Journal of Economic Perspectives—Volume 33, Number 2—Spring 2019—Pages 89–114

Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?

Valerie A. Ramey

When the financial crisis hit ten years ago and monetary policy interest rates fell to their near-zero percent lower bound, policymakers around the world turned to fiscal stimulus packages in order to prevent their economies from freefalling into another Great Depression. But then, as declining GDP and tax revenues led to deteriorating government budget deficits and worries about rising sovereign debt, numerous countries abandoned their fiscal stimulus packages and instead adopted fiscal consolidation measures. While attempting to forecast the impacts of these various fiscal programs, policymakers and academics were surprised to discover not only a lack of consensus about the size of the effects of fiscal policy, but also a dearth of research on the topic since the 1960s. A small army of researchers across many countries turned their attention to this important but long-neglected topic.

This paper takes a snapshot of the state of knowledge about the effects of fiscal policy ten years after the global financial crisis, during which time important progress has been made on theory, empirical methods, and data. The theoretical innovations include the analysis of the effects of sticky prices, hand-to-mouth consumers, lower bounds on policy interest rates, currency unions, the type of financing, and anticipations on the reactions of macroeconomic variables to fiscal policy. Contributions in empirical methods include new ways to identify exogenous variation in policy, standardization of methods for computing fiscal multipliers (defined as the ratio

■ *Valerie A. Ramey is Professor of Economics, University of California, San Diego, La Jolla, California, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Her email address is vramey@ucsd.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.33.2.89 doi=10.1257/jep.33.2.89

of the change in output to the change in spending or taxes that caused it), and the incorporation of state dependence. On the data front, researchers now have newly constructed historical and cross-sectional datasets, and are also exploiting the rich new data created by the variety of policymakers' fiscal responses to the crisis. These advancements offer the potential to estimate the effects of government spending with more precision and with a better understanding of how the effects depend on the particular context.

In 2011, I surveyed the pre-crisis and early crisis literature in the *Journal of Economic Literature*. In that paper, which focused only on temporary, deficit-financed increases in government purchases, I concluded based on the evidence available from US data at that time that the multiplier was probably between 0.8 to 1.5, but that the data did not reject a range from 0.5 to 2. The current paper refines those estimates and broadens the inquiry to consider the effects of tax and transfer policy, as well as the effects of fiscal consolidations, in developed countries. However, attention is still limited to the short- or medium-run effects, because the methods for estimating long-run effects are quite different.

My summary of the current state of knowledge about the effects of fiscal policies can be divided into three categories: government purchases multipliers, tax rate change multipliers, and fiscal multipliers in the wake of the financial crisis.

For multipliers on general government purchases, the evidence from developed countries suggests that they are positive but less than or equal to unity, meaning that government purchases raise GDP but do not stimulate additional private activity and may actually crowd it out. The bulk of the estimates across the leading methods of estimation and samples lie in a surprisingly narrow range of 0.6 to 1. However, this range widens once one distinguishes country characteristics, such as the exchange rate regime, and the type of government spending, such as infrastructure spending. The evidence for higher spending multipliers during recessions or times of high unemployment is fragile, and the most robust results suggest multipliers of one or below during these periods. The evidence for higher government spending multipliers during periods in which monetary policy is very accommodative, such as zero lower bound periods, is somewhat stronger. Recent time series estimates for the United States and Japan suggest that multipliers could be 1.5 or higher during those times. Estimated and calibrated New Keynesian models for the United States and Europe also imply higher multipliers under certain conditions.

For tax rate change multipliers, the estimates implied by the leading methods do not agree. Narrative methods (which use historical documents to find exogenous changes) for tax rate changes typically yield multiplier estimates that are surprisingly large and surprisingly uniform across a number of countries. The bulk of the empirical estimates vary between –2 and –3. In contrast, most calibrated and estimated New Keynesian dynamic stochastic general equilibrium (DSGE) models imply smaller multipliers, typically below unity for both labor and capital tax multipliers. Time series evidence, theory, and estimated New Keynesian DSGE models all point to tax multipliers being greater in magnitude during expansions than in recessions—that is, these measures suggest that tax multipliers may be *procyclical*.

Fiscal multipliers might be different in the wake of a financial crisis. However, the evidence for larger national multipliers on the 2009 Obama stimulus package is at best weak. Quantitative New Keynesian models do not find larger fiscal multipliers. Multipliers estimated on cross-state data appear larger at first, but shrink once they are adjusted to be nationally representative. The latest studies on multipliers during the fiscal consolidations in Europe suggest that they were not higher than usual, either.

This paper begins by reviewing how theory highlights the dependence of the size of the fiscal multipliers on numerous features of the policy and the economy. The next section summarizes strengths and weaknesses of the leading empirical approaches to identifying exogenous shifts in fiscal policy. The paper then overviews the innovations of the last ten years in estimating fiscal multipliers. One interesting finding is that the wide range of multipliers reported earlier narrows significantly once methods for calculating multipliers are standardized. The following section reviews the leading estimates of spending and tax multipliers, including those based on aggregate time series, estimated theoretical models, and subnational units and households. It also discusses the complexities of drawing aggregate inferences from parameters estimated on household data. The penultimate section asks what we know about whether multipliers were higher in the wake of the financial crisis. The final section offers some brief conclusions.

What Does Theory Predict about Fiscal Multipliers?

If we simply want to know how much GDP changes if we increase government spending by $1 or reduce tax rates by 1 percentage point, why do we need theory? Theory tells us that there is not just one government spending or tax multiplier. Rather, the effect of fiscal changes on output and other variables potentially depends on: 1) the persistence of the change; 2) the type of spending or taxes that changed; 3) how the policy was financed; 4) whether it was anticipated; 5) how the policy was distributed across potentially heterogeneous agents; 6) how monetary policy reacted; 7) the state of the economy when the policy took effect; and 8) other features that characterize the economy such as level of development, exchange rate regime, and openness. Because policymakers cannot conduct randomized control trials, virtually all multiplier estimates are based on time series, narrative, or natural experiment identification using samples determined by historical happenstance. To understand whether a particular estimate of fiscal effects is suitable for use in predicting the effects of a proposed policy, one must understand how the current circumstances differ from those present in the sample used to generate that estimate.

Most researchers and policymakers had their first exposure to the theoretical effects of fiscal policy in the Keynesian cross model of undergraduate textbooks, which assumes that GDP is demand-determined. This model further assumes that the government spending multiplier is the inverse of one minus the marginal propensity to consume: thus, a marginal propensity to consume of 0.5 yields a multiplier of 2. Because taxes enter the multiplier only through their effect on disposable income in this model, the tax multipliers are smaller than the spending multipliers. Expansion of the model to consider the marginal propensity to import, tax rates, and monetary policy reduces those simple multipliers.

Neoclassical models with variable labor supply and capital stock also predict positive spending multipliers and negative (distortionary) tax multipliers, but the mechanism is completely different from the one at the heart of the traditional Keynesian model. In these models, an increase in government spending has a negative wealth effect, because the government is extracting resources from the private sector. This negative wealth effect raises GDP because it causes households to work more. Distortionary tax rate changes can have potentially large effects in these models, but contrary to the simple Keynesian model, they work through "supply side" channels (for example, Baxter and King 1993).

New Keynesian dynamic stochastic general equilibrium (DSGE) models meld the insights from the traditional Keynesian and neoclassical approaches in a rigorous way (for example, Woodford 2011). The standard representative-agent sticky-price New Keynesian model with no financial frictions tends to produce multipliers below one for government spending. Models that add sticky wages and workers who are "off their labor supply curves" generate larger multipliers. In the last decade, representative agent models have been expanded to include heterogeneous agents and financial market frictions. In these models, either "rule-of-thumb" behavior or wealth held in illiquid assets leads agents to have much higher marginal propensities to consume than predicted by the permanent income hypothesis. These features can lead to spending multipliers above one when spending is deficit financed (for example, Galí, López-Salido, and Vallés 2007; Auclert, Rognlie, and Straub 2018). Alternatively, the models have explored the effects of fiscal policy when monetary policy deviates from the standard Taylor rule (higher interest rates when inflation is high and lower interest rates when unemployment is high) because interest rates are constrained by the zero lower bound. Both of these extensions result in higher multipliers, often above unity.

Clearly, when one is trying to estimate the effects of a specific fiscal policy, one must be aware of which macroeconomic model is being used, along with other factors like persistence of a path of government spending, how it is financed, and many other characteristics such as the exchange rate regime.

A Summary of Leading Empirical Approaches

Numerous empirical approaches have been used to estimate the effects of fiscal policies. I group these approaches into three broad categories: 1) aggregate country-level time series or panel estimates; 2) estimated or calibrated New Keynesian dynamic stochastic general equilibrium (DSGE) models; and 3) subnational geographic cross-section or panel estimates.

The first two categories—time series evidence at the national level and estimated/calibrated DSGE models—share the advantage that the estimates produced are directly informative about the national-level multipliers that are the focus of most policymakers. The time series approach has the advantage of not being tied to a particular structural model. On the other hand, the New Keynesian DSGE model approach can be used to perform counterfactuals because it seeks to estimate structural parameters.

However, these two approaches share some of the same weaknesses. Identification of macroeconomic parameters is always difficult, and the estimation of the aggregate effects of fiscal policy is no exception. The time series approach requires exogenous variation in policy. The leading approaches to identifying this exogenous variation are structural vector autoregressions and natural experiment methods, combined with narrative methods that use historical documents to create new data series of exogenous changes. Too often, though, the variations that turn out to be exogenous yield instruments that are not very *relevant*—that is, they have low correlation with the fiscal variable they are trying to explain—and the variations that are relevant are not always exogenous or are anticipated in advance.

Although many papers using estimated dynamic stochastic general equilibrium models never mention the word *identification*, identification is as crucial to this approach as it is to any other approach seeking to estimate a causal relationship. The New Keynesian DSGE approach identifies the effects of fiscal policy by using strong assumptions about the theoretical model structure and the time series processes driving the unobserved shocks. But such estimated quantitative models are not immune to weak identification (for discussion, see Canova and Sala 2009).

The third approach of estimating across subnational units, such as states or provinces, is more similar to applied microeconomics approaches. These approaches typically seek identification using a natural experiment approach or Bartik-style instrumental variables (which are based on interacting the distribution of industry shares across locations with national industry growth rates).[1] These analyses at lower levels of aggregation tend to have much stronger identification, in the sense that the necessary identifying assumptions are typically more plausible and the instruments are relevant. Moreover, these approaches can be used on a variety of datasets. However, this approach does not lead directly to macroeconomic estimates. Why? Any cross-sectional estimating equation includes a constant term, which means that the macroeconomic effects have been netted out and the parameters estimated are only *relative* effects. Such parameters answer the question: If State A is awarded $1 more in defense prime contracts than the average state, by how much does its employment change *relative* to the average state? In order to infer the implied national-level effects from such microeconomic estimates, researchers must then return to macroeconomic New Keynesian DSGE models, which, as discussed above, incorporate their own additional identifying assumptions. There is no "applied micro free lunch" for macroeconomists. Identification of macroeconomic effects must always depend on macroeconomic identification assumptions.

[1] For a description and critical analysis of Bartik instruments, see Goldsmith-Pinkham, Sorkin, and Swift (2018).

To summarize, there are several approaches to estimating the effects of fiscal policy. Each has its strengths and weaknesses. Moreover, some of the estimates are more appropriate for forecasting the effects of specific policies under certain conditions than others. For these reasons, it is useful to consider estimates across a range of different approaches.

Research Innovations and Lessons Learned during the Last Ten Years

Before the financial crisis, only a few isolated researchers studied the macroeconomic effects of fiscal policy and only a few conferences brought these researchers together. As a result, different researchers chose different methods and there was no agreement on a set of best practices. The situation has changed dramatically since the financial crisis, with many conferences devoted to the study of fiscal policies and much more interaction among researchers studying fiscal policy. As a result, the diffusion of knowledge among researchers has been much faster, and the literature has progressed at a very fast pace. In this section, I will highlight some of the new innovations and the lessons learned from this literature.

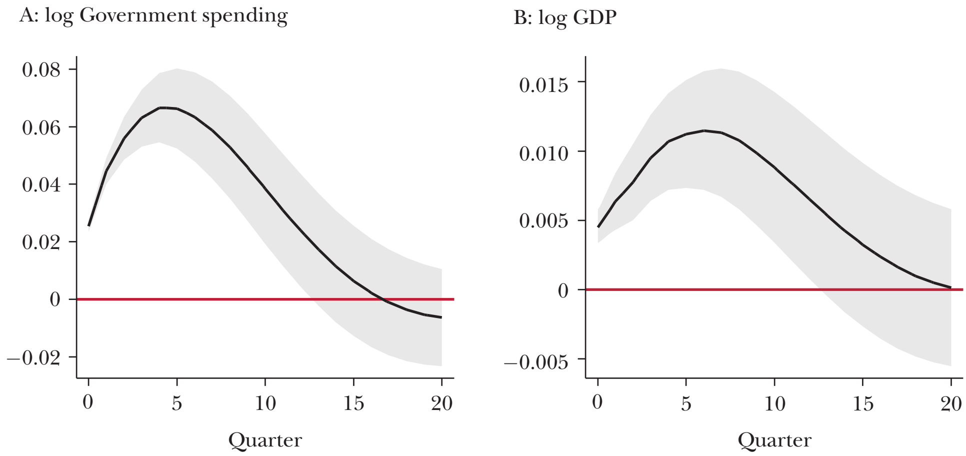
Calculating Multipliers in a Dynamic Environment

One often sees references to the "wide range" of multiplier estimates. The literature has come to realize that differences in reported multiplier estimates are often due not so much to differences in identification methods or samples, but to the methods used to construct multipliers from the raw estimates. In fact, what some researchers call "multipliers" have little to do with the multipliers of interest to policymakers. This section begins with some insights gained over the last decade regarding the computation of multipliers. I begin with spending multipliers and then address a further complication involved with tax multipliers.

Fiscal policy has dynamic effects on output and government budgets. A typical fiscal plan will set into motion a path of spending or taxes over time, and then GDP will respond dynamically to that path. The multiplier must take into account both the multi-year effects of the fiscal plan on the government budget, in order to count the costs fully, as well as the multi-year effects on GDP, in order to count the benefits fully.

Computation of fiscal multipliers was not a focus of research in the decades before the financial crisis. Indeed, in Ramey and Shapiro (1998), when discussing the effects of government spending two decades ago, we did not even mention the word "multiplier." When describing the patterns of the responses of GDP to spending and tax shocks, Blanchard and Perotti (2002) used the word "multiplier," but the quantities they calculated were not true dynamic multipliers; instead, Blanchard and Perotti calculated multipliers as the ratio of the output response at a particular horizon, or at its *peak*, to the *impact* effect of the shock on government spending. Many subsequent papers adopted their method, despite the fact that it did not take into account the multi-year path of spending or taxes. Mountford and Uhlig (2009) moved the literature forward by introducing the policy-relevant multipliers, calculated as the present

Figure 1
Estimated Impulse Response Functions for a Shock to Government Purchases



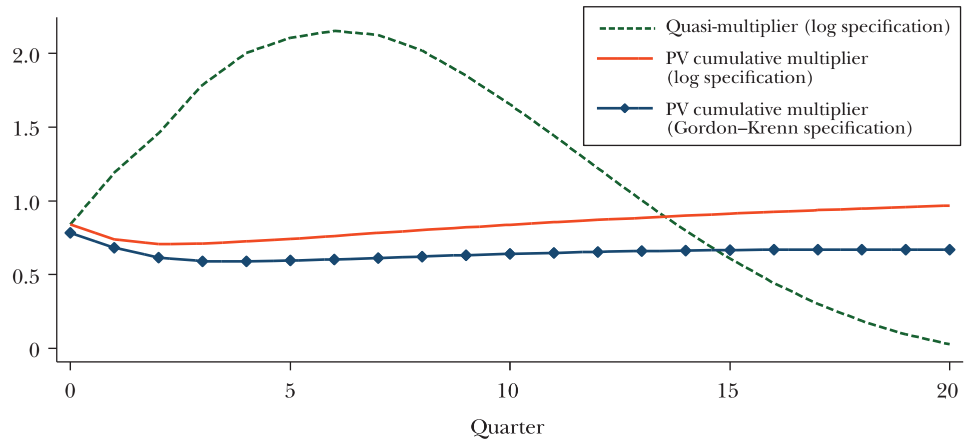
Source: Author.
Note: Estimated impulse responses based on structural vector autoregression (SVAR) estimates using quarterly data from 1939Q1– 2015Q4. The shaded area shows the 95-percent confidence bands. See the text and online appendix available with this paper at the journal website for more detail.

discounted value of the output response over time divided by the present discounted value of the government spending response over time to the shock. In most applications, different interest rates used for this present discounted value—including the use of a zero discount rate—give nearly identical multipliers because the timing of the government spending and output responses is very similar. These multipliers are often known as *present value* or *cumulative multipliers*.

How much do multiplier estimates differ across these various methods of calculating multipliers? It depends importantly on how much government spending rises after the initial impact. Here is one illustration of a situation in which it makes a big difference. I estimate a structural vector autoregression (SVAR) model of the Blanchard and Perotti (2002) type over the period 1939Q1–2015Q4 using the Ramey and Zubairy (2018) dataset. The model contains five endogenous macroeconomic variables: government spending, GDP, and federal tax receipts (with all three deflated by the GDP deflator, divided by population, and in logs), along with the three-month Treasury bill interest rate and inflation (measured as the log change in the GDP deflator). Four lags are included in order to model the dynamics. The exogenous shock to government spending is identified using Blanchard and Perotti's (2002) method, which assumes that any part of government spending not forecasted by lags of any of the variables included in the model is an exogenous shock to government spending.

Figure 1 shows the estimated impulse responses of the log of the government spending variable and the log of the GDP variable (notice that the vertical scales are not the same). The shaded area shows the 95-percent confidence bands. As the

Figure 2
Alternative Definitions of Multipliers: Multipliers by Horizon



Source: Author.
Note: The dotted and solid lines show multipliers calculated based on the log impulse responses shown in Figure 1. The dashed line shows the multiplier given by Blanchard-Perotti's (2002) method, which I call a *quasi-multiplier*. The solid line shows the the Mountford and Uhlig (2009) *present value (PV) cumulative multiplier*. The line with diamonds shows the PV cumulative multiplier using the impulse responses estimated using the Gordon–Krenn specification. See text and online appendix available with this paper at the journal website for more details.

graph illustrates, a positive shock to government spending leads both government spending and GDP to jump up on impact, but then to continue to rise, peaking after about a year. Because the variables are in log form, the impulse responses show elasticities, not the dollar changes required by multipliers, so multipliers cannot be read directly from the graphs. The standard practice until recently had been to use an *ad hoc* "conversion factor." That is, researchers who specified models using logarithms converted the elasticity estimates to multipliers by multiplying the elasticity estimates by the average of the ratio of GDP to total government spending, over the sample. In this illustration, the conversion factor, average Y/G over the sample, is 4.78. I will critique the use of these conversion factors shortly.

Figure 2 shows the multipliers calculated three different ways. The highest multiplier is given by Blanchard-Perotti's (2002) method for calculating a multiplier, which I will call a *quasi-multiplier*.[2] It is calculated as the ratio of the impulse response of output at horizon h to the initial jump in government spending at horizon 0 (multiplied by the average). Their method, shown by the dashed line, essentially traces out a renormalized version of the impulse response of output. In

[2]Note that the Blanchard–Perotti *identification* method is distinct from the Blanchard–Perotti method for calculating multipliers; their method for calculating multipliers could be applied to estimates using any identification method.

this case, it yields multipliers that peak at 2.2 at quarter 6. The Mountford and Uhlig (2009) *present value cumulative multiplier*, shown by the solid line, uses the ratio of the present value of the integral of impulse response of output to the present value of the integral of the impulse response of government spending up to each horizon h (again multiplied by the average *Y/G* factor). This multiplier varies between 0.7 and 1, depending on the horizon. The discounting for this multiplier uses the average three-month Treasury bill rate over the sample, 3.6 percent on an annual basis, but because of the timing of the shift, the simple cumulative version is almost identical.

Now let us return to the issues raised by the practice of converting elasticities with the *ad hoc* conversion factor, the average of *Y/G* over the sample. In Owyang, Ramey, and Zubairy (2013), we discovered biases that could arise from this practice. In our historical sample, *Y/G* varied significantly, from 2 to 24, with a mean of 8. Sims and Wolff (2018a, b) also discovered that this practice tends to bias multipliers differentially, making them seem much higher during recessions. The intuition is straightforward: because GDP is cyclical but government spending is not, the movement of *Y/G* is procyclical. However, the practice of using a sample average to convert elasticities to multipliers makes the multipliers appear more countercyclical than they really are. In Owyang, Ramey, and Zubairy, we avoided this problem by using the transformations employed by Hall (2010) and Barro and Redlick (2011): both the change in government spending and the change in GDP are divided by *lagged* GDP. Another transformation that overcomes the problem is Gordon and Krenn's (2010) approach, which divides both government spending and GDP by a measure of potential GDP.

To illustrate the effect of moving from a specification in logarithms that requires the *ad hoc* conversion factor to one that does not, I re-estimate the structural vector autogression (SVAR) model, replacing the logarithms of government spending, GDP, and taxes with the ratios of each of those variables to the Ramey and Zubairy (2018) polynomial trend estimate of potential GDP. The general shape of the estimated impulse responses (not shown) is very similar to those from the log specification, which were shown in Figure 1. The solid line with diamonds in Figure 2 shows the cumulative multiplier estimates based on the impulse responses from this alternative (Gordon–Krenn) specification. These multipliers, which do not rely on a conversion factor, are lower and range from 0.8 on impact down to 0.6.[3]

Thus, deceptively small changes in the method of calculation can make a very big difference in the resulting multipliers. For this application, using Blanchard and Perotti's (2002) quasi-multiplier for government spending on estimated elasticities requiring an *ad hoc* conversion factor produces a multiplier as high as 2.2. That multiplier falls below 0.8 when the fully dynamic Mountford and Uhlig (2009) cumulative multiplier is used on estimates based on data using the Gordon and

[3] This bias also affects the multipliers I reported in Ramey (2011a). The cumulative multipliers based on the elasticity estimates and conversion factor were 1.2. However, in Ramey (2013), I found evidence that private spending fell, which is inconsistent with a multiplier above 1.

Krenn (2010) transformation. Clearly, such differences could have important consequences for the decisions of policymakers.

In addition, even the cumulative multipliers do not fully reflect the consequences for the government budget. If an increase in government spending raises GDP, then we would expect a rise in tax revenues. Thus, even without an exogenous increase in tax rates, we would expect the government budget deficit to rise less than the total amount of government spending. This insight raises a complication when applying these same principles to the computation of tax multipliers. While there is strong feedback from GDP to tax revenue, there is little feedback from GDP to government spending. As a result, the negative effect of a tax cut on tax revenue is tempered by the feedback from the expansionary effect on output. Indeed, Mertens and Ravn (2013) were not able to compute a multiplier for corporate tax cuts because their large positive impact on GDP resulted in no net effect on tax revenues. Because of the presence of these "top of the Laffer curve" effects in some applications, most papers report multipliers using the tax changes measured as the legislative forecasts of the expected cumulative effect on tax revenues, *not* accounting for dynamic feedback from any potential induced GDP changes.

The Importance of Fiscal Foresight

An important innovation in the fiscal literature in the last decade is the recognition that many changes in government spending and taxes are announced in advance. In Ramey (2011a), I showed the importance of anticipations for estimating the effects of government spending shocks, particularly those involving military spending. For example, the responses of key variables such as consumption could change signs if researchers ignored the fact that many changes in government spending are anticipated by at least several quarters. A number of papers also show that "shocks" identified in standard ways are predicted by professional forecasts of government spending. On the tax front, House and Shapiro (2006) and Mertens and Ravn (2012) demonstrated the importance of distinguishing between changes in taxes implemented soon after legislation and changes in taxes implemented with a lag after legislation or phased in slowly. Both papers showed that while unanticipated tax cuts have expansionary effects on output, phased-in tax cuts depress output during the phase-in period because firms and consumers delay their activity until tax rates are lower. Leeper, Walker, and Yang (2013) derived the econometric biases that arise when there is this type of fiscal foresight. As a result of this work, most of the literature tries to address anticipation whenever feasible, either by constructing measures of news (from narratives or bond spreads) or by including professional forecasts of government spending to mitigate the problem.

Improvements in Fiscal Shock Identification

Any analysis that seeks to measure a causal effect must confront identification issues. An example of the problem that arises here is that if governments increase spending in response to a recession, then the simple correlation between government spending and GDP will confound the positive causal effect of government spending on GDP with the negative causal effect of GDP on government spending. In the past, the standard macro approach used to tease out the exogenous rise in government spending was a structural vector autoregression (SVAR). In most applications, this approach is based on the assumption that the exogenous part of government spending was the part of government spending not forecasted by lagged values of spending, GDP, and taxes. Alternatively, to identify exogenous movements in taxes, Blanchard and Perotti (2002) used external estimates of the elasticity of tax revenue to income, which allowed identification of the component of taxes that was not induced by movements in GDP. Several papers have highlighted potential problems with these widely used methods. First, as discussed above, failing to account for fiscal foresight could lead to biased estimates. Second, the tax multiplier estimates were very sensitive to the value of the external tax elasticity estimate used (for example, Mertens and Ravn 2014; Caldara and Kamps 2017). These concerns led to the development of other identification methods using natural experiments and narrative methods. As a result, the standard SVAR identification approach is no longer the first resort in the literature on fiscal multipliers.

In fact, long before structural vector autoregression methods were used, Hall (1980) and Barro (1981) used natural experiment methods to assess the effects of exogenous increases in government spending. Arguing that changes in US defense spending are typically driven by wars rather than the current state of the economy, they used war-induced government spending to estimate causal effects of government spending in US historical data. Ramey and Shapiro (1998) and numerous other follow-up papers built on treating wars as a natural experiment. This method works well for US data, but it does not export well to other countries. Most countries either do not have the substantial fluctuations in defense spending experienced by the United States or they have large variations that are accompanied by war-related destruction of the capital stock, which leads to confounding effects.

Other examples of recent fiscal research that use natural experiment methods abound. For example, Acconcia, Corsetti, and Simonelli (2014) used the central government response to Mafia infiltration as an exogenous change in government spending in Italian provinces. Many of the analyses of the Obama stimulus allocation of funds across states used natural experiment methods. Two analyses of marginal propensities to spend out of the temporary rebates of 2001 and 2008 exploited the randomized timing of the mailing of checks to households (Johnson, Parker, and Souleles 2006; Parker, Souleles, Johnson, and McClelland 2013). The application of these methods has shed significant light on the effects of fiscal policy, particularly at the local and household level.

Romer and Romer (2010) pioneered the use of narrative methods to identify tax changes that are exogenous to the state of the economy. For the post-World War II US economy, they read legislative records to identify whether tax changes were due either to inherited deficits or to beliefs about their ability to promote long-term growth. Their method is easily exported to other countries, and it has now become the standard method for assessing the effects of tax changes across a wide range of countries (for example, Guajardo, Leigh, and Pescatori 2014). Mertens and Ravn (2012) improved their measure by splitting their series into anticipated and unanticipated tax changes, so that the effects of fiscal foresight could be addressed. Alesina, Favero, and Giavazzi (2019) have added to the narrative analysis of fiscal consolidations by creating narrative series of fiscal plans. As they emphasize, most fiscal consolidations involve multi-year plans and those effects should be studied as a whole rather than as independent year-by-year isolated changes.

An additional innovation in the identification of fiscal shocks has been the recognition of the importance of instrument "relevance"—that is, whether the proposed instrument is actually correlated with the variable it is supposed to instrument. While early alarms about weak instruments were raised for macro studies by Nelson and Startz (1990) and for microeconomic studies by Bound, Jaeger, and Baker (1995), most macroeconomists began to pay attention to the issue only in the last five to ten years. The structural vector autoregression methodology hid the fact that the estimation of multipliers was actually an instrumental variables estimation. In Ramey (2016) and Ramey and Zubairy (2018), we showed that cumulative multipliers could be estimated in a one-step instrumental variables method based on local projections: cumulative GDP up to horizon h is regressed on cumulative government spending up to horizon h, using an SVAR shock or a narrative variable as an instrument. However, that recognition highlighted a widespread problem: many of the exogenous measures of fiscal policy are not very relevant instruments, at all or in some subsamples. For example, the military news variable I first introduced in Ramey (2011a) is a weak instrument for the post-1954 period, as are the alternative measures of defense news of Fisher and Peters (2010) and Ben Zeev and Pappa (2017). In contrast, the Blanchard and Perotti (2002) shock is a strong instrument by its nature, particularly at short horizons, since it is simply the one-step ahead forecast error of government spending.

In sum, research on the effects of fiscal policy has made significant strides in methodology. The literature now exploits many new datasets. It has imported some innovations from the applied microeconomics literature, and has extended them in important ways that account for anticipations and dynamics. Moreover, those estimates are now converted to multipliers defined in a way that is relevant for policymakers.

A Summary of Estimates of Spending and Tax Multipliers

This section summarizes the actual estimates of fiscal multipliers obtained from the leading methods. I begin with estimates based on aggregate data. I first review the estimated multipliers on government purchases, initially averages and then by state-dependence. Next, I move on to the effects of tax changes and transfer payments. I then discuss estimates of the effects of the American Recovery and Reinvestment Act of 2009 and the fiscal consolidations in Europe.

Government Spending Multipliers Based on Aggregate Data

Table 1 shows a sampling of estimates of government spending multipliers, grouped by method. Virtually all estimates shown are based on present value or undiscounted cumulative multipliers; in some cases, I updated the original estimates to apply best practices. As shown in Figure 2, the cumulative multipliers usually do not vary greatly across horizons up to five years, so there is little difference between average or peak multipliers. The estimates in Panel A show that the estimated multipliers are not very different across the various methods for identifying government spending shocks in time series. Panel B displays estimates based on New Keynesian DSGE models. The multiplier estimates from these models are similar to those from Panel A. On balance, the table shows that for a variety of samples, identification methods, and countries, most of the estimates are around one or below. A few estimates are noticeably above one, such as the Ben Zeev and Pappa (2017) estimate, but they tend to be less precise and are not statistically different from one. Not shown in the table are numerous multiplier estimates based on key features of a country. For example, Iltzetzki, Mendoza, and Végh (2010) estimate how multipliers change across various important features, such as whether an economy has fixed or flexible exchange rates. They find multipliers that vary between 0.1 on impact to 1.4 in the long run (with a 90-percent confidence interval from around 0.75 to 2.1) for fixed exchange rates and from 0.1 to –0.7 for flexible exchange rates. Thus, the range of estimated multipliers may become much wider when one begins to distinguish by key country characteristics.

The results shown in Table 1 are for total government spending or government consumption. Earlier work by Aschauer (1989), Pereira and Flores de Frutos (1999), and others found high returns to public investment. There is surprisingly little recent aggregate evidence on multipliers for public investment. As one example, Iltzetzki, Mendoza, and Végh (2010) found multipliers for public investment that ranged between 0.4 in the short-run to 1.6 in the long-run in their panel of countries.

Even if government spending multipliers are probably one or below on average, might they be higher during bad economic times? In estimating fiscal multipliers, some key states studied by recent papers are recessions or periods of excess slack (typically measured by unemployment rates), constraints on the monetary policy accommodation (such as the zero lower bound), and the ratio of public debt to GDP.

First consider multipliers during recessions or periods of slack. Auerbach and Gorodnichenko (2012), who conducted the pioneering study on this question, used a nonlinear time series model in which the parameters changed across expansions and recessions. They reported a multiplier of 2.2 in recessions and –0.3 in expansions (based on some simplifying assumptions about the state of the economy not changing after the shock). Various other studies have found high multipliers during recessions (for example, Auerbach and Gorodnichenko 2013; Fazzari, Morley, and Panovski 2015; Caggiano, Castelnuovo, Colombo, and Nodari 2015). However, subsequent research has found many of the state-dependent results to be very

Table 1
Estimates of Government Spending Multipliers Using Aggregate Data, No State Dependence
(almost all are cumulative multipliers, typically over horizons between 0 to 20 quarters)

Method/Sample	*Multipliers*	*Comments*
A: Time series analysis		
Updated implementation of Blanchard and Perotti (2002) identified SVAR		The tax response is positive for the 1939Q1–2015Q4 period, but is essentially 0 for the later periods.
1939Q1–2015Q4	0.6 to 0.8	
1947Q1–2015Q4	0.6 to 0.7	
Military news shocks, local projections Ramey and Zubairy (2018) military news		Tax response is positive for 1939Q1–2015Q4 period.
1889Q1–2015Q4	0.6 to 0.8	SE from 0.04 to 0.06
1939Q1–2015Q4	0.7 to 0.8	SE from 0.05 to 0.1
1947Q1–2015Q4	0.5 to 0.7	SE from 0.15 to 0.2
Ben Zeev and Pappa (2017) news, 1947Q1–2007Q4[a]	1.1 to 2	SE from 0.6 to 1
Hall (2019), Barro and Redlick (2011)—based on regressions using annual defense spending.	0.6 to 0.7	The Barro–Redlick analysis nets out effects of changes in tax rates.
Mountford and Uhlig (2009), SVAR with sign restrictions	0.65	Deficit-financed increase in government spending.
Iltzetzki, Mendoza, and Végh (2013), Blanchard–Perotti identification in SVAR, quarterly data, 1960–2007, 44 countries high-income countries	0.3 to 0.7	
Corsetti, Meier, and Müller (2012)	0.7	Based on unconditional model results reported in their Figure 1.
Leigh et al. (2010), Guajardo, Leigh, and Pescatori (2014), 17 OECD countries, 1978–2009, narrative method identification of spending-based fiscal consolidations	0.3	
Alesina, Favero, and Giavazzi (forthcoming). Narrative analysis of austerity plans, 16 OECD economies from 1978–2014.	0.3	
B: Estimated New Keynesian DSGE models		
Cogan et al. (2010), estimated Smets–Wouters DSGE model on US data	0.6 to 0.7	Based on my visual inspection of figures 2, 3, and 4.
Coenen et al. (2012), large-scale macro models used by central banks and IMF, United States and Europe	0.7 to 1	Based on the two-year cumulative multipliers shown in the upper left graph in figure 6.
Zubairy (2014), estimated medium-scale DSGE model on US data	0.7 to 1.05	Deficit financed, model features deep habits.
Leeper, Traum, and Walker (2017), estimated DSGE model on US data	0.7 to 1.36	Active monetary policy, table 7
Sims and Wolff (2018a)	1.07	The multiplier above 1 is due to estimated complementarity of government spending with private consumption.

Note: SVAR is structural vector autoregression. DSGE stands for "dynamic stochastic general equliibrium."
[a] These estimates are based on the analysis in Ramey (2016) using Ben Zeev and Pappa's estimated news series.

fragile to small changes in specification or to improvements in the methods for computing the multipliers from the basic estimates (Alloza 2017; Owyang, Ramey, and Zubairy 2013; Ramey and Zubairy 2018, and associated online appendix). The more robust methods generally fail to produce multipliers above one during recessions or times of slack.

Perhaps these empirical results should not be surprising, given some other results of theory and quantitative models. The only theoretical models that predict countercyclical markups are ones that include significant frictions. For example, Michaillat (2014) presents a stylized model with labor market frictions and finds that the aggregate employment effect of government hiring is countercyclical. However, the multipliers are always below one. Canzoneri, Collard, Dellas, and Diba (2016) present a model with financial frictions that does generate sizeable, though fleeting, multipliers during recessions. They find significantly higher-impact multipliers during recessions, near 2, but the cumulative multipliers fall below 1 after only a few quarters. Standard new Keynesian models do not predict higher multipliers during recessions. Indeed, Sims and Wolff (2018a) employ a medium-scale New Keynesian DSGE model with high-order terms in the approximations and find that this otherwise standard model implies mildly *procyclical* multipliers.

The situation is different with respect to periods when interest rates are near the zero lower bound or when monetary policy accommodates government spending increases (such as during World War II in the United States). Numerous New Keynesian DSGE models show that multipliers can be higher than one when monetary policy is constrained by the zero lower bound on interest rates. At the zero lower bound, an increase in government spending provides extra stimulus by increasing expected inflation, which lowers the real interest rate (Farhi and Werning 2016). Calibrated models such as the ones analyzed by Christiano, Eichenbaum, and Rebelo (2011) and Coenen et al. (2012) can produce multipliers that range between 2 and 3 when the period of monetary accommodation is sufficiently long. Some recent empirical work has found some evidence of higher multipliers, ranging from 1.5 to 2.5 at the zero lower bound for Japan (Miyamoto, Nguyen, and Sergeyev 2018) and around 1.5 for historical samples in the United States (Ramey and Zubairy 2018).

Finally, there is evidence that government spending multipliers may be negatively related to the public debt-to-GDP ratio. For example, Iltzetzki, Mendoza, and Végh (2013) find that countries with a government debt-to-GDP ratio above 60 percent have an impact multiplier of 0 and a long-run multiplier of –3 (estimated less precisely but still statistically below 0).

In summary, most estimates of government spending multipliers for general categories of government spending for averages over samples are in the range of 0.6 to 0.8, or perhaps up to 1. The evidence for multipliers above one during recessions or times of slack is typically not robust. However, some initial explorations suggest that government spending multipliers could be higher at times when monetary policy accommodates fiscal policy, such as during periods at the zero lower bound of interest rates or wartime.

Tax and Transfer Multipliers Based on Aggregate Data

I now turn to the leading estimates of tax and transfer multipliers at the aggregate level. Tax multipliers are generally negative since an increase in taxes lowers GDP. Table 2 shows the estimates from time series and New Keynesian DSGE estimates for tax rate changes. In contrast to government spending multipliers, which vary only a small amount across horizons, many estimates of tax multipliers start out low on impact but then build. Thus, I report the cumulative multipliers for the horizon where they peak. I should also note that most of the multipliers are calculated without allowing feedback from induced output changes to revenue but several (noted in the table) allow for dynamic feedback.

Most of the time series estimates based on narrative methods of identification are quite high (in absolute value), generally between –2 and –3. These narrative-based estimates are striking not only for their magnitudes, but also for their uniformity across countries and even across various methods of estimation. These estimates are much higher (in absolute value) than the tax multipliers reported by Blanchard and Perotti (2002). As discussed above, those estimates were based both on their assumed elasticity of tax revenue to output and on their unusual way of computing multipliers. Barro and Redlick (2011) estimate multipliers around –1.1. It may be that their use of various approximations and constraints on dynamics account for their smaller estimate. On the other hand, Mountford and Uhlig's (2009) estimates using sign restrictions are –5.

In contrast, the New Keynesian DSGE model estimates are much lower. Panel B of Table 2 shows that most New Keynesian model estimates yield multipliers that are below 1 in absolute value. Thus, there is a conflict between the narrative-based time series estimates and the New Keynesian estimates.

There is a small literature on whether tax multipliers differ by the state of the economy. So far, this literature offers fairly uniform answers. Eskandari (2015) and Demirel (2016) find, using the Romer and Romer (2010) narrative tax shocks, that tax multipliers are *greater* during times of low unemployment than times of high unemployment. Alesina, Azzalini, Favero, Giavazzi, and Miano (2018) also find higher multipliers in expansions using their narrative of fiscal plans across OECD countries. These results are consistent with the one New Keynesian analysis of this issue using the dynamic stochastic general equilibrium approach. Sims and Wolff (2018b) obtain estimates of tax multipliers that are procyclical: for example, their capital tax multiplier is 1 in recessions and almost 2 in expansions.

There has been very little work on the aggregate effects of transfers. Romer and Romer (2016) used changes in Social Security benefit increases to study the effects on macroeconomic variables. They found that permanent increases in benefits led to a roughly equal rise in consumption in the short-run, but the effect dissipated quickly. Temporary increases in benefits had no significant effect on aggregate consumption. Coenen et al. (2012) studied general transfers and directed transfers across the various New Keynesian DSGE models used at policy institutions. They found that general transfers had multipliers between 0.2 and 0.6, with the higher ones occurring with monetary accommodation. In contrast, targeted transfers (to

Table 2
Estimates of Tax Change Multipliers Using Aggregate Data, No State Dependence
(† denotes multipliers computed using the cumulative actual response of tax revenues or deficits in the denominator)

Method/Sample	*Largest cumulative multiplier within first 5 years*	*Comments*
A: Time Series Methods		
Mountford and Uhlig (2009), SVAR with sign restrictions, US data	–5†	
Romer and Romer (2010), narrative series of tax changes unrelated to current economy, US data, 1950 to 2007, dynamic single equation model or VAR	–2.5 to –3	The output effects take time to build.
Barro and Redlick (2011), historical annual US data, tax rate shocks.	–1.1	
Mertens and Ravn (2013, 2014), refinement of Romer and Romer series used in a proxy SVAR	–2.5 to –3†	The peak output effects occur in the first 18 months.
Cloyne (2013), narrative, UK	–2.5	
Hayo and Uhl (2013), narrative, Germany	–2.4	
Guajardo, Leigh, and Pescatori (2014), 17 OECD countries, 1978–2009, narrative taxed-based consolidations	–3	
Riera-Crichton, Vegh, and Vuletin (2016), narrative analysis of fiscal consolidations in 15 industrialized countries from 1980 to 2009, with focus on VAT rate changes	–3.5	
Alesina, Azzalini, Favero, Giavazzi, and Miano (2018), narrative analysis of austerity plans, 16 OECD economies from 1978 to 2014, taxed-based consolidations		
Based on static primary surplus	–1 to –1.6	
Based on actual response of primary surplus	–2.3 to –3.7†	
B: Estimated New Keynesian DSGE models		
Coenen et al. (2012), large-scale macro models used by central banks and IMF, United States, and Europe. Two-year cuts in tax, no monetary accommodation		
Consumption tax	–0.2 to –0.4	
Labor tax	–0.2 to –0.4	
Corporate income tax	0 to –0.15	
Zubairy (2014)		
Labor tax	–0.7 to –1	
Capital tax	–0.2	
Sims-Wolff (2018b), medium scale New Keynesian DSGE model that allows for higher-order terms.		Steady-state multipliers
Consumption tax	–0.6	
Labor tax	–1	
Capital tax	–1.5	

Note: SVAR is structural vector autoregression. VAR is vector autoregression. VAT is value added tax. DSGE stands for dynamic stochastic general equilibrium.

households that were financially constrained) yielded multipliers as high as 2 in some models when there was monetary accommodation.

In sum, most time series estimates of tax rate change multipliers indicate that they are very large, at least –2 to –3. This contrasts with the results from estimated New Keynesian dynamic stochastic general equilibrium, where the multipliers (in absolute value) are typically below 1 and never higher than 1.5. There is not much aggregate time series evidence for sizeable multipliers for temporary transfers, though calibrated New Keynesian models suggest they can be high if they are targeted and if monetary policy is accommodative.

Multiplier Estimates Based on Subnational Data

One of the important innovations in the fiscal multiplier literature, as mentioned earlier in this paper, has been the application of applied microeconomics-type identification methods to the estimation of parameters of use for macroeconomics. These include studies of panels or cross-sections of US states or provinces in other countries, as well as household-level estimates of marginal propensities to spend out of temporary transfers.

Chodorow-Reich (forthcoming) summarizes the panel and cross-section multipliers from individual studies, so I refer the reader to his tables. Many of the subnational multipliers for government purchases, temporary tax rebates, and transfers lie between 1.5 to 2. Thus, they tend to be higher than the aggregate-level estimates of multipliers.

As noted earlier, subnational multipliers are not the same as aggregate multipliers. The relationship between subnational multipliers and aggregate multipliers depends on many features, including how the spending is financed, whether there are spillovers across regions, whether there is a currency union, and whether the aggregate economy is at the zero lower bound. For discussion of some of the theoretical considerations when drawing implications from subnational multiplier estimates to aggregate estimates, see Nakamura and Steinsson (2014), Farhi and Werning (2016), and Chodorow-Reich (forthcoming). In some instances, the subnational multipliers are expected to be higher than the aggregate multipliers, whereas in other instances they are expected to be lower. There is no general rule. Dupor and Guerrero (2017) conduct an empirical investigation in which they directly compare estimates based on a state-level panel to those obtained when the state data are aggregated to the national level. They obtain similar multiplier estimates across the two datasets, though quite low, between 0 and 0.5.

Multipliers in the Wake of the Financial Crisis

A number of researchers and commentators have argued that the effects of the stimulus from the American Recovery and Reinvestment Act of 2009 and the subsequent fiscal consolidations in European countries were much larger than indicated by multipliers during average times. A common theme is that the high unemployment rates and lower bound on interest rates combined to raise the multipliers. But as shown in the previous sections, there is no robust evidence of higher multipliers during recessions or times of slack, for either spending or taxes. In fact, all studies of state dependence for tax multipliers find *higher* multipliers during expansions. However, there is evidence from historical periods in the United States and from Japan, as well as from New Keynesian models, that multipliers can be higher than one during periods of monetary accommodation such as the zero lower bound on interest rates. Thus, it is possible that multipliers could have been higher after the financial crisis.

Consider first the fiscal consolidations in Europe, aimed at reducing government deficits and debt. Blanchard and Leigh (2013, 2014) presented evidence that countries that implemented bigger fiscal consolidations grew more slowly than forecasted by the IMF and other organizations. They concluded that the models used by forecasters assumed values of multipliers that were too small. Górnicka, Kamps, Koster, and Leiner-Killinger (2018) gathered data on the forecasters' assumed values of multipliers and found that they were very low, around 0.25. They then calculated that the "true" multipliers were higher, though they never exceeded one.

The conclusions of Górnicka et al. (2018) are consistent with some other analyses of the size of multipliers in the European fiscal consolidations. For example, Alesina, Favero, and Giavazzi (2019) use their narrative dataset of fiscal consolidation plans across OECD countries to study whether fiscal multipliers were greater in the immediate post-financial crisis years. They find no evidence that multipliers were greater. At this point, the evidence does not suggest that multipliers were larger than normal for the fiscal consolidations in Europe.

The American Recovery and Reinvestment Act (ARRA) of 2009 was the leading stimulus program in the US economy. This program was a mix of spending and transfers to states and individuals. As Table 3 shows, none of the New Keynesian DSGE models find multipliers above 1 for this program, with the exception of one experiment by Coenen et al. (2012) that included two years of monetary accommodation. While interest rates were indeed at the zero lower bound during those years, Swanson and Williams (2014) present evidence that yields on one- and two-year Treasury bills were unconstrained from 2008 to 2010, "suggesting that monetary policy and fiscal policy were about as effective as usual during this period."

In contrast, the cross-state estimates of the effects of the American Recovery and Reinvestment Act are typically much higher. Chodorow-Reich (forthcoming) presents an extremely valuable standardization and synthesis of the leading estimates of the effects of the stimulus act on job creation across US states. This literature emphasizes employment effects, mainly because the employment data have less measurement error than gross state product. These estimates are based on strong applied microeconomic methods. His cross-state natural experiment estimates indicate multipliers from 1.7 to 2 for gross state product and $50,000 per job-year created. Building on Farhi and Werning's (2016) theoretical analysis, Chodorow-Reich (forthcoming) argues that these subnational multipliers are *lower bounds* on the national multipliers during a liquidity trap. Thus, he argues that the multiplier from the American Recovery and Reinvestment Act was at least 2.

Table 3
Multipliers for the American Recovery and Reinvestment Act (ARRA)

Method/Sample	*Peak cumulative multipliers within first 5 years*	*Comments*
Cogan et al. (2010)	0.6 to 0.7	
Coenen et al. (2012), large-scale macro models used by central banks and IMF, US, and Europe		From figure 7. These are the peak instantaneous multipliers.
No monetary accommodation	0.3 to 0.5	
1 year monetary accommodation	0.4 to 0.6	
2 years monetary accommodation	0.5 to 1.8	
Drautzburg and Uhlig (2015), medium-scale New Keynesian DSGE model, with ZLB, credit constraints	0.5	Multipliers become negative in the long run because of the necessary increase in taxation.
Chodorow-Reich (forthcoming), based on cross-state estimates and theoretical arguments about the relationship between subnational and national multipliers at the ZLB.		
Gross State Product multiplier	1.7 to 2	
Cost per job year	2 job-years per $100K	

Note: ZLB is zero lower bound.

But there is reason to suspect that the state-level estimates of the effects of the American Recovery and Reinvestment Act presented by Chodorow-Reich (forthcoming) are probably overestimates for the national-level multipliers. His cross-state estimates answer one question: "How much extra employment was induced in the average state by each *$1 of ARRA spending* by the federal government?" But the question relevant for the aggregate effects is a different one: "How much extra aggregate employment was generated by each *$1 of government spending induced by ARRA spending* by the federal government?" Chodorow-Reich uses per capita values of spending and employment in each state, and his cross-state estimates give equal weight to North Dakota and California, which is fine for answering the first cross-state question. But if there is heterogeneity in the treatment effects, the estimates will not give estimates that are nationally representative.[4] The data need to be weighted by population or in some other way to obtain nationally representative results. A second issue is that Chodorow-Reich's measure of spending is federal ARRA spending, which again is appropriate for measuring the first cross-state question. However, ARRA spending stimulated state and local spending more than dollar for dollar (Leduc and Wilson

[4] Most of the literature using cross-sectional estimates has used per capita estimates and has not weighted the estimates. However, Dupor and Mehkari (2016) started weighting the estimates and discovered that weighted estimates of the American Recovery and Reinvestment Act of 2009 are much lower than unweighted estimates.

Table 4
Conversion of Chodorow-Reich Estimates to Nationally Representative Estimates

	Cumulative employment multiplier estimates (number of job-years created per $100K of ARRA spending)
Chodorow-Reich (forthcoming) headline estimates (his table 1, column 4)	2.01 (0.59)
Weighted estimates (using December 2008 population of state)	1.15 (0.72)
Weighted estimates based on total spending, including induced spending by states	0.89 (0.45)

Note: ARRA is American Recovery and Reinvestment Act. Estimates presented in the last two rows are the author's estimates, based on Chodorow-Reich's programs and data in his forthcoming paper. See the text and online appendix for more detail and programs.

2017). Thus, multipliers that use only the ARRA transfers to the states will overestimate the multiplier per dollar spent across all levels of government.

Table 4 shows the effects of adjusting the employment response estimates to make them more suitable for answering the question about aggregate effects of federal government spending. The first row shows Chodorow-Reich's (forthcoming) preferred estimates, which use all three of the leading instruments for estimating cross-state effects of the American Recovery and Reinvestment Act: Medicaid formulas, Department of Transportation formulas, and a combination of multiple agency formulas. The estimates are for job-years created for each $100,000 of ARRA spending. Thus, the estimate of 2.01 implies that each $100,000 of ARRA spending creates two job-years of employment. The second row of Table 4 shows the results of my re-estimating Chodorow-Reich's model (using his replication files) but weighting by initial state population (in December 2008) to make the estimates representative of national data. The point estimate falls to 1.15 and the standard error is higher at 0.72. The third row of Table 4 shows the estimates when spending across the levels of government are substituted for the ARRA spending. Here, I use the Chodorow-Reich combination of instruments, and I weight by initial state population. The jobs multiplier estimate is now 0.89 with a standard error of 0.45. Chodorow-Reich's method for converting jobs multipliers to output multipliers is nearly one-for-one, so the 0.89 estimate also implies an output multiplier around 0.9. Thus, once the cross-state estimates are made nationally representative and include all spending, they look very much like the aggregate estimates and lie below unity.

Two important caveats about these adjusted estimates are in order. First, reweighting by population gives very large influence to just a few of the 50 states. Second, the great instrument relevance in Chodorow-Reich's analysis disappears once I add state and local spending to American Recovery and Reinvestment Act spending. In other words, the instruments that are so good at explaining ARRA spending are not very good at explaining total government spending in the state.

Thus, it appears that the natural experiments exploited by the ARRA literature are rich enough to answer questions about the effects of ARRA spending on a cross-state basis, but not to answer questions about the aggregate effects of government spending induced by the ARRA.

In sum, a number of commentators and researchers have argued that multipliers may have been higher than usual after the financial crisis. I interpret most of the evidence at this point as suggesting that they were not higher than usual.

Conclusion

The fiscal literature has made tremendous progress in the ten years since the start of the global financial crisis. The range of estimates for average fiscal multipliers has been reduced considerably, particularly for government purchases. On average, government purchases multipliers are likely to be between 0.6 and 1. Narrative-based time series estimates point to tax rate change multipliers between –2 and –3, though these are significantly greater in magnitude than those predicted by New Keynesian DSGE models. However, there is still ongoing debate about specific contexts, such as the size of fiscal multipliers during "bad" times and the effects of other characteristics, such as exchange rate regimes.

Across industrialized countries, most of the temporary stimulus packages enacted from 2007 to 2009 in response to the global financial crisis took the form of transfer payments or lump-sum tax rebates (Oh and Reis 2012). Policymakers were "flying blind" in that they had little research to guide them at that time. Had they known then some of the results now emerging from the literature, they might have fashioned the stimulus packages differently, perhaps relying more on tax rate cuts and less on expenditures.

I believe the literature would benefit from progress in three main areas. First, the literature needs to catch up to the current policy discussions by focusing more on the short-run and long-run effects of infrastructure investment. The few studies at the aggregate and subnational levels suggest that these multipliers can be very large in some contexts (for example, Leduc and Wilson 2013). Second, researchers need to be careful about their implementation decisions. Seemingly small changes, such as how multipliers are actually calculated, can make a big difference. Finally, researchers should continue to innovate along the lines they have pursued in the last ten years, exploiting new datasets, extending theoretical models, and improving estimation techniques. As part of this innovation, researchers should continue to analyze the link between micro estimates and aggregate effects.

■ *I am grateful for helpful comments from Alberto Alesina, Gabriel Chodorow-Reich, Martin Eichenbaum, Carlo Favero, Mark Gertler, Francesco Giavazzi, Gordon Hanson, Daniel Leff Yaffe, Karel Mertens, Maury Obstfeld, Timothy Taylor, Linda Tesar, Sarah Zubairy, and participants at the July 2018 NBER Conference "Global Financial Crisis @10."*

References

Acconcia, Antonio, Giancarlo Corsetti, and Saverio Simonelli. 2014. "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment." *American Economic Review* 104(7): 2185–2209.

Alesina, Alberto, Gualtiero Azzalini, Carlo Favero, Francesco Giavazzi, and Armando Miano. 2018. "Is it the 'How' or the 'When' that Matters in Fiscal Adjustments?" *IMF Economic Review* 66(1): 144–88.

Alesina, Alberto, Carlo Favero, and Francesco Giavazzi. 2019. *Austerity: When It Works and When It Doesn't.* Princeton University Press.

Alloza, Mario. 2017. "Is Fiscal Policy More Effective in Uncertain Times or During Recessions?" Banco de España Working Paper 1730.

Aschauer, David Alan. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23(2): 177–200.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2018. "The Intertemporal Keynesian Cross." NBER Working Paper 25020.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4(2): 1–27.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2013. "Fiscal Multipliers in Recession and Expansion." Chap. 2 in *Fiscal Policy After the Financial Crisis*, edited by Alberto Alesina and Francesco Giavazzi. University of Chicago Press.

Barro, Robert J. 1981. "Output Effects of Government Purchases." *Journal of Political Economy* 89(6): 1086–1121.

Barro, Robert J., and Charles J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126(1): 51–102.

Baxter, Marianne, and Robert G. King. 1993. "Fiscal Policy in General Equilibrium." *American Economic Review* 83(3): 315–34.

Ben Zeev, Nadav, and Evi Pappa. 2017. "Chronicle of a War Foretold: The Macroeconomic Effects of Anticipated Defence Spending Shocks." *Economic Journal* 127(603): 1568–97.

Blanchard, Olivier J., and Daniel Leigh. 2013. "Growth Forecast Errors and Fiscal Multipliers." *American Economic Review* 103(3): 117–20.

Blanchard, Olivier J., and Daniel Leigh. 2014. "Learning about Fiscal Multipliers from Growth Forecast Errors." *IMF Economic Review* 62(2): 179–212.

Blanchard, Olivier, and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117(4): 1329–68.

Bound, John, David A. Jaeger, and Regina M. Baker. 1995. "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90(430): 443–50.

Caggiano, Giovanni, Efrem Castelnuovo, Valentina Colombo, and Gabriela Nodari. 2015. "Estimating Fiscal Multipliers: News from a Nonlinear World." *Economics Journal* 125(584): 746–76.

Caldara, Dario, and Christophe Kamps. 2017. "The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers." *Review of Economic Studies* 84(3): 1015–40.

Canova, Fabio, and Luca Sala. 2009. "Back to Square One: Identification Issues in DSGE Models." *Journal of Monetary Economics* 56(4): 431–49.

Canzoneri, Matthew, Fabrice Collard, Harris Dellas, and Behzad Diba. 2016. "Fiscal Multipliers in Recessions." *Economic Journal* 126(590): 75–108.

Chodorow-Reich, Gabriel. Forthcoming. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *American Economic Journal: Economic Policy.*

Christiano, Lawrence, Martin Eichenbaum, and Sergio Rebelo. 2011. "When is the Government Spending Multiplier Large?" *Journal of Political Economy* 119(1): 78–121.

Cloyne, James. 2013. "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom." *American Economic Review* 103(4): 1507–28.

Coenen, Günter, Christopher J. Erceg, Charles Freedman, Davide Furceri, Michael Kumhof, René Lalonde, Douglas Laxton, Jesper Lindé, Annabelle Mourougane, Dirk Muir, Susanna Mursula, Carlos de Resende, John Roberts, Werner Roeger, Stephen Snudden, Mathias Trabandt, and Jan in't Veld. 2012. "Effects of Fiscal Stimulus in Structural Models." *American Economic Journal: Macroeconomics* 4(1): 22–68.

Cogan, John F., Tobias Cwik, John B. Taylor, and Volker Wieland. 2010. "New Keynesian versus Old Keynesian Government Spending Multipliers." *Journal of Economic Dynamics and Control* 34(3): 281–95.

Corsetti, Giancarlo, André Meier, and Gernot J. Müller. 2012. "What Determines Government Spending Multipliers?" *Economic Policy* 27(72): 521–65.

Demirel, U. Devrim. 2016. "The Short-Term Effects of Tax Changes—Evidence for State Dependence." Congressional Budget Office Working Paper 2016-03.

Drautzburg, Thorsten, and Harald Uhlig. 2015. "Fiscal Stimulus and Distortionary Taxation." *Review of Economic Dynamics* 18(4): 894–920.

Dupor, Bill, and Rodrigo Guerrero. 2017. "Local and Aggregate Fiscal Policy Multipliers." *Journal of Monetary Economics* 92: 16–30.

Dupor, Bill, and M. Saif Mehkari. 2016. "The 2009 Recovery Act: Stimulus at the Extensive and Intensive Labor Margins." *European Economic Review* 85: 208–28.

Eskandari, Ruhollah. 2015. "State-Dependent Macroeconomic Effects of Tax Changes: Estimates Based on Narrative Records in the US." In *Three Essays on Tax Policy, Macroeconomics, and Corporate Policies*. University of Milan dissertation.

Farhi, Emmanuel, and Iván Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." Chap. 31 in *Handbook of Macroeconomics*, Vol. 2, edited by John B. Taylor and Harald Uhlig, 2417–92. Amsterdam: Elsevier.

Fazzari, Steven M., James Morley, and Irina Panovska. 2015. "State-Dependent Effects of Fiscal Policy." *Studies in Nonlinear Dynamics and Econometrics* 19(3): 285–315.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *Economic Journal* 120(544): 414–36.

Gali, Jordi, J. David López-Salido, and Javier Vallés. 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association* 5(1): 227–270.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2018. "Bartik Instruments: What, When, Why, and How." NBER Working Paper 24408, revised July 2018.

Gordon, Robert J., and Robert Krenn. 2010. "The End of the Great Depression 1939–41: Policy Contributions and Fiscal Multipliers." NBER Working Paper 16380.

Górnicka, Lucyna, Christophe Kamps, Gerrit Koster, and Nadine Leiner-Killinger. 2018. "Learning about Fiscal Multipliers during the European Sovereign Debt Crisis: Evidence from a Quasi-Natural Experiment." European Central Bank Working Paper 2154, May.

Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. "Expansionary Austerity? International Evidence." *Journal of the European Economic Association* 12(4): 949–68.

Hall, Robert E. 1980 "Labor Supply and Aggregate Fluctuations." *Carnegie-Rochester Conference Series on Public Policy* 12: 7–33.

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

Hayo, Bernd, and Matthias Uhl. 2013. "The Macroeconomic Effects of Legislated Tax Changes in Germany." *Oxford Economic Papers* 66(2): 397–418.

House, Christopher L., Christian Proebsting, and Linda L. Tesar. 2017. "Austerity in the Aftermath of the Great Recession." NBER Working Paper 23147.

House, Christopher L., and Matthew D. Shapiro. 2006. "Phased in Tax Cuts and Economic Activity." *American Economic Review* 96(5): 1835–49.

Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh. 2013. "How Big (Small?) Are Fiscal Multipliers?" *Journal of Monetary Economics* 60(2): 239–254.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96(5): 1589–1610.

Leduc, Sylvain, and Daniel Wilson. 2013. "Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment." *NBER Macroeconomics Annual* 27(1): 89–142.

Leduc, Sylvain, and Daniel Wilson. 2017. "Are State Governments Roadblocks to Federal Stimulus? Evidence on the Flypaper Effect of Highway Grants in the 2009 Recovery Act." *American Economic Journal: Economic Policy* 9(2): 253–92.

Leeper, Eric M., Nora Traum, and Todd B. Walker. 2017. "Clearing up the Fiscal Multiplier Morass." *American Economic Review* 107(8): 2409–54.

Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. 2013. "Fiscal Foresight and Information Flows." *Econometrica* 81(3): 1115–45.

Leigh, Daniel, Pete Devries, Charles Freedman, Jaime Guajardo, Douglas Laxton, and Andrea Pescatori. 2010. "Will it Hurt? Macroeconomic Effects of Fiscal Consolidation." Chap. 3 of *World Economic Outlook: Recovery, Risk, and Rebalancing*, October 2010. International Monetary Fund.

Mertens, Karel, and Morten O. Ravn. 2012. "Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated US Tax Policy Shocks." *American Economic Journal: Economic Policy* 4(2): 145–81.

Mertens, Karel, and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103(4): 1212–47.

Mertens, Karel, and Morten O. Ravn. 2014. "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics* 68

(December Supplement): S1–S19.

Michaillat, Pascal. 2014. "A Theory of Countercyclical Government Multiplier." *American Economic Journal: Macroeconomics* 6(1): 190–217.

Miyamoto, Wataru, Thuy Lan Nguyen, and Dmitriy Sergeyev. 2018. "Government Spending Multipliers Under the Zero Lower Bound." *American Economic Journal: Macroeconomics* 10(3): 247–77.

Mountford, Andrew, and Harald Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24(6): 960–92.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions." *American Economic Review* 104(3): 753–92.

Nelson, Charles R., and Richard Startz. 1990. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica: Journal of the Econometric Society* 58(4): 967–76.

Oh, Hyunseung, and Ricardo Reis. 2012. "Targeted Transfers and the Fiscal Response to the Great Recession." *Journal of Monetary Economics* 59(Supplement): S50–S64.

Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data." *American Economic Review* 103(3): 129–34.

Parker, Jonathan A., Nicolas S. Souleles, David S. Johnson, and Robert McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103(6): 2530–53.

Pereira, Alfredo M., and Rafael Flores de Frutos. 1999. "Public Capital Accumulation and Private Sector Performance." *Journal of Urban Economics* 46(2): 300–322.

Ramey, Valerie A. 2011a. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126(1): 1–50.

Ramey, Valerie A. 2011b. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49(3): 673–85.

Ramey, Valerie A. 2013. "Government Spending and Private Activity." Chap. 1 in *Fiscal Policy after the Financial Crisis,* edited by Alberto Alesian and Francesco Giavazzi. University of Chicago Press.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." Chap 2 in *Handbook of Macroeconomics*, edited by John B. Taylor and Harald Uhlig. Vol. 2, 71–162. Amsterdam: Elsevier.

Ramey, Valerie A. Forthcoming. "Fiscal Policy: Tax and Spending Multipliers in the United States." In *Rethinking Macroeconomic Policy IV.* MIT Press.

Ramey, Valerie A., and Matthew D. Shapiro. 1998. "Costly Capital Reallocation and the Effects of Government Spending." *Carnegie Rochester Conference on Public Policy* 48(1): 145–194.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy* 126(2): 850–901.

Riera-Crichton, Daniel, Carlos A. Vegh, and Guillermo Vuletin. 2016. "Tax Multipliers: Pitfalls in Measurement and Identification." *Journal of Monetary Economics* 79: 30–48.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100(3): 763–801.

Romer, Christina D., and David H. Romer. 2016. "Transfer Payments and the Macroeconomy: The Effects of Social Security Benefit Increases, 1952–1991." *American Economic Journal: Macroeconomics* 8(4): 1–42.

Sims, Eric, and Jonathan Wolff. 2018a. "The Output and Welfare Effects of Government Spending Shocks over the Business Cycle." *International Economic Review* 59(3): 1403–35.

Sims, Eric and Jonathan Wolff. 2018b. "The State-Dependent Effects of Tax Shocks." *European Economic Review* 107: 57–85.

Swanson, Eric T., and John C. Williams. 2014. "Measuring the Effect of the Zero Lower Bound on Medium-and Longer-Term Interest Rates." *American Economic Review* 104(10): 3154–85.

Woodford, Michael. 2011. "The Simple Analytics of the Government Expenditure Multiplier." *American Economic Journal: Macroeconomics* 3(1): 1–35.

Zubairy, Sarah. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *International Economic Review* 55(1): 169–195.