Comments and Discussion

COMMENT BY

RICARDO REIS[1] Progress in the study of short-run economic fluctuations seems to come in three stages. First, macroeconomists become excited by the arrival of a new theoretical approach, a new set of principles to organize knowledge, or some new modeling tools. Second come the refiners, who explore how to apply the idea to an increasing number of markets and to tease out all of its implications. Third, a synthesis emerges, bringing together the progress in different areas into one large model that tries to capture many features of an aggregate economy. This last stage is always technically challenging and involves considerable ingenuity at fine tuning models to match the subtleties of the data.

One example of this evolution is the progress from Keynes's ideas on the role of aggregate demand, disequilibrium, and rigidities, to the refining work on the investment accelerator, the consumption function, money demand, and the Phillips curve, finally leading to the synthesis of these ideas in the large-scale MPS and Brookings models. Similarly, over the last 30 years, the ideas of Finn Kydland and Edward Prescott (1982) and Gregory Mankiw and David Romer (1991) were applied and refined, culminating in the 2000s in the dynamic stochastic general equilibrium (DSGE) synthesis of Lawrence Christiano, Martin Eichenbaum, and Charles Evans (2005) and Frank Smets and Raf Wouters (2003). For a subgroup of macroeconomists, work in the last few years has been solidly in the third, synthesis stage.

The DSGE approach has never lacked for criticism (for a recent critique, see Caballero 2010), but until recently these models seemed successful at empirically matching business cycle facts and producing short-run forecasts that were as good as those from vector autoregressions (VARs). However,

1. I am grateful to Betsy Feldman, Dylan Kotliar, and Benjamin Mills for comments.

the Great Recession dealt this body of work a heavy blow. The models not only failed to predict the crisis but also were unable to provide an interpretation of the events after the fact, because for the most part they omitted a financial sector. It is too early to tell whether this failure will lead to this class of DSGE models being refined or abandoned, but already it is clear that their empirical performance must be judged more carefully.

This is what Rochelle Edge and Refet Gürkaynak set out to do in this paper: to reassess empirically the forecasting performance of the Smets and Wouters DSGE model. They explore how this model would have forecasted, from 1 to 8 quarters ahead, movements in inflation, output growth, and interest rates between 1997 and 2006. Importantly, they do not give the model the unfair advantage of 20-20 hindsight. In 2000Q1, for example, their fictional econometrician produces estimates and forecasts using only the data available at the time.

The conclusions of their exercise are surprising, at least to this reader. On the positive side, the DSGE model's forecasts beat those from a Bayesian VAR as well as the Greenbook forecasts compiled by the staff of the Federal Reserve, and its forecasts are precise, as demonstrated by their small root mean squared errors (RMSEs). On the negative side, the forecasts themselves are terrible, worse than a simple naïve forecast of constant inflation (or constant output growth), and worse than a forecast that simply assumes that inflation equals its last available observation. In addition, the model's low RMSEs are much less impressive once one realizes that the variance of inflation was also quite small during this period. Rather, the forecasting power is close to zero, and trying to improve the forecasts through some second-stage "cleaning" regressions makes almost no difference.

Contemplating this outcome, the authors see the glass as half full. They argue that according to the model, if monetary policy was effective, then inflation *should* be difficult to predict and should have a low variance. I am considerably more skeptical of this point of view in light of the events of the last 2 years. Inflation and output growth have not been stable since 2008, but rather have fallen quite dramatically. At the same time, the model's forecast errors for 2008–10 are large and persistent, as figures 5 and 6 of the paper demonstrate. If the authors' explanation is correct, these two facts would have been highly unlikely, unless monetary policy suddenly became particularly ineffective during these last 2 years. I would argue instead that larger shocks during this period simply exposed the model's faults.

Beyond this general assessment, I will offer two comments on the paper, as well as on the broader literature on DSGEs and forecasting. First, I will quibble somewhat with the authors' methodology, in particular with their peculiar mix of Bayesian and frequentist elements. Second, I will argue more generally that by setting themselves the goal of unconditional forecasting of aggregate variables, macroeconomists are setting such a high bar that they are almost sure to fail. Instead I will argue, through reference to a practical example, that DSGE models can be useful at making predictions even when they fail at making forecasts.

FORECASTING METHODOLOGY. The problem of estimation and forecasting with a DSGE model (or indeed with most models) can be expressed in the following setup. Assume that a researcher has a model or structure, S, that postulates some relationships among variables. The model has a vector of parameters, $\boldsymbol{\theta}$, and some prior information is available about what their values might be, captured in a probability density function $p(\boldsymbol{\theta}|S)$. The sample of data that one is trying to explain at some date t, including current and past observations of many variables, is denoted by y_t, and its density is $p(y_t|S)$. Finally, the likelihood of having observed these data is the density $L(y_t|S,\boldsymbol{\theta})$, which is typically known and easy to calculate given certain assumptions about the normality of the distribution of shocks.

Edge and Gürkaynak use Bayes's rule to estimate the parameters:

$$p(\boldsymbol{\theta}|y_t,S) = \frac{L(y_t|S,\boldsymbol{\theta})\,p(\boldsymbol{\theta}|S)}{p(y_t|S)}. \tag{1}$$

The output is a posterior density that reflects the uncertainty about the parameters through the whole posterior distribution. Although conceptually simple, this estimation work can be computationally exhausting. Fortunately, there has been much progress on algorithms in this area, as evidenced by the fact that Edge and Gürkaynak's paper contains more than 300 estimates of the model for different subsamples.

BAYESIAN ESTIMATION BUT NOT BAYESIAN FORECASTING. When it comes to forecasting, the authors take a distinctly non-Bayesian approach. First, they pick the mode of the posterior density at a date t: $\boldsymbol{\theta}_t^* = \arg\max_\theta p(\boldsymbol{\theta}|y_t, S)$. Next, they use the model's law of motion to obtain the probability density for the variable to be forecasted j periods ahead: $p(y_{t+j}|y_t, S, \boldsymbol{\theta}^*)$. Finally, they take the average over this density to represent their model forecast as an expectation:

$$m_{t+j}(\boldsymbol{\theta}_t^*, S) = \int y_{t+j}\; p\left(y_{t+j}\middle|y_t, S, \boldsymbol{\theta}_t^*\right)\mathrm{d}y_{t+j}. \tag{2}$$

The common approach when taking a frequentist perspective is to take the mode of the density (akin to the maximum-likelihood estimator) and produce the unbiased point forecast. But this is unnatural to the Bayesian, who is careful to take into account parameter uncertainty in the estimation stage, and so does not want to ignore it by focusing on the mode when it comes to forecasting. Likewise, it is awkward for a Bayesian to focus on one average forecast rather than report that there is a distribution of possible forecasts, each with some probability of occurring.

As I see it, asked what the model predicts for inflation or output j periods out, the Bayesian forecaster would perform the following computation:

$$b\left(y_{t+j}\middle|y_t, S\right) = \int p\left(y_{t+j}\middle|y_t, S, \boldsymbol{\theta}\right)p\left(\boldsymbol{\theta}\middle|y_t, S\right)\mathrm{d}\boldsymbol{\theta}. \tag{3}$$

That is, she would consider both the uncertainty about the future due to the possible arrival of shocks, captured as a density, $p(y_{t+j}|y_t, S, \boldsymbol{\theta})$, and the uncertainty on the parameter estimates, captured as a posterior, $p(\boldsymbol{\theta}|y_t, S)$. Instead of producing a single average forecast, the Bayesian forecaster would integrate over all the possible parameter combinations, $\boldsymbol{\theta}$, and report not a single number but rather a density function of possible forecasts, $b(y_{t+j}|y_t, S)$, given the current data and the model at hand. To assess whether the model is good at forecasting, this econometrician might then ask, How often does the actual realization of y_{t+j} fall within the interquartile range of its prediction, $b(y_{t+j}|y_t, S)$? If this happens much less often than 75 percent of the time, then the model is not giving good forecasts.

WHAT IS IN THE MODEL, WHAT IS IN THE PARAMETERS? Another difficulty with the authors' methodology is that although they try very hard to keep future information from influencing their past forecasts, one can only push this pseudo-forecasting exercise so far. The authors are careful to try to use only data available up to date t to produce forecasts for date $t + j$. This care is evident in two ways. First, the forecast, $m_{t+j}(\boldsymbol{\theta}^*_t, S)$, depends on the posterior estimate of parameters, $\boldsymbol{\theta}^*_t$, which used only data up to date t. Second, the data are not the revised data that we have today for that period, but rather the data that forecasters had available at the time.

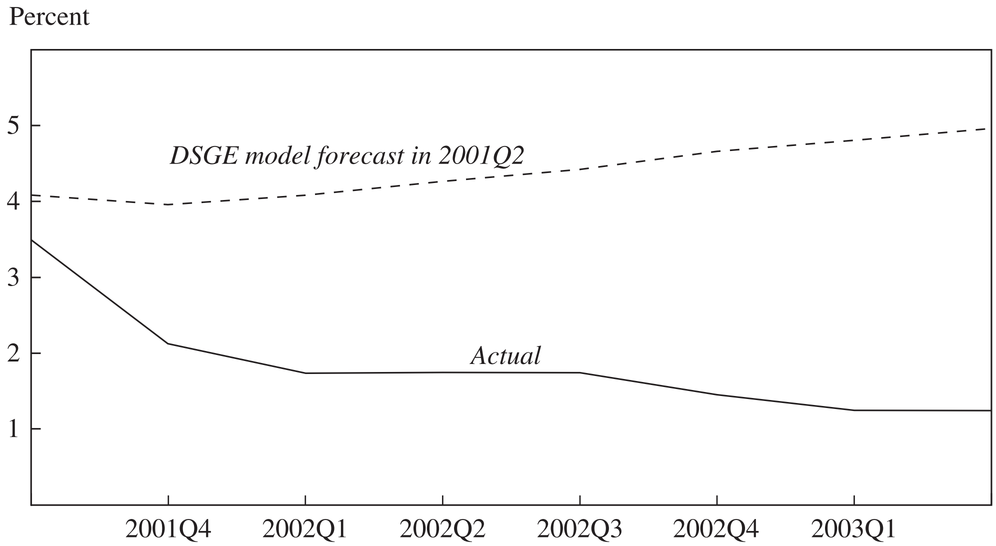
However, Edge and Gürkaynak use the model structure S at all dates, as given to them by Smets and Wouters (2007). As the opening paragraph of Smets and Wouters (2003) makes clear, this structure did not arise purely from theory. Rather, it assumes a particular utility function with a very peculiar habit term and a very specific law of motion. The Smets and Wouters model assumes adjustment costs for some actions but not for others, and it has sporadic updating, not of prices, but of prices relative to a backward-looking index. All of these elements and more arose because the Smets and Wouters model is the result of an iterative process between theorists and the data over the previous 20 years. Thus, even if the authors' estimates of the parameters in 1992Q1 use only information available then, the structure brought to the data was arrived at by researchers looking continuously at the data all the way into the 2000s and adjusting that structure to improve its fit and forecasting performance.

Moreover, the distinction between S and $\boldsymbol{\theta}$ is ultimately arbitrary. The Smets and Wouters model has a Cobb-Douglas production function (the S) for which the parameter is the labor share (the $\boldsymbol{\theta}$). But one can also see this as a production function with a constant elasticity of substitution (the S) and with the labor share and this elasticity of substitution (the $\boldsymbol{\theta}$) as parameters. Researchers used data covering all of the sample to agree on a strict prior that the elasticity of substitution is exactly equal to 1, and this knowledge has become embedded in the structure of the model, transitioning from $\boldsymbol{\theta}$ to S. In short, Edge and Gürkaynak make forecasts from the perspective of the 1990s using the structure S that researchers arrived at from interacting with the data in the 2000s.

THE HIGH, AND PERHAPS UNREALISTIC, EXPECTATIONS OF MACROECONOMISTS. Turning more generally to the goal of the broad literature that uses DSGE models in forecasting, I wonder whether macroeconomists are being unrealistically ambitious. At the same session of the Brookings Panel conference at which Edge and Gürkaynak presented this paper, two other papers were presented. In one, Thomas Dee and Brian Jacob build a regression model of educational outcomes to identify the effects of the No Child Left Behind policy. In the other, Gary Gorton and Andrew Metrick offer a theory of the role of shadow banks in the financial system and use it to justify a form of regulation. One could ask the authors of both papers, What are your unconditional forecasts for student achievement and total financial assets, respectively, in the United States for 2010–12?

If one attempted, literally, to use the models in those papers to make such forecasts, the results would likely be terrible. But it is not hard to guess that the authors would be puzzled that I would even be asking the question, and almost surely they would not endorse the forecasts thus arrived at. Nor, I would venture, would most, if not all, labor and financial economists. Most economists write models to capture some particular trade-offs and to make some limited predictions about what would happen if a particular policy were followed. To many economists, it is hard to imagine that one could know enough about any given market to

Figure 1. Federal Funds Rate: DSGE Model Forecast and Actual, 2001Q3–2003Q2



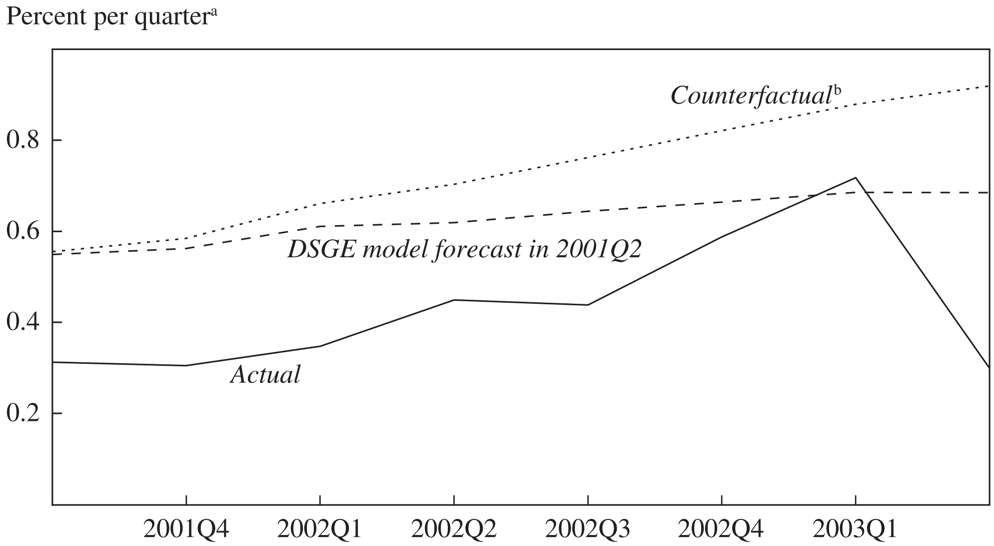
Source: Federal Reserve data and author's calculations.

make the type of unconditional forecasts sought in the question posed in the previous paragraph.

Some macroeconomists, however, do not shy away from producing unconditional forecasts. On the one hand, this is puzzling. If anything, our ability to forecast many aggregate variables at once is likely smaller than our ability to forecast outcomes in particular education or financial markets. On the other hand, it is understandable that macroeconomists produce these forecasts because there is an enormous demand for them from policymakers and the public at large. One consequence of this ambition to produce unconditional forecasts is that, with some regularity, the forecasts fail, sometimes in spectacular fashion. Forecasting is, simply put, a very hard thing to do.

PREDICTION INSTEAD OF FORECASTING. Even if unconditional forecasting may be too hard a task, a model can still make sharp *predictions* that are useful to policymakers. As an interesting illustration, consider the challenge facing the Federal Reserve at the start of 2001Q3. The economy was hit by a shock that economists did not predict (and, I would add, should not have predicted): the September 11 terrorist attacks. Imagine that the Federal Reserve at the time was using the Smets and Wouters model estimated by Edge and Gürkaynak to consider two possible policy responses to this shock. One response would be to ignore the shock, keeping to the same course of action as planned beforehand. This is displayed in my figure 1 as

Figure 2. Inflation: DSGE Model Forecast, Actual, and Post–September 11 Counterfactual, 2001Q3–2003Q2



Source: Bureau of Economic Analysis data and author's calculations.

a. Inflation is measured as the quarter-to-quarter change in the GDP deflator.

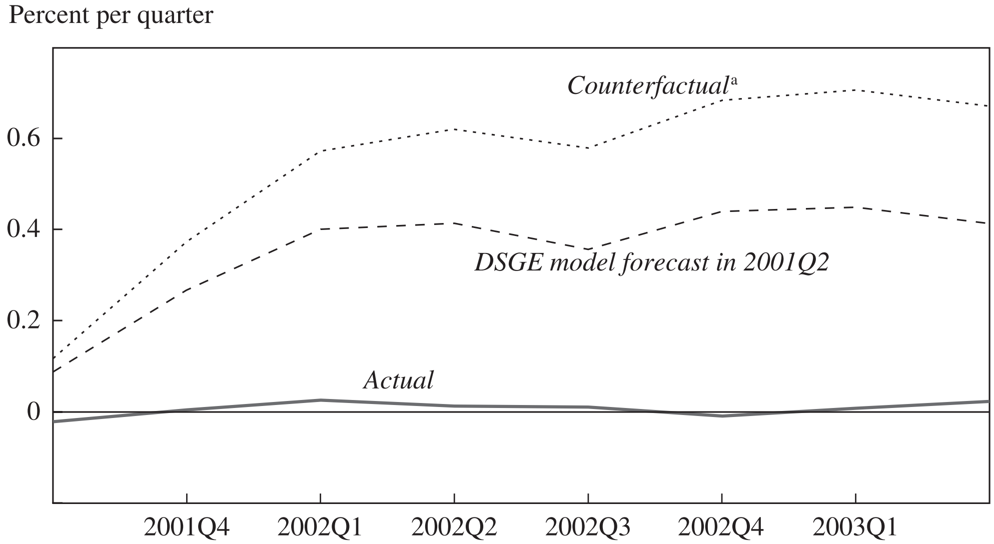
b. Inflation rate that would have prevailed had the Federal Reserve not changed its federal funds rate target after September 11, 2001.

the forecasted path for nominal interest rates before the terrorist attack. The other response would be to cut nominal interest rates aggressively. This is captured in the figure by the actual path of interest rates that the Federal Reserve followed. Figure 2 shows the effect of the two policies for inflation, and figure 3 for GDP. I obtained these by substituting the differences between the two paths in figure 1 and treating those as innovations that were then fed through the model. Because the solved Smets and Wouters model is linear, this delivers the right partial effect from considering what discretionary policy response to follow.

The model predicts that by aggressively cutting interest rates, the Federal Reserve generated higher inflation throughout the next 2 years, cumulating to a difference of almost 0.3 percentage point. That implies that whereas actual inflation in the United States was 0.3 percent in 2003Q2, if the Federal Reserve had not reacted to the shock, it would have been close to zero. Similarly, according to the model, GDP growth, instead of being close to zero between 2001Q3 and 2003Q2, would have been between –0.2 and –0.3 percent for most of 2002 and 2003.

This is the type of prediction that, I would conjecture, policymakers want from a model. It answers the following question: If some policy course is

Figure 3. GDP Growth: DSGE Model Forecast, Actual, and Post–September 11 Counterfactual, 2001Q3–2003Q2



Source: Bureau of Economic Analysis data and author's calculations.

a. GDP growth rate that would have prevailed had the Federal Reserve not changed its federal funds rate target after September 11, 2001.

followed, what will happen? Moreover, the DSGE model can confidently answer two further questions. First, why is the model predicting this? The impulse responses to monetary policy shocks in the model, and the trade-offs that agents face within it, provide a clear answer to this question. Second, how confident can we be about these predictions? This could be easily assessed by using the Bayesian approach I described in the previous section, rather than taking the modal estimate as I did for these plots.

This is where DSGE models excel. Indeed, few other types of models in economics can compete with them at answering these types of questions. DSGE models allow the researcher to provide precise quantitative predictions, to quantify the uncertainty around them, and to attach to the forecasts an internally coherent economic narrative. Considering more alternative scenarios is easy within the model, and more broadly, the information presented this way can be supplemented with that from other models as well as other subjective inputs.

If the models are going to be used this way, then one would like to know how good these predictions are. Unconditional forecasts do not answer this question, even if they give a strong hint (and the poor performance of the forecasts found by the authors suggests that the predictions may not be very trustworthy). As an alternative, researchers can (and do) compare the model's predictions with identified impulse responses from VARs or from natural experiments. Or they can use individual studies of the different mechanisms that the model is synthesizing, to see if the different parts of the story hold up on their own. I hope that more effort will go into refining the tests of models along this dimension. This would help in judging other DSGE models as well as in ultimately deciding whether the whole DSGE research agenda is useful.

REFERENCES FOR THE REIS COMMENT

Caballero, Ricardo J. 2010. "Macroeconomics after the Crisis: Time to Deal with the Pretense-of-Knowledge Syndrome." *Journal of Economic Perspectives* 24, no. 4: 85–102.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 2005. "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy* 113: 1–45.

Kydland, Finn E., and Edward C. Prescott. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica* 50, no. 6: 1345–70.

Mankiw, N. Gregory, and David Romer. 1991. *New Keynesian Economics,* vols. 1 and 2. MIT Press.

Smets, Frank, and Raf Wouters. 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association* 1, no. 5: 1123–75.

———. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97, no. 3: 586–606.

COMMENT BY

CHRISTOPHER A. SIMS It is important from time to time to look at the forecasting records of models used for policy analysis. This is how forecasters and users of models learn which ones are more reliable and discover ways to improve model specifications. Doing these evaluations is harder than it might appear. Data revisions are of the same order of magnitude as forecast errors, so it is essential to take a consistent view of what is to be forecast and to make sure that forecasts being compared are based on the same data. This is a formidable task if done carefully, and this paper by Rochelle Edge and Refet Gürkaynak has indeed done it carefully.

The paper says that the forecasts of dynamic stochastic general equilibrium models, like the other forecasts it considers, have been "poor" and "not very useful for policymaking" and that the DSGE model forecasts "do