Comment

Ricardo Reis, *Princeton University and NBER*

How should economists compare the predictions of a model with the data? A currently popular answer to this perennial question is to plot the impulse response functions of some variables to shocks and compare the responses predicted by the model to those estimated in the data. This approach is simple, intuitive, and even fairly comprehensive, since impulse responses contain a great deal of information. For a linear (or linearized) model with constant variances, the impulse response functions summarize all of the model's dynamics, and for covariance-stationary data, they capture all of the second-order properties of the data. Impulse responses have allowed economists to move from focusing solely on variances and covariances into assessing finer features like persistence, sluggishness, hump shapes, and lead-lag relations.

In practice, one difficulty with this methodology is how to estimate the empirical impulse responses. In the study of monetary policy, research has used vector autoregressions (VARs) and this is the recent growing approach in the study of fiscal policy. Perotti gives a thorough and insightful survey of this work, focusing on the impulse responses of output, hours, consumption, and real wages to government spending shocks. These have led to a debate and a challenge.

The debate is between Ramey and Shapiro on one side and Blanchard and Perotti on the other. All agree that output and hours rise following an exogenous expansion in government spending, but while Ramey and Shapiro find that consumption and real wages fall, Blanchard and Perotti find that they rise. Because these opposite results come from different empirical strategies to identify shocks to government spending—one is narrative and uses war buildups as exogenous dates, and the other is structural and assumes government spending responds with a lag to other shocks—this has led to a more general debate on the relative merits of these two methods. Furthermore, Ramey and Shapiro's results are used to support the neoclassical model, while Blanchard and Perotti's are used to support the Keynesian model, so they become involved in the more general debate of what is the best model of economic fluctuations.

The challenge is that if Blanchard and Perotti are right, it is hard—for two reasons—to understand the rise in consumption following the increase in government spending. First, since more government consumption uses resources and lowers private wealth, any model with a significant role for the permanent-income hypothesis will predict a fall in consumption. Second, since with standard parameters the marginal rate of substitution between leisure and consumption rises significantly with the increase in hours but the wage only slightly changes, so for households to be on their labor supply and the two to be equal, consumption must fall. One answer to this challenge is, of course, the old-fashioned IS-LM model, since it violates the permanent-income hypothesis and has hours determined by labor demand, not supply. But more modern models, whether neoclassical or neo-Keynesian, whether real or monetary, fail the challenge. One exception is Galí, López-Salido, and Vallés's (2007) "truly Keynesian" model, where there are not only pricing frictions but also a large group of Keynesian hand-to-mouth consumers (who consume more with the rise in income counteracting the wealth effect) and Keynesian labor markets, where unions set hours and wages (so these are determined by labor demand, not supply).

In this comment, I discuss the use of VAR evidence to test models via impulse responses. There has been an intense debate on the merits and flaws of VARs at estimating impulse responses.[1] Here, I am actually going to assume that Perotti's estimates are exactly right. Instead, I will focus on the use of these estimates to distinguish between models.

1 An Aside into Monetary Policy and Anticipated Policy

Before I start, it is worthwhile taking a short detour into the literature on monetary policy to make an observation inspired by Cochrane (1998). Imagine that three researchers estimated the response of output and a policy variable (say an interest rate) to an exogenous contraction in policy. All three found the same output response, in the left panel of figure 3C1.1, but each found a different response of the policy variable, in the right panel of the figure. Would they reach the same conclusion if they were interested in testing a theory of output fluctuations?

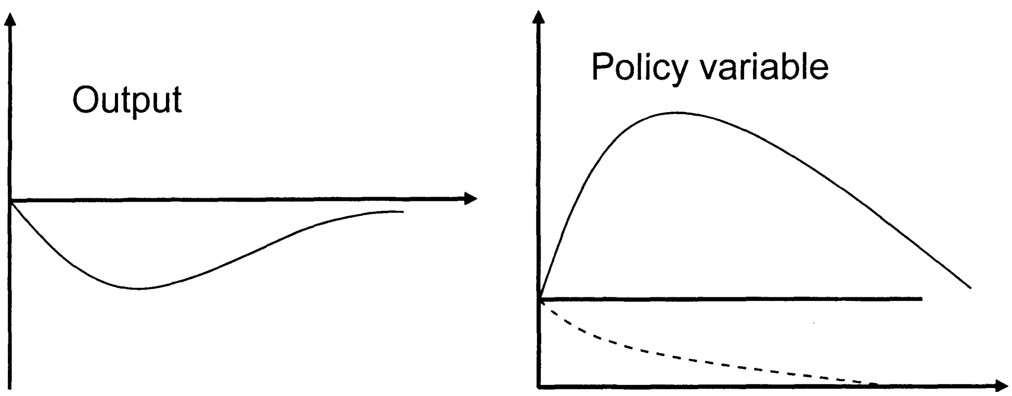

Figure 3C1.1
Fictional Impulse Responses to Policy Shock

If that theory stated that only unanticipated policy matters, as in the classical models of Lucas and Barro, the answer is yes. All three estimated the same instantaneous impact on the policy variable, and that is all that matters for output. The path of policy afterward is anticipated so it is neutral, whether it goes up, down, or stays the same.

If, however, they were examining a modern sticky-price model, the answer is no. In this model, the anticipated policy path after the shock affects by how much adjusting firms change their prices, which in turn affects by how much output falls. Each of the responses of the policy variable on the right side of the figure would lead to a different response of output, so only one (if any) could be consistent with the output response in the left side. In modern models of nominal rigidities, *policy rules matter,* and the response of policy variables to policy shocks provides information on these policy rules.

In the study of fiscal policy, anticipated policy matters even more. Most fiscal policy changes are announced a few quarters in advance and they tend to persist, so fiscal policy is quite predictable. Moreover, changes in government spending typically come with future changes in fiscal policy to balance the budget (and intense debates on the best way to do it). And last, in models with intertemporal substitution, future fiscal policy affects relative trade-offs and therefore behavior in the present.

2 A Neoclassical Model of Fiscal Policy

Consider a simple neoclassical model of fiscal policy and the economy. Households maximize:

$$E_t \sum_{s=0}^{\infty} \beta^s \left(\frac{C_t^{1-\theta} - 1}{1 - \theta} - \frac{\kappa N_t^{1+\psi}}{1 + \psi} \right)$$

$$s.t.: (1 + \tau_t^C)C_t + K_{t+1} = R_{t+1}K_t + W_t N_t + T_t,$$

where C_t is consumption and τ_t^C a consumption tax, N_t hours worked and W_t their after-tax wage, K_t the capital stock and R_t the after-tax return on renting it, and T_t are lump-sum transfers from the government. Firms produce private output to maximize profits:

$$\max_{N,K} \left\{ K_t^{\alpha} N_t^{1-\alpha} - \frac{W_t N_t}{1 - \tau_t^N} - \frac{[R_t - (1 - \delta)(1 - \tau_t^K)]K_t}{1 - \tau_t^K} \right\}$$

where τ_t^N is the tax rate on labor and τ_t^K the tax rate on capital (with no depreciation exemption). Finally, the economy's resource constraint and total output Y_t are:

$$K_{t+1} = (1 - \delta)K_t + Y_t - C_t - G_t$$
$$Y_t = K_t^{\alpha} N_t^{1-\alpha} + \sigma G_t$$

Total government spending is G_t and a fraction σ of it is used in the public sector to generate output, while the remaining $1 - \sigma$ is wasted or provides welfare through some additive extra term in the utility function.[2] The government chooses $(G_t, \tau_t^C, \tau_t^N, \tau_t^K)$, and T_t ensures a balanced budget every period.

This model has a steady state where all variables are constant. It is described by four nonlinear equations relating the four endogenous variables that Perotti wants to focus on, (Y, N, C, W), to the four policy variables, $(G, \tau^C, \tau^N, \tau^K)$. My only assumption on the parameters is that the steady-state endogenous variables are positive. Evaluating the Jacobian determinant of this system at the point where all policy variables are zero.[3]

Proposition 1: *There is (locally) a one-to-one relation between (Y, N, C, W) and $(G, \tau^C, \tau^N, \tau^K)$.*

Therefore, given an appropriate choice of fiscal policy, the neoclassical model can generate *any* steady state that you want. This result is not surprising: observations of average output, hours, consumption, and wages (properly scaled with growth) convey no information on the validity of the neoclassical model.

To study the predicted response to government spending shocks, one must specify the dynamics of the shocks and the fiscal policy rules. I assume the shock follows an AR(1), $a_t = \rho a_{t-1} + \varepsilon_t$. Letting small letters denote the log of the respective capital letter relative to its steady state, the fiscal policy rules are:

$$g_t = \gamma^G c_t + (1 + \lambda^G) a_t,$$

$$\tau_t^C = \gamma^C c_t + \lambda^C g_t,$$

$$\tau_t^N = \gamma^N c_t + \lambda^N g_t,$$

$$\tau_t^K = \gamma^K c_t + \lambda^K g_t.$$

Total government spending responds to a 1 percent fiscal shock by $(1 + \lambda^G)$ percent, and it is cyclical, adjusting to the level of consumption. Tax rates are also cyclical and respond to movements in government spending. These fiscal policy rules may not be optimal or realistic for developed economies, but they are plausible and roughly capture the cyclicality of fiscal policy and the interaction between taxes and spending.[4] There are eight policy-rule parameters: $\pi = (\gamma^G, \gamma^C, \gamma^N, \gamma^K, \lambda^G, \lambda^C, \lambda^N, \lambda^K)$.

The log-linear approximate solution of the model implies an ARMA(2,1) structure for the impulse response to an ε_t shock:

$$[1 - \eta(\pi)L](1 - \rho L)x_t = \mu_x(\pi)\varepsilon_t + \nu_x(\pi)\varepsilon_{t-1}$$

where x_t is either y_t, n_t, c_t, or w_t. The autoregressive coefficients are common to all variables, so differences in dynamics depend on the eight moving-average coefficients $\phi(\pi) = (\mu_y, \nu_y, \mu_n, \nu_n, \mu_c, \nu_c, \mu_w, \nu_w)$, which are functions of the policy parameters.

The neoclassical model's predictions for the variables for interest are fully described by $\phi(\pi)$. If Perotti's estimates fit this ARMA(2,1) structure, then he has effectively estimated $\hat{\phi}$. Asking if the neoclassical model fits the data then amounts to asking whether $\hat{\phi}$ is close to $\phi(\pi)$. Roberto finds that $\hat{\mu}_c > 0$ and $\hat{\mu}_w > 0$ and argues that the neoclassical model predicts the opposite signs, so he concludes against it. However, in the neighborhood of the point where all the elements of π of zero, and for conventional parameter values:[5]

Proposition 2: *There is (locally) a one-to-one relation between π and $\phi(\pi)$.*

That is, *whatever* Perotti's estimated impulse responses of output, hours, consumption, and wages, they are consistent with a neoclassical model with an appropriate choice of policy rules. Perotti's conclusion comes from arbitrarily assuming that all the elements of π are zero. But, with freedom to pick the policy-rule parameters in π, the result on steady states applies also to the model's dynamics. No set of impulse responses could ever reject the model.

It is important to not overstate this result. This is not a claim that anything goes in the neoclassical model, nor is it necessarily specific to the neoclassical versus other dynamic models. The point is instead that looking only at a few impulse responses and having a lot of freedom to pick policy rules gives so much freedom that it leads to no predictions. This problem is familiar to empirical VARs, but here is turned on its head to apply to theoretical models: identification.

3 Identification in the Neoclassical Model

In principle, identification in a theoretical model can follow the same strategies used in empirical estimation. For instance, it is popular in the literature on VARs to impose timing restrictions. These have a direct counterpart in the model. To see how they work, note that the neoclassical model has two static optimality conditions, one from the household's intratemporal allocation of labor and consumption, and the other from labor demand by firms:

$$\kappa N_t^{\psi} C_t^{\theta} = W_t/(1 \; \tau_t^C),$$

$$W_t = (1 - \tau_t^N) Y_t / N_t.$$

Now, imagine imposing the restrictions that the tax rates on consumption and labor income adjust only with a one-quarter delay to changes in spending. Then, these two conditions will pin down the impact response of two of (Y_t, N_t, C_t, W_t) as a function of the other two, independently of the policy-rule parameters. Proposition 2 will no longer hold, and the model has testable predictions on the impact response to spending shocks.

Another approach is to use institutional restrictions, using the details of how taxes are set in a country to learn about some of the policy-rule parameters directly (Blanchard and Perotti 2002). In principle, one could impose exactly the same identifying restrictions on both the VAR and the model, solving both the empirical and theoretical identification problems in a coherent way.

I would like to propose a third approach to identification that uses the impulse responses of policy variables to policy shocks. These responses trace out the policy dynamics. The researcher can use them to pin down the policy-rule parameters, tying his or her hands before looking at the impulse responses of the nonpolicy variables. In this model, this would amount to using the estimated impulse responses of $(g_t, \tau_t^C, \tau_t^N, \tau_t^K)$ to pin down the policy-rule parameters. The resulting $\hat{\pi}$ can then be fed into $\phi(\hat{\pi})$ and compared with the empirical estimates $\hat{\phi}$.

This strategy accomplishes the coherence in identification between estimates and model, because the estimated impulse responses of the policy variables respect the empirical identifying assumptions by construction. When it is hard to map the empirical identifying restrictions to their theoretical counterparts, this procedure accomplishes it directly. Moreover, when the empirical identifying restrictions are not sufficient to identify the model, the policy-variables impulse responses include new information from the data to achieve identification.

To see this approach in action, I pursue an example using Perotti's baseline SVAR estimates with U.S. data from 1947. Because there are only two policy variables in his baseline VAR—government spending and an income tax—I consider a simpler version of the above neoclassical model where there is only an income tax (so $\tau_t^C = 0$ and $\tau_t^N = \tau_t^K = \tau_t$) and consider only the impulse responses of output and consumption. I solve the model for the theoretical impulse responses of g_t and τ_t, which follow the ARMA(2,1) structure above with four moving-average parameters. I pin down the four policy-rule parameters to match as closely as possible the first 16 elements of the empirical impulse responses of g_t and τ_t. Figure 3C1.2 shows the reasonably good match.

Using these policy-rule parameters, I then solve for the theoretical impulse responses of y_t and c_t and compare them to their empirical counterparts in figure 3C1.3. There are three results to note. First, after an expansion in spending, consumption rises on impact. Contrary to Perotti's claim, rising consumption is consistent with the neoclassical model. The reason is that in Perotti's estimates in figure 3C1.2, when spending rises, taxes rise and are expected to fall in the future. Households therefore realize it is relatively less rewarding to work today rather than in the future and so cut hours. Since consumption and leisure are complements, this pushes consumption up.

The second thing to note is that output also falls on impact. This example illustrates the perils of not taking into account the identification of the model. Perotti contrasted his estimates with the predictions of falling consumption and rising output coming from a neoclassical

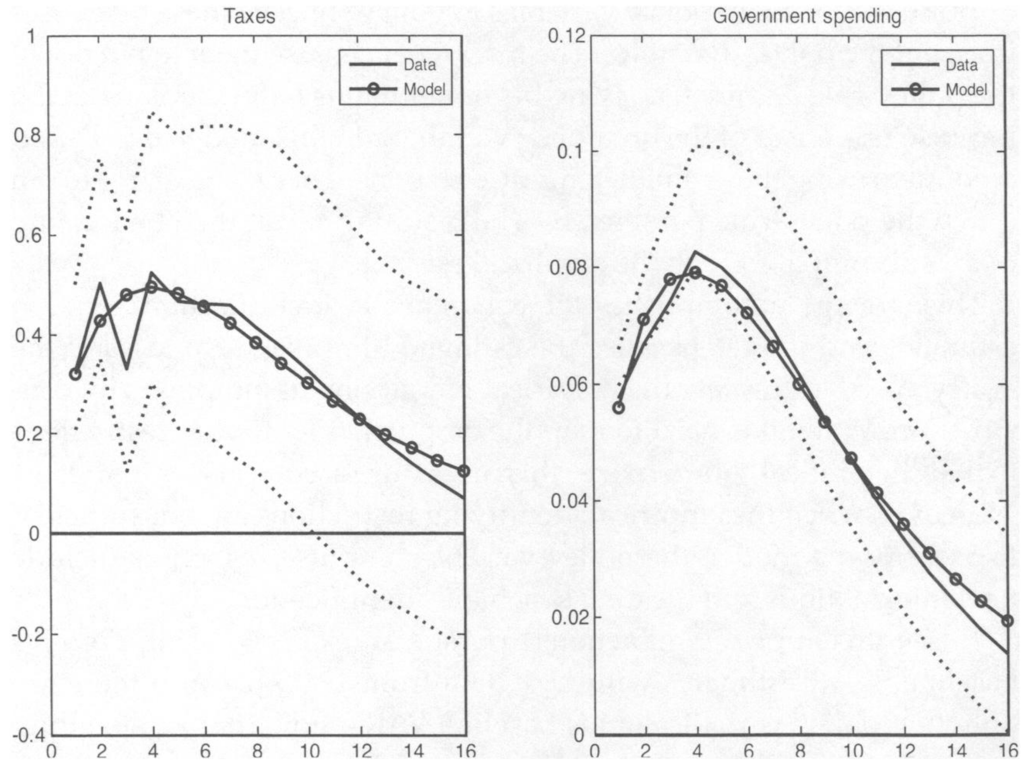

Figure 3C1.2
Impulse responses of taxes and government spending

model where all the policy-rule parameters are equal to zero. In fact, given the policy rules for government spending and income taxes that he estimated, the neoclassical model predicts the opposite—a fall in output and a rise in consumption on impact.

The third result is that the neoclassical model is at odds with the facts. While consumption rises on impact in both data and theory, it stays positive in the former but falls to negative in the latter. And the output response is positive in the data but negative in the theory at all horizons. In general, the theory predictions are quite far from the empirical confidence bands.

4 Conclusion

Perotti has performed a tour de force on the difficult and important issue of estimating and identifying empirical impulse responses to government spending shocks. He used these estimates, in part, to test models, and this comment focuses on this application.

I have tried to make two points that apply more generally than to his

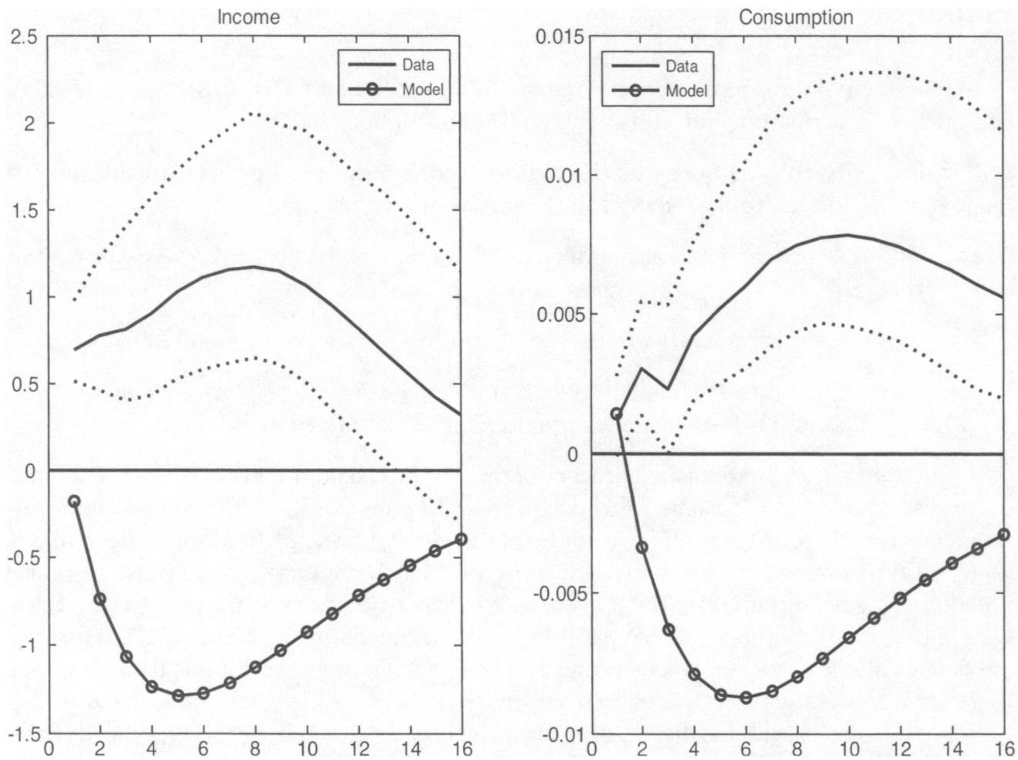

Figure 3C1.3
Impulse responses of income and consumption

paper. The first is well known: policy rules and anticipated policy matter for the dynamics of intertemporal models. The second is perhaps less appreciated: theoretical models can suffer from identification problems that are as serious as those in empirical estimates. The theorist has many degrees of freedom in building his or her model, and some of the most important are the most difficult to pin down, the policy rules.

To be constructive, I propose an approach to identify the theoretical model. It uses the empirical impulse responses of the policy variables to the policy shocks as a summary of both the data and the VAR's identification conditions to identify the policy rules in the model. Then, it compares the theoretical impulse responses for the nonpolicy variables with their empirical counterparts.[6] When I applied this method to compare Perotti's empirical estimates with those of a neoclassical model, I agreed with him that they seem inconsistent, but for very different reasons.

The typical debate on structural VARs focuses on how one can use information from models to help estimate and identify VARs. But, sometimes, the reverse can also be true: one can use information from VARs to help formulate and identify models.

Endnotes

1. In the Macro Annual conference alone, see Galí and Rabanal (2005), McGrattan (2005), Christiano, Eichenbaum, and Vigfussonl (2007), and Kehoe (2007).

2. For simplicity, this assumes that the public sector's output is a perfect substitute with the private sector's output, so there is only one consumption good.

3. All results are proven in an appendix available at my web site: http://www.princeton.edu/~rreis.

4. For a careful empirical study of this interaction, see Romer and Romer (2007).

5. The parameter values are $\beta = 0.99$, $\theta = 1$, $\psi = 4$, $\alpha = 0.34$, $\delta = 0.025$, $\sigma G/Y = 0.12$, $G/Y = 0.21$, $\tau^K = 0.54$, and $\rho = 0.8$. See the appendix for explanations.

6. Aside from the well-established practice of picking model parameters to fit estimated impulse responses (e.g., Christiano, Eichenbaum, and Evans 2005), there are two closer antecedents to this approach. Both also abide by the general principle that the policymaker's policy-rule parameters in the model are chosen to match the empirical impulse response function of the policy variables, but they impose stricter restrictions on the policy rules. Edelberg, Eichenbaum, and Fisher (1999) and Burnside, Eichenbaum, and Fisher (2004) assume the policy rules for government spending and taxes are a moving average of the exogenous fiscal shocks, without any feedback from endogenous variables, and pick the moving-average parameters to match their VAR empirical estimates. Rotemberg and Woodford (1996) make timing assumptions on monetary policy that ensure that the policy rule parameters can be identified from the VAR estimates without having to specify the rest of the model.

References

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113 (1): 1–45.

Christiano, Lawrence J., Martin Eichenbaum, and Robert Vigfusson. 2007. Assessing structural VARs. In *NBER Macroeconomics Annual 2006*, ed., D. Acemoglu, K. Rogoff, and M. Woodford, 1–72. Cambridge: MIT Press.

Cochrane, John. 1998. What do the VARs mean? Measuring the output effects of monetary policy. *Journal of Monetary Economics* 41:277–300.

Galí, Jordi, and Pau Rabanal. 2005. Technology shocks and aggregate fluctuations: How well does the RBC model fit postwar U.S. data? In *NBER Macroeconomics Annual 2004*,ed. M. Gertler and K. Rogoff, 225–88. Cambridge: MIT Press.

Kehoe, Patrick J. 2007. Comment. In *NBER Macroeconomics Annual 2006*, ed. D. Acemoglu, K. Rogoff, and M. Woodford, 73–96. Cambridge: MIT Press.

McGrattan, Ellen. 2005. Comment. In *NBER Macroeconomics Annual 2004*, ed. M. Gertler and K. Rogoff, 289–308.

Romer, Christina D., and David H. Romer. 2007. Starve the beast or explode the deficit? The effects of tax cuts on government spending. Unpublished working paper, UC Berkeley.

Rotemberg, Julio J., and Michael Woodford. 1997. An optimization-based econometric framework for the evaluation of monetary policy. In *NBER Macroeconomics Annual 1997*, ed. B. Bernanke and J. Rotemberg, 297–346.