

## A CALL FOR OUT-OF-SAMPLE TESTING IN MACROECONOMICS

Robert Gelfond  
*MQS Management LLC*

Ryan H. Murphy\*  
*Southern Methodist University*  
[rhmurphy@smu.edu](mailto:rhmurphy@smu.edu)

### ABSTRACT

Social science has come under a firestorm of criticism following the discovery that many findings cannot be replicated. This paper argues that increasing the use of out-of-sample tests would improve the reliability of findings in all fields, including “complex” fields such as macroeconomics. It illustrates the paucity of examples of this method by reviewing the literature on the government spending multiplier. Two out-of-sample tests in academic papers from the 1960s were identified, along with a procedure similar to an out-of-sample test in a recent paper by Jorda and Taylor (2016).

### 1. Introduction

Out-of-sample testing should be a useful item in the social scientist’s toolkit. When a contentious issue arises, social science models, specified and parameterized in advance and tested against new data, circumvents many of the issues present in empirical work. The contentious social scientific issue this paper focuses on is the value of the government spending multiplier, and the lack of out-of-sample tests in support of the empirical results related to it. This paper’s primary contribution is to document this fact within the literature. The lack of empirical research of this nature should raise concerns of the multiplier’s external validity. But this is also illustrative of the problems in the ongoing Replication Crisis (see Gelman and Loken 2014) as well.

Concerns regarding the validity of statistical research from within econometrics itself dates to at least Leamer (1983), and Ioannidis (2005) offered an early warning that “[m]ost published research findings are false” prior to the true onset of the Replication Crisis. The Reproducibility Project, in *Science*, reexamined 100 psychology studies published in elite journals and found that only one third to one half could be reproduced (Open Science Collaboration 2015). A similar project found that only around half of the results in economics articles could be reproduced (Chang and Li 2015). Some fields within economics have attempted to combat these issues, for example the academic journal *Public Finance Review* has itself dedicated space and intellectual support for research replicating earlier results (Alm and Reed 2015).

While we recommend out-of-sample testing, others have proposed other solutions to the issues facing social science today. Hanson (1995) early on recommended that betting could drastically improve honest assessments of evidence. Nosek et al. (2012) emphasizes the importance of keeping all data, materials, and workflow open. Ioannidis et al. (2017) suggests that statistical power, and not simply statistical significance, is a superior means of demarcating effective statistical tests. Silver (2012: 42-44), whom we follow, advocates out-of-sample testing. We see value in out-of-sample testing, including in fields as complex as macroeconomics. For the purposes of this paper, we define an out-of-sample test as one which uses a subset of data as a training sample to parameterize the model, with another independent subsample withheld to test the model against a benchmark (the null hypothesis). This framework can be read in terms of Friedman (1953), Popper ([1959] 2009), and Tetlock and Gardner (2015).

The debate this paper will focus on is the government spending multiplier. The primary contribution of this paper will be an extensive literature review of 87 empirical papers on the multiplier, documenting that

\*Corresponding author

Research Assistant Professor, the O’Neil Center for Global Markets and Freedom, SMU Cox School of Business P.O. Box 750333, Dallas, TX 75275.

This research was performed with the support of Americans for Limited Government.

almost none have performed an out-of-sample test.<sup>1</sup> Silver (2012: 40-42) underlines that the uncertainty associated with such research, combined with the 2007-2008 Financial Crisis's distance from previous observations – that is, distantly out-of-sample – may have led to the subsequent difficulties and failures of policy worldwide. It seems straightforward that such a crucial question deserves rudimentary methods that are at least suggestive of external validity on the topic.

Before commencing, we should first credit the existent literature within macroeconomics that does attempt out-of-sample testing. Some recent macroeconomic research does perform out-of-sample testing, and complex modern models are not outperformed by, for example, AR(1). Notably, Smets and Wouters (2007) offer an out-of-sample test which outperforms a baseline vector autoregression and, admirably, make this fact clear in the context of their article. Rotemberg and Woodford (1992) show that the assumption of oligopolistic pricing outperforms the assumption of perfectly competitive pricing regarding the path of an economy following a change in military spending, but the manner in which it was parameterized prevents its interpretation as a proper out-of-sample test.<sup>2</sup> Similar in spirit to this paper, but not quite in content, Wieland et al. (2012) allow several macroeconomic models to compete, with none doing well in predicting the recession and recovery. This last approach is essential for further advances in macroeconomics; out-of-sample testing should be far more commonplace, especially with respect to the size of the multiplier.

The structure of this paper follows. Section II presents the philosophical case for testing of this type. Section III provides detail on the literature review, the complete results of which can be found in Appendix A. The literature review demonstrates the issue as well as provides the primary finding of the paper. Section IV concludes.

## 2. Philosophical and Practical Issues with Out-of-Sample Testing in Macroeconomics

While the Replication Crisis has opened up social scientists' minds to changing certain empirical practices, out-of-sample testing may appear inappropriate in fields as complex as macroeconomics. Testing whether a model outperforms a simple baseline out-of-sample is rare; one counterexample being Smets and Wouters (2007). There is a certain amount of resistance to subjecting theories to such harsh tests. A theory which allows itself a sufficient number of moving parts, unspecified in advance, quickly becomes unfalsifiable in the face of any evidence. For example, the historical Hayek-Keynes debate can be interpreted as never having been sufficiently specified to allow either theory to be possibly falsified. In the eyes of Hayek or Keynes themselves, the evidence presented against their theory by their respective opponents would be unlikely to ever be taken seriously (see Scheall 2015).

On the one hand, this is true. The idea that no "true" falsification exists (the "Duhem-Quine problem") is well-known, and was clearly evident to Popper even in the first chapter of *The Logic of Scientific Discovery* ([1959] 2009: 19-20).<sup>3</sup> But this is not an issue per se with the desire to conduct empirical tests.<sup>4</sup> A scatterplot is a test, just a bad one. A randomized controlled trial is a better one. All theories will be falsified, sometimes incorrectly by imperfect tests. The question ultimately is which theory, stated in terms which actually make predictions, is *least* falsified, or in other words, predicts better than extant theories, even if it is far from perfect. The solution to complex and difficult areas is neither to cherrypick nor to question whether economics can be studied scientifically (as in Roberts 2010). It is to embrace the fact that all theories are, to some extent, falsified.<sup>5</sup>

Out-of-sample tests can thusly be thought of in this way. Suppose one macroeconomic theory predicts growth following government spending while another predicts contraction. Both theories ahead of time make predictions for the next five years, conditional on several variables (e.g. population growth). However, halfway through the prediction period, a significant deposit of oil is unexpectedly discovered, leading to growth. Neither theory included oil in their conditional predictions. The opponent of government spending could then invoke the "auxiliary hypothesis" (i.e., Duhem-Quine) of the effect of oil discoveries on GDP to defend his model even

<sup>1</sup> A previous journalistic account of this literature review can be found in Gelfond (2014).

<sup>2</sup> This prevents us from interpreting it as an out-of-sample test, not to mention that Rotemberg and Woodford are addressing more directly a theoretical dispute, as opposed to testing the actual value of the multiplier.

<sup>3</sup> For this point we thank Lee Kelly.

<sup>4</sup> Our position here is in opposition to Cross (1982).

<sup>5</sup> Some philosophers have since given some reason to allow for "double counting" – allowing the same data to generate the theory and to test the data – but the larger point still stands. In this frame, testing of macroeconomic theories nearly *always* involves double counting. See, for example, Steele et al. (2018).

though it was outperformed in terms of root-mean squared error. However, the point is that, as long as the world is not systematically biased so as to disproportionately harm the predictions of one type of hypothesis as opposed to another, the better model will win out more often as more out-of-sample tests take place.

It is not that out-of-sample tests guarantee a model's performance going forward, but that those which best survive them offer the best guidance in the future. By their nature, noisy, highly complex systems such as those studied in macroeconomics, will never be able to be perfectly modeled. Incorporating all possibly relevant variables is impossible. The virtue of the out-of-sample test is that it offers a standard of comparison which is not susceptible to issues of data mining and publication bias (among others), which plague so much research today. Yet more extensive uses of out-of-sample tests will only improve comparisons between competing theories, even if knowledge in them fails to be formally justified.

This perspective, taken seriously, does somewhat contradict recent trends in empirical economics. At times it seems like economists are focused like a laser on establishing that the relevant coefficient estimation in their model is consistent. This can be thought of as searching for a testing methodology that is immune from the criticism of Duhem-Quine. But not only is it impossible to find such a test, modern methodologies are often so idiosyncratic that subsequent out-of-sample testing is quite literally impossible. This may not always be the case, but often it is (e.g., the use of past Medicaid allocations to correct for the endogeneity of ARRA spending). On the other hand, certain quite conventional econometric techniques have ambiguous interpretations in terms of out-of-sample tests. For example, when predicting future years in a model which used fixed effects, what value should the fixed effect take? In any case, this is not strictly specific to the credibility revolution.

Out-of-sample testing also offers a complement to the prescriptions of Leamer (1983). Leamer's solution to avoiding researcher degrees of freedom is to run all possible specifications of a regression model to determine robustness. However, Leamer's dictum still grants researchers some ability to game results by searching for alternative data sources, selectively remembering tests performed, or otherwise privileging their preferred hypothesis. The pared down version of Leamer's dictum, the use of robustness checks, is common but has apparently failed to stem the "Con in Econometrics," with the "credibility revolution" now having made the claim to end the "conning" instead (see Harford 2010). Out-of-sample testing circumvents this by forcing researchers to commit to specifications, preferably before even seeing the data (this can be reinforced further via betting).

One interesting side point is that the Lucas Critique may be interpreted differently under this frame. The Lucas Critique may simply be a rule of thumb for theories to improve the performance of theories in out-of-sample tests. Failing to recognize that changes in deep parameters may shift observed historical patterns in the data is one obvious way a theory optimized within sample would subsequently be falsified out of sample. But it is also true that whether deep parameters are important is itself a falsifiable proposition. If Old Keynesian models which ignore deep parameters outperform those which include them in out-of-sample tests – though this does not appear to be the case (Wieland et al. 2012) – then it is best to ignore the Lucas Critique.

There is little preventing macroeconomists from pitting their theories against one another (or against AR(1)) in an out-of-sample test. Pre-committing to specifications and parameters offers a partial remedy to the Duhem-Quine problem, even if concerns can always still be raised after fact. Econometric methods and issues like the Lucas Critique, viewed through this lens, ultimately concern how poorly designed tests or poorly specified theories may later fail out-of-sample; the out-of-sample test is still the arbiter of how much any of these issues matter or how much they may be ignored.

### 3. Literature Review on the Government Spending Multiplier

The size of the government spending multiplier may be the most contentious and important question in macroeconomics. The historical length of the debate, and the intellectual heft invested in it, makes it a natural starting point for investigating the use of out-of-sample tests in macroeconomics. Eighty-seven papers from 1956 through 2016 are included in the literature review, weighted heavily towards the most recent. A truly complete literature on the topic would be impossible. The last known publicly available version of fifteen of the papers remain in the Working Paper stage. Another 6 papers are book chapters. The papers examined, inevitably, are disproportionately in elite journals or disproportionately highly cited. Of the remaining 66 papers, 34 (52%) are in a top 25 journal according to RePEc. Nothing like an out-of-sample test appears in 84 of the 87 papers examined (97%).

The first borderline out-of-sample test case is Jorda and Taylor (2016). The authors construct a “counterfactual forecast” of the UK economy from data which excludes the Global Financial Crisis. Their model, which typically implies multipliers greater than 2, explains much of the divergence between actual RGDP data and both that which was forecasted by the UK Office for Budget Responsibility and what would have occurred on average historically following a high leverage financial crisis recession. But, as the authors say, “this exercise has a flavour of an out-of-sample evaluation,” not a true out-of-sample test. Still, it is admirable.

A truer out-of-sample test can be found in Morishima and Saito (1964), using an Old Keynesian model. The authors use a sample period of 1902-1952 to build their model and find that it is “tolerably effective” in predicting the period 1953-1960; although this is not evaluated in rigorous terms, it is still an out-of-sample test. Similarly, Smith (1967) creates two IS-LM models incorporating the money supply. In Table 2 of this paper, the author forecasts the following eight quarters of data (1962 and 1963), though he notes that the period in the out-of-sample test is of limited use since no “turning points” occurred.

Per our reading of the literature, no other example of an out-of-sample test exists. Far more than eighty-seven empirical papers measuring the size of the multiplier exist, but it appears likely to find more in the distant past of macroeconomics than many others hiding in plain sight in the present. Wieland et al. (2012) give us pause as to whether it is correct to claim that these antiquated models are better than modern DSGE models, but the apparent unwillingness to employ out-of-sample testing for an issue as controversial as spending multipliers does not speak well of modern macroeconomic methodology, to say the least.

#### 4. V. Conclusion

A rather obvious method to determine the external validity of an empirical result is to perform an out-of-sample test. With the increasing concerns about replicability of results in social sciences, it is a surprise that few, besides Silver (2012), have focused in on out-of-sample testing as a methodological standard (if not a litmus test or a line of demarcation). This is illustrated by the near-absence of out-of-sample testing in the literature on the size of the government spending multiplier, as studied in 87 empirical papers. Out-of-sample tests, to the extent that they exist, are to be found primarily in two academic papers published in the 1960s. More broadly in macroeconomics, effective out-of-sample testing can be found in Smets and Wouters (2007) and the approach of Wieland et al. (2012) is especially laudable.

Tournaments amongst various models should more readily be undertaken, even if they take years or even decades to play out. Replication studies also play a clear role in this (for instance, by identifying the most elite models and their final parameterizations as of the year 2000, and have them compete using data since 2000). The specifics of the implementations could be widely varied, so long as the spirit of the out-of-sample test is maintained.

Philosophical or practical concerns raised by researchers regarding these types of testing are generally unhelpful; however complex an issue is, a model-based forecast is still possible, preferably buttressed by bets, as in Hanson (1995). Encouraging a norm of betting on beliefs may reduce the inflammatory rhetoric associated with issues like the effectiveness of fiscal stimulus (e.g. Krugman 2015); “A Bet is a Tax on Bullshit” (Tabarrok 2012). Out-of-sample tests may not be the only factor in evaluating competing hypotheses, just as p-values should not be. But basic reporting of root-mean square error relative to some simple baseline should be performed at least occasionally.

## 5. References

- [1] Alm, James and W. Robert Reed. 2015. "The Need for Replications." *Public Finance Review*, 43, no. 2: 139-142.
- [2] Chang, Andrew C. and Phillip Li. 2015. "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journal Journals Say 'Usually Not.'" *Finance and Economics Discussion Paper Series* 2015-083. Washington, DC: Board of Governors of the Federal Reserve System.
- [3] Cross, Rod. 1982. "The Duhem-Quine Thesis, Lakatos and the Appraisal of Theories in Macroeconomics." *The Economic Journal* 92: 320-340.
- [4] Friedman, Milton. 1953. "The Methodology of Positive Economics." In *Essays in Positive Economics*. Chicago: University of Chicago Press.
- [5] Gelfond, Robert. 2014. "The Myth of the Keynesian Multiplier." *Forbes*, <http://www.forbes.com/sites/realspin/2014/11/26/the-myth-of-the-keynesian-multiplier/>
- [6] Gelman, Andrew and Eric Loken. 2014. "The Statistical Crisis in Science." *The American Scientist* 102, no. 6.
- [7] Hanson, Robin. 1995. "Could Gambling Save Science? Encouraging an Honest Consensus." *Social Epistemology: A Journal of Knowledge, Culture, and Policy* 9, no. 1: 3-33.
- [8] Harford, Tim. 2010. "At Last the Con Has Been Taken Out of Econometrics." *Financial Times*, <http://www.ft.com/intl/cms/s/2/ea23cc58-3589-11df-963f-00144feabdc0.html>
- [9] Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2, no. 8: e124.
- [10] Ioannidis, John P. A., T.D. Stanley, and Hristos Doucouliagos. 2017. "The Power of Bias in Economics Research." *The Economic Journal* 127, no. 605: F236-F265.
- [11] Jorda, Oscar and Alan M. Taylor. 2016. "The Time For Austerity: Estimating the Average Treatment Effect of Fiscal Policy." *The Economic Journal* 126, no. 590: 219-255.
- [12] Krugman, Paul. 2015. "Destructive Long-Termism." *The New York Times*, <http://krugman.blogs.nytimes.com/2015/12/28/destructive-long-termism/>
- [13] Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73, no. 1: 31-43.
- [14] Meese, Richard A. and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14, no. 1-2: 3-24.
- [15] Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. 2012. "Scientific Utopia II. Restructuring Incentives and Practices to Promote Truth Over Publishability." *Perspectives on Psychological Science* 7, no. 6: 615-631.
- [16] Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349, no. 6251.
- [17] Popper, Karl. [1959] 2009. *The Logic of Scientific Discovery*. New York: Routledge.
- [18] Roberts, Russ. 2010. "Is the Dismal Science Really a Science?" *The Wall Street Journal* <http://www.wsj.com/articles/SB10001424052748704804204575069123218286094>
- [19] Rotemberg, Julio and Michael Woodford. 1992. "Oligopolistic Pricing and Its Effects of Aggregate Demand on Economic Activity." *Journal of Political Economy* 100, no. 6: 1153-1207.
- [20] Scheall, Scott. 2015. "Slaves of the Defunct: the Epistemic Intractability of the Hayek-Keynes Debate." *Journal of Economic Methodology* 22, no. 2: 215-234.

- [21] Silver, Nate. 2012. *The Signal and the Noise: Why So Many Predictions Fail – But Some Don't*. New York, New York: The Penguin Press.
- [22] Smets, Frank and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97, no. 3: 586-606.
- [23] Smith, Paul E. 1967. "Money Supply and Demand: A Cobweb?" *International Economics Review* 8 (1): 1-12.
- [24] Steele, Katie and Charlotte Werndl. 2018. "Model-Selection Theory: The Need for a More Nuanced Picture of Use-Novelty and Double Counting." *The British Journal for the Philosophy of Science*, 69, no. 2: 351-375.
- [25] Tabarrok, Alex. 2012. "A Bet is a Tax on Bullshit." *Marginal Revolution*, <http://marginalrevolution.com/marginalrevolution/2012/11/a-bet-is-a-tax-on-bullshit.html>
- [26] Tetlock, Philip E. and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction*. New York: Crown Publishers.
- [27] Wieland, Volker, Tobias Cwik, Gernot J. Muller, Sebastian Schmidt, and Maik Wolters. 2012. "A New Comparative Approach to Macroeconomic Modeling and Policy Analysis." *Journal of Economic Behavior & Organization* 83, no. 3: 523-541.

## 6. APPENDIX A. List of Papers Examined

- [1] Acconia, Antonio, Giancarlo Corsetti, and Saverio Simonelli. 2014. "Mafia and Public Spending: Evidence on the Fiscal Multiplier from a Quasi-Experiment." *American Economic Review* 104(7): 2185-2209.
- [2] Almunia, Miguel, Agustin Benetrix, Barry Eichengreen, Kevin H. O'Rourke, and Gisela Rua. 2010. "From Great Depression to Great Credit Crisis: Similarities, Differences, and Lessons." *Economic Policy* 25 (62): 219-265.
- [3] Anderson, Leonall C. and Jerry J. Jordan. 1968. "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization." *Federal Reserve Bank of St. Louis Review* (April): 11-23.
- [4] Auerbach, Alan J. and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy* 4(2): 1-27.
- [5] Auerbach, Alen and Yuriy Gordonichenko. 2013. "Fiscal Multipliers in Recession and Expansion." In *Fiscal Policy after the Financial Crisis*. Chicago: University of Chicago Press.
- [6] Auerbach, Alen and Yuriy Gorodnichenko. 2014. "Fiscal Multipliers in Japan." *NBER Working Paper* no. 19911.
- [7] Bachmann, Rudiger and Eric Sims. 2012. "Confidence and the Transmission Mechanism of Government Spending Shocks." *Journal of Monetary Economics* 59: 235-49.
- [8] Ball, R.J. 1963. "The Significance of Simultaneous Methods of Parameter Estimation in Econometric Models." *Applied Statistics* 12 (1): 14-25.
- [9] Barro, Robert. 1981. "Output Effects of Government Purchases." *Journal of Political Economy* 89 (6).
- [10] Barro, Robert J. and Charles J. Redlick. 2011. "Macroeconomic Effects from Government Purchases and Taxes." *Quarterly Journal of Economics* 126: 51-102.
- [11] Beetsma, Roel and Massimo Giuliodori. 2011. "The Effects of Government Purchases Shocks: Review and Estimates for the EU." *The Economic Journal* 121 (February): 4-32.
- [12] Beetsma, Roel, Frank Klassan, and Massimo Giuliodori. 2008. "The Effects of Public Spending Shocks on Trade Balances and Budget Deficits in the European Union." *Journal of the European Economic Association* 6 (2-3): 414-23.
- [13] Belinga, Vincent and Constant Lonkeng Ngouana. 2015. "(Not) Dancing Together: Monetary Policy Stance and the Government Spending Multiplier." *IMF Working Paper* 15/114.
- [14] Benetrix, Agustin S. and Philip R. Lane. 2009. "The Impact of Fiscal Shocks on the Irish Economy." *The Economic and Social Review* 40 (4): 407-34.
- [15] Bilbie, Florin O., Andre Meier, and Gernot J. Mueller. 2008. "What Accounts for the Changes in U.S. Fiscal Policy Transmission." *Journal of Money, Banking, and Credit* 40 (7).
- [16] Blanchard, Olivier J. and Daniel Leigh. 2013. "Growth Forecast Errors and Fiscal Multipliers." *American Economic Review* 103 (3): 117-20.
- [17] Blanchard, Olivier and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117 (4): 1329-68.
- [18] Brown, E. Cary. 1956. "Fiscal Policy in the 'Thirties: A Reappraisal." *American Economic Review* 46 (5): 857-879.
- [19] Burriel, Pablo, Francisco de Castro, Daniel Garrote, Esther Gordo, Joan Paredes, and Javier J. Perez. 2010. "Fiscal Policy Shocks in the Euro Area and the US: An Empirical Assessment." *Fiscal Studies* 31 (2): 251-85.

- [20] 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 No. 877*.
- [21] Carmino, Gerald and Robert P. Inman. 2014. "Macro Fiscal Policy in Economic Unions: States as Agents." *Federal Reserve Bank of Philadelphia Working Paper no. 14-20*.
- [22] Chen, Yvonne. 2008. "Assessing the Effects of Monetary and Fiscal Shocks: The Polish Experience." *International Research Journal of Finance and Economics* 17: 49-56.
- [23] Chodorov-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston. 2012. "Does State Fiscal Relief During Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4 (3): 118-45.
- [24] Cimadomo, Jacopo and Agnes Benassy-Quere. 2012. "Changing Patterns of Fiscal Policy Multipliers in Germany, the UK, and the US." *Journal of Macroeconomics* 34: 845-73.
- [25] Clemens, Jeffrey and Stephen Miran. 2012. "Fiscal Policy Multipliers on Subnational Government Spending." *American Economic Journal: Economic Policy* 4 (2): 46-68.
- [26] Coenen, Gunther, Christopher J. Erceg, Charles Freedman, Davide Furceri, Michael Kumhof, Rene Lalonde, Douglas Laxton, Jesper Linde, Annabelle Mourougane, Dirk Muir, Susanna Mursula, Carlos de Resende, John Roberts, Werner Roeger, Stephen Snudden, Mathias Trabandt, and Jan In't Veld. 2012. "Effects of Fiscal Stimulus in Structural Models." *American Economic Journal: Macroeconomics* 4(1): 22-68.
- [27] Cohen, Lauren, Joshua Coval, and Christopher Malloy. 2011. "Do Powerful Politicians Cause Corporate Downsizing?" *Journal of Political Economy* 119 (6): 1015-60.
- [28] Corbi, Raphael, Elias Papaioannou, and Paolo Surico. 2014. "Federal Transfer Multipliers. Quasi-Experimental Evidence from Brazil." *NBER Working Paper no. 20751*.
- [29] Corsetti, Giancarlo, Andre Meier, and Gernot J. Muller. 2012. "What Determines Government Spending Multipliers?" *Economic Policy* 27 (72): 521-65.
- [30] Corsetti, Giancarlo and Gernot J. Muller. 2006. "Twin Deficits: Squaring Theory, Evidence and Common Sense." *Economic Policy* (October): 597-638.
- [31] De Castro, Francisco. 2006. "The Macroeconomic Effects of Fiscal Policy in Spain." *Applied Economics* 38: 913-24.
- [32] Evans, Michael K. 1966. "Multiplier Analysis of a Post-war Quarterly U.S. Model and a Comparison with Several Other Models." *The Review of Economic Studies* (33) 4: 337-60.
- [33] Evans, Michael K. 1969. "Reconstruction and Estimation of the Balanced Budget Multiplier." *Review of Economics and Statistics* 51 (1): 14-25.
- [34] Fatas, Antonio and Ilian Mihov. 2001. "The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence." *CEPR Discussion Paper No. 2760*.
- [35] Fazzari, Stephen M., James Morley, and Irina Panovska. 2015. "State-Dependent Effects of Fiscal Policy." *Studies in Nonlinear Dynamics & Econometrics* 19(3): 285-315.
- [36] Fishback, Price V. and Valentina Kachanovskaya. 2010. "In Search of the Multiplier for Federal Spending in the States During the New Deal." *NBER Working Paper no. 16561*.
- [37] Fishback, Price and Valentina Kachanovskaya. 2015. "The Multiplier for Federal Spending in the States During the Great Depression." *Journal of Economic History* 75 (1): 125-62.
- [38] Fisher, Jonas D.M. and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *The Economic Journal* 120 (May): 414-36.
- [39] Gali, Jordi, Javier Valles, and J. David Lopez-Salido. 2007. "Understanding the Effects of Government Spending on Consumption." *Journal of the European Economic Association* 5 (1): 227-70.

- [40] Giordano, Raffaele, Sandro Momigliano, Stefano Neri, Roberto Perotti. 2007. "The Effects of Fiscal Policy in Italy: Evidence from a VAR model." *European Journal of Political Economy* 23: 707-33.
- [41] Gordon, Robert J. and Robert Krenn. 2013. "The End of the Great Depression 1939-41: Fiscal Multipliers, Capacity Constraints, and Policy Contributions." *NBER Working Paper No. 16380*.
- [42] Guajardo, Jaime, Daniel Leigh, and Andrea Pescatori. 2014. "Expansionary Austerity: New International Evidence." *Journal of the European Economic Association* 12, no. 4: 949-968.
- [43] Hall, Robert E. 1986. "The Role of Consumption in Economic Fluctuations." In *The American Business Cycle : Continuity and Change*, ed. Roger J Gordon, 237-55. Chicago and London: University of Chicago Press.
- [44] Hooker, Mark A. and Michael M. Knetter. 1997. "The Effects of Military Spending on Economic Activity: Evidence from State Procurement Spending." *Journal of Money, Credit, and Banking* 29 (3): 400-21.
- [45] Ilzetzki, Ethan, Enrique G. Mendoza, Carlos A. Vegh. 2013. "How big (small?) are fiscal multipliers?" *Journal of Monetary Economics* 60: 239-54.
- [46] International Monetary Fund. 2014 "Is It Time for an Infrastructure Push? The Macroeconomic Effects of Public Investment." *IMF World Economic Outlook, October 2014: Legacies, Clouds, and Uncertainties*, ch. 3.
- [47] Jalil, Andrew J. 2012. "Comparing Tax and Spending Multipliers: It's All About Controlling for Monetary Policy." Working Paper.
- [48] Jordà, Oscar and Alan M. Taylor. 2016. "The Time For Austerity: Estimating the Average Treatment Effect of Fiscal Policy." *The Economic Journal*, forthcoming.
- [49] Kim, Soyoung and Nouriel Roubini. 2008. "Twin Deficit or Twin Divergence? Fiscal Policy, Current Account, and Real Exchange Rate in the US." *Journal of International Economics* 74: 362-83.
- [50] Kirchner, Marcus, Jacopo Cimadomo, and Sebastian Hauptmeier. 2010. "Transmission of Government Spending Shocks in the Euro Area: Time Variation and Driving Forces." *Tinbergen Institute Discussion Paper*.
- [51] Kraay, Aart. 2012. "How Large is the Government Spending Multiplier? Evidence from World Bank Lending." *Quarterly Journal of Economics* 127: 829-87.
- [52] Kuttner, Kenneth N. and Adam S. Posen. 2002. "Fiscal Policy Effectiveness in Japan." *Journal of the Japanese and International Economies* 16: 536-58.
- [53] Leduc, Sylvain and Daniel Wilson. 2012. "Highway Grants: Roads to Prosperity?" FRBSF Economic Letter 2012-35.
- [54] Leeper, Eric M., Todd B. Walker, and Shin-Chun S. Yang. 2010. "Government Investment and Fiscal Stimulus." *Journal of Monetary Economics* 57: 1000-12.
- [55] Leigh, Daniel, Pete Devries, Charles Freedman, Jaime Guajardo, Douglas Laxton, and Andrea Pescatori. 2010. "Will It Hurt? Macroeconomic Effects of Fiscal Consolidation." In *World Economic Outlook: Recovery, Risk, and Rebalancing*, 93-124. Washington, DC: International Monetary Fund.
- [56] Liu, Ta-Cheung. 1963. "An Exploratory Quarterly Econometric Model of Effective Demand in the Postwar U.S. Economy." *Econometrica* 31 (3).
- [57] Mittnik, Stefan and Willi Semmler. 2012. "Regime Dependence of the Fiscal Multiplier." *Journal of Economic Behavior and Organization* 83: 502-22.
- [58] Monacelli, Tommaso, Roberto Perotti, and Antonella Trigari. 2010. "Unemployment Fiscal Multipliers." *Journal of Monetary Economics* 57: 531-53.

- [59] Morishima, Michio and Mitsuo Saito. 1964. "A Dynamic Analysis of the American Economy, 1902-1952." *International Economic Review* 5 (2): 125-64.
- [60] Mountford, Andrew and Harald Uhlig. 2009. "What are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24: 960-992.
- [61] Nakamura, Emi and Jon Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from U.S. Regions." *American Economic Review* 104(3): 753-792.
- [62] Nekarda, Christopher J. and Valerie A. Ramey. 2011. "Industry Evidence on the Effects of Government Spending." *American Economic Journal: Macroeconomics* 3 (January): 36-59.
- [63] Nitsch, Thomas O. 1972. "A Further Adjustment in a Test of the Relative Importance of Monetary and Fiscal Actions in Economic Stabilization." *Nebraska Journal of Economic and Business* 11 (1): 11-24.
- [64] Owyang, Michael T., Valerie A. Ramey, and Sarah Zubairy. 2013. "Are Government Spending Multipliers Greater During Periods of Slack? Evidence from 20<sup>th</sup> Century Historical Data." *American Economic Review* 103(2): 129-34.
- [65] Pereira, Manuel C. 2008. "Empirical Evidence on the Stabilizing Role of Fiscal and Monetary Policies in the US." *MPRA Paper No. 19675*.
- [66] Pereira, Manuel C. and Arthur S. Lopes. 2014. "Time-Varying Fiscal Policy in the U.S." *Studies in Nonlinear Dynamics & Econometrics* 18(2): 157-184.
- [67] Perotti, Roberto. 2005. "Estimating the Effects of Fiscal Policy in OECD Countries." *Center for Economic and Policy Research Discussion Paper No. 4842*.
- [68] Perotti, Roberto. 2007. "In Search of the Transmission Mechanism of Fiscal Policy." In *NBER Macroeconomics Annual 2007, Volume 22*. Chicago: University of Chicago Press.
- [69] Roberto, Perotti. 2014. "Defense Spending is Contractionary, Civilian Spending is Expansionary." *NBER Working Paper No. 20179*.
- [70] Ramey, Valerie A. 2011. "Identifying Government Spending Shocks: It's All About the Timing." *Quarterly Journal of Economics* 126 (1).
- [71] Ramey, Valerie A. 2013. "Government Spending and Private Activity." In *Fiscal Policy After the Financial Crisis*, eds. Alberto Alesina and Francesco Giavazzi. Chicago: University of Chicago Press.
- [72] Ramey, Valerie A. and Matthew D. Shapiro. 1998. "Costly capital reallocation and the effects of government spending." *Carnegie-Rochester Conference Series on Public Policy* 48: 145-94.
- [73] Ramey, Valerie A. and Sarah Zubairy. 2014. "Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data." Working Paper.
- [74] Ravn, Morten O., Stephanie Schmitt-Grohe, and Martin Uribe. 2012. "Consumption, Government Spending, and the Real Exchange Rate." *Journal of Monetary Economics* 59(3): 215-234.
- [75] Ravnik, Rafael and Ivan Zilic. 2011. "The Use of SVAR Analysis in Determining the Effects of Fiscal Shocks in Croatia." *Financial Theory and Practice* 35 (1): 25-58.
- [76] Riera-Crichton, Daniel, Carlos A. Végh, and Guillermo Vuletin. 2015. "Procyclical and Countercyclical Fiscal Multipliers: Evidence from OECD Countries." *Journal of International Money and Finance* 52: 15-31.
- [77] Serrato, Juan Carlos Suarez and Phillippe Wingender. 2011. "Estimating Local Fiscal Multipliers." Manuscript.
- [78] Shoag, Daniel. 2013. "Using State Pension Shocks to Estimate Fiscal Multipliers Since the Great Recession." *American Economic Review: Papers & Proceedings* 103 (3): 121-24.
- [79] Smith, Paul E. 1967. "Money Supply and Demand: A Cobweb?" *International Economics Review* 8 (1): 1-12.

- [80] Stevans, Lonnie K. and David N. Sessions. 2010. "Calculating and Interpreting Multipliers in the Presence of Non-Stationary Time Series: The Case of US Federal Infrastructure Spending." *American Journal of Social and Management Sciences* 1 (1): 24-38.
- [81] Suits, Daniel B. 1962. "Forecasting and Analysis with an Econometric Model." *American Economic Review* 52 (1): 104-32.
- [82] Tagkalakis, Athanasios. 2008. "The Effects of Fiscal Policy on Consumption in Recessions and Expansions." *Journal of Public Economics* 92: 1486-1508.
- [83] Turrini, Alessandro, Werner Roeger, and Istvan P. Szekely. 2012. "Banking Crises, Output Loss, and Fiscal Policy." *CESifo Economic Studies* 58 (1): 181-219.
- [84] Wang, Xin and Yi Wen. 2013. "Is Government Spending a Free Lunch? Evidence from China." *Federal Reserve Bank of St. Louis Working Paper Series No. 2013-013A*.
- [85] Wilson, Daniel J. 2012. "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4 (4): 251-82.
- [86] Wren-Lewis, Simon. 2000. "The Limits to Discretionary Fiscal Stabilization Policy." *Oxford Review of Economic Policy* 16 (4): 92-105.
- [87] Zubairy, Sarah. 2014. "On Fiscal Multipliers: Estimates from a Medium Scale DSGE Model." *International Economic Review* 55(1): 169-195.