with the rest of the nation. She notes that an understanding of the potential benefits from regional consolidation requires specific information about how each service is provided—not just a tally of the number of local governments—because there is not a one-to-one correspondence between service units and local governments. The author applies regression analysis to actual data from other states to gauge how much Massachusetts and Connecticut governments could save by consolidating service provision for each of the three services. She estimates cost functions for each service category, based on the available nationwide data on expenditures, scale, and additional information that affects how much is spent per unit of service. The resulting shape of the cost curve shows the range over which the economies of scale are most pronounced, enabling policymakers to identify service units that are inefficiently small. The author uses these estimates to examine the potential cost savings in Massachusetts and Connecticut from consolidating provision of each of the three services. These two states are distinguished by their relatively high numbers of service units, so the computed percentage savings can serve as upper bounds for the remaining New England states.

The study also discusses policies that other states have used to promote regionalization or consolidation of these services, including direct mandates and financial incentives, contrasting these policies with policies currently in place in New England. Although the direct evidence focuses on three specific service types, states may be able to accelerate regionalization of additional local services using similar tools.

Key Findings
- Evidence from the existing literature indicates that many services can be provided as cost effectively by smaller units as by larger units of government. However, some services exhibit economies of scale, indicating that local governments may be able to achieve savings

through regionalized service delivery. Moreover, for a subset of these services, examples of successful regionalization are available to guide cities and towns that continue to provide these same services locally. For a limited number of services, there is also evidence that regionalization would likely lead to improvements in service quality.

- It appears that regional consolidation efforts in New England should target the roughly 20 percent of local government spending that is characterized by demonstrated economies of scale, in situations where loss of local control does not seriously compromise service quality. While 20 percent or so appears to be an appropriate upper bound for the region as a whole, the portion of local budgets that may be amenable to some form of consolidation across city and town borders likely varies both across and within states, depending, in part, on variation in spending allocations.

- Although the services studied in this report tend to be delegated to local governments or authorities, in fact the New England states differ in the degree to which service areas cross geographic or political boundaries. For all three functions Maine has extensive service sharing and centralization. The remaining two northern New England states (New Hampshire and Vermont) tend to have more service sharing and centralization than the southern New England states (Connecticut, Massachusetts, and Rhode Island). The major exception is that Rhode Island has only one public health department serving the entire state.

- The estimates in this study indicate that the potential cost savings from consolidation vary, depending on the service and the hypothetical consolidation scenario considered. Regional consolidation of emergency call handling and dispatch yields the most cost reduction—over 50 percent in the scenarios considered for both Massachusetts and Connecticut. Moving to larger-scale public health departments offers somewhat smaller but still substantial cost reduction for these two states. Consolidating the administration of public pension plans would bring about much larger percentage savings in Massachusetts than in Connecticut, owing to the greater existing degree of fragmentation in Massachusetts. While these are the statewide conclusions for two of the six New England states, the framework suggests that there may be smaller areas within each state that could achieve substantial percentage cost reduction from regionalization. The study finds evidence confirming that these types of services should be prioritized for regional consolidation.

Implications

On the one hand, New England is a good target for regional consolidation efforts. Many local governments in New England serve small populations or land areas. On the other hand, agreeing on how to coordinate the delivery of specific public services is complicated and cannot be accomplished as a "quick fix" in the midst of a budget crisis. Consolidating services across jurisdictions offers the potential for saving costs in the long run and should be considered seriously if the alternative is permanent reductions in the scope or quality of public services provided by cities and towns. And while consolidation cannot be completed quickly, a local fiscal crisis, particularly one that is serious enough to prompt state intervention, can serve as a catalyst.

While the methodology underlying these estimates undoubtedly leaves out many of the details that would be needed to examine specific cases, it at least indicates that local control for 911 call handling and dispatch, public health, and some administrative and financial functions

comes at a nontrivial cost to taxpayers. (In the case of public pensions, the higher costs may fall largely on plan participants rather than taxpayers.)

In summary, the study comes to three broad conclusions. First, policymakers should not expect regionalization to offer immediate major relief from the budgetary stresses that many local governments are experiencing. Rather, policymakers should consider regional consolidation in addition to other measures that could bring local budgets into structural balance over the medium to long term. Second, based on both cost and quality considerations, a strong case can be made for sharing or centralizing some services that are currently provided mostly at the local level throughout much of New England, particularly in the three southern states. Third, in states with fragmented public service provision, state legislatures could encourage further regionalization by adopting stronger and more targeted regulations and fiscal incentives. Such measures would likely result in accelerated regionalization, compared with the situation in which local governments pursue intermunicipal partnerships and service sharing without these types of intervention.

About the Author

Yolanda K. Kodrzycki is a vice president and the director of the New England Public Policy Center in the research department of the Federal Reserve Bank of Boston. The Policy Center conducts research on key economic and policy issues in New England and engages with regional partners in advancing identified policy options.

FEDERAL RESERVE
BANK OF BOSTON™

Research Department
Federal Reserve Bank of Boston
600 Atlantic Avenue
Boston, MA 02210
www.bostonfed.org/economic/index.htm