Unpacking the Multiplier: Making Sense of Recent Assessments of Fiscal Stimulus Policy

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation	Marglin, Stephen A., and Peter M Spiegler. 2013. Unpacking the Multiplier: Making Sense of Recent Assessments of Fiscal Stimulus Policy. Social Research: An International Quarterly 80 (3): 819-854.
Published Version	https://muse.jhu.edu/journals/social_research/ v080/80.3.marglin.pdf
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:25658990
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http:// nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of- use#LAA

Stephen A. Marglin and Peter Spiegler

Unpacking the Multiplier: Making Sense of Recent Assessments of Fiscal Stimulus Policy

POLICYMAKERS ACROSS THE GLOBE RESPONDED DIFFERENTLY TO the Great Recession, some with harsh austerity, others with activist income support and job growth strategies. This diversity offers a good laboratory to assess the relative merits of stimulus and austerity responses. Much of the answer depends on the values of "the multiplier"—the ratio of change in GDP to the resources expended due to the policy (that is, how much increases in government spending or decreases in taxes affect GDP). Ever since Keynes (1936) made the multiplier a cornerstone of the analysis laid out in *The General Theory of Employment, Interest, and Money*, it has been standard shorthand for discussing the impact of exogenous "shocks" of various kinds to the macroeconomy. This is true even for those whose vision of the macroeconomy is very different from that of Keynes. For instance, Real Business Cycle theorists often express their position that government spending will not affect the level of output by arguing that the government spending multiplier has a low or zero value. For better or worse, discourse over fiscal policy seems destined to be undertaken in the vocabulary of the multiplier.

While it is not surprising that assessments of stimulus policy should center around the multiplier, it may well be undesirable. The

government spending multiplier is a nebulous and contingent concept. It is nebulous in the sense that there are many possible values for the multiplier for any single case of stimulus depending, for example, "on the type of government spending, its persistence, and how it is financed" (Ramey 2011, 673). The multiplier is contingent in the sense that its meaning depends upon the ostensible goals of the stimulus policy and the counterfactual path against which its performance is assessed. We cannot answer the question "does this value of the multiplier mean that the stimulus was a success?" without reference to the questions "what was it supposed to accomplish?" and "relative to what are we assessing its success?"

In light of this, it is little wonder that recent multiplier-based assessments of the wisdom of fiscal stimulus have been all over the map—ranging from the highly pessimistic (the 0.64 multiplier of Cogan et al. [2011]) to the highly optimistic (Gordon and Krenn's 2011 value of 1.8).[1] Each of these individual assessments was calculated on a particular set of assumptions, using a particular methodology, at various levels of aggregation, examining a particular time frame. Under such circumstances, it would be much more surprising if there had been general agreement on "the" multiplier's value. Various multipliers are measuring different relationships between stimulus and output, and if we want to understand just what each is actually measuring, how the various concepts are related to each other, and what their significance is for fiscal stimulus policy, we will need to exhume the assumptions upon which they are built and examine the role of these assumptions in their proper interpretation.

In what follows, we shed some light on the meaning of empirical estimates of the multiplier and their proper use. On the general level, we will undertake a critique of the multiplier, explaining the role that various assumptions play in the three leading methodologies for calculating multiplier values. We find that two types of assumptions are crucial: "counterfactual assumptions" that specify the baseline against which the impact of the stimulus is judged and "behavioral assumptions" about the decision-making processes of economic agents. On the specific level, we

apply what we learn from the critique to a sample of recent work claiming that certain aspects of the 2009 American Recovery and Reinvestment Act's (ARRA) stimulus were ineffective—in particular, work by Stanford's John Cogan and John Taylor claiming that ARRA funds funneled through state governments had no effect because states saved rather than spent the funds. We argue that the conclusions of these studies are highly sensitive to counterfactual and behavioral assumptions that are in some cases questionable and in others clearly implausible. We conclude with some general thoughts on the proper use and interpretation of the multiplier in assessing fiscal stimulus programs.

THE MULTIPLIER IN THEORY

The concept of the multiplier as we know it originated with Richard Kahn, a student of Keynes. In a pamphlet coauthored with D. H. Henderson to support the Liberal election campaign in 1929, Keynes had argued that ripple effects from government spending would enhance the impact of the original outlay on the economy, and he assigned Kahn the task of developing a model to quantify these ripple effects. The basic idea is that a new purchase calls forth not only an immediate addition to production but also an immediate increase in income for the producer, and therefore a subsequent increase in his purchases. These purchases in turn represent new income for some other producers, and new spending on their part. In principle the chain continues indefinitely.

The question Kahn set out to answer was how much additional spending and income could be expected from an initial expenditure of one pound. Kahn's insight was that though the number of rounds might be infinite, each round of spending would be smaller because some of the income would "leak" into saving and imports, not to mention taxes. So from £1 of government spending, the workers, contractors, and other direct recipients of income might spend only half a pound on consumption, creating only 50 pence of additional income. If in turn the recipients of this 50p also spend only half, the next round of spending will produce only 25p of new output and income.

In this analysis the crucial determinant of the size of ripple effects is the proportion of new income that individuals spend—in Keynes's vocabulary, their "marginal propensity to consume." That is, the increase in GDP that occurs from each extra dollar of government spending (over and above the direct effect of that dollar on GDP) is positively proportional to and solely dependent on the Marginal Propensity to Consume (see Appendix A).

The world is obviously a more complicated place than the stripped-down expository model of the General Theory, and the multiplier will require four qualifications. First, there is a conceptual difference between the multiplier as it is applied in present-day models and Keynes's original exposition. In Keynes's simple model, the original expenditure—his example used a change in private investment rather than in government spending—stimulates the economy until eventually people's saving (what they don't spend on consumption) just balances the original investment. For expositional simplicity, Keynes focused on the chain of consumption expenditures, but what really matters for the multiplier is the fraction of newly created income that goes to purchase domestically produced goods and services. This enlarges the scope of the multiplier since the expenditure chain includes investment and government spending. So what we are calling the "marginal propensity to consume" actually represents the "marginal propensity to spend on anything reflected in GDP."

A second qualification is that the multiplier formula implicitly assumes an exact equivalence between expenditure and the creation of new goods and services in response.[2] This is likely to be the case if there is considerable slack in the economy but much less likely if the economy is already near to fully utilizing the available resources. In the second case, crowding out may prevent the original stimulus spending from generating output and income on a dollar for dollar basis.[3] Neoclassical and New Keynesian models often reflect this by including a negative relationship between the interest rate and investment. In these models, stimulus financed through borrowing will drive interest rates up, effectively crowding the private sector out of the credit

markets. Alternatively, if private demand is not sufficiently curbed by rising interest rates, the pressure on resources will be reflected in higher prices. In both cases, output would rise by a factor of less than the original multiplier (see Appendix B).

A third qualification introduces two reasons other than crowding out to explain why the numerator of the multiplier might be less than 1, resulting in smaller multipliers. Unlike crowding out, these reasons cannot be assumed to affect the initial and subsequent rounds of spending symmetrically. In Keynes's expository models, the original spending goes to purchases of goods and services that go into capital formation—plant, equipment, railroads, and houses, for instance. The corresponding element of fiscal stimulus policies would be direct purchases by the federal government of newly produced goods and services, but fiscal stimulus often involves using funds in ways other than such direct purchases. Most of the $800 billion dollar stimulus in the 2009 American Recovery and Reinvestment Act, as John Cogan and John Taylor (2012, 89–91) remind us, actually took the form of transfers and tax breaks to individuals and businesses, as well as grants to states to supplement the already massive grants-in-aid that have been part of our fiscal system for the past generation. This would not matter to the calculation of the multiplier if the beneficiaries of federal largesse were themselves to spend all the stimulus money they receive or, in the case of tax breaks, all the money they do not have to pay to the IRS. If this were the case, the numerator of the multiplier formula would continue to be determined by the degree to which expenditure simply crowds out other production. But suppose the direct beneficiaries take a more conservative approach to spending. If none of the original stimulus is spent, the multiplier would be zero.

Both these extremes bend the original logic of the multiplier, according to which the fraction of income spent is neither 0 nor 1. Indeed, the textbook multiplier for taxes and transfers conventionally assumes that, at least as a first approximation, the direct beneficiaries spend a fraction equal to the average for the economy, namely, the marginal propensity to consume. This means that the first-round impact of each

dollar of tax cuts (or transfer increases) will not be one dollar, but one dollar multiplied by the marginal propensity to consume: the numerator is less than one with or without crowding out (see Appendix C).

Crowding out apart, the logic of the generic tax multiplier works only if the spending of direct beneficiaries of the stimulus mirrors the average spending pattern in the economy, which brings us to the fourth qualification. Although we can justify using an economy-wide average for the second, third, and subsequent rounds of spending—since there is no way of tracing out the expenditure of each income recipient in the chain—we can surely do better in measuring the impact on the spending of the direct beneficiaries (that is, in the first round). The direct beneficiaries are responding to specific tax cuts, transfers, and grants, and we normally have information not only about the specifics of tax cuts but about the characteristics of the beneficiaries, including their particular circumstances and constraints. This information allows us to multiply our multiplier by the fraction of tax reductions (transfers and grants) spent by beneficiaries (see Appendix D). We can think of the fraction of tax reductions spent by beneficiaries as a valve controlling the flow of the initial stimulus into the economy. If the fraction is zero then the valve is shut and none of the tax reductions are spent by beneficiaries. The spending never makes it into the economy, so the multiplier has nothing to multiply.

This list of qualifications shows there is great variability in the multiplier, which, in turn, would seem to counsel a nuanced use of multipliers when assessing the prospects of fiscal stimulus. If a particular stimulus program is targeted toward low-income individuals at a time when the economy is in recession, or is directed at large corporations in more prosperous times, we would want to know what the multiplier is for these particular situations. An imagined generic value of "the multiplier" that covers all situations at all times may be a legitimate simplification for introductory textbooks, but is not, or at least ought not to be, the stuff of policymaking.

This point is well understood by practitioners. In evaluating the Obama stimulus, the Congressional Budget Office (CBO), for example,

used a variety of multipliers, and indeed a range of values rather than a single point estimate for each. The ranges varied from 0 to 0.4 for certain corporate tax breaks and 0.1 to 0.6 for tax cuts for high-income people to 0.3 to 1.5 for tax cuts for middle-income folks and 0.4 to 2.1 for transfer payments like food stamps and unemployment compensation. For payments to states to supplement education and Medicaid budgets the CBO multiplier ranged from 0.4 to 2.1 (Congressional Budget Office 2012, 6–7, table 2).

The different ranges of multiplier estimates for different elements of the stimulus raises an obvious question: If the point was to add demand to a weakening private sector and thereby maintain prerecession levels of employment and output, why would stimulus money take the form of tax breaks directed to corporations and high-income individuals? A much greater bang for the buck was available from tax cuts for middle-income people and transfers to the poor and unemployed, not to mention transfers to the states. The answer is that tax breaks for the rich and tax breaks for corporations were never intended to stimulate. One reading suggests that perhaps as much as one-quarter of the total stimulus went toward paying the political price of the stimulus, rather than toward the stimulus itself (Marglin and Spiegler 2013a). Tax breaks for the wealthy represent the price of getting a Congress dominated by special interests to act. On another, more generous reading, these tax breaks were not intended to stimulate spending directly but rather to help private agents get their balance sheets in order after the excesses of the Bush years. "Stimulus," or at least a substantial part of it, was, like the Toxic Asset Relief Program (TARP), really about swapping high-quality federal government debt for the tarnished (if not absolutely toxic) debt of private individuals and businesses, as well as state and local governments. A case can be made that these agents would not be in a position to spend until their own financial houses were in order. This might qualify as indirect stimulus under an elastic definition of the term, but not as stimulus is conventionally defined.

Notwithstanding the obvious advantages of allowing the multiplier to have more than one value, recent arguments against the effi-

cacy of the stimulus treat the multiplier as singular, claiming that the fraction of new spending that displaces existing production is always at or near one, or the fraction of tax reductions spent by beneficiaries is always at or near zero. Robert Barro (2009), for example, focuses on crowding out. If the fraction of new spending that displaces existing production is at one,[4] the multiplier is reduced to zero, regardless of the value of the other parameters. John Cochrane (2009) agrees with Barro and also argues that the fraction of tax reductions spent by beneficiaries is zero on the grounds that any rational agent who receives a tax cut, transfer, or grant will take into account the debt the federal government incurs to finance the stimulus.[5] If she does her arithmetic, she will, according to the "Ricardian Equivalence" theory developed by Barro in the 1970s and 1980s (Barro 1989),[6] put the stimulus money into a bank account to repay her share of the new taxes that will be required to pay off the debt. Of course, the rational agents who do not benefit from the stimulus will still recalculate their spending to take account of their future tax obligations. The result is a tie: according to the theory, any new spending by stimulus recipients will be just canceled out by spending reductions elsewhere in the economy.

The very general prescriptions of these arguments regarding the wisdom of fiscal stimulus, however, are unwarranted. An argument that crowding out will absorb the multiplier is surely relevant to stimulus programs launched in times of high-capacity utilization but not otherwise. High-capacity utilization was the reality in the World War II-era—the time period covered in Barro (2009)—but has little bearing on the Great Recession. By mid-2009, when the ARRA stimulus kicked in, the unemployment rate had climbed to almost 10 percent. There was, accordingly, plenty of spare capacity and an abundance of available labor; crowding out was hardly an issue.[7]

The argument that the fraction of tax reductions spent by beneficiaries is zero due to Ricardian Equivalence is questionable even on the individual level, let alone as a general behavioral assumption for all economic actors. How many of us could actually *do* the calculations implied in Ricardian equivalence? Moreover, if debt-financed stimulus

succeeds in fostering economic growth (or at least preventing it from falling as much as it would have in the absence of stimulus), then tax revenues would rise without a future lump-sum tax or rise in the tax rate. It is telling that Barro (2009) himself, the architect of Ricardian Equivalence, did not see fit to emphasize this line of thought in attacking the stimulus.

Another argument why the multiplier is zero is harder to dismiss. John Taylor, along with John Cogan (Taylor 2011a; Cogan and Taylor 2012), argue that the fraction of tax reductions spent by beneficiaries is zero because beneficiaries of tax rebates and transfers, recognizing the temporary nature of federal largesse, did not treat it as they would a regular source of income but rather as a one-time addition to their assets, to be doled out in little bits over the long-term future. Significantly, Cogan and Taylor attribute this caution not to Ricardian Equivalence but rather to "rational expenditure smoothing."

The basic idea is very familiar to economists: in an optimal spending plan, a rational agent will insulate spending from fluctuations in income by laying aside, investing, or lending surplus (in times of plenty) and borrowing (in times of dearth). This is the kernel of the theory of household spending developed independently by Milton Friedman (1957) with the permanent income hypothesis, and by Franco Modigliani (Modigliani and Brumberg 1954; Ando and Modigliani 1963) with the life-cycle hypothesis. These theories became a central pillar of the counterrevolution against Keynesian economics. Despite an array of refinements to Friedman and Modigliani's original formulations, the essence of the theory remains the expenditure smoothing that rational agents engage in when income fluctuates.

The rational expenditure smoothing assumption plays a dual role in arguments against fiscal stimulus, as it pertains both to potentially observable behavior *and* to unobservable (counterfactual) behavior. In principle, it should be possible to ascertain what proportion of the stimulus funds is actually spent by recipients. But rational expenditure smoothing also implies the counterfactual assumption that in the absence of stimulus, individuals would have maintained their prereces-

sion expenditures by borrowing and dipping into savings. Both of these facets of the assumption are important in the Cogan and Taylor argument against stimulus: not only do individuals spend little of what they receive, but the proper baseline against which to compare this small or zero change is stable spending rather than a drop in spending.

A relatively novel feature of Cogan and Taylor's argument is the idea that the same logic that applies to households also applies to state and local governments. Cogan and Taylor note Edward Gramlich's pioneering work on the effects of federal grants on state budgets. Gramlich is skeptical of the efficacy of trying to stimulate the economy through grants to states, arguing as Cogan and Taylor do, a generation later, that grants end up fortifying state balance sheets (Gramlich 1978; 1979).

Before Gramlich, the terrain of how government spending is determined had been left mainly to students of politics. As early as the 1960s, Aaron Wildavsky argued the position that would later inform Gramlich's work: last year's expenditures are the primary determinant of this year's expenditures. An important difference between Wildavsky and Gramlich, Taylor, and other economists who invoke expenditure smoothing is that Wildavsky claimed no rational basis—on the contrary—for the workings of the budgetary process (Wildavsky 1964; Davis, Dempster, and Wildavsky 1966, 529–547; 1974, 419–452), nor did he apply his arguments to the operation of state and local government. His focus was instead on the process that determined agency budgets within the federal government, an altogether different environment from the states and cities. (For starters, no balanced budget constraints operate at the federal level.)

For the purposes of this paper, however, the question is not merely whether expenditure smoothing is a plausible behavioral assumption. The pertinent questions for us are whether and to what extent the initial recipients of fiscal stimulus actually engage in rational expenditure smoothing, whether institutions (such as state and local governments) behave the same or differently in this regard, and how we would empirically distinguish a rational expenditure smooth-

ing motive from other mechanisms that might generate similar behavior.

As with crowding out and Ricardian Equivalence, there are at least prima facie reasons to be skeptical about rational expenditure smoothing as a general assumption. First, many agents are simply unable to engage in expenditure smoothing—they have little or no savings and equally little access to credit markets. This is the focus of the literature on what are called "liquidity constrained households."

Second, in our view the economist's notion of "rationality" in "rational consumption smoothing" makes untenable demands on decision makers with respect both to their intertemporal utility functions and their intertemporal budget constraints. Most people simply do not know enough about their future needs and wants, much less about their future incomes, for the framework of the standard theory of consumer choice to make sense.[8] Instead, people fall back on habit, rules of thumb, and other perhaps less elegant but more realistic ways of coping than what the economist's ideas of optimal planning dictate (Marglin 2008, 119–122). Moreover, real world rationality may suggest a higher premium on solidarity and sharing than the economist's paradigm of individual choice allows. A poor person embedded in a community may feel that sharing a tax rebate with her less fortunate neighbors, particularly the neighbor faced with eviction if the rent goes unpaid or a blackout if paying the electricity bill is put off, is a higher priority than replenishing her own bank account. She knows that someday it will be her turn to rely on the community (Stack 1975, quoted in Marglin 2008, 23).

THE MULTIPLIER IN USE: ESTIMATION, SIMULATION, AND ASSUMPTIONS

Theory, however, can only take us so far. Ideally, we would like to be able to adjudicate disagreements over the appropriate measure and use of the multiplier empirically by appealing to the data. Unfortunately, there are significant challenges to doing so. The greatest of these is the task of isolating the effect of the stimulus from other contempo-

raneous macroeconomic activity—a standard difficulty of econometric analyses but one that is made particularly acute in the case of estimating the multiplier due to the relative dearth of adequate data. Not only are fiscal stimulus programs relatively rare, they are also nonuniform. This turns a relatively small number of historical stimulus programs into a collection of even smaller pools of different types of programs deployed under different circumstances. Concretely, this means that when we are prospectively assessing the wisdom of a particular stimulus program, we cannot draw on a large sample of similar past episodes as a guide to its likely impact.

The response of economists to this state of affairs has been to use estimation and simulation techniques that lean heavily on theoretical structure and various kinds of assumptions to draw sharp inferences about the multiplier, the lack of data notwithstanding. These techniques have yielded a wealth of estimates of the multiplier, especially during the recent upsurge in interest in fiscal stimulus engendered by the global recession. In a recent review of the literature, Ramey (2011) reports estimates from 18 such studies.

In many cases, however, the sharpness of the inference is purchased at the expense of flexibility and applicability. The estimates are useful only in situations in which the underlying assumptions that generated the estimates hold. Worse, as Taylor emphasizes (2011a, 687), the failure to explicitly acknowledge these assumptions *as* assumptions often leads to a circularity of logic in arguments employing the multiplier estimates to assess fiscal policy: the multipliers are justified in part on the basis of the correctness of the underlying assumptions and the faith in the underlying assumptions is attributed to the multiplier estimates. In the remainder of this section, we will illustrate the importance of these issues by reviewing three leading techniques of multiplier estimation and highlighting the importance of identifying assumptions in applying the estimates to fiscal policy assessment.

Although there is significant variation in multiplier estimation techniques, it is useful to separate them into three broad categories according to the identification strategies they use to isolate the effect

of fiscal stimulus: (1) the sophisticated statistical technique of Vector Autoregression (VAR); (2) the construction of behavioral models of the micro foundations of macroeconomic activity (Dynamic Stochastic General Equilibrium, or DSGE, models); and (3) detailed simulation of the economy via Large-Scale Macroeconomic (LSM) models.

The use of VAR in macroeconomic modeling was first suggested by Christopher Sims (1980) as a way of improving on existing macroeconometric models that estimated aggregate macroeconomic dynamics by combining many separately estimated partial equilibrium models. This approach was inadequate, according to Sims, because the individual restrictions and assumptions used in the separate models were often ad hoc and did not aggregate well into valid restrictions and assumptions for the general model. That is, the models were somewhat awkward patchworks rather than seamless wholes. VAR methods were meant to unify the analysis by simultaneously regressing a vector of all of the variables of interest (such as government spending, tax revenues, and GDP) against a matrix of lagged values of those variables.[9] The promise of VAR techniques was that they would capture the complex structure of interactions between and among the variables of interest—both contemporaneously and through time—to give a statistically robust picture of their joint evolution.

One problem with this approach was that, because all of the variables were simultaneously mutually determining, it was difficult to draw causal stories from the results. The solution to this was to add "structure" to the VAR by imposing restrictions on how the residuals of each of the variables (unexpected movements in them) related to each other. This allowed economists to specify which of the variables was to be interpreted as the first mover and which were responding. The resulting estimates of the coefficients on the lagged variables allow economists to construct "impulse response functions" (IRFs) that trace the change in a given variable over time in response to an initial shock ("impulse") to another variable. When the impulse is a shock to government spending, the IRF of GDP can be used to construct an estimate of the multiplier.

The primary advantage of these structured VAR (SVAR) models is that they have proved superior to alternative macroeconometric techniques in fitting the data. They are now in regular and widespread use in macroeconomics, including in estimation of multipliers, and are often used as benchmarks against which to assess the accuracy of non-VAR models (Smets and Wouters 2007, 595–6).

The strengths of SVAR are, however, a double-edged sword. The statistical and computational complexity of the technique necessitates certain simplifications that diminish their effectiveness as guides to fiscal policy. Two factors in particular are significant. The first is the parsimony required with respect to the number and level of aggregation of the dependent variables. Each new variable that is added to an SVAR entails a large number of new estimation tasks, a number that grows with the number of lags included. As such, SVARs typically include only a few variables. Blanchard and Perotti's (2002) seminal paper, for example, includes only government spending, tax revenues, and GDP. Because of this parsimony, SVARs can provide little guidance to policymakers in the complex and important issue of fiscal stimulus design—for example, how the stimulus should be targeted, and what channels it should be targeted through.

The second limiting factor is what Jonathan Parker (2011) has referred to as the "linearity" of SVAR models. They are linear in the sense that the calculated impact of a fiscal shock on GDP "*by assumption . . . is constrained to be the same independent of the state of the business cycle*" (Parker 2011, 709, emphasis in original). That is, SVARs do not take into account the fluctuation of the business cycle. This assumption of linearity makes SVAR estimates of the multiplier particularly inappropriate for prospective assessment of fiscal stimulus during a recession. It biases the estimate in two ways. First, it would not be able to capture any difference between the impact of government spending on GDP in a recession versus an expansion. To the extent that the multiplier would be greater in the former state (due, for example, to more slack in the economy or binding liquidity constraints) the SVAR estimate would be biased downward. Second, by treating recession and expansion

symmetrically, the estimate implicitly assumes that the proper baseline for evaluating the impact of government spending on GDP is some measure of the average performance of the economy over the course of the business cycle. If, on the other hand, we expect that counterfactually (that is, absent increased government spending) economic performance would be significantly worse in a recession than in a boom, then a given increase in GDP due to government spending would actually represent a higher impact in a recessionary period versus a prosperous one.

Recognition of these potential biases has recently led to attempts to construct SVARs that control for variations in the state of the economy. Auerbach and Gorodnichenko (2012), for example, explicitly incorporate the possibility of switching between recessionary and expansionary regimes into their SVAR and estimate significantly higher government spending multipliers in the recessionary regime. But such innovations do not address the first criticism, the general insensitivity of SVARs to differential impacts of stimulus among different populations of stimulus recipients. For this, we need estimation methods that model the spending behavior(s) of heterogeneous populations. It is to two such methods that we now turn.

Dynamic Stochastic General Equilibrium models are detailed structural models built on micro foundations of macroeconomic activity. A typical DSGE model consists of a system of equations depicting demand and supply relations and the lower level relations that inform demand and supply—for example, price- and wage-setting relations, a consumption function, an arbitrage condition for the value of capital, and a monetary policy reaction function, inter alia (Smets and Wouters 2007). The functional forms of the relations are determined by economic theory. In most DSGE models, the theoretical base is "New Keynesian," which is essentially neoclassical theory augmented with various adjustment frictions (for example, sticky wages and prices). The agents in these models are forward-looking, incorporating expectations into their current decision-making. This is usually reflected by the assumption that they engage in some level of expenditure smoothing along the lines of the permanent income or life cycle hypotheses.

In contrast to SVAR, the high degree of structure imposed in DSGE models facilitates detailed economic interpretation of the resulting parameter estimates. For this reason, DSGE models are widely used by both academic and policy-oriented macroeconomists. But as with SVAR, the strength of DSGE is also a weakness when it comes to applying its multiplier estimates. The intricate structure imposed on the data leads to estimates of the parameters *on the assumption that the economy actually is structured in the manner specified in the model*. This assumption is not tested by the model but is rather a putative fact necessary for interpreting the results. Consequently, as Christopher Sims (2007, 153) argues, we should not understand DSGE models as tools for assessing the likely impact of a particular policy on the actual economy but rather as "storytelling devices" about what would be true of the economic data if it had been generated by the kind of behavior depicted in the model. To the extent that the particular situation we are concerned with understanding strays in important ways from the DSGE model's depiction of the economy, the model will be inherently inappropriate as a guide to policy. For these reasons, Sims concludes that "making forecasts, policy projections, and (especially) welfare evaluations of policies with these models as if their behavioral interpretation were exactly correct is a mistake" (Sims 2007, 153). But this is precisely what one implicitly does when interpreting a DSGE-based multiplier estimate as reflective of the actual multiplier.

It is important to be clear about precisely what kind of mistake this is—since it is a serious problem and one that is distressingly common among recent critiques of fiscal stimulus. Essentially, the mistake lies in treating one's identifying assumptions as though they were hypotheses to be tested by the model. For example, in a DSGE model that imputes rational consumption smoothing behavior to economic agents, we should not interpret parameter estimates that seem to indicate that consumption smoothing is exhibited in the data as evidence that rational consumption smoothing is a general feature of economic agents. Rather, the proper interpretation of such a result is that to the extent that rational consumption smoothing is a general

feature of economic agents and that all agents engage in it uniformly, then the particular parameter values we have estimated measure that effect. What the DSGE model is doing is *calibrating* the extent of rational consumption smoothing, assuming that it takes the assumed form. The assumption itself is not, and cannot be, tested by the process used to estimate the model because that process depends on the assumption.[10] This general point is just as true for SVAR (and of any estimation technique for that matter) as it is for DSGE, but is particularly important in the latter case because DSGE's identifying assumptions include so many restrictions on the structure of economic behavior. If we were to interpret the mere fact of being able to estimate parameter values for these models as evidence of the accuracy of the structure they posit for the economy, we would be taking much more on board than is warranted.[11]

With respect to fiscal stimulus policy, one particularly important restriction of DSGE models is that they depict the behavior of a "representative agent" and so do not allow for the consideration of differential effects of a policy targeted to different groups with different behavior and circumstances. For example, while the assumption of rational expenditure smoothing behavior *may* be broadly applicable, DSGE models obscure the importance of the fact that the ability to forestall spending from windfalls or to borrow in lean times differs among those at different points in the income scale. For the reasons discussed above, it would be wrong to interpret a DSGE model that fits the data well as an inherent refutation of this, but such interpretations are not uncommon. Cogan et al. (2009, 3–4, 15), for example, make precisely this claim in pointing to the good data-fitting performance of an influential DSGE model (Smets and Wouters 2007) as evidence that credit and informational constraints on consumers do not play a significant role in the performance of fiscal stimulus.

The ability of DSGE models to relax the many restrictions imposed by their identifying assumptions is limited by their computational complexity. Typically, parameter estimates are only possible for linearized versions of the full model, and this means that all of the problems of linearity noted by Parker (2011) with respect to SVAR models apply

to DSGE as well. The study of optimal fiscal policy in a linearized DSGE, he writes, "is based on the answer to the question 'can the government raise model-based utility by conditioning government spending linearly on the state of the economy given that its effects are always the same?' and not 'can the government raise output or consumption more by increasing government spending in a recession than a boom and so should it?'" (Parker 2011, 708) The typical representation of monetary policy in DSGE models (a simple reaction function called a "Taylor Rule"), for example, cannot contemplate fundamental shifts in monetary policy regimes. They therefore obscure the importance of constraints on monetary policy in severe recessions—for example, the constraint that nominal interest rates cannot go lower than zero (the "zero interest lower bound"), which can be a binding constraint for central banks especially during long or deep recessions. Several recent studies have suggested that multiplier estimates are significantly higher in such cases (Krugman 1998; Christiano, Eichenbaum and Rebelo 2011).

Large-Scale Macroeconomic models are similar to DSGE models but, as the name suggests, are much larger in scale. A typical LSM model comprises anywhere from hundreds to thousands of equations and accounting identities. The US Federal Reserve system's FRB/US model, which is used to analyze and forecast domestic macroeconomic activity, consists of roughly 300 behavioral equations and accounting identities. It is part of the larger FRB/Global model, which comprises roughly 2,000 equations and identities. The aim of such models is essentially to replicate the economy in as much detail as possible, subject to relevance and to computing constraints. Because of their extremely high computational and maintenance demands, LSMs are generally created and run only by organizations that can dedicate large economic research staffs to the task—for example, central banks and private macroeconomic research firms such as Moody's Analytics, Macroeconomic Advisers, and IHF Global Insight.

The high level of detail in LSM models uniquely equips them to simulate, evaluate, and forecast the effects of (among other things)

fiscal policies in a highly disaggregated manner. In this sense, they are ideal vehicles for estimating more nuanced multipliers than are possible with SVAR or DSGE models and for using these nuanced multiplier estimates to provide guidance regarding the wisdom of fiscal stimulus or austerity policies. This capacity has been used in several recent studies of the potential impact of the American Recovery and Reinvestment Act stimulus. Blinder and Zandi (2010) use the Moody's Analytics LSM model to simulate the effects of the ARRA stimulus on the economy, as well as the effects of several alternative scenarios to form the baseline for their assessment of the impact of ARRA. This exercise generates specific multiplier estimates for a range of different tax cuts and spending increases. The multipliers are higher for stimulative programs aimed at those lower on the income scale—for example, 1.24 for a payroll tax holiday versus 0.37 for making dividend and capital gains tax cuts permanent. Romer and Bernstein (2009) use a disaggregated set of historical multipliers generated by the FRB/US model and that of a private firm to estimate the impact of ARRA by straightforwardly applying the historical multipliers to the appropriate parts of the stimulus program. On this basis, they estimate a government spending multiplier that would reach 1.57 by the eighth quarter of the program.

Despite the gain in nuance, however, LSM models do not entirely escape the problems associated with DSGE models. The fact that LSM models break economic activity down into such small pieces does mean that they are not dependent upon the kind of strict, sweeping assumptions of DSGE models since the more expressions one uses to represent a given aspect of the economy, the less one needs to assume homogeneity. All the same, some identifying assumptions will always be necessary and, to the extent that these are significantly inaccurate, so will be the multiplier estimates. Indeed, this is one way of interpreting the breakdown of the first generation of LSM models during the 1970s and the subsequent "Lucas Critique" (Lucas 1976)[12].

The foregoing discussion highlights the central importance of scrutinizing the assumptions underlying estimates of the multiplier when using these estimates to assess the wisdom of fiscal stimulus

policy. The important questions for economists and policymakers in such cases must be *specific* questions—about the likely impact of various possible stimulus designs on the particular populations at which they are targeted in the specific economic circumstances in which the policy will be implemented. Determination of the usefulness of multiplier estimates, then, must include investigation of the plausibility of their underlying assumptions within these specific circumstances.

REMEDIES

Crucially, effectively investigating the assumptions underlying multiplier estimates will require evidence from outside of the model itself. In this section, we will present an example of two such strategies we employed in two recents papers (Marglin and Spiegler 2013a, 2013b) investigating the assumption that recipients of fiscal stimulus save rather than spend it.[13] Our particular focus is recent work by John Taylor and John Cogan claiming that the portion of the ARRA stimulus that went to state and local governments—accounting for roughly $250 billion of the total $800 billion program—was ineffective because states had engaged in the same kind of rational expenditure smoothing that economic theory ascribes to individuals (Taylor 2011a; 2011b; Cogan and Taylor 2011; 2012).[14] Specifically, they claim that the states saved rather than spent the stimulus funds, and that in the absence of the stimulus states would have maintained their previous levels of expenditures by borrowing and/or drawing down savings.

Their evidence for this was drawn from a time-series analysis of state government data from 1969 to 2010 that separately regressed (1) current purchases of goods and services against lagged purchases, federal receipts without the ARRA stimulus, and the ARRA stimulus itself; and (2) transfer payments against lagged transfers, federal receipts without the ARRA, and the ARRA stimulus itself. In both equations, they found that the lagged variables explained most of the variation over time in state government spending, and that the coefficients on revenues were correspondingly low. The killer result was that the coefficients on the ARRA stimulus indicated that the impact of ARRA

money on overall spending was, statistically, nil. They interpret these results as demonstrating that states engage in rational consumption smoothing because they tend to hold expenditures steady (thus giving the significantly positive coefficients on the lagged variables), and adjust to current windfalls or shortfalls by adjusting net borrowing and lending. Stimulus channeled through the states, then, has no positive impact on the economy because the states do not use it to increase net spending.

We challenged these findings on many grounds. First, we found reason to be skeptical that the high t-values on the coefficients in the Cogan-Taylor regressions and the high R^2's were convincing evidence for their chosen specification. It is a cliché that correlation does not prove causality, but in the specific case of Cogan and Taylor's use of lagged dependent variables, the reliance on correlation is more misleading than usual. Suppose that in fact—a messenger of God told us so—it is the other variables (that is, revenues in the Cogan-Taylor equations) that are driving expenditures. Nonetheless, lagged dependent variables will still show up with high t-values and bias the estimates of the true drivers of expenditures downward, provided that in the correct specification (the one that God's messenger vouched for) the independent variables (revenues) and the error term are both serially correlated (Achen 2001). This does not disprove Cogan and Taylor's interpretation of the data, but it does suggest that their econometric evidence should not be taken as support for their claims.

The problems with the Cogan-Taylor analysis were not merely econometric. A closer look at the structure of interpretation in the Cogan-Taylor analysis provides a good illustration of the dependence of multiplier estimates on untested assumptions that we discussed above. In order for the zero coefficient on ARRA expenditures in the Cogan-Taylor regressions to be taken as evidence of the failure of the ARRA, it must be the case that states would have and could have engaged in expenditure smoothing—that is, borrowing to maintain prerecession expenditure patterns. This corresponds to a counterfactual assumption of steady spending in the absence of stimulus. The entire weight of their

interpretation rests on this assumption. If, on the contrary, we assume that the recession had pushed states into an expenditure *cutting* regime, then the zero coefficient would indicate that the ARRA had indeed been effective in supporting otherwise unsupportable levels of spending. Evaluating the plausibility of this crucial assumption requires learning something about states' likely actions. Cogan and Taylor, however, rest the plausibility of the counterfactual assumption on a behavioral assumption with regard to expenditure smoothing. They support this assumption by appealing to the results of their analysis: the positive coefficient on the lagged variables is evidence that states engage in consumption smoothing. But this logic is circular. If the proper counterfactual assumption is expenditure cutting, then Cogan and Taylor's coefficient estimates actually *refute* the assumption of expenditure smoothing. To evaluate their results we need a way to break out of this circularity and examine the assumptions directly.

We did so in two ways. First, we looked behind the aggregate numbers by comparing the individual states' spending responses to the ARRA stimulus. This removed the counterfactual assumption that the proper baseline against which to compare actual post-ARRA spending is steady spending at pre-ARRA levels. Our cross-sectional analysis sought to determine, instead, the impact of an extra dollar of ARRA funds on a state's spending relative to other states, controlling for the level of financial solvency of the state. Our results showed that the marginal dollar of ARRA funds was associated with an additional $0.66 in spending and $0.34 in decreases in borrowing (again, controlling for financial solvency). This comports with other recent cross-state studies of the effects of ARRA. Chodorow-Reich et al. (2012), for example, estimate that ARRA Medicaid assistance to states produced 38 job-years per $1 million. Wilson (2011) estimates that ARRA created or saved roughly 2 million nonfarm jobs in its first year, and 3.4 million by 2011.[15]

These cross-state studies cast some doubt on the counterfactual assumption underlying Cogan and Taylor's interpretation of their results, but they do so in a relatively indirect way: by demonstrating that an alternative specification that drops that counterfactual assump-

tion produces different results. A more direct approach would be to make the assumption itself the subject of the study, as recommended by Parker (2011). In two recent papers, for example, Parker and his coauthors examined the rational consumption smoothing assumption directly by studying the effects of tax rebates issued during the 2001 and 2008 recessions on consumption, allowing for variation in response along income level and different types of consumption (Johnson, Parker, and Souleles 2006; Parker et al. 2011). They found that the effect was larger than would be expected under the assumption of rational expectations behavior and that it was larger for low-income and low-asset households.

But microeconometric studies are not the only way to glean information about the plausibility of behavioral assumptions and perhaps not even the best way. Despite their considerable advantages, econometric analyses will always be limited in their ability to reveal the true nature of the subjects' lived experience—most notably by the need to convert that experience into data amenable to econometric analysis and by the restricting effects of identifying assumptions. Another valuable source of information, and one that suffers less from these drawbacks, is direct acquaintance with the on-the-ground realities of economic agents.

Of course, direct appeal to agents has traditionally been more within the province of anthropology than economics. Economists generally resist asking agents for information about why they do what they do or what they would have done if the circumstances had been different.[16] Often, there is good reason for this reluctance: there are too many agents, it is hard to get a representative sample, and agents may have trouble reconstructing the circumstances of their decisions well enough to answer, especially when the questions involve a counterfactual.

With respect to the question of whether state governments engaged in rational expenditure smoothing during the Great Recession, however, none of these reasons apply. There are only 50 states, and state budget officers are a highly professional group of men and women. A

priori, then, it seemed sensible to ask these officers what they would have done had there been no ARRA funds to offset lost revenues and increased demands for expenditure that were the twin results of the Great Recession.

We devised a relatively open set of interview/survey questions to pose to the 50 state budget officers, with the goal of ascertaining the extent to which they engage in expenditure smoothing in general and the extent to which, in the particular case of the period during which ARRA funds were administered, they would have been able to maintain their expenditure at the observed levels in the absence of ARRA funding.[17] The results of the interviews/surveys indicated that while the behavioral assumption of an expenditure smoothing motive was valid in some respects, the counterfactual assumption that states would have been able to spend at the observed levels in the absence of ARRA was invalid.[18]

With the exception of the few states that had significant revenues from fossil fuels, the respondents were virtually unanimous in stating that significant expenditure cuts and revenue increases would have been necessary in the absence of ARRA. Michigan's State Office of the Budget, for example, reported that "had no ARRA funding come to Michigan, general fund reductions of approximately 18% would have been required each fiscal year and would have been in addition to measures taken to close a $1.4 billion budget gap for fiscal 2009, and $1.8 billion in general fund reductions enacted for fiscal 2010." Moreover, many of the respondents commented that it was likely that the balance of the adjustment to lower revenues would likely have been weighted heavily toward spending cuts rather than tax or fee increases due to political considerations. These responses cast serious doubt on Cogan and Taylor's interpretation of relatively flat spending data in the wake of the ARRA stimulus as an indication that it was ineffective. If significant cuts in expenditure was the proper counterfactual, as our interviews/surveys suggest, then the correct interpretation of the data is precisely the opposite: that ARRA had the positive effect of avoiding a significant decline in spending.

The sentiment that lower operating expenditures would have been necessary without ARRA was not sensitive to political party—it was voiced equally by those states with Democratic or Republican governors. There was, however, some difference along political lines with respect to the attitude toward the maintenance of spending that was enabled by ARRA. Several officials from Republican states told us that while their states would likely have enacted more spending cuts in the absence of ARRA, this would have been a positive rather than a negative for economic health.[19] We heard this comment both with respect to spending in general, and specifically with respect to Medicaid and education—two areas where ARRA money came with maintenance of effort ("MOE") provisions. In general, the theme of these comments was that ARRA allowed the state government to put off dealing with budgetary problems, some of which were structural and would still have to be dealt with once the ARRA funds dried up. Many of the budget officials commented that they were wary of creating a "fiscal cliff" by using ARRA money to continue to fund programs at levels that would likely be unsustainable post-ARRA. However, despite these differences in political attitudes toward the stimulus, the verdict with respect to the Cogan and Taylor counterfactual assumption was uniformly negative.

Of course, some of the doubt cast on Cogan and Taylor's counterfactual assumption comes simply from formal legal and institutional facts about state government budgeting. Constitutional and statutory provisions prevent all 50 state governments from borrowing to fund operating budget deficits. As such, one might argue that it would have been possible to judge the plausibility of Cogan and Taylor's counterfactual assumption simply by determining whether the savings states had built up prior to 2009 would have been sufficient to fill their operating shortfalls in the absence of ARRA. In fact, we performed this very calculation using public data and found the so-called rainy-day funds woefully inadequate to withstand the shock of the Great Recession (Marglin and Spiegler 2013a, 2013b). But direct interaction was needed to determine whether the actual practice of state budgeting included

strategies for creatively circumventing the limits of rainy-day funds in times of crisis. We learned (1) that some such strategies are available to budget officers, (2) that there is significant variation across states in their nature and usage, and (3) that even taking such strategies into consideration, it still would not have been possible for the vast majority of the responding states to have avoided significant expenditure cuts in the absence of ARRA.

Our study of state budget officers' experience with the ARRA stimulus highlights the importance not only of assessing the assumptions underlying multiplier calculations, but also of matching the choice of investigative methods to the level and type of information necessary to perform such an assessment. Microeconometric and partial-equilibrium macroeconometric studies can elicit some types of information about behavioral responses to stimulus payments, but they are constrained to do so by imposing structure on agents' experience to make it amenable to econometric analysis. This is the well-known "looking under the lamppost" problem of economic methodology (because it is capable only of analyzing problems and information of a certain form, economists are constrained to ignore problems and information that are not of this form). Direct appeal to agents using qualitative methods can help us to shed light on the dark areas beyond the bailiwick of our formal methods. By exploring the meaning of the agents' circumstances and behavior from their own perspective, we allow for the possibility that relevant types of information and conceptions of the problem might be other than we had initially imagined. In light of the high stakes involved in evaluating the possibilities of fiscal stimulus, and the heavy dependence of current multiplier estimation techniques on untested assumptions, these advantages of qualitative, interpretive techniques seem to us to outweigh their logistical costs.

CONCLUSION

Empirical estimates of the multiplier are imperfect tools for assessing the wisdom of fiscal stimulus policy. To solve the enormous identification problem of isolating the impact of government spending in a

dynamic economy, restrictive assumptions must be made that severely limit the applicability of the resulting estimate. The leading identification techniques in the recent literature produce estimates that are either implausibly general—for example, they posit a single multiplier that is insensitive to variation in economic conditions or the composition of the economic agents receiving the stimulus—or that are implausibly specific—they, for instance, are generated by highly structured models that can only produce estimates of what the multiplier would be if their assumptions about economic behavior and dynamics were correct. In the case of DSGE models, the estimates actually suffer from both of these problems.

Assessment of the wisdom of fiscal stimulus policy, if it is to be done responsibly, requires paying attention to the actual (or potential) design features of the policy and the actual characteristics of the economy into which the stimulus is being introduced. If we are to use multipliers as part of our assessment toolkit, then it would be most desirable to prospectively design studies to estimate specific multipliers for specific fiscal stimulus episodes—that is, studies that reflect the characteristics and conditions relevant to the particular fiscal stimulus we want to assess. If we are to use historical estimates of the multiplier responsibly, we must at the very least subject their identifying assumptions to rigorous scrutiny before claiming that the estimates are relevant to the particular fiscal stimulus program we aim to assess.

Our study of recent work by John Taylor and John Cogan reveals the dangers of not doing so. Taylor and Cogan's conclusion that one of the most substantial elements of the ARRA stimulus had been a failure was fundamentally dependent on untested assumptions that were at best questionable and at worst clearly implausible. Nonetheless, Cogan and Taylor claimed that both the finding and the underlying assumptions had been vindicated by the data, and their claim was given the implicit imprimatur of the discipline by its publication in one of the economics profession's elite journals (Taylor 2011a). The stakes surrounding the policy debate over fiscal stimulus are too high to allow such porous standards for assessing multiplier-based arguments.

Subjecting the underlying assumptions of such arguments to rigorous scrutiny would go some distance toward encouraging a more responsible debate and better policy advice.

ACKNOWLEDGEMENTS

APPENDIX A

In mathematical notation, the ultimate effect on output, Y, of an initial impulse—say an increase in government spending, G, would be:

$$\Delta Y = \Delta G + MPC\Delta G + MPC^2\Delta G + MPC^3\Delta G + \ldots = \Delta G\left(\frac{1}{1 - MPC}\right)$$

This sum gives a value for the generic government spending multiplier—the ratio of the change in output to the value of the initial impulse—of $1 / (1 - MPC)$; in symbols, the ratio of the overall increase in output and income to the original stimulus is

$$\Delta Y / \Delta G = 1 / 1MPC$$

APPENDIX B

We can reflect crowding out in the multiplier formula by adjusting the multiplicand. If CO represents the fraction of new spending which displaces existing production and $m = 1 - CO$, the initial spending of $1 now generates only m of new output. Assuming subsequent rounds of spending are subject to the same degree of crowding out, the second round generates m x (m x MPC), the third round m x (m x MPC) x (m x MPC), and so on. The multiplier sum becomes

$$\frac{Y}{G} = m\left(1 + mMPC + m^2MPC^2 + m^3MPC^3 + \ldots\right)$$

and the multiplier, instead of being $1 / (1 - MPC)$, is

$$\text{Spending Multiplier with Crowding Out} = m / 1 - mMPC)$$

Observe that m appears both in the numerator and the denominator if, as we assume, crowding out is assumed to affect every round of spending equally.

APPENDIX C

The multiplier sum becomes

$$\text{Generic Tax Multiplier} = \$MPC (1 + MPC + MPC\ 2 + MPC\ 3 + \cdots) = MPC/(1 - MPC)$$

Here we are assuming no crowding out, so that the fraction of income spent on purchases of goods and services actually leads to an equal increase in new production.

Observe that government spending exactly offset by taxes leads to a multiplier of 1. The formula is

$$\text{Generic Spending Multiplier} - \text{Generic Tax Multiplier} =$$
$$\frac{1}{1 - MPC} - \frac{MPC}{1 - MPC} = 1 - t$$

This is the so-called *balanced-budget multiplier.* With crowding out, the tax multiplier becomes

Generic Tax Multiplier with Crowding Out = $mMPC / (1 - mMPC)$

APPENDIX D

Taking this qualification into account, we have the tax multiplier as

Specific Tax Multiplier with Crowding Out = $mv / (1 - mMPC)$

where v = fraction of tax reductions (transfers, grants) spent by beneficiaries. The parameter v is an additional adjustment to the multiplicand that we can think of as a valve controlling the flow of the initial stimulus into the economy. If $v = 0$, then the valve is shut. The spending never makes it into the economy, and so the multiplier has nothing to multiply.

NOTES

1. The government spending multiplier measures the marginal impact of a dollar of extra government spending on GDP. A value below 1 indicates an impact on GDP less than the initial dollar spent and, therefore, a negative effect of government spending to values of GDP.
2. $\Delta Y / \Delta G = 1 / (1 - MPC)$.
3. "Crowding out" refers to the process by which investment spending by one sector of the economy (in this case, the government) reduces opportunities for other sectors (in this case, private investment).
4. Refer to Appendix B. $m = 1 - CO$, when $CO = 1$, $m = 0$, Spending Multiplier with Crowding Out = $m / (1 - mMPC) = 0$.
5. Refer to Appendix C, $v = 0$.
6. J. Bradford DeLong and Lawrence Summers (2012) argue that deficit spending may be self-financing because a higher level of economic activity may forestall a decline in potential GDP associated with the existence of underutilized capacity.
7. For a contrary view, see Conley and Dupor (2011).

8. Most people, not everybody: the late James Duesenberry once quipped that the life-cycle hypothesis is exactly the theory one would expect from a middle-aged college professor, thus demonstrating that some people's quips are as profound as other people's theories.
9. The matrix would have i rows and j columns, where i is the number of dependent variables and j the number of lags included.
10. This is true not only for empirical models but also for theoretical models. For elaboration of this point with reference to the history and present of economics, see Spiegler (2012). For an analysis of the mischief caused by this issue in recent literature in institutional economics, see Spiegler and Milberg (2009).
11. One could object that it is not simply estimating parameter values that should give us confidence in the identifying assumptions but rather estimating parameters values that give the model a close fit to the data. But we can never be certain what the parameter estimates and data fit are telling us about the identifying assumptions because of the interdependence of the two.
12. The Lucas Critique asserts that the effects of a change in economic policy cannot be predicted by observing historical data.
13. That is, that $v = 0$.
14. The site www.recovery.gov puts the cumulative total of ARRA expenditures at $804 billion, as of February 2013. The $250 billion figure for funds flowing to state governments is our own calculation (Marglin and Spiegler 2013a), based on data from the Bureau of Economic Analysis and www.recovery.gov. It includes supplemental Medicaid assistance and funds going to local governments.
15. Wilson issues a caveat to his findings that supports our general argument in this paper: "It should be emphasized that the stimulus effects estimated in this paper correspond to the effects of one particular stimulus program enacted in a unique economic environment" (Wilson 2011, 31).
16. There are a few notable exceptions. For example, Henderson (1938) and Meade and Andrews' (1938) use of interviews with businessmen to explore the impact of the interest rate in the determination of

investment; and Blinder et al. (1998) and Bewley's (1999) discussions with relevant economic actors to explore the reasons behind the stickiness of prices and wages, respectively.

17. We sent the questions along with a cover letter explaining our research to all of the state budget officers via email, offering them the options of answering the questions in writing or through a telephone interview. In one case (Massachusetts) our interview was conducted in person.

18. Of the 50 state budget directors we contacted, we received written responses from or had phone interviews with 29. Obviously, our aim was to collect information from all of the states, and we made efforts over a five-month period to collect a comprehensive set of responses. Despite these efforts, however, we received no response from 21 states. Nonetheless, we feel that our group of respondents is large and comprehensive enough and similar enough to the nonresponse set in many important demographic aspects to give us some confidence that the responses are not tainted with selection bias. The responding states accounted for 64 percent of US GDP, 61 percent of the population, 64 percent of total state government expenditures, and had an average GDP per capita of $47,603 versus the national figure of $45,457 (Marglin and Spiegler 2013b; all figures from 2009).

19. This sentiment was expressed to us by state budget officers whose current administration is Republican—in particular, those from Ohio, Wyoming, and Kansas—or whose state was under a Republican administration during the years in question—in particular, Minnesota.

REFERENCES

Achen, C. 2001. "Why Lagged Dependent Variables Can Suppress the Explanatory Power of Other Independent Variables." Unpublished working paper. Department of Political Science and Institute for Social Research, University of Michigan.

Ando, A. K. and F. Modigliani. 1963. "The 'Life Cycle' Hypothesis of

Saving: Aggregate Implications and Test." *American Economic Review* 53 (1): 55–86.

Auerbach, A. J. and Y. Gorodnichenko. 2010. "Measuring the Output Responses to Fiscal Policy." Working Paper 16311, National Bureau of Economic Research.

Barro, R. J. 1989. "The Ricardian Approach to Budget Deficits." *Journal of Economic Perspectives* 3 (2): 37–54.

———. 2009. "Voodoo Multipliers." *Economists' Voice* 6 (2).

Bewley, T. F. 1999. *Why Wages Don't Fall During a Recession.* Cambridge: Harvard University Press.

Blanchard, O. and R. Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *The Quarterly Journal of Economics* 117 (4): 1329–1368.

Blinder, A. S., E. R. D. Canetti, D. E. Lebow, and J. B. Rudd.. 1998. *Asking About Prices: A New Approach to Understanding Price Stickiness.* New York: Russell Sage Foundation.

Blinder, A. S. and M. Zandi. 2010. "How the Great Recession Was Brought to an End." http://www.economy.com/mark-zandi/documents/end-of-great-recession.pdf.

Chodorow-Reich, G., L. Feiveson, Z. Liscow, and W. G. Woolston. 2012. "Does State Fiscal Relief During Recessions Increase Employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy* 4 (3): 118–145.

Christiano, L. J., M. Eichenbaum, and S. Rebelo. 2011. "When Is the Government Spending Multiplier Large?" *Journal of Political Economy* 119 (1): 78–121.

Cochrane, J. H. 2009. "Fiscal Stimulus, Fiscal Inflation, or Fiscal Fallacies?" Accessed June 24 2012. faculty.chicagobooth.edu/john.cochrane/research/papers/fiscal2.htm.

Cogan, J. F., T. Cwik, J. B. Taylor, and V. Wieland. 2009. "New Keynesian Versus Old Keynesian Government Spending Multipliers." Working Paper 14782, National Bureau of Economic Research.

Cogan, J .F. and J. B. Taylor. 2011. "Stimulus Has Been a Washington Job Killer." *Wall Street Journal*, October 4: A21.

————. 2012. "What the Government Purchases Multiplier Actually Multiplied in the 2009 Stimulus Package." In *Government Policies and the Delayed Economic Recovery*, edited by Lee. E. Ohanion, John B. Taylor, and Ian J. Wright. Stanford: Hoover Institution Press.

Congressional Budget Office. 2012. Estimated Impact of the American Recovery and Reinvestment Act on Employment and Economic Output from October 2011 Through December 2012: A CBO Study. CBO Publication No. 4435. Washington, D.C.: U.S. Government Printing Office. http://www.cbo.gov/sites/default/files/cbofiles/attachments/02-22-ARRA.pdf.

Conley, T. and B. Dupor. 2011. "The American Recovery and Reinvestment Act: Public Sector Jobs Saved, Private Sector Jobs Forestalled." Accessed November 23, 2011. http://web.econ.ohio-state.edu/dupor/arra10_may11.pdf.

Davis, O. A., M. A. H. Dempster, and A. Wildavsky. 1966. "A Theory of the Budgetary Process." *The American Political Science Review* 60 (3): 529–547.

————. 1974. "Towards a Predictive Theory of Government Expenditure: US Domestic Appropriations." *British Journal of Political Science* 4 (4): 419–452.

DeLong, J. B. and L. Summers. 2012. "Fiscal Policy in a Depressed Economy." *Brookings Papers on Economic Activity* 2012 (1): 233–297.

Friedman, M. 1957. *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Gordon, R. J. and R. Krenn. 2010. "The End of the Great Depression, 1939–41: Policy Contributions and Fiscal Multipliers." Working Paper 16380, National Bureau of Economic Research.

Gramlich, E. H. 1978. "State and Local Budgets the Day after It Rained: Why Is the Surplus So High?" *Brookings Papers on Economic Activity* 1978 (1): 191–216.

————. 1979. "Macro Policy Responses to Price Shocks." *Brookings Papers on Economic Activity* 1979 (1): 125–166.

Henderson, H. D. 1938. "The Significance of the Rate of Interest." *Oxford Economic Papers* 1: 1–13.

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

Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*. London: Macmillan.

Keynes, J. M. and H. D. Henderson. 1929. *Can Lloyd George Do It? An Examination of the Liberal Pledge*. London: The Nation and Athenaeum.

Krugman, P. R. 1998. "It's Baaack: Japan's Slump and the Return of the Liquidity Trap." *Brookings Papers on Economic Activity* 2: 137–187.

Lucas, Jr., R. E. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie Rochester Conference Series on Public Policy* 1: 19–46.

Marglin, S. A. 2008. *The Dismal Science: How Thinking Like an Economist Undermines Community*. Cambridge: Harvard University Press.

Marglin, S. A. and P. Spiegler. 2013a. "Did the States Pocket the Obama-Stimulus Money? (I): Limits of Time-Series Analysis." Working Paper. Department of Economics, Harvard University.

———. 2013b. "Did the States Pocket the Obama-Stimulus Money? (II): Lessons of Cross-Section Regression and Interviews with State Officials." Working Paper. Department of Economics, Harvard University.

Meade, J. E. and P. W. S. Andrews. 1938. "Summary of Replies to Questions on Effects of Interest Rates." *Oxford Economic Papers* 1: 14–31.

Modigliani, F. and R. H. Brumberg. 1954. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In *Post-Keynesian Economics*, edited by Kenneth K. Kurihara, 388–436. New Brunswick, NJ: Rutgers University Press.

Parker, J. A. 2011. "On Measuring the Effects of Fiscal Policy in Recessions." *Journal of Economic Literature* 49 (3): 703–718.

Parker, J. A., N. S. Souleles, D. S. Johnson and R. McClelland. 2011. "Consumer Spending and the Economic Stimulus Payments of 2008." Working Paper 16684, National Bureau of Economic Research.

Ramey, V. 2011. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49 (3): 673–685.

Romer, C. and J. Bernstein. 2009. "The Job Impact of the American Recovery and Reinvestment Plan." http://otrans.3cdn.net/45593e8ecbd339d074_l3m6bt1te.pdf.

Sims, C. A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (1): 1–48.

——. 2007. "On the Fit of New Keynesian Models: Comment." *Journal of Business and Economic Statistics* 25 (2): 152–154.

Smets, F. and R. Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *The American Economic Review* 97 (3): 586–606.

Spiegler, P. 2012. "The Unbearable Lightness of the Economics-Made-Fun Genre." *Journal of Economic Methodology* 19 (3): 283–301.

Spiegler, P. and W. Milberg. 2009. "The Taming of Institutions in Economics: The Rise of the 'New, New Institutionalism'." *Journal of Institutional Economics* 5 (3): 289–313.

Stack, C. 1975. *All Our Kin: Strategies for Survival in a Black Community.* New York: Harper and Row.

Taylor, J. B. 2011a. "An Empirical Analysis of the Revival of Fiscal Activism in the 2000s." *Journal of Economic Literature* 49 (3): 686–702.

——. 2011b. "The 2009 Stimulus Package: Two Years Later." Testimony before the Committee on Oversight and Government Reform, Subcommittee on Regulatory Affairs, Stimulus Oversight and Government Spending, U.S. House of Representatives. February 16. Accessed June 24, 2012. http://media.hoover.org/sites/default/files/ documents/2009-Stimulus-two-years-later.pdf.

Wildavsky, A. 1964. *The Politics of the Budgetary Process.* Boston: Little Brown.

Wilson, D. J. 2011. "Fiscal Spending Jobs Multipliers: Evidence from the 2009 American Recovery and Reinvestment Act." Working Paper 2010-17, Federal Reserve Bank of San Francisco. http://www.frbsf.org/publications/ economics/papers/2010/wp10-17bk.pdf.