EJW
ECON JOURNAL WATCH
Scholarly Comments on Academic Economics

ECON JOURNAL WATCH 12(1)
January 2015: 72-83

Unconventional Confidence Bands in the Literature on the Government Spending Multiplier

Ryan H. Murphy[1]

LINK TO ABSTRACT

In recent decades, vector autoregression, especially structural vector autoregression, has been used to study the size of the government spending multiplier (Blanchard and Perotti 2002; Fatás and Mihov 2001; Mountford and Uhlig 2009). Such methods are used in a significant proportion of empirical research designed to estimate the multiplier (see Ramey 2011a). Despite being published in respected journals and cited by prominent members of the profession, much of this literature does not use the conventional standard of statistical significance that economists are accustomed to in empirical research.

Results in the literature on the fiscal multiplier are typically communicated using a graph of the estimated impulse-response functions. For instance, the effect of government spending on output may be reported by reproducing a graph of an impulse-response function of a one-unit (generally, one percentage point or one standard error) change in government spending. The graph would show the percent change in output over time following the change in government spending. To report statistical significance, authors of these studies may then draw confidence bands around the impulse response function. Ostensibly, if zero lies outside the confidence band, it is statistically distinguishable from zero. But very frequently in this literature the confidence bands correspond to only one standard error. In other words, instead of representing what corresponds to rejecting the null hypothesis at a 90% level or 95% level, the confidence bands correspond to rejecting the null hypothesis at a 68% level. By conventional standards, this confidence band is insufficient for hypothesis testing. Not every useful empirical study must achieve

1. O'Neil Center for Global Markets and Freedom, Southern Methodist University, Dallas, TX 75275.

significance at the 95% level to be considered meaningful, of course, but a pattern of studies which do not use and reach the conventional benchmark is a cause for attention and perhaps concern. Statistical significance is not the only standard by which we should judge empirical research (Ziliak and McCloskey 2008). It is, however, a useful standard, and still an important one.

Here I examine papers in the fiscal multiplier literature which apply vector autoregression methods. Sixteen of the thirty-one papers identified use narrow, one-standard-error confidence bands to the exclusion of confidence bands corresponding to the conventional standard of 90% or 95% confidence. This practice will often not be clear to the reader of a paper unless its text is read rather carefully.

Methodological preliminaries

It is a mistake for economists to take statistical significance too seriously. Without understanding the limits of data availability, or by too stubbornly adhering to arbitrary rules, worthwhile research may never reach the audience it deserves. The mere fact that one regression fails to achieve a certain p-value does not mean that a result is worthless, even though that notion is taught all too frequently via the frequentist interpretation of statistics.

On the other hand, statistical significance is a useful standard by which to judge empirical research. In some ways statistical significance is an arbitrary construct, but it offers one potential way for applied economists to decide which empirical results they find most credible. It is not to be ignored, or perhaps worse, obscured. And while economists should not use $p < 0.05$ as a lazy litmus test, it remains an important point of reference. The onus is on the practitioners of a subfield whose favored econometric method does not pass the conventional standard to demonstrate why they should not be suspected of simply moving the goalposts.

In certain contexts, when a lack of robustness has been made especially obvious, economists will change their minds. An example of this is the work of William Easterly in casting doubt on the then-fashionable belief that foreign aid could encourage growth in developing nations when the developing nations adopt good economic policies. Craig Burnside and David Dollar (2000) had previously developed evidence across many specifications that such an interaction term was statistically significant and in the 'correct' direction. Easterly (2003) showed that this result was contingent on the definitions of aid and of good policy; when these definitions were adjusted, the statistical significance of the results disappeared. Easterly's evidence led economists to be far more skeptical of the work of Burnside and Dollar, even though in most of Easterly's specifications *the point estimate of the*

aid-policy interaction term remained the same in sign and magnitude. Several of Easterly's specifications even resulted in t-statistics for the redefined aid-policy interaction greater than one, as high as 1.41 (2003, 41-44). Yet showing the fragility of the reported statistical significance made development economists more circumspect in their support of aid.

Unconventional standard error bands in studies of the fiscal multiplier

> The standard error bands shown are only 68% bands, based on bootstrap standard errors. Although this is common practice in the government spending literature, it has no theoretical justification. … Some have appealed to Sims and Zha (1999) for using 68% bands. However, there is no formal justification for this particular choice. It should be noted that most papers in the monetary literature use 95% error bands. (Ramey 2011b, 11, 11 n.11)

> The numbers in brackets are the one standard error confidence bands from the bootstrap distribution of multipliers. (Bachmann and Sims 2012, 244)

> The asterisks indicate significance within the one-standard-deviation bandwidth. (Burriel et al. 2010, 265)

As Valerie Ramey (2011b, 11) notes, the decision to use confidence bands of one standard error is an ad hoc departure from professional standards. There are sometimes, to be sure, good reasons to include more bands than only the ones corresponding to two standard errors. The inclusion of bands corresponding to the 68% level, in addition to 95% bands, may communicate the "shape" of the results better than the latter alone (Sims and Zha 1999).[2] This paper takes no issue with such a practice of using multiple bands, but the *exclusion* of those bands corresponding to conventional notions of statistical significance is problematic. Authors who employ the narrower bands should provide a rationale for doing so and make their use of unconventional confidence levels obvious to the casual reader. It is not cynical to expect that most readers of journal articles skim the

2. "We argue that the conventional pointwise bands common in the literature should be supplemented with measures of shape uncertainty… [F]or characterizing likelihood shape, bands that correspond to 50% or 68% posterior probability are often more useful than 95% or 99% bands" (Sims and Zha 1999, 1113, 1118).

articles, reading only to the point that they believe they can understand the model or results. Economists are trained to perceive confidence bands as implicit hypothesis testing, and so there is a danger that unconventional confidence bands may project a false sense of the power of the result. Figure 1 and Figure 2 are two examples of graphs that at first glance may seem to convey statistical significance as conventionally understood; however, the confidence bands employed correspond to 68%, not the usual 90% or 95%.

Regrettably, some economists have further muddled the matter by using one standard error as the threshold for deploying the term "statistically significant." Raffaela Giordano et al. (2007) do so, despite providing both one- and two-sigma bands graphically.[3] Giordano and collaborators provide no rationale for this use of "statistically significant" save for the remark that the same was done "in most previous studies" (2007, 716). But the argument that providing the 68% bands is desirable for certain technical reasons, such as the capacity to convey notions of shape with greater precision, does not tell us that the 68% level is also an acceptable standard for hypothesis testing. If one wishes to argue that the two-sigma confidence level is too high of a standard to apply to one's empirical research, one should offer a compelling reason why the field of study in question can be held to a lower standard of significance or explain why economists in other fields should reevaluate their methods.

Literature review and results

To investigate the prevalence of these issues, I studied 31 papers which applied some variation of vector autoregression in measuring the size of the multiplier.[4] These papers varied from unpublished manuscripts to publications in the *Quarterly Journal of Economics*. A similar exercise could be performed for the tax multiplier. I consider only the debate on the size of the government spending multiplier because the debate over that is most in vogue. Also, the most prominent line of contemporary research on the tax multiplier does not emphasize vector autoregression.[5]

3. Giordano et al. (2007) cite Sims and Zha (1999) for the idea that "error bands corresponding to 0.50 or 0.68 probability…provide a more precise estimate of the true coverage probability" (Giordano et al. 2007, 716 n.8).

4. These papers were collected from work from an in-progress larger literature review which attempts to review all existing empirical literature on the size of the government spending multiplier. The thirty papers are those which employ vector autoregression as described in this paper. One additional paper was identified by a referee.

5. A recent and important paper in the literature on the tax multiplier, Romer and Romer (2010), includes a vector autoregression that uses the one-standard-error confidence bands.

A table of the 31 papers can be found in Appendix A, giving the number of papers citing them (according to Google Scholar) and whether the result was published in a top-100 journal (according to RePEc's "simple" journal rankings). Such metadata allows me to evaluate whether, say, only the less prominent papers use the unconventional confidence bands. Full bibliographic information on the 31 papers is given in Appendix B.

Sixteen of the thirty-one papers use the unconventionally narrow confidence bands at the exclusion of other confidence bands. Papers which provide the 90% or 95% confidence bands, even once in the paper, are not included among the sixteen. Five of the six papers with at least 500 citations use the unconventionally narrow bands.[6] As for papers with between 100 and 500 citations, there are eleven, and four of these used the unconventionally narrow bands. Looking at papers with less than 100 citations, eight of fourteen used the unconventionally narrow bands. Regarding journal placement, thirteen of the thirty-one papers were published in a top-100 journal. Of those thirteen, six use the unconventionally narrow bands and seven do not. Of the eighteen papers not published in a top-100 journal, ten use the unconventionally narrow bands and eight do not.

A narrative consistent with these facts is that the papers on the frontier of empirical macroeconomic methodology used the unconventionally narrow bands and others followed. In the middle (still respectably cited) tier of the literature, the one-standard-error practice is less pervasive. Finally, there is no strong evidence that the top journals are playing an important role in limiting the use of these narrower bands. It seems fair to say that the usage is pervasive within the literature and that those who employ the practice are following the most cited papers in the field.

Final remarks

Further research may bring even further into question whether this literature is held to the same standard as other areas of applied econometric analysis. As Ramey (2011a) has shown, the disagreement among macroeconomists today is whether the government spending multiplier is less than 1 or greater than 1, not whether it is greater than zero. That being the case, an important hypothesis to test is whether a multiplier point estimate is statistically different than 1, not zero, but that test is rarely performed. Where it has become common practice to lower the threshold to one standard error, it is hard to imagine point estimates are commonly

6. These five are Blanchard and Perotti (2002), Fatás and Mihov (2001), Galí et al. (2007), Mountford and Uhlig (2009), and Perotti (2005).

statistically distinguishable from 1. Another obfuscating issue is the lack of standardization in reporting the results of an impulse-response function, as quite often papers will merely report the first period ('impact') multiplier or the value of the impulse-response function at its highest point ('peak'). Beyond the narrow issue of confidence bands, statistical significance generally is an issue with which the literature struggles.

Proponents of vector autoregression should be more transparent about its limitations. While in some ways it may be uniquely useful for macroeconomic topics (Sims 2010), its inability to pass conventional standards is demonstrative of its low power. When a headline point estimate is not statistically significant as conventionally defined, that fact should be made clear, not obscured. Otherwise, we could be holding this method to a lower standard than we hold competing methods. If that is the case, then other empirical methods tending to have smaller standard errors and tighter confidence intervals/bands, such as those pioneered by Valerie Ramey and Robert Barro (see, e.g., Barro and Redlick 2011), or those using clever research designs (e.g., Wilson 2012; Shoag 2013), should be held in higher relative esteem than they are presently.

Figure 1. Reproduction of "Figure 1" in Jordi Galí et al. (2007, 232)

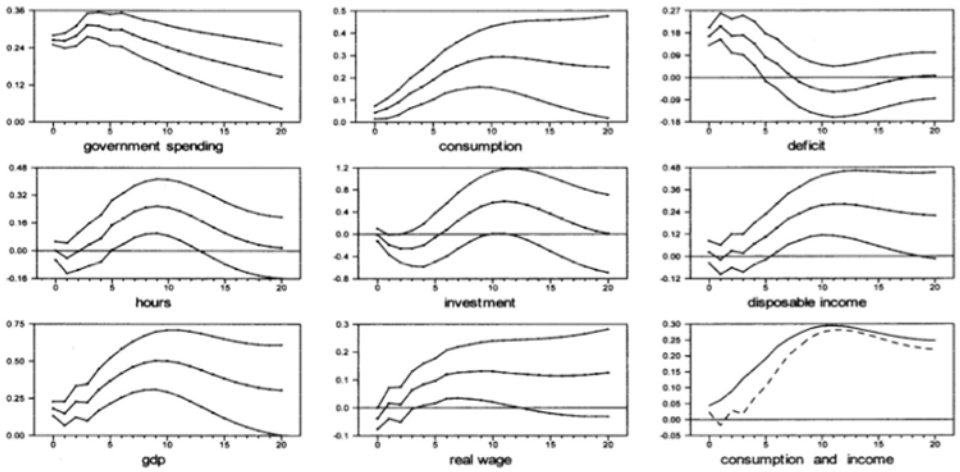

FIGURE 1. The dynamic effects of a government spending shock.

Note: Estimated impulse responses to a government spending shock in the large VAR. Sample Period 1954:I–2003:IV. The horizontal axis represents quarters after the shock. Confidence intervals correspond ±1 standard deviations of empirical distributions, based on 1,000 Monte Carlo replications. The right bottom panel plots the point estimates of both consumption (solid line) and disposable income (dashed line).

Figure 2. Reproduction of "Figure V" in Olivier Blanchard and Roberto Perotti (2002, 1348)

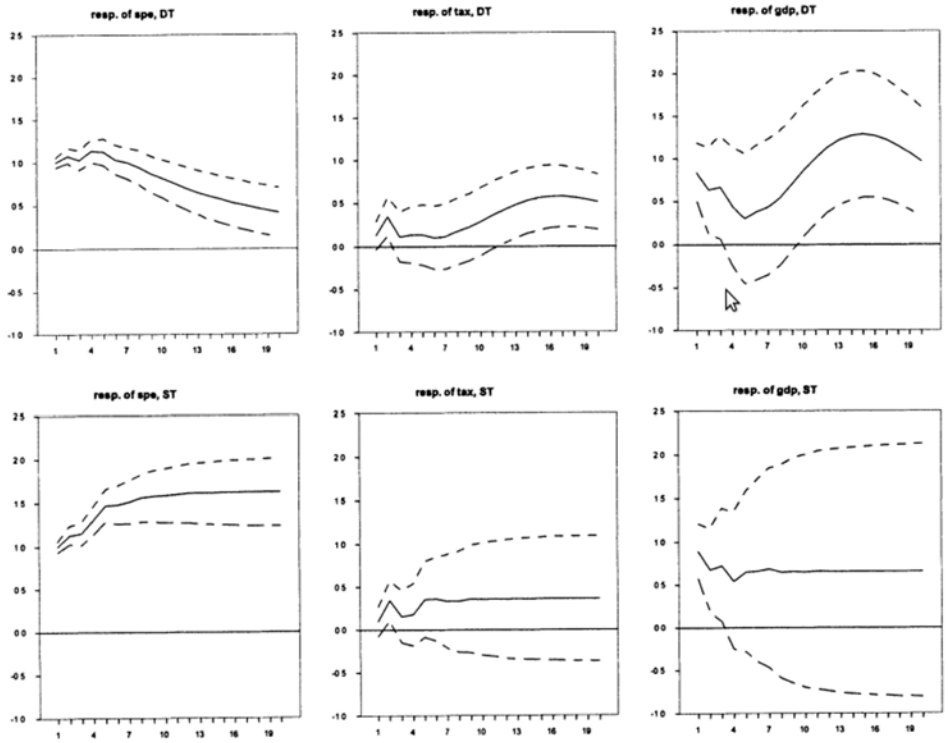

FIGURE V
Response to a Spending Shock

Appendix A.
Use of unconventionally narrow confidence bands in the literature on the fiscal multiplier

Paper	Use 1 s.e. confidence bands	Google Scholar citations	Top-100 Journal (RePEc)
Bachmann and Sims (2012)	Yes	50	Yes
Beetsma and Giuliodori (2011)	No	65	Yes
Beetsma et al. (2008)	No	128	Yes
Bénétrix and Lane (2009)	Yes	28	No
Blanchard and Perotti (2002)	Yes	1999	Yes
Burnside et al. (2004)	No	372	Yes
Burriel et al. (2010)	Yes	50	No
Caldara and Kamps (2008)	Yes	102	No
Cimadomo and Bénassy-Quéré (2012)	No	3	No
Corsetti and Müller (2006)	Yes	183	Yes
Corsetti et al. (2012)	Yes	130	Yes
De Castro (2006)	Yes	62	No
Fatás and Mihov (2001)	Yes	529	No
Fazzari et al. (2013)	No	33	No
Galí et al. (2007)	Yes	1050	Yes
Giordano et al. (2007)	No	151	No
Ilzetzki et al. (2013)	No	309	Yes
Kirchner et al. (2010)	Yes	65	No
Kuttner and Posen (2002)	No	67	No
Monacelli et al. (2010)	No	103	Yes
Mountford and Uhlig (2009)	Yes	933	Yes
Pereira (2008)	Yes	1	No
Pereira and Lopes (2010)	Yes	22	No
Perotti (2005)	Yes	927	No
Perotti (2007)	Yes	330	No
Ramey (2011b)	No	733	Yes
Ramey (2012)	No	33	No
Ravn et al. (2007)	No	132	No
Ravnik and Žilić (2011)	No	7	No
Stevans and Sessions (2010)	No	1	No

Appendix B. Bibliographical information for the papers listed in Appendix A

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

Bachmann, Rüdiger, and Eric R. Sims. 2012. Confidence and the Transmission Mechanism of Government Spending Shocks. *Journal of Monetary Economics* 59: 235-249.

Beetsma, Roel, and Massimo Giuliodori. 2011. The Effects of Government Purchases Shocks: Review and Estimates for the EU. *Economic Journal* 131(550): F4-F32.

Beetsma, Roel, Frank Klassen, and Massimo Giuliodori. 2008. The Effects of Public Spending Shocks on Trade Balance and Budget Deficits in the European Union. *Journal of the European Economic Association* 6(2-3): 414-423.

Bénétrix, Agustin S., and Philip R. Lane. 2009. The Impact of Fiscal Shocks on the Irish Economy. *Economic and Social Review* 40(4): 407-434.

Blanchard, Olivier, and Roberto Perotti. 2002. An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics* 117(4): 1329-1368.

Burnside, Craig, Martin Eichenbaum, and Jonas D. M. Fischer. 2004. Fiscal Shocks and Their Consequences. *Journal of Economic Theory* 115: 89-117.

Burriel, Pablo, Francisco de Castro, Daniel Garrote, Esther Gordo, Joan Paredes, and Javier J. Pérez. 2010. Fiscal Policy Shocks in the Euro Area and the US: An Empirical Assessment. *Fiscal Studies* 31(2): 251-285.

Caldara, Dario, and Christophe Kamps. 2008. What Are the Effects of Fiscal Policy Shocks? A VAR-Based Comparative Analysis. *European Central Bank Working Paper Series* 877. European Central Bank (Frankfurt am Main). **Link**

Cimadomo, Jacopo, and Agnès Bénassy-Quéré. 2012. Changing Patterns of Fiscal Policy Multipliers in Germany, the UK, and the US. *Journal of Macroeconomics* 34: 845-873.

Corsetti, Giancarlo, and Gernot J. Müller. 2006. Twin Deficits: Squaring Theory, Evidence and Common Sense. *Economic Policy* 48: 597-633.

Corsetti, Giancarlo, Andre Meier, and Gernot J. Müller. 2012. What Determines Government Spending Multipliers? *Economic Policy* 72: 558-565.

de Castro, Francisco. 2006. The Macroeconomic Effects of Fiscal Policy in Spain. *Applied Economics* 38(8): 913-924.

Fatás, Antonio, and Ilian Mihov. 2001. The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence. *CEPR Discussion Paper* 2760. Centre for Economic Policy Research (London). **Link**

Fazzari, Steven M., James Morley, and Irina Panovska. 2013. State-Dependent Effects of Fiscal Policy. Working paper. **Link**

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

Giordano, Raffaela, Sandro Momigliano, Stefano Neri, and Roberto Perotti. 2007. The Effects of Fiscal Policy in Italy: Evidence from a VAR Model. *European Journal of Political Economy* 23(3): 707-733.

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

Kirchner, Marcus, Jacopo Cimadomo, and Sebastian Hauptmeier. 2010. Transmission of Government Spending Shocks in the Euro Area: Time Variation and Driving Forces. *European Central Bank Working Paper Series* 1219. European Central Bank (Frankfurt am Main). **Link**

Kuttner, Kenneth N., and Adam S. Posen. 2002. Fiscal Policy Effectiveness in Japan. *Journal of the Japanese and International Economies* 16(4): 536-558.

Monacelli, Tommaso, Roberto Perotti, and Antonella Trigari. 2010. Unemployment Fiscal Multipliers. *Journal of Monetary Economics* 57(5): 531-553.

Mountford, Andrew, and Harald Uhlig. 2009. What Are the Effects of Fiscal Policy Shocks? *Journal of Applied Econometrics* 24(6): 960-992.

Pereira, Manuel C. 2008. Empirical Evidence on the Stabilizing Role of Fiscal and Monetary Policies in the US. Working paper. **Link**

Pereira, Manuel C., and Artur S. Lopes. 2010. Time-Varying Fiscal Policy in the U.S. *Banco de Portugal Estudos e Documentos de Trabalho* 21-2010. Banco de Portugal (Lisbon). **Link**

Perotti, Roberto. 2005. Estimating the Effects of Fiscal Policy in OECD Countries. *CEPR Discussion Paper* 4842. Centre for Economic Policy Research (London). **Link**

Perotti, Roberto. 2007. In Search of the Transmission Mechanism of Fiscal Policy. *NBER Macroeconomics Annual* 22: 169-226.

Ramey, Valerie A. 2011b. Identifying Government Spending Shocks: It's All In the Timing. *Quarterly Journal of Economics* 126(1): 51-102.

Ramey, Valerie A. 2012. Government Spending and Private Activity. *NBER Working Paper* 17787. National Bureau of Economic Research (Cambridge, Mass.). **Link**

Ravn, Morten O., Stephanie Schmitt-Grohé, and Martín Uribe. 2007. Explaining the Effects of Government Spending on Consumption and the Real Exchange Rate. *NBER Working Paper* 13328. National Bureau of Economic Research (Cambridge, Mass.). **Link**

Ravnik, Rafael, and Ivan Žilić. 2011. The Use of SVAR Analysis in Determining the Effects of Fiscal Shocks in Croatia. *Financial Theory and Practice* 35(1): 25-58.

Stevans, Lonnie K., and David N. Sessions. 2010. Calculating and Interpreting Multipliers in the Presence of Non-Stationary Time Series: The Case of U.S. Federal Infrastructure Spending. *American Journal of Social and Management Science* 1(1): 24-38.

References

Bachmann, Rüdiger, and Eric R. Sims. 2012. Confidence and the Transmission Mechanism of Government Spending Shocks. *Journal of Monetary Economics* 59(3): 235-249.

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

Blanchard, Olivier, and Roberto Perotti. 2002. An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *Quarterly Journal of Economics* 117(4): 1329-1368.

Burnside, Craig, and David Dollar. 2000. Aid, Policies, and Growth. *American Economic Review* 90(4): 847-868.

Burriel, Pablo, Francisco de Castro, Daniel Garrote, Esther Gordo, Joan Paredes, and Javier J. Pérez. 2010. Fiscal Policy Shocks in the Euro Area and the US: An Empirical Assessment. *Fiscal Studies* 31(2): 251-285.

Easterly, William. 2003. Can Foreign Aid Buy Growth? *Journal of Economic Perspectives* 17(3): 23-48.

Fatás, Antonio, and Ilian Mihov. 2001. The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence. *CEPR Discussion Paper* 2760. Centre for Economic Policy Research (London). **Link**

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

Giordano, Raffaela, Sandro Momigliano, Stefano Neri, and Roberto Perotti. 2007. The Effects of Fiscal Policy in Italy: Evidence from a VAR Model. *European Journal of Political Economy* 23(3): 707-733.

Mountford, Andrew, and Harald Uhlig. 2009. What Are the Effects of Fiscal Policy Shocks? *Journal of Applied Econometrics* 24: 960-992.

Perotti, Roberto. 2005. Estimating the Effects of Fiscal Policy in OECD Countries. *CEPR Discussion Paper* 4842. Centre for Economic Policy Research (London). **Link**

Ramey, Valerie A. 2011a. Can Government Purchases Stimulate the Economy? *Journal of Economic Literature* 49(3): 673-685.

Ramey, Valerie A. 2011b. Identifying Government Spending Shocks: It's All In the Timing. *Quarterly Journal of Economics* 126(1): 51-102.

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

Shoag, Daniel. 2013. Using State Pension Shocks to Estimate Fiscal Multipliers Since the Great Recession. *American Economic Review* 103(3): 121-124.

Sims, Christopher A. 2010. But Economics Is Not an Experimental Science. *Journal of Economic Perspectives* 24(2): 59-68. **Link**

Sims, Christopher A., and Tao Zha. 1999. Error Bands for Impulse Response Functions. *Econometrica* 67(5): 1113-1155.

Wilson, Daniel J. 2012. Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act. *American Economic Journal: Economic Policy* 4(3): 251-282.

Ziliak, Stephen T., and Deirdre N. McCloskey. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor, Mich.: University of Michigan Press.

About the Author

Ryan Murphy is a Research Associate at the O'Neil Center for Global Markets and Freedom at the Southern Methodist University Cox School of Business. He received his Ph.D. in economics from Suffolk University in 2013. His email address is rhmurphy@smu.edu.

Go to archive of Economics in Practice section
Go to January 2015 issue