COMMENT ON DEL NEGRO, SCHORFHEIDE, SMETS AND WOUTERS

CHRISTOPHER A. SIMS

1. Why this approach has been successful

This paper sets out to blend the advantages of VAR models, which forecast well, with those of dynamic stochastic general equilibrium (DSGE) models, which have fewer free parameters, allow prior information to be brought to bear more directly, and can be used for counterfactual policy simulations. They do this by modeling the data as a VAR — that is, without the tight parametric restrictions implied by a DSGE — but using a DSGE, and prior beliefs about the parameters of the DSGE, to generate a prior distribution for the parameters of the VAR. This approach was originated by Del Negro and Schorfheide, though it had precedents in earlier work they cite.

The most widely used priors for VAR's (with a prior, a VAR becomes a BVAR, or Bayesian VAR) are variants on the Minnesota prior. The details of that prior we need not trace out here. What is important about it is that it expresses beliefs only about the lengths of lags and degrees of persistence implied by the model; it treats all variables symmetrically and therefore incorporates no behavioral interpretations of parameters or equations. Macroeconomists have views, though, on how variables are related and how their properties differ. These views are most easily expressed as views about behavioral parameters in DSGE models. So the Del Negro Schorfheide approach is appealing.

Date: November 4, 2006.

Work on this comment was funded in part by NSF grant SES-0350686.

Another approach is that originated by the other two co-authors, Smets and Wouters, who use relatively richly parameterized DSGE's, together with priors on the parameters, to arrive at a DSGE model that fits and forecasts relatively well. The Del Negro and Schorfheide (DS) approach is probably the right one, though, for situations where the model is to be used in forecasting and policy analysis. This is in part because VAR models do still fit better than DSGE's when they are applied to real data (and not to processed data that has had trend removed by filtering or regression). But more importantly, aggregate DSGE models are story-telling devices, not hard scientific theories. We know there is no aggregate capital stock and no aggregate consumption good. We know that the real economy has a rich array of financial markets we do not include in our DSGE models. These and many other simplifications that go into the construction of aggregate behavioral models do not prevent them from helping us think about the way the economy works, but it does not make sense to require these models to match in fine detail the dynamic behavior of the accounting constructs and proxy variables that make up our data. When we do so, we find ourselves adding to the DSGE mechanisms for friction and inertia, or ad hoc "measurement error", with little empirical foundation or even intuitive plausibility. Making forecasts, policy projections, and (especially) welfare evaluations of policies with these models as if their behavioral interpretation were exactly correct is a mistake.

The fact that their approach generates a prior for a VAR, not a DSGE model fit to the data, was at the forefront in earlier papers on their methods by Del Negro and Schorfheide. This paper's exposition emphasizes the DSGE, but a careful reading makes it clear that the setup is still the same: the DSGE is a mechanism for generating a prior, not a model of the data.

Another approach has been to use VAR's as a standard of comparison for DSGE's, with Bayesian posterior odds ratios or pseudo-out-of-sample forecasting performance used to check whether the DSGE is close to matching the fit of a BVAR or VAR. While such comparisons are helpful, they can be hard to interpret. The methdology assumes that the models being considered are an exhaustive list of possible true models, when in fact they are usually representative points in a continuum of possible models. Furthermore, this approach leaves us with two extreme models — a BVAR with no substantive information incorporated in it, or a DSGE with tight and unbelievable parametric restrictions. The DS approach blends substantive prior information from the DSGE with the VAR model, introducing a continuous paramater to control the weight on the DSGE prior. This is more realistic and more easily interpreted.

2. IMPROVEMENTS I: A PROPER SYMMETRIC PRIOR ON THE VAR

Though this paper emphasizes the possibility of using the weight parameter λ as an indicator of the reliability of the DSGE, the DS methods in their current form cannot give a clear indication that the DSGE is useless, even if it is in fact useless. In models with more than two or three variables, unrestricted VAR's — which is what emerges from estimation with a flat prior — generally forecast very badly. These models have many free parameters, and estimating them all at once, without restrictions, induces sampling error that makes forecast errors large. BVAR's produce better results by introducing a prior favoring persistence, weak cross-variable connections, and smaller coefficients on more distant lags. The DS approach does not make any use of such symmetric, economics-free priors. The *only* way to bring in prior information of any kind is via putting some weight on the DSGE. But we know that with a flat prior a VAR will not fit well. What we would really like to know is whether the DSGE's behaviorally-based priors are helping beyond what could be achieved with symmetric priors.

The procedure could easily be improved by use of a proper, but "economics-free" prior on the VAR, e.g. some version of the "Minnnesota prior" This would make monotonicity of the marginal on λ with the peak at the VAR a realistic possibility, and thereby let us see whether the economics in the DSGE, as opposed to its serial correlation, is proving helpful.

3. Improvements II: Less ad hockery in identifying the structural VAR

The DS setup includes a reduced form VAR and also a structural VAR, related to each other in the usual way. In the structural VAR, the disturbances are interpreted behaviorally. Most importantly, there is one shock or set of shocks that are interpreted as stochastic shifts in policy behavior, and these correspond to equations that describe policy behavior. The interpretation of these structural shocks is the same as that of corresponding shocks in the DSGE model. Thus in the structural VAR it is possible to carry out counterfactual policy projections, holding policy variables on a given path and projecting other variables conditional on the policy actions that are required to produce that path for the policy variables. The DS notation for these two models is

$$\begin{aligned} \text{RF}: &\qquad y = \Phi\,(L)y + u\,, \quad \operatorname{Var}(u) = \Sigma_u \\ \text{SVAR}: &\qquad C(L)y = \varepsilon\,, \quad \operatorname{Var}(\varepsilon) = I\,, \\ \text{Connection}: &\qquad A_0A_0' = \Sigma_u\,, \quad A_0^{-1}\cdot(I - \Phi\,(L)) = C(L)\,. \end{aligned}$$

The reduced form and structural VAR's are connected via the relation above between A_0 and Σ_u. The DSGE implies a matrix $A_0(\theta)$ that connects the DSGE's implied reduced form VAR to its implied SVAR. One could imagine generating a prior on the SVAR, conditional on the DSGE parameters θ, by generating a prior conditional on θ on the reduced form coefficients Φ as DS do, and then asserting dogmatically that in the SVAR $A_0 = A_0(\theta)$. This is unappealing, though, because there seems to be no good reason to treat $A_0 = C_0^{-1}$ as deterministic conditional on θ when C_s for $s > 0$ are all treated as uncertain conditional on θ. This would amount to trusting completely the DSGE assertions about contemporaneous relations among variables, while treating its assertions about lagged effects as uncertain.

So DS do something else, which does treat A_0 as random conditional on θ. They apply a QR transformation to $A_0(\theta)$, expressing it as

$$A_0(\theta) = \Sigma^*_{tr}(\theta)\Omega(\theta)\,,$$

where $\Sigma^*_{tr}(\theta)$ is triangular and $\Omega(\theta)$ is orthonormal. They then write the SVAR A_0 as

$$A_0 = \Sigma_{tr}\Omega(\theta)\,, \quad \text{where } \Sigma_{tr} = \text{chol}(\Sigma_u)\,. \tag{$*$}$$

(Here $\text{chol}(X)$ is the Choleski factor of X.) In other words, the "rotation" matrix Ω is treated as non-stochastic, conditional on θ, while the lower triangular part of the QR decomposition is treated as a priori random, and with its distribution derived from their prior on the reduced form. Conditional on θ, a realization of the prior distribution for the SVAR is obtained by first obtaining a draw of the reduced form parameters (including Σ_u), then calculating the QR decomposition of $A_0(\theta)$, then applying ($*$).

But while this method does make A_0 random conditional on θ, it only sometimes treats the identifying restrictions embodied in the $A_0(\theta)$ matrix as stochastic. If it happens that $A_0(\theta)$ is triangular, for example, the Ω matrix is the identity and the SVAR is identified using exactly the restrictions that deliver triangularity of A_0. But if $A_0(\theta)$ is triangular only after a re-ordering of the variable list, then the SVAR generated by the DS prior conditional on θ will not exactly satisfy the re-ordered triangularity restrictions. This means that identifying restrictions from the DSGE may or not be applied deterministically. The QR decomposition, on which the DS procedure is based, gives results that depend on the ordering of the variables, and this is the source of this somewhat arbitrary behavior.

This could be fixed, though at some cost in complexity of the procedure.

3.1. **Improvements III: More emphasis on low frequencies.** Use of DSGE's as "core" models, insulated from the data, by central bank modelers suggests a lack of confidence in "statistical models" at low frequencies, but also lack of confidence in high frequency behavior of DSGE's This is quite explicit in the Bank of England's monograph rationalizing its recently developed BEQM model, but is also present in the Fed's FRBUS and the Bank of Canada's QPM, on which quite a few other central bank models have been based. One of the primary objections to the new "DSGE's that fit" is that to fit well they have to be equipped with many sources of inertia and friction that seem arbitrary (i.e., more uncertain a priori than is acknowledged by the model), yet may have important implications for evaluating policy.

The DS procedure does use cointegrating restrictions from the DSGE (non-stochastically). But otherwise it mimics information from a modest sized sample. Such a prior inherently is more informative about short than long run behavior. This also could be fixed. One could use dummy observations in the style of the Minnesota prior, centering on the DSGE implied VAR coefficients, but making beliefs tighter at low frequencies than at high frequencies.

3.2. **Conclusion.** The DS approach is already practically useful, and appears to be the most promising direction to follow in developing models that combine accurate probability modeling of the behavior of economic time series with insights from stochastic general equilibrium models. Of course the approach requires both a good time series model and a good DSGE to work with, so there is plenty of room for futher research on both these lines as well as on improving the DS methodology itself.