Commentary

Christopher A. Sims

THE CONTEXT

Galí's paper (2005) is difficult to understand unless one places it in the context of the series of papers, set off by Galí's 1999 paper, which investigates how much of the business cycle is accounted for by "technology shocks." The 1999 paper found that little of the business cycle was accounted for by technology shocks and that technology shocks caused productivity and labor input to move in opposite directions, contrary to the pattern of most business cycle fluctuations. Subsequently, in the paper Galí labels CEV, Christiano, Eichenbaum, and Vigfusson (2003) showed that in a two-variable or multivariable vector autoregression (VAR) identified by long-run restrictions, one could obtain a quite different result if one used data in levels rather than first differences. Chari, Kehoe, and McGrattan (2004) have attacked structural VARs in general, apparently motivated by their disbelief in the original Galí result. Galí and Rabanal have surveyed this literature and connected it to related literature. The Galí and Rabanal paper (2004; henceforth GR) is the place to start if one is interested in this literature, as it considers a wide variety of previous work and makes some nicely executed contributions of its own.

The conclusion in GR is still that technology shocks are not the main cause of business cycles. The most convincing evidence in GR is the results from the multivariate equilibrium model estimated in the paper and from the similar multivariate equilibrium models that have been fitted to U.S. and European data by Smets and Wouters (2003a,b). The Smets and Wouters models are validated by careful comparison of their statistical fit to that of Bayesian VARs. These models suggest a contribution of technology shocks of about 15 to 35 percent of business cycle variance, in contrast to the under-10 percent estimates in Galí's paper for this conference (2005) and his original (1999) paper. GR show that the estimated contribution of neutral technology shocks does not rise to the high levels suggested by the early real business cycle (RBC) literature unless all of a long list of frictional mechanisms are shut down. They do not compute posterior odds ratios, and indeed the error bands they display for impulse responses are so narrow that it seems likely that their model does not compete with Bayesian VARs in fit. Still, their results roughly match those of the Smets and Wouters models, in which we know that shutting down these frictions seriously impairs the fit.

Though one can read GR as confirming the original Galí paper's conclusions, GR does represent some movement away from the original paper's conclusions and a major step away from its methodology. GR acknowledge that there can be technology shocks that do not have a long-run impact on productivity, and indeed that such shocks emerge in estimated dynamic stochastic general equilibrium (DSGE) models as important and as inducing positive comovement between labor input and productivity. The possibility of drifting productivity due to nontechnology shocks

Christopher A. Sims is a professor of economics at Princeton University and a visiting scholar at the Federal Reserve Banks of New York and Atlanta.

Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 487-92.

is also acknowledged, with capital taxes discussed explicitly; drifts in time preference, which they don't discuss explicitly, might in principle be equally important. And as we have already noted, the estimated contribution of technology shocks to variance is in GR much greater than in the earlier Galí paper and in this paper.

So the substantive conclusion from GR, as I read it, is that technology shocks are very likely important enough that Keynesians of the early 1970s would have found the results surprising, even if they are also very likely not nearly as important as suggested in the earliest RBC models. In fact, it is a bit ironic that the fitted DSGE models imply similar nontrivial but modest roles for monetary policy shocks and technology shocks. It is as if the data are telling us that extremist monetarists, Keynesians, and RBC calibrators are all wrong, but all have a piece of the truth.

The methodological conclusion from the GR paper, as I read it, is that models with one or two shocks and/or one or two variables should be set aside. The estimated DSGE models do not imply that such small models can cleanly separate two meaningful categories of disturbances to the economy. There is plenty of room for methodological improvement, however, and this could change substantive conclusions. The entire literature, including GR, struggles with issues of detrending, differencing or not differencing, Hodrick-Prescott (HP) filtering, etc. Even Smets and Wouters, who do the best job so far of integrating various sources of uncertainty into substantive conclusions, use ad hoc detrending methods and do not fully incorporate uncertainty about low-frequency behavior into their analysis.

The paper at hand, though it does not reach this conclusion explicitly, is in fact a confirmation of the points that bivariate models are inadequate and that uncertainty about low frequencies is central. It documents drifting behavior in consumption share and hours per worker in all the countries it studies. It provides potential explanations for these drifts that amount to postulating additional sources of disturbance, which implies that a larger model would be useful, or perhaps even necessary, to sort out sources of variation. And its results are inconclusive and variable across countries.

LONG- AND SHORT-RUN INFERENCE

From one perspective, the RBC innovation was to insist that we should integrate the theoretical frameworks in which we analyzed growth and business cycle fluctuations. It has always seemed to me paradoxical, therefore, that the convention in the RBC literature has been to filter low-frequency variation out of the data before proceeding to analyze business cycle variations. The practical reason for this is similar to that underlying the use of deseasonalized data: The low-frequency data are *extremely informative* about parameters of simple growth models, so that if we fit freely to all the data, the resulting model would be essentially determined by the very-low-frequency data, leaving a poor fit to the cyclical frequencies. Of course, this is not a necessary outcome in principle. It is because of unrealistic simplicity in model dynamics that they cannot at the same time match low-frequency and higher-frequency data.

With seasonality, the additional model complexity required to fit seasonal and cyclical variation simultaneously is arguably a poor trade-off, because seasonality involves phenomena—like weather and holidays—that bring in new parameters unrelated to our central interests. With growth, though, the additional model complexity required would essentially be just more flexible and realistic modeling of sources of inertia, not fundamentally new structural parameters. We could handle such models now.

The literature has persisted in using ad hoc detrending methods and ignoring the effects of the detrending on uncertainty. The best treatment of low-frequency variation so far is probably in the Smets and Wouters U.S. model, where they remove a common linear trend from most (logged) variables and account for uncertainty about the trend parameter in constructing their posterior distributions. The trend is not treated directly as a structural parameter, however, and sample means are extracted in advance with no accounting for uncertainty about them. In their model of the European Economic and Monetary Union, Smets and Wouters extracted a separate trend from each variable and did not account for uncertainty about the trends. Unsurprisingly, in light of these differences in treatment of low frequencies, the variance decomposition for the European data is quite different from that for the U.S. data, even though the impulse responses are qualitatively similar in the short run.

Galí's early paper and the one at hand use differenced log data. CEV pointed out that conclusions differed with variables in levels. GR agree, pointing also to work by other authors, that results are sensitive to whether data are differenced and to what kind of trend-removal is applied.

It appears to me that there are two reasons, beyond the fact cited above that standard models are not crafted to match low- and business cycle–frequencies simultaneously, for the persistence of detrending and differencing adhockery. One is the use of the HP filter in the early RBC literature and the strong tendency in the economics literature for the methodology of widely read papers to be imitated uncritically. The other is that, until recently, few economists understood Bayesian reasoning and hence most were inhibited by the formidable conceptual problems for inference about low frequencies in a frequentist approach. GR, though, do estimate a multivariate DSGE model using Bayesian methods. Bayesian methods can easily accommodate inference about means, trends, and orders of differencing. It is therefore disappointing that GR follow the rest of the literature in using a preliminary ad hoc detrending approach.

UNIT ROOTS, IDENTIFICATION

There is a fundamental problem, recognized years ago, with identification of VARs by means of long-run restrictions. Sums of coefficients in MA or AR operators are weakly identified—indeed identified only by means of lag length restrictions—unless the variables driving the operator are nonstationary. Without the nonstationarity, one can fix sums of coefficients arbitrarily while achieving fits arbitrarily close to that of the true model.

The sum of coefficients that drives identification in this paper's structural VAR exercise is a sum of coefficients on a lag operator that applies to a stationary variable—the nontechnology shocks in the MA form, the Δn variable in the AR form. It is therefore likely that the identification is fragile. Just to illustrate, suppose Δn is i.i.d. and that the true lag distribution on Δn in the equation defining the technology shock is a sequence of zeros. Suppose further that there is a rotation of the model that makes the lag distribution on Δn in that equation $\gamma(L) = L$, i.e., only lagged Δn enters the equation, with a coefficient of 1. How close could the mistaken rotation, in which $\gamma(1) \neq 0$, come to the fit of the true model while still satisfying the identifying assumption $\gamma(1) = 0$? The answer depends on lag length. If we fit a model in which $\gamma(L)$ is restricted to order 5, setting $\gamma(L) = 0.8L - 0.2L^2 - 0.2L^3 - 0.2L^4 - 0.2L^5$ gives a predictor with zero sum of coefficients that has $R^2 = 0.8$ with the false $\gamma(L) = 1$ predictor. With a lag length of n we can achieve an R^2 of $1 - 1/n$.

That identification in this setup depends on our treating lag length as known a priori is made very clear in the framework used by Shapiro and Watson (1988) and followed also in CEV. The estimation proceeds by using an equation of the form

$$\Delta f_t = \beta(L)\Delta f_{t-1} + \gamma(L)\Delta X_t + \varepsilon_t^z,$$

where f_t is productivity and X_t is a list of other, stationary variables. Current X is allowed in the equation, but is assumed to be possibly correlated with ε_t^z. β contains powers 1 to q of L and γ contains powers 0 to $q-1$. The solution to the simultaneity problem is to use lags 1 to q of Δf_t and lags 1 to q of X_t as instruments. It may be initially puzzling as to how this makes sense. The proposed instruments all seem to appear directly in the equation. But this is not quite true. X appears in the restricted equation only as ΔX. The instruments are the levels of X. It is true that all the lagged ΔX terms are exact linear combinations of the instruments, but current ΔX_t is not—quite. Identification conditions are formally satisfied because the best predictor of ΔX_t based on $q-1$ lags of ΔX_t is not as good as the best predictor based on q lagged levels of X_t. But as the little example above should make clear, the difference between these two predictors, and hence the firmness of the identification, quickly shrinks toward zero as q increases. Because in fact we do not know q, but adjust it by checking fit with various values of q—i.e., we estimate q—the zero sum of coefficients restriction is arguably no restriction at all.

Once we understand this knife-edge identification, it is unsurprising that apparently minor differences in specification can have major effects on results.

The SVAR work in this paper is conditioned on there being two, nonrepeated unit roots in the joint productivity and hours process. The data are consistent with this assumption. When I run for Italy and the United States a levels version of the reduced form VAR underlying this paper's panel of country SVARs, a weak Minnesota prior, which pulls gently toward the two-unit-roots hypothesis, manages to pull the point estimates to nearly exactly satisfying the two-unit-roots hypothesis.

But when I run that reduced form without a prior, the sums of coefficients matrices emerge as follows:

	Italy		**United States**	
	Production	Hours	Production	Hours
Production	0.9555	–0.0109	0.9823	0.0225
Hours	–0.0179	0.6975	0.4634	0.6028

The roots associated with the coefficient matrix for Italy are obviously about 0.956 and 0.698. For the U.S. matrix they are 1.008 and 0.577. Clearly a single-unit-root hypothesis is *also* consistent with the data, indeed even more consistent with it (in terms of likelihood) than the two-unit-roots hypothesis. Hours in the United States are estimated as nonstationary only because of cointegration with productivity. In Italy they are estimated as stationary, with the paper's identifying assumption very nearly satisfied by the reduced form, taking the nontechnology shock as simply the productivity innovation.

It is possible to deal directly with the uncertainty about whether roots are exactly 1 and how many roots are nonstationary. Bayesian inference in these models has no need for preliminary tests to determine stationarity or for identifying cointegrating vectors and conditioning inference on them. From a Bayesian perspective, all scientific reporting of inference is best regarded as helping readers to understand the shape of the likelihood function, on which any decision making use of the results ought to be based. The likelihood function for these models shows no special behavior as we cross from stationary to nonstationary regions of the parameter space.

I have reestimated the paper's structural VAR using priors that pull toward unit root behavior with varying intensity and toward forms with one or two unit roots. In most countries, the qualitative results are similar to those shown in the paper, with unambiguously negative responses of hours to identified technology shocks in the same countries where the paper finds them. In the United States, though, results are sensitive to how insistent one is about there being two distinct unit roots.

The prior I use (documented in Sims and Zha, 1998, and also in the comments to the code in mgnldnsty.R or mgnldnsty.m available at sims.princeton.edu/yftp/VARtools) has one component indexed by the parameter λ that pulls estimates toward at least one unit root and zero constant term, or else toward stationarity (with a nontrivial constant). Another component, indexed by μ, pulls toward independent unit roots in all variables. When μ is even moderately large, the posterior peaks at two roots very close to 1. But for any given value of μ, as λ increases, the estimates eventually flip to showing a positive response of hours to productivity shocks at all horizons. Figure 1 shows the posterior modal response for $\lambda = 0.5$, $\mu = 0$, together with the 90 percent error band. The error band is not so different from that shown in Galí's paper, but the location of the modal response is very different. This prior is very weak, but not so weak as to imply low posterior odds. The marginal data density corresponding to the plot is within a factor of 10 of the highest marginal density I have found by varying parameters of the prior. It is possible to get the same pattern in the modal responses, higher marginal data densities, and narrower implied error bands, by tightening up the prior somewhat.

Figure 1

Response of Employment to Technology Shock: λ = 0.5, μ = 0

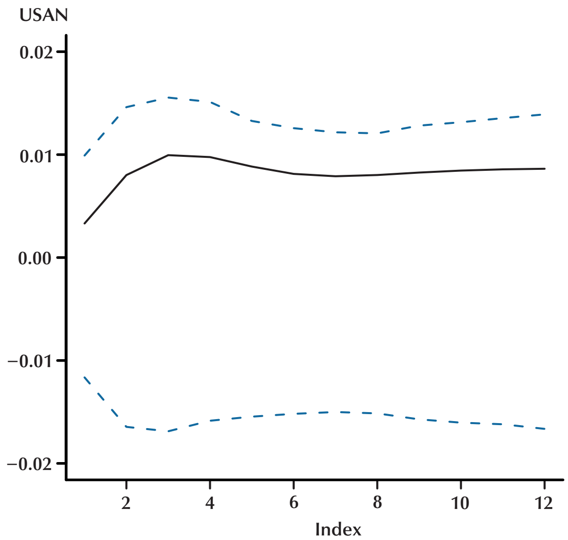


Figure 2

Largest Root, λ = 0.5, μ = 0

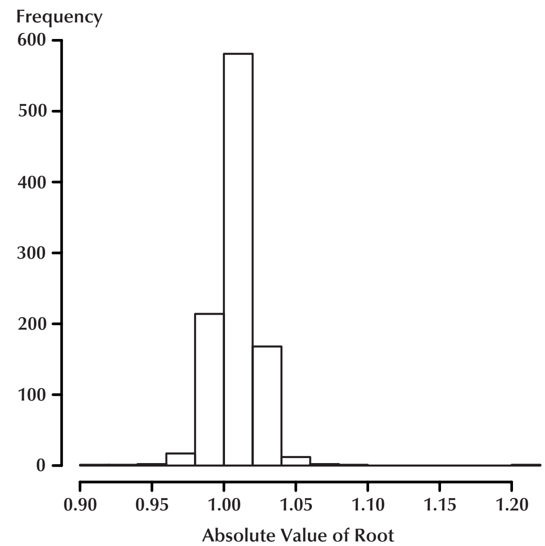
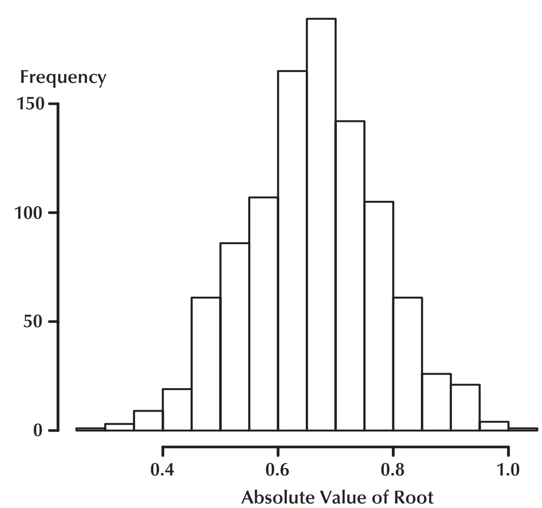


Figure 3

Second-Largest Root, λ = 0.5, μ = 0



The degree to which the data leave the sizes of roots indeterminate can be seen from the posterior probability density functions for the absolute values of the smallest and largest roots, shown in Figures 2 and 3, computed with the same fairly diffuse prior as was used to generate Figure 1.

These results only strengthen a conclusion already available from the impulse response graphs at the end of Galí's paper. Those figures show that the response of hours to a technology shock is estimated as significantly positive in one country (Japan), significantly negative in two countries (United Kingdom and Italy), and indeterminate in four (United States, Canada, France, and Germany). Reworking the U.S. data allowing for uncertainty in the number of roots has made the U.S. results even more uncertain and, with one reasonable prior, made the modal response positive rather than negative.

In fact, the responses of gross domestic product (GDP) to a technology shock are equally unstable across countries. Four countries (Italy, United States, Canada, and Germany) show indeterminate responses of GDP to a technology shock, one shows a negative response (United Kingdom), and two show a positive response (France and Japan).

It is natural then, I would say, to question whether the paper's methodology is isolating the same two structural shock in all of these countries. Note that at least for the United States and Europe, Smets and Wouters found in their larger model patterns of impulse responses that were quite stable across countries.

CONCLUSION

The paper is carefully done and thought-provoking. To get the most out of it, one should not let the clash between one-dimensional RBC models and two-dimensional SVAR models that occupies the foreground of the paper hide the background issues that the paper illuminates:

- Uncertainty about stationarity matters.
- If we are to integrate our modeling of long- and short-run macro-dynamics, it appears we need to go beyond one- and two-dimensional models.

REFERENCES

Chari, V.V.; Kehoe, Patrick J. and McGrattan, Ellen. "An Economic Test of Structural VARs." Federal Reserve Bank of Minneapolis, Research Department Staff Report 345, 2004.

Christiano, Lawrence; Eichenbaum, Martin and Vigfusson, Robert. "What Happens after a Technology Shock?" NBER Working Paper No. 9819, National Bureau of Economic Research, 2003.

Galí, Jordi. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 1999, *89*(1), pp. 249-71.

Galí, Jordi. "Trends in Hours, Balanced Growth, and the Role of Technology in the Business Cycle." Federal Reserve Bank of St. Louis *Review*, July/August 2005, *87*(4), pp. 459-86.

Galí, Jordi and Rabanal, Pau. "Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data?" NBER Working Paper No. 10636, National Bureau of Economic Research, 2004 (*NBER Macroeconomics Annual*, forthcoming).

Shapiro, Matthew D. and Watson, Mark W. "Sources of Business Cycle Fluctuations." *NBER Macroeconomics Annual 1988*. Cambridge, MA: MIT Press, 1989, pp. 111-48.

Sims, Christopher A. and Zha, Tao. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review*, November 1998, *39*(4), pp. 949-68.

Smets, Frank and Wouters, Raf. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association*, September 2003a, *1*, pp. 1123-75.

Smets, Frank and Wouters, Raf. "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." Discussion paper, European Central Bank and National Bank of Belgium, May 2003b.