

## Identification in Macroeconomics

Emi Nakamura and Jón Steinsson

**A**ny scientific enterprise needs to be grounded in solid empirical knowledge about the phenomenon in question. Many of the main empirical questions in macroeconomics are the same as they have been since at least the Great Depression. What are the sources of business cycle fluctuations? How does monetary policy affect the economy? How does fiscal policy affect the economy? Why do some countries grow faster than others? Those new to our field may be tempted to ask, “How can it be that after all this time we don’t know the answers to these questions?”

The reason is that identification in macroeconomics is difficult. Take the case of monetary policy. Unfortunately for us as empirical scientists, the Federal Reserve does not randomize when setting interest rates. Quite to the contrary, the Federal Reserve employs hundreds of PhD economists to pore over every bit of data about the economy so as to make monetary policy as endogenous as it possibly can be. This fact means that quite a bit of ingenuity and careful research is required to identify a component of monetary policy that is plausibly exogenous to future output and can, thus, be used to measure directly the effects of policy on output.

An important strand of empirical work in macroeconomics attempts the challenging task of identifying plausibly exogenous variation in macroeconomic policy and using this variation to assess the effects of the policy. We refer to this type of work as direct causal inference. Later in this article, we provide a critical assessment of several of the main methods that have been used to assess the effects of macroeconomic policies

■ *Emi Nakamura and Jón Steinsson are both Chancellor’s Professors of Economics at the University of California at Berkeley, Berkeley, California. Their email addresses are [enakamura@berkeley.edu](mailto:enakamura@berkeley.edu) and [jsteinsson@berkeley.edu](mailto:jsteinsson@berkeley.edu).*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.32.3.59>

doi=10.1257/jep.32.3.59

in the academic literature—including identified vector autoregression. We do this in the context of asking what the best available evidence is on monetary nonneutrality.

A serious challenge faced by researchers attempting direct causal inference in macroeconomics is that the natural experiments we can find in the data are rarely exactly the experiments we would need to answer the policy questions in which we are interested. This “external validity” problem can be illustrated by thinking about monetary and fiscal policy. One issue is that the dynamic nature of monetary and fiscal policy makes these policies very high dimensional. Some monetary policy announcements only affect expectations about policy in the very short run (for example, whether the Federal Reserve will tighten this month or next month), while others affect policy expectations both in the short run and longer run, and still others only affect expectations about policy several years in the future (for example, when the policy interest rate is at the zero lower bound and the Fed makes a commitment to keep it there for longer than previously expected). The same is true of fiscal policy. The recent theoretical literature has emphasized that the future time profile of a policy action greatly affects its impact on current output and inflation. Identifying the effects of a policy shock with one time profile, therefore, does not necessarily identify the effects of a policy shock with a different time profile. A second external validity issue is that the effects of fiscal shocks depend on the response of monetary policy (for example, whether it is constrained by the zero lower bound). The effects of monetary policy, of course, also depend on the response of fiscal policies. A third issue is that the effects of monetary and fiscal policy may differ depending on the level of slack in the economy and how open the economy is. A fourth issue is that the degree to which a policy action is a surprise can affect both how strongly and when the economy reacts to it.

These external validity issues (and others) mean that even very cleanly identified monetary and fiscal natural experiments give us, at best, only a partial assessment of how *future* monetary and fiscal policy actions—which may differ in important ways from those in the past—will affect the economy. One response to these issues is to gather direct causal evidence about each and every different case. This however, may not be feasible. Even if it is feasible, it seems that one should be able to learn something about one case from evidence on another.

Due to the challenges described above, much empirical work in macroeconomics is more structural in nature. Such work often takes the form of researchers focusing on a set of moments in the data and arguing that these moments can discriminate between different models of how the economy works. Estimates of causal effects (that is, the response to identified structural shocks) can play an important role in this type of inference. The causal effects estimates can be viewed as target moments that models should match in much in the same way as unconditional means, variances, and covariance. We will use the term “identified moments” as a shorthand for “estimates of responses to identified structural shocks”—what applied microeconomists would call “causal effects.”<sup>1</sup>

<sup>1</sup>The term “identified moments” may seem odd to some. In econometrics, it is parameters that are identified, not moments. We use the term “moment” in a broad sense to refer to a target statistic that

We argue that identified moments are often particularly informative moments for distinguishing between important classes of macroeconomic models. As a first example, consider the case of distinguishing between real business cycle and New Keynesian models—which is important for many policy questions. One approach is to use full information structural estimation methods such as maximum likelihood or Bayesian methods. Another approach is to match unconditional means, variances, and covariances in the tradition of the real business cycle calibration literature. In contrast, Rotemberg and Woodford (1997) and Christiano, Eichenbaum, and Evans (2005) estimate the response of output and inflation to identified monetary policy shocks and use these responses to discriminate between different business cycle models. Similarly, Galí (1999) and Basu, Fernald, and Kimball (2006) use the response of output and hours to identified productivity shocks to distinguish between models.

A second example is the recent literature on regional fiscal multipliers, and more generally the growing literature that aims to shed light on macroeconomic questions using cross-sectional identification strategies. Estimates of the regional government spending multiplier do not *directly* answer the policy question macroeconomists are most interested in—the effect of fiscal stimulus at the national level. However, the regional fiscal multiplier turns out to have a great deal of power in distinguishing between different models of the business cycle (Nakamura and Steinsson 2014). Models that can match a large regional multiplier typically imply that output responds strongly to demand shocks. In these models the aggregate multiplier is large when monetary policy is accommodative (for example, at the zero lower bound). For this reason, the recent literature on the regional fiscal multiplier has been able to provide powerful indirect evidence on the effectiveness of aggregate fiscal stimulus.

A third example is the use of estimates of the marginal propensity to consume (MPC) from a transitory fiscal rebate to distinguish between competing models of consumption dynamics. One approach uses truly random variation in the timing of fiscal stimulus checks to estimate a quarterly MPC of roughly 0.25 for nondurable consumption (for example, Johnson, Parker, Souleles 2006; Parker, Souleles, Johnson, and McClelland 2013). Kaplan and Violante (2014) use these estimates to distinguish between competing models of consumption dynamics. Their favored model adds illiquid assets that earn high returns to an otherwise standard model with uninsurable income risk and borrowing constraints. Angeletos, Laibson, Rebetto, Tobacman, and Weinberg (2001) argue that models in which households face self-control problems can help match the estimated MPC in the data.

The identifying assumptions that identified moments rely on are typically controversial. So, why use such moments? What is the upside of this approach relative to, for example, targeting simple unconditional moments? An important advantage

---

a researcher wants the model to match. We use the term “identified” because the target statistics we have in mind are estimates of causal effect parameters—or what macroeconomists would call estimated responses to “identified structural shocks”—as opposed to simple unconditional moments such as means, variances, and covariances.

is that in some cases the ability of a model to match identified moments may depend primarily on a particular sub-block of the model and be relatively insensitive to model misspecification in other sub-blocks of the model.<sup>2</sup> In the examples above, one does not, for example, have to take a stand on whether business cycles are driven by demand or supply shocks. In contrast, inference based on unconditional moments is typically highly sensitive to this. Moreover, Kaplan and Violante's (2014) inference based on the marginal propensity to consume seems unlikely to depend heavily on which frictions are included in the model outside of the consumption block, for example, price or wage rigidity, or investment and labor adjustment costs.

## The Power of Portable Statistics

One important innovation of the early real business cycle literature was a move away from using likelihood-based empirical methods towards empirical evaluation based on matching moments (Kydland and Prescott 1982; Prescott 1986). An advantage of this approach is that it leads to the creation of “portable statistics” that can be used over and over again by researchers to discipline and test models. This allows for a division of labor and a fruitful back-and-forth between the theoretical and empirical parts of the field (and other fields). The equity premium is a good example of a portable statistic. Mehra and Prescott (1985) consider whether the equity premium is consistent with one specific class of models. A generation of subsequent researchers has then used this same statistic to evaluate a host of new models. The result has been an enormously influential literature on the nature of risk and risk aversion.<sup>3</sup>

It is useful to distinguish several types of moments that have been influential in empirical macroeconomics. We first consider the distinction between “micro moments” and “macro moments,” and then the distinction between what we call “identified moments” and simpler moments.

### Micro and Macro Moments

Micro moments are constructed using microeconomic data on the behavior of individuals and firms. A prominent example is the frequency of price change and related statistics on price rigidity (Bils and Klenow 2004; Nakamura and Steinsson 2008; Klenow and Kryvtsov 2008). These statistics help to discipline models that

<sup>2</sup>Our argument for moment matching using identified moments is related to Chetty's (2009) argument for sufficient statistics. However, we argue that identified moments can help answer many policy questions by distinguishing between models—that is, by learning about deep structural parameters. Chetty's emphasis is on combining identified moments into a formula (a “sufficient statistic”) that answers a particular policy question.

<sup>3</sup>What is an example of a statistic that is not portable? The score of the likelihood function of a particular model (the moment that maximum likelihood estimation seeks to match) is very informative about that particular model. But it is not a very intuitive statistic to use to evaluate other models and is rarely (if ever) reused as a moment that other researchers (with other models) seek to match.

are designed to understand the effects of monetary policy. Another prominent example is the change in time spent shopping as well as the quantity and quality of food intake at the time of retirement (Aguiar and Hurst 2005). These statistics help distinguish between competing life-cycle models of household consumption and savings behavior.

Macro moments use aggregated data to identify equilibrium outcomes that are informative about what type of world we live in. The equity premium is an example of a highly influential macro moment, as are facts about changes in real wages and hours worked per person over the past century. The fact that real wages have risen by a large amount while hours worked have been stable or fallen slightly strongly rejects models without income effects on labor supply and, in fact, suggests that income effects are slightly larger than substitution effects in the long run. This motivates the use of “balanced growth preferences” in macroeconomic models (King, Plosser, and Rebelo 1988; Boppart and Krusell 2016).

There is a rich tradition in macroeconomics of using simple micro and macro moments to make inferences about how the world works. In many cases, these types of statistics can yield powerful inference. Prominent examples in addition to those discussed above include the real business cycle literature (Kydland and Prescott 1982; King and Rebelo 1999), the Shimer (2005) puzzle literature, the misallocation literature (Hsieh and Klenow 2009), the literature on exchange rate disconnect (Meese and Rogoff 1983; Itskhoki and Mukhin 2017), and the literature on “wedges” (Chari, Kehoe, and McGrattan 2008; Shimer 2009).

### **Causal Effects as Identified Moments**

Here, we contrast simple statistics such as means, variances, and covariances with more complex statistics derived from empirical strategies designed to uncover what applied microeconomists would call causal effects, but macroeconomists would call responses to structural shocks. The last quarter century has seen a “revolution of identification” in many applied fields of economics (Angrist and Pischke 2010). This revolution has increased emphasis on identifying causal effects using credible research designs based on the use of instrumental variables, difference-in-difference analysis, regression discontinuities, and randomized controlled trials. It is these types of causal effects estimates that we refer to as identified moments.

In some cases, there is a one-to-one mapping between identified moments and a deep structural parameter. For example, there is a large literature in labor economics that estimates the labor supply elasticity (Chetty 2012; Chetty, Guren, Manoli, and Weber 2013). Macroeconomists have long made use of causal effects estimates of this kind to discipline the models that they work with. In the jargon of macroeconomics, we frequently “calibrate” certain parameters of our models (such as the labor supply elasticity) based on external estimates.

Many identified moments, however, do not correspond directly to a deep structural parameter. Two prominent examples we discussed earlier are estimates of the marginal propensity to consume out of a transitory fiscal rebate and estimates of the regional fiscal multiplier. In these cases, a theoretical framework is required to

go from the identified moment to the macroeconomic questions of interest. These types of identified moments are valuable because they can be used as empirical targets in a structural moment-matching exercise aimed at distinguishing between competing models that differ in their implications about the macroeconomic question of interest.<sup>4</sup>

There is a prevalent view in macroeconomics that if your empirical strategy is to calculate the same moment in real-world data as in data from a set of models, you might as well focus on very simple unconditional moments. A moment is a moment, the argument goes. But relative to simple unconditional moments, the advantage of identified moments is that they can provide evidence on specific causal mechanisms of a model and may be relatively invariant to other model features.

Consider the recent debate on the role of changes in house prices in causing the Great Recession of 2007–2009. Mian, Rao, and Sufi (2013) and Mian and Sufi (2014) compare changes in consumption and employment in metropolitan areas that experienced larger or smaller house price changes. Of course, causation may run both ways: increases in house prices may stimulate economic activity, but a local boom may also increase house prices. To isolate the causal effect of house prices on consumption and employment, these authors propose to instrument for changes in house prices with estimates of housing supply elasticities constructed by Saiz (2010), which in turn are constructed from data on regional topology and land-use regulation. The idea for identification is that national shocks will lead house prices to increase more in metropolitan areas where housing supply is less elastic.<sup>5</sup> Using this identification strategy, Mian, Rao, and Sufi (2013) find that the elasticity of consumption with respect to housing net worth is 0.6 to 0.8, while Mian and Sufi (2014) find that the elasticity of nontradable employment with housing net worth is 0.37.

These estimates do not directly answer the macroeconomic question of how much aggregate house prices affect economic activity because this empirical approach is based on comparing one metropolitan area to another and therefore “differences out” aggregate general equilibrium effects. However, these estimates are quite informative about the “consumption block” of macroeconomic models. They strongly reject simple complete-markets models of consumption such as the influential model of Sinai and Souleles (2005)—in which the elasticity of consumption to house prices is zero—in favor of models with life-cycle effects, uninsurable income risk, borrowing constraints, and in which households can substitute away from housing when its price rises (for a discussion of such a model, see Berger, Guerrieri, Lorenzoni, and Vara 2018).

<sup>4</sup>Formally, the idea is to use indirect inference with limited-information empirical models designed to estimate causal effects (for example, an instrumental variables regression, a difference-in-difference design, or a regression-discontinuity design) as auxiliary models. See Smith (2008) for an introduction to indirect inference.

<sup>5</sup>The housing supply elasticity is obviously not randomly assigned (for example, land availability is correlated with whether a city is on the coast). However, the argument for identification is that whatever makes these coastal (and otherwise land-constrained) locations different does not affect their response to aggregate shocks directly, except through the implications for the housing market.

The identification strategy used by Mian, Rao, and Sufi (2013) and Mian and Sufi (2014) is by no means uncontroversial, as is often the case with identified moments. However, focusing instead on simple moments like the raw correlation between house prices and consumption or house prices and employment has downsides too. These simple moments are likely to be sensitive to assumptions that have little to do with consumption behavior, such as what shocks drive the business cycle, and to the strength of general equilibrium effects, such as the response of prices and wages to demand shocks (which, in turn, may depend on virtually everything about how the model works). In other words, using these simple moments results in a joint test of all the parameters and assumptions in the model, while identified moments can focus in on the consumption block of the model and provide inference that is robust to the specification of other parts of the model.<sup>6</sup>

Another prominent example of an identified moment matching exercise is recent work that seeks to determine the role of unemployment insurance extensions in delaying recovery from the Great Recession. Hagedorn, Manovskii, and Mitman (2015) and Chodorow-Reich, Cogleanese, and Karabarbounis (2017) estimate the response of unemployment to variation in unemployment benefit extensions across states using discontinuity-based identification and an instrumental variables approach, respectively. They then use these cross-sectional identified moments to determine key parameters in a labor market search model and use the resulting model to determine how unemployment insurance extensions will affect the economy at the aggregate level. Here, again, it is likely possible to pin down these same parameters using simple moments (such as the variance and covariance of unemployment, vacancies, wages, and benefit extensions) in a fully specified structural model. However, such an exercise would likely be sensitive to many auxiliary features of the model being used outside the “labor market block” of the model.

In the examples above, identified moments provide information primarily on a particular “block” or mechanism of a macroeconomic model. This “piecemeal” form of inference will, therefore, result in partial identification on the model space. It is inevitable that any given statistic or set of statistics will not be able to pick out a single model and reject all others. The fact that several models are consistent with a statistic is not grounds for rejecting the statistic as being uninteresting. We should instead think in reverse: If a statistic has power to reject an important set of models in favor of another set of models, the statistic is useful.

<sup>6</sup>Similar methods have been applied to assess the importance of financial frictions on firms. Catherine, Chaney, Huang, Sraer, and Thesmar (2017) use instrumental variable estimates of the response of firm investment to changes in real estate collateral to pin down parameters in a structural model with financial frictions and then use the model to quantify the aggregate effects of relaxing collateral constraints. Chodorow-Reich (2014) uses bank shocks to identify the effects of financial constraints on firm employment. In an appendix, he uses these micro-level estimates to pin down parameters in a general equilibrium model of the effect of bank shocks on the economy as a whole. Huber (2018) estimates direct and indirect firm effects as well as regional effects of a large bank shock in Germany. Other important papers in this literature include Peek and Rosengren (2000), Calomiris and Mason (2003), Ashcraft (2005), Greenstone, Mas, and Nguyen (2017), and Mondragon (2018).

## Aggregate versus Cross-Sectional Identification

The increased use of cross-sectional identification approaches has been an exciting development in empirical macroeconomics. In this work, researchers use geographically disaggregated panel datasets—often disaggregated to the level of the state or metropolitan statistical area—to identify novel causal effects. The use of regional data typically multiplies the number of data points available by an order of magnitude or more. It also allows for difference-in-difference identification and makes possible the use of a powerful class of instrumental variables: differential regional exposure to aggregate shocks. Prominent examples of this approach include: Mian and Sufi (2014) on the role of the housing net worth channel in the Great Recession; Autor, Dorn, and Hanson (2013) on the effects of Chinese imports on US employment; Beraja, Hurst, and Ospina (2016) on wage rigidity during the Great Recession; Martin and Philippon (2017) on the effects of debt during the Great Recession in the eurozone; Carvalho et al. (2016) on supply chain disruptions after the Great East Japan Earthquake of 2011; and a large literature on fiscal stimulus discussed below.

A key challenge for this literature is how to translate regional responses into aggregate responses. A common approach is to add up regional responses with the implicit assumption that the least affected region is unaffected by the shock and report this sum as the aggregate response. However, this approach ignores general equilibrium effects that influence the aggregate response but are absorbed by time fixed effects in the cross-sectional regressions used in this literature. As an example, Mian and Sufi show that the dramatic fall in house prices between 2006 and 2009 did not *differentially* affect tradables employment in areas with larger house price declines. However, this does not mean that this shock had no effect on tradables employment in the aggregate. (To be clear, they don't make any such claim.) Typically, regional responses can only be translated into aggregate responses through the lens of a fully specified general equilibrium model (Nakamura and Steinsson 2014).

A common critique of estimates based on cross-sectional identification in macroeconomics is that they don't answer the right question. While it is true that these estimates don't *directly* provide estimates of aggregate responses, they often provide a great deal of indirect evidence by helping researchers discriminate between different theoretical views of how the world works. In the language of the previous section, these cross-sectional estimates are examples of identified macro moments. They can be used as moments in a moment-matching exercise that is aimed at distinguishing between important classes of general equilibrium structural models of the economy that have different implications about the primary question of interest. This combination of theory and empirics can yield very powerful inference. The literature on the stimulative effects of government spending provides a nice case study to illustrate these ideas.

### Fiscal Stimulus: Aggregate Evidence

Direct aggregate evidence on the government spending multiplier is far from conclusive. This evidence largely comes in two forms: evidence from wars and from



vector autoregressions. Barro and Redlick (2011) regress changes in output on changes in defense spending and a few controls for a US sample period including several major wars. Their main conclusion is that the government purchases multiplier is between 0.6 and 0.7. Virtually all of the identification in their sample comes from World Wars I and II, and to a lesser extent the Korean War. When they restrict attention to data from after the Korean War, the confidence interval for their estimate includes all remotely plausible values.

Barro and Redlick (2011) assume that war-related defense spending is exogenous to output. Conceptually, this is like using wars as an instrument for government spending. The strength of this approach is that reverse causation is not likely to be a problem. World War I, World War II, and the Korean War did not happen because the US economy was in a recession or a boom. However, for war-related spending to be exogenous, wars must only affect output through spending. This is unlikely to be true. Barro and Redlick are aware of this issue and discuss some potential confounding factors. On one side, patriotism likely increased labor supply during major wars and thus results in an upward bias of the multiplier. On the other hand, wartime rationing and price controls likely result in a downward bias. Barro and Redlick argue that patriotism is likely the dominant bias, while Hall (2009) argues that the effects of wartime controls result in a net downward bias.

Blanchard and Perotti (2002) regress government spending on four quarterly lags of itself, taxes, output, a quadratic time trend, and a dummy variable for 1975:Q2. They view the residual from this regression as exogenous shocks to government spending.<sup>7</sup> They construct an impulse response function for output, consumption, and other variables to the government spending shocks by iterating forward a vector autoregression. Their baseline sample period is 1960–1997. The estimated response of output swings up, then down, then up again, with a peak response of output of 1.3 times the initial response of spending 15 quarters after the impulse.

Blanchard and Perotti's (2002) approach to identification makes the strong assumption that controlling for four lags of taxes, spending, and output eliminates all endogenous variation in spending. They argue that this approach to identification is more plausible for fiscal policy than for monetary policy, because output stabilization is not as dominant a concern of fiscal policy and because implementation lags are longer in fiscal policy. Ramey (2011) argues, however, that Blanchard and Perotti miss an important part of the response to fiscal shocks because news about fiscal shocks—especially those associated with major wars—arrives well ahead of the main increase in spending. Another concern is that estimates of the fiscal multiplier based on post–World War II aggregate data—such as Blanchard and Perotti's estimates—have such large standard errors that few interesting hypotheses can be rejected. Also, their estimates are highly sensitive to the sample period (like whether the Korean War is included) and to which controls are included (Galí, López-Salido, and Vallés 2007; Ramey 2016).

<sup>7</sup>This methodology is equivalent to performing a Cholesky decomposition of the reduced form errors from the vector autoregression with government spending ordered first.

Finally, aggregate estimates of the government spending multiplier are subject to an important external validity problem having to do with the response of monetary policy. Blanchard and Perotti's estimates come from a time period when monetary policy could "lean against the wind" by raising real interest rates in response to a government spending shock. The aggregate fiscal multiplier is potentially quite different in the midst of a deep recession when monetary is constrained by the zero lower bound on nominal interest rates. Because fiscal stimulus packages tend to be discussed in recessions, evidence on the government spending multiplier for this circumstance is particularly valuable. Unfortunately, direct aggregate evidence on the effectiveness of government spending when monetary policy is constrained is even less well established than on the simpler question of the average aggregate multiplier (Ramey and Zubairy 2018; Miyamoto, Nguyen, and Sergeev 2018).

An indirect way to infer the effectiveness of government spending when monetary policy is constrained is to amass evidence about whether a New Keynesian or neoclassical model is a better description of the world. In the New Keynesian model, the aggregate multiplier from government spending can be quite low if monetary policy is responsive, but when monetary policy is unresponsive—say, at the zero-lower-bound—the aggregate multiplier can be quite large. In contrast, in a neoclassical model, the aggregate multiplier is small independent of monetary policy. Unfortunately, evidence on the aggregate fiscal multiplier is not very helpful in this regard. Estimates between 0.5 and 1.0—which is where most of the more credible estimates based on US data lie—are consistent with both of these models. As we explain below, cross-sectional evidence on the effects of government spending yields much sharper inference on this point.

### **Fiscal Stimulus: Cross-Sectional Evidence**

In recent years, researchers have used a wide array of cross-sectional identification strategies to estimate the effects of government spending. Examples include windfall returns on state pension plans (Shoag 2015), differential state sensitivity to military buildups (Nakamura and Steinsson 2014), crackdowns on Mafia-infiltrated municipalities in Italy (Acconcia, Corsetti, and Simonelli 2014), formulas used to allocate spending across states from the American Recovery and Reinvestment Act (Chodorow-Reich, Feiveson, Liscow, and Woolston 2012; Wilson 2012; Dupor and Mehkari 2016), and spending discontinuities associated with decadal population estimate revisions (Suárez Serrato and Wingender 2016). Chodorow-Reich (2017) surveys this literature in detail.<sup>8</sup>

Estimates of the regional spending multiplier from this literature tend to cluster in the range of 1.5–2.0, which is substantially larger than typical estimates of the aggregate multiplier. However, these two sets of estimates are not necessarily inconsistent. After all, they measure different things. To understand this point, consider the identification strategy in Nakamura and Steinsson (2014). The basic

<sup>8</sup>Nekarda and Ramey (2011) consider variation in government spending across industries, as opposed to variation across regions as in the literature discussed above.

idea is that national military buildups (like the Carter–Reagan buildup following the Soviet invasion of Afghanistan in 1979) result in much larger changes in military spending in some states than others because the plants that build military hardware are unevenly distributed across the country. For example, national military buildups imply that spending rises much more in California than in Illinois. We then ask whether this translates into bigger increases in output in California than in Illinois. Our conclusion is that output in California rises by roughly \$1.5 relative to output in Illinois for each extra \$1 of spending in California relative to Illinois.

Importantly, our specification includes time fixed effects, which implies that our multiplier estimates are only identified off of the response of California relative to Illinois, not the response of all states to the aggregate buildup. This also means that aggregate general equilibrium effects are absorbed by the time fixed effects. These include any tightening of monetary policy that may occur as a consequence of the military buildup and the change in federal taxes needed to finance the buildup. Estimates of aggregate multipliers include these effects. Regional multiplier effects therefore do not provide direct evidence on the aggregate multiplier.

However, the regional multiplier provides a powerful diagnostic tool for distinguishing between competing macroeconomic models, and thereby for indirectly learning about the effectiveness of aggregate fiscal stimulus. In Nakamura and Steinsson (2014), we write down several multi-region business cycle models, simulate the same policy experiment in these models as we identify in the data, and compare the regional multipliers generated by each model with estimates of regional multipliers from real-world data. The textbook real business cycle model generates regional multipliers that are substantially smaller than our empirical estimate. In this model, a “foreign” demand shock (the federal government demanding military goods), leads people to cut back on work effort in other areas, which implies that the regional multiplier is less than one. We conclude that the regional multiplier evidence favors models in which output responds more strongly to a foreign demand shock than this model implies. We present an example of a Keynesian model in which the aggregate multiplier can be large (for example, when monetary policy is constrained at the zero lower bound). The regional multiplier does not uniquely identify a correct model (and no single statistic will). However, large regional multiplier estimates suggest that researchers should put more weight on models in which demand shocks can have large effects on output.

## **Monetary Policy: What Is the Best Evidence We Have?**

What is the most convincing evidence for monetary nonneutrality? When we ask prominent macroeconomists this question, the three most common answers have been: the evidence presented in Friedman and Schwartz (1963) regarding the role of monetary policy in the severity of the Great Depression; the Volcker disinflation of the early 1980s and accompanying twin recession; and the sharp break in the volatility of the US real exchange rate accompanying the breakdown of the Bretton

Woods system of fixed exchange rates in 1973 (first emphasized by Mussa 1986).<sup>9</sup> It is interesting that two of these pieces of evidence—the Great Depression and Volcker disinflation—are large historical events often cited without reference to modern econometric analysis, while the third is essentially an example of discontinuity-based identification. Conspicuous by its absence is any mention of evidence from vector autoregressions, even though such methods have dominated the empirical literature for quite some time. Clearly, there is a disconnect between what monetary economists find convincing and what many of them do in their own research.

### **Large Shocks**

The holy grail of empirical science is the controlled experiment. When it comes to monetary policy, for obvious reasons, we cannot do controlled experiments. We must instead search for “natural experiments”—that is, situations in which we can argue that the change in policy is large relative to potential confounding factors. Much empirical work takes the approach of seeking to control for confounding factors as well as possible. A different approach is to focus on large policy actions for which we can plausibly argue that confounding factors are drowned out. Such policy actions are, of course, rare. But looking over a long period and many countries, we may be able to piece together a body of persuasive evidence on the effects of monetary policy.

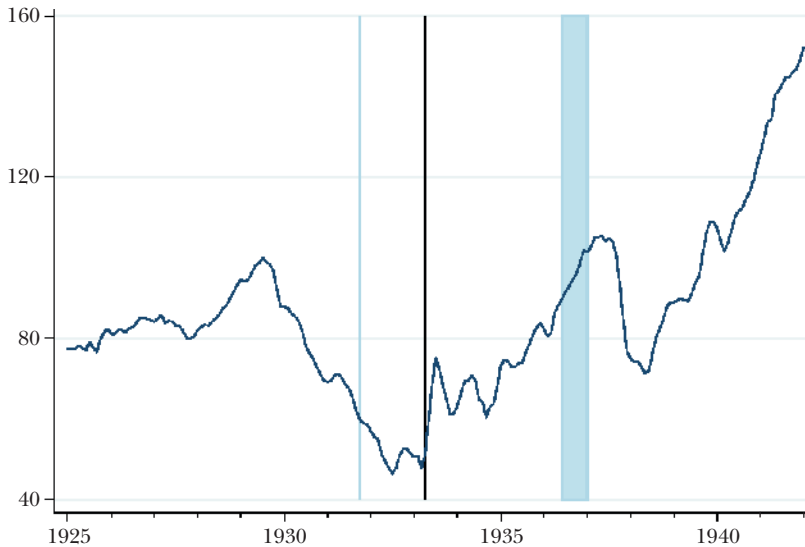
Friedman and Schwartz (1963, p. 688) argue in the final chapter of their monumental work on US monetary history that three policy actions taken by the Federal Reserve in the interwar period were 1) “of major magnitude” and 2) “cannot be regarded as necessary or inevitable economic consequences of contemporary changes in money income and prices.” They furthermore argue that “like the crucial experiments of the physical scientist, the results are so consistent and sharp as to leave little doubt about their interpretation.” The dates of these events are January–June 1920, October 1931, and July 1936–January 1937. We will focus on the latter two of these events, which occurred during the Great Depression.

Figure 1 plots the evolution of US industrial production from 1925 to 1942. The fall from the July 1929 peak to March 1933 was a staggering 53 percent. The recovery between March 1933 and the subsequent peak in May 1937 was rapid and large. But then a second very large downturn occurred. From May 1937 to May 1938, industrial production fell by 33 percent.

The light vertical bars in Figure 1 show the times of the two policy mistakes that Friedman and Schwartz (1963) highlight. In October 1931, the Federal Reserve raised the rediscount rate (the policy rate of that time) sharply from 1.5 to 3.5 percent in response to a speculative attack on the US dollar that followed Britain’s decision to leave the gold standard. The Fed drastically tightened policy despite the fact that industrial production was in free fall and a wave of bank failures was underway. At first pass, it may seem reasonable to interpret this as a clean

<sup>9</sup>Of course, a significant fraction say something along the lines: “I know in my bones that monetary policy has no effect on output.”

Figure 1

**Industrial Production from 1925 to 1942***(index equals 100 in July 1929)*

*Note:* The figure plots an index for industrial production in the US economy from January 1925 to January 1942. The index is equal to 100 in July 1929 (the peak month). The shaded bar and vertical line are the periods of policy mistakes identified by Friedman and Schwartz (1963). The dark vertical line is the time at which Roosevelt took the United States off the gold standard.

monetary shock. However, the subsequent fall in industrial production is not very different from the fall in the previous two years. It is not clear how much of the subsequent fall in industrial production is due to this policy shock as opposed to other developments that led to the equally rapid fall in the previous two years.

The second monetary shock emphasized by Friedman and Schwartz is more promising in this regard. From July 1936 to January 1937, the Fed announced a doubling of reserve requirements (fully implemented by May 1937) and the Treasury engaged in sterilization of gold inflows. Before this period, industrial production had been rising rapidly. Shortly after, it plunged dramatically. Friedman and Schwartz argue that the Fed's policy actions caused this sharp recession. However, a closer look reveals important confounding factors. Fiscal policy tightened sharply in 1937 because of the end of the 1936 veterans' bonus and the first widespread collection of Social Security payroll taxes, among other factors. In fact, prior to Friedman and Schwartz's (1963) work, Keynesians often held up the 1937 recession as an example of the power of fiscal policy. Romer and Romer (1989) also emphasize that 1937 was a year of substantial labor unrest. For these reasons, it is perhaps not clear that this episode is "so consistent and sharp as to leave little doubt about [its] interpretation."

A more general argument runs through the Friedman and Schwartz (1963) narrative of the Great Depression period. They argue that the Fed failed to act from early 1930 and March 1933, and instead allowed the money supply to fall and a substantial fraction of the banking system to fail. Eichengreen (1992) argues that an important reason why the Fed did not act during this period was that effective action would have been inconsistent with remaining on the gold standard. One of President Franklin Roosevelt's first policy actions was to take the United States off the gold standard in April 1933, shown by the black vertical line in Figure 1. The dollar quickly depreciated by 30 percent. Industrial production immediately skyrocketed. But, of course, the Roosevelt administration changed a number of policies. Whether it was going off gold or something else that made the difference is not entirely clear.<sup>10</sup>

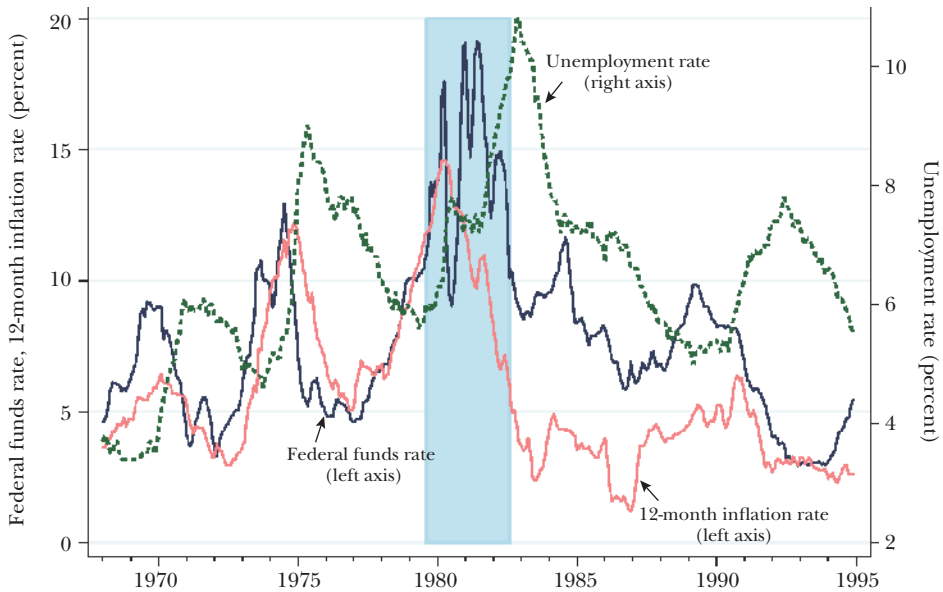
The Volcker disinflation and accompanying twin recessions of the late 1970s and early 1980s are another often-cited piece of evidence on monetary nonneutrality. US inflation had been low and stable since the end of the Korean War in the early 1950s, but then started to rise in the late 1960s. In the 1970s, inflation was both higher and much more volatile than before, as shown in Figure 2. Monetary policy during the 1970s is often described as “stop-go”—tight when the public was exercised about inflation and loose when the public was exercised about unemployment (Goodfriend 2007).

In August 1979, Paul Volcker became chairman of the Federal Reserve. Under Volcker's leadership, policy interest rates rose dramatically between October 1979 and March 1980. However, as shown in Figure 2, the Fed then eased policy such that rates fell even more dramatically (by 9 percentage points) in the spring and summer of 1980 as it became clear that the economy was contracting strongly. In the fall of 1980, inflation was no lower than a year earlier, and the Fed's credibility was, if anything, worse than before. Goodfriend and King (2005) argue that it was only at this point—in November 1980—that Volcker truly broke with prior behavior of the Fed and embarked on a sustained, deliberate disinflation. Interest rates rose dramatically to close to 20 percent and the Fed kept policy tight even in the face of the largest recession the US had experienced since the Great Depression.

The behavior of output during this period is consistent with the view that monetary nonneutrality is large. Output fell dramatically in the spring and summer of 1980 shortly after the Fed raised interest rates sharply. Output then rebounded in late 1980 shortly after the Fed reduced interest rates sharply. Output then fell by a large amount for a sustained period in 1981–1982 while the Fed maintained high interest rates to bring down inflation. Finally, output started recovering when the Fed eased monetary policy in late 1982.

<sup>10</sup>Eichengreen and Sachs (1985) offer related evidence that supports a crucial role for going off gold; they show that countries that went off the gold standard earlier, recovered earlier from the Great Depression. Also, Eggertsson and Pugsley (2006) and Eggertsson (2008) present a model and narrative evidence suggesting that the turning points in 1933, 1937, and 1938 can all be explained by a commitment to reflate the price level (1933 and 1938) and an abandonment of that commitment (1937). Going off gold was an important element of this commitment in 1933.

Figure 2

**Federal Funds Rate, Inflation, and Unemployment from 1965 to 1995**

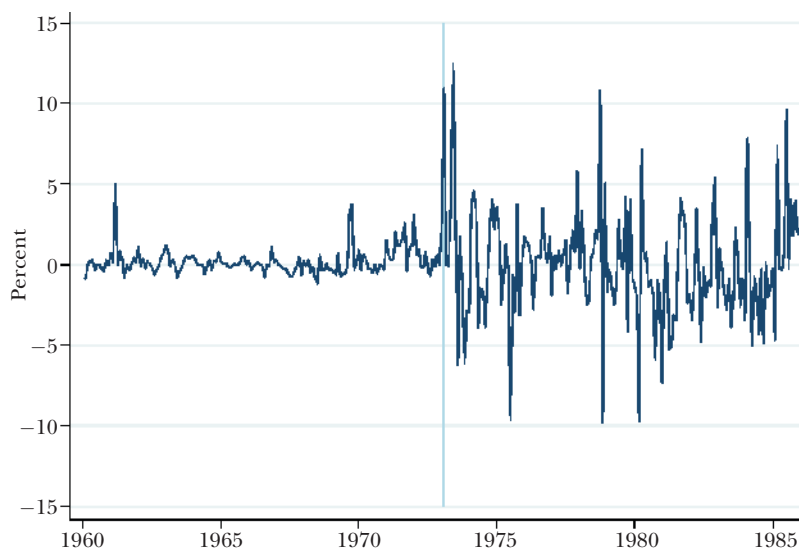
*Note:* The figure plots the federal funds rate (dark solid line, left axis), the 12-month inflation rate (light solid line, left axis), and the unemployment rate (dashed line, right axis). The Volcker disinflation period is the shaded bar (August 1979 to August 1982).

Many economists find the narrative account above and the accompanying evidence about output to be compelling evidence of large monetary nonneutrality.<sup>11</sup> However, there are other possible explanations for these movements in output. There were oil shocks both in September 1979 and in February 1981 (described in Table D.1 in the online Appendix). Credit controls were instituted between March and July of 1980. Anticipation effects associated with the phased-in tax cuts of the Reagan administration may also have played a role in the 1981–1982 recession (Mertens and Ravn 2012).

While the Volcker episode is consistent with a large amount of monetary nonneutrality, it seems less consistent with the commonly held view that monetary policy affects output with “long and variable lags.” To the contrary, what makes the Volcker episode potentially compelling is that output fell and rose largely in sync with the actions of the Fed. If not for this, it would have been much harder to attribute the movements in output to changes in policy.

<sup>11</sup> The Volcker disinflation has also had a profound effect on beliefs within academia and in policy circles about the ability of central banks to control inflation. Today the proposition that inflation is “always and everywhere a monetary phenomenon” is firmly established; so firmly established that it is surprising to modern ears that this proposition was doubted by many in the 1970s (Romer and Romer 2002; Nelson 2005).

Figure 3  
**Monthly Change in the US–German Real Exchange Rate**



*Note:* The figure plots the monthly change in the US–German real exchange rate from 1960 to 1990. The vertical line marks February 1973, when the Bretton Woods system of fixed exchange rates collapsed.

### Discontinuity-Based Identification

There is incontrovertible, reduced-form evidence that monetary policy affects relative prices. The evidence on this point is strong because it can be assessed using discontinuity-based identification methods. The pioneering paper on this topic is Mussa (1986). He argued that the abrupt change in monetary policy associated with the breakdown of the Bretton Woods system of fixed exchange rates in February 1973 caused a large increase in the volatility of the US real exchange rate. Figure 3 plots the monthly change in the US–German real exchange rate from 1960 to 1990. There is a clear break in the series in February 1973: its standard deviation rose by more than a factor of four. The switch from a fixed to a flexible exchange rate is a purely monetary action. In a world where monetary policy has no real effects, such a policy change would not affect real variables like the real exchange rate. Figure 3 demonstrates dramatically that the world we live in is not such a world.

As with any discontinuity-based identification scheme, the identifying assumption is that other factors affecting the real exchange rate do not change *discontinuously* in February 1973. The breakdown of the Bretton Woods system was caused by a gradual build-up of imbalances over several years caused by persistently more inflationary macroeconomic policy in the United States than in Germany, Japan, and other countries. There were intense negotiations for over a year before the system finally collapsed. It is hard to think of plausible alternative



explanations for the *discontinuous* increase in the volatility of the US real exchange rate that occurred at this time.<sup>12</sup>

There is also strong discontinuity-based evidence that monetary policy affects real interest rates. A large amount of monetary news is revealed discretely at the time of the eight regularly scheduled meetings of the Federal Open Market Committee (FOMC) of the Federal Reserve. In Nakamura and Steinsson (forthcoming), we construct monetary shocks using changes in interest rates over a 30-minute window surrounding FOMC announcements. Over such a short window, movements in interest rates are dominated by the monetary announcement. Furthermore, if financial markets are efficient, any systematic response of the Fed to information about the economy that is public at the time of the announcement is already incorporated into financial markets, and, therefore, does not show up as spurious variation in the monetary shocks we construct. We show that nominal and real interest rates respond roughly one-for-one to these monetary shocks several years out into the term structure of interest rates, while expected inflation responds little. For example, the three-year nominal and real forward rates respond by very similar amounts. Hanson and Stein (2015) present similar empirical results.

Direct high-frequency evidence of the effect of monetary policy on *output* is much weaker than for relative prices (Cochrane and Piazzesi 2002; Angrist, Jordà, and Kuersteiner 2017). The reason is that high-frequency monetary shocks are quite small, as is often the case with very cleanly identified shocks. This implies that the statistical power to assess their effect on output several quarters in the future is limited (because many other shocks also affect output over longer time periods).

The effect of monetary shocks on output can be broken into two separate questions: First, how much do monetary shocks affect various relative prices? Second, how much do these various relative prices affect output? Viewing things this way helps clarify that it is the first question that is *the* distinguishing feature of models in which monetary policy has effects on output. All models—neoclassical and New Keynesian—imply that relative prices affect output. However, only in some models does monetary policy affect relative prices, and these models typically imply that money is nonneutral.

A complicating factor regarding the high-frequency evidence from announcements of the Federal Open Market Committee is that these announcements may lead the private sector to update its beliefs not only about the future path of monetary policy, but also about other economic fundamentals. If the Fed reveals greater optimism about the economy than anticipated, the private sector may revise its own beliefs about where the economy is headed. We find evidence for such “Fed information effects” in Nakamura and Steinsson (forthcoming). A surprise tightening of policy is associated with an *increase* in expected output growth in the Blue Chip survey of professional forecasters (the opposite of what standard analysis of monetary shocks

<sup>12</sup>Velde (2009) presents high-frequency evidence on the effects of a large monetary contraction in 18th-century France. He shows that domestic prices responded sluggishly and incompletely to this shock. Burstein, Eichenbaum, and Rebelo (2005) show that large, abrupt devaluations lead to large changes in real exchange rates that are mostly due to changes in the relative price of nontradable goods.

would imply). We show that a New Keynesian model that incorporates Fed information effects can match this fact as well as facts about the response of real interest rates and expected inflation. In this model, a Fed tightening has two effects on the economy: a traditional contractionary effect through increases in real interest rates relative to natural rates and a less traditional expansionary effect coming from the Fed's ability to increase optimism about the economy. Earlier evidence on Fed information effects is presented by Romer and Romer (2000), Faust, Swanson, and Wright (2004), and Campbell, Evans, Fisher, and Justiniano (2012).

The Fed information effect implies that an external validity problem arises whenever researchers use responses to monetary shocks to make inferences about the effects of systematic monetary policy actions. Surprise monetary actions lead to information effects, while the systematic response of the Fed to new data do not. This implies that surprise monetary shocks are less contractionary than the systematic component of monetary policy.

### **Using the Narrative Record to Identify Shocks**

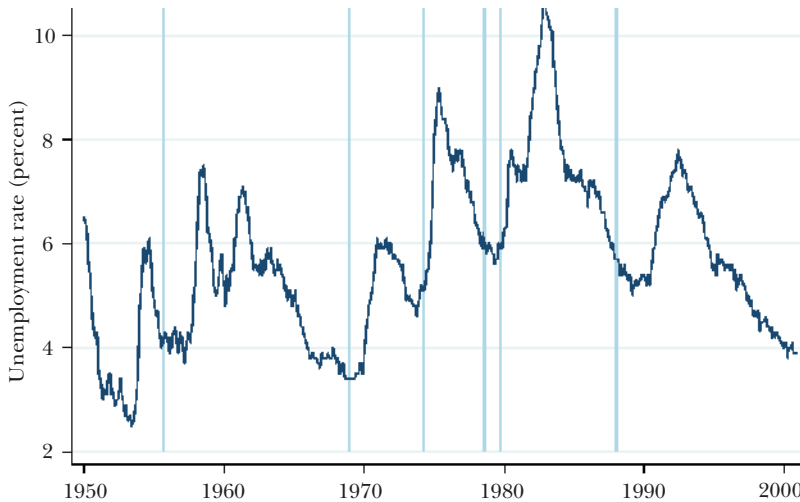
Romer and Romer (1989) argue that contemporaneous Federal Reserve records can be used to identify natural experiments. They use such records to identify “episodes in which the Federal Reserve attempted to exert a contractionary influence on the economy in order to reduce inflation” in the post-World War II period. They identify six such episodes and subsequently added a seventh (Romer and Romer 1994). Figure 4 shows that after each of the Romer–Romer dates, marked by vertical lines, unemployment rises sharply. Pooling the data from these seven episodes, Romer and Romer argue that together they constitute strong evidence for substantial real effects of monetary policy.

While this “narrative approach” to identification is clearly valuable, it faces several challenges. First, narrative shocks are selected by an inherently opaque process. This raises the concern that the results are difficult to replicate. Second, with only seven data points, it may happen by chance that some other factor is correlated with the monetary shocks. In cases when one has dozens or hundreds of shocks, any random correlation with some other factor is likely to average to zero. But with seven data points, this may not happen. In fact, Hoover and Perez (1994) argue that Romer and Romer’s monetary dates are strikingly temporally correlated with dates of oil shocks (see Table D.1 in the appendix).

Third, narrative shocks are often found to be predictable, suggesting the possibility of endogeneity. In the case of Romer and Romer’s (1989) analysis, this concern was raised by Shapiro (1994) and Leeper (1997). However, it can be difficult to convincingly establish predictability due to overfitting concerns. The cumulative number of regressions run by researchers trying to assess the predictability of narrative shocks may be very large. Even by chance, some of these should turn up estimates that are statistically significant.<sup>13</sup>

<sup>13</sup>In the case of Leeper’s (1997) results, Romer and Romer (1997) present a simple pseudo-out-of-sample procedure demonstrating dramatic overfitting. They redo Leeper’s analysis seven times, in each case leaving

Figure 4  
Unemployment from 1950 to 2000



Note: The figure plots the unemployment rate from 1950 to 2000. The light vertical lines indicate the dates identified by Romer and Romer (1989, 1994) as “episodes in which the Federal Reserve attempted to exert a contractionary influence on the economy in order to reduce inflation.”

### Controlling for Confounding Factors

The most prevalent approach to identifying exogenous variation in monetary policy is to attempt to control for confounding factors. Much of the vector autoregression literature takes this approach. A common specification is to regress the federal funds rate on contemporaneous values of several variables (such as output and inflation), as well as several lags of itself and these other variables, and then to view the residuals from this regression as exogenous monetary policy shocks.<sup>14</sup>

This approach to identifying monetary policy shocks is often described as involving “minimal identifying assumptions.” In our view, however, the implicit assumptions are very strong. What is being assumed is that controlling for a few lags of a few variables captures all endogenous variation in policy. This seems highly unlikely to be true in practice. The Fed bases its policy decisions on a huge amount of data. Different considerations (in some cases highly idiosyncratic) affect policy at different times. These include stress in the banking system, sharp changes in commodity prices, a recent stock market crash, a financial crisis in emerging markets, terrorist attacks, temporary investment tax credits, and the Y2K computer glitch. The list goes on and on. Each of these considerations may only affect policy in a meaningful way on a small number of dates, and the number of such influences is so large that it is not feasible

---

out one of the monetary shock dates. For each set of estimates, they then see if the model can predict the shock date that is left out. Using this pseudo-out-of-sample procedure, they find no predictability.

<sup>14</sup>A common way of describing this procedure is as performing a Cholesky decomposition of the reduced form errors from the vector autoregression with the federal funds rate ordered last.

to include them all in a regression. But leaving any one of them out will result in a monetary policy “shock” that the researcher views as exogenous but is in fact endogenous. Rudebusch (1998) is a classic discussion of these concerns.

To demonstrate this point, consider an example. Following Cochrane and Piazzesi (2002), Figure 5 plots the evolution of the federal funds rate target of the Federal Reserve as well as the one-month Eurodollar rate—that is, a one-month interbank interest rate on US dollars traded in London—around the time of the September 11, 2001, terrorist attacks. The Eurodollar rate may be viewed as the average expected federal funds rate over the next month. On September 10, the Eurodollar rate traded at 3.41 percent, quite close to the target federal funds rate of 3.5 percent. Markets did not open on September 11 due to the terrorist attacks in New York. Before markets reopened on September 17, the Fed announced a 50 basis points drop in the target federal funds rate to 3 percent and the one-month Eurodollar rate traded at 2.97 percent that day. The futures contract for the federal funds rate in September 2001 reveals that this 50 basis points drop in the federal funds rate was completely unanticipated by markets as of September 10. The one-month Eurodollar rate was trading below the target for the federal funds rate on September 10 because markets were anticipating an easing of 25–50 basis points at the regularly scheduled meeting of the Federal Open Market Committee on October 2.

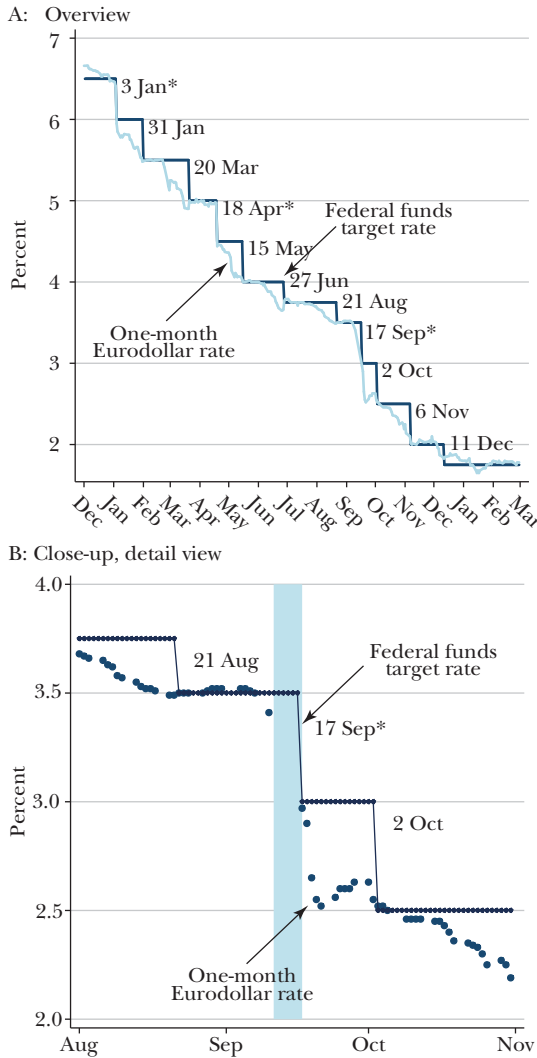
The easing on September 17 was obviously due to the terrorist attacks and therefore obviously endogenous: the terrorist attacks caused the Fed to revise its assessment about future growth and inflation, leading to an immediate drop in interest rates. However, standard monetary vector autoregressions treat this policy easing as an exogenous monetary shock. The reason is that the controls in the policy equation of the vector autoregression are not able to capture the changes in beliefs that occur on 9/11. Nothing has yet happened to any of the controls in the policy equation (even in the case of a monthly vector autoregression). From the perspective of this equation, therefore, the drop in interest rates in September 2001 looks like an exogenous easing of policy. Any unusual (from the perspective of the vector autoregression) weakness in output growth in the months following 9/11 will then, perversely, be attributed to the exogenous easing of policy at that time. Clearly, this is highly problematic.

In our view, the way in which identification assumptions are commonly discussed in the vector autoregression literature is misleading. It is common to see “the” identifying assumption in a monetary vector autoregression described as the assumption that the federal funds rate does not affect output and inflation contemporaneously. This assumption sounds innocuous, almost like magic: You make one innocuous assumption, and voilà, you can estimate the dynamic causal effects of monetary policy. We remember finding this deeply puzzling when we were starting off in the profession.

In fact, the timing assumption that is usually emphasized is not the only identifying assumption being made in a standard monetary vector autoregression. The timing assumption rules out reverse causality. Output and interest rates are jointly determined. An assumption must be made about whether the contemporaneous correlation between these variables is taken to reflect a causal influence of one on the other or the reverse. This is what the timing assumption does.

Figure 5

**Federal Funds Rate Target and 1-Month Eurodollar Rate in 2001 and Early 2002**



*Note:* The figure plots the federal funds rate target (the steps) and the one-month Eurodollar rate (smooth line in left panel and the dots in right) at a daily frequency (beginning of day) from December 2001 to March 2002. Dates of changes in the federal funds rate target are indicated in the figure. The dates marked with an asterisk (\*) are unscheduled Federal Open Market Committee conference calls. The other dates are scheduled FOMC meetings. The shaded bar in Figure 5B indicates September 11, 2001 (day of the New York terrorist attacks) to September 17 (the day the markets reopened). Figure 5A is very similar to the top panel of Figure 1 in Cochrane and Piazzesi (2002).

But reverse causality is not the only issue when it comes to identifying exogenous variation in policy. Arguably, a much bigger issue is whether monetary policy is reacting to some other piece of information about the current or expected future state of the economy that is not included in the vector autoregression (that is, omitted variables

bias). The typical vector autoregression includes a small number of variables and a small number of lags (usually one year worth of lagged values). Any variable not sufficiently well-proxied by these variables is an omitted variable. The omission of these variables leads endogenous variation in policy to be considered exogenous. In the econometrics literature on structural vector autoregression, this omitted variables issue is referred to as the “non-invertibility problem” (Hansen and Sargent 1991; Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson 2007; Plagborg-Møller 2017).

The extremely rich nature of the Fed’s information set means that it is arguably hopeless to control individually for each relevant variable. Romer and Romer (2004) propose an interesting alternative approach. Their idea is to control for the Fed’s own “Greenbook” forecasts. The idea is that the endogeneity of monetary policy is due to one thing and one thing only: what the Fed thinks will happen to the economy. If one is able to control for this, any residual variation in policy is exogenous. For this reason, the Fed’s forecasts are a sufficient statistic for everything that needs to be controlled for.<sup>15</sup>

Romer and Romer’s (2004) approach helps answer the question: What is a monetary shock? The Fed does not roll dice. Every movement in the intended federal funds rate is in response to something. Some are in response to developments that directly affect the change in output in the next year. These are endogenous when changes in output over the next year are the outcome variable of interest. But the Fed may also respond to other things: time variation in policymakers’ preferences and goals (for example, their distaste for inflation), time variation in policymakers’ beliefs about how the economy works, political influences, or pursuit of other objectives (for example, exchange rate stability). Importantly, changes in policy need *not* be unforecastable as long as they are orthogonal to the Fed’s forecast of the dependent variable in question. However, changes in policy that are forecastable (for example, forward guidance) are more complicated to analyze since they can have effects both upon announcement and when they are implemented.

### What Do We Do with These Shocks?

Any exercise in dynamic causal inference involves two conceptually distinct steps: 1) the construction of the shocks, and 2) the specification used to construct an impulse response once one has the shocks in hand. We have discussed the first of these steps in detail above. But the second step is also important. A specification that imposes minimal structure (apart from linearity) is to directly regress the variable of interest (say, the change in output over the next year) on the shock, perhaps controlling for some variables determined before the shock occurs (pre-treatment controls). This is the specification advocated by Jordà (2005). To construct an

<sup>15</sup>Suppose we are interested in the effect of a change in monetary policy at time  $t$ , denoted  $\Delta r_t$ , on the change in output over the next  $j$  months,  $\Delta^j y_{t+j} = y_{t+j} - y_{t-1}$ . The potential concern is that  $\Delta r_t$  may be correlated with some other factor that affects  $\Delta^j y_{t+j}$ . But this can only be the case if the Fed knows about this other factor, and to the extent that it does, this should be reflected in the Fed’s time  $t$  forecast of  $\Delta^j y_{t+j}$ . As Cochrane (2004) emphasizes, controlling for the Fed’s time  $t$  forecast of  $\Delta^j y_{t+j}$  should therefore eliminate all variation in policy that is endogenous to the determination of  $\Delta^j y_{t+j}$ .

impulse response using this approach, one must run a separate regression for each time horizon that one is interested in plotting.

Standard vector autoregressions construct impulse response functions using a different approach that imposes much more structure: they use the estimated dynamics of the entire vector autoregression system to iterate forward the response of the economy to the shock. This method for constructing impulse responses embeds a new set of quite strong identifying assumptions. In a standard monetary vector autoregression, whether the shocks truly represent exogenous variation in monetary policy only relies on the regression equation for the policy instrument being correctly specified. In contrast, the construction of the impulse response relies on the entire system of equations being a correct representation of the dynamics of all the variables in the system—that is, it relies on the whole model being correctly specified.

It is well-known that the solution to any linear rational expectations model can be represented by a vector autoregression (Blanchard and Kahn 1980; Sims 2002). This idea is the usual defense given regarding the reasonableness of the impulse response construction in a standard vector autoregression. However, to estimate the true vector autoregression, all state variables in the economy must be observable so that they can be included in the system. If this is not the case, the vector autoregression is misspecified and the impulse responses that it yields are potentially biased.<sup>16</sup>

Coibion (2012) has drawn attention to the fact that in Romer and Romer's (2004) results, the peak responses of industrial production and unemployment to a change in the federal funds rate are roughly six times larger than in a standard monetary vector autoregression. In the online appendix to this paper, we revisit this issue by estimating the response of industrial production and the real interest rate to monetary shocks in six different ways. First, we use two different shock series: Romer and Romer's (2004) shock series (as updated and improved by Wieland and Yang 2017) and a shock series from a standard monetary vector autoregression. Second, we estimate the impulse response using three different methods: iterating the vector autoregression dynamics, direct regressions, and the single-equation autoregressive model employed by Romer and Romer (2004). This analysis shows that both the shocks and the method for constructing an impulse response can lead to meaningful differences and help explain the difference in results between Romer and Romer (2004) and standard monetary vector autoregressions. In addition, this analysis shows—as Coibion (2012) emphasizes—that about half of the difference is due to the fact that Romer and Romer's shocks are bigger, that is, they result in larger responses of the real interest rate.

<sup>16</sup>Suppose one of the state variables in the system is not observable. One strategy is to iteratively solve out for that variable. The problem with this is that it typically transforms a vector autoregression of order  $p$  (VAR( $p$ )) into an infinite order vector autoregression moving average system (VARMA( $\infty, \infty$ )) in the remaining variables. Thus, the estimation of standard vector autoregressions relies on the assumption that the true infinite-order vector autoregression moving average system, in the variables that the researcher intends to include in the analysis, can be approximated with a vector autoregression of order  $p$ . This is a strong assumption that we fear is unlikely to hold in practice. In the online appendix to this paper, we present an example (in the form of a problem set) that illustrates these ideas.

A recent innovation in dynamic causal inference is the use of “external instruments” in vector autoregressions (Stock and Watson 2012; Mertens and Ravn 2013; Stock and Watson 2018). Gertler and Karadi (2015) use this method to estimate the effects of exogenous monetary shocks on output, inflation, and credit spreads. The strength of this method is that it allows researchers to use instrumental variables to identify monetary shocks within the context of a vector autoregression. However, it does not relax the assumptions embedded in using the vector autoregression system to construct the impulse response. We discuss this method in more detail in the online appendix. The use of sign restrictions is another recent development in this area (for example, Uhlig 2017).

## Conclusion

Macroeconomics and meteorology are similar in certain ways. First, both fields deal with highly complex general equilibrium systems. Second, both fields have trouble making long-term predictions. For this reason, considering the evolution of meteorology is helpful for understanding the potential upside of our research in macroeconomics. In the olden days, before the advent of modern science, people spent a lot of time praying to the rain gods and doing other crazy things meant to improve the weather. But as our scientific understanding of the weather has improved, people have spent a lot less time praying to the rain gods and a lot more time watching the weather channel.

Policy discussions about macroeconomics today are, unfortunately, highly influenced by ideology. Politicians, policymakers, and even some academics hold strong views about how macroeconomic policy works that are not based on evidence but rather on faith. The only reason why this sorry state of affairs persists is that our evidence regarding the consequences of different macroeconomic policies is still highly imperfect and open to serious criticism. Despite this, we are hopeful regarding the future of our field. We see that solid empirical knowledge about how the economy works at the macroeconomic level is being uncovered at an increasingly rapid rate. Over time, as we amass a better understanding of how the economy works, there will be less and less scope for belief in “rain gods” in macroeconomics and more and more reliance on convincing empirical facts.

■ *We thank Miguel Acosta, Juan Herreño, and Yeji Sung for excellent research assistance. We would like to thank Joshua Angrist, John Cochrane, Gregory Cox, Gauti Eggertsson, Mark Gertler, Adam Guren, Gordon Hanson, Jonathon Hazell, Juan Herreño, Ethan Iizetzki, Alisdair McKay, Edward Nelson, Serena Ng, Mikkel Plagborg-Møller, Valerie Ramey, David Romer, Timothy Taylor, David Thesmar, Jonathan Vogel, Johannes Wieland, Christian Wolf, and Michael Woodford for valuable comments and discussions. We thank the National Science Foundation (grant SES-1056107) and the Alfred P. Sloan Foundation for financial support.*



## References

- Acconica, Antonio, Giancarlo Corsetti, and Saverio Simonelli.** 2014. "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment." *American Economic Review* 104(7): 2185–2209.
- Aguar, Mark, and Erik Hurst.** 2005. "Consumption versus Expenditure." *Journal of Political Economy* 113(5): 919–48.
- Angeletos, George-Marios, David Laibson, Andrea Rebetto, Jeremy Tobacman, and Stephen Weinberg.** 2001. "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation." *Journal of Economic Perspectives* 15(3): 47–68.
- Angrist, Joshua D., Òscar Jordà, and Guido M. Kuersteiner.** Forthcoming. "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited." *Journal of Business and Economic Statistics*.
- Angrist, Joshua D., and Jörn-Steffen Pischke.** 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24(2): 3–30.
- Ashcraft, Adam B.** 2005. "Are Banks Really Special? New Evidence from the FDIC-Induced Failure of Healthy Banks." *American Economic Review* 95(5): 1712–30.
- Autor David H., David Dorn, and Gordon H. Hanson.** 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103(6): 2121–68.
- Barro, Robert J., and Charles J. Redlick.** 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126(1): 51–102.
- Basu, Susanto, John G. Fernald, and Miles S. Kimball.** 2006. "Are Technology Improvements Contractionary?" *American Economic Review* 96(5): 1418–48.
- Beraja, Martin, Erik Hurst, and Juan Ospina.** 2016. "The Aggregate Implications of Regional Business Cycles." NBER Working Paper 21956.
- Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vara.** 2018. "House Prices and Consumption Spending." *Review of Economic Studies* 85(3): 1502–42.
- Bils, Mark, and Peter J. Klenow.** 2004. "Some Evidence on the Importance of Sticky Prices." *Journal of Political Economy* 112(5): 947–85.
- Blanchard, Olivier Jean, and Charles M. Kahn.** 1980. "The Solution of Linear Difference Models under Rational Expectations." *Econometrica* 48(5): 1305–11.
- Blanchard, Olivier, and Roberto Perotti.** 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117(4): 1329–68.
- Boppart, Timo, and Per Krusell.** 2016. "Labor Supply in the Past, Present, and Future: A Balanced-Growth Perspective." NBER Working Paper 22215.
- Burstein, Ariel, Martin Eichenbaum, and Sergio Rebelo.** 2005. "Large Devaluations and the Real Exchange Rate." *Journal of Political Economy* 113(4): 742–84.
- Calmoris, Charles W., and Joseph R. Mason.** 2003. "Consequences of Bank Distress during the Great Depression." *American Economic Review* 93(3): 937–47.
- Campbell, Jeffrey R., Charles L. Evans, Jonas D. M. Fisher, and Alejandro Justiniano.** 2012. "Macroeconomic Effects of Federal Reserve Forward Guidance." *Brookings Papers on Economic Activity* 43(1): 1–80.
- Carvalho, Vasco M., Makoto Nirei, Yukiko U. Saito, and Alirez Tahbaz-Salehi.** 2016. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." Working Paper 2017-01, Becker Friedman Institution for Research in Economics, University of Chicago.
- Catherine, Sylvain, Thomas Chaney, Zongbo Huang, David Alexandre Sraer, and David Thesmar.** 2017. "Quantifying Reduced-Form Evidence on Collateral Constraints." Available at SSRN: <https://ssrn.com/abstract=2631055>.
- Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan.** 2008. "Business Cycle Accounting." *Econometrica* 75(3): 781–836.
- Chetty, Raj.** 2009. "Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-Form Methods." *Annual Review of Economics* 1: 451–87.
- Chetty, Raj.** 2012. "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica* 80(3): 969–1018.
- Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber.** 2013. "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities." *NBER Macroeconomics Annual* 27(1): 1–56.
- Chodorow-Reich, Gabriel.** 2014. "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008–9 Financial

Crisis." *Quarterly Journal of Economics* 129(1): 1–59.

**Chodorow-Reich, Gabriel.** 2017. "Geographic Cross-Sectional Fiscal Multipliers: What Have We Learned?" Working Paper, Harvard University OpenScholar.

**Chodorow-Reich, Gabriel, John Coglianesi, and Loukas Karabarbounis.** 2017. "The Limited Macroeconomic Effects of Unemployment Benefit Extensions." Unpublished paper.

**Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston.** 2012. "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4(3): 118–45.

**Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans.** 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113(1): 1–45.

**Cochrane, John H.** 2004. "Comments on 'A New Measure of Monetary Shocks: Derivation and Implications,' by Christina Romer and David Romer." Presented at NBER EF&G meeting, July 17.

**Cochrane, John H., and Monika Piazzesi.** 2002. "The Fed and Interest Rates: A High-Frequency Identification." *American Economic Review* 92(2): 90–95.

**Coibion, Olivier.** 2012. "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics* 4(2): 1–32.

**Dupor, Bill, and M. Saif Mehkari.** 2016. "The 2009 Recovery Act: Stimulus at the Extensive and Intensive Labor Margins." *European Economic Review* 85: 208–228.

**Eggertsson, Gauti B.** 2008. "Great Expectations and the End of the Depression." *American Economic Review* 98(4): 1476–1516.

**Eggertsson, Gauti B., and Benjamin Pugsley.** 2006. "The Mistake of 1937: A General Equilibrium Analysis." *Monetary and Economic Studies* 24(S1): 151–208.

**Eichengreen, Barry.** 1992. *Golden Fetters: The Gold Standard and the Great Depression 1919–1939*. Oxford University Press.

**Eichengreen, Barry, and Jeffrey Sachs.** 1985. "Exchange Rates and Economic Recovery in the 1930s." *Journal of Economic History* 45(4): 925–46.

**Faust, Jon, Eric Swanson, and Jonathan H. Wright.** 2004. "Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy?" *The BE Journal of Macroeconomics* 4.

**Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson.** 2007. "ABCs (and Ds) of Understanding VARs." *American Economic Review* 97(3): 1021–26.

**Friedman, Milton, and Anna Jacobson Schwartz.** 1963. *A Monetary History of the United States, 1867–1960*. Princeton University Press.

**Gali, Jordi.** 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations." *American Economic Review* 89(1): 249–71.

**Gali, Jordi, David López-Salido, and Javier Vallés.** 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association* 5(1): 227–70.

**Gertler, Mark, and Peter Karadi.** 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7(1): 44–76.

**Goodfriend, Marvin.** 2007. "How the World Achieved Consensus on Monetary Policy." *Journal of Economic Perspectives* 21(4): 47–68.

**Goodfriend, Marvin, and Robert G. King.** 2005. "The Incredible Volcker Disinflation." *Journal of Monetary Economics* 52(5): 981–1015.

**Greenstone, Michael, Alesandre Mas, and Hoai-Luu Nguyen.** 2017. "Do Credit Market Shocks Affect the Real Economy? Quasi-Experimental Evidence from the Great Recession and 'Normal' Economic Times." Available at SSRN: <https://ssrn.com/abstract=2187521>.

**Hagedorn, Marcus, Iourii Manovskii, and Kurt Mitman.** 2015. "The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle?" NBER Working Paper 20884.

**Hall, Robert E.** 2009. "By How Much Does GDP Rise If the Government Buys More Output?" *Brookings Papers on Economic Activity* 40(2): 183–231.

**Hansen, Lars Peter, and Thomas J. Sargent.** 1991. "Two Difficulties in Interpreting Vector Autoregressions." In *Rational Expectations Econometrics*, edited by L. P. Hansen and T. J. Sargent, 77–119. Boulder, Co.: Westview Press

**Hanson, Samuel G., and Jeremy C. Stein.** 2015. "Monetary Policy and Long-Term Real Rates." *Journal of Financial Economics* 115(3): 429–48.

**Hoover, Keven D., and Stephen J. Perez.** 1994. "Post Hoc Ergo Propter Once More: An Evaluation of 'Does Monetary Policy Matter?' in the Spirit of James Tobin." *Journal of Monetary Economics* 34(1): 47–73.

**Hsieh, Chang-Tai, and Peter J. Klenow.** 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124(4): 1403–48.

**Huber, Kilian.** 2018. "Disentangling the Effects of a Banking Crisis: Evidence from German Firms and Counties." *American Economic Review* 108(3): 868–98.

- Itskhoki, Oleg, and Dmitry Mukhin.** 2017. "Exchange Rate Disconnect in General Equilibrium." NBER Working Paper 23401.
- Johnson, David S., Jonathan A. Parker, and Nicolas S. Souleles.** 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96(5): 1589–1610.
- Jordà, Óscar.** 2005. "Estimation and Inference of Impulse Response by Local Projection." *American Economic Review* 95(1): 161–82.
- Kaplan, Greg, and Giovanni L. Violante.** 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica* 82: 1199–1239.
- King, Robert G., Charles I. Plosser, and Sergio T. Rebelo.** 1988. "Production, Growth and Business Cycles: I. The Basic Neoclassical Model." *Journal of Monetary Economics* 21(2–3): 195–232.
- King, Robert G., and Sergio T. Rebelo.** 1999. "Resuscitating Real Business Cycles." In *Handbook of Macroeconomics*, edited by John B. Taylor and Michael Woodford, 927–1007. Amsterdam, Holland: Elsevier.
- Klenow, Peter J., and Oleksiy Kryvtsov.** 2008. "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation." *Quarterly Journal of Economics* 123: 863–904.
- Kydland, Finn E., and Edward C. Prescott.** 1982. "Time to Build and Aggregate Fluctuations." *Econometrica* 50(6): 1345–70.
- Leeper, Eric M.** 1997. "Narrative and VAR Approaches to Monetary Policy: Common Identification Problems." *Journal of Monetary Economics* 40(3): 641–57.
- Martin, Pilippe, and Thomas Philippon.** 2017. "Inspecting the Mechanism: Leverage and the Great Recession in the Eurozone." *American Economic Review* 107(7): 1904–37.
- Meese, Richard A., and Kenneth Rogoff.** 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14(1–2): 3–24.
- Mehra, Rajnish, and Edward C. Prescott.** 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15(2): 145–61.
- Mertens, Karel, and Morten O. Ravn.** 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks." *American Economic Journal: Economic Policy* 4(2): 145–81.
- Mertens, Karel, and Morten O. Ravn.** 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103(4): 1212–47.
- Mian, Ataf, Kamalesh Rao, and Amir Sufi.** 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics* 128(4): 1687–1726.
- Mian, Atif, and Amir Sufi.** 2014. "What Explains the 2007–2009 Drop in Employment?" *Econometrica* 82(6): 2197–2223.
- Miyamoto, Wataru, Thuy Lan Nguyen, and Dmitry Sergeyev.** 2018. "Government Spending Multipliers under the Zero Lower Bound: Evidence from Japan." *American Economic Journal: Macroeconomics* 10(3): 247–77.
- Mondragon, John.** 2018. "Household Credit and Employment in the Great Recession." Nielsen Dataset Paper I-025, Kilts Center for Marketing at Chicago Booth. Available at SSRN: <https://ssrn.com/abstract=2521177>.
- Mussa, Michael.** 1986. "Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications." *Carnegie-Rochester Conference Series on Public Policy* 25: 117–214.
- Nakamura, Emi, and Jón Steinsson.** 2008. "Five Facts about Prices: A Reevaluation of Menu Cost Models." *Quarterly Journal of Economics* 123(4): 1415–64.
- Nakamura, Emi, and Jón Steinsson.** 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104(3): 753–92.
- Nakamura, Emi, and Jón Steinsson.** Forthcoming. "High Frequency Identification of Monetary Non-Neutrality: The Information Effect." *Quarterly Journal of Economics*.
- Nekarda, Christopher J., and Valerie A. Ramey.** 2011. "Industry Evidence on the Effects of Government Spending." *American Economic Journal: Macroeconomics* 3(1): 36–59.
- Nelson, Edward.** 2005. "The Great Inflation of the Seventies: What Really Happened?" *The B.E. Journal of Macroeconomics* 5(1).
- Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland.** 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103(6): 2530–53.
- Peek, Joe, and Eric S. Rosengren.** 2000. "Collateral Damage: Effects of the Japanese Bank Crisis on Real Activity in the United States." *American Economic Review* 90(1): 30–45.
- Plagborg-Møller, Mikkel.** 2017. "Bayesian Inference on Structural Impulse Response Functions." Unpublished paper, Princeton University.
- Prescott, Edward C.** 1986. "Theory Ahead of Business-Cycle Measurement." *Carnegie-Rochester Conference Series on Public Policy* 25: 11–44.
- Ramey, Valerie A.** 2011. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126(1): 1–50.
- Ramey, Valerie A.** 2016. "Macroeconomic Shocks and Their Propagation." Chap. 3 of *Handbook of Macroeconomics*, vol. 2A. edited by John B.

- Taylor and Harald Uhlig, 71–162. Amsterdam, Holland: Elsevier.
- Ramey, Valerie A., and Sarah Zubairy.** 2018. “Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data.” *Journal of Political Economy* 126(2): 850–901.
- Romer, Christina D., and David H. Romer.** 1989. “Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz.” In *NBER Macroeconomics Annual*, edited by O. J. Blanchard and S. Fischer, 121–170. Cambridge, MA: MIT Press.
- Romer, Christina D., and David H. Romer.** 1994. “Monetary Policy Matters.” *Journal of Monetary Economics* 34(1): 75–88.
- Romer, Christina D., and David H. Romer.** 1997. “Identification and the Narrative Approach: A Reply to Leeper.” *Journal of Monetary Economics* 40(3): 659–65.
- Romer, Christina D., and David H. Romer.** 2000. “Federal Reserve Information and the Behavior of Interest Rates.” *American Economic Review* 90(3): 429–57.
- Romer, Christina D., and David H. Romer.** 2002. “The Evolution of Economic Understanding and Postwar Stabilization Policy.” In *Rethinking Stabilization Policy*, 11–78. Kansas City: Federal Reserve Bank of Kansas City.
- Romer, Christina D., and David H. Romer.** 2004. “A New Measure of Monetary Shocks: Derivation and Implications.” *American Economic Review* 94(4): 1055–84.
- Rotemberg, Julio J., and Michael Woodford.** 1997. “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy.” In *NBER Macroeconomics Annual 1997*, edited by B. S. Bernanke and J. J. Rotemberg, 297–346. Cambridge, MA: MIT Press.
- Rudebusch, Glenn D.** 1998. “Do Measures of Monetary Policy in a VAR Make Sense?” *International Economic Review* 39(4): 907–931.
- Saiz, Albert.** 2010. “The Geographic Determinants of Housing Supply.” *Quarterly Journal of Economics* 125(3): 1253–96.
- Shapiro, Michael D.** 1994. “Federal Reserve Policy: Cause and Effect.” In *Monetary Policy*, edited by N. G. Mankiw, 307–334. University of Chicago Press.
- Shimer, Robert.** 2005. “The Cyclical Behavior of Equilibrium Unemployment and Vacancies.” *American Economic Review* 95(1): 25–49.
- Shimer, Robert.** 2009. “Convergence in Macroeconomics: The Labor Wedge.” *American Economic Journal: Macroeconomics* 1(1): 280–97.
- Shoag, Daniel.** 2015. “The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns.” Working Paper, Harvard University.
- Sims, Christopher.** 2002. “Solving Linear Rational Expectations Models.” *Journal of Computational Economics* 20(1–2): 1–20.
- Sinai, Todd, and Nicholas S. Souleles.** 2005. “Owner-Occupied Housing as a Hedge against Rent Risk.” *Quarterly Journal of Economics* 120(2): 763–89.
- Smith, Anthony A.** 2008. “Indirect Inference.” In *The New Palgrave Dictionary of Economics*, Second Edition, edited by S. Durlauf and L. E. Blume. London: Palgrave Macmillan. Also available at: <http://www.econ.yale.edu/smith/palgrave7.pdf>.
- Stock, James H., and Mark W. Watson.** 2012. “Disentangling the Channels of the 2007–09 Recession.” *Brookings Papers on Economic Activity* 43(1): 81–135.
- Stock, James H., and Mark W. Watson.** 2018. “Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments.” NBER Working Paper 24216.
- Suárez Serrato, Juan Carlos, and Philippe Wingender.** 2016. “Estimating Local Fiscal Multipliers.” NBER Working paper 22455.
- Uhlig, Harald.** 2017. “Shocks, Sign Restrictions, and Identification.” Chap. 4 in *Advances in Economics and Econometrics: Eleventh World Congress*, vol. 2, edited by B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, 97–127. Cambridge University Press.
- Velde, François R.** 2009. “Chronicle of a Deflation Unforetold.” *Journal of Political Economy* 117(4): 591–634.
- Wieland Johannes F., and Mu-Jeung Yang.** 2017. “Financial Dampening.” Unpublished paper, University of California, San Diego.
- Wilson, Daniel J.** 2012. “Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act.” *American Economic Journal: Economic Policy* 4(3): 251–82.