

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

By

Valerie A. Ramey
University of California, San Diego and NBER

July 3, 2018

Abstract

This paper takes stock of what we have learned from the “Renaissance” in fiscal research in the ten years since the financial crisis. I first summarize the new innovations in methodology and discuss the various strengths and weaknesses of the main approaches. Reviewing the estimates, I come to the surprising conclusion that the bulk of the estimates for *average* spending and tax change multipliers lie in a fairly narrow range, 0.6 to 0.8 for spending multipliers and -2 to -3 for tax change multipliers. However, I identify economic circumstances in which multipliers lie outside those ranges. I conclude by reviewing the debate on whether multipliers were higher on the stimulus spending in the U.S. and the fiscal consolidations in Europe.

* Prepared for the NBER Conference “Global Financial Crisis @10.”

I. Introduction

When the financial crisis hit ten years ago and interest rates fell to their lower bound, policymakers around the world turned to fiscal stimulus packages in order to prevent their economies from freefalling into another Great Depression. But as declining GDP and tax revenues led to deteriorating government budgets and worries about sovereign debt, numerous countries abandoned stimulus packages and instead adopted fiscal consolidation measures. While attempting to forecast the impacts of these various fiscal programs, policymakers and even most academics were surprised to learn that not only was there no consensus about the size of the effects of fiscal policy but that there had not even been much research on the topic since the 1960s. This situation changed quickly as armies of researchers across many countries turned their attention to this important, but long-neglected, topic. Indeed, a positive by-product of the financial crisis has been a renaissance in fiscal research.

In the last ten years, important progress has been made on all three important methodological fronts: theory, empirical methods, and data. On the theory front, we now have a much better understanding of how deviations from the classic Baxter-King (1992) neoclassical benchmark affect multipliers. The theoretical innovations include the analysis of the effects of sticky prices, hand-to-mouth consumers, lower bounds on policy interest rates, currency unions, fiscal outlays financed by outside entities, and anticipations on the reactions of macroeconomic variables to fiscal policy. The contributions in the realm of empirical methods include important new ways of identifying exogenous variation in policy, standardization of methods for computing multipliers, and the incorporation of state dependence. On the data front, researchers now have much more data that can be used to estimate effects. In addition to newly constructed historical and cross-sectional data sets, researchers are also exploiting the rich new data created

by the variety of policymakers' fiscal responses to the crisis. All of 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.

This paper takes a snapshot of the state of knowledge about the effects of fiscal policy ten years after the global financial crisis. 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 theory and evidence available from U.S. data at that time that the multiplier was “probably between 0.8 and 1.5” though “reasonable people” could argue “that the data do not reject 0.5 or 2.” The current paper 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 since the methods for estimating long-run effects are quite different.

My summary of the current state of knowledge about the effects of fiscal policies is as follows:

Government purchases multipliers

- On average, multipliers on general government purchases in developed countries are positive but less than or equal to unity, meaning that government purchases raise GDP but often crowd out some private activity. The bulk of the estimates across the leading methods of estimation and samples lie in a surprisingly narrow range of 0.6 to 0.8. However, this range widens once one distinguishes country characteristics, such as the exchange rate regime.

- There is far less evidence for infrastructure investment, but the few estimates available suggest multipliers that may be well above unity. However, some of the highest estimates are for infrastructure products whose duplication would not be expected to have the same high marginal product in already-developed economies (e.g. the U.S. interstate highway system).
- The evidence for higher spending multipliers during recessions or times of high unemployment is fragile. The most robust results suggest multipliers of one or below during these periods.
- Both empirical time series estimates for the U.S. and Japan and estimated New Keynesian DSGE models for the U.S. and Europe suggest that multipliers could be 1.4 to 2.6 during periods in which monetary policy is very accommodative, such as zero lower bound periods. The higher estimates are less precisely estimated and not necessarily robust to the empirical sample selection or DSGE specification, so more research needs to be conducted before these results can be considered definitive.

Tax rate change multipliers

- On average, multipliers for tax changes involving tax rate changes are surprisingly large and surprisingly uniform across a number of countries. The bulk of the estimates vary between -2 and -3. A few studies find that the tax multipliers are greater in magnitude during expansions than in recessions.
- Economic activity reacts strongly to tax rate changes that are implemented with delays or phased in slowly. News of a future tax rate cut lowers economic activity in the period between announcement and implementation and boosts it after the implementation.

Multipliers in the Wake of the Financial Crisis

- The debate on whether multipliers were higher for the stimulus packages in the U.S. depend on aggregate versus subnational or household evidence. I present arguments and evidence that the current estimates at the subnational and household level cannot be applied to the national level.
- Estimates of multipliers on fiscal consolidation packages after the financial crisis depend on whether the packages were more tax-based or spending-based and the methodology used. Narrative estimates and case studies suggest that multipliers were not larger. Indirect evidence based on the correlation between forecast errors suggests that multipliers were larger than those incorporated into the large macro models of the time.

The outline of the paper is as follows. Section II briefly reviews how theory highlights the dependence of the size of the multipliers on numerous features of the policy and the economy. Section III summarizes strengths and weaknesses of each of the leading approaches to identifying exogenous shifts in fiscal policy. Section IV highlights the innovations of the last ten years in estimating fiscal multipliers. One of the interesting findings is that the wide range of multipliers reported earlier is reduced significantly once methods for calculating multipliers are standardized. Section V reviews the leading estimates of spending and tax multipliers, including those from aggregate time series, estimated theoretical models, and from subnational units and households. It also discusses the complexities of drawing aggregate inferences from parameters estimated on household data. Section VI asks what we know about whether multipliers were higher in the wake of the financial crisis. Section VII concludes.

II. What Does Theory Predict?

If we simply want to know how much GDP changes if we increase government spending by one dollar or reduce tax rates by one percentage point, why do we need theory? As this section explains, the description of the policy in the previous sentence is incomplete. Theory tells us that there is not one government spending or tax multiplier. Rather, the impact on output and other variables potentially depends on (i) the time horizon, (ii) the type of spending or taxes that changed, (iii) how it was financed, (iv) the persistence of the change, (v) whether it was anticipated, (vi) what individuals expected about the future, (vii) how the policy was distributed across potentially heterogeneous agents, (viii) how monetary policy reacted, (ix) what was the state of the economy when the policy took effect, and (x) what other features characterize the economy in question, such as the level of development, the exchange rate regime, and the openness. And this list is incomplete. Because policymakers cannot conduct randomized control trials, virtually all estimates are based on time series, narrative, or natural experiment identification using convenience samples determined by historical happenstance. In order to understand whether a particular estimate of fiscal effects is suitable for use in predicting the impact of a proposed policy, one must understand how the current circumstances differ from those present in the sample that was used to generate the estimate.

Most researchers and policymakers' first exposure to the theoretical effects of fiscal policy is the Keynesian-Cross closed economy model of undergraduate textbooks. In this simple model, which assumes that GDP is demand-determined, the government spending multiplier is the inverse of one minus the marginal propensity to consume. A marginal propensity to consume of 0.5 yields a multiplier of two. Since taxes enter only through disposable income, the tax multipliers are smaller than the spending multipliers. Expansion of the model to consider the

marginal propensity to import, taxes, 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. An increase in government spending has a negative wealth effect since the government is extracting resources from the private sector. This negative wealth effect raises GDP because it causes the poorer households to supply more labor (but decrease consumption). Distortionary tax rate changes can have potentially large effects. Contrary to the simple Keynesian model, they work through “supply side” channels.

The New Keynesian dynamic stochastic general equilibrium (DSGE) models meld the insights from the traditional Keynesian and neoclassical approaches. The standard representative agent New Keynesian model (RANK) with non-accommodative monetary policy tends to produce multipliers below one for government spending. In the last decade, the RANK models have been expanded to include heterogeneity (two agent or heterogeneous agent) TANK and HANK models. In these models, a share of the consumers behave like the traditional Keynesian consumers rather than being permanent income consumers. These agents instead consume according to “rule-of-thumb” or “hand-to-mouth” rules, immediately consuming all income they receive, whether it is temporary or permanent. Alternatively, the models have explored the effects of fiscal policy when monetary policy is constrained by the zero lower bound on interest rates. Both of these extensions result in higher multipliers, often above unity.

However, all of these models show that the particular multipliers depend crucially on the other aspects listed above. The persistence of a path of government spending and how it is financed is crucial, as are many other characteristics such as the exchange rate regime. These results must be kept in mind when one is trying to forecast the effects of a particular policy.

III. A Summary of the Leading Empirical Approaches

A variety of empirical approaches have been used to estimate the effects of fiscal policies. I group these approaches into four broad categories: (1) aggregate country-level time series or panel estimates; (2) estimated or calibrated dynamic stochastic general equilibrium (DSGE) models; (3) subnational geographic cross-section or panel estimates; and (4) individual industry, firm or household estimates of key parameters (such as marginal propensities to consume, MPCs). Table 1 shows these categories along with the strengths and weaknesses associated with estimation of each type of data.

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 DSGE model approach has the additional advantage that since it estimates structural rather than reduced form parameters, one can use the estimates to perform counterfactuals. These two approaches share some of the same weaknesses, though. 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 VARS (SVARs) and natural experiment methods combined with narrative methods that use historical documents to create new data series that are argued to be uncorrelated with the existing economic conditions. Too often, though, the variations that are exogenous yield instruments that are not very relevant (i.e. they have low correlation with the fiscal variable) and the variations that are relevant are not always exogenous or are anticipated.

Although many papers using estimated DSGE 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 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. Unfortunately, estimated DSGE models are not immune to weak identification, as pointed out by Fabio Canova and Luca Sala (2009).

The third and fourth approaches, estimates across subnational units, such as states or provinces, and estimates of individual firm or household parameters, are more akin to applied microeconomics approaches. These approaches typically use the natural experiment approach or Bartik-style instruments that are widely used in applied microeconomics contexts. Similar to the microeconomic context, these analyses at lower levels of aggregation tend to have much stronger identification; the necessary identifying assumptions are typically more plausible and the instruments are relevant. Moreover, these approaches can be used on a variety of datasets. Both approaches, however, suffer from the same key weakness: the estimates produced are not macroeconomic estimates. Why? First, any cross-sectional estimating equation includes a constant term, which means that the macroeconomic effects have been netted out. This means that the parameters estimated are only *relative* effects, i.e., it answers 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? Second, some of the estimation equations, particularly at the household level, have an implicit control group that makes it very hard to translate the parameters to the aggregate level. I will discuss this weakness in more detail later in the article. In order to infer the implied national-level effects from the microeconomic estimates, researchers must turn to macroeconomic DSGE models, which, as discussed above,

incorporate their own additional identifying assumptions. Thus, there is no “applied micro free lunch” for macroeconomists: identification of macroeconomic effects must always depend on macroeconomic identification assumptions.

To summarize, there are a number of empirical approaches to estimating the effects of fiscal policy. Each has its strengths, but also its share of weaknesses. Moreover, some of the estimates are more appropriate for forecasting the effects of particular policies under certain conditions than others. For these reasons, it is useful to consider the estimates across the different approaches. One gains more confidence in the estimates if very different approaches and different data sets give similar estimates for like policies.

IV. 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 there were few conferences that brought the researchers together. As a result, different researchers chose different estimation methods and different ways of calculating multipliers from those estimates and there was no agreed upon 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 some of the lessons learned from this literature.

A. Calculating Multipliers in a Dynamic Environment

A number of papers discuss the “wide range” of multiplier estimates. What many do not realize is that differences in reported multiplier estimates are often due not so much to differences in identification methods or samples, but methods used to construct multiplier 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, which need to be taken into account in the computation of multipliers. Serious computation of multipliers was typically not a focus of the research before the financial crisis. Indeed, in my work with Shapiro on the effects of government spending, we did not even mention the word “multiplier” (Ramey and Shapiro (1998)). 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 multipliers. In particular, 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. While their method is a useful description of impulse responses, it does not yield multipliers of use for policymakers for the following reason. Estimated shocks to government spending typically lead to hump-shaped responses of government spending involving elevated levels of government spending for several years. Since the government spending path generates the output responses, the dynamic path of government spending *must* be incorporated when calculating multipliers in order to account for the impact on government budgets. Fortunately, Mountford and Uhlig (2009) moved the literature forward by introducing the policy-relevant multipliers calculated as

the present discounted value of the *integral* of the output response divided by the present discounted value of the *integral* of the government spending response to the shock. In most applications, the undiscounted integral gives 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 generally depends on how much government spending rises after the initial impact. I will offer an illustration of a typical situation in which it makes a big difference. To this end, I estimate a Blanchard-Perotti (2002) type structural vector autoregression (SVAR) model over the period 1939:1 – 2015:4 using Ramey and Zubairy's (2018) update of Ramey's (2011) quarterly data set. The SVAR model contains five endogenous variables: log real total government spending per capita, log real GDP per capita, log real federal tax receipts per capita, the 3-month Treasury bill interest rate, and inflation measured as the log change in the GDP deflator. Four lags are included, as well as a quadratic trend. The shock to government spending is identified using Blanchard and Perotti's (2002) method, which assumes that any part of government spending not forecasted by any of the other variables included in the model is an exogenous shock to government spending.

Panel A of Figure 1 shows the estimated impulse responses of the log of the government spending variable and the log of the GDP variable. The shaded area shows the 95-percent confidence bands. As the 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. Note that 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

off of the graphs. The standard practice has been to use an ad hoc conversion factor. Researchers convert the elasticities $d\ln(Y)/d\ln(G)$ to multipliers dY/dG by multiplying the elasticities by the average of the ratio of GDP to total government spending, Y/G , over the sample. For this sample, that ratio is 4.78. I will have more to say about the use of conversion factors shortly.

Panel B of Figure 1 shows the multipliers calculated three different ways. The highest multiplier is given by Blanchard-Perotti's method, which I will call a *quasi-multiplier*. 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 Y/G). Their method, shown by the dashed line, essentially traces out a renormalized version of the impulse response of output. In this case it yields multipliers that peak at 2.2 at quarter 6. The Mountford and Uhlig (2009) *cumulative multiplier*, shown by the solid line, uses the ratio of the integral of impulse response of output to 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 present discounted value version, using the average 3-month Treasury bill rate over the sample, is almost identical and is therefore not shown.

A second issue, however, is the practice of converting elasticities with the ad hoc conversion factor, the average of Y/G over the sample. Owyang, Ramey, and Zubairy (2013) pointed out biases that could arise from this practice. In their historical sample, Y/G varied significantly, from 2 to 24, with a mean of 8. They determined that the way to avoid the problem was to transform the government spending and output variables in a way that puts them in the same units. Owyang, Ramey and Zubairy (2013) used the transformations employed by Hall (2009) and Barro and Redlick (2011) that overcome the problem; the disadvantage, however, is

that they cannot be used in an SVAR framework.¹ Fortunately, Gordon and Krenn's (2010) transformation, which divides both government spending and GDP by a measure of potential GDP, can be used in an SVAR.

To illustrate the effect of moving from a specification in logarithms that requires the ad hoc conversion to one that does not, I re-estimated the model using Gordon and Krenn's transformation for government spending and GDP, employing Ramey and Zubairy's (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. The solid line with diamonds in Panel B of Figure 1 shows the cumulative multiplier estimates based on the impulse responses from this alternative specification. These multipliers, which do not rely on the ad hoc conversion factor, tend to be 5 to 35 percent smaller than those that rely on the ad hoc conversion factor, depending on the horizon. The horizon with the highest multiplier is horizon 0, with a multiplier of only 0.78.

Thus, deceptively small changes in the way the multipliers are calculated can make a very big difference. Using Blanchard and Perotti's quasi multiplier for government spending on estimates requiring an ad hoc conversion factor produces a multiplier as high as 2.2. That multiplier falls to 0.78 when the fully dynamic Mountford and Uhlig cumulative multiplier is used on estimates based on data using the Gordon-Krenn transformation. These simple changes have important consequences for the decisions of policymakers.²

¹ Hall and Barro and Redlick regressed $[Y(t+h) - Y(t-1)]/Y(t-1)$ on $[G(t+h) - G(t-1)]/Y(t-1)$, where Y is GDP, G is government spending, and h is the horizon.

² In Ramey (2011a), I reported a cumulative five-year multiplier of 1.2 based on the data in logs and the ad hoc conversion factor. A multiplier above one implies that private spending, $Y - G$, rises. Yet in Ramey (2013), I used the same log model, but substituted private spending for GDP and was surprised to find a statistically significant decline in log private spending. At the time, I did not realize that my use of the standard ad hoc conversion factor was biasing my multiplier estimates upward.

The same principles apply to the computation of tax multipliers, with one additional complication. There is little feedback from GDP to government spending, but strong feedback from GDP to tax revenue. 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 the multiplier for corporate tax cuts because their large positive effect 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, virtually all 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. Thus, the reported multipliers are an underestimate of the multipliers relative to actual tax revenue changes.

B. 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 involving military spending. For example, I showed that 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. 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 unanticipated tax cuts had expansionary effects on output, but news about future tax cuts had

contractionary effects on output in the short-run. Their estimates implied that the depth of the 1981-82 recession and the slow recovery from the 2001 recession owed much to the phase-in of tax cuts. 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.

C. Improvements in Fiscal Shock Identification

Any analysis that seeks to measure a causal effect must confront identification issues because of classic simultaneous equations bias. Initially, the standard macro approach used was a structural vector autoregression (SVAR). In most applications, it was assumed that the exogenous part of government spending was simply the part of government spending not forecasted by lagged values of spending, GDP, and taxes. To identify exogenous movements in taxes, Blanchard and Perotti (2002) brought in external estimates of the tax revenue elasticity to income, which allowed identification of the component of taxes that was not induced by movements in GDP. Several papers highlighted potential problems with these widely-used method. The first problem, discussed above, was the realization that failing to account for fiscal foresight could lead to seriously biased estimates. The second was the demonstration that the tax multiplier estimates were very sensitive to the value of the external tax elasticity used (e.g. 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 has moved from being the first resort to the last resort identification method in the fiscal literature.

In fact, long before SVAR methods were used, Hall (1980) and Barro (1981) used natural experiment methods to assess the effects of exogenous increases in government spending. Arguing that defense spending in the U.S. is 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 U.S. historical data. Ramey and Shapiro (1998), and numerous follow-up papers, built on that natural experiment insight about wars, but refined the measures to account for the importance of news. To create series of news about future government spending induced by current military events, they turned to narrative methods using business and other periodicals. However, this method that works quite well for U.S. data does not export well to other countries since most countries either do not have the substantial fluctuations in defense spending experienced by the U.S. 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 uses 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. Johnson, Parker, and Souleles (2006) and Parker, Souleles, Johnson, and McClelland's (2013) analysis 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. 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 were exogenous to the state of the economy. For the post-WWII U.S., they read legislative

records to identify tax changes that were due either to inherited deficits or 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 (e.g. Guajardo, Leigh, and Pescatori (2014)). Mertens and Ravn (2012) improved their measure by splitting their series into anticipated and unanticipated tax changes for the effects of fiscal foresight could be addressed. Alesina, Favero, and Giavazzi (forthcoming) has 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. While early alarms about weak instruments were raised for macro studies by Nelson and Startz (1990) and for microeconomic studies by Bound, Baker, and Jaeger (1995), macroeconomists began to pay attention to the issue only in the last five to ten years. The SVAR methodology hid the fact that ultimately the estimation of multipliers was actually an instrumental variables estimation, in which cumulative GDP up to horizon h was regressed on cumulative government spending, using an SVAR shock or a narrative variable as an instrument. In fact, Ramey (2016) and Ramey and Zubairy (2018) explicitly estimate multipliers this way. 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, my military news variable 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 shock, by its nature, is a strong instrument, particularly at short horizons.

In sum, research on the effects of fiscal policy has made significant strides in methodology. The literature now exploits many new data sets. It has imported some of the innovations from the applied microeconomics literature, but 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.

IV. A Summary of Estimates of Spending and Tax Multipliers

This section summarizes fiscal multiplier estimates 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, and then move on to the effects of tax changes and transfer payments. I then discuss estimates of the effects of the ARRA and the fiscal consolidations in Europe.

A. Government Spending Multipliers based on Aggregate Data

I begin by summarizing estimates of government spending multipliers based on models that do not distinguish by the state of the economy. I discuss mostly multipliers that are based on present value cumulative or cumulative methods, as discussed in the last section, and that do not have known weak instrument problems.

Table 2 shows the estimates, grouped by method and country to some extent. Panel A shows estimates for the U.S. based on a variety of time series implementations. The first two lines show the cumulative multipliers based on my estimation of the SVAR discussed in the last section. In particular, this is a five-variable SVAR that uses the Gordon-Krenn transformation of the fiscal and GDP variables. The first row shows the multipliers using the Blanchard and

Perotti identification scheme. As shown in Figure 1B of the previous section, the multipliers range from 0.6 to 0.8 over a 20 quarter horizon for the estimation sample from 1939q1 – 2015q4. The estimates are similar for the post-WWII sample, though the underlying estimates of the impulse responses are less precise. Interestingly, while the results suggest a significant increase in taxes for the sample including WWII, thus complicating the interpretation of the multiplier for that period, the results show no change in taxes for the post-WWII sample. Thus, these estimates presumably give multipliers for government spending increases that were on average deficit-financed.

The second row uses the same set of variables, but employs the Ramey-Zubairy (2018) military news variable as the shock. This variable is added to the system and ordered first. For the sample starting in 1939q1, the multiplier estimates range from 0.7 to 0.8.³ For the sample starting in 1947q1, the estimates become less precise and the range expands from 0.4 to 1. A comparison of the first row and second row illustrates another recent finding from the fiscal literature: contrary to earlier received wisdom, the Blanchard and Perotti identification scheme does not produce larger multipliers. As explained in the previous section, the earlier higher estimates of multipliers using their scheme were due not to their identification method but to their idiosyncratic method for calculating multipliers. As the rest of the rows of Panel A show, these results are surprisingly similar to most of the other results in the literature.

Table 2, Panel B shows some of the leading estimates based on estimated New Keynesian DSGE models. Cogan et al. (2010) estimate the Smets-Wouters model, including an extension with rule-of-thumb consumers, and simulate the effect of the Obama stimulus program. Their multiplier estimates range from 0.6 to 0.7. Coenen et al. (2012) compare estimates across large-

³ This range excludes estimates from the first few horizons. Because news about future government spending causes GDP to jump immediately but government spending to rise with a delay, the short horizon multiplier estimates are not well-defined because the denominator is near 0.

scale New Keynesian DSGE models used by central banks and international policy organizations for a two-year temporary increase in government consumption. For the case of no monetary accommodation, they find cumulative multipliers between 0.7 and 1. Zubairy (2014) estimates a rich DSGE model and finds spending multipliers that range from 0.7 to 1.1. Leeper, Traum, and Walker (2017) estimate a DSGE model estimates multipliers varying between 0.7 and 1.3 for the regime when monetary policy is active. They show that the higher multipliers depend crucially on an extremely high rate of habit formation (their baseline estimate of the key parameter is 0.99 and lowering it to 0.8 significantly reduces the multiplier). Overall, these estimates are quite similar to those from the SVAR, natural experiment, and narrative literature, though the Leeper et al. multiplier are notable in that the upper part of the range is above one.

On balance, the estimated multipliers are not very different across the two leading methods for identifying government spending shocks nor for estimated DSGE New Keynesian models.⁴ With the exception of the Leeper, et al. multiplier, most of the estimates tend to indicate government spending multipliers that are less than or equal to one for the U.S.

Panel C of Table 2 shows government spending multipliers estimated for a number of other countries. Here again, both the Blanchard and Perotti identification and the narrative identification tend to give similar results, typically 0.3 to 0.7 for spending. For example, Iltzetzki, Mendoza, Vegh (2010) used the Blanchard and Perotti identification in an SVAR on a panel of 44 countries and found multipliers for developed countries between 0.3 to 0.7. The various narrative approaches to studying fiscal consolidations used by Leigh et al. (2010),

⁴ Fisher and Peters (2010) and Ben Zeev and Pappa (2017) find multipliers between 1.5 and 2 for their innovative measures of defense spending news. Fisher and Peters construct a measure of news using excess returns on defense company stocks. Ben Zeev and Pappa construct a measure of news by imposing both short- and medium-horizon restrictions. Both of these measures are very promising, but as discussed in Ramey (2016), they both seem to have weak instrument problems for government spending.

Guajardo, Leigh and Pescatori (2014) and Alesina, Favero, and Giavazzi (2018, forthcoming) all produce multipliers around 0.5 or below when they focus on spending-based consolidations.

Not shown in the table are numerous multiplier estimates based on key features of a country. For example, Iltetzki, Mendoza, Vegh (2010) estimate how multiplier change across various important feature, such as fixed or flexible exchange rates. They find multipliers that vary between 0.1 on impact to 1.4 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. Distinguishing by debt burden, they 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). Thus, the evidence they present suggests that the range of multipliers is much wider when one begins to distinguish by key country characteristic.

The results shown in Table 2 are for total government spending or government consumption. There is surprisingly little aggregate evidence on multipliers for public investment. Even in a neoclassical model, such as Baxter and Kings's (1992) model, the multiplier effects of government spending can be quite large, particularly in the long-run. The returns to infrastructure investment gained attention with a series of papers of David Aschauer that documented a very close relationship between the trend in TFP growth and infrastructure investment in the post-WWII U.S. (e.g. Aschauer (1989)). Pereira and Flores de Frutos (1999) used SVAR analysis on U.S. time series data and found that a \$1 increase in spending on public capital resulted in a permanent rise in private output of 65 cents. The implied cumulative multiplier for this permanent effect is very large. Looking at other countries, Iltetzki, Mendoza, Vegh (2010) found multipliers for public investment that ranged between 0.4 in the short-run to

1.6 in the long-run. Much more attention should be devoted to the estimation of the short- and long-run multipliers on infrastructure investment.

A rough consensus has developed that government spending multipliers are probably below unity on average. However, the key question that has emerged is whether multipliers are higher during bad economic times. The key states studied by recent papers are recessions, periods of excess slack (typically measured by unemployment rates), and constraints on the monetary policy accommodation, such as at the zero lower bound.

Panel A of Table 3 shows the estimates of government spending multipliers conditional on recession or slack. The pioneering estimates were by Auerbach and Gorodnichenko (2012), who used the Blanchard and Perotti (2002) identification method on U.S. data. They extended the standard SVAR by allowing the parameters estimates to vary according to their definition of recession. In the baseline model, they reported a multiplier of 2.2 in recessions and -0.3 in expansions.

Subsequent research has found these results to very fragile. For example, Ramey and Zubairy (2018, associated online appendix) showed that application of the state-dependent Jordà (2005) local projections method, pioneered by Auerbach and Gorodnichenko (2013) in their subsequent work using OECD data, overturned the previous Auerbach and Gorodnichenko results. Ramey and Zubairy explored further and determined that the source of the differences was not the method used to estimate the reduced form parameters, but rather Auerbach and Gorodnichenko's (2012) use of the data-inconsistent assumption that recessions last at least five years; in their data set, the average recession lasted three quarters and no recession lasted more than a couple of years. Separately, Alloza (2017) showed that their findings of higher multipliers during recessions was due to their use of future information to define the current state. In

particular, their definitions of the recession and expansion states depended on a *centered* 7-quarter moving average of GDP growth, meaning that future GDP growth was used to define the current state. It makes no sense to include future GDP growth in the definition of the current state. Alloza found that simply reducing the 7-quarter moving average to a 5-quarter moving average reversed their results, so that multipliers during recessions were actually estimated to be negative. These negative multipliers during recessions line up with Alloza's (2017) findings using the NBER definition of recessions.

In their follow-up work on OECD data, Auerbach and Gorodnichenko (2013) continued to find higher multipliers during recessions, despite using the more robust local projection method. However, the only multipliers they reported were ratios of the average output response or the peak output response to the initial response of government spending, which as I discussed above are not the correct dynamic multipliers. In addition, they specified their model in logs and hence must use an ad hoc conversion factor. As discussed above, those two methods are known to bias multipliers significantly upward.

Ramey and Zubairy (2018) conducted a detailed study of multipliers during periods of slack using historical U.S. data. They estimated effects using the two leading identification methods and explored hosts of robustness checks. They found no evidence of multipliers above unity during high unemployment periods or during recessions.

In sum, the estimates of higher multipliers during recession are exceedingly fragile; simple improvements in the specification result in multipliers that are less than unity. In contrast, the low estimates are robust to plausible permutations. Thus, at this point there is no clear evidence of higher multipliers during recessions.

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 WWII in the U.S.) Earlier work by Ramey (2011a) and Crafts and Mills (2013) could find no evidence of higher multipliers when they explored subsamples of their historical data for the U.S. and U.K. However, Ramey and Zubairy (2018) estimated a state dependent model (using the same techniques they used for exploring the effects of slack states) but for zero lower bound states and found some evidence for higher multipliers in their sample from 1889 through 2015. While the entire sample results did not show higher multipliers, the results estimated on a sample that excluded the rationing periods during WWII indicated higher multipliers, around 1.4 (standard error of 0.15) at two years. The estimate was even higher, 1.6, when both the military news variable and the Blanchard and Perotti shock were used as instruments, but the standard error of the estimate was higher (0.5). Miyamoto, Nguyen, and Sergeyev (2018) used the same estimation method, but somewhat different identification, on Japanese data and found significantly higher multipliers during Japan's extended zero lower bound period. They found cumulative multipliers of 1.4 on impact and 2.7 at four quarters, though neither was statistically different from one using conventional significance levels. Those contrast with non-ZLB multipliers ranging from -0.6 to 0.6.

These higher multipliers during ZLB periods are qualitatively consistent with the predictions of New Keynesian DSGE models at the zero lower bound. Panel B of Table 3 shows some of the multipliers that emerge from these models. For example, Coenen et al. (2012) find a number of estimates that range between 2 and 3 when the period of monetary accommodation is sufficiently long.

To summarize, the few papers that have estimated multipliers during zero lower bound periods provide some evidence of multipliers higher than unity, meaning the government spending stimulates private spending. Moreover, the estimated large-scale DSGE models produce similar estimates.

B. Tax and Transfer Multipliers based on Aggregate Data

I now turn to the leading estimates of tax and transfer multipliers at the aggregate level. Table 4 shows the time series estimate in U.S. data, estimated DSGE models, and other country data. Note that in contrast to the government spending multipliers, I report the peak cumulative multiplier (in absolute value), typically within the first three years. In many of the cases, the multiplier starts out low on impact but then builds so the range across horizons is wide.

It is also important to note the following feature, discussed earlier with respect to the computation of tax multipliers. Virtually all of the literature computes multipliers with a static forecast of the tax revenue effect in the denominator; it is not based on the actual response of tax revenues. To the extent that a tax cut leads to higher GDP, which raises tax revenue, this method will *underestimate* the tax multiplier for budgetary purposes. I am aware of only three instances in which the actual response of tax revenue is used in the multiplier calculation. First, Mountford and Uhlig (2009) use the actual response, which might account for the fact that their estimate of the tax multiplier of -5 is at the top of the range (in absolute value). Second, Mertens and Ravn (2013) calculate the multiplier this way in their paper comparing the effects of personal income tax changes relative to corporate tax changes. They find a multiplier of -2.5 by the third quarter for personal income tax rate changes. They cannot calculate a multiplier for corporate income tax rate changes *because the net effect of the change on actual tax revenue is*

zero. That is, the multiplier looks infinite because they appear to be on the top of the Laffer curve. Third, Alesina, Azzalini, Favero, Giavazzi, and Miano (2018) calculate tax multipliers both ways. As Table 4 shows, the multipliers based on the actual response of tax revenue are substantially larger.

Table 4 shows that the other multiplier estimates based on narrative approaches, which unless noted are computed relative to the forecast of tax revenue without dynamic feedbacks, are generally between -2 and -3. These are much higher (in absolute value) than the tax multipliers reported by Blanchard and Perotti (2002). As discussed above, their 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.

Eskandari (2015) and Demirel (2016) studied whether the Romer and Romer (2010) narrative tax shocks had different effects if they hit during bad or good times. Both authors use local projections and consider measures of slack based on the unemployment rate. Both find that the effects of tax shocks 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. Thus, the few papers that have studied state dependence for taxes find greater effects during expansions than during recessions.

There is very little work on the aggregate effects of transfers. Oh and Reis (2012) called attention to the fact that transfers were an important part of the stimulus packages and discussed possible effects in a theoretical model. Recently, 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 equal to one percent of consumption led to a roughly equal rise in aggregate consumption in the short-run, which then dissipated after five months. Temporary increases in benefits had no significant effect on aggregate consumption. They also compared the responses to their permanent transfer payment increases versus their narrative series of tax changes. They found that tax changes had far greater effects than transfer payments.

In sum, most empirical estimates of tax multipliers indicate that they are very large, at least -2 to -3. Thus, it appears that contrary to the simple Keynesian model, tax multipliers are greater than spending multipliers. An explanation is offered by Mertens and Olea-Montiel (2018), who construct a new narrative series of exogenous changes in average marginal tax rates and study their effects on macroeconomic variables. They find that tax changes have their effects not through an after-tax income channel, but rather appear to be through the channel emphasized in theoretical models of incentive effects and forward-looking behavior. Thus, it is the supply-side effects, and not the Keynesian demand-side effects, that seem to matter.

C. Subnational Multipliers and Micro-Level Estimates

As discussed earlier in this paper, one of the important innovations in the fiscal literature 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 U.S. 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 or cross-section multipliers from individual studies, so I refer the reader to his tables. As he notes, and as I also noted in my earlier literature review (Ramey (2011b)), the bulk 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. I will discuss the ARRA cross-state estimates in more detail in the next section.

The estimates of the household level responses to temporary tax rebates and transfers have also been interpreted as implying much higher national-level multipliers. For example, the studies by Shapiro and Slemrod (2003, 2009), Johnson, Parker and Souleles (2006), Sahm, Shapiro and Slemrod (2012), Parker, Souleles, Johnson, McClelland (2013), and Broda, and Parker (2014) of the effects of the temporary tax rebates in 2001 and 2008 are exemplars of the use of natural experiments, along with the creation of new data, to obtain estimates of some of key micro parameters of interest to macroeconomists. In particular, they exploited the fact that the 2001 and 2008 tax rebate checks were announced several months in advance and then distributed over a several month period, with timing randomized by the last two digits of Social Security numbers. Moreover, in some of the best examples of entrepreneurial data collection, they added special questions to leading household surveys in order to match the household behavior to the receipt of the rebate. In the case of the Shapiro and Slemrod work, they asked respondents in the University of Michigan Survey of Consumers qualitative questions about what they did with their rebate check. In the case of the work by Parker and co-authors, they added questions about when the household received the rebate, how much it was, and in what form it arrived to the Consumer Expenditure survey or Nielsen survey. Shapiro and co-authors tended to find smaller marginal propensities to spend, around 0.3, but Parker and co-authors found some very high marginal propensities to spend. For example, Parker, Souleles, Johnson, and McClelland (2013) found a marginal propensity to spend out the temporary tax rebate of 50 to 90 percent within three months of receiving the 2008 tax rebate. Estimates from these studies have

been used to calibrate recent heterogeneous agent models, such as by Kaplan and Violante (2014) and to argue that temporary tax rebates can have large multipliers.

I now question whether researchers have been too quick to export the parameters estimated on household data to macroeconomic models. As an illustration, consider the Parker, Souleles, Johnson, and McClelland (2013) paper, which finds very high marginal propensities to spend out of the 2008 tax rebate. As they note, much is driven by expenditures on motor vehicles. I present two pieces of evidence that their estimates could not possibly apply to the aggregate level.

The first piece of evidence concerns expenditures on new motor vehicles and is based on the numbers in Table 14 of Sahm, Shapiro and Slemrod (2012). In this calculation, Sahm et al. applied the Parker et al. estimate of the marginal propensity to spend on new vehicles to the actual pattern of 2008 rebate disbursements in order to compute a partial equilibrium counterfactual of what aggregate new vehicle expenditures would have been without the rebate. The only changes I made to their calculation were to put them on a monthly basis and to use updated NIPA data.

Figure 2A shows actual spending on new vehicles and the counterfactual based on Parker et al.'s estimates. The counterfactual path implies that had there been no rebate, expenditures on new vehicles would have fallen from around \$17 billion per month in March 2008 to a trough of only \$2.6 billion in June 2008, only to rise back up to \$14 billion and then fall to \$12 billion in the wake of the collapse of Lehman Brothers. Even taking into account the high oil prices in the summer of 2008, this counterfactual implication is extremely implausible.

The second piece of evidence, which concerns the overall marginal propensity to consume estimate by Parker et al. (2013), extends arguments made earlier by Feldstein (2008,

2009), Taylor (2009), and Shapiro and Slemrod (2009). Figure 2B uses aggregate monthly NIPA data and rebate data to compare two rates. Both rates use aggregate personal disposable income, excluding the rebate, in the denominator. The “rebate rate” uses the aggregate rebate in the numerator; the “saving rate” uses aggregate personal saving in the numerator. In order to be able to compare the heights, I renormalized the actual saving rate to be zero in January by subtracting 3.3. I chose January for the normalization because that was the month before the tax rebate was passed. It should be understood, though, that this normalization is not innocuous. The actual saving rate was 0.5 percentage points higher in February and March and normalizing by those months would shift the line down more.

The solid green line shows the rebate as a percent of non-rebate disposable income. The rebate was very big relative to income. The dashed blue line shows that the actual saving rate jumped almost as much as the rebate rate, suggesting that most of the rebate was saved. A back-of-the-envelope calculation based on the cumulative difference between the two graphs during the rebate months suggests that only 0.13 of the rebate was spent.

The lower red dotted line is the saving rate implied by the lower bound of Parker et al.’s (2013) estimates of the marginal propensity to spend. That lower bound is 50 percent, so it implies a saving rate of 50 percent out of the rebate. If I used their estimated higher marginal propensities to spend of 70 or 90 percent, the implied saving rate line would be even lower. This figure shows that there is a substantial gap in the behavior of the actual saving rate relative to the estimates implied by Parker et al. In work-in-progress, I show how their estimating equation can over-estimate the aggregate marginal propensity to consume if individual consumers bunch their consumption because of liquidity constraints that bind only at high frequency.

In sum, while the estimates based on detailed micro studies have the potential to shed significant light on the macroeconomic effects of various fiscal policies, I believe that much more study is required before they are used to reach macroeconomic conclusions.

V. Multipliers in the Wake of the Financial Crisis

A number of researchers and commentators have argued that the effects of the stimulus program in the U.S., the American Recovery and Reinvestment Act (ARRA), and the subsequent fiscal consolidations in European countries were much larger than indicated by multipliers during average times. They argue that the high unemployment rates and lower bound on interest rates combined to raise the multipliers.

As shown in the last two sections, there is not good 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 U.S. 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 from 2009 through 2011.

Table 5 shows some of the evidence for the fiscal consolidations and the ARRA. Consider first the fiscal consolidations.

Blanchard and Leigh (2013, 2014) offer some evidence that they interpret as multipliers being higher than one for the European fiscal consolidations. They show in a cross-section of countries that the subsequent forecast error for GDP growth for 2010 and 2011 was negatively correlated with the initial forecast for the fiscal consolidation. The forecasts used were the IMF

forecasts based on their models. They conclude that the models used to generate the forecasts incorporated fiscal multipliers that were too small and suggest that actual multipliers were substantially above one.

House, Proebsting, and Tesar (2017) expanded on their analysis by studying the relationship between forecast errors for government spending and forecast errors for output in a set of European countries. They construct their forecast errors differently from Blanchard and Leigh. In particular, they estimate a forecasting equation through 2005 based on trend movements and convergence and use that to forecast the variables from 2010 to 2014. They regress the forecast error in GDP on the government purchases forecast error and interpret the coefficient as the government spending multiplier. Depending on the specification, they find a multiplier as high as 2.5.

These forecast error results are very intriguing. The implied multipliers are not necessarily at odds with the previous literature I surveyed, since government spending multipliers did appear to be higher at the zero lower bound and tax multipliers were always high. However, it is important to keep in mind that there is an alternative interpretation of the results. The timing and size of fiscal consolidations are not randomly distributed. Suppose that policymakers in a country have better information about the future path of the economy than the IMF does or than is forecasted by the simple model used by House et al. (2017). If those policymakers foresee a steeper decline in trend growth in the future, they might be more likely to implement bigger fiscal consolidations. Such actions would lead to exactly the correlations highlighted by Blanchard and Leigh (2013, 2014) and House, Proebsting, and Tesar (2017).

Alesina et al. (2018) and Alesina, Favero, and Giavazzi(forthcoming) use their narrative data set 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 and discuss at length why they disagree with Blanchard and Leigh's conclusion.

In sum, there is still a lively, ongoing debate on whether multipliers were greater for the fiscal consolidations in Europe.

The effects of the earlier ARRA stimulus in the U.S. are also still being debated. Table 5 shows some estimates of the effects of the ARRA stimulus program in the U.S. This program was a mix of spending and transfers to states and individuals. As the table shows, none of the New Keynesian DSGE models finds multipliers above unity for this program. However, as synthesized by Chodorow-Reich (forthcoming), the cross-state natural experiment estimates indicate higher multipliers from 1.7 to 2 for gross state product and \$50K per job-year created.

The fiscal literature realized early on that subnational multipliers are not the same as national multipliers for a host of reasons. However, building on Farhi and Werning's (2016) analysis, Chodorow-Reich (forthcoming) argues that the subnational multipliers are *lower bounds* on the national multipliers during a liquidity trap. Thus, he argues that the ARRA multiplier was at least 2.

Why, then, are the answers from the subnational estimates so much greater than the aggregate times series and estimated New Keynesian DSGE estimates? I will now argue that the estimates and arguments presented by Chodorow-Reich are overestimates for the national-level multipliers. Chodorow-Reich presents an extremely valuable synthesis of the leading estimates of the effects of the ARRA on job creation across U.S. states. This literature emphasizes employment effects because the employment data have less measurement error than gross state product. He considers several of the leading instruments from the literature as well as their combination. The first-stage F-statistics are very high and suggest strong instruments. The

results across the separate instruments are very similar and he cannot reject the over-identifying restrictions. Thus, the estimates are based on strong applied microeconomic methods.

I will begin by highlighting how big his estimates are by conducting a counterfactual similar to the ones I conducted earlier with respect to the marginal propensity to spend estimates. In particular, I repeat the argument from Ramey (forthcoming) by using Chodorow-Reich's estimated impulse responses of employment to calculate a partial equilibrium counterfactual of what the unemployment rate would have been had there been no ARRA. I add the estimated induced employment to the actual number of unemployed to create the counterfactual unemployment rate.

Figure 3 shows the results. The actual unemployment rose from under 5 percent in 2007 to 10 percent by late 2009. The counterfactual implied by Chodorow-Reich's estimates imply that the unemployment rate would have risen to 15.5 percent had the ARRA not passed. This counterfactual is far above the estimates used by economists within the Obama administration.

Thus, we are left with a counterfactual, while not as implausible as the previous one involving the Parker et al. marginal propensity to consume estimates, that still gives one pause. I now offer some easy explanations for why Chodorow-Reich's synthesis estimates are not lower bounds on the national level estimates. These explanations are based on both empirics and theory.

On the empirical part, I now highlight that two adjustments must be made to the Chodorow-Reich's estimates before they can be used at the national level. The first issue is weighting. Chodorow-Reich's uses per capita values of each state as observations, meaning that he gives equal weight to North Dakota and California. If there is heterogeneity in the treatment effects, the unweighted results will not give the average results needed to apply the estimates to

the national level.⁵ To derive the average treatment effect relevant to the national level, I re-estimate Chodorow-Reich's model (using his replication files) but weight by initial (Dec. 2008) state population.

Table 6 shows three rows. All of them are estimates of the employment response. The first row, which is from Table 1, column 4 of his paper shows his preferred estimates, which use all three of the leading ARRA instruments. The estimates are for job-years created for each \$100K of the ARRA spending. The estimate of 2 implies that each \$100K of ARRA spending creates 2 job years.

The second row of Table 6 shows that weighting the estimates by state population leads the point estimate to fall to 1.15. The standard error is higher at 0.72. Thus, the point estimate falls by over 40 percent once state observations are weighted by their population weights in order to make them representative of national data.

The second adjustment that must be made to make the estimate relevant for the national-level multiplier is the measure of spending. Chodorow-Reich's measure of spending is federal ARRA spending. The estimates based on this measure answer the question of how many job-years are created per federal dollar transferred to the state. However, as Leduc and Wilson (2017) show, the ARRA spending stimulated state and local spending more than dollar for dollar. Thus, multipliers that use only the federal transfers to the states will overestimate the multiplier per total dollar spent. To obtain a multiplier based on total spending at the state level, we need to use instead the cumulative state and local spending. Table B3 of Chodorow-Reich's online appendix shows that each \$1 of ARRA spending raised state and local spending by \$1.22 and lowered taxes by \$0.11. Rather than calculating a back-of-the-envelope estimate by combining

⁵ 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 ARRA are much lower than unweighted estimates.

the various estimates, I instead obtain a direct estimate by reestimating his model, substituting the state and local cumulative spending for the ARRA spending. I use his combination of instruments and I weight by initial state population.

Row 3 of Table 6 shows the jobs multiplier based on total state and local spending. That estimate is 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.

These new results show that there is no longer a discrepancy between estimates based on national data or DSGE models and cross-state estimates. 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.

Finally, on the theoretical side, it should be recalled that the theoretical results of Farhi and Werning that the cross-state estimates are lower bounds on national multipliers only holds in certain cases, in particular when monetary policy is constrained by the zero lower bound. While interest rates were indeed at the zero lower bound during those years, Swanson and Williams (2014) present evidence that yields on 1-year and 2-year treasury bills were unconstrained through 2008 to 2010, "suggesting that monetary policy and fiscal policy were about as effective as usual during this period." (abstract). Thus, it is not even clear whether the circumstances prevailed for the national multiplier to be above the subnational multipliers during the time of the ARRA.

In sum, whether multipliers were higher than usual during the Great Recession and after the financial crisis is still being debated. I have offered arguments in the U.S. context that the multipliers on the ARRA were in the range of those typically estimated on aggregate data.

VI. Conclusions

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 multipliers has been reduced considerably. On average, government purchases multipliers are likely to be between 0.6 and 0.8 whereas tax rate change multipliers are likely to be between -2 and -3. However, there is still ongoing debate about specific contexts, such as the size of multipliers during the Great Recession and its aftermath and various exchange rate regimes.

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. A few tantalizing studies at the aggregate and subnational levels suggest that these multipliers can be very large in some contexts. Second, researchers need to be careful about their implementation decisions. As I have shown, seemingly small changes, such as how multipliers are actually calculated, can make a big difference. Finally, researchers should continue to innovate along the same lines they have pursued in the last ten years, exploiting new data sets, extending theoretical models, and improving estimation techniques. As part of this innovation, researchers need to think more about the link between micro estimates and aggregate effects.

References

- Acconcia, Antonio, Giancarlo Corsetti, and Saverio Simonelli. "Mafia and public spending: Evidence on the fiscal multiplier from a quasi-experiment." *American Economic Review* 104, no. 7 (2014): 2185-2209.
- Alesina, Alberto, Carlo Favero, and Francesco Giavazzi, *Austerity: When It Works and When It Does Not*, forthcoming Princeton University Press
- Alesina, Alberto, Gualtiero Azzalini, Carlo Favero, Francesco Giavazzi, and Armando Miano. "Is it the "How" or the "When" that Matters in Fiscal Adjustments?." *IMF Economic Review* 66, no. 1 (2018): 144-188.
- Alloza, Mario, "Is Fiscal Policy More Effective in Uncertain Times or During Recessions?" Centre for Macroeconomics Working Paper CFM-DP2016-31. Banco de España Working Paper No. 1730, 2017.
- Aschauer, David Alan. "Is public expenditure productive?." *Journal of monetary economics* 23, no. 2 (1989): 177-200.
- Auerbach, Alan and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4 (2): 1-27.
- Auerbach, Alan and Yuriy Gorodnichenko. 2013. "Fiscal Multipliers in Recession and Expansion" *Fiscal Policy After the Financial Crisis*, eds. Alberto Alesina and Francesco Giavazzi, University of Chicago Press.
- Barro, Robert J., "Output Effects of Government Purchases," *The Journal of Political Economy*, 89 (1981), 1086-1121.
- Barro, Robert J., and Charles J. Redlick. "Macroeconomic effects from government purchases and taxes." *The Quarterly Journal of Economics* 126, no. 1 (2011): 51-102.
- Baxter, Marianne, and Robert G. King, "Fiscal Policy in General Equilibrium," *American Economic Review*, 83, no. 3 (1993), 315–334.
- Ben Zeev, Nadav, and Evi Pappa. "Chronicle of a war foretold: The macroeconomic effects of anticipated defence spending shocks." *The Economic Journal* 127, no. 603 (2017): 1568-1597.
- Blanchard, Olivier J., and Daniel Leigh. "Growth forecast errors and fiscal multipliers." *American Economic Review* 103, no. 3 (2013): 117-20.
- Blanchard, Olivier J., and Daniel Leigh. "Learning about fiscal multipliers from growth forecast errors." *IMF Economic Review* 62, no. 2 (2014): 179-212.

- Blanchard, Olivier, and Roberto Perotti, "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output," *Quarterly Journal of Economics*, 117, No. 4 (2002), 1329-1368.
- Broda, Christian, and Jonathan A. Parker. "The economic stimulus payments of 2008 and the aggregate demand for consumption." *Journal of Monetary Economics* 68 (2014): S20-S36.
- Bound, John, David A. Jaeger, and Regina M. Baker. "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, no. 430 (1995): 443-450.
- Caldara, Dario, and Christophe Kamps. "The analytics of SVARs: a unified framework to measure fiscal multipliers." *The Review of Economic Studies* 84, no. 3 (2017): 1015-1040.
- Canova, Fabio, and Luca Sala. "Back to square one: Identification issues in DSGE models." *Journal of Monetary Economics* 56, no. 4 (2009): 431-449.
- Chodorow-Reich, Gabriel. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" forthcoming *American Economic Journal: Economic Policy*.
- Cloyne, James, "Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom," *The American Economic Review*, Volume 103, Number 4, June 2013, pp. 1507-1528(22).
- Coenen, Günter, Christopher J. Erceg, Charles Freedman, Davide Furceri, Michael Kumhof, René Lalonde, Douglas Laxton et al. "Effects of fiscal stimulus in structural models." *American Economic Journal: Macroeconomics* 4, no. 1 (2012): 22-68.
- Cogan, John F., Tobias Cwik, John B. Taylor, and Volker Wieland. "New Keynesian versus old Keynesian government spending multipliers." *Journal of Economic dynamics and control* 34, no. 3 (2010): 281-295.
- Corsetti, Giancarlo, Andre Meier, and Gernot J. Müller. "What determines government spending multipliers?." *Economic Policy* 27, no. 72 (2012): 521-565.
- Crafts, Nicholas, and Terence C. Mills. "Rearmament to the rescue? New estimates of the impact of "Keynesian" policies in 1930s' Britain." *The Journal of Economic History* 73, no. 4 (2013): 1077-1104.
- Demirel, Ufuk Devrim, "The Short-Term Effects of Tax Changes—Evidence for State Dependence", 2016, Congressional Budget Office Working Paper 2016-03.

- Dupor, Bill, and M. Saif Mehkari. "The 2009 Recovery Act: Stimulus at the extensive and intensive labor margins." *European Economic Review* 85 (2016): 208-228.
- Eskandari, Ruhollah. "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*, 2015, University of Milan dissertation.
- Farhi, Emmanuel, and Iván Werning. "Fiscal Multipliers☆: Liquidity Traps and Currency Unions." In *Handbook of Macroeconomics*, vol. 2, pp. 2417-2492. Elsevier, 2016.
- Feldstein, Martin "The Tax Rebate Was a Flop. Obama's Stimulus Plan Won't Work Either." *The Wall Street Journal*, August 6, 2008.
- Feldstein, Martin. "Rethinking the role of fiscal policy." *American Economic Review* 99, no. 2 (2009): 556-59.
- Fisher, Jonas D.M., and Ryan Peters, "Using Stock Returns to Identify Government Spending Shocks," *The Economic Journal*, 120 (May 2010): 414-436.
- Gordon, Robert J., and Robert Krenn. "The end of the Great Depression: VAR insight on the roles of monetary and fiscal policy." *NBER Working paper* 16380 (2010): 183-231.
- Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. "Expansionary austerity? International evidence." *Journal of the European Economic Association* 12, no. 4 (2014): 949-968.
- Hall, Robert E., "Labor Supply and Aggregate Fluctuations," *Carnegie-Rochester Conference Series on Public Policy*, 12 (1980), 7-33.
- Hall, Robert E. "By How Much Does GDP Rise If the Government Buys More Output?" *Brookings Papers on Economic Activity* No. 2 (2009).
- Hayo, Bernd, and Matthias Uhl. "The macroeconomic effects of legislated tax changes in Germany." *Oxford Economic Papers* 66, no. 2 (2013): 397-418.
- House, Christopher L., and Matthew D. Shapiro. "Temporary investment tax incentives: Theory with evidence from bonus depreciation." *American Economic Review* 98, no. 3 (2008): 737-68.
- House, Christopher L., Christian Proebsting, and Linda L. Tesar. *Austerity in the Aftermath of the Great Recession*. No. w23147. National Bureau of Economic Research, 2017.
- Ilzetzki, Ethan, Enrique G. Mendoza, and Carlos A. Végh. "How big (small?) are fiscal multipliers?." *Journal of monetary economics* 60, no. 2 (2013): 239-254.

- Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. "Household expenditure and the income tax rebates of 2001." *American Economic Review* 96, no. 5 (2006): 1589-1610.
- Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review* 95 (1): 161-82.
- Kaplan, Greg, and Giovanni L. Violante. "A model of the consumption response to fiscal stimulus payments." *Econometrica* 82, no. 4 (2014): 1199-1239.
- Leduc, Sylvain and Daniel Wilson. "Roads to Prosperity or Bridges to Nowhere? Theory and Evidence on the Impact of Public Infrastructure Investment," (*NBER Macroeconomics Annual* 27, no. 1 (2012), pp. 89-142.
- Leduc, Sylvain, and Daniel Wilson. "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, no. 2 (2017): 253-92.
- Leeper, Eric M., Nora Traum, and Todd B. Walker. "Clearing up the fiscal multiplier morass." *American Economic Review* 107, no. 8 (2017): 2409-54.
- Leeper, Eric M., Todd B. Walker, and Shu-Chun Susan Yang. "Fiscal foresight and information flows." *Econometrica* 81, no. 3 (2013): 1115-1145.
- Leigh, Daniel, Pete Devries, Charles Freedman, Jaime Guajardo, Douglas Laxton, and Andrea Pescatori. "Will it hurt? Macroeconomic effects of fiscal consolidation." *World Economic Outlook* 93 (2010): 124.
- Mertens, Karel and Jose Montiel-Olea. "Marginal Tax Rates and Income: New Time Series Evidence". *Quarterly Journal of Economics* (November 2018).
- Mertens, Karel, and Morten O. Ravn. "Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks." *American Economic Journal: Economic Policy* 4, no. 2 (2012): 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., "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers," *Journal of Monetary Economics* 68 (2014): S1-S19.
- Miyamoto, Wataru and Nguyen, Thuy Lan, "Government Spending Multipliers Under the Zero Lower Bound," forthcoming *American Economic Journal: Macroeconomics*.

- Mountford, Andrew, and Harald Uhlig. "What are the effects of fiscal policy shocks?." *Journal of applied econometrics* 24, no. 6 (2009): 960-992.
- Nelson, Charles R., and Richard Startz. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator." *Econometrica: Journal of the Econometric Society* 58, no. 4 (1990): 967-976.
- Oh, Hyunseung, and Ricardo Reis. "Targeted transfers and the fiscal response to the great recession." *Journal of Monetary Economics* 59 (2012): S50-S64.
- Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy. "Are Government Spending Multipliers Greater during Periods of Slack? Evidence from Twentieth-Century Historical Data." *The American Economic Review* 103.3 (2013): 129-134.
- Parker, Jonathan A, Nicolas S. Souleles, David S. Johnson, Robert McClelland, "Consumer Spending and the Economic Stimulus Payments of 2008," *The American Economic Review*, 103 (October 2013): 2530-2553.
- Pereira, Alfredo M., and Rafael Flores De Frutos. "Public capital accumulation and private sector performance." *Journal of Urban economics* 46, no. 2 (1999): 300-322.
- Ramey, Valerie A. and Matthew Shapiro, "Costly Capital Reallocation and the Effects of Government Spending," *Carnegie Rochester Conference on Public Policy* Vol. 38, no. No. 1, (1998).
- Ramey, Valerie A. "Identifying government spending shocks: it's all in the timing." *The Quarterly Journal of Economics* 126, no. 1 (2011a): 1-50.
- Ramey, Valerie A. "Can government purchases stimulate the economy?" *Journal of Economic Literature* 49, no. 3 (2011b): 673-85.
- Ramey, Valerie A. "Government Spending and Private Activity." *Fiscal Policy after the Financial Crisis* (2013), edited by Alberto Alesian and Francesco Giavazzi, pp. 19–55. University of Chicago Press.
- Ramey, Valerie A. "Macroeconomic shocks and their propagation." In *Handbook of Macroeconomics*, vol. 2, pp. 71-162. Elsevier, 2016.
- Ramey, Valerie A., and Sarah Zubairy. "Government spending multipliers in good times and in bad: evidence from US historical data." *Journal of Political Economy* 126, no. 2 (2018): 850-901.
- Ramey, Valerie A., "Fiscal Policy: Tax and Spending Multipliers in the United States," forthcoming in *Rethinking Macroeconomic Policy IV*, Peterson Institute for International Economics (PIIE), MIT Press.

- Romer, Christina D., and David H. Romer, "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks," *American Economic Review*, 100 (June 2010): 763-801.
- Romer, Christina D., and David H. Romer. "Transfer payments and the macroeconomy: The effects of social security benefit increases, 1952-1991." *American Economic Journal: Macroeconomics* 8, no. 4 (2016): 1-42.
- Sahm, Claudia R., Matthew D. Shapiro, Joel Slemrod, "Check in the Mail or More in the Paycheck: Does the Effectiveness of Fiscal Stimulus Depend on How it is Delivered?" *American Economic Journal: Economic Policy*, 4 (August 2012):216-250.
- Shapiro, Matthew D., and Joel Slemrod. "Did the 2001 tax rebate stimulate spending? Evidence from taxpayer surveys." *Tax policy and the economy* 17 (2003): 83-109.
- Shapiro, Matthew D., and Joel Slemrod. "Did the 2008 tax rebates stimulate spending?." *American Economic Review* 99, no. 2 (2009): 374-79.
- Swanson, Eric T., and John C. Williams. "Measuring the effect of the zero lower bound on medium-and longer-term interest rates." *American Economic Review* 104, no. 10 (2014): 3154-85.
- Zubairy, Sarah. "On fiscal multipliers: Estimates from a medium scale DSGE model." *International Economic Review* 55, no. 1 (2014): 169-195.

Table 1

Strengths and Weaknesses of Various Empirical Approaches

Type	Strengths	Weaknesses
1. Aggregate country-level time series or panel-level estimates	- Estimates are directly informative about national-level multipliers. - Estimates are not tied to a particular theoretical model.	- Identification of exogenous policy shocks is often challenging. - Estimates are based on averages in a historical sample. - Difficult to construct counterfactuals.
2. Estimated or calibrated DSGE models	- Estimates are directly informative about national-level multipliers. - Estimates can be used to form counterfactuals.	- Identification is based on strong assumptions about the particular model structure and the driving force processes.
3. Subnational geographic cross-sectional or panel estimates	- Identification is often much easier and stronger – uses applied micro identification. - Many different possible data sets.	- Estimates only <i>relative effects</i>, so not directly informative about national multipliers since national factors are differenced out. - Requires additional identification assumptions in the form of a DSGE model to translate subnational to national multipliers.
4. Individual industry, firm or household estimates of key parameters (such as marginal propensities to consume (MPCs)).	- Identification is often much easier and stronger – uses applied micro identification. - Many different possible data sets.	- Only estimates some key micro parameters, so it is not directly informative about national multipliers. - A DSGE model is required to translate the micro parameter estimates to national predictions about multipliers.

Table 2

Estimates of Government Spending Multipliers using Aggregate Data, No State Dependence

Method/Sample	Multipliers	Comments Cumulative multipliers, typically over horizon 0 to 20 quarters
A. U.S. data time series analysis		
Updated implementation of Blanchard-Perotti identified SVAR 1939q1 – 2015q4 1947q1 – 2015q4	0.6 to 0.8 0.6 to 0.7	The tax response is positive for the 1939q1-2015q4 period, but is essentially 0 for the later periods.
Updated implementation of military news identified SVAR 1939q1 – 2015q4 1947q1 – 2015q4	0.7 – 0.8 0.4 to 1*	Tax response is positive for 1939q1-2015q4 period. *Underlying estimates are less precise.
Hall (2010), Barro and Redlick (2011) – annual defense spending	0.6 to 0.7	The Barro-Redlick analysis nets out effects of changes in tax rates.
Ramey-Zubairy (2018), local projections on historical data using military news, 1889q1–2015q4	0.6 to 0.8	The standard error of the estimates range from 0.04 to 0.06.
Mountford and Uhlig (2009), SVAR with sign restrictions	0.65	Deficit-financed increase in government spending.
B. Estimated New Keynesian DSGE models		
Cogan et al. (2010), estimated Smets-Wouters DSGE model on U.S. 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, U.S. and Europe	0.7 to 1	Based on the two year cumulative multipliers shown in the upper left graph in their Figure 6.
Zubairy (2014), estimated medium scale DSGE model estimated on U.S. data.	0.7 to 1.05	Deficit financed
Leeper, Traum, and Walker (2017), estimated DSGE model on U.S. data	0.7 to 1.36	Active monetary policy, Table 7
C. Other country time series data		
Iltzetzki, Mendoza, Vegh (2010), BP identification in SVAR, quarterly data, 1960-2007, 44 countries High income countries	0.3 to 0.7	
Corsetti, Meier, and Mueller (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 fiscal consolidations		They find that the lower multiplier for spending-based consolidations is partly due to more monetary policy accommodation.

Spending-based consolidation	0.3	
Alesina, Favero, and Giavazzi (forthcoming). Narrative analysis of austerity plans, 16 OECD economies from 1978 - 2014.	0.3	They find that the smaller spending multipliers are not due to monetary accommodation.

Table 3

Some Estimates of Government Spending Multipliers using Aggregate Data, Conditional on Key States of the Economy.

Method/Sample	Cumulative multiplier estimates, typically over horizons 0 to 20 quarters	Comments
A. Recessions or High Unemployment		
Auerbach and Gorodnichenko (2012), BP identification, STVAR, post-WWII U.S.	Expansions -0.3 Recessions 2.2	Multipliers calculated under the assumptions (1) government spending cannot get the economy out of a recession; (2) recessions last five years.
Ramey-Zubairy (2018) online appendix, application of Jorda local projection to Auerbach-Gorodnichenko (2012) data.	Expansions -0.6 Recessions 0.8	
Alloza (2017), U.S. data	Expansions small positive Recessions negative	
Ramey-Zubairy (2018), historical U.S. data from 1889 – 2015, local projections		
Military news shock:	Low unemp state 0.6 to 0.7 High unemp state 0.6 to 0.7	
Blanchard-Perotti shock:	Low unemp state 0.3 to 0.4 High unemp state 0.7 to 0.8	
Alesina et al. (2018), Narrative analysis of austerity plans, 16 OECD economies from 1978 - 2014.	Expansion 0.75 Recession 0.58	From Table 10, expenditure based plans.
B. ZLB or Monetary Accommodation		
Estimates based on 1930s-1940s		
Ramey (2011a) – US	0.7	
Crafts-Mills (2013) - UK	0.3 to 0.8	
Ramey-Zubairy (2018), U.S., military news		For full sample, the standard errors are 0.03 to 0.1; for the reduced sample they are 0.1 to 0.15.
1889 – 2015 sample	0.6 to 0.8	
1889 – 2015, omitting period of WWII rationing	1 to 1.4	

Miyamoto-Nguyen-Sergeyev (2018), Japan, Blanchard-Perotti identification	1.4 to 2.6	Standard errors vary between 0.43 and 1.1.
Coenen et al. (2012), large-scale macro models used by central banks and IMF, U.S. and Europe	1 to 3	From Fig. 6 – 2 year increase in government spending and 2 years of monetary accommodation.

Table 4

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	Peak cumulative multipliers within first 3 years.	Comments
U.S. data		
Mountford and Uhlig (2009), SVAR with sign restrictions, U.S. data	-5†	
Romer and Romer (2010), narrative series of tax changes unrelated to current economy, U.S. 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 U.S. data, tax rate shocks.	-1.1	
Mertens and Ravn (2012, 2013) – refinement of Romer and Romer series used in a proxy SVAR	-2.5 to -3	The peak output effects occurs in the first 18 months.
Other country data		
Cloyne (2013), narrative, U.K.	-2.5	
Hayo-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 – 2009, with focus on VAT rate changes	-3.5	
Alesina, Azzalini, Favero, Giavazzi, Miano (2018) Narrative analysis of austerity plans, 16 OECD economies from 1978 – 2014, taxed based consolidations		
Static primary surplus in denominator	-1 to -1.6	
Actual response of primary surplus in denominator	-2.3 to -3.7†	

Table 5

Multipliers in the Wake of the Global Financial Crisis

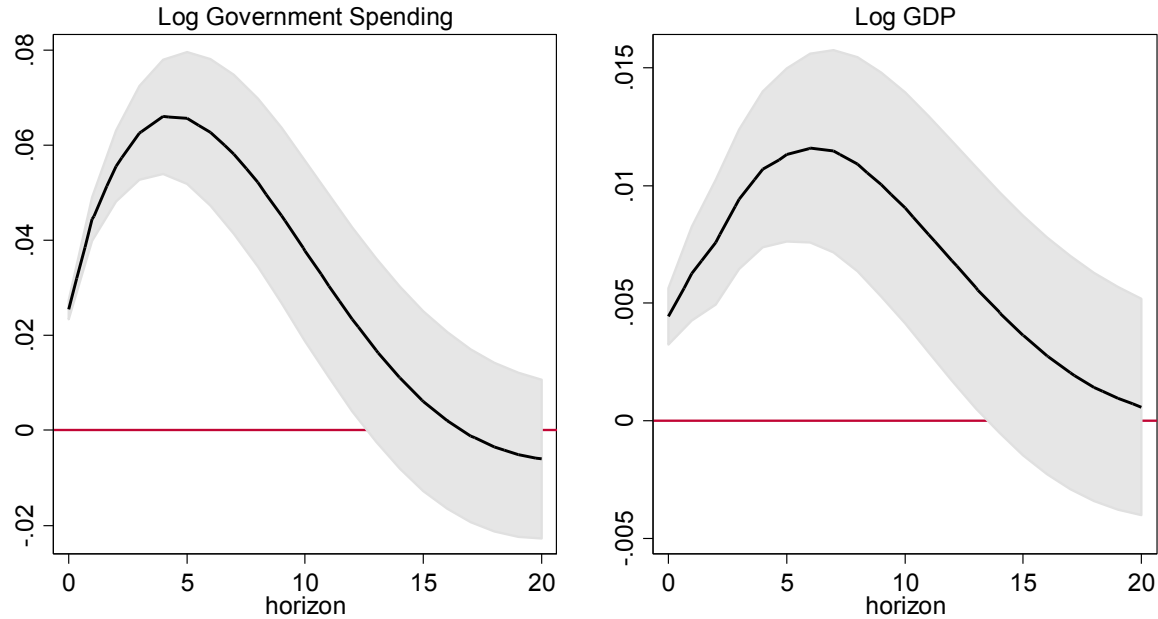
A. Fiscal Consolidations in the Wake of the Financial Crisis		
Indirect evidence based on correlations using forecast errors Blanchard and Leigh (2013, 2014) House, Proebsting, Tesar (2017)	Substantially above 1 2.5	
Alesina et al. (forthcoming), simulations based on pre-crisis estimates and case studies	Multipliers were no different in the wake of the financial crisis.	
B. American Recovery and Reinvestment Act (ARRA)		
Cogan et al. (2010)	0.6 to 0.7	
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. GSP multiplier Cost per job year	 1.7 to 2 2 job-years per \$100K	

Table 6. Conversion of Chodorow-Reich Estimates to National Estimates

	Cumulative Employment Multiplier Estimates – Number of Job Years Created per \$100K of ARRA
Chodorow-Reich headline estimates (his Table 1, column 4)	2.01 (0.59)
Weighted estimates (using Dec. 2008 population of state)	1.15 (0.72)
Weighted estimates	
Estimates based on total spending, including induced spending by states	0.89 (0.45)

Figure 1
Calculating Multipliers from Impulse Response Functions

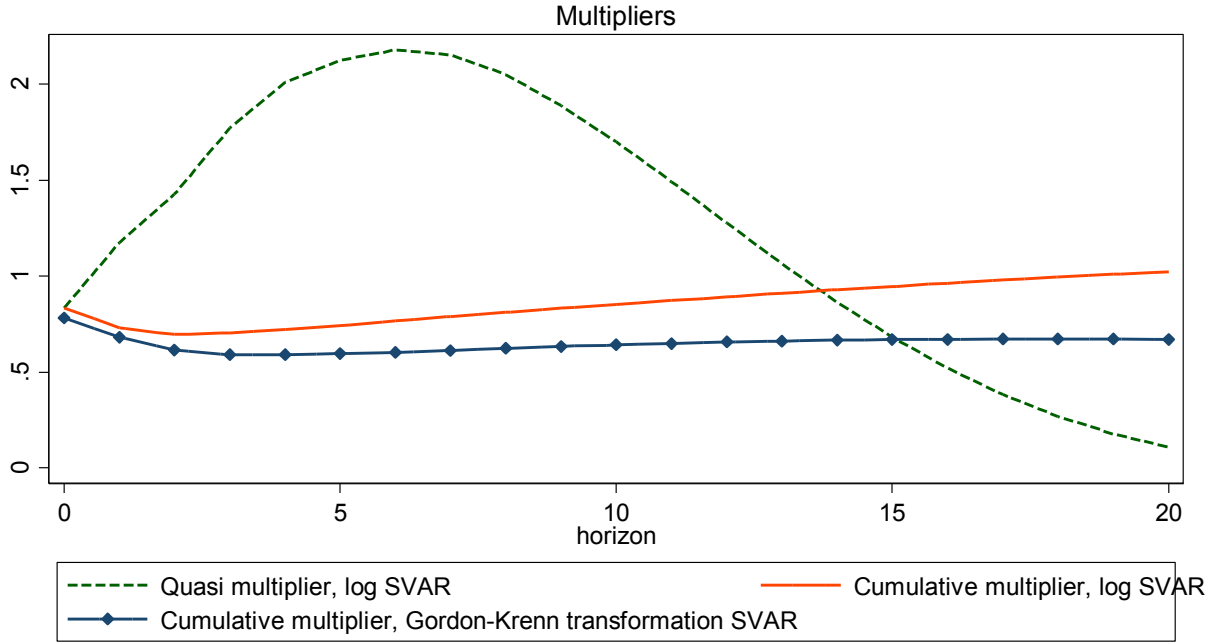
A. Estimated Impulse Response Functions



Source: Author.

Note: Estimated impulse responses based on SVAR estimates using the Ramey-Zubairy (2018) data. The sample is quarterly, 1939:1 – 2015:4. The SVAR contains five endogenous variables: log real total government spending per capita, log real GDP per capita, log real federal tax receipts per capita, the 3-month Treasury bill interest rate, inflation measured as the log change in the GDP deflator. The SVAR includes four quarterly lags of variables, as well as a quadratic trend. The shock to government spending is identified using Blanchard and Perotti's (2002) method, which orders government spending first. The shaded area shows the 95-percent confidence bands.

B. Alternative definitions of multipliers



Source: Author.

Note: The dotted and solid lines show multipliers calculated based on the log impulse responses shown in Panel A. Let $y(j)$ denote the value of the impulse response of $\log(\text{GDP})$ at horizon j and $g(j)$ denote the impulse response of $\log(\text{government spending})$.

Quasi-multiplier(h) = $\left(\overline{Y/G}\right) \cdot y(h) / g(0)$, where $\left(\overline{Y/G}\right) = 4.78$.

Cumulative multiplier(h) = $\left(\overline{Y/G}\right) \cdot \sum_{j=0}^h y(j) / \sum_{j=0}^h g(j)$ for log SVAR

The line with diamonds shows the multiplier using the Gordon-Krenn transformation, where

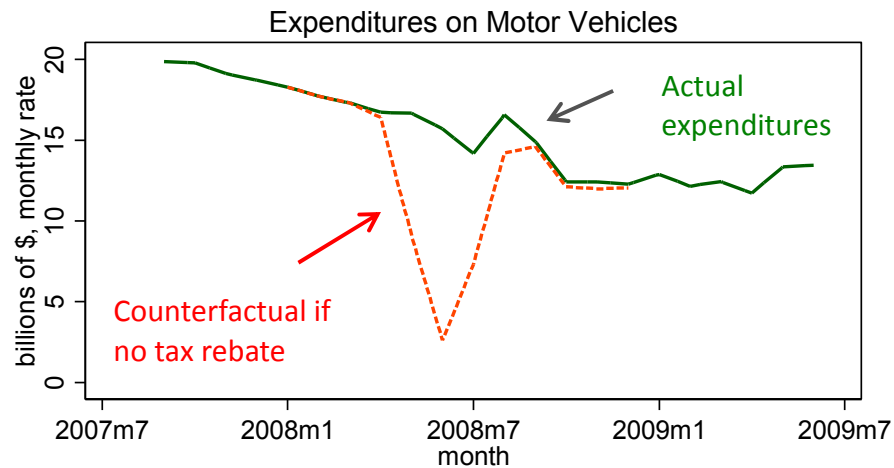
Cumulative multiplier(h) = $\sum_{j=0}^h Y_{GK}(j) / \sum_{j=0}^h G_{GK}(j)$, where Y_{GK} and G_{GK} are the responses of

the Gordon-Krenn transformed variables and the estimates are from the alternative SVAR which uses those variables.

Figure 2. Evidence that Household MPCs Do Not Aggregate

A. Counterfactual Expenditures on New Motor Vehicles

Sahm, Shapiro, and Slemrod (2012) Counterfactual of New Motor Vehicle Expenditures, based on Parker, Souleles, Johnson, McClelland (2013) Estimates of the Marginal Propensity to Spend on New Vehicles from the 2008 Tax Rebate



B. Aggregate Saving and Rebate Rates Relative to Disposable Income without Rebate

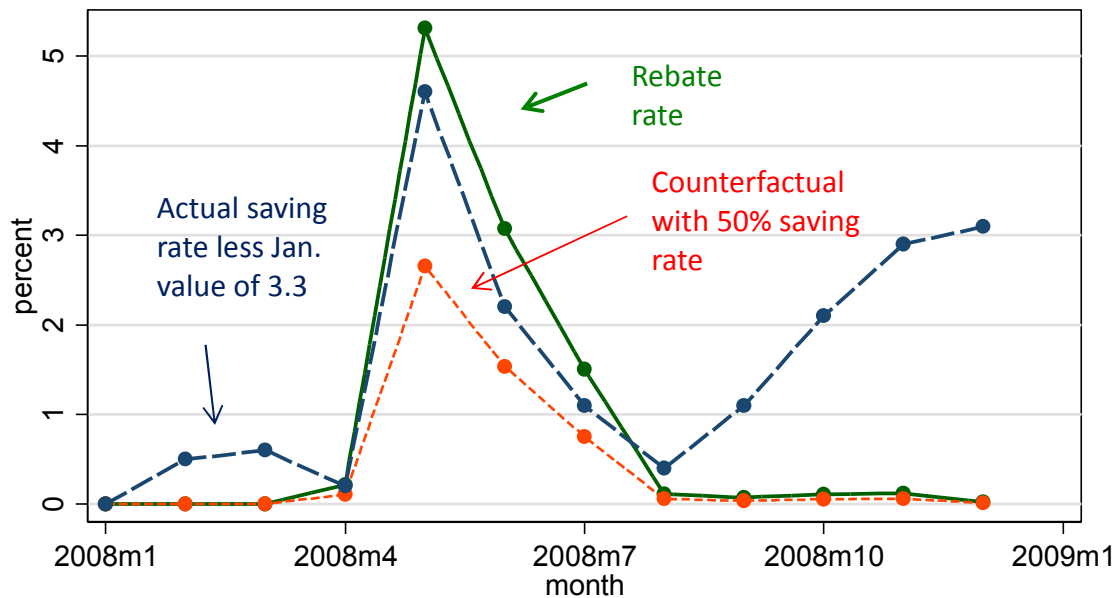


Figure 3. Aggregate counterfactual unemployment rate implied by Chodorow-Reich's Estimates

