

Identification in Macroeconomics

Emi Nakamura and Jón Steinsson*

Columbia University

October 25, 2017

Abstract

This paper discusses empirical approaches macroeconomists use to answer questions like: What does monetary policy do? How large are the effects of fiscal stimulus? What caused the Great Recession? Why do some countries grow faster than others? Identification of causal effects plays two roles in this process. In certain cases, progress can be made using the direct approach of identifying plausibly exogenous variation in a policy and using this variation to assess the effect of the policy. However, external validity concerns limit what can be learned in this way. Carefully identified causal effects estimates can also be used as moments in a structural moment matching exercise. We argue that such “identified moments” are often powerful diagnostic tools for distinguishing between important classes of models (and thereby learning about the effects of policy). To illustrate these notions we discuss the growing use of cross-sectional evidence in macroeconomics and consider what the best existing evidence is on the effects of monetary policy.

JEL Classification: E00

*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, Adam Guren, Jonathon Hazell, Juan Herreño, Ethan Ilzetzki, Alisdair McKay, Edward Nelson, Serena Ng, Valerie Ramey, David Romer, David Thesmar, Jonathan Vogel, Johannes Wieland, 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.

1 Introduction

Any scientific enterprise needs to be grounded in solid empirical knowledge about the phenomenon in question. Milton Friedman put this well in his Nobel lecture in 1976: “In order to recommend a course of action to achieve an objective, we must first know whether that course of action will in fact promote the objective. Positive scientific knowledge that enables us to predict the consequences of a possible course of action is clearly a prerequisite for the normative judgment whether that course of action is desirable.”

Many of the main empirical questions in macroeconomics are the same as they were 80 years ago when macroeconomics came into being as a separate sub-discipline of economics in the wake of the Great Depression. These are questions such as: 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 or viewing it from afar 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. To be able to accumulate direct positive knowledge about the effects of monetary policy on the economy, we need exogenous variation in 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 little bit of incoming data about the economy so as to make monetary policy as endogenous as it possibly can be. Indeed, the main purpose of the Federal Reserve is to vary monetary policy in a way that reacts to other developments that are affecting output and inflation. This fact, of course, makes our lives as empirical macroeconomists very difficult. Quite a bit of ingenuity and careful research is required to identify a component of monetary policy that is plausibly exogenous to the main outcome variables of interest such as future output and inflation.

This challenge can most clearly be seen by considering an example. The Federal Reserve lowered interest rates aggressively in 2008 as evidence mounted that the economy was heading for a severe downturn. Suppose one sought to use this variation in monetary policy to estimate the effect of monetary policy on the economy by estimating an OLS regression of the change in output on the change in interest rates. Doing this might lead one to conclude that reductions in interest rates lead to *decreases* in output. Would this constitute convincing evidence on the effects of monetary policy? Of course not. The reason the Fed was lowering interest rates was because other

factors—such as rapidly falling home prices and their effects on the balance sheets of households, firms, and banks—were negatively affecting the economy. In a simple OLS regression, these other factors would confound the effects of the change in monetary policy. This approach would, therefore, uncover not the pure effect of the interest rate reduction, but rather, the combined effect of the interest rate reduction and the adverse macroeconomic factors that led the Fed to undertake it.¹

The challenging nature of identification in macroeconomics as well as the modest number of data points macroeconomists typically have to work with has meant that macroeconomists have attempted an eclectic approach to identification. As wars between the forces of structural and non-structural methods have raged in applied microeconomics, empirical macroeconomists have borrowed from both camps but also advanced ideas that in some cases seem puzzling or even downright backwards to empirical researchers outside our field. In our view, this eclectic approach has served our field well and while empirical macroeconomists still have much to learn from their colleagues in applied microeconomics, we feel that applied microeconomists could also learn a few things from our approach to identification.

Empirical work in macroeconomics may be divided into several strands. One of these strands attempts to directly answer some of the field’s big questions by identifying plausibly exogenous variation in macroeconomic policy and using this variation to assess the effects of the policy. Since truly randomized experiments are not available when it comes to macroeconomics policies, studies in this vein typically seek to identify natural experiments of various types. The approaches macroeconomists use are by and large similar to the work-horse methods of applied microeconomics. The analysis looks a bit different in macroeconomics mainly because of the greater importance of dynamics. In section 4, we ask what the best available evidence is on monetary non-neutrality. In the context of answering this question, we provide a critical assessment of how several of the main methods of direct causal inference are used in macroeconomics. In particular, we discuss the use of VARs as a method for identifying exogenous variation in policy and constructing dynamic causal estimates.

Researchers attempting direct causal inference in macroeconomics, however, face two important challenges. The first challenge is the one we discuss above. Namely, it is difficult to identify exogenous variation in macroeconomic policy. The second challenge is that the natural experiments we can identify in the data are rarely exactly the experiments we would need to answer the policy question we are interested in. This “external validity” problem can be illustrated by think-

¹An analogous argument can be made regarding the fiscal stimulus that was enacted during the Great Recession.

ing about monetary and fiscal policy. One such 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 (e.g., will the Fed 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 (e.g., when policy 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 degree of persistence and more generally the future time profile of a policy action greatly affects its impact on current output and inflation.²

A second external validity issue is that the effects of fiscal shocks depend on the response of monetary policy (i.e., whether it is constrained by the zero-lower-bound) and the response of tax policy (i.e., whether a war is financed with a sharp contemporary increase in taxes or with debt that is whittled down over many subsequent years). The effects of monetary policy, of course, also depend on the response of fiscal and tax 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 various other characteristics of the economy, such as how open it 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, at best, give us a partial assessment of how future monetary and fiscal policy actions will affect the economy. One response to these issues is to simply gather direct causal evidence about each and every different case. This may, however, not be feasible. And 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 typically takes the form of researchers identifying a set of moments in the data and arguing that these moments can discriminate between different models of how the economy works. All too often, this structural mode of inference is viewed as completely separate from empirical work seeking to uncover causal effects. In fact, however, estimates of causal effects (i.e., the response to structural shocks) are often particularly informative moments for distinguishing between important classes of macroeconomic models. The reason for this is that the value of such “identified moments” often provide evidence on a specific causal mechanism and

²For example, a highly persistent monetary policy shock has much larger effects on current output and inflation than a transitory shock. Standard models also imply that news about a future monetary policy action of equal magnitude (in terms of its effect on real interest rates) has larger effects if the news is about policy further in the future.

are insensitive to other aspects of the economy.³

Consider the analysis of [Rotemberg and Woodford \(1997\)](#) and [Christiano, Eichenbaum, and Evans \(2005\)](#). These authors use identified responses of output and inflation to monetary shocks to discriminate between different business cycle models. This analysis provides support for New Keynesian models in comparison to RBC models (i.e., the null hypothesis of no nominal rigidities is rejected). The reason for this is that RBC models imply that monetary shocks should affect inflation but not output, while the empirical evidence suggests a substantial response of output. This conclusion is insensitive to many model features, such as how investment is modeled, the structure of the tax system, and whether the labor market is competitive or not.

Similarly, [Gali \(1999\)](#) and [Basu, Fernald, and Kimball \(2006\)](#) argue that identified responses of output and hours to productivity shocks reject RBC models in favor of New Keynesian models. In RBC models, an increase in productivity increases output and hours. Gali and Basu et al.'s empirical evidence however suggests that improvements in productivity lead firms to fire workers in the short run. New Keynesian models can match this implication under the assumption that monetary policy does not react strongly enough to productivity shocks (perhaps because they are hard to observe in real time). While this type of analysis can be very informative about how the world works, it clearly relies heavily on the identifying assumptions made when the causal effects are estimated.

Estimates of the marginal propensity to consume (MPC) from a transitory fiscal rebate is another example of an identified moment. Recent work has used truly random variation in the timing of fiscal stimulus checks to estimate a quarterly MPC of roughly 0.25 for non-durable consumption (see, e.g., [Johnson, Parker, Souleles, 2006](#); [Parker et al., 2013](#)). Simple models in which households have standard preferences and access to complete markets cannot match these estimates. In such models households choose to smooth spending out of lump sum payments over their remaining lifetime, implying a quarterly MPC close to zero. Models where households face self-control problems and incomplete markets models with uninsurable income risk, borrowing constraints, and illiquid assets are better able to match this evidence ([Angeletos et al., 2001](#); [Kaplan and Violante, 2014](#)). This inference is insensitive to a long list of other features of these two

³The 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 a researcher wants his or her 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. In other words, we are using the term “identified moments” as short-hand for “target statistics derived from estimators of responses to identified structural shocks.”

model classes such as whether prices and wages are sticky, whether business cycles are driven by demand or supply shocks, whether search frictions are important, etc., etc. In this sense, inference based on the the MPC is quite robust.⁴

In recent years, macroeconomists have increasingly turned to analyzing regional data. Doing this multiplies the number of data points available to macroeconomists by an order of magnitude or more. It also allows researchers to identify responses to various types of shocks that are hopefully confounded by other factors at the aggregate level. One example of this type of work is the recent literature on regional fiscal multipliers (see, e.g., [Nakamura and Steinsson, 2014](#)). Such estimates are a good example of identified moments. They do not *directly* answer the policy question macroeconomists are most interested in—the effect of fiscal stimulus at the national level—since various types of general equilibrium effects are “differenced out” by time fixed effects. However, the regional fiscal multiplier turns out to have a great deal of power in distinguishing between different models of the business cycle. Models that can match a large regional multiplier imply that output responds strongly to demand shocks. In these models the aggregate multiplier is large when monetary policy is accommodative, e.g., 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. Similar arguments have been made in a number of other areas in macroeconomics in recent years. We will discuss these ideas in more detail in [section 3](#).

The state of empirical macroeconomics is frequently critiqued. These critiques tend to focus on battles of decades past. Our view is much more positive. There has been a great deal of growth in empirical macroeconomics over the past decade or so with new approaches being used and the standards for identification rising rapidly. A particularly exciting development has been the growing literature estimating what we call identified moments and relating them to fully formed general equilibrium structural models. This growing synergy between well identified empirical work and serious theoretical work has the potential to advance our knowledge about the core questions in macroeconomics substantially.

The rest of this article is organized as follows. [Section 2](#) discusses the value of identified moments in structural estimation. It also discusses the distinction between micro moments and macro

⁴Our argument for moment matching using identified moments is related to [Chetty’s \(2009\)](#) argument for sufficient statistics. Our emphasis is, however, on using the identified moments to distinguish between models, i.e., learn about deep structural parameters. The hope is that these structural parameter estimates will help answer many policy questions. Chetty’s emphasis is on combining identified moments into a formula (a “sufficient statistic”) that answers a particular policy question.

moments and the value of coming up with portable statistics. Section 3 discusses the growing use of cross-sectional identification strategies in macroeconomics. A key issue is how to go from cross-sectional moments to aggregate implications. In this context, we survey the empirical literature on the government spending multiplier. Section 4 provides a critical assessment of methods used for direct causal inference in macroeconomics. We discuss VARs in detail (including the use of external instruments), but we also discuss the use of large shocks, the narrative record, and discontinuity-based identification. We do this in the context of surveying the empirical literature on monetary non-neutrality.

2 The Power of Portable Statistics

One of the important innovative features of the early RBC literature was a move away from using likelihood-based empirical methods towards empirical evaluation based on matching moments (Kydland and Prescott, 1982; Prescott, 1986).⁵ 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).⁶ The equity premium is a good example of a portable statistic.⁷ 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.

2.1 Micro and Macro Moments

It is useful to distinguish between several different types of moments that have been influential in empirical macroeconomics. One distinction is between “micro moments” and “macro moments”. Micro moments are constructed using microeconomic data on the behavior of individuals and

⁵This literature also adopted an informal approach to estimation which is often referred to as calibration. Calibration is simply informal GMM or SMM without standard errors. Macroeconomists tend to be more worried about “model error” than sampling error compared to researchers in some other subfields of economics (e.g., IO). We therefore tend to use a larger share of our computational budget on exploring the model space than on computing standard errors, relative to researchers in these other subfields.

⁶On this point, our argument is similar to Summers (1991).

⁷What is an example of a statistic that is not portable? The score of the likelihood function of a particular model (the moment 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.

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 discipline models designed to understand the effects of monetary policy on the economy. The fact that prices change infrequently provides evidence for the key friction in such models. 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. Another example is 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 inference 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 RBC literature (Kydland and Prescott, 1982; King and Rebelo, 1999), the Shimer puzzle literature (Shimer, 2005), the misallocation literature (Hsieh and Klenow, 2009), the literature on exchange rate disconnect (Meese and Rogoff, 1983; Itkhoki and Mukhin, 2017), and the literature on “wedges” (Chari, Kehoe, and McGrattan, 2008; Shimer, 2009).

2.2 Causal Effects as Identified Moments

A second distinction is between what we call “identified moments” and “unidentified moments”. Our idea in making this distinction is to contrast simple statistics such as means, variances, and correlations, with more complex statistics that are 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 in-

strumental 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, identified moments correspond directly to a particular deep structural parameter. For example, there is a large literature in labor economics that estimates the labor supply elasticity (see, e.g., [Chetty, 2012](#); [Chetty et al., 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 are estimates of the MPC 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 (e.g., the MPC out of transitory fiscal rebates) to the macroeconomic questions of interest (e.g. the aggregate effects of fiscal rebates). These types of identified moments can be used as empirical targets in a structural moment matching exercise aimed at distinguishing between competing models of the world.⁸

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. So, we might as well pick simple moments such as means, variances, and covariances. Why go through the trouble of constructing “identified” moments, that are both more complicated and more controversial (since they rely on identifying assumptions)?

The reason why identified moments may be of particular value in a structural moment matching exercise is that they can be designed to provide evidence on a specific causal mechanism of a model, and may be 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 \(2014\)](#) seek to shed light on these questions by comparing changes in consumption and employment in cities that experienced large house price changes with cities that experienced smaller house price changes. An important empirical challenge is that causation may run both ways: increases in house prices may stimulate economic activity, but a local boom may also increase house prices. The raw correlation between changes in house prices and economic

⁸Formally, the idea is to use indirect inference with limited information empirical models designed to estimate causal effects (e.g., 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.

activity will be due to both effects. To isolate the causal effect of house prices on consumption and employment, Mian et al. propose to instrument for changes in house prices with estimates of housing supply elasticities constructed by Saiz (2010). The idea for identification is that national shocks will lead house prices to increase more in cities where housing supply is less elastic.⁹

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 non-tradable 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 since Mian et al.’s empirical approach is based on comparing one city to another and therefore “differences out” aggregate general equilibrium effects.¹⁰ The empirical value of Mian et al.’s estimates is that they 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 (see, e.g., Berger et al. (2017) for a discussion of such a model). This latter class of models has quite different implications about the effects of macroeconomic policies such as fiscal and monetary policy—but also housing policy—than the former class of models.¹¹

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. A much simpler alternative empirical approach would be to consider instead the raw correlation between house prices and consumption or house prices and employment. However, a downside of focusing on these simple unidentified moments is that their value in a model is likely to be sensitive to assumptions that

⁹Saiz’s (2010) estimate of the housing supply elasticity is constructed based on data on city topology and land-use regulation. It is obviously not randomly assigned (for example, land availability is correlated with whether a city is on the coast). However, the argument for identification, like in many natural experiments, is that whatever makes these coastal (and otherwise land constrained) locations different does not affect their response to aggregate shocks directly, except for through the implications for the housing market.

¹⁰The the whole question of how house prices affect consumption and employment many strike some macroeconomists as being ill-posed. In general equilibrium, house prices are an endogenous variable. They respond to various exogenous disturbances, which may affect consumption directly (as opposed to only through house prices). However, since Mian et al.’s estimates focus on the cross-sectional relationship between house prices and economic activity and thereby difference out aggregate general equilibrium effects. In some simple cases, it is possible to show that this type of cross-sectional instrumental variables regression is equal to the partial equilibrium response of consumption to a change in house prices or the response to a foreign housing demand shock (see Guren et al., 2017). More generally, these estimates must be interpreted through the lens of a fully formed general equilibrium model.

¹¹Midrigan and Philippon (2016) use evidence similar to that of Mian and Sufi (2014) to pin down key parameters of a household liquidity model. They use the model to argue that household deleveraging cannot explain the large drop in employment during the Great Recession because its effect on the natural rate of interest is too small.

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 unidentified 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.¹²

Another prominent example of an identified moment matching exercise is recent work that seeks to determine the role of unemployment insurance (UI) extensions in delaying recovery from the Great Recession. During this recession, temporary federal programs raised the duration of UI benefits from 26 weeks to 99 weeks in some states, and an important policy question is the extent to which this led to reduced job search effort by households and reduced job posting by employers. [Hagedorn, Manovskii, and Mitman \(2015\)](#) and [Chodorow-Reich and Karabarbounis \(2016\)](#) estimate the response of unemployment to variation in UI benefit extensions across states (the former paper uses discontinuity based identification, while the latter uses an instrumental variables approach). 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 UI extensions will affect the economy at the aggregate level. Here, again, it is likely possible to pin down these same parameters using unidentified 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.

Notice that in the examples above, the identified moments give us information primarily on a particular “block” or mechanism of a macroeconomic model. This “piecemeal” form of inference will, therefore, result in partial identification of 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, this makes that statistic useful.

¹²Similar methods can be applied to assess the importance of financial frictions on the firm side. For example, [Catherine et al. \(2017\)](#) use of an IV estimate of the response of firm investment to changes in real estate collateral. They use this cross-section identified moment to pin down parameters in a structural model 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 the appendix to his paper, he uses these micro-level estimates pin down parameters in a general equilibrium model of the effect of bank shocks on the economy as a whole.

3 Aggregate Versus Cross-Sectional Identification

The increased use of cross-sectional identification approaches has been an exciting development in empirical macroeconomics over the past several years. In this work, researchers use geographically disaggregated panel data sets—often disaggregated to the state or metropolitan statistical area level—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’s \(2014\)](#) work on the role of the housing net worth channel in the Great Recession, [Autor, Dorn, and Hanson’s \(2013\)](#) work on the effects of Chinese imports on US employment over the past two decades, [Beraja, Hurst, and Ospina’s \(2016\)](#) work on wage rigidity during the Great Recession, [Martin and Philippon’s \(2017\)](#) work on the effects of debt on the Great Recession in the Eurozone, 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. This approach is typically invalid since it 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 (see, e.g., [Nakamura and Steinsson, 2014](#)).

Since an aggregate response is usually the object of most direct interest, a common critique of estimates based on cross-sectional identification is that they don’t answer the right question. While it is true that these estimates don’t *directly* answer the question of interest, they often provide a great deal of indirect evidence regarding this question. The way in which they do this is by helping researchers discriminate between different theoretical views of how the world works. In the language of section 2, 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 kind of 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.

3.1 Fiscal Stimulus: Aggregate Evidence

Direct aggregate evidence on the government spending multiplier is far from conclusive. This evidence largely comes in two forms: 1) evidence from wars, and 2) evidence from VARs. 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 comes from the major wars in their sample (WWI, WWII, 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 assume that war-related defense spending is exogenous to output. Conceptually, this is like using wars as an instrument for spending. The strength of this approach is that reverse causation is not likely to be a problem. WWI, WWII, and the Korean War did not happen because the US was in a recession or a boom. However, for war-related spending to be exogenous, it must be that wars only affect output through spending. This is unlikely to be true. Barro and Redlick are aware of this and discuss some of the main potential confounding factors. Two important confounding factors are patriotism and war-time controls on the economy such as rationing and price controls. Patriotism likely increased labor supply during major wars such as WWI and WWII. It therefore potentially results in an upward bias of the multiplier. On the other hand, war-time controls such as rationing and price controls likely result in a downward bias. The overall direction of bias is unclear. Barro and Redlick argue that patriotism is likely the dominant bias, while Hall (2009) argues that the effects of war-time 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 1975:2.¹³ They view the residual from this regression as exogenous shocks to government spending and construct an impulse response function for output, consumption, and other variables to the government spending shocks by iterating forward a VAR.¹⁴ Their baseline sample period is 1960-1997. They highlight a peak

¹³The dummy variable is included due to a large temporary tax cut in 1975:2. They allow for quarter dependence of the coefficients in their VAR. They also consider a version estimated in differences.

¹⁴This is equivalent to performing a Cholesky decomposition of the reduced form errors from the VAR with government spending ordered first.

response of output of 1.29 times the initial response of spending. This response occurs 15 quarters after the impulse (after the VAR response of output has swung up, then down, then back up). Unfortunately, they do not report an estimate of the cumulative response of output divided by the cumulative response of spending, which would be more informative.¹⁵

Blanchard and Perotti's approach to identification makes the strong assumption that controlling for four lags of taxes, spending, and output eliminates all endogenous variation in spending. (See section 4 for a more detailed discussion of VAR identification.) They argue that this approach is more likely to be valid than in the case of 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—arrive well ahead of the main increase in spending.¹⁶ A different concern with Blanchard and Perotti's estimates is lack of precision. Their results come with standard errors so large that few interesting hypotheses can be rejected.¹⁷ Also, their estimates are highly sensitive to the sample period and to which controls are included (Gali et al., 2007). For example, including the Korean War yields substantially lower estimates.

For the purposes of guiding policy, the most serious weakness of the literature discussed above is, however, arguably the fact that these estimates are not very informative about the effectiveness of fiscal stimulus in a deep recession when monetary policy is constrained by the zero lower bound on nominal interest rates. In normal times, monetary policy “leans against the wind” by raising real interest rates in response to expansionary shocks such as a government spending shock. Tax policy has at times also responded aggressively to government spending shocks (e.g., during the Korean war). The aggregate fiscal multiplier is potentially quite different in a setting in which monetary and tax policy are more accommodative, such as in the midst of a large recession. Since fiscal stimulus packages tend to be discussed exactly in such large recessions, evidence on the government spending multiplier for this circumstance is particularly valuable from a policy perspective.

Direct aggregate evidence on the effectiveness of government spending when monetary policy

¹⁵Ramey (2016) and Ramey and Zubairy (2017) provide multiplier estimates of this kind for the Blanchard and Perotti shocks as well as novel defense news shocks for a somewhat longer simple period. These estimates are substantially below one.

¹⁶The wealth effect associated with a fiscal shock should occur at the time of announcement. Theory suggests that an empirical approach that fails to capture this part of the effect of the fiscal shock will yield a response of output that is downward biased, but a response of consumption that is upward biased.

¹⁷Readers should be aware that it is standard in the fiscal VAR literature to report one standard error bands, presumably because two standard error bands are very wide.

is constrained is, as one would expect, even less well established than on the simpler question of the average aggregate multiplier (see [Ramey and Zubairy, 2017](#); [Miyamoto et al., 2017](#)). An indirect way to make inference about this question is, however, 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 is highly dependent on the responsiveness of monetary policy. If monetary policy is responsive, the aggregate multiplier can be quite low. But in circumstances when monetary policy is unresponsive—e.g., at the zero-lower-bound—the aggregate multiplier can be quite large. In contrast, in the Neoclassical model, the aggregate multiplier is small independent of monetary policy. Unfortunately, the evidence on the aggregate fiscal multiplier is not very helpful in this regard. In particular, 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.

3.2 Fiscal Stimulus: Cross-Sectional Evidence

Since the Great Recession, there has been an explosion of work that estimates the effects of government spending in the cross-section. Researchers in this literature have used a wide array of—in some cases very creative—identification strategies. 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 et al., 2014](#)), formulas used to allocate spending across states from the American Recovery and Reinvestment Act ([Chodorow-Reich et al., 2012](#); [Wilson, 2012](#); [Dupor and Mehkari, 2016](#)), and spending discontinuities associated with decadal population estimate revisions ([Suarez Serrato and Wingender, 2016](#)). [Chodorow-Reich \(2017\)](#) surveys this literature in detail.

Estimates of the regional spending multiplier from this literature tend to cluster in the range of 1.5-2.0. They are therefore substantially larger than typical estimates of the aggregate multiplier. This does not, however, mean that these two sets of estimates are inconsistent. After all, they measure different things. To understand this better, let's consider the identification strategy we used in [Nakamura and Steinsson \(2014\)](#). The basic idea in our identification strategy is that national military buildups (e.g., 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. This is because the plants that build the hardware used by the military 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 at these times.¹⁸ Our conclusion is that output in California rises by roughly \$1.5 relative to output in Illinois for each extra dollar of spending in California relative to Illinois.

Importantly, our specification includes time fixed effects implying 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 all 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 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.

It turns out, however, that the regional multiplier a powerful diagnostic tool for distinguishing between competing macroeconomic models, and thereby 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 (i.e., regional spending shocks financed with federal taxes), and compare the resulting regional multiplier estimates generated by each model with the regional multiplier estimate we estimate in real-world data. We find that both the textbook RBC model and the textbook New Keynesian model generate regional multipliers that are substantially smaller than our empirical estimate. The reason is that in these models a foreign demand shock (the federal government demanding military goods), leads people to cut back on work effort in other areas. The output response to such a shock is therefore less than one. We conclude that the regional multiplier evidence favors models in which output responds more strongly to a foreign demand shock than either the textbook RBC or New Keynesian models imply. We present an example of such a model and show that in this model the aggregate multiplier can be large (e.g., when monetary policy is constrained at the zero lower bound). The regional multiplier obviously does not point identify the correct model (no single statistic will). However, estimates of it should result in a posterior over models that puts less weight on a large class of models in which demand shocks have relatively small effects on output.

¹⁸We regressed changes in regional output on changes in regional prime military contract spending controlling for region fixed effects and time fixed effects. We instrumented for changes in military spending using national prime military contract spending interacted with a region dummy.

4 Monetary Policy: What Is the Best Evidence We Have?

What is the most convincing evidence we have for monetary non-neutrality? When we have asked prominent macroeconomists this question, the most common answers have been: [Friedman and Schwartz \(1963\)](#), the Volcker disinflation, and [Mussa \(1986\)](#).¹⁹ [Friedman and Schwartz \(1963\)](#) famously argued that the severity of the Great Depression was due to bad policy by the Federal Reserve. They also point to a doubling of reserve requirements by the Federal Reserve between June 1936 and January 1937 and sterilization of gold inflows by the Treasury as a major cause for the sharp recession of 1937. Paul Volcker dramatically tightened monetary policy after taking office as chairman of the Federal Reserve Board in August 1979. These actions—which lasted well into 1982 (with a brief easing in the spring of 1980)—are frequently cited as a major cause of the large twin recessions that occurred in 1980 and 1981-1982. Finally, international economists often point to 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 \(1986\)](#)—as dramatic evidence of monetary non-neutrality.

The fact that monetary economists point to these three pieces of evidence as most convincing is interesting and informative regarding the types of empirical methods that are influential in macroeconomics. Two of these three pieces of evidence are large historical events often cited without reference to modern econometric analysis: the Great Depression and the Volcker disinflation. The third is essentially an example of discontinuity-based identification. Conspicuous by its absence is any mention of evidence from Vector Autoregressions (VARs) even though such methods have dominated the empirical literature for quite some time. Clearly, there is a disconnect between what monetary economists find most convincing and what many of them do in their own research.

4.1 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”, i.e., situations in which we can argue that the change in policy is large relative to potential confounding factors that cannot be controlled for. Much empirical work takes the approach of seeking to control for confounding factors as well as possible. A different approach is

¹⁹Of course, a significant fraction say something along the lines of “I know it in my bones that monetary policy has no effect on output.”

to focus on large policy actions for which we can (hopefully) transparently argue that confounding factors are drowned out. Such policy actions are of course few and far between. But looking over a long period of history and many countries, one may be able to find events like these that constitute convincing pieces of evidence on the effects of monetary policy.

Friedman and Schwartz (1963, p. 688) argue in the summary of their monumental work on U.S. 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.

The Great Depression is by a wide margin the largest economic downturn in the United States in the last 150 years (if not ever). Figure 1 plots the evolution of industrial production from 1925 to 1942. Industrial production peaked in July of 1929. It then fell almost continuously until March 1933. The cumulative fall in industrial production was as staggering 53%. The recovery between March 1933 and the subsequent peak in May 1937 was rapid and large. Industrial production rose at an annual rate of 8.5% per year during that time to a level above the 1929 peak. But then a second very large downturn occurred. From May 1937 to May 1938 industrial production fell by 33%.

We have indicated with light blue vertical bars in Figure 1 the times of the two policy mistakes Friedman and Schwartz highlight. The first of these is October 1931. At this time, the Fed raised the rediscount rate sharply from 1.5% to 3.5% 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 at this time even despite the fact that industrial production was in free fall and a wave of bank failures was underway. At first pass, this may seem like a very clean monetary shock. But the difficulty with it is that 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 fall in industrial production between October 1931 and March 1933 is due to this policy shock as opposed to other developments that led to the the equally rapid fall in the two years before October 1931.

The second monetary shock emphasized by Friedman and Schwartz is more promising in this regard. During the period July 1936-January 1937 the Fed announced a doubling of reserve requirements (fully implemented by May 1937) and the Treasury engaged in sterilization of gold

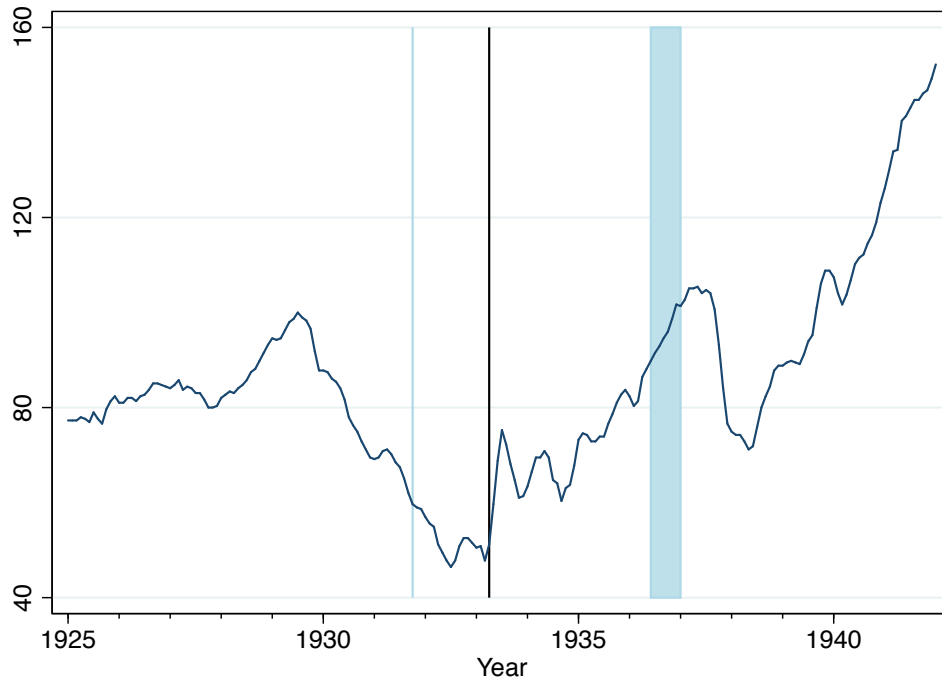


Figure 1: Industrial Production from 1925 to 1942

Note: The figure plots an index for industrial production in the US from January 1925 to January 1942. The index is equal to 100 in July 1929 (the peak month). The light blue shaded area and vertical line are the periods of policy mistakes identified by [Friedman and Schwartz \(1963\)](#). The black vertical line is the time at which Roosevelt took the US off the gold standard.

inflows. Before this period, industrial production had been rising rapidly. Shortly after, however, it plunged dramatically by 33%. Friedman and Schwartz argue that the increase in reserve requirements and sterilization of gold inflows caused a sharp slowdown in money creation, which caused the large 1937 recession. A closer look however reveals important confounding factors. Fiscal policy was tightened sharply in 1937.²⁰ In fact, prior to Friedman and Schwartz’s 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.”

There is a more general argument that runs through Friedman and Schwartz’s whole narrative of the Great Depression period. They argue that throughout the period between early 1930 and March 1933, the Fed failed to act and instead allowed the money supply to fall by large amounts

²⁰The budget balance swung substantially towards surplus because of the end of the 1936 veterans’ bonus and the first widespread collection of social security payroll tax in 1937 among other factors.

and allowed a substantial fraction of the banking system to fail.²¹ [Eichengreen \(1992\)](#) argues that an important reason why the Fed did not act was that it felt constrained by the gold standard. The idea is that effective action almost surely would have been inconsistent with remaining on the gold standard and those that favored remaining on the gold standard prevailed until Roosevelt took office. One of President Roosevelt's first policy actions was to take the US off the gold standard in April 1933. The dollar quickly depreciated by 30%. The black vertical line in [Figure 1](#) indicates April 1933. Industrial production immediately sky-rocketed. The weakness of this observation is that the Roosevelt administration, of course, changed a number of policies. It seems clear that the switch from Hoover to Roosevelt mattered. Whether it was going off gold or something else that made the difference is, however, not entirely clear.²²

The Great Inflation of the late 1970's and early 1980's is another landmark event in US macroeconomic history. After having been low and stable since the end of the Korean War in 1953, inflation started to rise in the late 1960's. In the 1970's, inflation was both higher and much more volatile than before (see [Figure 2](#)). In 1974, President Ford declared inflation—then running at around 10% per year—“public enemy number one”. Inflation subsequently fell back to about 5% per year. But in 1978 and 1979 it rapidly rose back into double digits. Monetary policy during the 1970's 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 the Chairman of the Federal Reserve Board, after having been appointed by President Carter. Under the leadership of Volcker, the Fed first conducted policy such that interest rates rose dramatically between October 1979 and March 1980. However, 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 (see [Figure 2](#)). [Goodfriend and King \(2005\)](#) argue that this first year under Volcker was therefore not very different from earlier stop-go attempts by the Fed to reign in inflation. 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

²¹They trace this failure to act to a power struggle within the Fed between the New York Fed that wanted to act and other parts of the Fed that favored inaction. They conclude that if Benjamin Strong—the powerful President of the New York Fed, who died in 1928—had lived the Fed might have ended the depression in 1930.

²²A related piece of evidence that supports the case for going off gold as being crucial is that [Eichengreen and Sachs \(1985\)](#) 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.

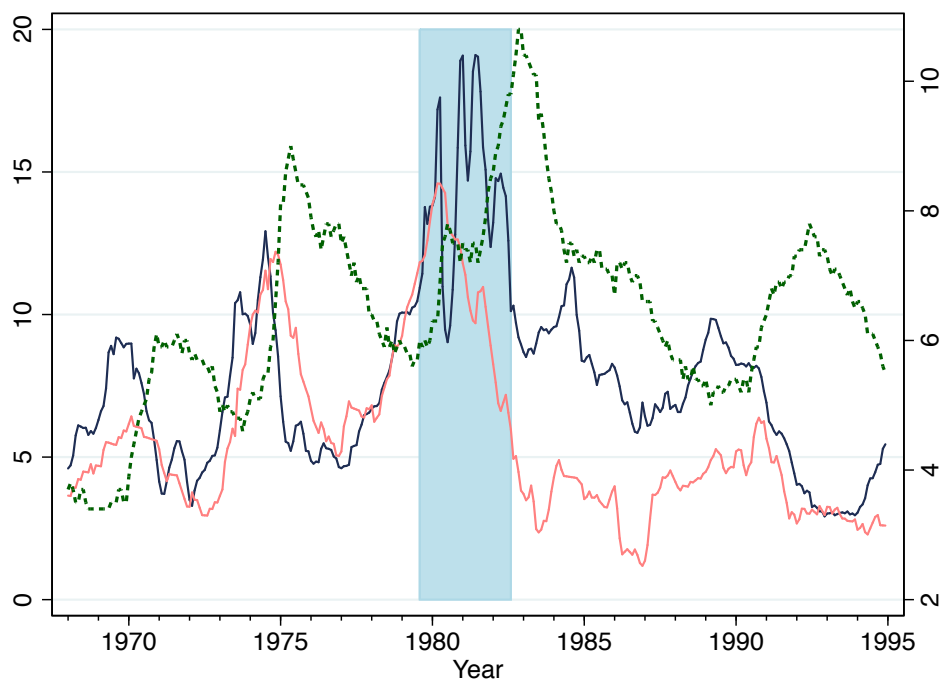


Figure 2: Federal Funds Rate, Inflation, and Unemployment from 1965 to 1995

Note: The figure plots the federal funds rate (solid blue, left axis), the 12-month inflation rate (solid pink, left axis), and the unemployment rate (dashed green, right axis). Volcker disinflation period shaded in blue (August 1979 to August 1982).

Volcker truly broke with prior behavior of the Fed and embarked on a sustained, deliberate disinflation. He conducted policy such that rates rose dramatically to close to 20% and kept policy tight even in the face of the largest recession the US had experienced since the Great Depression. Only when inflation had fallen to 5% did Volcker let up and allow interest rates to drop substantially.

The behavior of output during this period is consistent with the view that monetary non-neutrality 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. Many economists find the narrative account above and the accompanying evidence about output to be compelling evidence of large monetary non-neutrality.²³ However, there are other possible explanations for these movements in output.

²³The 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. The idea that inflation is “always and everywhere a monetary phenomenon” was a proposition that (surprisingly to modern ears) was doubted by many in the 1970’s (Romer and Romer, 2002; Nelson, 2005). Estimates of the “sacrifice ratio” associated with bringing down inflation—i.e., the cost in terms of output of a permanent 1 percentage point reduction in inflation—were huge in the late 1970’s (Okun, 1978). Also, the

There were oil shocks both in September 1979 and in February 1981 (see Table A.1 in the appendix). Credit controls were instituted between March and July of 1980. It may even be that anticipation effects associated with the phased-in tax cuts of the Reagan administration played a role in the 1981-1982 recession (Mertens and Ravn, 2012).

While the Volcker episode is consistent with a large amount of monetary non-neutrality, it seems much 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 compelling is that output fell and then rose and then fell again 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.

4.2 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.²⁴ 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 (comparing two year periods before and after). The switch from a fixed exchange rate to a flexible exchange rate is a purely monetary action. In a world in which monetary policy has no real effects, such a policy change will not affect real variables such as 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 being made 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. The root cause was persistently more inflationary macroeconomic policy in the United States than in Germany, Japan, and other countries. By early 1973, the US exchange rate was overvalued and the US had a substantial current account deficit, while Germany and other countries had substantial surpluses. There were intense negotiations about the future of the idea that credible announcements to disinflate can reduce these costs was not well understood or widely believed.

²⁴The Bretton Woods system officially broke down on March 1st 1973. However, we date the de facto breakdown as February 1973 because the exchange rate of the dollar vis-a-vis major European countries and Japan was allowed to move substantially in that month in response to mounting pressure. For example, on February 12th 1973 the US devalued the dollar by 10%. See Bordo, Humpage, and Schwartz (2011) for a detailed discussion of the breakdown of the Bretton Woods system.

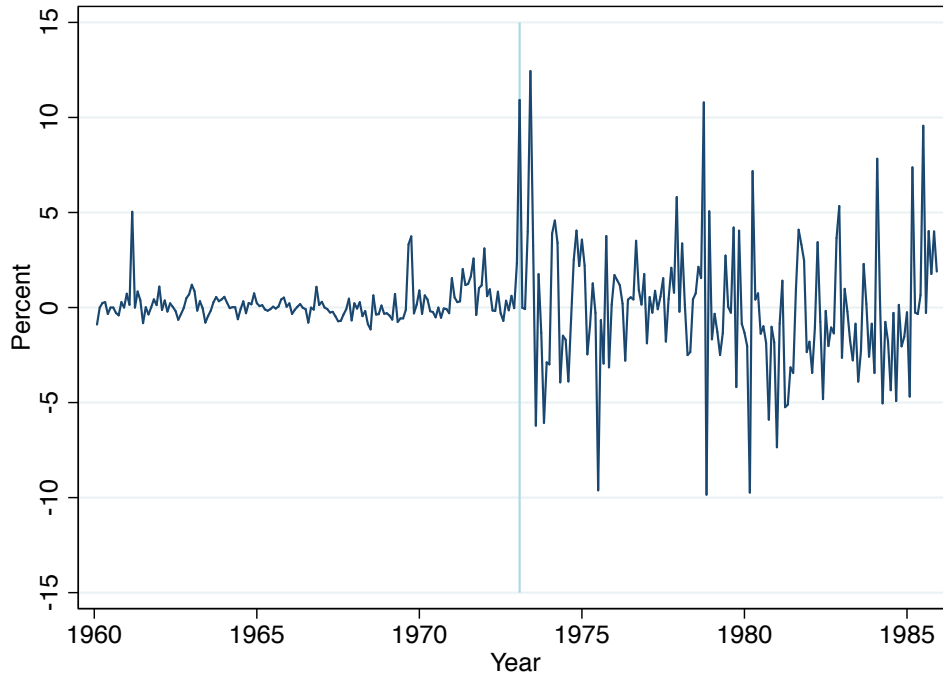


Figure 3: Monthly Change in the US-German Real Exchange Rate

system for over a year before it finally collapsed. It is hard to think of plausible alternative candidate explanations for the *discontinuous* increase in the volatility of the US real exchange rate that occurred at this time.²⁵

There is also strong evidence that monetary policy affects real interest rates. This can be established at very high frequencies using the response of the yields on Treasury Inflation Protected Securities (TIPS)—which provide a measure of relatively risk-free real interest rates—to monetary announcements. Importantly, 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. This allows for a discontinuity-based identification scheme.

In Nakamura and Steinsson (2017), 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.²⁶ Furthermore, 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

²⁵Velde (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. He also presents evidence on output. As with more modern evidence, Velde’s on output is less conclusive.

²⁶We verify this using a heteroskedasticity-based identification scheme building on the work of Rigobon (2003) and Rigobon and Sack (2004).

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 3-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, Jorda, and Kuersteiner, 2017](#)). The reason for this 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 (since many other shocks affect output over such a long time period). However, the effect of monetary shocks on output can be broken into two separate questions: 1) How much do monetary shocks affect various relative prices? and 2) How much do these various relative prices affect output? Viewing things this way makes clear that it is the first of these two questions 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. Only in some models does monetary policy affect relative prices, and these models typically imply that money is non-neutral. The use of discontinuity-based evidence regarding the effect of monetary shocks on relative prices to make inference about monetary non-neutrality is therefore a good example of the identified moment matching approach we discuss earlier in this article.

A complicating factor regarding the high-frequency evidence from FOMC announcements discussed above 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 Chair reveals that (s)he is more optimistic about the economy than the private sector had anticipated, this may lead the private sector to revise its own beliefs about where the economy is headed. In fact, we find evidence for such “Fed information effects” in [Nakamura and Steinsson \(2017\)](#). 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 would imply).²⁷ We show that a New Keynesian model that incorporates Fed information effects can match this fact as well as the facts about real interest rates and expected inflation. In this model, Fed tightenings have two effects on the economy: traditional contrac-

²⁷See [Romer and Romer \(2000\)](#), [Faust, Swanson, and Wright \(2004\)](#), and [Campbell et al. \(2012\)](#) for earlier evidence of Fed information effects.

tionary effects through increases in real interest rates relative to natural rates and a less traditional expansionary effects coming from the Fed's ability to increase optimism about the economy.

The existence of this Fed information effect implies that there is an "external validity" problem whenever researchers try to use responses to monetary shocks to make inference about the effects of systematic monetary policy actions. Surprise monetary actions lead to information effects. Changes in policy that are expected because they are part of the systematic response of the Fed to developments in the economy do not lead to information effects. This implies that the responses to monetary shocks estimated in the literature likely underestimate the effect of systematic monetary policy (but less so in periods when information effects were less likely to be important).

4.3 Using the Narrative Record to Identify Shocks

Identification involves the search for natural experiments. [Romer and Romer \(1989\)](#) argue that contemporaneous Federal Reserve records can be used to identify such natural experiments. Specifically, 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-WWII period. They identify six such times and have subsequently added a seventh ([Romer and Romer, 1994](#)). [Figure 4](#) plots the US unemployment rate with the seven Romer-Romer dates marked by vertical lines. After each of the Romer-Romer dates, 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 very valuable, it faces several challenges. One challenge is the inherent opacity of the process by which the narrative shocks are selected. This raises the concern that data are (perhaps unconsciously) reverse-engineered to generate favored conclusions.²⁸ Another challenge is that with only seven data points, it may happen by chance to be the case that some other factor is correlated with the monetary shocks. In cases when one has dozens or hundreds of shocks that are plausibly orthogonal to other factors, 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 A.1](#) in the appendix). A third concern that is often raised with narrative analysis is that the shocks might be

²⁸Clearly, this concern applies to all research. But it applies with particular force to narrative analysis because of the high costs associated with attempting to replicate such analysis. In general, the reverse-engineering concern favors using simple, transparent, empirical strategies rather than more complicated strategies in which researchers can make various seemingly innocuous decisions without raising concerns on the part of the reader.

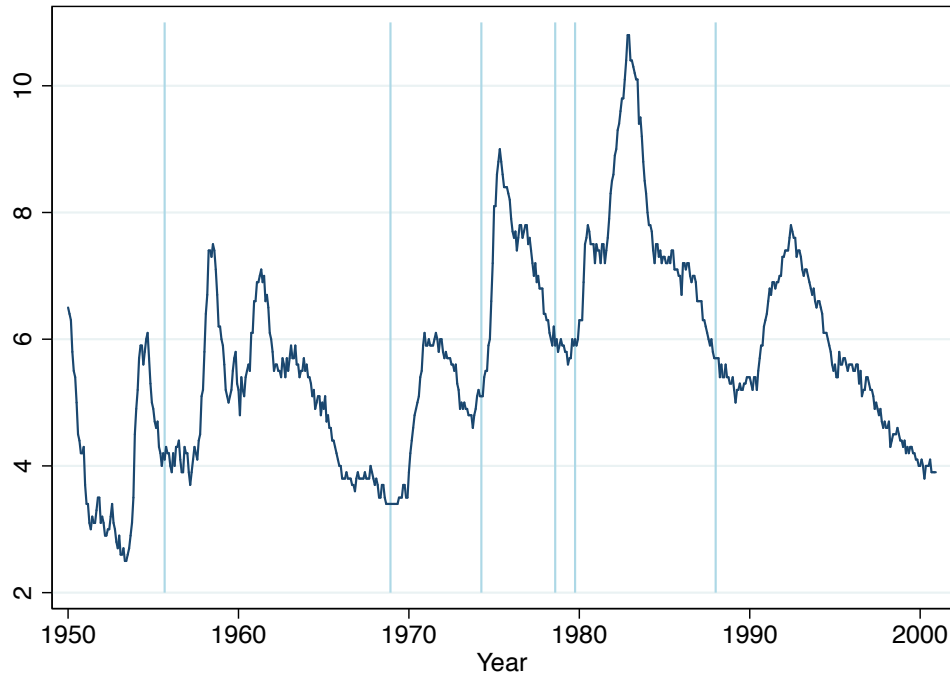


Figure 4: Unemployment from 1950 to 2000

Note: The figure plots the unemployment rate from 1950 to 2000. The light-blue vertical lines indicate Romer-Romer dates.

predictable suggesting endogeneity. In the case of [Romer and Romer's \(1989\)](#) analysis, this type of concern was raised by [Shapiro \(1994\)](#) and [Leeper \(1997\)](#). However, it is 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 and is hard to estimate. Even by chance, some of these should turn up estimates that are statistically significant based on standard significance cutoffs.²⁹

4.4 Controlling for Confounding Factors

By far, the most prevalent approach to identifying exogenous variation in monetary policy is to attempt to control for confounding factors. This is the approach taken in much of the VAR literature. A common specification is to regress the federal funds rate on contemporaneous values of several variables (output, inflation, etc.), as well as several lags of itself and these other variables and to view the residual from this regression as exogenous monetary policy shocks. A common

²⁹In the case of Leeper's 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 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.

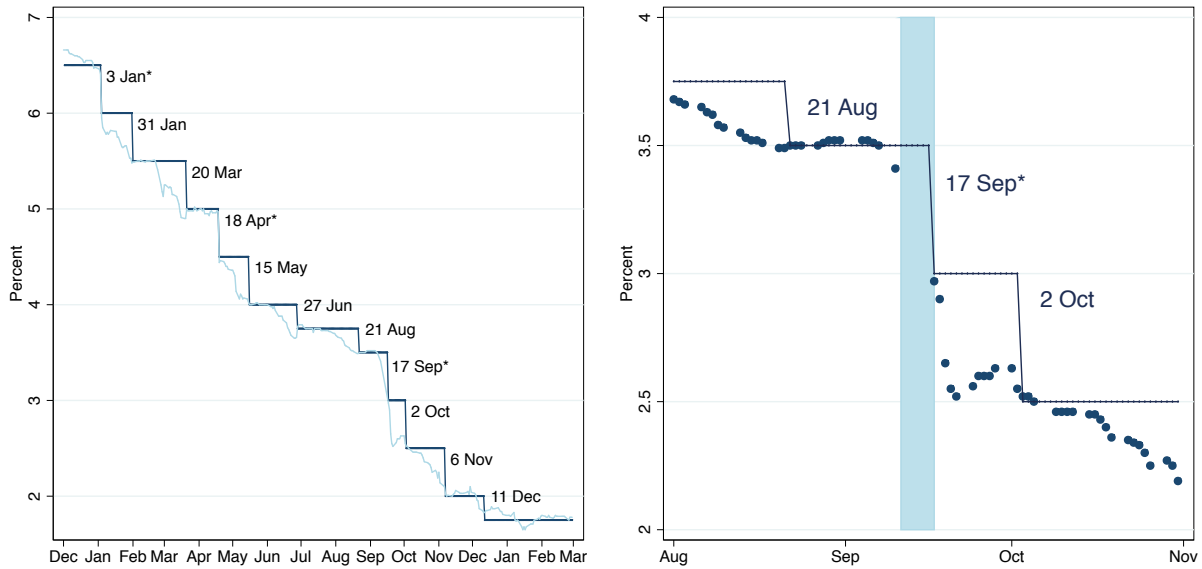


Figure 5: Federal Funds Rate Target and 1-Month Eurodollar Rate

Note: The figure plots the federal funds rate target (blue line) and the 1-month Eurodollar rate (red line in left panel and blue dots in right) at a daily frequency (beginning of day). Dates of changes in the federal funds rate target are indicated in the figure. The dates marked with a * are occurred at unscheduled FOMC conference calls. The other dates are scheduled FOMC meetings. The left panel is very similar to the top panel of Figure 1 in [Cochrane and Piazzesi \(2002\)](#).

way of describing this procedure is as performing a Cholesky decomposition of the reduced form errors from the VAR with the federal funds rate ordered last.

This approach to identifying monetary policy shocks is often describes as involving “minimal identifying assumptions.” In our view, however, these are very strong assumptions. 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. Stress in the banking system affects policy in some periods. In other periods, sharp changes in commodity prices affect policy. In yet other periods, a recent stock market crash affects policy, or a financial crisis in emerging markets, or terrorist attacks, or temporary investment tax credits, or 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 to include them all in a regression without risking massive over-fitting concerns or literally run out of degrees of freedom. Any one of these factors that is left 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 concretely demonstrate this point, it is useful to 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 1-month Eurodollar rate—i.e., a 1-month interbank interest rate on US dollars traded in London—over the period December 2000 to February 2002. The Eurodollar rate on any given day may be viewed as the average expected federal funds rate over the next month. On September 10th 2001, the Eurodollar rate traded at 3.41%, quite close to the target federal funds rate of 3.5%. Markets did not open on September 11th 2001 due to the terrorist attacks in New York and stayed closed until Monday September 17th 2001. Before markets opened on that day, the Fed announced a 50bp drop in the target federal funds rate to 3% and the 1-month Eurodollar rate traded at 2.97% that day. The futures contract for the federal funds rate in September 2001 reveals that this 50bp drop in the federal funds rate was completely unanticipated by markets as of September 10th.³⁰ The easing on September 17th 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.

How do standard monetary VARs treat 9/11? As an exogenous monetary shock. The reason is that the controls in the policy equation of the VAR 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 VAR). From the perspective of this equation, therefore, the drop in interest rates in September 2001 looks like an exogenous easing of policy.³¹ Any unusual (from the perspective of the VAR) 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.

The way in which identification assumptions are commonly discussed in the VAR literature is very misleading, in our view. It is common to see “the” identifying assumption in a monetary VAR described as the assumption that the federal funds rate does not affect output and inflation contemporaneously (or, in the fiscal multiplier case, that government spending is not affected by contemporaneous changes in output). These assumptions sound quite innocuous. To an outsider, the whole thing sounds like magic: You make one pretty innocuous assumption, and voilà, you can estimate the dynamic causal effects of monetary and fiscal policy. We remember finding this deeply puzzling when we were starting off in the profession.

³⁰The 1-month Eurodollar rate was trading below the target for the federal funds rate on September 10th because markets were anticipating an easing of 25-50bp at the regularly scheduled FOMC meeting on October 2nd. When markets opened on September 17th, markets anticipated an easing of an additional 25-50bp on October 2nd (over and above the 50bp that had just occurred).

³¹We have verified this. The monthly VAR used by [Coibion \(2012\)](#) yields a monetary shock of -48bp in September 2001.

In fact, the timing assumption that is usually emphasized is not the only identifying assumption being made in a standard monetary or fiscal VAR. 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 reflects a causal influence of one on the other or the reverse. This is what the timing assumption does.

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 VAR, i.e., omitted variables bias. The typical VAR 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.

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 forecasts (the so called Greenbook forecasts). The logic for this 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. It is useful to be concrete. Suppose we are interested in the effect of a change in monetary policy at time t , denoted Δr_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 Δr_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 $\Delta^j y_{t+j}$. 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 $\Delta^j y_{t+j}$.³²

Romer and Romer's approach helps answer the question: What is a monetary shock? As [Cochrane \(2004\)](#) notes, 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 $\Delta^j y_{t+j}$. These are endogenous, when $\Delta^j y_{t+j}$ is the dependent variable. But others are in response to other things. Examples include time variation in policy makers' preferences and goals (e.g., distaste for inflation), time variation in policy makers' beliefs about how the economy works, political

³²This discussion is based on [Cochrane \(2004\)](#).

influences, pursuit of other objectives (e.g., exchange rate stability), etc.³³ 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.³⁴

4.4.1 What Do We Do with These Shocks?

Any exercise in dynamic causal inference involves two conceptually distinct steps: 1) the construction of the shocks, 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 potentially very important. A specification that imposes minimal structure (apart from linearity) is to directly regress the variable of interest ($\Delta^j y_{t+j}$, say) on the shock, perhaps controlling for some pre-treatment variables (i.e., variables determined before date t). This is the specification advocated by [Jorda \(2005\)](#). To construct an impulse response using this approach, one must run a separate regression for each horizon j that one is interested in plotting.

Standard VARs construct impulse response functions using a very different approach that imposes much more structure: they use the estimated dynamics of the entire VAR system to iterate forward the response of the economy to the shock. This method for constructing impulse responses embeds a whole new set of quite strong identifying assumptions. In a standard monetary VAR, 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, i.e., the whole model being correctly specified.

It is well known that the solution to any linear rational expectations model can be represented by a VAR. This idea is the usual defense given regarding the reasonableness of VAR impulse response construction. However, to estimate the true VAR, all state variables in the economy must be observable (so that they can be included in the VAR). If this is not the case, the VAR is misspecified and the impulse responses that it yields are potentially biased. 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 VAR(p) system into a VARMA(∞, ∞) sys-

³³The main threat to validity for this approach is that some variable affects both Δr_t and $\Delta^j y_{t+j}$ conditional on the Greenbook forecast. If the Greenbook forecast at time t is an efficient forecast of $\Delta^j y_{t+j}$, no such variable can exist (since no variable forecasts $\Delta^j y_{t+j}$ conditional on the Fed's forecast, in that case). If the Fed's forecast is inefficient, such a variable may exist. However, for the variable to be a problem, it must influence both Δr_t and $\Delta^j y_{t+j}$ conditional on the Greenbook forecast.

³⁴Changes in policy that are forecastable (e.g., forward guidance) are more complicated to analyze since they can have effects both upon announcement and when they are implemented.

tem in the remaining variables. The estimation of VARs therefore relies on the assumption that the true VARMA(∞, ∞) in the variables the researcher intends to include in his/her analysis can be approximated with a VAR(p). This is a strong assumption that we fear is unlikely to hold in practice.³⁵

A recent innovation in dynamic causal inference is the use of “external instruments” in VARs (Stock and Watson, 2012; Mertens and Ravn, 2013; Stock and Watson, 2017). Gertler and Karadi (2015) use this method to estimate the effects of exogenous monetary shocks on output, inflation, and credit spreads. Their measure of exogenous variation in monetary policy is the surprise movement in the 3-month-ahead fed funds rate futures contract in a 30-minute window around FOMC announcements. They then run a monthly VAR and instead of viewing the reduced form error in the interest rate equation as the monetary shock, they regress all the reduced form residuals on their “external” monetary policy shock. This yields a vector of contemporaneous responses of the variables in the VAR to the monetary shock (non of which is constrained to be zero). Finally, they construct an impulse response by iterating forward the VAR dynamics starting with this contemporaneous response. Figure 6 reproduces Figure 1 from Gertler and Karadi (2015). The column on the left gives impulse responses using external instruments, while the column on the right uses a standard Cholesky decomposition (see the figure note for details).

An important advantage of the external instruments approach relative to standard monetary VARs that rely on a Cholesky ordering is that it is possible to include fast-moving financial variables—such as stock prices, exchange rates, and credit spreads—in the VAR. In a standard monetary VAR, one must make a stark choice regarding the direction of contemporaneous causation of each variable with the policy variable: either it runs only from the policy variable to the variable in question or it runs only the other way. This is clearly unsatisfactory for variables such as stock prices, exchange rates, and credit spreads. The notion that the contemporaneous values

³⁵Problem set for interested readers: Consider the following simple New Keynesian model. Phillips curve: $\pi_t = \beta E_t \pi_{t+1} + \kappa(y_t - y_t^n)$. Aggregate demand equation: $\Delta M_t = \Delta y_t + \pi_t$. Monetary shock: $\Delta M_t = \rho \Delta M_{t-1} + \epsilon_t$. Productivity shock: $y_t^n = \eta_t$. Assume ϵ_t and η_t are i.i.d. normal. Set the parameters to $\beta = 0.99$, $\kappa = 0.13$, $\rho = 0.8$, $\sigma_\epsilon = 0.00066$, $\sigma_\eta = 0.007$ (the last two are chosen so that the contribution of ΔM_t and y_t^* to the variance of y_t is equal). 1) Show that the solution for output in this model takes the form $y_t = ay_{t-1} + b\Delta M_{t-1} + c\epsilon_t + d\eta_t$ (hint: method of undetermined coefficients). 2) Calculate the true impulse response of output to a monetary shock. 3) Simulate 500 time series from the model each of length 500 data points. Estimate the following three misspecified empirical models for output for each of these series: a model with the contemporaneous monetary shock and one lag of output, a model with the contemporaneous monetary shock and four lag of output, a model with the contemporaneous monetary shock and twelve lag of output. Plot the median impulse response for each misspecified empirical model. Notice that adding 12 lags of output doesn’t help at all in matching the true impulse response even though the true impulse response “looks like” an AR(2). 4) Do the same for a model with one lag of output, the contemporaneous monetary shock, and 6 lags of the monetary shock. Notice that this matches the true impulse response out to horizon 6 (but is biased after that). 5) Finally, do the same using the Jorda specification (with or without controls). Notice that this matches the true impulse response almost perfectly.

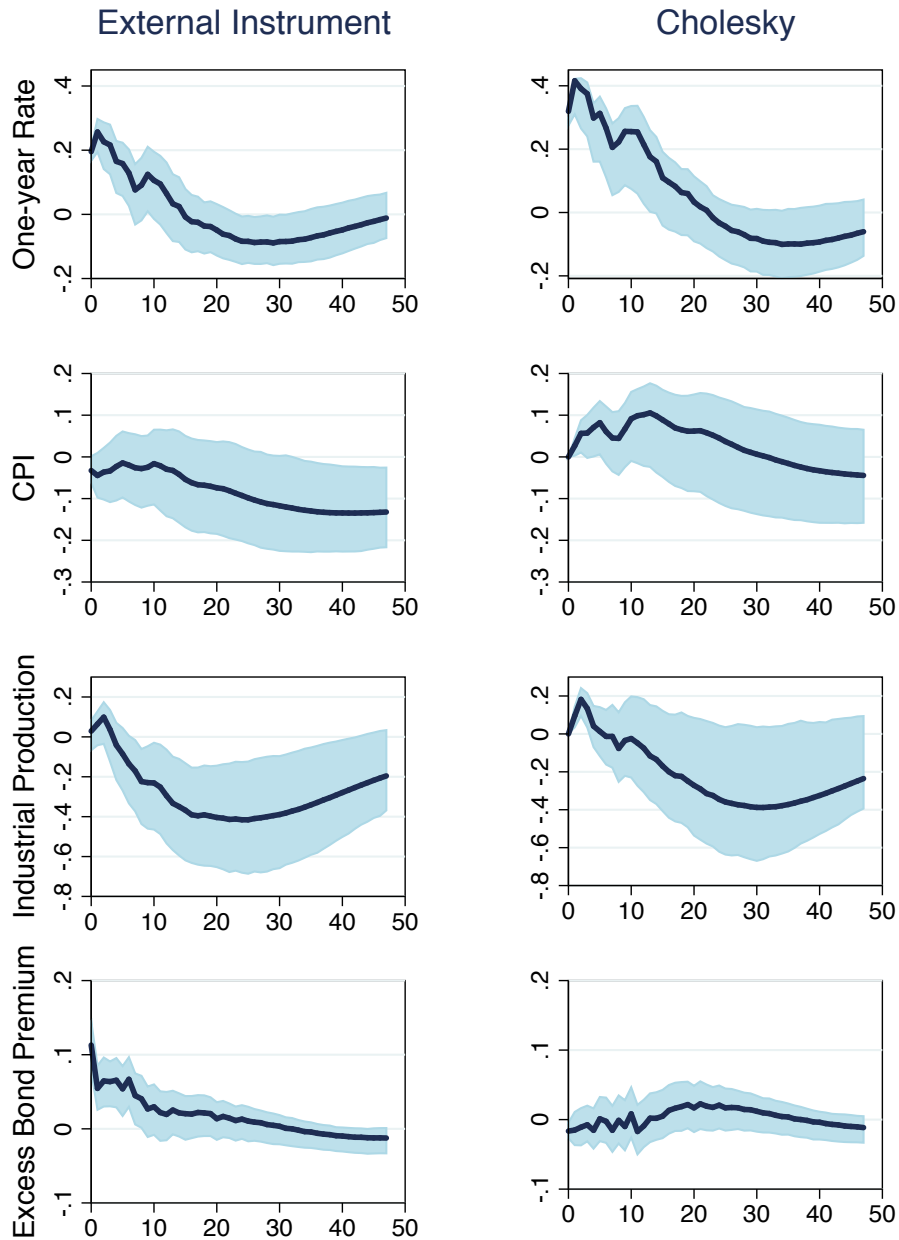


Figure 6: Responses to Monetary Shocks from [Gertler and Karadi \(2015\)](#)

Note: This figure replicates Figure 1 in [Gertler and Karadi \(2015\)](#). The figure plots the response of the one-year Treasury bond yield, the CPI, industrial production, and a measure of the excess bond premium in response to monetary policy shocks identified in two different ways. The left column uses external instruments to identify the contemporaneous response to the monetary shocks. The right column uses a Cholesky decomposition with the one-year yield considered the policy instrument and this variable ordered second to last with the excess bond premium ordered last. In both cases, the responses after the initial period are calculated by iterating forward a VAR with these four variables.

of stock prices, exchange rates, and credit spreads do not contain useful information about the endogenous component of monetary policy (even conditional on the other variables in the VAR) is highly dubious. This would suggest including them in the policy equation so as to get a cleaner measure of exogenous policy actions. However, if this is done, one cannot use the VAR dynamics to iterate forward the impulse response without assuming that these variables do not react to exogenous policy actions contemporaneously, which is inconsistent with much evidence (see, e.g., [Gurkaynak, Sack, and Swanson, 2005](#); [Bernanke and Kuttner, 2005](#), for stock prices).

However, while VARs identified with external instruments relax the Cholesky timing assumptions for contemporaneous responses, they do not relax the assumptions embedded in using the VAR system to construct the impulse response. This methodology still must assume that the VAR is a correct representation of the dynamics of all the variables in the system (the model is correct). Consider the response of industrial production in [Figure 6](#). The external instruments approach allows the contemporaneous response of industrial production to be non-zero. In fact, however, it is estimated to be very close to zero (i.e., not very different from the Cholesky case). The fact that the response of industrial production is different at later horizons is therefore mainly due to dynamic influences of other variables on industrial production as estimated in the VAR (most notably differences in the response of the excess bond premium and its estimated effect on output in the VAR). In other words, while the contemporaneous responses are estimated more freely, the VAR dynamics are still doing a lot of the work when it comes to the inference about responses at future horizons.

4.4.2 An Example: Romer and Romer Versus VARs

We have discussed different ways to identify monetary shocks and different ways to construct impulse responses using these shocks. We close this section with an example where we contrast these different methods in a setting in which they yield quite different results. [Coibion \(2012\)](#) has drawn attention to the fact that [Romer and Romer's \(2004\)](#) results about the impact of monetary shocks contrast sharply with those of standard monetary VARs. The peak responses of industrial production and unemployment to a change in the federal funds rate are roughly six times larger in [Romer and Romer \(2004\)](#) than in a standard monetary VAR. Furthermore, while the contribution of monetary policy shocks to fluctuations in unemployment and inflation is quite modest according to a standard monetary VAR, monetary policy shocks account for a very large portion of fluctuations in these variables according to Romer and Romer's results.

[Figure 7](#) presents six different estimates of the response of industrial production and the real

interest rate to monetary shocks. The two columns present results based on different monetary shocks: results in the left column are based on monetary shocks from a standard monetary VAR (the VAR used in [Coibion \(2012\)](#)), while results in the right column are based on [Romer and Romer's \(2004\)](#) monetary shocks.³⁶ In the three rows, different methods are used to construct impulse responses: in the top row impulse responses are constructed using VAR dynamics, in the middle row using the Jorda specification, and in the bottom row using the single-equation method employed by [Romer and Romer \(2004\)](#).³⁷ The sample period is 1970-1996—the same sample period as in [Romer and Romer \(2004\)](#).³⁸

There are two things we would like to emphasize about the results presented in [Figure 7](#). First, both the shocks used and the method used to construct impulse responses matters a great deal for the conclusions reached. The Romer-Romer shocks generate much larger effects than the VAR shocks. The VAR impulse responses suggest that the standard errors are substantially smaller than the other methods for constructing impulse responses. Finally, the Romer-Romer method for constructing impulse response generated much larger effects than the other two. Clearly, it matters a great deal which methods are used both for constructing shocks and estimating impulse responses.

The second important lesson illustrated by [Figure 7](#) has to do with how to best measure the size of the response. The direct comparison between [Romer and Romer's \(2004\)](#) results and those from a standard monetary VAR can be seen by comparing the top-left panel (VAR) with the bottom-right panel (Romer-Romer). The response of industrial production is clearly much bigger for Romer-Romer than VAR. But notice that the response of the real interest rate is also much bigger. Loosely speaking, this means that the “treatment” in the Romer-Romer panel is much bigger. It stands to reason that this contributes substantially to the bigger response of industrial production in the Romer-Romer panel.

³⁶[Coibion's \(2012\)](#) VAR is a monthly VAR that includes the logarithm of industrial production, the unemployment rate, the logarithm of CPI, the logarithm of a commodity price index, and the effective federal funds rate, in that order. Twelve lags are included. Standard errors are constructed using a wild bootstrap ([Goncalves and Kilian, 2004](#); [Mertens and Ravn, 2013](#)). For [Romer and Romer's \(2004\)](#) shocks, we use a version of compiled by [Wieland and Yang \(2017\)](#).

³⁷Romer and Romer regress industrial production on 36 lags of their monetary shocks as well as 24 lags of industrial production and month dummies. They then construct an impulse response function by iterating forward the response of industrial production to a shock (including the effects of the lagged dependent variables). In the Jorda specification, we include two lags of the five variables in [Coibion's VAR](#) as controls. The standard errors in the top row are constructed using a wild bootstrap ([Goncalves and Kilian, 2004](#); [Mertens and Ravn, 2013](#)). In the middle panel, we use [Newey and West \(1987\)](#) standard errors with a lag length equal to 1 plus the horizon in question. The standard errors in the bottom two panels are constructed based on estimates of the asymptotic distribution of the parameters as in [Romer and Romer \(2004\)](#).

³⁸See [Figure A.1](#) in the appendix for an analogous figure plotting the response of the CPI and the nominal interest rate.

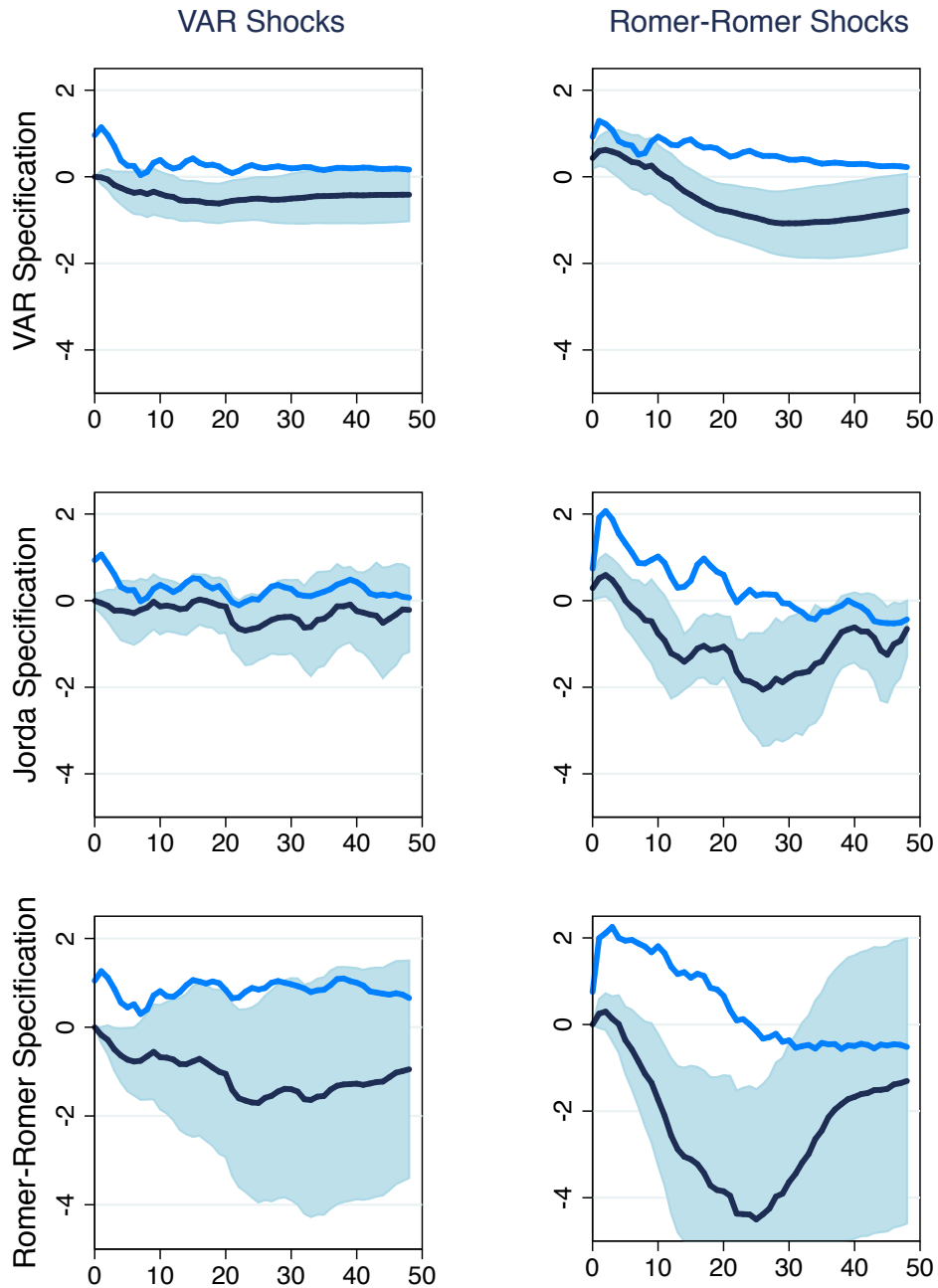


Figure 7: Response of Industrial Production and the Real Interest Rate to Monetary Shocks

Note: The figure plots the response of industrial production (black line) and the real interest rate (blue line) to monetary shocks calculated in six different ways. The light-blue area represents a 95% confidence band for industrial production. The three panels on the left use the monetary shocks produced by Coibion's (2012) monthly VAR, while the three panels on the right use Romer and Romer's (2004) monetary shocks. The top two panels use Coibion's (2012) VAR specification to construct the impulse response, while the middle two panels use the Jorda specification and the bottom two panels use Romer and Romer's (2004) specification. The sample period is 1970 to 1996, the same as the sample period in Romer and Romer (2004). The top panel on the right uses the Romer-Romer shocks as external instruments in Coibion's VAR.

Table 1: Scaled Cumulative Response of Industrial Production
to Monetary Shocks over 36 Months

	VAR Shocks	Romer-Romer Shocks
VAR Specification	1.4	0.7
Jorda Specification	1.0	2.0
Romer-Romer Specification	1.3	3.5

Notes: Each number in the figure is the (negative of the) cumulative response of industrial production divided by the cumulative response of the real interest rate over the first 36 months after a monetary shock.

To get a meaningful measure of the effect of a treatment, one must divide the size of the response with the size of the treatment in each case. One way to do this is to sum up the response of industrial production over some time horizon, sum up the response of the real interest rate over the same horizon, and divide the cumulative response of one by the other. This is what is done in Table 1 using a horizon of 36 months. While the peak response of industrial production is roughly 6 times larger in Romer-Romer than VAR, the cumulative response of industrial production scaled by the cumulative response of real interest rates is only 2.5 times larger (3.5 versus 1.4). This illustrates well the point we made earlier in this article that different dynamics of different shocks can make it tricky to apply the estimates from one setting to another without the aid of a model.

5 Conclusion

Macroeconomics and meteorology are similar in certain ways. First, both fields deal with highly complex general equilibrium systems. Second, both field 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, policy makers, 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 macroe-

conomic 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 solid scientific facts.

The time it will take to drive ideology out of macroeconomic debates depends critically on the quantity and quality of empirical work in macroeconomics. But for good empirical work to be done in macroeconomics, macroeconomists must take care to train students in the methods needed to do good empirical work. Many of the advances in research design that are currently mainly associated with work in applied microeconomics are highly relevant for developing identified moments in macroeconomics. Understanding these methods and how they are applied within macroeconomics is of great value for macroeconomics students. Likewise, many of the ideas macroeconomists have developed about how to use identified moments to draw inference about policy questions, are also relevant and useful for applied microeconomics.

Table A.1: Romer-Romer Dates and Oil-Shock Dates

Romer and Romer Dates	Oil Shock Dates
October 1947	December 1947
	June 1953
September 1955	June 1956
	February 1957
December 1968	March 1969
	December 1970
April 1974	January 1974
August 1978	March 1978
October 1979	September 1979
	February 1981
	January 1987
December 1988	December 1988
	August 1990

Notes: Romer-Romer dates are dates are identified by [Romer and Romer \(1989\)](#) and [Romer and Romer \(1994\)](#). Oil-shock dates up to 1981 are taken from [Hoover and Perez \(1994\)](#), who refine the narrative identification of these shocks by [Hamilton \(1983\)](#). The last three oil shock dates are from [Romer and Romer \(1994\)](#).

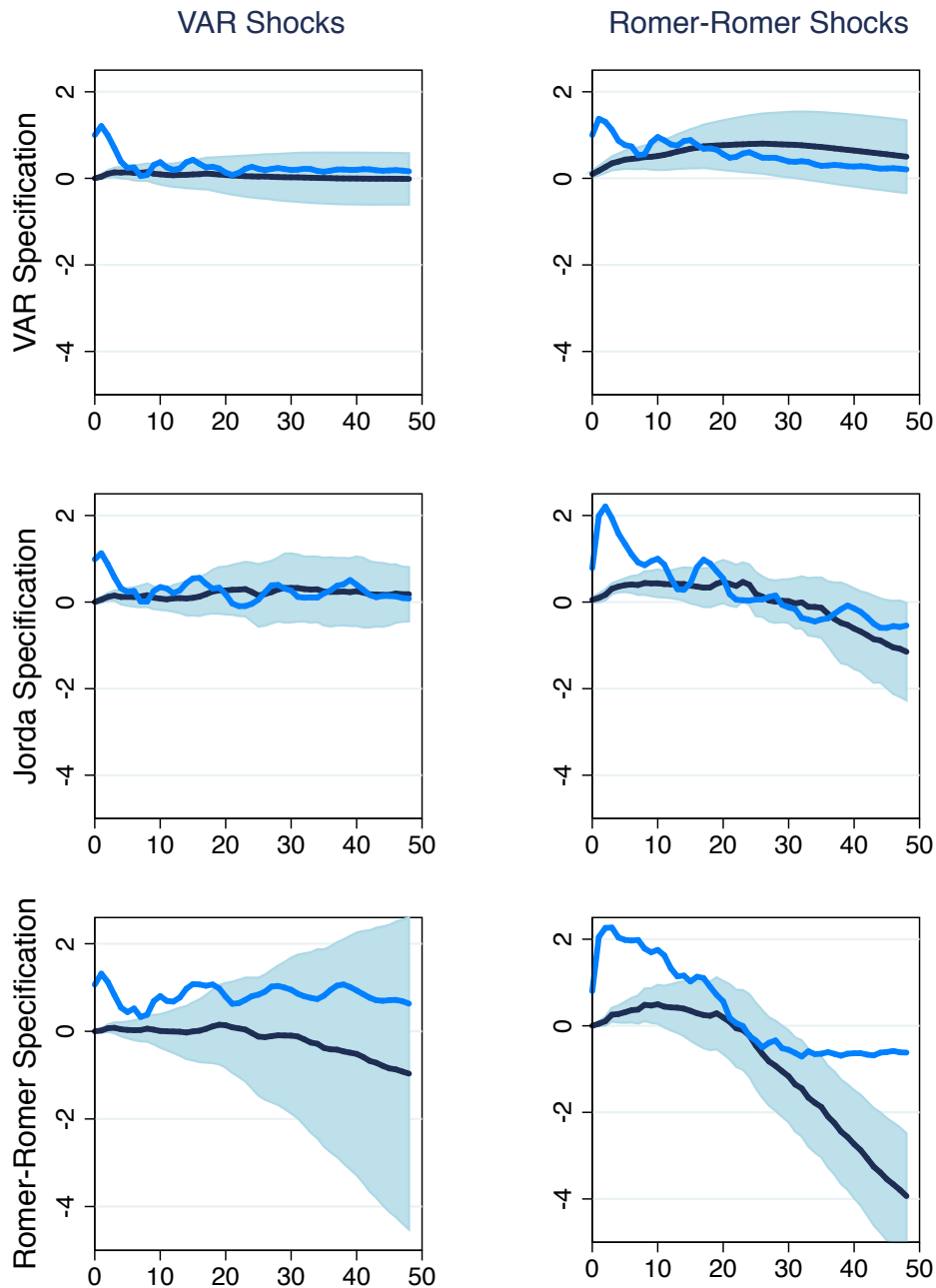


Figure A.1: Response of the CPI and the Nominal Interest Rate to Monetary Shocks

Note: The figure plots the response of the CPI (black line) and the nominal interest rate (blue line) to monetary shocks calculated in six different ways. The light-blue area represents a 95% confidence band for the CPI. The three panels on the left use the monetary shocks produced by Coibion's (2012) monthly VAR, while the three panels on the right use Romer and Romer's (2004) monetary shocks. The top two panels use Coibion's (2012) VAR specification to construct the impulse response, while the middle two panels use the Jorda specification and the bottom two panels use Romer and Romer's (2004) specification. The sample period is 1970 to 1996, the same as the sample period in Romer and Romer (2004). The top panel on the right uses the Romer-Romer shocks as external instruments in Coibion's VAR.

References

- ACCONCIA, A., G. CORSETTI, AND S. SIMONELLI (2014): "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-experiment," *American Economic Review*, 104, 2185–2209.
- AGUIAR, M. AND E. HURST (2005): "Consumption versus Expenditures," *Journal of Political Economy*, 113, 919–948.
- ANGELETOS, G.-M., D. LAIBSON, A. REBETTO, J. TOBACMAN, AND S. WEINBERG (2001): "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 15, 47–68.
- ANGRIST, J. D., O. JORDA, AND G. M. KUERSTEINER (2017): "Semiparametric Estimates of Monetary Policy Effects: String Theory Revisited," *Journal of Business and Economic Statistics*, forthcoming.
- ANGRIST, J. D. AND J.-S. PISCHKE (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.
- AUTOR, D. H., D. DORN, AND G. H. HANSON (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–2168.
- BARRO, R. J. AND C. J. REDLICK (2011): "Macroeconomic Effects from Government Purchases and Taxes," *Quarterly Journal of Economics*, 126, 51–102.
- BASU, S., J. G. FERNALD, AND M. S. KIMBALL (2006): "Are Technology Improvements Contractionary?" *American Economic Review*, 96, 1418–1448.
- BERAJA, M., E. HURST, AND J. OSPINA (2016): "The Aggregate Implications of Regional Business Cycles," Working Paper, University of Chicago.
- BERGER, D., V. GUERRIERI, G. LORENZONI, AND J. VAVRA (2017): "House Prices and Consumption Spending," Working Paper, Northwestern University.
- BERNANKE, B. S. AND K. N. KUTTNER (2005): "What Explains the Stock Market's Reaction to Federal Reserve Policy," *Journal of Finance*, 60, 1221–1257.
- BILS, M. AND P. J. KLENOW (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.
- BLANCHARD, O. AND R. PEROTTI (2002): "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *Quarterly Journal of Economics*, 117, 1329–1368.
- BOPPART, T. AND P. KRUSELL (2016): "Labor Supply in the Past, Present, and Future: A Balanced-Growth Perspective," NBER Working Paper No. 22215.
- BORDO, M., O. F. HUMPAGE, AND A. J. SCHWARTZ (2011): "U.S. Intervention During the Bretton Woods Era: 1962-1973," NBER Working Paper No. 16946.
- CAMPBELL, J. R., C. L. EVANS, J. D. FISHER, AND A. JUSTINIANO (2012): "Macroeconomic Effects of Federal Reserve Forward Guidance," *Brookings Papers on Economic Activity*, 2012, 1–80.
- CATHERINE, S., T. CHANEY, Z. HUANG, D. A. SRAER, AND D. THESMAR (2017): "Aggregate Effects of Collateral Constraints," SSRN Working Paper.

- CHARI, V. V., P. J. KEHOE, AND E. R. MCGRATTAN (2008): "Business Cycle Accounting," *Econometrica*, 75, 781–836.
- CHETTY, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–487.
- (2012): "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80, 969–1018.
- CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2013): "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macroeconomics Annual*, 27, 1–56.
- CHODOROW-REICH, C., L. FEIVESON, Z. LISCOW, AND W. G. WOOLSTON (2012): "Does State Fiscal Relief during Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act," *American Economic Journal: Economic Policy*, 4, 118–145.
- CHODOROW-REICH, G. (2014): "The Employment Effects of Credit Market Disruptions: Firm-Level Evidence from the 2008-9 Financial Crisis," *Quarterly Journal of Economics*, 129, 1–59.
- (2017): "Geographic Cross-Sectional Fiscal Multipliers: What Have We Learned?" Working Paper, Harvard University.
- CHODOROW-REICH, G. AND L. KARABARBOUNIS (2016): "The Limited Macroeconomic Effects of Unemployment Benefit Extensions," NBER Working Paper No. 22163.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 115, 1–45.
- COCHRANE, J. H. (2004): "Comments on 'A New Measure of Monetary Shocks: Derivation and Implications'," Presented at NBER EF&G meeting, July 17 2004.
- COCHRANE, J. H. AND M. PIAZZESI (2002): "The Fed and Interest Rates: A High-Frequency Identification," *American Economic Review*, 92, 90–95.
- COIBION, O. (2012): "Are the Effects of Monetary Policy Shocks Big or Small?" *American Economic Journal: Macroeconomics*, 4, 1–32.
- DUPOR, B. AND M. S. MEHKARI (2016): "The 2009 Recovery Act: Stimulus at the Extensive and Intensive Labor Margins," *European Economic Review*, 85, 208–228.
- EGGERTSSON, G. B. (2008): "Great Expectations and the End of the Depression," *American Economic Review*, 98, 1476–1516.
- EGGERTSSON, G. B. AND B. PUGSLEY (2006): "The Mistake of 1937: A General Equilibrium Analysis," *Monetary Economic Studies*, 24, 151–208.
- EICHENGREEN, B. (1992): *Golden Fetters: The Gold Standard and the Great Depression 1919-1939*, Oxford, UK: Oxford University Press.
- EICHENGREEN, B. AND J. SACHS (1985): "Exchange Rates and Economic Recovery in the 1930s," *Journal of Economic History*, 45, 925–946.
- FAUST, J., E. T. SWANSON, AND J. H. WRIGHT (2004): "Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy?" *Contributions to Macroeconomics*, 4, 1–29.
- FRIEDMAN, M. (1977): "Nobel Lecture: Inflation and Unemployment," *Journal of Political Economy*,

85, 451–472.

- FRIEDMAN, M. AND A. J. SCHWARTZ (1963): *A Monetary History of the United States, 1867-1960*, Princeton, NJ: Princeton University Press.
- GALI, J. (1999): “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations,” *American Economic Review*, 89, 249–271.
- GALI, J., D. LOPEZ-SALIDO, AND J. VALLES (2007): “Understanding the Effects of Government Spending on Consumption,” *Journal of the European Economic Association*, 5, 227–270.
- GERTLER, M. AND P. KARADI (2015): “Monetary Policy Surprises, Credit Costs, and Economic Activity,” *American Economic Journal: Macroeconomics*, 7, 44–76.
- GONCALVES, S. AND L. KILIAN (2004): “Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form,” *Journal of Econometrics*, 123, 89–120.
- GOODFRIEND, M. (2007): “How the World Achieved Consensus on Monetary Policy,” *Journal of Economic Perspectives*, 21, 47–68.
- GOODFRIEND, M. AND R. G. KING (2005): “The Incredible Volcker Disinflation,” *Journal of Monetary Economics*, 52, 981–1015.
- GUREN, A. M., A. MCKAY, E. NAKAMURA, AND J. STEINSSON (2017): “Housing Wealth Effects: The Long View,” Working Paper, Boston University.
- GURKAYNAK, R. S., B. SACK, AND E. T. SWANSON (2005): “Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements,” *International Journal of Central Banking*, 1, 55–93.
- HAGEDORN, M., I. MANOVSKII, AND K. MITMAN (2015): “The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle?” NBER Working Paper No. 20884.
- HALL, R. E. (2009): “By How Much Does GDP Rise If the Government Buys More Output?” *Brookings Papers on Economic Activity*, 2009, 183–231.
- HAMILTON, J. D. (1983): “Oil and the Macroeconomy since World War II,” *Journal of Political Economy*, 91, 228–248.
- HANSON, S. G. AND J. C. STEIN (2015): “Monetary Policy and Long-Term Real Rates,” *Journal of Financial Economics*, 115, 429–448.
- HOOVER, K. D. AND S. 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, 47–73.
- HSIEH, C.-T. AND P. J. KLENOW (2009): “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 124, 1403–1448.
- ITSKHOKI, O. AND D. MUKHIN (2017): “Exchange Rate Disconnect in General Equilibrium,” Working Paper, Princeton University.
- JOHNSON, D. S., J. A. PARKER, AND N. S. SOULELES (2006): “Household Expenditure and the Income Tax Rebates of 2001,” *American Economic Review*, 96, 1589–1610.
- JORDA, O. (2005): “Estimation and Inference of Impulse Response by Local Projection,” *American Economic Review*, 95, 161–182.

- KAPLAN, G. AND G. L. VIOLANTE (2014): "A Model of the Consumption Response to Fiscal Stimulus Payments," *Econometrica*, 82, 1199–1239.
- KING, R. G., C. I. PLOSSER, AND S. T. REBELO (1988): "Production, Growth and Business Cycles: I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, 195–232.
- KING, R. G. AND S. REBELO (1999): "Resuscitating Real Business Cycles," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, Amsterdam, Holland: Elsevier, 927–1007.
- KLENOW, P. J. AND O. KRYVTSOV (2008): "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation," *Quarterly Journal of Economics*, 123, 863–904.
- KYDLAND, F. E. AND E. C. PRESCOTT (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370.
- LEEPER, E. M. (1997): "Narrative and VAR Approaches to Monetary Policy: Common Identification Problems," *Journal of Monetary Economics*, 40, 641–657.
- MARTIN, P. AND T. PHILIPPON (2017): "Inspecting the Mechanism: Leverage and the Great Recession in the Eurozone," *American Economic Review*, 107, 1904–1937.
- MEESE, R. A. AND K. ROGOFF (1983): "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, 14, 3–24.
- MEHRA, R. AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.
- MERTENS, K. AND M. O. RAVN (2012): "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks," *American Economic Journal: Economic Policy*, 4, 145–181.
- (2013): "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States," *American Economic Review*, 103, 1212–1247.
- MIAN, A., K. RAO, AND A. SUFI (2013): "Household Balance Sheets, Consumption, and the Economic Slump," *Quarterly Journal of Economics*, 128, 1449–1496.
- MIAN, A. AND A. SUFI (2014): "What Explains the 2007–2009 Drop in Employment?" *Econometrica*, 82, 2197–2223.
- MIDRIGAN, V. AND T. PHILIPPON (2016): "Household Leverage and the Recession," Working Paper, New York University.
- MIYAMOTO, W., T. L. NGUYEN, AND D. SERGEYEV (2017): "Government Spending Multipliers under the Zero Lower Bound: Evidence from Japan," Working Paper, Santa Clara University.
- MUSSA, M. (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, E. AND J. STEINSSON (2008): "Five Facts About Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123, 1415–1464.
- (2014): "Fiscal Stimulus in a Monetary Union: Evidence from US Regions," *American Economic Review*, 104, 753–792.
- (2017): "High Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, forthcoming.

- NELSON, E. (2005): "The Great Inflation of the Seventies: What Really Happened?" *Advances in Macroeconomics*, 5.
- NEWKEY, W. K. AND K. D. WEST (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.
- OKUN, A. M. (1978): "Efficient Disinflation Policies," *American Economic Review*, 68, 348–352.
- PARKER, J. A., N. S. SOULELES, D. S. JOHNSON, AND R. MCCLELLAND (2013): "Consumer Spending and the Economic Stimulus Payments of 2008," *American Economic Review*, 103, 2530–2553.
- PRESCOTT, E. C. (1986): "Theory Ahead of Business-Cycle Measurement," *Carnegie-Rochester Conference Series on Public Policy*, 25, 11–44.
- RAMEY, V. AND S. ZUBAIRY (2017): "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data," *Journal of Political Economy*, forthcoming.
- RAMEY, V. A. (2011): "Identifying Government Spending Shocks: It's All in the Timing," *Quarterly Journal of Economics*, 126, 1–50.
- (2016): "Macroeconomic Shocks and Their Propagation," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Amsterdam, Holland: Elsevier, 71–162.
- RIGOBON, R. (2003): "Identification through Heteroskedasticity," *The Review of Economics and Statistics*, 85, 777–792.
- RIGOBON, R. AND B. SACK (2004): "The impact of monetary policy on asset prices," *Journal of Monetary Economics*, 51, 1553–1575.
- ROMER, C. D. AND D. H. ROMER (1989): "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," in *NBER Macroeconomics Annual*, ed. by O. J. Blanchard and S. Fischer, Cambridge, MA: MIT Press, 121–170.
- (1994): "Monetary Policy Matters," *Journal of Monetary Economics*, 34, 75–88.
- (1997): "Identification and the Narrative Approach: A Reply to Leeper," *Journal of Monetary Economics*, 40, 659–665.
- (2000): "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 90, 429–457.
- (2002): "The Evolution of Economic Understanding and Postwar Stabilization Policy," in *Rethinking Stabilization Policy*, Kansas City: Federal Reserve Bank of Kansas City, 11–78.
- (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.
- ROTEMBERG, J. J. AND M. WOODFORD (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in *NBER Macroeconomics Annual 1997*, ed. by B. S. Bernanke and J. J. Rotemberg, Cambridge, MA: MIT Press, 297–346.
- RUDEBUSCH, G. D. (1998): "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review*, 39, 907–931.
- SAIZ, A. (2010): "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, 125, 1253–1296.

- SHAPIRO, M. D. (1994): "Federal Reserve Policy: Cause and Effect," in *Monetary Policy*, ed. by N. G. Mankiw, Chicago, IL: University of Chicago Press, 307–334.
- SHIMER, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49.
- (2009): "Convergence in Macroeconomics: The Labor Wedge," *American Economic Journal: Macroeconomics*, 1, 280–297.
- SHOAG, D. (2015): "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns," Working Paper, Harvard University.
- SINAI, T. AND N. S. SOULELES (2005): "Owner-Occupied Housing as a Hedge against Rent Risk," *Quarterly Journal of Economics*, 120, 763–789.
- SMITH, A. A. (2008): "Indirect Inference," in *The New Palgrave Dictionary of Economics, Second Edition*, ed. by S. Durlauf and L. E. Blume, London: Palgrave Macmillan, available at: <http://www.econ.yale.edu/smith/palgrave7.pdf>.
- STOCK, J. H. AND M. W. WATSON (2012): "Disentangling the Channels of the 2007-09 Recession," *Brookings Papers on Economic Activity*, 2012, 81–135.
- (2017): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics," Working Paper, Harvard University.
- SUAREZ SERRATO, J. C. AND P. WINGENDER (2016): "Estimating Local Fiscal Multipliers," Working Paper, Duke University.
- SUMMERS, L. H. (1991): "The Scientific Illusion in Empirical Macroeconomics," *Scandinavian Journal of Economics*, 93, 129–148.
- VELDE, F. R. (2009): "Chronicle of a Deflation Unforetold," *Journal of Political Economy*, 117, 591–634.
- WIELAND, J. AND M.-J. YANG (2017): "Financial Dampening," Working Paper, University of California, San Diego.
- WILSON, D. J. (2012): "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act," *American Economic Journal: Economic Policy*, 4, 251–282.