

On the Fit of New Keynesian Models

Marco DEL NEGRO

Federal Reserve Bank of Atlanta, Atlanta, GA 30309 (marcodelnegro@frbatlanta.org)

Frank SCHORFHEIDE

Department of Economics, University of Pennsylvania, Philadelphia, PA 19104 (schorf@ssc.upenn.edu)

Frank SMETS

European Central Bank, D-60311 Frankfurt, Germany (Frank.Smets@ecb.int)

Rafael WOUTERS

National Bank of Belgium, B-1000 Bruxelles, Belgium (Rafael.Wouters@nbb.be)

This article provides new tools for the evaluation of dynamic stochastic general equilibrium (DSGE) models and applies them to a large-scale new Keynesian model. We approximate the DSGE model by a vector autoregression, and then systematically relax the implied cross-equation restrictions and document how the model fit changes. We also compare the DSGE model's impulse responses to structural shocks with those obtained after relaxing its restrictions. We find that the degree of misspecification in this large-scale DSGE model is no longer so large as to prevent its use in day-to-day policy analysis, yet is not small enough to be ignored.

KEY WORDS: Bayesian analysis; Dynamic stochastic general equilibrium model; Model evaluation; Vector autoregression.

1. INTRODUCTION

Dynamic stochastic general equilibrium (DSGE) models are not only attractive from a theoretical perspective, but also are emerging as useful tools for forecasting and quantitative policy analysis in macroeconomics. Due to improved time series fit, these models are gaining credibility in policy making institutions, such as central banks. Up until recently, DSGE models had the reputation of being unable to track macroeconomic time series. In fact, an assessment of their forecasting performance was typically considered futile (an exception being DeJong, Ingram, and Whiteman 2000). Apparent model misspecifications were used as an argument in favor of informal calibration approaches to the evaluation of DSGE models along the lines of work by Kydland and Prescott (1982). Subsequently, researchers have developed econometric frameworks that formalize aspects of the calibration approach (see, e.g., Canova 1994; DeJong et al. 1996; Diebold, Ohanian, and Berkowitz 1998; Geweke 1999b; Schorfheide 2000; Dridi, Guay, and Renault 2007). A common feature of many evaluation procedures is that DSGE model predictions are either implicitly or explicitly compared with those from a reference model. Much of the applied work related to monetary models has proceeded by, for instance, assessing DSGE models based on discrepancies between impulse response functions obtained from the DSGE model and those obtained from the estimation of identified vector autoregressions (VARs). However, adopting Bayesian language, such an evaluation is sensible only if the VAR attains a higher posterior probability than the DSGE model, as was pointed out by Schorfheide (2000).

Smets and Wouters (2003) developed a large-scale monetary DSGE model in the new Keynesian tradition based on work by Christiano, Eichenbaum, and Evans (2005) and estimated it on

Euro-area data. One of their remarkable empirical results was that posterior odds favored their DSGE model relative to VARs estimated with a fairly diffuse training sample prior. Previous studies using more stylized DSGE models always found that even simple VARs dominate DSGE models. On the methodological side, Smets and Wouters' finding challenges the practice of assessing DSGE models based on their ability to reproduce VAR impulse response functions without carefully documenting that the VAR indeed fits better than the DSGE model. On the substantive side, it poses the question of whether researchers now should be less concerned about misspecification of DSGE models.

The contributions of this article are twofold, one methodological and the other substantive. First, we develop a set of tools that is useful for assessing the time series fit of a DSGE model. In particular, we systematically relax the implied cross-coefficient restrictions of the DSGE model to obtain a VAR specification that is guaranteed to fit better than the DSGE model yet simultaneously stays as close as possible to the DSGE restrictions. We use this specification as a benchmark to characterize and understand the degree of misspecification of the DSGE model. Second, we apply these tools to a variant of the model of Smets and Wouters and document its fit and forecasting performance based on postwar U.S. data. We find that model misspecification remains a concern.

Our model evaluation approach is related to work on DSGE model priors for VARs by Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004), as well as the idea of indirect inference developed by Gourieroux, Monfort, and Renault

(1993) and Smith (1993) and recently applied in a Bayesian setting by Gallant and McCulloch (2004). We use the VAR as an approximating model for the DSGE model and construct a mapping from the DSGE model to the VAR parameters. This mapping leads to a set of cross-coefficient restrictions for the VAR. Deviations from these restrictions are interpreted as evidence for DSGE model misspecification. In particular, we specify a prior distribution for deviations from the DSGE model restrictions. The prior tightness is scaled by a hyperparameter λ . The values $\lambda = \infty$ and $\lambda = 0$ correspond to the two polar cases in which the cross-coefficient restrictions are strictly enforced and completely ignored (unrestricted VAR). The marginal likelihood function of $\lambda \in (0, \infty]$ provides an overall assessment of the DSGE model restrictions that is more robust and informative than a comparison of the two polar cases, which is widespread practice in literature.

We denote the peak of the marginal likelihood function as $\hat{\lambda}$. We have evidence of misspecification whenever the marginal likelihood ratio of $\lambda = \hat{\lambda}$ versus $\lambda = \infty$ indicates that model fit improves substantially if the DSGE restrictions are relaxed. The resulting VAR specification, which we label DSGE-VAR($\hat{\lambda}$), can be used as a benchmark for evaluating the dynamics of the DSGE model. We ask the question: In which dimension do the impulse response functions change as we relax the cross-coefficient restriction? To facilitate impulse response function comparisons, we provide a coherent identification scheme for the DSGE-VAR. By coherent, we mean that in the absence of DSGE model misspecification and VAR approximation error, the impulse responses of the DSGE model and DSGE-VAR to all structural shocks would coincide. Thus, in constructing a benchmark for the evaluation of the DSGE model, we are trying to stay as close to the original specification as possible.

The empirical findings are as follows. The marginal likelihood function of the hyperparameter λ has an inverse U-shape, indicating that the fit of the AR system can be improved by relaxing the DSGE model restrictions. The shape of the posterior also implies that the restrictions should not be completely ignored when constructing a benchmark for the model evaluation, because VARs with very diffuse priors are clearly dominated by the DSGE-VAR($\hat{\lambda}$). This finding is confirmed in the pseudo-out-of-sample forecasting experiment. According to a widely used multivariate forecast error statistic, the DSGE model and the VAR with diffuse prior perform about equally well in terms of one-step-ahead forecasts but are clearly worse than the DSGE-VAR($\hat{\lambda}$).

Comparing impulse responses between the DSGE model and the DSGE-VAR($\hat{\lambda}$), we find that the DSGE model misspecification does not translate into differences among impulse response functions to technology or monetary policy shocks. The latter result is important from a policy perspective, because it confirms that despite its deficiencies, the predictions of the effects of unanticipated changes in monetary policy derived from the new Keynesian DSGE model are not contaminated by its dynamic misspecification. However, responses to some of the other shocks differ between the DSGE model and DSGE-VAR($\hat{\lambda}$), particularly in the long run, suggesting that some low-frequency implications of the model are at odds with the data. We also use the DSGE-VAR framework to make comparisons

across DSGE model specifications. In particular, we consider a version of the model without habit formation and another version without price and wage indexation. We find that the evidence from the DSGE-VAR analysis against the no-indexation specification is not nearly as strong as the evidence against the model without habit formation.

The article is organized as follows. Section 2 presents the DSGE model, and Section 3 discusses the DSGE model evaluation framework. Section 4 describes the data, and Section 5 presents empirical results. Section 6 concludes.

2. THE DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODEL

This section describes our DSGE model, which is a slightly modified version of the DSGE model developed and estimated for the Euro area by Smets and Wouters (2003). In particular, we introduce stochastic trends into the model so that it can be estimated with unfiltered time series observations. The DSGE model is based on work of Christiano et al. (2005) and contains numerous nominal and real frictions. To make this article self-contained, we subsequently describe the structure of the model economy and the decision problems of the agents in the economy.

2.1 Final Goods Producers

The final good, Y_t , is a composite made of a continuum of intermediate goods, $Y_t(i)$, indexed by $i \in [0, 1]$,

$$Y_t = \left[\int_0^1 Y_t(i)^{1/(1+\lambda_{f,t})} di \right]^{1+\lambda_{f,t}}, \quad (1)$$

where $\lambda_{f,t} \in (0, \infty)$ follows the exogenous process

$$\ln \lambda_{f,t} = (1 - \rho_{\lambda_f}) \ln \lambda_f + \rho_{\lambda_f} \ln \lambda_{f,t-1} + \sigma_{\lambda_f} \epsilon_{\lambda,t}, \quad (2)$$

where $\epsilon_{\lambda,t}$ is an exogenous shock with unit variance that in equilibrium affects the markup over marginal costs. The final goods producers are perfectly competitive firms that buy intermediate goods, combine them to get the final good Y_t , and resell the final good to consumers. The firms maximize profits,

$$P_t Y_t - \int P_t(i) Y_t(i) di,$$

subject to (1). Here P_t denotes the price of the final good and $P_t(i)$ is the price of intermediate good i . From their first-order conditions and the zero-profit condition, we obtain that

$$Y_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-(1+\lambda_{f,t})/\lambda_{f,t}} Y_t \quad \text{and} \quad (3)$$

$$P_t = \left[\int_0^1 P_t(i)^{-1/\lambda_{f,t}} di \right]^{-\lambda_{f,t}}.$$

2.2 Intermediate Goods Producers

Good i is made using the technology

$$Y_t(i) = \max\{Z_t^{1-\alpha} K_t(i)^\alpha L_t(i)^{1-\alpha} - Z_t \mathcal{F}, 0\}, \quad (4)$$

where the technology shock Z_t (common across all firms) follows a unit root process and \mathcal{F} represents fixed costs faced by the firm. Based on preliminary estimation results, we decided to set $\mathcal{F} = 0$ in the empirical analysis. We define technology growth, $z_t = \log(Z_t/Z_{t-1})$, and assume that z_t follows the AR process

$$z_t = (1 - \rho_z)\gamma + \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}. \quad (5)$$

All firms face the same prices for their labor and capital inputs. Hence profit maximization implies that the capital-to-labor ratio is the same for all firms,

$$\frac{K_t(i)}{L_t(i)} = \frac{\alpha}{1 - \alpha} \frac{W_t}{R_t^k}, \quad (6)$$

where W_t is the nominal wage and R_t^k is the rental rate of capital. Following Calvo (1983), we assume that in every period a fraction of firms ζ_p is unable to reoptimize their prices $P_t(i)$. These firms adjust their prices mechanically according to

$$P_t(i) = (\pi_{t-1})^{\iota_p} (\pi_*)^{1-\iota_p}, \quad (7)$$

where $\pi_t = P_t/P_{t-1}$, π_* is the steady-state inflation rate of the final good, and $\iota \in [0, 1]$. Those firms that are able to reoptimize prices choose the price level $\tilde{P}_t(i)$ that solves

$$\begin{aligned} \max_{\tilde{P}_t(i)} \mathbb{E}_t \left[\sum_{s=0}^{\infty} \zeta_p^s \beta^s \Xi_{t+s}^p \left(\tilde{P}_t(i) \left(\prod_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p} \right) - MC_{t+s} \right) \right. \\ \left. \times Y_{t+s}(i) \right] \\ \text{s.t. } Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i) \left(\prod_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p} \right)}{P_{t+s}} \right)^{-(1+\lambda_{f,t})/\lambda_{f,t}} Y_{t+s}, \\ MC_{t+s} = \frac{\alpha^{-\alpha} W_{t+s}^{1-\alpha} R_{t+s}^{\alpha}}{(1-\alpha)^{(1-\alpha)} Z_{t+s}^{1-\alpha}}, \end{aligned} \quad (8)$$

where $\beta^s \Xi_{t+s}^p$ is today's value of a future dollar for the consumers and MC_t reflects marginal costs. We consider only the symmetric equilibrium at which all firms will choose the same $\tilde{P}_t(i)$. Thus from (3), we obtain the following law of motion for the aggregate price level:

$$P_t = [(1 - \zeta_p) \tilde{P}_t^{-1/\lambda_{f,t}} + \zeta_p (\pi_{t-1}^{\iota_p} \pi_*^{1-\iota_p} P_{t-1})^{-1/\lambda_{f,t}}]^{-\lambda_{f,t}}. \quad (9)$$

2.3 Labor Packers

There is a continuum of households, indexed by $j \in [0, 1]$, each supplying a differentiated form of labor, $L(j)$. The labor packers are perfectly competitive firms that hire labor from the households and combine it into labor services, L_t , that are offered to the intermediate goods producers,

$$L_t = \left[\int_0^1 L_t(j)^{1/(1+\lambda_w)} di \right]^{1+\lambda_w}, \quad (10)$$

where $\lambda_w \in (0, \infty)$ is a fixed parameter. From first-order and zero-profit conditions of the labor packers, we obtain the labor demand function and an expression for the price of aggregated labor services L_t ,

$$L_t(j) = \left(\frac{W_t(j)}{W_t} \right)^{-(1+\lambda_w)/\lambda_w} L_t \quad (11a)$$

and

$$W_t = \left[\int_0^1 W_t(j)^{-1/\lambda_w} di \right]^{-\lambda_w}. \quad (11b)$$

2.4 Households

The objective function for household j is given by

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s b_{t+s} \left[\log(C_{t+s}(j) - hC_{t+s-1}(j)) - \frac{\phi_{t+s}}{1 + v_l} L_{t+s}(j)^{1+v_l} + \frac{\chi}{1 - v_m} \left(\frac{M_{t+s}(j)}{Z_{t+s} P_{t+s}} \right)^{1-v_m} \right], \quad (12)$$

where $C_t(j)$ is consumption, $L_t(j)$ is labor supply, and $M_t(j)$ is money holdings. A household's preferences display habit persistence. The preference shifters, ϕ_t , which affects the marginal utility of leisure, and b_t , which scales the overall period utility, are exogenous processes common to all households that evolve as

$$\ln \phi_t = (1 - \rho_\phi) \ln \phi + \rho_\phi \ln \phi_{t-1} + \sigma_\phi \epsilon_{\phi,t} \quad (13)$$

and

$$\ln b_t = \rho_b \ln b_{t-1} + \sigma_b \epsilon_{b,t}. \quad (14)$$

Real money balances enter the utility function deflated by the (stochastic) trend growth of the economy, so as to make real money demand stationary.

The household's budget constraint, written in nominal terms, is given by

$$\begin{aligned} P_{t+s} C_{t+s}(j) + P_{t+s} I_{t+s}(j) + B_{t+s}(j) + M_{t+s}(j) + T_{t+s}(j) \\ \leq R_{t+s-1} B_{t+s-1}(j) + M_{t+s-1}(j) + A_{t+s-1}(j) \\ + \Pi_{t+s} + W_{t+s}(j) L_{t+s}(j) \\ + (R_{t+s}^k u_{t+s}(j) \bar{K}_{t+s-1}(j) - P_{t+s} a(u_{t+s}(j)) \bar{K}_{t+s-1}(j)), \end{aligned} \quad (15)$$

where $I_t(j)$ is investment, $B_t(j)$ represents holdings of government bonds, $T_t(j)$ represents lump-sum taxes (or subsidies), R_t is the gross nominal interest rate paid on government bonds, $A_t(j)$ is the net cash inflow from participating in state-contingent securities, Π_t is the per capita profit that the household gets from owning firms (households pool their firm shares, and they all receive the same profit), and $W_t(j)$ is the nominal wage earned by household j . The term within parentheses represents the return to owning $\bar{K}_t(j)$ units of capital. Households choose the utilization rate of their own capital, $u_t(j)$.

Households rent to firms in period t an amount of effective capital equal to

$$K_t(j) = u_t(j) \bar{K}_{t-1}(j), \quad (16)$$

and receive $R_t^k u_t(j) \bar{K}_{t-1}(j)$ in return. However, they must pay a cost of utilization in terms of the consumption good equal to $a(u_t(j)) \bar{K}_{t-1}(j)$. Households accumulate capital according to the equation

$$\bar{K}_t(j) = (1 - \delta) \bar{K}_{t-1}(j) + \mu_t \left(1 - s \left(\frac{I_t(j)}{I_{t-1}(j)} \right) \right) I_t(j), \quad (17)$$

where δ is the rate of depreciation and $S(\cdot)$ is the cost of adjusting investment, with $S(e^\gamma) = 0$ and $S'(\cdot) > 0$. The term μ_t is a stochastic disturbance to the price of investment relative to consumption (see Greenwood, Hercowitz, and Krusell 1998), which follows the exogenous process

$$\ln \mu_t = (1 - \rho_\mu) \ln \mu + \rho_\mu \ln \mu_{t-1} + \sigma_\mu \epsilon_{\mu,t}. \quad (18)$$

The households' wage setting is subject to nominal rigidities as used by Calvo (1983). In each period, a fraction ζ_w of households is unable to readjust wages. For these households, the wage $W_t(j)$ will increase at a geometrically weighted average of the steady-state rate increase in wages (equal to steady-state inflation π_* times the steady-state growth rate of the economy e^γ) and of last period's inflation times last period's productivity ($\pi_{t-1} e^{z_{t-1}}$). These weights are $1 - \iota_w$ and ι_w . Those households that are able to reoptimize their wage solve the problem

$$\max_{\tilde{W}_t(j)} \mathbb{E}_t \sum_{s=0}^{\infty} \zeta_w^s \beta^s b_{t+s} \left[-\frac{\phi_{t+s}}{1 + \nu_l} L_{t+s}(j)^{1+\nu_l} \right] \quad (19)$$

s.t. eq. (15) for $s = 0, \dots, \infty$, (11a), and

$$W_{t+s}(j) = \left(\prod_{l=1}^s (\pi_* e^\gamma)^{1-\iota_w} (\pi_{t+l-1} e^{z_{t+l-1}})^{\iota_w} \right) \tilde{W}_t(j).$$

We again consider only the symmetric equilibrium in which all agents solving (19) will choose the same $\tilde{W}_t(j)$. From (11b), it follows that

$$W_t = \left[(1 - \zeta_w) \tilde{W}_t^{-1/\lambda_w} + \zeta_w \left((\pi_* e^\gamma)^{1-\iota_w} (\pi_{t-1} e^{z_{t-1}})^{\iota_w} W_{t-1} \right)^{-1/\lambda_w} \right]^{-\lambda_w}. \quad (20)$$

Finally, we assume that there is a complete set of state contingent securities in nominal terms, which implies that the Lagrange multiplier $\Xi_t^p(j)$ associated with (15) must be the same for all households in all periods and across all states of nature. This in turn implies that in equilibrium, households will make the same choice of consumption, money demand, investment, and capital utilization. Because the amount of leisure will differ across households due to wage rigidity, separability between labor and consumption in the utility function is key for this result.

2.5 Government Policies

The central bank follows a nominal interest rate rule by adjusting its instrument in response to deviations of inflation and output from their respective target levels,

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*} \right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*} \right)^{\psi_1} \left(\frac{Y_t}{Y_t^*} \right)^{\psi_2} \right]^{1-\rho_R} e^{\sigma_R \epsilon_{R,t}}, \quad (21)$$

where $\epsilon_{R,t}$ is the monetary policy shock, R^* is the steady-state nominal rate, Y_t^* is the target level of output, and the parameter ρ_R determines the degree of interest rate smoothing. This specification of the Taylor rule is more standard than that used by Smets and Wouters (2003), who introduced a time-varying inflation objective that varies stochastically according to a random walk. The random-walk inflation target may help the model fit the medium- and long-frequency fluctuations in inflation. In this article we are interested in assessing the model's fit of inflation without the extra help from the exogenous inflation target shocks. We set the target level of out-

put Y_t^* in (21) equal to the trend level of output $Y_t^* = Z_t Y^*$, where Y^* is the steady state of the model expressed in terms of detrended variables. The central bank supplies the money demanded by the household to support the desired nominal interest rate. We also considered an alternative specification in which the central bank targets the level of output that would have prevailed in absence of nominal rigidities; unreported results indicate that this alternative specification leads to a deterioration of fit.

The government budget constraint is of the form

$$P_t G_t + R_{t-1} B_{t-1} + M_{t-1} = T_t + M_t + B_t, \quad (22)$$

where T_t are total nominal lump-sum taxes (or subsidies), aggregated across all households. Government spending is given by

$$G_t = (1 - 1/g_t) Y_t, \quad (23)$$

where g_t follows the exogenous process

$$\ln g_t = (1 - \rho_g) \ln g + \rho_g \ln g_{t-1} + \sigma_g \epsilon_{g,t}. \quad (24)$$

2.6 Resource Constraint

The aggregate resource constraint

$$C_t + I_t + a(u_t) \bar{K}_{t-1} = \frac{1}{g_t} Y_t \quad (25)$$

can be derived by integrating the budget constraint (15) across households and combining it with the government budget constraint (22) and the zero profit conditions of both labor packers and final good producers.

2.7 Model Solution

As in the work of Altig, Christiano, Eichenbaum, and Lindé (2004) our model economy evolves along stochastic growth path. Output Y_t , consumption C_t , investment I_t , the real wage W_t/P_t , physical capital \bar{K}_t , and effective capital K_t all grow at the rate Z_t . Nominal interest rates R_t , inflation π_t , and hours worked L_t are stationary. The model can be rewritten in terms of detrended variables. We find the steady states for the detrended variables and use the method of Sims (2002) to construct a log-linear approximation of the model around the steady state. All subsequent statements about the DSGE model are statements about its log-linear approximation. We collect all of the DSGE model parameters in the vector θ , stack the structural shocks in the vector ϵ_t , and derive a state-space representation for the $n \times 1$ vector Δy_t ,

$$\Delta y_t = [\Delta \ln Y_t, \Delta \ln C_t, \Delta \ln I_t, \ln L_t, \Delta \ln (W_t/P_t), \pi_t, R_t]',$$

where Δ denotes the temporal difference operator.

3. DSGE–VARs AS TOOLS FOR MODEL EVALUATION

In addition to the DSGE model, we consider a VAR specification for y_t . The VAR is written in vector error-correction form as

$$\Delta y_t = \Phi_0 + \Phi_\beta (\beta' y_{t-1}) + \Phi_1 \Delta y_{t-1} + \dots + \Phi_p \Delta y_{t-p} + u_t. \quad (26)$$

We assume that the vector of reduced-form innovations is normally distributed conditional on past information, $u_t \sim \mathcal{N}(0, \Sigma_u)$. The normality assumption is common in the likelihood-based analysis of VARs, albeit mostly for convenience. According to the DSGE model, the technology process Z_t generates common trends in output, consumption, investment, and real wages. We impose this common-trend structure on the VAR by including the error-correction term

$$\beta' y_{t-1} = [\ln C_{t-1} - \ln Y_{t-1}, \ln I_{t-1} - \ln Y_{t-1}, \ln(W_{t-1}/P_{t-1}) - \ln Y_{t-1}]'$$

on the right side of (26). We denote the dimension of Δy_t by n , define the $k \times 1$ vector $x_t = [1, (\beta' y_{t-1})', \Delta y_{t-1}', \dots, \Delta y_{t-p}']'$, and let $\Phi = [\Phi_0, \Phi_\beta, \Phi_1, \dots, \Phi_p]'$.

VARs are widely used in empirical macroeconomics and often serve as benchmarks for the evaluation of dynamic equilibrium economies. We borrow from the indirect inference literature (e.g., Gouriéroux et al. 1993; Smith 1993) and use the VAR as an approximating model for the DSGE model. We construct a mapping from the DSGE model parameters to the VAR parameters. As is well known, the DSGE model leads to a restricted VAR approximation. We interpret deviations of the VAR parameters from the cross-coefficient restrictions as DSGE model misspecification. Although the approach described here is also applicable if the DSGE model is solved with nonlinear techniques, we use a log-linear approximation in our empirical analysis, as discussed in Section 2.7.

So far, the VAR in (26) is written in reduced form. To obtain a structural VAR, we express the one-step-ahead forecast errors u_t as a function of the shocks ϵ_t that appear in the DSGE model described previously,

$$u_t = \Sigma_{\pi} \Omega \epsilon_t, \quad (27)$$

where Σ_{π} is the (unique) Cholesky decomposition of Σ_u and Ω is an orthonormal matrix. It is well known that Ω is not identifiable from the data, because the likelihood function of the VAR depends only on the covariance matrix $\Sigma_u = \Sigma_{\pi} \Sigma_{\pi}'$.

Broadly speaking, the goals of our analysis are to obtain estimates of the DSGE model and the VAR parameters, to assess the magnitude of the DSGE model misspecification, and to learn from the discrepancy between restricted and unrestricted impulse response dynamics how to improve the specification of the DSGE model. The analysis is conducted in a Bayesian framework. Starting from a prior distribution for the DSGE model parameters θ , we use the mapping from θ to the VAR coefficients Φ and Σ_u to obtain a prior for the VAR parameters. Our prior is centered at the VAR approximation of the DSGE model, which we denote by $\Phi_*(\theta)$ and $\Sigma_u^*(\theta)$, but allows for deviations from DSGE model restrictions to account for potential misspecification. The precision of the prior is scaled by

a hyperparameter, λ . This hyperparameter generates a continuum of models, which we call DSGE–VAR(λ), that essentially has an unrestricted VAR at one extreme (λ is near 0) and the VAR approximation of the DSGE model at the other extreme ($\lambda = \infty$). (By “model,” we mean a joint probability distribution for the data and parameters.)

To obtain a prior distribution for Ω , we define a function $\Omega^*(\theta)$ of the DSGE model parameters in Section 3.5. Roughly speaking, this function has the following property: Combining $\Omega^*(\theta)$ with the reduced-form VAR approximation of the DSGE model results in a structural VAR that mimics the impulse response dynamics of the DSGE model. Unlike for the reduced-form VAR parameters, we do not allow Ω to deviate from $\Omega^*(\theta)$. Thus, conditional on the DSGE model parameters, our prior for Ω degenerates to a point mass. This implies that we take the DSGE model literally in the directions of the VAR parameter space in which the data are uninformative.

Overall, we are constructing a joint prior distribution for the VAR and DSGE model parameters that has the following hierarchical structure:

$$p(\theta, \Phi, \Sigma_u, \Omega | \lambda) = p(\theta) p(\Phi, \Sigma_u, \Omega | \theta, \lambda). \quad (28)$$

This prior is combined with the VAR likelihood function $p(Y | \Phi, \Sigma_u)$ to obtain a joint posterior distribution

$$p(\theta, \Phi, \Sigma_u, \Omega | Y, \lambda) = \frac{p(Y | \Phi, \Sigma_u) p(\theta) p(\Phi, \Sigma_u, \Omega | \theta, \lambda)}{p(Y | \lambda)}. \quad (29)$$

We specify a grid $\Lambda = \{l_1, \dots, l_q\}$ for the hyperparameter λ . If we assign prior probabilities $\pi_{j,0}$ to the grid points l_j , then posterior odds are given by

$$\frac{\pi_{i,0} p(Y | \lambda = l_i)}{\pi_{j,0} p(Y | \lambda = l_j)}.$$

We use Markov chain Monte Carlo (MCMC) methods to conduct posterior inference. Rather than specify an explicit prior distribution for λ , we simply interpret the marginal likelihood function of λ , $p(Y | \lambda)$, as an overall measure of fit and denote its peak by $\hat{\lambda}$. A large value of $\hat{\lambda}$ and a likelihood ratio of $\lambda = \hat{\lambda}$ versus $\lambda = \infty$ close to 1 is interpreted as evidence in favor of the DSGE model restrictions. Impulse response comparisons of DSGE–VAR(∞) and DSGE–VAR($\hat{\lambda}$) can generate insights into the sources of DSGE model misspecification.

Our approach is related to recent work by Gallant and McCulloch (2004), who proposed a Bayesian framework for indirect inference. In their analysis, the approximating model is mainly a device for obtaining a likelihood function in a setting where it is computationally cumbersome to evaluate the underlying structural model. In our analysis we use the approximating model mainly as a tool to relax DSGE model restrictions and obtain an empirical specification that fits well and can serve as a benchmark for impulse response comparisons.

In the remainder of this section we define the VAR approximation of the DSGE model at which our prior is centered (Sec. 3.1), motivate the specification of the prior distribution $p(\Phi, \Sigma, \Omega | \theta, \lambda)$ as a summary of beliefs about potential DSGE model misspecification (Sec. 3.2), characterize the posterior distribution of VAR and DSGE model parameters (Sec. 3.3), explore the properties of the marginal likelihood function $p(Y | \lambda)$

(Sec. 3.4), and propose a mapping $\Omega^*(\theta)$ to obtain identification and enable construction of identified impulse responses from the DSGE-VAR (Sec. 3.5). Because the likelihood function is invariant to Ω , the choice of $\Omega^*(\theta)$ does not affect the joint posterior distribution of θ , Φ , and Σ_u . Therefore, we drop Ω from the notation in Sections 3.1–3.4 and begin with the analysis of the reduced-form specification.

3.1 Vector Autoregressive Approximation of the Dynamic Stochastic General Equilibrium Model

Assuming that under the DSGE model, the distribution of x_t is stationary with a nonsingular covariance matrix (both conditions are satisfied for the model specified in Sec. 2), we define the moments $\Gamma_{YY}(\theta) = \mathbb{E}_\theta^D[\Delta y_t \Delta y_t']$, $\Gamma_{XX}(\theta) = \mathbb{E}_\theta^D[x_t x_t']$, and $\Gamma_{XY}(\theta) = \mathbb{E}_\theta^D[x_t \Delta y_t']$ and use a population regression to obtain the mapping from DSGE model to VAR parameters,

$$\begin{aligned} \Phi^*(\theta) &= \Gamma_{XX}^{-1}(\theta) \Gamma_{XY}(\theta) \quad \text{and} \\ \Sigma_u^*(\theta) &= \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta) \Gamma_{XX}^{-1}(\theta) \Gamma_{XY}(\theta). \end{aligned} \quad (30)$$

Here $\Gamma_{YX} = \Gamma'_{XY}$. We refer to $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ as restriction functions used to center the prior distribution $p(\Phi, \Sigma_u | \theta, \lambda)$.

3.2 Misspecification and Bayesian Inference

If the VAR representation of Δy_t deviates from the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$, then the DSGE model is misspecified. A key step in our analysis is the formulation of a prior distribution for the discrepancy between Φ and $\Phi^*(\theta)$, which we denote by Φ^Δ . We use a prior with density decreasing in Φ^Δ , implying that large misspecifications have low probabilities. This assumption reflects the belief that the DSGE model provides a good (albeit not perfect) approximation of reality. We use an information-theoretic metric to assess the magnitude of Φ^Δ . This metric allows us to develop a fairly general evaluation procedure, subsequently keeping the computational burden manageable.

To fix ideas, we begin by (a) ignoring the dependence of Φ^* on θ and (b) imposing that $\Sigma_u = \Sigma_u^*$. Suppose that we generate a sample of λT observations from the DSGE model, collected in the matrices Y_* and X_* . Our prior for Φ^Δ has the property that its density is proportional to the expected likelihood ratio of $\Phi^* + \Phi^\Delta$ versus Φ^* . The log-likelihood ratio is given by

$$\begin{aligned} & \ln \left[\frac{\mathcal{L}(\Phi^* + \Phi^\Delta, \Sigma_u^* | Y_*, X_*)}{\mathcal{L}(\Phi^*, \Sigma_u^* | Y_*, X_*)} \right] \\ &= -\frac{1}{2} \text{tr} \left[\Sigma_u^{*-1} (\Phi^{\Delta'} X_*' X_* \Phi^\Delta + 2 \Phi^{*'} X_*' X_* \Phi^\Delta \right. \\ & \quad \left. - 2(\Phi^* + \Phi^\Delta)' X_*' Y_* + 2 \Phi^{*'} X_*' Y_*) \right], \end{aligned} \quad (31)$$

where Y_* denotes the $\lambda T \times n$ matrix with rows $y_t^{*'}$ and X_* is the $\lambda T \times k$ matrix with rows $x_t^{*'}$. Taking expectations under the distribution generated by the DSGE model yields

$$\begin{aligned} & \mathbb{E}_\theta^D \left[\ln \left[\frac{\mathcal{L}(\Phi^* + \Phi^\Delta, \Sigma_u^* | Y_*, X_*)}{\mathcal{L}(\Phi^*, \Sigma_u^* | Y_*, X_*)} \right] \right] \\ &= -\frac{1}{2} \text{tr} \left[\Sigma_u^{*-1} (\Phi^{\Delta'} \lambda T \Gamma_{XX} \Phi^\Delta) \right]. \end{aligned} \quad (32)$$

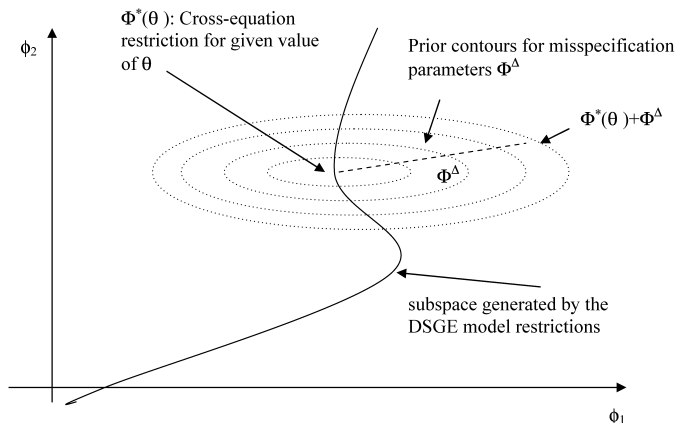


Figure 1. Stylized View of DSGE Model Misspecification. $\Phi = [\phi_1, \phi_2]$ can be interpreted as the VAR parameters, and $\Phi^*(\theta)$ is the restriction function implied by the DSGE model.

We now choose a prior density that is proportional (\propto) to the expected likelihood ratio,

$$p(\Phi^\Delta | \Sigma_u^*) \propto \exp \left\{ -\frac{1}{2} \text{tr} \left[\lambda T \Sigma_u^{*-1} (\Phi^{\Delta'} \Gamma_{XX} \Phi^\Delta) \right] \right\}. \quad (33)$$

As the sample size λT increases, the prior places more mass on misspecification matrices that are close to 0. A graphical illustration is provided in Figure 1.

In the empirical application, we allow for uncertainty about θ by specifying a prior with density $p(\theta)$ and take potential misspecification of the covariance matrix $\Sigma_u^*(\theta)$ into account. T will correspond to the size of the actual sample, and λ is a hyperparameter that controls the expected magnitude of the deviations from the DSGE model restrictions. Conditional on θ , our prior for the VAR coefficients takes the form

$$\Sigma_u | \theta, \lambda \sim \mathcal{IW}(\lambda T \Sigma_u^*(\theta), \lambda T - k), \quad (34)$$

$$\Phi | \Sigma_u, \theta, \lambda \sim \mathcal{N} \left(\Phi^*(\theta), \frac{1}{\lambda T} [\Sigma_u^{-1} \otimes \Gamma_{XX}(\theta)]^{-1} \right),$$

where \mathcal{IW} denotes the inverted Wishart distribution. This prior distribution is proper (i.e., has mass 1) provided that $\lambda T \geq k + n$. Thus we restrict the domain of λ to the interval $[(k + n)/T, \infty]$. The prior is identical to that used in earlier work (Del Negro and Schorfheide 2004), but its motivation is different. The earlier work focused on the improvement of VARs and emphasized mixed estimation based on artificial data from a DSGE model and actual data. In this article we ask the opposite question: How can we relax DSGE model restrictions and evaluate the extent of their misspecification?

3.3 Posterior Distributions

The posterior density is proportional to the product of the prior density and the likelihood function. We factorize the posterior into the conditional density of the VAR parameters given the DSGE model parameters and the marginal density of the DSGE model parameters,

$$p(\Phi, \Sigma_u, \theta | Y, \lambda) = p(\Phi, \Sigma_u | Y, \theta, \lambda) p(\theta | Y, \lambda). \quad (35)$$

The actual observations are collected in the matrices Y and X , with the subscript λ indicating the dependence of the posterior

on the hyperparameter. We use $\hat{\Gamma}_{XX}$, $\hat{\Gamma}_{XY}$, and $\hat{\Gamma}_{XX}$ to denote the sample autocovariances such as $\frac{1}{T} \sum x_t x_t'$. It is straightforward to show (e.g., Zellner 1971) that the posterior distribution of Φ and Σ is also of the inverted Wishart–normal form,

$$\begin{aligned} \Sigma_u | Y, \theta, \lambda \\ \sim \mathcal{IW}(T(\lambda + 1) \hat{\Sigma}_{u,b}(\theta), T(\lambda + 1) - k), \end{aligned} \quad (36)$$

$$\begin{aligned} \Phi | Y, \Sigma_u, \theta, \lambda \\ \sim \mathcal{N}(\hat{\Phi}_b(\theta), \Sigma_u \otimes [T(\lambda \Gamma_{XX}(\theta) + \hat{\Gamma}_{XX})]^{-1}), \end{aligned}$$

where $\hat{\Phi}_b(\theta)$ and $\hat{\Sigma}_{u,b}(\theta)$ are given by

$$\hat{\Phi}_b(\theta) = (\lambda \Gamma_{XX}(\theta) + \hat{\Gamma}_{XX})^{-1} (\lambda \Gamma_{XY}(\theta) + \hat{\Gamma}_{XY})$$

and

$$\begin{aligned} \hat{\Sigma}_{u,b}(\theta) = \frac{1}{(\lambda + 1)} [(\lambda \Gamma_{YY}(\theta) + \hat{\Gamma}_{YY}) - (\lambda \Gamma_{YX}(\theta) + \hat{\Gamma}_{YX}) \\ \times (\lambda \Gamma_{XX}(\theta) + \hat{\Gamma}_{XX})^{-1} (\lambda \Gamma_{XY}(\theta) + \hat{\Gamma}_{XY})]. \end{aligned}$$

Thus the larger the weight λ of the prior, the closer the posterior mean of the VAR parameters is to $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$, the values that respect the cross-equation restrictions of the DSGE model. On the other hand, if λ equals the lower bound $(n + k)/T$, then the posterior mean is close to the ordinary least squares (OLS) estimate $\hat{\Gamma}_{XX}^{-1} \hat{\Gamma}_{XY}$. The formula for the marginal posterior density of θ and the description of a MCMC algorithm that generates draws from the joint posterior of Φ , Σ_u , and θ have been provided in earlier work (Del Negro and Schorfheide 2004), where we also demonstrated (prop. 2) that under certain conditions, the estimate of θ can be interpreted as the minimum distance estimate obtained by projecting the VAR coefficient estimates back onto the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$.

3.4 The Marginal Likelihood Function of λ

We study the fit of the DSGE model by examining the marginal likelihood function of the hyperparameter λ , defined as

$$p(Y|\lambda) = \int p(Y|\theta, \Sigma, \Phi) p(\theta, \Sigma, \Phi|\lambda) d(\theta, \Sigma, \Phi). \quad (37)$$

We use Geweke's (1999a) modified harmonic mean estimator to obtain a numerical approximation of the marginal likelihood function based on the output of the MCMC computations. For computational reasons, we consider only a finite set of values $\Lambda = \{l_1, \dots, l_q\}$, where $l_1 = (n + k)/T$ and $l_q = \infty$. If we assign equal prior probabilities to the elements of Λ , then the posterior probabilities for the hyperparameter are proportional to the marginal likelihood. Thus we also refer to $p(Y|\lambda)$ as the posterior of λ and denote its mode by

$$\hat{\lambda} = \arg \max_{\lambda \in \Lambda} p(Y|\lambda). \quad (38)$$

It is common in the literature (e.g., Smets and Wouters 2003) to use marginal data densities to document the fit of DSGE models relative to VARs with diffuse priors. In our framework this approach corresponds (approximately) to comparing $p(Y|\lambda)$ for the extreme values of λ , that is, $\lambda = \infty$ (DSGE model) and $\lambda = (k + n)/T$ (VAR with a nearly flat prior). It is preferable to report the entire marginal likelihood function $p(Y|\lambda)$ rather than

just its endpoints. The function $p(Y|\lambda)$ summarizes the time series evidence on model misspecification and documents by how much the restrictions of the DSGE model must be relaxed to balance in-sample fit and model complexity.

To illustrate the properties of the marginal likelihood function $p(Y|\lambda)$, it is instructive to consider the following univariate example. Suppose that the VAR takes the special form of an AR(1) model,

$$y_t = \phi y_{t-1} + u_t, \quad u_t \sim \text{iid } \mathcal{N}(0, 1), \quad (39)$$

and the DSGE model restricts ϕ to be equal to ϕ^* . We denote the DSGE model implied autocovariances of order 0 and 1 by γ_0 and γ_1 . Moreover, $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are sample autocovariances based on T observations. The prior in (34) simplifies to

$$\phi \sim \mathcal{N}\left(\phi^*, \frac{1}{\lambda T \gamma_0}\right). \quad (40)$$

For this simple model, the marginal likelihood of λ takes the form

$$\ln p(Y|\lambda, \phi^*) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \tilde{\sigma}^2(\lambda, \phi^*) - \frac{1}{2} c(\lambda, \phi^*). \quad (41)$$

The term $\tilde{\sigma}^2(\lambda, \phi^*)$ measures the in-sample one-step-ahead forecast error and can be written as

$$\tilde{\sigma}^2(\lambda, \phi^*) = \hat{\gamma}_0 + \lambda \gamma_0 - \frac{(\hat{\gamma}_1 + \lambda \gamma_1)^2}{\hat{\gamma}_0 + \lambda \gamma_0} - \lambda \left(\gamma_0 - \frac{\gamma_1^2}{\gamma_0} \right). \quad (42)$$

It can be verified that as λ approaches 0, $\tilde{\sigma}^2(\lambda, \phi^*)$ converges to the OLS forecast error, whereas as $\lambda \rightarrow \infty$, we obtain the in-sample forecast error under the restriction $\phi = \phi^*$. Formally,

$$\begin{aligned} \lim_{\lambda \rightarrow 0} \tilde{\sigma}^2(\lambda, \phi^*) &= \frac{1}{T} \sum (y_t - \hat{\phi} y_{t-1})^2 \quad \text{and} \\ \lim_{\lambda \rightarrow \infty} \tilde{\sigma}^2(\lambda, \phi^*) &= \frac{1}{T} \sum (y_t - \phi^* y_{t-1})^2, \end{aligned}$$

where $\hat{\phi} = \hat{\gamma}_1 / \hat{\gamma}_0$. Moreover, $\tilde{\sigma}^2(\lambda, \phi^*)$ is monotonically increasing in λ ; that is, the larger the λ , the worse the in-sample fit. The third term in (41) can be interpreted as a penalty for model complexity and is of the form

$$c(\lambda, \phi^*) = \ln \left(1 + \frac{\hat{\gamma}_0}{\lambda \gamma_0} \right). \quad (43)$$

In the context of a standard regressor selection problem, model complexity is tied to the number of included regressors, and the penalty is an increasing function of the number of parameters being estimated. In our setup, model complexity is a continuous function of the hyperparameter λ . If $\lambda = \infty$, then there is no parameter to estimate in the AR(1) example, and the complexity (or, alternatively, the dimensionality) of the model is 0. If $\lambda = 0$, then the autoregressive parameter is completely unrestricted, and the dimensionality is 1. Accordingly, the penalty term (43) is monotonically decreasing in λ . As λ approaches 0 and the prior becomes more diffuse, the penalty diverges to infinity.

Several features of the marginal data density are noteworthy. First, the marginal likelihood function is monotonically decreasing, is increasing, or has an interior maximum. If an interior maximum exists, it is given by

$$\hat{\lambda} = \frac{\gamma_0 \hat{\gamma}_0^2}{T(\hat{\gamma}_0 \gamma_1 - \gamma_0 \hat{\gamma}_1)^2 - \gamma_0^2 \gamma_0}. \quad (44)$$

Thus if the sample autocovariances differ significantly from the autocovariances derived under the restriction $\phi = \phi^*$, then the marginal likelihood peaks at a small value of λ . As the discrepancy between sample and DSGE model autocovariances decreases, $\hat{\lambda}$ increases, and the marginal likelihood will eventually attain its maximum at $\hat{\lambda} = \infty$.

Second, as λ approaches 0, the marginal log-likelihood function tends to minus infinity. In the context of high-dimensional VARs this feature of the marginal likelihood function enforces parsimony and prevents the use of overparameterized specifications that cannot be estimated precisely based on the fairly small samples available to macroeconomists. In these cases, a naive posterior odds comparison of VAR and DSGE model based on the endpoints of the marginal likelihood function, corresponding to a VAR with diffuse prior (small λ) and a VAR with DSGE model restrictions imposed, may not be very informative, because it tends to favor the restricted specification. This phenomenon arises more generally in Bayesian posterior odds comparisons and is called Lindley's paradox. Rather than limiting the attention to extremes, our procedure creates a continuum of prior distributions and evaluates the marginal likelihood function for a range of hyperparameter values. The magnitudes of $\hat{\lambda}$ and $p(Y|\lambda = \hat{\lambda}, \phi^*)/p(Y|\lambda = \infty, \phi^*)$ provide measures of overall fit of the DSGE model.

Third, consider the comparison of two models \mathcal{M}_1 and \mathcal{M}_2 . In the context of our univariate example, these models correspond to different restrictions, say $\phi_{(1)}^*$ and $\phi_{(2)}^*$. In our empirical analysis we compare the marginal likelihood functions associated with different DSGE model specifications. For small values of λ , the goodness-of-fit terms $\tilde{\sigma}^2(\lambda, \phi_{(1)}^*)$ and $\tilde{\sigma}^2(\lambda, \phi_{(2)}^*)$ are essentially identical, and differences in marginal likelihoods are due to differences in the penalty terms. For large values of λ , in contrast, penalty differentials are less important, and the marginal likelihood comparison is driven by the relative in-sample fit of the two restricted specifications. If the autocovariances associated with \mathcal{M}_1 are closer to the sample autocovariances than to the \mathcal{M}_2 autocovariances, then, according to (44), $\hat{\lambda}_{(1)}$ tends to be larger than $\hat{\lambda}_{(2)}$.

3.5 Impulse Response Function Comparisons

The goal of our impulse response function comparisons is to document in which dimensions the DSGE model dynamics are (in)consistent with the data. An extensive literature evaluates DSGE models by comparing their impulse responses to those obtained from VARs (e.g., Cogley and Nason 1994; Rotemberg and Woodford 1997; Schorfheide 2000; Boivin and Giannoni 2006; Christiano et al. 2005).

Impulse response comparisons face two challenges. First, for the VAR to be a meaningful benchmark, it must attain a higher posterior probability than the DSGE model. In a Bayesian framework, the odds of a VAR versus the DSGE model are updated by the ratio of marginal likelihoods for the two specifications. Marginal likelihood functions in turn measure the time series fit of a model, adjusted for its complexity. Many authors are using simple least squares techniques to estimate unconstrained, high-dimensional VAR systems. Due to their complexity, these VARs typically attain much lower marginal likelihoods than DSGE models, and it would be incoherent from

a Bayesian perspective to use them as a benchmark for DSGE model evaluations. From a frequentist perspective, the imprecise VAR coefficient estimates translate into impulse response function estimates that in a mean squared error sense are worse than the estimates obtained directly from the DSGE model.

Second, for the comparison to be insightful from an economic perspective, the VAR must be expressed in terms of structural shocks. It is typically difficult to find identification schemes that are consistent with the DSGE model and simultaneously identify an entire vector of structural shocks in a high-dimensional VAR.

In the DSGE-VAR procedure, the benchmark is given by DSGE-VAR($\hat{\lambda}$), the model that attains the highest marginal likelihood. Therefore, by construction, our procedure meets the first challenge: the benchmark model attains a better fit—penalized for model complexity to avoid overparameterization—and tends to deliver more reliable impulse response estimates than the restrictive DSGE model. The spirit of our evaluation is to keep the autocovariance sequence associated with the benchmark model as close as possible to the DSGE model without sacrificing the ability to track the historical time series. Next, we describe how the DSGE-VAR analysis can address the second challenge, identification.

To compare impulse response functions, we need to characterize the matrix Ω that appears in (27) and provides the link between reduced-form and structural innovations in the VAR. We follow our earlier work (Del Negro and Schorfheide 2004) and construct a restriction function $\Omega^*(\theta)$ as follows. The state-space representation of the DSGE model is identified in the sense that for each value of θ , there is a unique matrix $A_0(\theta)$ that determines the contemporaneous effect of ϵ_t on Δy_t . Using a QR factorization of $A_0(\theta)$, the initial response of Δy_t to the structural shocks can be uniquely decomposed into

$$\left(\frac{\partial \Delta y_t}{\partial \epsilon_t'} \right)_{DSGE} = A_0(\theta) = \Sigma_{tr}^*(\theta) \Omega^*(\theta), \quad (45)$$

where $\Sigma_{tr}^*(\theta)$ is lower-triangular and $\Omega^*(\theta)$ is orthonormal. According to (26), the initial impact of ϵ_t on the endogenous variables Δy_t in the VAR is given by

$$\left(\frac{\partial \Delta y_t}{\partial \epsilon_t'} \right)_{VAR} = \Sigma_{tr} \Omega. \quad (46)$$

To identify the DSGE-VAR, we maintain the triangularization of its covariance matrix Σ_u and replace the rotation Ω in (46) with the function $\Omega^*(\theta)$, which appears in (45).

Using the rotation matrix $\Omega^*(\theta)$, we turn the reduced-form DSGE-VAR into an identified DSGE-VAR. Conditional on θ , our prior for Ω takes the form of a point mass at $\Omega^*(\theta)$. The marginal distribution of Ω is updated indirectly, as we learn about the DSGE model parameters θ from the data. Because beliefs about the VAR parameters are centered around the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$, our prior implies, roughly speaking, that beliefs about impulse responses to structural shocks are centered around the DSGE model responses, even for small values of the hyperparameter λ . However, the smaller the λ , the wider the probability intervals for the response functions. Our approach differs from much of the empirical literature on identified VARs because it closely ties identification to the underlying DSGE model. We do not view this feature as

a shortcoming. Because the premise of our analysis is that the DSGE model provides a good (albeit not perfect) approximation of reality, strong views about the identification of particular structural shocks can and should be directly incorporated into the underlying DSGE model.

Two pairwise comparisons of impulse responses are interesting: (a) the DSGE model versus DSGE-VAR(∞) and (b) DSGE-VAR(∞) versus DSGE-VAR($\hat{\lambda}$). In our application we are working with a log-linearized DSGE model that can be expressed a vector autoregressive moving average (VARMA). The first comparison provides insight into the accuracy of the VAR approximation, whereas the second comparison provides insight into the dimensions in which the DSGE model is misspecified. If the DSGE model's moving average (MA) polynomial is noninvertible or has roots near the unit circle, then the approximation by a finite-order VAR could be poor. In contrast, if the MA polynomial is well approximated by a few AR terms, then our identification procedure for the DSGE-VAR is able to recover the DSGE model responses associated with the VARMA representation. Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2007) provided necessary and sufficient conditions for the invertibility of the MA components of linear state-space models. In our application we find that for parameter values of θ near the posterior mode, the discrepancy between DSGE and DSGE-VAR(∞) responses is fairly small, particularly in the short run. However, as in the indirect inference literature, our analysis remains coherent and insightful even if the VAR provides only an approximation to the underlying DSGE model.

Comparing the DSGE-VAR($\hat{\lambda}$) and DSGE-VAR(∞) responses illustrates the discrepancy between the coefficient estimates that optimally relax the DSGE model restrictions and the restricted estimates. If the posterior estimates of the VAR parameters are close to the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$, then the DSGE-VAR($\hat{\lambda}$) and DSGE-VAR(∞) will be very similar. If, on the other hand, the posterior estimates strongly deviate from the restriction function, then the discrepancy between the impulse responses potentially provides valuable insight into how to improve the underlying DSGE model.

4. THE DATA

All data were obtained from Haver Analytics (Haver mnemonics are in italics). Real output, consumption of nondurables and services, and investment (defined as gross private domestic investment plus consumption of durables) are obtained by dividing the nominal series (GDP , $C - CD$, and $I + CD$) by population 16 years and older ($LN16N$) and deflating using the chained-price GDP deflator ($JGDP$). The real wage is computed by dividing compensation of employees ($YCOMP$) by total hours worked and the GDP deflator. Note that compensation per hours includes wages as well as employer contribution; it accounts for both wage and salary workers and proprietors. Our measure of hours worked is computed by taking total hours worked reported in the National Income and Product Accounts (NIPA), which is at an annual frequency, and interpolating it using growth rates computed from hours of all persons in the nonfarm business sector ($LXNFH$). Our broad measure of hours worked is consistent with our definition of both wages and output in the economy. We divide hours worked by $LN16N$ to convert them into per capita terms. We then take the

log of the series multiplied by 100, so that all figures can be interpreted as percentage changes in hours worked. All growth rates are computed using quarter-to-quarter log differences and then multiplied by 100 to convert them into percentages. Inflation rates are defined as log differences of the GDP deflator and converted into annualized percentages. The nominal rate corresponds to the effective Federal funds rate ($FFED$), also in percent. Data are available for QIII:1954–QI:2004.

5. EMPIRICAL RESULTS

The empirical analysis is presented in four parts. The first part reports on the prior and posterior distributions for the DSGE model parameters. The second part discusses the evidence of misspecification in the new Keynesian model. We calculate marginal likelihood functions for the hyperparameter λ and study the discrepancy in the impulse responses to monetary and technology shocks between the DSGE-VAR($\hat{\lambda}$) and the DSGE-VAR(∞). In the third part, we use the DSGE-VAR framework for the comparison of different DSGE model specifications. We strip the baseline model of some of its frictions (habit formation and price/wage indexation) and investigate to what extent the time series fit suffers as a consequence. Finally, we report some results on pseudo-out-of-sample forecasting accuracy.

Unless noted otherwise, all results are based on 30 years of observations ($T = 120$), starting in QII:1974 and ending in QI:2004. We used the same sample size in the pseudo-out-of-sample forecasting exercise. Beginning in QIII:1954, we constructed 58 rolling samples of 120 observations, estimate the DSGE-VARs as well as the state-space representation of the DSGE model for each sample, and compute forecast error statistics. All MCMC results are based on 110,000 draws from the relevant posterior distribution, discarding the first 10,000. We checked whether 110,000 draws were sufficient by repeating the MCMC computations from overdispersed starting points, verifying that we obtained the same results for parameter estimates and log-marginal likelihood functions.

The lag length p of the DSGE-VAR is 4. To make the DSGE-VAR estimates comparable to the estimates of the state-space representation of the DSGE model, in both cases we used likelihood functions that condition on the four observations needed to initialize lags in period $t = 1$ as well as on the cointegration vector $\beta'y_0$. Because DSGE-VAR(∞) is not equivalent to the state-space representation of the DSGE model, we adopt the convention that whenever we refer to the estimation of the DSGE model, we mean its state-space representation.

5.1 Priors for the Dynamic Stochastic General Equilibrium Parameters

Priors for the DSGE model parameters are provided in the first four columns of Table 1. All intervals reported in the text are 90% probability intervals. The priors for the degree of price and wage stickiness, ζ_p and ζ_w , are both centered at .6, implying that firms and households reoptimize their prices and wages on average every two-and-half quarters. The 90% interval is very wide and encompasses findings in microlevel studies of price adjustments, such as that of Bils and Klenow (2004). The

Table 1. The DSGE Model Parameter Estimates

Distribution		Prior			DSGE-VECM($\hat{\lambda}$) posterior		DSGE posterior	
		P(1)	P(2)	Interval	Mean	Interval	Mean	Interval
α	\mathcal{B}	.33	.10	[.16, .49]	.23	[.20, .26]	.26	[.23, .29]
ζ_p	\mathcal{B}	.60	.20	[.29, .93]	.79	[.72, .86]	.83	[.79, .87]
ι_p	\mathcal{B}	.50	.28	[.08, .95]	.75	[.53, 1.00]	.76	[.57, .97]
s''	\mathcal{G}	4.00	1.50	[1.60, 6.28]	4.57	[2.60, 6.61]	5.70	[3.34, 7.90]
h	\mathcal{B}	.70	.05	[.62, .78]	.75	[.70, .81]	.81	[.77, .85]
a'	\mathcal{G}	.20	.10	[.05, .35]	.27	[.10, .43]	.19	[.07, .32]
v_l	\mathcal{G}	2.00	.75	[.81, 3.15]	1.69	[.66, 2.74]	2.09	[.95, 3.19]
ζ_w	\mathcal{B}	.60	.20	[.29, .94]	.79	[.70, .87]	.89	[.84, .93]
ι_w	\mathcal{B}	.50	.28	[.05, .93]	.45	[.04, .80]	.70	[.47, .96]
r^*	\mathcal{G}	2.00	1.00	[.49, 3.49]	1.36	[.41, .28]	1.52	[.48, 2.50]
ψ_1	\mathcal{G}	1.50	.40	[.99, 2.09]	1.80	[1.42, 2.19]	2.21	[1.79, 2.63]
ψ_2	\mathcal{G}	.20	.10	[.05, .35]	.16	[.09, .22]	.07	[.03, .10]
ρ_r	\mathcal{B}	.50	.20	[.18, .83]	.76	[.70, .83]	.82	[.78, .86]
π^*	\mathcal{N}	3.01	1.50	[.56, 5.46]	2.98	[.89, 5.19]	5.98	[4.61, 7.38]
γ	\mathcal{G}	2.00	1.00	[.46, 3.47]	1.08	[.39, 1.80]	.94	[.40, 1.43]
λ_f	\mathcal{G}	.15	.10	[.01, .29]	.35	[.29, .42]	.29	[.24, .34]
g^*	\mathcal{G}	.30	.10	[.14, .46]	.19	[.13, .24]	.23	[.20, .26]
L_{adj}	\mathcal{N}	252.0	10.0	[235.5, 268.4]	257.6	[244.3, 271.5]	245.2	[233.5, 255.3]
ρ_z	\mathcal{B}	.20	.10	[.04, .35]	.20	[.08, .32]	.20	[.09, .31]
ρ_ϕ	\mathcal{B}	.60	.20	[.29, .93]	.38	[.20, .58]	.25	[.11, .37]
ρ_{λ_f}	\mathcal{B}	.60	.20	[.28, .93]	.11	[.03, .21]	.12	[.02, .21]
ρ_μ	\mathcal{B}	.80	.05	[.72, .88]	.74	[.68, .81]	.87	[.81, .94]
ρ_b	\mathcal{B}	.60	.20	[.29, .93]	.80	[.68, .92]	.92	[.86, .97]
ρ_g	\mathcal{B}	.80	.05	[.72, .88]	.90	[.85, .96]	.95	[.93, .97]
σ_z	\mathcal{IG}	.75	2.00	[.31, 2.34]	.57	[.48, .65]	.82	[.72, .91]
σ_ϕ	\mathcal{IG}	4.00	2.00	[1.64, 12.57]	11.83	[4.41, 19.84]	40.54	[18.21, 64.09]
σ_{λ_f}	\mathcal{IG}	.75	2.00	[.31, 2.34]	.21	[.18, .25]	.24	[.21, .28]
σ_μ	\mathcal{IG}	.75	2.00	[.30, 2.33]	.55	[.43, .67]	.66	[.54, .78]
σ_b	\mathcal{IG}	.75	2.00	[.30, 2.33]	.32	[.24, .41]	.54	[.36, .71]
σ_g	\mathcal{IG}	.75	2.00	[.31, 2.34]	.30	[.26, .34]	.38	[.34, .42]
σ_r	\mathcal{IG}	.20	2.00	[.08, .62]	.18	[.15, .21]	.28	[.25, .31]

NOTE: See Section 2 for a definition of the DSGE model parameters, and Section 4 for a description of the data. \mathcal{B} represents beta; \mathcal{G} , gamma; \mathcal{IG} , inverse gamma; and \mathcal{N} , normal distribution. $P(1)$ and $P(2)$ denote means and standard deviations for the \mathcal{B} , \mathcal{G} , and \mathcal{N} distributions; s and v do so for the \mathcal{IG} distribution, where $p_{\mathcal{IG}}(\sigma|v, s) \propto \sigma^{-v-1} e^{-vs^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region, and the prior probability interval reflects this truncation. All probability intervals are 90% credible. The following parameters are fixed: $\delta = .025$, $\lambda_w = .3$, and $\mathcal{F} = 0$. Estimation results are based on the sample period Q1:1974–Q1:2004.

priors for the degree of price and wage indexation, ι_p and ι_w , are nearly uniform over the unit interval. The prior for the adjustment cost parameter s'' is taken from Smets and Wouters (2003) and is consistent with the values that Christiano et al. (2005) used when matching DSGE impulse response functions to consumption and investment, among other variables, to VAR responses.

Our prior for the habit persistence parameter h is centered at .7, which is the value used by Boldrin, Christiano, and Fisher (2001). Those authors found that $h = .7$ enhances the ability of a standard DSGE model to account for key asset market statistics. The prior for a' implies that in response to a 1% increase in the return to capital, utilization rates rise by .1% to .3%. These numbers are considerably smaller than that used by Christiano et al. (2005). The 90% interval for the prior distribution on v_l implies that the Frisch labor supply elasticity lies between .3 and 1.3, reflecting the microlevel estimates at the lower end and the estimates of Kimball and Shapiro (2003) and Chang and Kim (2006) at the upper end.

We use a presample of observations from Q1:1960–Q1:1974 to choose the prior means for the parameters that determine steady states. The prior mean for the technology growth rate is 2% per year. The annualized steady-state inflation rate lies between .5% and 5.5%, and the prior for the inverse of the discount factor r^* implies a growth-adjusted real interest rate of 4% on average. The prior means for the capital share α , the substitution parameter λ_f , and the steady-state government

share $1 - 1/g$ are chosen to capture the labor share of .57, the investment-to-output ratio of .24, and the government share of .21 in the presample. The distribution for ψ_1 and ψ_2 is approximately centered at Taylor's (1993) values, whereas the smoothing parameter lies in the range .18–.83.

Because we model the level of technology Z_t as a unit root process, the prior for ρ_z , which measures the serial correlation of technology growth z_t , is centered at .2. The priors for ρ_μ (shocks to the capital accumulation equation) and ρ_g (government spending) are quite tight around .8 to prevent these parameters from hitting the boundary. The priors for the remaining autocorrelation coefficients of the structural shocks— ρ_ϕ (preferences of leisure), ρ_b (overall preference shifter), and ρ_{λ_f} (price markup shocks)—are fairly diffuse and centered around .6. Finally, the priors for the standard deviation parameters are chosen to obtain realistic magnitudes for the implied volatility of the endogenous variables. Throughout the analysis, we fix the capital depreciation rate $\delta = .025$ and $\lambda_w = .3$. The parameter λ_w affects the substitution elasticity between different types of labor. Unlike λ_f , it is not identifiable from the steady-state relationships. We introduce a parameter, L_{adj} , that captures the units of measured hours worked. In our model we choose ϕ such that in steady state, each household supplies one unit of labor. A prior for L_{adj} is chosen based on quarterly per capita hours worked in the presample. We assume that the parameters are a priori independent. Although this assumption is common in the literature, we make it mostly for convenience.

5.2 Posteriors for the Dynamic Stochastic General Equilibrium Parameters

The remaining columns of Table 1 report on the posterior estimates of the DSGE model parameters for both the DSGE model and the estimation of the DSGE- $\text{VAR}(\hat{\lambda})$. As described later in detail, for the sample beginning in QII:1974, the value of $\hat{\lambda}$ is 1.25. We start by focusing on the parameter estimates for the state-space representation of the DSGE model. The comparison of the 90% coverage intervals suggests that likelihood contains information about most of the parameters. Three exceptions are the parameters a' , v_l , and ρ_z , for which prior and posterior intervals roughly overlap. The parameter estimates for the DSGE model are also generally in line with those of Smets and Wouters (2005), which is not surprising because our model specification and choice of prior are similar to theirs. In particular, the model displays a relatively high degree of price and wage stickiness, as measured by the probability that firms (wage setters) cannot change their price (wage) in a given period. The posterior means of ζ_p and ζ_w are .83 and .89. The estimated degree of indexation is about .7 for both prices and wages. For some of the structural shocks, notably ϕ_t and $\lambda_{f,t}$, the degree of persistence is not as high as that given by Smets and Wouters (2005).

We now turn to the parameter estimates obtained from the DSGE- $\text{VAR}(\hat{\lambda})$. In earlier work (Del Negro and Schorfheide 2004) we showed that as the prior on the VAR parameters becomes more diffuse, information about the DSGE model parameters accumulates more slowly. In the limit, when $\lambda = 0$, the DSGE- $\text{VAR}(\lambda)$ likelihood contains no information about the parameter vector θ , and the posterior will be identical to the prior. Thus in general, we expect that for $\hat{\lambda} < \infty$, the DSGE- $\text{VAR}(\hat{\lambda})$ posteriors will be closer to the prior than the DSGE model posterior. Table 1 confirms that for many of the parameters (including the degree of price and wage stickiness, the policy parameters, and some of the autocorrelation coefficients), the DSGE- $\text{VAR}(\hat{\lambda})$ estimates indeed lie between the DSGE posterior and the prior distribution. One exception are the standard deviations of the structural shocks, which are estimated to be lower under DSGE- $\text{VAR}(\hat{\lambda})$ than under the DSGE model regardless of the prior.

5.3 Evidence of Misspecification in the New Keynesian Model

Smets and Wouters (2003, table 2) found that for Euro-area data, a large-scale new Keynesian DSGE models can attain a larger marginal likelihood than VARs with training sample prior and specific versions of the Minnesota prior. This result has had a considerable impact on applied macroeconomists and policy-makers, because it suggests that new Keynesian DSGE models have achieved a degree of sophistication that makes them competitive with more densely parameterized models, such as VARs. In this section we revisit the findings of Smets and Wouters using the DSGE- VAR procedure. We make three distinct points based on marginal likelihood functions and impulse response comparisons. First, the posterior odds of a DSGE model versus a VAR with a fairly diffuse prior do not provide a particularly robust assessment of fit. Small changes in the sample period can lead to reversals of the model ranking. The

DSGE- VAR analysis, on the other hand, is much less sensitive to changes in the sample period. Second, there is strong evidence of misspecification in the new Keynesian model, suggesting that forecasts and policy recommendations obtained from this class of models should be viewed with some degree of skepticism. Finally, on the positive side, we find that accounting for misspecification by optimally relaxing the DSGE model restrictions does not alter the responses to a monetary policy and technology shocks in any significant way, either qualitatively or quantitatively. Thus, despite its deficiencies, the new Keynesian DSGE model can indeed generate realistic predictions of the effects of unanticipated changes in monetary policy and technology shocks.

5.3.1 The Marginal Likelihood Function of λ . Figure 2 shows the logarithm of the marginal likelihood of DSGE- $\text{VAR}(\lambda)$ for different values of λ , as well as for the DSGE model. The values of λ considered are .33 (the smallest λ value for which we have a proper prior), .5, .75, 1, 1.25, 1.5, 2, 5, Inf, DSGE.

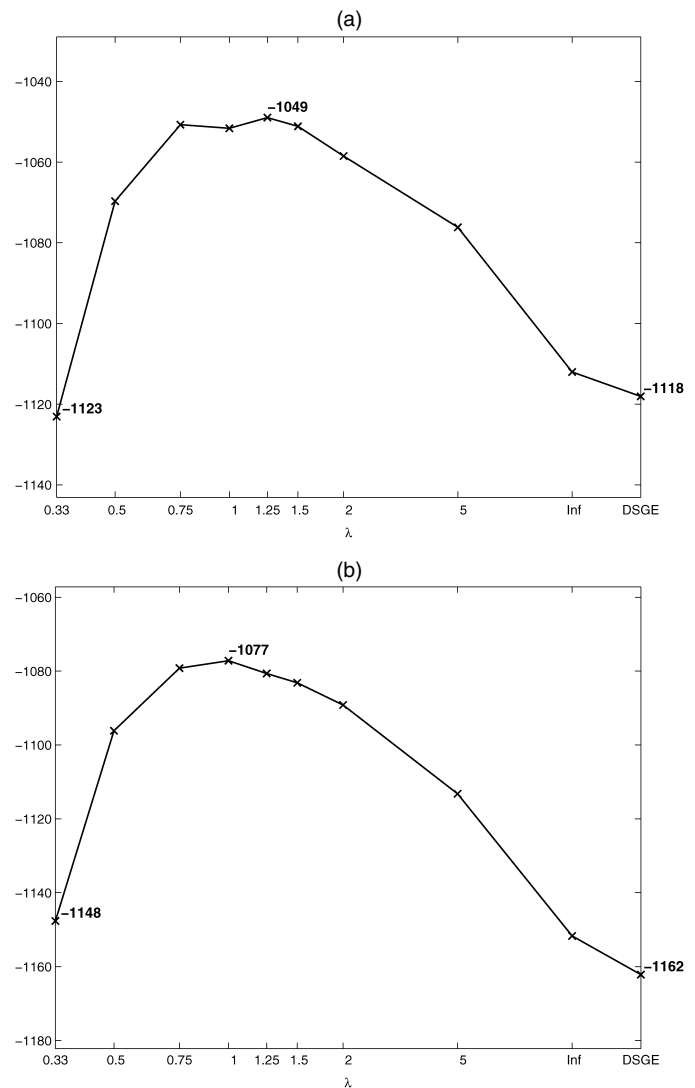


Figure 2. Marginal Likelihood as a Function of λ . (a) 30-year sample: QII:1974–QI:2004. (b) 30-year sample: QII:1970–QI:2000. The two panels depict the log-marginal likelihood function on the y-axis and the corresponding value of λ , rescaled between via the transformation $\lambda/(1+\lambda)$, on the x-axis. The right endpoint depicts the log-marginal likelihood for the state-space representation of the DSGE model.

and ∞ . We rescale the x -axis according to $x = \lambda / (1 + \lambda)$. Figure 2(a) depicts the marginal likelihood function for the 30-year sample beginning in QII:1974, which is the sample used for most of the subsequent analysis. Figure 2(b) is based on a 30-year sample starting 4 years earlier, in QII:1970.

The comparison between the two extremes—the VAR with loose prior on the left side of the plot and the DSGE model on the right side—leads to opposite conclusions depending on the sample period. In the QII:1974–QI:2004 sample, the difference in log-marginal likelihoods between the DSGE model and DSGE–VAR(.33) is 5, which translates into posterior odds of roughly 150 to 1 in favor of the DSGE model. Conversely, for the QII:1970–QI:2000 sample, the difference is -14 , overwhelmingly against the DSGE model. This result confirms Sims' (2003) conjecture that marginal likelihood comparisons among far-apart models are not robust. The four years of difference between the two samples are very unlikely to contain major shifts in the economy and thus should not cause a change in the DSGE model's assessment.

The lack of robustness in the comparison between the two extremes contrasts with the robustness of the overall shape of the marginal likelihood function. In both panels, this function has an inverted U-shape. The marginal likelihood increases sharply as λ moves from .33 to .75, is roughly flat for values between .75 and 1.25, and subsequently decreases, first gradually and then more rapidly, as λ exceeds 1.5. The substantial drop in marginal likelihood between DSGE–VAR($\hat{\lambda}$) and DSGE–VAR(∞) is strong evidence of misspecification for the new Keynesian model: As the prior tightly concentrates in the neighborhood of the cross-equation restrictions imposed by the DSGE model, the in-sample fit of the DSGE–VAR deteriorates. Earlier (Del Negro and Schorfheide 2006) we showed that the shape of the posterior distribution of λ is roughly the same for all of the 58 30-year rolling samples considered in the forecasting exercise in Section 5.4. Therefore, the evidence of misspecification for the new Keynesian model is robust to the choice of the sample.

This inverted-U shape with peaks between .75 and 1.25 contrasts with the pattern that we would expect were the data generated by the DSGE model. The AR(1) example in Section 3.4 suggests that if the sample autocovariances were close to the population autocovariances implied by the DSGE model, then the marginal likelihood function would peak at a much larger value of λ and possibly be monotonically increasing. This is confirmed by simulation results reported by An and Schorfheide (2006), who generated observations from a small-scale DSGE model and then calculated marginal likelihood functions for λ that are indeed monotone in λ .

5.3.2 Impulse Response Function Comparisons. To gain further insight into the misspecification of the DSGE model, we proceed by comparing impulse responses from the DSGE–VAR(∞) to our benchmark specification DSGE–VAR($\hat{\lambda}$). It turns out that in our application, the approximation error of the DSGE–VAR(∞) relative to the state-space representation of the DSGE model is small (see the App.). Consequently, the impulse responses from the DSGE–VAR(∞)—in particular to a technology and a monetary policy shock—are very similar to those from the DSGE model.

We subsequently focus on the impulse response functions that have received the most attention in the literature: responses

to monetary policy and technology shocks. The full set of 49 response functions is given in the Appendix. Figure 3 depicts mean responses to one-standard-deviation shocks for the DSGE–VAR(∞) (gray solid lines), the DSGE–VAR($\hat{\lambda}$) (dark dashed–dotted lines), and 90% bands (dark dotted lines) for DSGE–VAR($\hat{\lambda}$). The responses are computed based on the respective posterior draws for the DSGE–VAR(∞) and DSGE–VAR($\hat{\lambda}$).

Figure 3(a) shows that the impulse response functions with respect to a monetary policy shock for DSGE–VAR(∞) match those for DSGE–VAR($\hat{\lambda}$), not only qualitatively but also, by and large, quantitatively. Both in the DSGE–VAR(∞) and in the DSGE–VAR($\hat{\lambda}$) output, consumption, investment, and hours display a hump-shaped response to the policy shock, although quantitatively, the hump for investment is more pronounced in the data than it is in the DSGE model. Unlike that of Christiano et al. (2005), our DSGE model implies that monetary policy shocks are observed contemporaneously. Yet, thanks to various sources of inertia, including habit formation, the initial impact of the shock on real variables is very small. The response of inflation is the only dimension in which DSGE model and data disagree; according to the DSGE–VAR($\hat{\lambda}$), it is more sluggish than in the DSGE model. In summary, as reported by Christiano et al. (2005), we find that the DSGE model's impulse response to a policy shock is in agreement with the data. On the one hand, this finding may not be too surprising, given that this specific model was written with this purpose in mind. On the other hand, unlike Christiano et al., we do not estimate the DSGE model by minimizing the discrepancy between the DSGE and the VAR's impulse responses, and, moreover, we use a different benchmark and identification procedure. Yet we find that their result is robust.

Figure 3(a) shows that the responses to a technology shock have similar shapes for the DSGE–VAR(∞) and DSGE–VAR($\hat{\lambda}$), but they appear to be quantitatively different. The technology shock seems to have a greater effect in the DSGE–VAR(∞). The amplification is due to a larger estimate of the shock standard deviation caused by poorer in-sample fit of the DSGE–VAR(∞) relative to the DSGE–VAR($\hat{\lambda}$). The differences between the response functions disappear if the technology shocks in the two models are renormalized to have the same long-run effect on output.

According to the analysis of Altig et al. (2004), inflation in the DSGE model essentially does not move in response to a permanent technology shock. We find that it does. Moreover, the inflation response is consistent with our benchmark impulse response function obtained from the DSGE–VAR($\hat{\lambda}$). We conjecture that this difference is due to the estimation procedure used. Altig et al. estimated their DSGE model by matching impulse response functions. Technology shocks in their VAR are identified through long-run restrictions which tend to be imprecisely estimated; thus, when minimizing the discrepancy between VAR and DSGE responses, more weight is placed on the responses to the monetary shocks. But, as Figure 3(a) shows, in the data inflation reacts with a delay to the monetary shock; therefore, a sluggish response of inflation is wired into their estimates, translating into a sluggish response to a technology shock as well. Our likelihood-based estimation implicitly places more weight on reproducing the response of inflation to a technology shock.

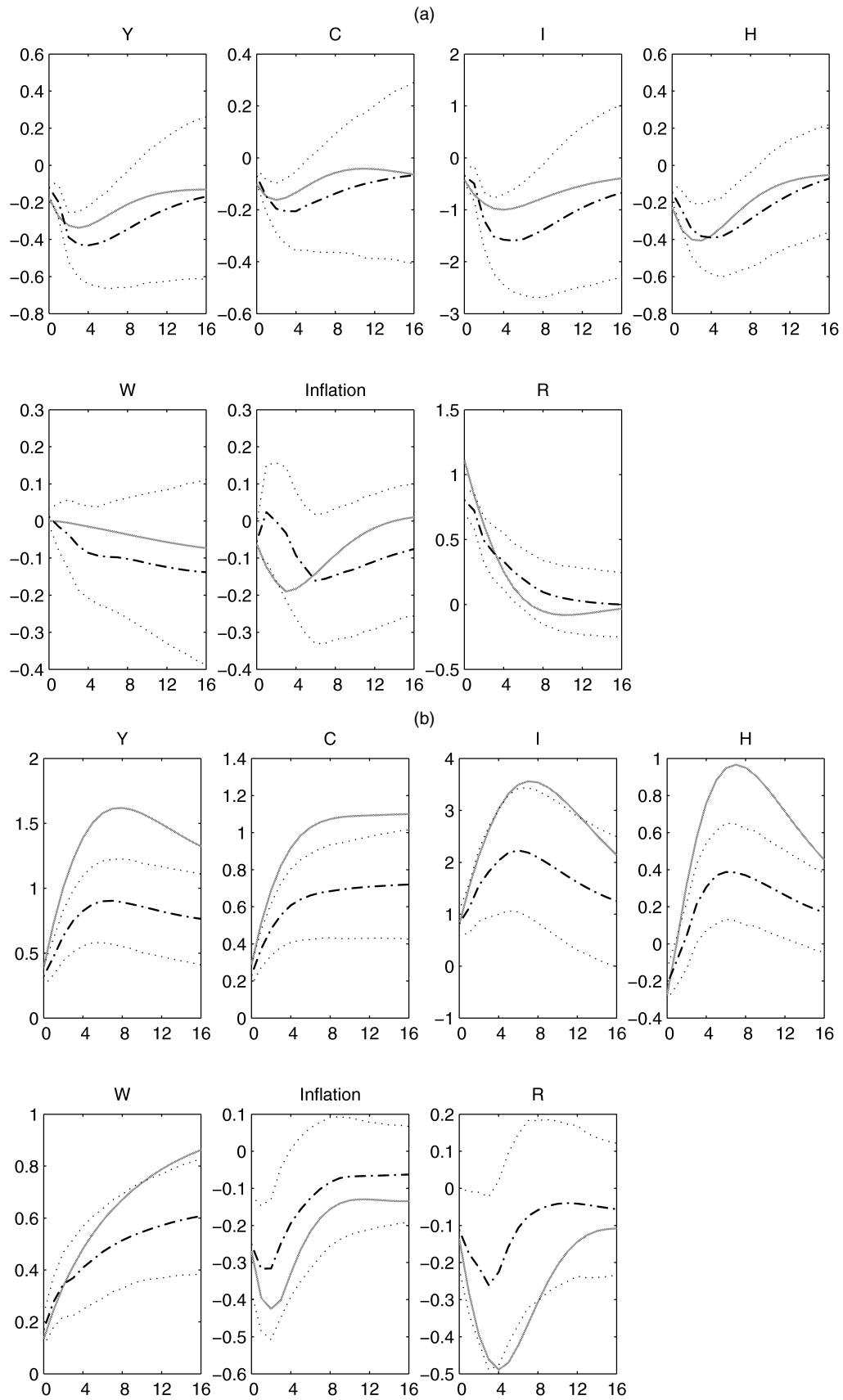


Figure 3. Impulse Response Functions: DSGE- $\text{VAR}(\hat{\lambda})$ versus DSGE- $\text{VAR}(\infty)$. (a) Monetary policy shocks. (b) Technology shocks. This figure depicts posterior mean responses for the DSGE- $\text{VAR}(\infty)$ (gray solid lines) and the DSGE- $\text{VAR}(\hat{\lambda})$ (dark dashed-dotted lines), and 90% bands (dark dotted lines) for DSGE- $\text{VAR}(\hat{\lambda})$. Y, C, I, and W denote the percentage quarterly growth rates in real output, consumption, investment, and real wages. Inflation is annualized inflation. H is the log level of per-capita hours (times 100), and R is the Fed funds rate in percent. For Y, C, I, and W, the impulse responses are cumulative.

In conclusion, we find that the DSGE model's misspecification does not translate into impulse responses to monetary policy or technology shocks differ greatly between the DSGE model and the benchmark DSGE- $\text{VAR}(\hat{\lambda})$. Many macroeconomists believe that these two shocks provide a very important source of business cycle fluctuations. Our results suggests that business cycle research has to a large extent been successful in developing a model that can produce realistic responses to these shocks; however, a nonnegligible fraction of fluctuations is attributed to the remaining five shocks in the model. We document in the Appendix that for some of the shocks, such as μ_t , which affects the shadow price of installed capital, DSGE- $\text{VAR}(\infty)$ and DSGE- $\text{VAR}(\hat{\lambda})$ differ substantially, particularly in the long run, suggesting that some low-frequency implications of the model are at odds with the data.

5.4 Comparing Dynamic Stochastic General Equilibrium Model Specifications

The DSGE model used in this article is rich in terms of nominal and real frictions. An important aspect of the empirical analysis of Smets and Wouters (2003) and Christiano et al. (2005) is assessing which of these frictions are important to fit the data. Smets and Wouters (2003) used marginal likelihood comparisons, eliminating one friction at a time and computing posterior odds relative to the baseline specification. Christiano et al. (2005) studied whether the impulse responses of a model without a specific friction can match the VAR's impulse responses as well as the baseline model.

In this article we use DSGE-VARs to assess the importance of two particular features of the DSGE model: price and wage indexation and habit formation. We refer to the model without wage and price indexation as the *no indexation* model and to the model without habit formation as the *no habit* model, whereas we call the standard DSGE model used up to now the *baseline* model. We document that habit formation is important to fit the data, whereas the evidence in favor of indexation is weak.

We compare the marginal likelihood of λ for the baseline model with that of the two alternative specifications. Our example given in Section 3.4 suggests that as the mismatch between sample autocovariances and population autocovariances implied by the DSGE model increases, $\hat{\lambda}$ decreases, and the marginal likelihood function shifts downward. Therefore, we can infer from the magnitude of the south-west shift in the marginal likelihood function the extent to which a specific friction is useful in fitting the data.

We emphasized previously that in the absence of a more elaborate DSGE model, a comparison of impulse responses between the DSGE- $\text{VAR}(\infty)$ and DSGE- $\text{VAR}(\hat{\lambda})$ can generate important insight into how to improve the model specification. Using the hindsight from our analysis of the *baseline* model, we subsequently examine whether such a comparison for the *no indexation* and *no habit* models reveals the directions in which these models need to be augmented.

5.4.1 Evidence From the Marginal Likelihood Functions. Figure 4 resembles Figure 2(a), except that we overlay the marginal likelihood functions for the *baseline* (solid line), the *no indexation* (dashed line), and the *no habit* (dashed-dotted line) model. Smets and Wouters (2003) dogmatically enforced the

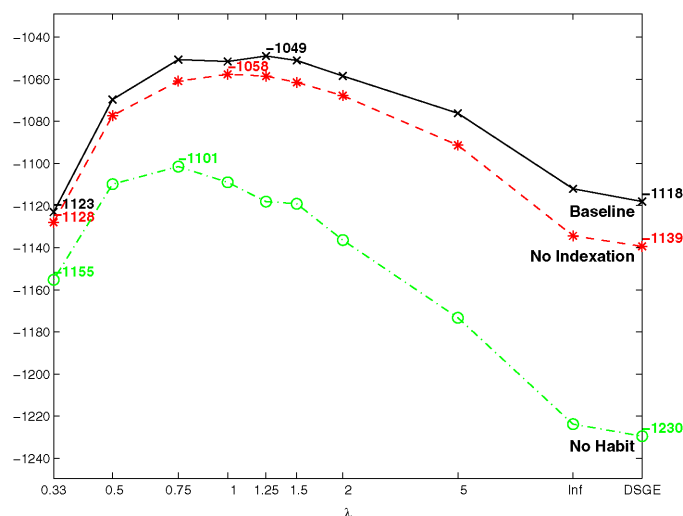


Figure 4. Marginal Likelihood as a Function of λ : Comparison Across Models. See Figure 2 for an explanation.

cross-equation restrictions of the DSGE model specifications, which leads to a comparison of the three marginal likelihood values on the right edge of Figure 4. Both alternative specifications are strongly rejected in favor of the *baseline*, even though the rejection for the *no indexation* is not as stark as that for the *no habit* model.

The evidence contained in the overall posterior distribution of λ against the *no habit* model is equally strong. Figure 4 shows that relative to the *baseline* model, the marginal likelihood of λ shifts not only down, but also to the left. Translating the marginal likelihood values into posterior probabilities, for the *no habit* model there is very little probability mass associated with values of $\lambda > 1$. Conversely, the leftward shift for the *no indexation* model is much less pronounced, and the marginal likelihood remains fairly flat for values of λ between .75 and 2.

5.4.2 Evidence From Impulse Response Functions. Suppose that all we have available is the *no habit* (*no indexation*) model. Can we see from the impulse response comparison between the DSGE- $\text{VAR}(\infty)$ and DSGE- $\text{VAR}(\hat{\lambda})$ that some important feature is missing from the structural model? Figure 5 depicts the mean impulse responses to monetary policy [Fig. 5(a)] and technology shocks [Fig. 5(b)] for DSGE- $\text{VAR}(\infty)$ (gray solid line) and DSGE- $\text{VAR}(\hat{\lambda})$ (dark dash-and-dotted lines), as well as the 90% bands (dark dotted lines) for DSGE- $\text{VAR}(\hat{\lambda})$. Figure 5 is obtained based on the *no habit* model; therefore, the benchmark DSGE- $\text{VAR}(\hat{\lambda})$ in Figure 5 differs from that in Figure 3 for two reasons. First, the value of $\hat{\lambda}$ is lower, as can be appreciated from Figure 4. Second, the prior for the VAR coefficients is based on the *no habit* model as opposed to the *baseline* model.

Comparing Figures 5 and 3 indicates that the initial responses to a monetary policy shock of output, consumption, and hours for the *no habit* DSGE model look very different from those of the *baseline* DSGE model. All real variables, with the exception of investment and real wages, now display a strong initial reaction to the monetary shock, which contrasts with the hump-shaped responses in the DSGE- $\text{VAR}(\hat{\lambda})$. Even if Figure 3 were not available to the researcher, the comparison between the impulse responses for $\lambda = \infty$ and $\lambda = \hat{\lambda}$ in Figure 5 would reveal

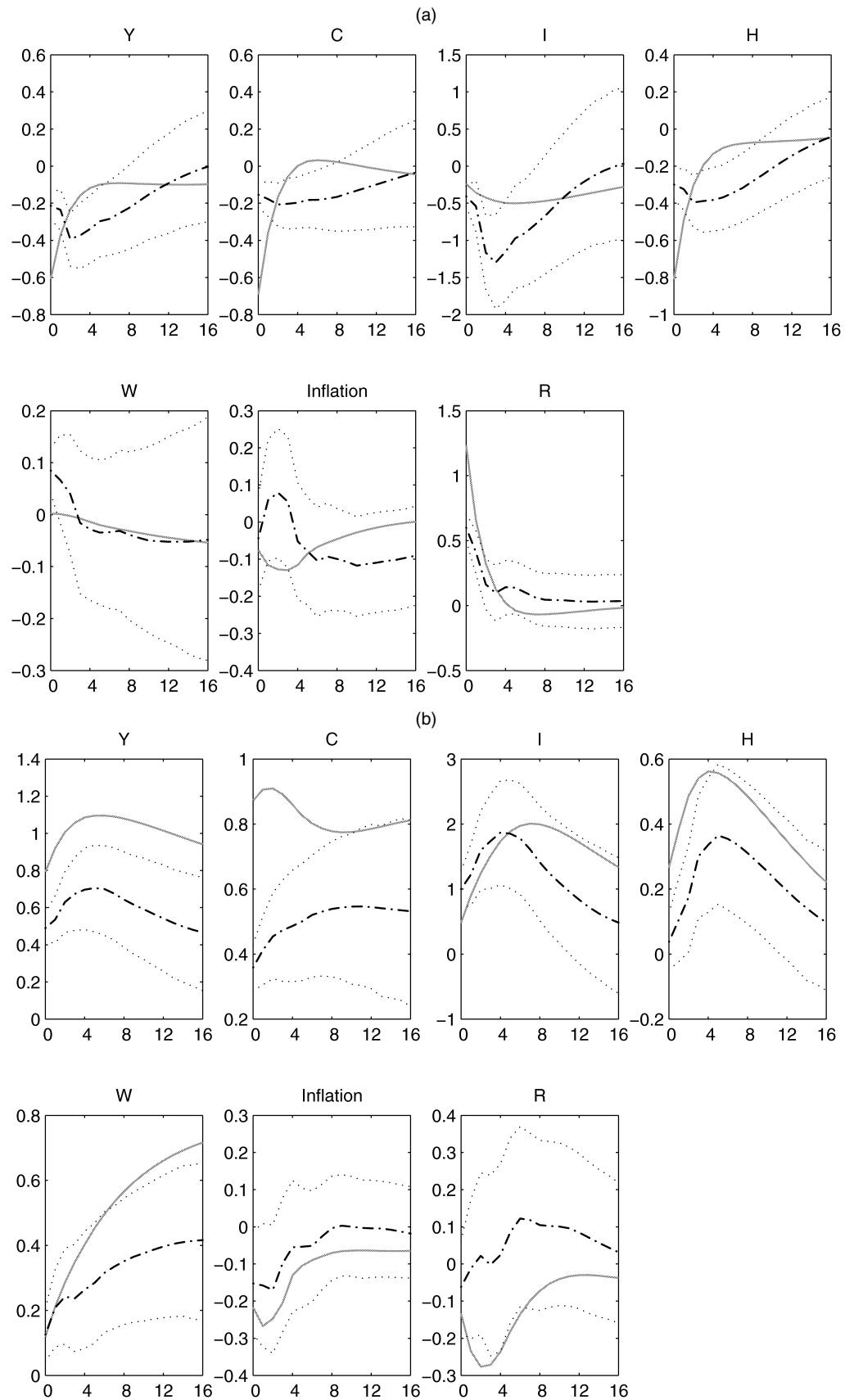


Figure 5. Impulse Response Functions for the no Habit Model: $DSGE-VAR(\hat{\lambda})$ versus $DSGE-VAR(\infty)$. (a) Monetary policy shocks. (b) Technology shocks. See Figure 3 for an explanation.

that something is amiss in DSGE model without habit formation. A similar analysis applies to the responses to a technology shock [Fig. 5(b)], where consumption reacts strongly on impact according to DSGE- $\text{VAR}(\infty)$, compared with the more gradual response in the DSGE- $\text{VAR}(\hat{\lambda})$. Importantly, the benchmark responses in Figures 5 and 3 are similar, both qualitatively and quantitatively, despite the fact that the underlying set of cross-equation restrictions is different. Thus, even under *no habit*, the DSGE- $\text{VAR}(\hat{\lambda})$ provides a reasonable benchmark, although the DSGE model misspecification is seemingly stronger than for the *baseline* model.

Figure 6 shows the impulse responses for the *no indexation* model. Unlike Figure 5, Figure 6 shows no stark divergence between DSGE- $\text{VAR}(\infty)$ and the benchmark, DSGE- $\text{VAR}(\hat{\lambda})$. Indeed, the impulse response functions in Figure 6 are quite similar to those of Figures 3. The change in the cross-equation restrictions does not seem to translate into an appreciable change in the transmission mechanism of monetary policy and technology shocks. Perhaps the main difference consists of the response of inflation to technology shocks, which is somewhat hump-shaped in Figure 3 but not in Figure 6. Quantitatively, however, this difference does not amount to much, because the hump is small.

In conclusion, the evidence from the DSGE-VAR procedure against the *no indexation* model is not nearly as strong as that against the *no habit* model. These findings suggest that habit persistence in preferences substantially improves the fit of the DSGE model. Thus those who believe that habit persistence is not a “structural” feature may have to introduce alternative mechanisms that deliver similar effects. Simply eliminating habit persistence comes at a cost in terms of fit. In contrast, the evidence in favor of price and wage indexation is not nearly as strong, despite the fact that the marginal likelihood comparison between DSGE models (Fig. 4), if taken literally, rejects the *no indexation* model in favor of the *baseline* model.

5.5 Pseudo-Out-of-Sample Forecast Accuracy

We now discuss the pseudo-out-sample fit of DSGE- $\text{VAR}(\infty)$ and compare it with that of the DSGE- $\text{VAR}(\hat{\lambda})$ and an unrestricted VAR. The out-of-sample forecasting accuracy is assessed based on a rolling sample starting in QIV:1985 and ending in QI:2000, for a total of 58 periods. At each date of the rolling sample, we use the previous 120 observations to reestimate the models and the following eight quarters to assess forecasting accuracy, which is measured by the root mean squared error (RMSE) of the forecast. For the variables that enter the VAR in growth rates (output, consumption, investment, real wage) and inflation, we forecast cumulative changes. For instance, the RMSE of inflation for eight-quarters-ahead forecasts measures the error in forecasting cumulative inflation over the next 2 years (in essence, average inflation), as opposed to quarter-to-quarter inflation in 2 years. The DSGE-VARs are reestimated for each of the 58 samples. As discussed earlier, the value of $\hat{\lambda}$ hovers between .75 and 1.25.

Table 2 documents for each series and forecast horizon the RMSE of the unrestricted VAR, as well as the percentage improvement in forecasting accuracy (whenever positive) of DSGE- $\text{VAR}(\hat{\lambda})$ and DSGE- $\text{VAR}(\infty)$ relative to the VAR. The

last three rows of the table report the corresponding figures for the multivariate statistic, a summary measure of joint forecasting performance, which is computed as the converse of the log-determinant of the variance-covariance matrix of forecast errors, divided by 2 to convert from variance to standard error and by the number of variables to obtain an average figure. The percentage improvement in the multivariate statistic across models is computed by taking the difference multiplied by 100.

Table 2 shows that for the multivariate statistic, and for most variables, DSGE- $\text{VAR}(\hat{\lambda})$ improves over the VAR for all forecasting horizons. Short-run consumption forecasts and long-run investment forecasts are exceptions. Interestingly, there seems to be a trade-off between forecasting consumption and investment. This trade-off reflects the fact that all three models considered in Table 2 are error-correction models with the same long-run cointegrating restrictions on output, consumption, investment, and real wages. These cointegrating restrictions are at odds with the data. Thus accurate forecasts for some of these variables result in inaccurate forecasts for others, given that not all series grow proportionally in the long run as the model predicts. Another manifestation of this phenomenon is the fact that DSGE- $\text{VAR}(\infty)$ outperforms the other two models in forecasting the real wage in the long run, but performs very poorly in forecasting both output and investment. In summary, the fact that the DSGE model imposes these long-run cointegrating restrictions results in a serious limitation of its forecasting ability. To the extent that DSGE-VAR inherits the same long-run restrictions, its accuracy suffers as well.

For the remaining variables, DSGE- $\text{VAR}(\hat{\lambda})$ is roughly as accurate as the unrestricted VAR in terms of hours per capita, whereas DSGE- $\text{VAR}(\infty)$ is far worse, especially in the long run. Conversely, DSGE- $\text{VAR}(\infty)$ performs well in terms of the nominal variables, inflation, and interest rate. For inflation, the forecasting accuracy of DSGE- $\text{VAR}(\infty)$ is inferior to that of DSGE- $\text{VAR}(\hat{\lambda})$, but far better than that of the unrestricted VAR. For the nominal interest rate, DSGE- $\text{VAR}(\infty)$ outperforms DSGE- $\text{VAR}(\hat{\lambda})$ for longer forecast horizons, whereas in the short run, the two models have roughly the same forecasting performance.

Extending the analysis of Section 5.4, we now discuss the comparison of the out-of-sample forecasting performance across models. Figure 7 shows the one-quarter-ahead percentage improvement in the multivariate forecast statistic relative to the unrestricted VAR for the *baseline* (solid line), *no indexation* (dashed line), and the *no habit* (dash-and-dotted line) models, as a function of λ . Note that the benchmark used for the computation of the percentage improvement—the unrestricted VAR—is the same for all three models. Figure 7 focuses on one-period-ahead forecasting accuracy to facilitate a comparison with the results in Figure 4, which were based on the marginal likelihood.

The results in Figure 7 agree in a number of dimensions with those in Figure 4. The inverted-U shape that characterized the posterior distribution of λ for each of the model in Figure 4 also describes the improvement in forecasting accuracy relative to the VAR. Results documented earlier (Del Negro and Schorfheide 2006) showed that this inverted-U shape characterizes the improvement in forecasting accuracy for all forecasting horizons from one to eight quarters ahead. Relaxing, but not ignoring the cross-equation restrictions leads to an improvement

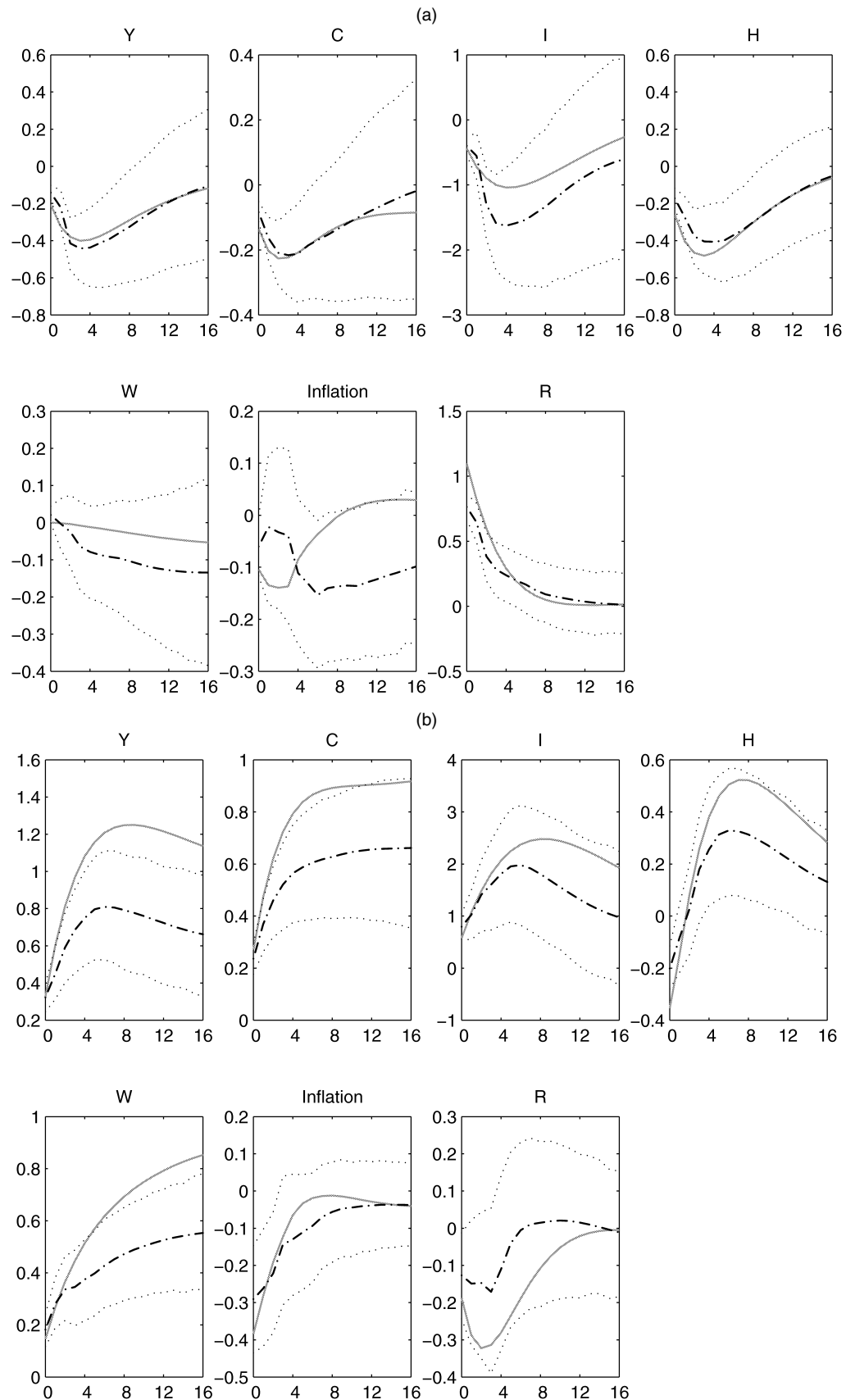


Figure 6. Impulse Response Functions for the no Indexation Model: DSGE-VAR($\hat{\lambda}$) versus DSGE-VAR(∞). (a) Monetary policy shocks. (b) Technology shocks. See Figure 3 for an explanation.

Table 2. Pseudo-Out-of-Sample RMSEs: Percentage Improvement Relative to VAR

				Forecast horizon				
				1	2	4	6	8
Y	DSGE-VAR($\hat{\lambda}$)			16.3	14.1	12.5	13.5	13.6
	DSGE-VAR(∞)			.9	-17.6	-56.5	-82.5	-102.9
	VAR,	RMSE:		.67	.97	1.68	2.38	2.98
C	DSGE-VAR($\hat{\lambda}$)			-6.8	-7.6	7.1	16.6	21.5
	DSGE-VAR(∞)			-15.7	-21.4	-8	11.3	12.0
	VAR,	RMSE:		.42	.62	1.06	1.56	2.03
I	DSGE-VAR($\hat{\lambda}$)			17.8	8.0	-5.0	-11.5	-17.2
	DSGE-VAR(∞)			-4.2	-41.2	-101.0	-135.3	-157.8
	VAR,	RMSE:		2.67	3.98	6.59	9.14	11.45
H	DSGE-VAR($\hat{\lambda}$)			10.0	10.9	-.6	-0	.7
	DSGE-VAR(∞)			-13.6	-37.9	-95.4	-116.5	-127.2
	VAR,	RMSE:		.58	.92	1.56	2.26	2.88
W	DSGE-VAR($\hat{\lambda}$)			8.2	11.7	11.1	14.9	18.4
	DSGE-VAR(∞)			6.7	12.7	18.1	27.0	36.6
	VAR,	RMSE:		.65	1.06	1.72	2.28	2.82
Inflation	DSGE-VAR($\hat{\lambda}$)			10.7	10.9	22.9	31.0	36.6
	DSGE-VAR(∞)			8.4	4.2	10.4	21.1	29.6
	VAR,	RMSE:		.25	.47	.98	1.68	2.42
R	DSGE-VAR($\hat{\lambda}$)			27.3	23.4	9.2	7.0	9.1
	DSGE-VAR(∞)			27.7	17.8	3.2	8.2	17.1
	VAR,	RMSE:		.68	1.14	1.63	2.11	2.64
Multivariate statistic	DSGE-VAR($\hat{\lambda}$)			11.0	8.8	6.1	9.4	9.4
	DSGE-VAR(∞)			3.8	-2.1	-6.9	-2.7	-.2
	VAR,	RMSE:		.68	.23	-.18	-.47	-.65

NOTE: Results are based on 58 rolling samples of 120 observations. For each rolling sample, we estimate DSGE model and DSGE-VARs, compute $\hat{\lambda}$, and calculate pseudo-out-of-sample forecast errors for the subsequent eight periods. For each variable, the table reports RMSE of the forecast from the VAR and improvements in forecast accuracy obtained by the DSGE model and the DSGE-VAR($\hat{\lambda}$). Improvements (positive entries) are measured by the percentage reduction in RMSE. The multivariate statistic is computed as the converse of the log-determinant of the variance-covariance matrix of forecast errors divided by 2 to convert from variance to standard error and by the number of variables to obtain an average figure. Percentage improvements are computed by taking the difference times 100. Y, C, I, and W denote the percentage quarterly growth rates in real output, consumption, investment, and real wages. H is the log level of per capita hours (times 100), and R is the Fed funds rate in percent. For Y, C, I, W, and Inflation, the RMSE is computed using the cumulative forecast error over the relevant horizon. The forecast horizon is measured in quarters.

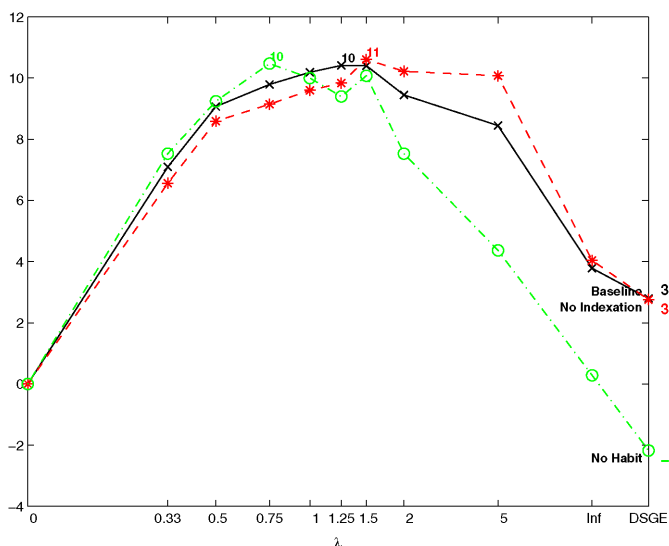


Figure 7. One-Period-Ahead RMSE Summary: Model Comparison. This figure depicts the improvement in the one-period-ahead multivariate statistic relative to an unrestricted VAR as a function of λ for three different models, the baseline model (solid line), the no indexation model (dashed line), and the no habit model (dashed-dotted line). The multivariate statistic is computed as the converse of the log-determinant of the variance-covariance matrix of forecast errors divided by 2 to convert from variance to standard error and by the number of variables to obtain an average figure. Percentage improvements are computed by taking the difference times 100.

in fit and forecasting performance. Consistent with the overall message from the previous section, the *no indexation* and the *baseline* models perform roughly as well in terms of multivariate statistic, whereas the forecasting accuracy worsens considerably for the *no habit* model relative to the *baseline* model as the DSGE prior becomes too tight.

6. CONCLUSION

Smets and Wouters (2003) showed that large-scale new Keynesian models with real and nominal rigidities can fit as well as VARs estimated under diffuse priors, possibly better. This result implies that these models are tools for quantitative analysis by policy making institutions. In addition, it implies that VARs estimated with simple least squares techniques or, from a Bayesian perspective, estimated under a very diffuse prior many not provide a reliable benchmark. This in turn suggests that more elaborate tools for model evaluation are necessary. Using techniques that we developed earlier (Del Negro and Schorfheide 2004), we constructed a reliable benchmark by systematically relaxing the restrictions that the DSGE model poses on a VAR to optimize its fit measured by the marginal likelihood function. We argued that comparing the impulse response functions of the DSGE model's and the benchmark's can shed light on the nature of the DSGE model's misspecification.

Our substantive findings are as follows. First, the posterior odds of a DSGE model versus a VAR with a fairly diffuse

prior do not provide a particularly robust assessment of fit. Small changes in the sample period can lead to reversals of the model ranking. The DSGE–VAR analysis, on the other hand, is much less sensitive to changes in the sample period. Second, there is strong evidence of misspecification in the new Keynesian model, suggesting that forecasts and policy recommendations obtained from this class of models should be viewed with some degree of skepticism. Finally, on the positive side, we find that accounting for misspecification by optimally relaxing the DSGE model restrictions does not alter the responses to a monetary policy and technology shocks in any significant way, both qualitatively and quantitatively. Thus, despite its deficiencies, the new Keynesian DSGE model indeed can generate realistic predictions of the effects of unanticipated changes in monetary policy and technology shocks.

ACKNOWLEDGMENTS

The authors thank Sungbae An, Ron Gallant, Jim Hamilton, Giorgio Primiceri, Chris Sims, Tony Smith, and seminar participants at the FRB Atlanta, Bank of England, Duke University,

London School of Economics, the 2005 NBER Summer Institute, New York University, Northwestern University, Nuffield College, the FRB Richmond, Stanford University, University of Chicago, University of Southern California, University of Virginia, Yale University, the 2004 FRB Cleveland Workshop on Empirical Methods and Applications for DSGE Models, 2003 Euro Area Business Cycle Network Meeting, 2004 SCD Meeting, 2004 SED Meeting, and the 2005 FRB Chicago Conference on Price Stability for useful comments. Schorfheide gratefully acknowledges financial support from the Alfred P. Sloan Foundation. The views expressed in this article are solely the authors' and do not necessarily reflect those of the Federal Reserve Bank of Atlanta, the Federal Reserve System, the European Central Bank, or the National Bank of Belgium.

APPENDIX: THE FULL SET OF IMPULSE RESPONSE FUNCTIONS

Figure A.1 shows the impulse responses of the endogenous variables to one-standard-deviation shocks for the DSGE–VAR(∞) (dotted lines) and the state-space representation of the

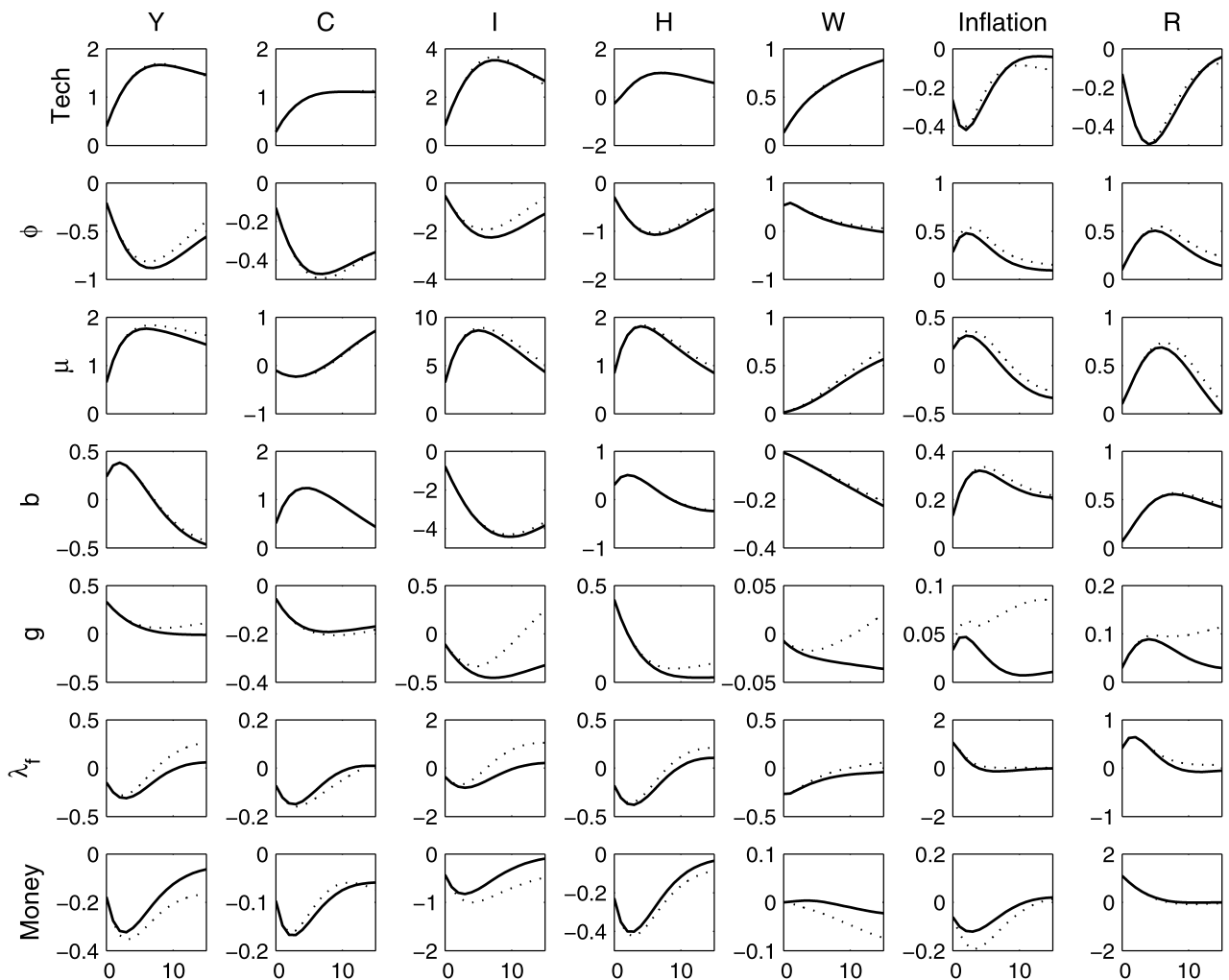


Figure A.1. Baseline Model Impulse Response Functions: DSGE Model versus DSGE–VAR(∞). This figure depicts the impulse responses of the endogenous variables to one standard deviation shocks for the DSGE–VAR(∞) (dotted lines) and for the state-space representation of the DSGE model (solid lines).

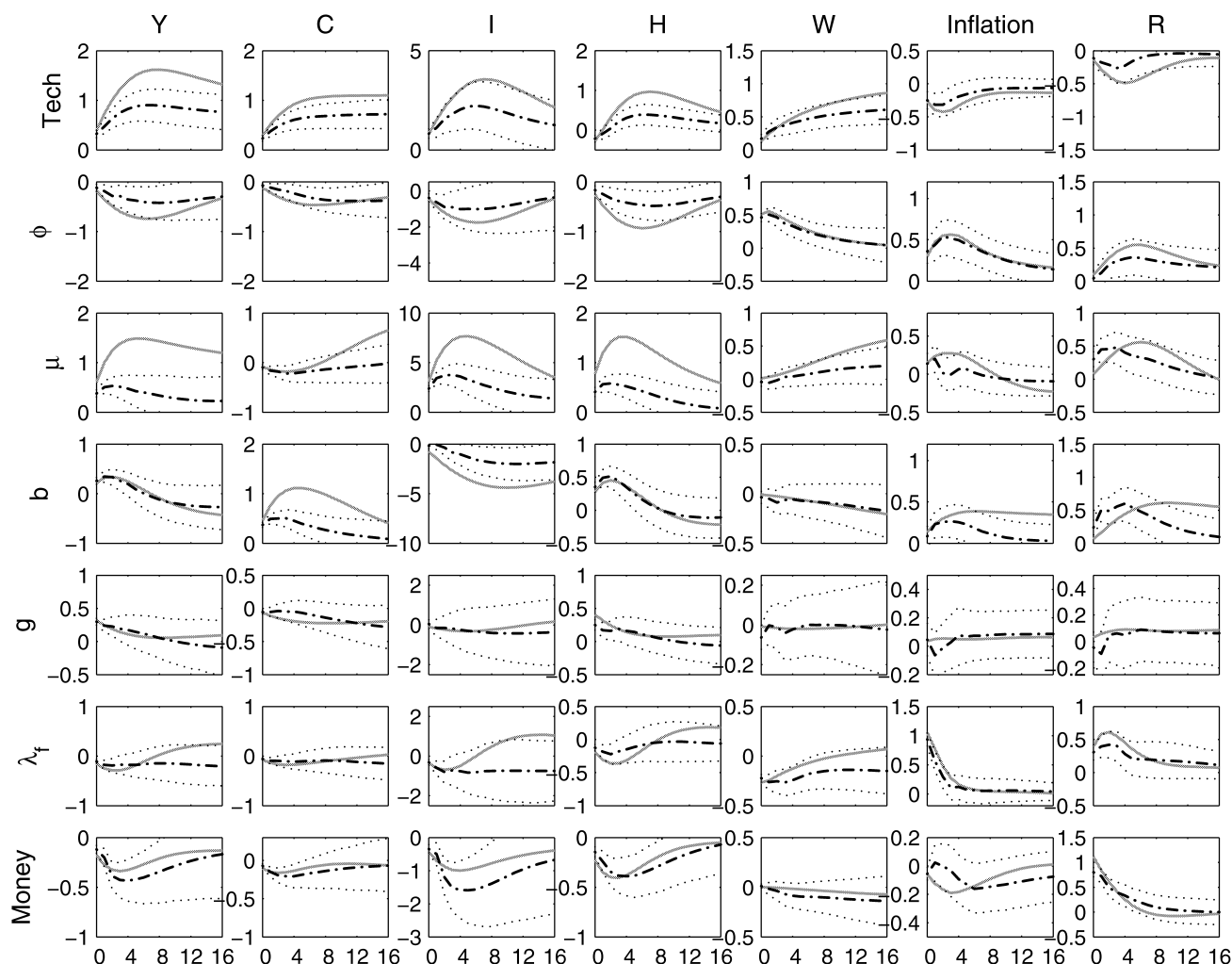


Figure A.2. Baseline Model Impulse Response Functions: $DSGE-VAR(\hat{\lambda})$ versus $DSGE-VAR(\infty)$. This figure depicts mean responses of the endogenous variables to one standard deviation shocks for the $DSGE-VAR(\infty)$ (gray solid lines), the $DSGE-VAR(\hat{\lambda})$ (dark dash-and-dotted lines), and 90% bands (dark dotted lines) for $DSGE-VAR(\hat{\lambda})$.

DSGE model (solid lines). Both impulse responses are computed using the same set of DSGE model parameters, namely the mean estimates for the DSGE model reported in Table 1.

Figure A.2 depicts mean responses of the endogenous variables to one-standard-deviation shocks for the $DSGE-VAR(\infty)$ (gray solid lines), the $DSGE-VAR(\hat{\lambda})$ (dark dashed-dotted lines), and 90% bands (dark dotted lines) for $DSGE-VAR(\hat{\lambda})$.

The impulse responses are computed with respect to the following shocks: technology growth z_t (*tech*), labor/leisure preference (ϕ), capital adjustment (μ), intertemporal preference (b), government spending (g), markup (λ_f), and monetary policy (*money*).

[Received February 2006. Revised August 2006.]

REFERENCES

- Altig, D., Christiano, L., Eichenbaum, M., and Lindé, J. (2004), "Firm-Specific Capital, Nominal Rigidities, and the Business Cycle," Working Paper 04-16, Federal Reserve Bank of Cleveland, Research Department.
- An, S., and Schorfheide, F. (2006), "Bayesian Analysis of DSGE Models," Working Paper 06-5, Federal Reserve Bank of Philadelphia.
- Bils, M., and Klenow, P. (2004), "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, 112, 947–985.
- Boivin, J., and Giannoni, M. (2006), "Has Monetary Policy Become More Effective," *Review of Economics and Statistics*, 88, 445–462.
- Boldrin, M., Christiano, L., and Fisher, J. (2001), "Habit Persistence, Asset Returns, and the Business Cycle," *American Economic Review*, 91, 149–166.
- Calvo, G. (1983), "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12, 383–398.
- Canova, F. (1994), "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, 9, S123–S144.
- Chang, Y., and Kim, S.-B. (2006), "From Individual to Aggregate Labor Supply: A Quantitative Analysis Based on Heterogeneous Agent Macroeconomy," *International Economic Review*, 47, 1–27.
- Christiano, L., Eichenbaum, M., and Evans, C. (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.
- Cogley, T., and Nason, J. (1994), "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, 9, S37–S70.
- DeJong, D., Ingram, B., and Whiteman, C. (1996), "A Bayesian Approach to Calibration," *Journal of Business Economics & Statistics*, 14, 1–9.
- (2000), "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics*, 98, 203–223.
- Del Negro, M., and Schorfheide, F. (2004), "Priors From General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.
- (2006), "How Good Is What You've Got? DSGE-VAR as a Toolkit for Evaluating DSGE Models," *Federal Reserve Bank of Atlanta Economic Review*, 91 (2), 21–37.
- Diebold, F., Ohanian, L., and Berkowitz, J. (1998), "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies*, 65, 433–452.

- Dridi, R., Guay, A., and Renault, E. (2007), "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136, 397–430.
- Fernandez-Villaverde, J., Rubio-Ramirez, J., and Sargent, T. (2007), "A, B, C (and D's) for Understanding VARs," *American Economic Review*, forthcoming.
- Gallant, R., and McCulloch, R. (2004), "On the Determination of General Scientific Models," manuscript, Duke University, Fuqua School of Business.
- Geweke, J. (1999a), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, 18, 1–126.
- (1999b), "Computational Experiments and Reality," manuscript, University of Iowa, Dept. of Economics.
- Gourieroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.
- Greenwood, J., Hercowitz, Z., and Krusell, P. (1998), "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review*, 87, 342–362.
- Ingram, B., and Whiteman, C. (1994), "Supplanting the Minnesota Prior: Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, 34, 497–510.
- Kimball, M., and Shapiro, M. (2003), "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" manuscript, University of Michigan, Dept. of Economics.
- Kydland, F., and Prescott, E. (1982), "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370.
- Rotemberg, J., and Woodford, M. (1997), "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomics Annual*, 12, 297–246.
- Schorfheide, F. (2000), "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, S645–S670.
- Sims, C. (2002), "Solving Rational Expectations Models," *Computational Economics*, 20, 1–20.
- (2003), "Probability Models for Monetary Policy Decisions," manuscript, Princeton University, Dept. of Economics.
- Smets, F., and Wouters, R. (2003), "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.
- (2005), "Comparing Shocks and Frictions in U.S. and Euro Area Business Cycles: A Bayesian DSGE Approach," *Journal of Applied Econometrics*, 20, 161–183.
- Smith, A. (1993), "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.
- Taylor, J. (1993), "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.
- Zellner, A. (1971), *Introduction to Bayesian Inference in Econometrics*, New York: Wiley.

Comment

Lawrence J. CHRISTIANO

Department of Economics, Northwestern University, Evanston, IL 60208, and National Bureau of Economic Research (lchristiano@northwestern.edu)

1. INTRODUCTION

I am very grateful to have been given the opportunity to discuss this important and influential article by Del Negro, Schorfheide, Smets, and Wouters (DSSW hereinafter). It represents a notable step forward in the ongoing enterprise of introducing Bayesian ideas into the analysis of macroeconomic time series. As dynamic stochastic general equilibrium (DSGE) models become more useful from an empirical standpoint, we need increasingly sophisticated methods to diagnose how well they fit. Because these developments in DSGE modeling are relatively recent and have occurred rather suddenly, we are short on diagnostic methods. The article's main contribution is to present and apply such a method, building on the work of Del Negro and Schorfheide (2006).

I begin with a brief review of DSSW's procedure. That procedure works with a "hybrid model" that is a combination of an unrestricted vector autoregression (VAR) for the data and the VAR implied by the econometrician's DSGE model. The combination is indexed by a scalar parameter, λ , where the hybrid model reduces to the unrestricted VAR when λ is small and to the DSGE model as $\lambda \rightarrow \infty$. The best hybrid model is the one associated with $\hat{\lambda}$, the value of λ that results in the highest marginal likelihood for the data. If $\hat{\lambda}$ is large, then the DSGE model is a good one. If $\hat{\lambda}$ is sufficiently small, then this is evidence that the researcher needs to go back to the drawing board to improve the DSGE model.

My comment focuses on two questions: (1) What is the rationale for using the marginal likelihood to assess alternative values of λ ?, and (2) What should the cutoff values of $\hat{\lambda}$ be for

deciding whether a DSGE model is good or bad? After addressing these questions, I ask whether there are other procedures for evaluating model fit. I turn to this question in the conclusion.

The two basic ingredients in the marginal likelihood are the likelihood of the data, which is assumed to be normal, and the priors over model parameters. In practice, the choices made on both dimensions are controversial. Based on the skewness and kurtosis properties of residuals in an estimated VAR, I find strong evidence against the normality assumption. In addition, the choice of priors is as heavily influenced by computational tractability as by plausibility. The marginal likelihood is compelling only to the extent that its two ingredients are compelling.

I report the results of computational experiments with simple examples that suggest that the magnitude of deviation from normality, which is statistically very significant, is not large enough to distort the DSSW analysis. Regarding the choice of priors, in my comment I merely question the appropriateness of the DSSW priors. I suggest a way to construct an alternative set of priors that may better capture a researcher's actual priors over VAR parameters. However, it is beyond the scope of this comment to investigate whether a DSSW-style analysis is robust to such an alternative specification of priors.

Next, I turn to the question of how large is a "large" and how small is "small" in the case of $\hat{\lambda}$. I construct two Monte Carlo experiments in which artificial data are generated by a DSGE model and the econometrician correctly specifies the

model. This allows me to assess how small $\hat{\lambda}$ must be for the econometrician to conclude that something is wrong with the DSGE model. Not surprisingly, I find that the answer depends on two things: (1) the details of the DSGE model and (2) the number of free parameters in the unrestricted VAR relative to the number of free parameters in the DSGE model. This suggests that the DSSW method could be made even more useful if explicit guidance could be provided to link the lower cutoff value of $\hat{\lambda}$ to the model used in the analysis and to the number of degrees of freedom. I also construct a Monte Carlo experiment in which the econometrician's DSGE model is misspecified. The DSSW method is shown to have power in that it discovers with very high probability that the DSGE model is a poor one. All of the experiments suggest one simple improvement to the DSSW method that would help it better identify weaknesses in model fit. In addition to reporting $\hat{\lambda}$ itself, there should be an analysis of the rate at which the marginal likelihood declines for $\lambda > \hat{\lambda}$. The experiments suggest that such a measure would help sharply differentiate between good-fitting and bad-fitting DSGE models. On the whole, the Monte Carlo experiments support DSSW's conclusion that there is information about model fit in their method.

DSSW argue that the best hybrid model that emerges from their analysis is of independent interest. The idea is that it can serve as a basis for thinking about how to improve the DSGE model in cases where $\hat{\lambda}$ is small. This is possible, though I am skeptical. As emphasized by DSSW, the marginal likelihood penalizes models with a large number of free parameters. In practice, the parameters of the hybrid model are those of the unrestricted VAR, modified to resemble those implied by the DSGE model. Such a hybrid model can lead to improvement in the marginal likelihood simply because the DSGE model has substantially fewer free parameters, not because the hybrid model is necessarily closer to the "true" reduced form in a sense relevant to the economic analyst. Nonetheless, it is possible to evaluate the DSSW idea that the hybrid model is useful for identifying directions for improvement in the DSGE model by constructing the type of experiments analyzed in this comment. In such an experiment, the econometrician would be modeled as analyzing artificial data generated from a wrong model and use the hybrid model to identify directions for improvement.

The following section briefly reviews the DSSW procedure. Section 3 evaluates the marginal likelihood as a measure of model fit. Section 4 investigates how the magnitude of $\hat{\lambda}$ should be interpreted. The final section concludes.

2. THE DSSW PROCEDURE

Let the constants and the parameters on the lag coefficients in the VAR representation of the data, Y , be denoted by Φ . Let the variance-covariance matrix of the one-step-ahead forecast errors in this VAR be denoted by Σ . The mapping from the DSGE model parameters, θ , to the VAR representation of Y is denoted by $\Phi(\theta)$ and $\Sigma(\theta)$. DSSW assume that the data have a normal distribution, so that the likelihood of the data is a function only of Φ and Σ ,

$$L(Y|\Phi, \Sigma).$$

Evaluating the marginal likelihood requires integration over the model parameters. One possibility for doing this is to replace

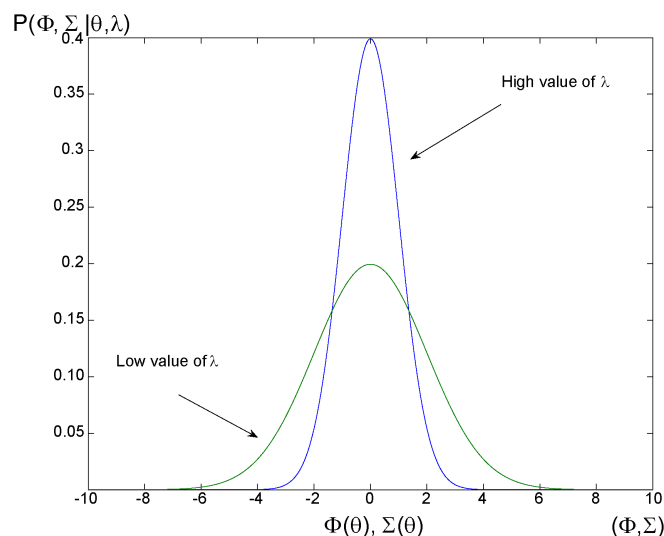


Figure 1. Priors Over VAR Parameters.

Φ and Σ by $\Phi(\theta)$ and $\Sigma(\theta)$ and specify a prior over θ . But this does not serve DSSW's purpose, because this presumes that the DSGE model is true and that the only thing not known about it is the values of θ . In their assessment of the fit of the DSGE model, DSSW wanted to be open to the possibility that the model does not fit well. To see how DSSW proceeded, consider Figure 1. On the horizontal axis are the VAR parameters, reduced to a single dimension for the sake of the discussion. In the middle of the horizontal axis are the values of the VAR parameters implied by the DSGE model, with parameter values θ . If the DSGE model were true, then the prior over Φ and Σ conditional on θ would be a single spike above 0 on the horizontal axis. But this would defeat a basic objective of DSSW, which is to evaluate the fit of the DSGE model and in particular to entertain the possibility that the fit of the DSGE model is poor. For this reason, DSSW construct a prior distribution on Φ and Σ that assigns a positive probability to the state of the world in which the DSGE model is false. Conditional on a value for θ , the prior has mode $\Phi(\theta)$, $\Sigma(\theta)$ and is denoted by

$$P(\Phi, \Sigma|\theta, \lambda),$$

where the value of the scalar, λ , controls how quickly the prior drops to zero (see Fig. 1). As λ goes to infinity, the prior converges to a single spike over $\Phi(\theta)$ and $\Sigma(\theta)$ and corresponds to the case in which the DSGE model is believed to be true. With small values of λ , the prior becomes increasingly diffuse, and a sufficiently small value of λ captures the view that the DSGE model provides very little prior information on Φ and Σ . The marginal likelihood of the data, conditional on the priors and on a value for λ , is denoted by $\mathcal{L}(Y, \lambda)$, which is defined as

$$\mathcal{L}(Y, \lambda) = \int_{\theta} \int_{(\Phi, \Sigma)} L(Y|\Phi, \Sigma) P(\Phi, \Sigma|\theta, \lambda) P(\theta) d(\Phi, \Sigma) d\theta.$$

The DSSW procedure computes $\hat{\lambda}$ as the solution to

$$\hat{\lambda} = \arg \max_{\lambda \geq 0} \mathcal{L}(Y, \lambda). \quad (1)$$

In principle, evaluating $\mathcal{L}(Y, \lambda)$ requires solving a massive numerical integration problem. For example, suppose that we had an $m = 10$ variable VAR with four lags and a constant term

in each equation, so that $k = 41$. Then the number of elements in Φ would be 410, and the number of elements in Σ would be 55. In this case the parameters Φ and Σ alone contribute 465 dimensions to the integration problem, whereas the parameters θ , contribute another 30–40 dimensions. Numerical integration in such a high-dimensional space, although not impossible, would be a major impediment to implementing the DSSW procedure. To avoid this, DSSW specify $P(\Phi, \Sigma|\theta, \lambda)$ to be conjugate with the normal likelihood; that is, $P(\Phi, \Sigma|\theta, \lambda)$ is the product of the inverse-Wishart density for Σ and the multivariate normal density for Φ conditional on Σ . The scalar, λ , controls how tightly concentrated this density is about $\Phi(\theta)$, $\Sigma(\theta)$. With this specification of the prior, the product,

$$\int_{(\Phi, \Sigma)} L(Y|\Phi, \Sigma) P(\Phi, \Sigma|\theta, \lambda) d(\Phi, \Sigma),$$

can be evaluated analytically for given values of θ and λ . With this dramatic reduction in the dimension of the integration problem, evaluating $\mathcal{L}(Y, \lambda)$ becomes computationally feasible. Nevertheless, the computational problem is quite cumbersome, however, so that in practice the maximization problem in (1) must be limited to a coarse grid of λ 's.

For a specific value of λ , the mode of the posterior distribution of Φ and Σ is DSSW's hybrid VAR. Thus a side product of the calculations is a "best" hybrid parameterization. If that parameterization is far from the DSGE model (i.e., $\hat{\lambda}$ is small), this is an indication that the DSGE model fits poorly. If the hybrid parameterization corresponds closely to that implied by the DSGE model (i.e., $\hat{\lambda}$ is large), this is an indicator of good fit.

3. THE DSSW STRATEGY: A PRIORI CONSIDERATIONS

In this section I raise some questions about the a priori appeal of using the marginal likelihood to evaluate model fit. That the strategy in principle has some appeal is suggested by the fact that we use it in ordinary day-to-day conversation. For example, in a discussion about the reason that one's car will not start or about the solution to a murder mystery, the hypothesis that best explains the pertinent facts commands the most attention. However, although in principle the marginal likelihood may seem an attractive way to select among models, in practice significant compromises must be made for the approach to be tractable. This is a shortcoming of the strategy, because the impact of these compromises on the outcome of the analysis is hard to judge. I now review these considerations by examining the two basic ingredients in constructing the marginal likelihood: the likelihood function and the priors.

DSSW follow convention in specifying the likelihood function of the data to be normal. I investigated the plausibility of this specification by fitting a four-lag, seven-variable VAR using monthly U.S. data for the period 1955Q4–2006Q1. The following variables are used:

$$y_t = [\log C_t/Y_t, \log I_t/Y_t, \text{inflation}_t,$$

$$\log Y_t/L_t - \log W_t/P_t, R_t, \log L_t, \text{GDP growth}_t].$$

Here C_t denotes real per capita nondurables and services consumption Y_t denotes real, per capita GDP; L_t denotes per capita hours worked, inflation denotes inflation in the personal consumption expenditure index, W_t/P_t denotes real labor compensation for the whole economy (e.g., farmers and government included), R_t denotes federal funds rate, and I_t denotes real per capita gross private domestic investment plus household purchases of durable goods. All real variables were obtained by deflating by the GDP chain price index. Skewness and kurtosis statistics were computed for each of the seven VAR disturbances and their p value were computed relative to the null hypothesis that the underlying disturbances are normal. The p values were computed by (1) simulating 1,000 artificial datasets using the fitted VAR and drawing the disturbances, u_t , from the multivariate normal distribution with mean 0 and variance-covariance matrix equal to its estimated sample analog and (2) computing the percentage of times that the empirically estimated statistic is exceeded by its analog computed across the artificial samples. The results are reported in Table 1. The kurtosis statistics are particularly large, and all kurtosis statistics but the one on the disturbance in the inflation equation have p values $< .01$. The kurtosis statistic on the interest rate has a p value of 0 after rounding. The kurtosis statistic on GDP growth is also particularly large, with a p value of .02. The skewness statistic on the interest rate and on inflation are both very large, with p values $< .01$. We can conclude that the evidence against the normality assumption is substantial.

Next, consider the priors over the VAR parameters. In practice, it is difficult to deviate from the usual normal/Wishart assumption. This is because doing so results in a prior that is not conjugate with the normal distribution, which then forces one to confront the formidable numerical integration problem discussed in the previous section. But for the DSSW strategy to be compelling requires that we accept the normal/Wishart specification as a reasonable specification of one's priors over VAR parameters. DSSW do take some steps in the direction of a defense. They show that under their assumptions, the height of the prior in Figure 1 is inversely proportional to how surprised one expects to be at VAR parameterizations different from $\Phi(\theta)$,

Table 1. Empirical Skewness and Kurtosis Statistics, and Corresponding p Values

Disturbance	Kurtosis		Skewness	
	Statistic, s	$Pr(s > s normal)$	Statistic, s	$Pr(s > s normal)$
$\log C_t/Y_t$	1.40	.18%	-.17	84.64%
$\log I_t/Y_t$	1.59	.12%	.07	33.04%
PCE inflation _t	.79	2.00%	.41	.68%
$\log Y_t/L_t - \log W_t/P_t$	1.04	.66%	.39	1.22%
R_t	11.01	0%	1.62	0%
$\log I_t$	1.27	.40%	.11	26.46%
GDP growth _t	1.85	.02%	-.02	55.00%

$\Sigma(\theta)$. The surprise is measured by the expected drop in the likelihood, conditional on the DSGE model and on θ .

The DSSW defense of their specification of the prior over the VAR parameters may have some appeal. Perhaps it is consistent with the notion that in case one's most preferred DSGE model is wrong, the most likely alternative is somewhere nearby. However, for this type of argument to rationalize the type of normal/Wishart prior distribution used here would seem to require an extraordinary coincidence. One could assess how well the normal/Wishart represents a researcher's priors over VAR parameters by the following exercise. Assign a set of probabilities to a range of models and a subprobability distribution over the parameters of each model, conditional on that model being true. This specification of model priors induces a prior distribution over VAR parameters. If each DSGE model were not too similar, then it seems safe to speculate that these priors over VAR parameters would have a very different shape—possibly with multiple local peaks—than what we see in Figure 1.

In sum, using the marginal likelihood to assess the plausibility of alternative hypotheses requires a number of detailed assumptions. The normality assumption on the likelihood function seems to be outright inconsistent with the data. The primary advantage of the prior distribution over VAR parameters seems to lie with computational tractability. Using the marginal likelihood to select models has some a priori appeal. But this appeal rests on two propositions: that the right likelihood is used and that the prior distribution corresponds to the priors held by actual researchers. Evidence has been presented that the first proposition is false; the second proposition remains to be established.

4. THE DSSW STRATEGY: HOW IT WORKS IN PRACTICE

DSSW do not provide guidance on how exactly $\hat{\lambda}$ should be used to evaluate model fit. How big should $\hat{\lambda}$ be for one to feel comfortable about a DSGE model? How small does $\hat{\lambda}$ have to be to justify going back to the drawing board to redesign the model? DSSW present a value of $\hat{\lambda}$ in the region of .75 and 1.5. Does this mean that the model fits well, or poorly, or something in between?

To shed light on these questions, I implemented Monte Carlo experiments using artificial datasets of 200 observations generated using three DSGE models. One DSGE model, which I call the RBC model, is the Long–Plosser real business cycle model. The other two DSGE models are different versions of the Clarida, Gali, and Gertler (2000) (CGG) sticky price model (CGG1 and CGG2) (see also Gali, Lopez-Salido, and Valles 2003). Three experiments were done. The first two use RBC and CGG1. In these experiments, the econometrician computes $\hat{\lambda}$ knowing the true model, although not the values of some of its parameters. These experiments provide a sense of the sort of $\hat{\lambda}$'s to expect when the right model is in hand. In the third experiment, the true model is CGG2, but the econometrician mistakenly believes that the true model is a version of the RBC model. This experiment provides a sense of the type of $\hat{\lambda}$'s to expect when the econometrician's model is wrong. Section 4.1 describes the technical aspects of the data-generating mechanisms and experiments. (This could be skipped in a quick read.) Section 4.2 summarizes the Monte Carlo results.

4.1 The Experiments

To help economize on computer time, each Monte Carlo experiment is designed so that the mapping from the DSGE model parameters estimated by the econometrician to a VAR representation for the data $\Phi(\theta)$ and $\Sigma(\theta)$ is trivial. Multiple artificial datasets of length 200 observations are generated from a specific true DSGE model, and the $\hat{\lambda}$ that solves (1) is computed in each dataset. In each dataset, the integral in $\mathcal{L}(Y, \lambda)$ is computed using Geweke's (1999) modified harmonic mean method and 100,000 Markov chain Monte Carlo (MCMC) trials. The procedure was tuned so that the acceptance rate in the MCMC trials averaged approximately 30%. Del Negro and Schorfheide kindly supplied their MATLAB code for the calculations.

As an additional step to keep the required computer time down, the maximization in (1) is restricted to the following set of possible values of λ :

$$\begin{array}{ccccc} .11 & .25 & .43 & .67 & 1.00 \\ 1.5 & 2.33 & 4.00 & 9.00 & \infty. \end{array}$$

When transformed into $\lambda/(1 + \lambda)$, this corresponds to the 10 equally spaced grid points, .1, .2, ..., 1.0. The quantity $\lambda/(1 + \lambda)$ is of interest because it corresponds to the relative weight assigned in the hybrid model to the DSGE model.

The RBC Model. The preferences, technology, and shocks in the Long–Plosser model are as follows:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[\log C_t - \frac{\exp(\tau_t)}{1 + \psi} l_t^{1+\psi} \right],$$

$$C_t + K_{t+1} \leq K_t^\alpha (\exp(z_t) l_t)^{(1-\alpha)} = Y_t,$$

and

$$\tau_t, z_t: \text{iid mean-0 random variables, variance } \sigma_\tau^2 \text{ and } \sigma_z^2,$$

where C_t denotes consumption, l_t denotes labor, K_{t+1} denotes capital, z_t is a technology shock, and τ_t is a preference shock. I set $\alpha = 1/3$ and $\beta = .99$. The parameters estimated by the econometrician are $\theta = (\psi, \sigma_\tau^2, \sigma_z^2)$, the true values of which are $(1, .02^2, .02^2)$. The econometrician's prior distribution for each parameter is inverted gamma, with mode equal to the corresponding true value. Specifically, denote the inverted gamma density for the random variable, x , by $f(x)$. Then

$$f(x) = \frac{\zeta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\zeta}{x}\right),$$

where Γ is a gamma function. I assume that $\alpha = 10$ and that ζ is determined by the assumption on the mode of the distribution.

Artificial datasets on $k_{t+1} = \log(K_{t+1})$ and $\log(Y_t/l_t)$ are generated by the RBC model and provided to the econometrician, whose mapping from θ to a VAR representation is defined by

$$\begin{pmatrix} k_{t+1} \\ \log(Y_t/l_t) \end{pmatrix} = \begin{pmatrix} \gamma_k \\ \gamma_a \end{pmatrix} + \begin{bmatrix} \alpha & 0 \\ \alpha & 0 \end{bmatrix} \begin{pmatrix} k_t \\ \log(Y_{t-1}/l_{t-1}) \end{pmatrix} + \begin{pmatrix} (1-\alpha)(z_t - \frac{1}{1+\psi}\tau_t) \\ (1-\alpha)z_t + \frac{\alpha}{1+\psi}\tau_t \end{pmatrix},$$

where

$$\gamma_k = \frac{1-\alpha}{1+\psi} \log\left(\frac{1-\alpha}{1-\beta\alpha}\right) + \log \beta\alpha$$

and

$$\gamma_a = -\frac{\alpha}{1+\psi} \log\left(\frac{1-\alpha}{1-\beta\alpha}\right).$$

This VAR representation can be derived from the well-known fact that the solution to this model is given by $K_{t+1} = \beta\alpha Y_t$.

The CGG1 Model. In the CGG model, the equilibrium allocations under a specific monetary policy rule (the “equilibrium”) are expressed as a deviation from the best equilibrium achievable when the monetary policy rule is dropped (the “Ramsey equilibrium,” or “natural equilibrium”). In the Ramsey equilibrium inflation, π_t , is always 0, and the nominal rate of interest is given by

$$rr_t^* = \log \frac{1}{\beta} + \rho \Delta a_t + \frac{1-\zeta}{1+\varphi} \tau_t, \quad (2)$$

where β is the discount rate of the representative household (see App. A for the specification of preferences and technology underlying this economy). The equilibrium conditions are

$$\beta E_t \pi_{t+1} + \kappa x_t - \pi_t = 0 \quad (\text{Calvo pricing equation}), \quad (3)$$

$$-[r_t - E_t \pi_{t+1} - rr_t^*] + E_t x_{t+1} - x_t = 0 \quad (\text{intertemporal Euler equation}), \quad (4)$$

and

$$u_t + \phi_\pi \pi_t - r_t = 0 \quad (\text{monetary policy rule}), \quad (5)$$

where r_t is the equilibrium nominal rate of interest and $x_t \equiv y_t - y_t^*$ is the deviation of equilibrium output, y_t , from Ramsey equilibrium output, y_t^* . In addition,

$$\kappa = \frac{(1-\xi_p)(1-\beta\xi_p)(1+\varphi)}{\xi_p},$$

where ξ_p is the probability that an intermediate goods producer is not able to reoptimize its price in any given period. The monetary policy shock u_t , the growth rate of technology Δa_t , and a labor preference shock τ_t are assumed to have the following scalar first-order autoregressive representations:

$$u_t = \delta u_{t-1} + \eta_t,$$

$$\Delta a_t = \rho \Delta a_{t-1} + \varepsilon_t,$$

and

$$\tau_t = \zeta \tau_{t-1} + \varepsilon_t^\tau.$$

In these expressions the innovations are iid and have variances σ_η^2 , σ_ε^2 , and $\sigma_{\varepsilon^\tau}^2$. I adopt the following model parameterization:

$$\begin{aligned} \phi_\pi &= 1.5, & \beta &= .99, & \varphi &= 1, & \rho &= .2, \\ \xi_p &= .75, & \delta &= .2, & \zeta &= .5, & \lambda_f &= 1.25, \\ \sigma_\eta &= .005, & \sigma_\varepsilon &= .01, & \sigma_{\varepsilon^\tau} &= .006. \end{aligned}$$

To solve this model, first write the four equilibrium conditions, (2), (3), (4), and (5) in matrix form:

$$E_t[\alpha_0 z_{t+1} + \alpha_1 z_t + \beta_0 s_{t+1} + \beta_1 s_t] = 0,$$

where

$$z_t = \begin{pmatrix} \pi_t \\ x_t \\ r_t - \log(1/\beta) \\ rr_t^* - \log(1/\beta) \end{pmatrix}, \quad s_t = \begin{pmatrix} \Delta a_t \\ u_t \\ \tau_t \end{pmatrix},$$

$$s_t = P s_{t-1} + \varepsilon_t,$$

and $\alpha_0, \alpha_1, \beta_0, \beta_1$, and P are functions of the model parameters. Then the solution is

$$z_t = B s_t,$$

where the 4×3 matrix B uniquely solves the following linear system of equations:

$$(\beta_0 + \alpha_0 B)P + \beta_1 + \alpha_1 B = 0.$$

The vector of DSGE model parameters, θ , estimated by the econometrician has eight dimensions and is composed of φ , ξ_p , and the two parameters associated with each of the three shocks. In the case of each element of θ , the prior is an inverted-gamma distribution with mode equal to the true value of the parameter and standard deviation equal to the true value of the parameter, divided by 2. The econometrician estimates a three-variable VAR using data on x_t , π_t , and r_t . Appendix B verifies that with this DSGE, the data satisfy a first-order VAR.

The CGG2 Model. In this version of the CGG model, I replace the technology shock process by the following stationary representation:

$$a_t = \rho a_{t-1} + \varepsilon_t,$$

where $\rho = .95$ and all other parameters are as in the CGG1 model. With this change, the equilibrium condition for the Ramsey rate of interest, (2), is replaced by

$$rr_t^* = \log \frac{1}{\beta} + (1-\rho)a_t + \frac{1-\zeta}{1+\varphi} \tau_t,$$

and the log of the Ramsey level of output satisfies

$$y_t^* = \gamma + a_t - \frac{1}{1+\varphi} \tau_t, \quad \gamma = -\frac{\log \lambda_f}{1+\varphi}. \quad (6)$$

Here $\lambda_f (= 1.25)$ is a parameter that controls the elasticity of demand for intermediate goods and corresponds to the markup earned in steady state by monopolists in the model. Equilibrium output, Y_t , is obtained as

$$\log Y_t = x_t + y_t^*,$$

and equilibrium employment, l_t , is obtained from

$$\begin{aligned} \log l_t &= \log Y_t - a_t \\ &= x_t + \gamma - \frac{1}{1+\varphi} \tau_t. \end{aligned}$$

Artificial data on $\log Y_t$ and $\log l_t$ are generated using CGG2 and provided to the econometrician, who mistakenly assumes that the data were generated by a version of the RBC model. In this version, the preference shock has a first-order autocorrelation structure

$$\tau_t = \rho_\tau \tau_{t-1} + \varepsilon_{\tau,t},$$

where $\rho_\tau = .9$ and $E\varepsilon_{\tau,t}^2 = \sigma_{\varepsilon^\tau}^2$. The first-order bivariate VAR representation that the econometrician (falsely) deduces for the

data is

$$\begin{pmatrix} \log Y_t \\ \log l_t \end{pmatrix} = \begin{pmatrix} \gamma_Y \\ \gamma_l \end{pmatrix} + \begin{bmatrix} \alpha & (1-\alpha)\rho_\tau \\ 0 & \rho_\tau \end{bmatrix} \begin{pmatrix} \log Y_{t-1} \\ \log l_{t-1} \end{pmatrix} + \begin{pmatrix} (1-\alpha)(z_t - \frac{1}{1+\psi}\varepsilon_{\tau,t}) \\ -\frac{1}{1+\psi}\varepsilon_{\tau,t} \end{pmatrix},$$

where

$$\gamma_Y = \alpha \log(\alpha\beta) + (1-\alpha) \frac{1-\rho_\tau}{1+\psi} \log\left(\frac{1-\alpha}{1-\beta\alpha}\right)$$

and

$$\gamma_l = \frac{1-\rho_\tau}{1+\psi} \log\left(\frac{1-\alpha}{1-\beta\alpha}\right).$$

To derive the first row of this representation, first note that the log of the production function is

$$\log Y_t = \frac{1}{3}k_t + \frac{2}{3}(z_t + \log(l_t)).$$

Then use the solution of the model to express k_t as a function of $\log Y_{t-1}$ and to express $\log(l_t)$ as a function ε_t^τ and $\log(l_{t-1})$.

The DSGE model parameters estimated by the econometrician are $\theta = (\psi, \sigma_z, \sigma_{\varepsilon_\tau})$. It is assumed that the econometrician uses an inverted gamma prior on each of the three parameters with mode (1, .02, .02) and standard deviation equal to the mode, divided by 7.

4.2 The Results

This section presents results of the Monte Carlo experiments described earlier. Consider first the results when the data-generating mechanism is RBC and the DSSW method is implemented with a bivariate unrestricted VAR that has 1 lag; see RBC (lag 1) in Table 2. In this case the unrestricted model has nine free parameters (six parameters associated Φ and three with Σ), whereas the econometrician's DSGE model has three free parameters. Note that even though the econometrician has the true model in hand, a substantial fraction (6%) of artificial datasets result in a $\hat{\lambda}$ that assigns a relative weight of 1/2 or less to the DSGE model. The reason why $\hat{\lambda}$ sometimes provides evidence against the DSGE model when it is true is that there is a positive probability of datasets in which the unrestricted VAR fits better than the true VAR. To see this, consider the likelihood ratio statistic formed from twice the difference

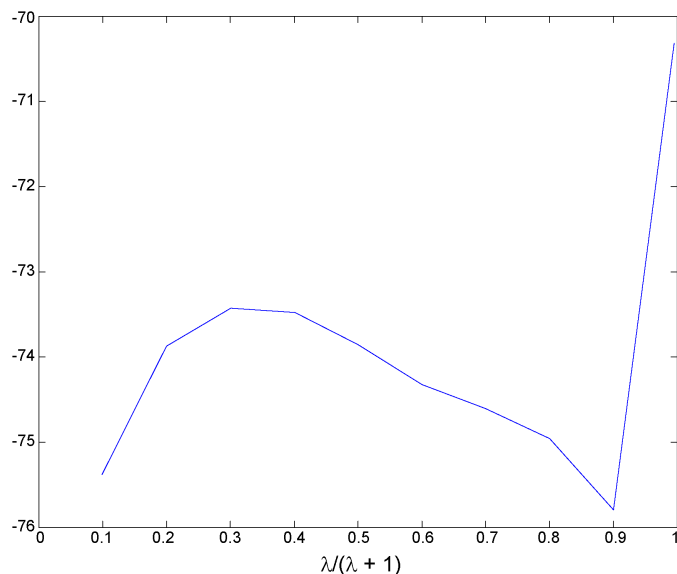


Figure 2. Realization of log-Marginal Likelihood, RBC (lag 1).

of the log-likelihood associated with the estimated unrestricted VAR and the log-likelihood associated with the true VAR implied by the DSGE. Asymptotically, this is a realization from a chi-squared distribution with 9 degrees of freedom. The average value of this statistic over all datasets with the indicated value of $\hat{\lambda}$ is reported in the row beneath the results for RBC (lag 1). Note how these likelihood ratio statistics tend to be higher in datasets associated with a low $\hat{\lambda}$. With some exceptions, this general pattern is also a feature of the CGG1 experiment. The exceptions may reflect Monte Carlo sampling uncertainty.

Two features of the log-marginal likelihood in the RBC (lag 1) results are worth emphasizing. One is illustrated in Figure 2. This shows a type of shape that occurs a nontrivial fraction of times in data generated by RBC (lag 1). In these cases the log-marginal likelihood is concave over most values of λ and then rises sharply for λ near ∞ . The abrupt change in the behavior of the log-marginal likelihood for large values of λ seems puzzling. A consequence of this shape is that results are sensitive to the specification of the set of λ 's over which the maximization in (1) is done. For example, if the upper bound is $\lambda = 200$ rather than $\lambda = \infty$, then the percentage of artificial

Table 2. Cumulative Distribution of $\hat{\lambda}$

Experiment	$\hat{\lambda}/(1+\hat{\lambda})$									
	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
RBC (lag 1)	0	0	.5	2.5	6.0	10.0	15.0	20.5	26.5	100
Likelihood ratio	NA	NA	18	21.4	17.3	16.2	14.4	13.3	13.0	8.8
RBC (lag 4)	0	0	0	0	0	0	0	0	.5	100
Likelihood ratio	NA	NA	NA	NA	NA	NA	NA	NA	27.5	22.5
CGG1 (lag 1)	.5	1.0	1.0	1.0	1.5	4.0	11.0	24.5	53.0	100
Likelihood ratio	14.3	12.5	12.5	12.5	15.2	20.7	21.7	18.9	16.9	15.3
CGG1 (lag 4)	0	0	0	0	0	2.0	9.5	23.5	50.5	100
Likelihood ratio	NA	NA	NA	NA	NA	70.5	59.4	54.3	49.1	45.1
CGG2	100	100	100	100	100	100	100	100	100	100

NOTE: Entries indicate the percentage out of 200 simulations that $\hat{\lambda}$ is less than or equal to value indicated in column head. (lag x): x indicates the number of lags in the unrestricted VAR. Likelihood ratio represents the average likelihood ratio (LR) statistic over all artificial datasets having $\hat{\lambda}$ indicated in the column head, where LR is twice the difference of log-likelihood of unrestricted VAR versus log likelihood of true VAR. NA means not applicable, because there were no $\hat{\lambda}$'s in this entry.

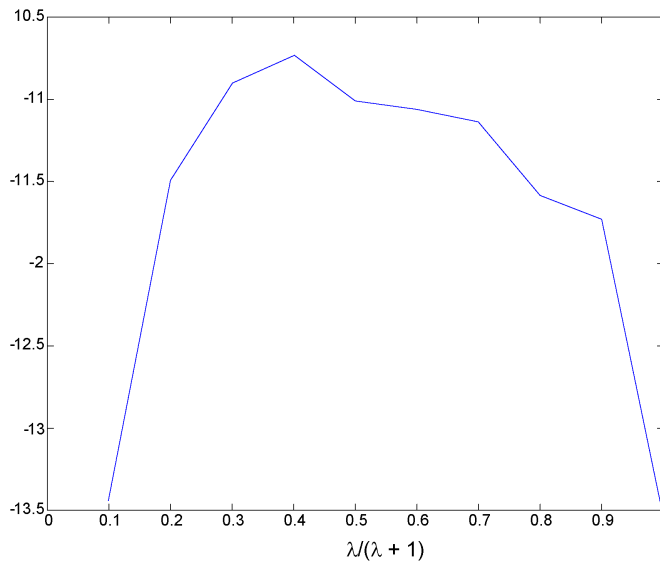


Figure 3. Realization of log-Marginal Likelihood, RBC (lag 1).

datasets with $\hat{\lambda}/(1 + \hat{\lambda}) \leq .9$ is 64%, rather than the 26.5% reported in Table 1. In the calculations for the oral presentation of this comment, the upper bound on λ was $\lambda = 5$, and the reported frequency of small $\hat{\lambda}$'s was even greater. To see this, note from Figure 2 that $\hat{\lambda}$ is .43 if the upper bound on the λ 's considered in the maximization in (1) is 9, and $\hat{\lambda} = \infty$ if the upper bound on is ∞ . The puzzling shape of the log-marginal likelihood in the case of RBC (lag 1) was not observed in the other Monte Carlo experiments.

A second notable feature of the log-marginal likelihood corresponding to RBC (lag 1) is that it exhibits very little variation across different values of λ . For example, the average difference between the log-marginal likelihood at the smallest value of λ and that associated with $\hat{\lambda}$ is only 7.7, which is considerably smaller than that reported in the empirical example presented in DSSW. This finding is illustrated in Figure 3, which displays the log-marginal likelihood associated with a different artificial dataset from the one underlying Figure 2. It bears emphasis that the log-likelihoods in Figures 2 and 3 are chosen for illustrative purposes. Both are atypical, in that they imply very low values of $\hat{\lambda}$.

The lack of variation in the log-marginal likelihood motivated me to consider an unrestricted VAR with additional lags. The results are reported in Table 2, in the row labeled "RBC (lag 4)." In this case $\hat{\lambda} = \infty$ in almost all of the artificial datasets. In addition, the mean difference of the log-marginal likelihood at the lowest value of λ and at $\hat{\lambda}$ is now 32.4. This degree of variation in the log-marginal likelihood is similar to that reported by DSSW. In RBC (lag 4), the number of free parameters in the unrestricted VAR jumps to 21, versus the 3 free parameters in the econometrician's DSGE model. As emphasized by DSSW, the log-marginal likelihood assigns a substantial penalty to free parameters, which is manifest here in the form of a sharp preference in favor of the DSGE model over the unrestricted VAR.

Now consider the CGG1 experiment with 1 lag. In this case the VAR has 3 variables, and so the number of unrestricted parameters is 18, compared with the 8 free parameters of the econometrician's DSGE model. In this experiment, small $\hat{\lambda}$'s are still

possible, although this is less likely than it is in the case of RBC (lag 1). Thus in 4% of the datasets, the relative weight assigned to the DSGE model is 60% or less in the case of CGG1 (lag 1), versus 10% for RBC (lag 1). Presumably, the improved performance of the DSSW method reflects the greater number of parameters in the unrestricted VAR in the CGG1 (lag 1) experiment than in the RBC (lag 1) experiment.

In CGG1 (lag 1), the average difference between the log-marginal likelihood at the lowest value of λ and at $\hat{\lambda}$ is 16.6, which is smaller than the value given for the empirical example reported by DSSW. This led me to consider the case with 4 lags in the unrestricted VAR, raising the number of free parameters from 18 to 45, while the number of free parameters in the DSGE model remains at 8. As in the case of the RBC model, the frequency of low values of $\hat{\lambda}$ declines, although less dramatically than we saw in RBC (lag 4). In particular, 9.5% of the $\hat{\lambda}$'s assign a weight of 70% or less to the DSGE model in CGG1 (lag 4) versus 0% for RBC (lag 4). In CGG1 (lag 4), the average difference between the log-marginal likelihood at the lowest value of λ and at $\hat{\lambda}$ is 35.4, which is closer to the empirical example of DSSW.

In the lag 4 versions of both RBC and CGG1, there is a substantial difference between the log-marginal likelihood at the lowest value of λ and at $\lambda = \hat{\lambda}$, as found by DSSW. However, in my examples there is relatively less difference between the log-marginal likelihood at $\lambda = \hat{\lambda}$ and at $\lambda = \infty$. For example, in CGG2 the mean decline in the log-marginal likelihood from $\lambda = \hat{\lambda}$ to $\lambda = \infty$ when $\hat{\lambda} < \infty$ is slightly less than unity. This is substantially smaller than the sharp drop in the log-marginal likelihood reported by DSSW.

My next experiment, CGG2, investigates the behavior of $\hat{\lambda}$ and the log-marginal likelihood when the econometrician's DSGE model is false by construction. In this case, the data are generated by a version of the CGG model, but the econometrician's DSGE model is a version of the RBC model. The unrestricted VAR is estimated with one lag. Because two variables are included in the analysis, the unrestricted VAR has nine free parameters. The econometrician's DSGE model has three free parameters. The results for this experiment are dramatic. In each artificial dataset, $\hat{\lambda}$ is the lowest value of λ . Thus DSSW's method correctly reveals, with probability 1, that the DSGE model is misspecified. Moreover, the slope of the log-marginal likelihood is very steep, with the difference between the log-marginal likelihood at the lowest and highest values of λ being on the order of 300–500. I presume that the finding that the unrestricted VAR is always the best model in this experiment is an artifact of the specification that it has only one lag. If more lags had been permitted in the unrestricted VAR, then $\hat{\lambda}$ would have exceeded the lowest value of λ at least occasionally. The features of this example that I expect to be robust are that $\hat{\lambda}$ is substantially less than infinity, and that the log-marginal likelihood declines steeply for $\lambda > \hat{\lambda}$.

Of the examples considered, the only one that can replicate DSSW's finding that the log-marginal likelihood declines steeply for $\lambda > \hat{\lambda}$ is the one in which the econometrician's model is false. The two examples in which the econometrician's DSGE model is true do occasionally produce a $\hat{\lambda}$ substantially less than infinity. However, it is rare for the slope of the log-marginal likelihood to be steeply negative for $\lambda > \hat{\lambda}$.

These findings suggest that the DSSW method would have even greater power to identify evidence against DSGE models if the method formally integrated the slope of the log-marginal likelihood for $\lambda > \hat{\lambda}$ into the diagnostic procedure.

Finally, recall the empirical evidence of leptokurtosis reported in the previous section. In principle, using the normal likelihood in Bayesian analysis entails specification error. To investigate whether this error distorts the DSSW analysis, I redid the RBC experiment using disturbances that exhibit the amount of kurtosis observed in the data. My results were essentially unchanged from what is reported in Table 2, consistent with the proposition that the amount of leptokurtosis observed in the data is not sufficient to distort Bayesian analyses that use the normal likelihood. Of course, these findings (like all other findings in Table 2) are only indicative and need to be substantiated by similar additional experiments.

5. CONCLUSION

DSSW have provided a valuable service in describing and implementing a measure of fit for DSGE models. The Monte Carlo evidence presented in this comment suggests four ways in which DSSW's measure of fit could be made even more useful:

1. It would be useful if a lower cutoff value of $\hat{\lambda}$ were provided, such that for smaller $\hat{\lambda}$, the researcher knows with high probability, there is a problem with the DSGE model. The Monte Carlo experiments in my comment suggest that such a cutoff would be a function of, among other things, the difference between the number of free parameters in the unrestricted VAR and in the DSGE model.
2. The rate at which the marginal likelihood declines for $\lambda > \hat{\lambda}$ should be formally integrated into the DSSW procedure. The Monte Carlo experiments suggest that a steep rate of decline is a reliable signal that the econometrician's DSGE model fits poorly. This rate of decline can be measured in various ways. One way would be to report Bayesian probability intervals for λ .
3. In the absence of a stronger defense for the priors used in the DSSW analysis, it would be useful to have evidence that results based on the DSSW priors are robust to plausible alternatives. A practical impediment to evaluating robustness is that priors that deviate from DSSW's are unlikely to have convenient conjugacy properties. As a result, the numerical integration problem in (1) would be computationally very burdensome in practical situations. Nonetheless, robustness could be studied in the type of simple examples considered in my comment, where computational limitations are less binding.
4. I have provided evidence that the DSSW results are robust to the kind of evidence against normality observed in the data. A more systematic investigation of robustness would be useful.

DSSW compare their procedure with alternative measures of model fit based on out-of-sample forecasting performance. Further comparisons of this type would be of interest. Measures of out-of-sample forecasting performance appear to offer at least four advantages over the DSSW method:

1. The computational burden is minimal compared with the substantial resources required to evaluate (1).
2. Computational tractability limits the range of model comparisons that can be done with the DSSW method, whereas there is no limit to the models that can be compared under out-of-sample forecasting criteria. For example, the forecasting performance of the DSGE model can be compared with that of a Bayesian VAR, as done by Ingram and Whiteman (1994). In practice, Bayesian VARs are more useful for forecasting than unrestricted VARs because of parameter parsimony. Alternatively, DSSW's hybrid model under alternative specification of λ could be compared with a Bayesian VAR under the out-of-sample forecasting criterion.
3. Classical sampling theory offers some assistance in determining whether differences in the out-of-sample root mean squared error (RMSE) performance of alternative models are statistically significant (see, e.g., Christiano 1989, app. D). This contrasts with the DSSW method, in which a small $\hat{\lambda}$ suggests the presence of evidence against a DSGE model, but there is no guidance as yet on how small such a $\hat{\lambda}$ must be (see item 1 in the previous list).
4. The out-of-sample forecast performance criterion is transparent and is of obvious interest to everyone. In contrast, for the marginal likelihood to be compelling, one must first confront several difficult—and in some cases, possibly unresolvable—questions. What likelihood is appropriate for the data? Are the researcher's priors faithfully captured by the choice of prior distribution? Implicit in the DSSW procedure is the assumption that one prior is suitable for everyone. But why should researchers with different backgrounds and experiences use the same prior?

DSSW show that an out-of-sample forecast RMSE criterion produces fit results similar to what the DSSW procedure produces. In view of the advantages of the out-of-sample forecasting approach, DSSW's findings would appear to be a powerful argument in its favor.

ACKNOWLEDGMENTS

The author is particularly grateful for the extensive advice and expert research assistance of Cosmin Ilut; for conversations with Joshua Davis, Martin Eichenbaum, and Giorgio Primiceri; and for the research assistance of Patrick Higgins. My comment has benefited from the reactions and advice of Marco Del Negro and Frank Schorfheide. Financial support was provided by the National Science Foundation.

APPENDIX A: PREFERENCES AND TECHNOLOGY UNDERLYING THE CLARIDA–GALI–GERTLER MODEL

Although the preferences and technology underlying the CGG model are well known, we include them here for completeness. In particular, the representative household's preferences are

$$E_0 \sum_{t=0}^{\infty} \beta^t \left(\log C_t - \exp(\tau_t) \frac{l_t^{1+\varphi}}{1+\varphi} \right), \quad \varphi > 0,$$

where C_t and l_t denote consumption and employment and τ_t is a labor supply shock. A budget constraint allows the household to finance consumption by participating in a competitive labor market and by participating in a loan market in which the log of the gross nominal rate of interest is r_t . In equilibrium, the loan market must clear with zero trade.

Final output is produced by competitive firms using intermediate inputs, $Y_t(i)$, $i \in (0, 1)$, using the following technology:

$$Y_t = \left(\int_0^1 Y_t(i)^{1/\lambda_f} di \right)^{\lambda_f}, \quad \lambda_f \geq 1.$$

The technology for producing $Y_t(i)$ is

$$Y_t(i) = A_t l_t(i), \quad a_t = \log(A_t),$$

where $l_t(i)$ is employment by the i th intermediate good producer. This producer, subject to Calvo sticky price frictions, is able to reoptimize its price with probability $1 - \xi_p$. With the complementary probability, the intermediate good producer cannot change its price. In steady state, equilibrium inflation is zero, $Y_t = C_t$, and $l_t = \int_0^1 l_t(i) di$. In the text, Δa_t denotes the first difference of a_t .

APPENDIX B: FIRST-ORDER VECTOR AUTOREGRESSION REPRESENTATION OF THE CGG1 MODEL

Let \tilde{z}_t denote the 3×1 vector composed of the first three elements of z_t , and let \tilde{B} denote the first three rows of B , so that

\tilde{B} is a square matrix. B is invertible in the numerical examples that I considered. The solution for \tilde{z}_t is written $\tilde{z}_t = \tilde{B}s_t$, or

$$\tilde{B}^{-1}\tilde{z}_t = s_t.$$

Then multiply on the left by the matrix lag operator, $I - PL$, to obtain

$$(I - PL)\tilde{B}^{-1}\tilde{z}_t = \epsilon_t$$

or

$$\tilde{B}^{-1}\tilde{z}_t = P\tilde{B}^{-1}\tilde{z}_{t-1} + \epsilon_t.$$

Multiply on the left by \tilde{B} to obtain the first-order VAR representation for \tilde{z}_t ,

$$\tilde{z}_t = \tilde{B}P\tilde{B}^{-1}\tilde{z}_{t-1} + \tilde{B}\epsilon_t.$$

Then $\Phi(\theta)$ is constructed from $\tilde{B}P\tilde{B}^{-1}$, and $\Sigma(\theta)$ corresponds to $\tilde{B}V\tilde{B}'$, where V is the variance-covariance matrix of ϵ_t . This establishes that the variables in the CGG1 model has a first-order VAR representation.

ADDITIONAL REFERENCES

- Christiano, L. J. (1989), "P*: Not the Inflation Forecaster's Holy Grail," *Federal Reserve Bank of Minneapolis Quarterly Review*, 13, 3–18.
- Clarida, R., Gali, J., and Gertler, M. (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–180.
- Gali, J., Lopez-Salido, J. D., and Valles, J. (2003), "Technology Shocks and Monetary Policy: Assessing the Fed's Performance," *Journal of Monetary Economics*, 50, 723–743.

Comment

A. Ronald GALLANT

Duke University, Fuqua School of Business, Durham, NC 27708 (arg@duke.edu)

My overall impression is that this is an important and very well-written article. It establishes a standard that future empirical work in macroeconomics should meet to be taken seriously by the scientific community. The empirical results are relevant and persuasive: (1) Dynamic stochastic general equilibrium (DSGE) models fit the data about as well as feasible reduced-form models, (2) DSGE models are as reliable for some policy purposes as feasible reduced-form models (e.g., monetary policy), and (3) some neo-Keynesian features are essential to fit the data (e.g., habit persistence).

The fundamental problem that this article addresses is that macroeconomic data are sparse. It is difficult to make progress in empirical work without serious use of prior information. This fact compels the use of Bayesian methods. Two works that drive this point home are those of Bansal, Gallant, and Tauchen (2004) and Gallant and McCulloch (2005). In the former, one is compelled to use relatively low-quality dividend data and make a counterfactual assumption that the data are conditionally homoscedastic to estimate the parameters of two general equilibrium models. In the latter, the use of prior information allows

dividends to be treated as unobserved and the conditional heteroscedasticity of the data to be incorporated into the analysis. The statistical methods used in these two works are, loosely speaking, the frequentist and Bayesian nonlinear analogs of the methods advocated in the present article.

There are five key ingredients to the methodology advocated in this article: (1) an auxiliary model, which is a vector autoregression (VAR); (2) a structural model, which is a DSGE; (3) a prior that forces the auxiliary model to mimic the structural model for large values of a hyperparameter λ and which produces a model that is a blend of the two for smaller λ ; (4) posterior probabilities of model plausibility indexed by λ ; and, to my mind the most important, (5) model adequacy expressed in terms of a posteriori values of relevant functionals of the model indexed by λ (e.g., impulse response curves). This last feature

allows us to deal with the admitted fact that the models involved are approximations and yet assess their adequacy and usefulness in terms relevant to the underlying scientific discipline rather than dismiss them out of hand by failure to adequately approximate some aspect of the data that is effectively irrelevant to the scientific discipline.

As stated earlier, my overall impression is that the article is important and very well written, that it establishes a standard for empirical work in macroeconomics to be taken seriously by the scientific community, and that the empirical results are relevant and persuasive. Nonetheless, a discussant's job is to quibble. When evaluating the quibbles that follow, the reader should bear in mind that a response by the authors, although feasible, would entail substantial additional computational effort, because the quibbles essentially advocate the use of nonlinear methods instead of linear methods.

The auxiliary model, the VAR, is conditionally homoscedastic. For the data used here, this is counterfactual. The data are conditionally heteroscedastic. This fact was documented by, for instance, Bansal and Lundblad (2002). The bias caused by ignoring conditional heterogeneity can be large in the sorts of models considered in this article, as documented by Gallant and McCulloch (2005). The variance mismatch shown in table 1 suggests that the DSGE can generate conditional heteroscedasticity, and thus it should be matched to an auxiliary model that can also accomplish this so that the DSGE is allowed to track this important feature of the data. It is in fact practicable using parallel equipment to compute the binding function from a DSGE to a semiparametric auxiliary model with a conditionally heteroscedastic variance function (Gallant and McCulloch 2005).

The DSGE is not actually solved, but rather the solution is approximated by a log-linear function whose coefficients are nonlinear functions of model parameters. What passes for the DSGE model is actually the driving processes passed through a filter. How accurate is this approximation? Probably reasonably accurate for reasonable values of θ . But MCMC subjects the DSGE to unreasonable values of θ . It is worth pointing out that there is a Bayesian method analogous to generalized method of moments that avoids the need to solve the model and also

can handle high-dimensional observations (Gallant and Hong 2006). But probably if that method were used here, then some parameters of the DSGE would not be identified by the data used here.

Bayesian inference is subjective; therefore, the authors have an absolute right to their own choice of prior. This is nonnegotiable. The author's prior does appear to have some congruency and computational advantages. Nonetheless, it is customary for discussants of Bayesian applications to criticize the prior, and I do not want to break with tradition. At first glance, the treatment of the scale parameter of the prior appears to be absurd because of the way in which λ enters it. It takes about four pages of discussion to convince the reader that the prior is not absurd. A simple discrepancy prior stated as λ times some norm of the difference between the location and scale given by the binding function and the location and scale of the auxiliary model would be understood instantly. A simple discrepancy prior can be made scale-invariant and is practicable (Gallant and McCulloch 2005).

To close, let me repeat that this is an important article that establishes a standard that empirical work in macroeconomics must meet to be taken seriously by those who work at some distance from that field.

ACKNOWLEDGMENT

This research was supported by National Science Foundation grant SES 0438174.

ADDITIONAL REFERENCES

- Bansal, R., Gallant, A. R., and Tauchen, G. (2004), "Rational Pessimism, Rational Exuberance, and Markets for Macro Risks," manuscript, Duke University, Fuqua School of Business, available at www.duke.edu/~arg.
- Bansal, R., and Lundblad, C. (2002), "Market Efficiency, Asset Returns, and the Size of the Risk Premium in Global Equity Markets," *Journal of Econometrics*, 109, 195–237.
- Gallant, A. R., and Hong, H. (2006), "A Statistical Inquiry Into the Plausibility of Recursive Utility," manuscript, Duke University, Fuqua School of Business, available at www.duke.edu/~arg.
- Gallant, A. R., and McCulloch, R. E. (2005), "On the Determination of General Statistical Models With Application to Asset Pricing," manuscript, Duke University, Fuqua School of Business, available at www.duke.edu/~arg.

Comment

Christopher A. Sims

Department of Economics, Princeton University, Princeton, NJ 08544 (sims@princeton.edu)

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 para-

metric 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 VARs (with a prior, a VAR becomes a BVAR, or Bayesian VAR) are variants on the Minnesota prior. We need not provide the details of that prior 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 thus incorporates no behavioral interpretations of parameters or equations. Macroeconomists have views on how variables are related and how their properties differ, however. These views are most easily expressed as views about behavioral parameters in DSGE models. Thus the Del Negro–Schorfheide (DS) approach is appealing.

Another approach is that originated by the other two co-authors, Smets and Wouters, who use relatively richly parameterized DSGEs, together with priors on the parameters, to arrive at a DSGE model that fits and forecasts relatively well. The 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 fit better than DSGEs when they are applied to real data (not to processed data that have had trend removed by filtering or regression). But more importantly, aggregate DSGE models are story-telling devices, not hard scientific theories. We know that there is no aggregate capital stock and no aggregate consumption good. We know that the real economy has a rich array of financial markets that 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 that these models 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 and not a DSGE model fit to the data was at the forefront in earlier work by Del Negro and Schorfheide. The present articles 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 VARs as a standard of comparison for DSGEs, 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. Although such comparisons are helpful, they can be hard to interpret. The methodology 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 and 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 parameter 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 VECTOR AUTOREGRESSION

Although this article 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 VARs—which is what emerge 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. BVARs 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 by 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 using of a proper but “economics-free” prior on the VAR (e.g. some version of the “Minnesota 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, prove 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 interpreted as stochastic shifts in policy behavior, which 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 required to produce that path for the policy variables. The DS notation for these two models is

$$\text{RF: } y = \Phi(L)y + u, \quad \text{var}(u) = \Sigma_u,$$

$$\text{SVAR: } C(L)y = \varepsilon, \quad \text{var}(\varepsilon) = I,$$

$$\text{Connection: } A_0 A_0' = \Sigma_u, \quad A_0^{-1} \cdot (I - \Phi(L)) = C(L).$$

The reduced-form and structural VARs are connected through the foregoing relation between A_0 and Σ_u . The DSGE implies a matrix $A_0(\theta)$ that connects the DSGEs implied reduced-form VAR to its implied SVAR. We 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, $A_0 = A_0(\theta)$. But this is unappealing, 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 completely trusting the DSGE assertions about contemporaneous relations among variables, while treating its assertions about lagged effects as uncertain.

So DS do something else: 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

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

[Here $\text{chol}(X)$ is the Choleski factor of X .] In other words, the “rotation” matrix Ω is treated as nonstochastic, conditional on θ , whereas the lower triangular part of the QR decomposition is treated as a priori random, 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), calculating the QR decomposition of $A_0(\theta)$, then applying (*).

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

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

4. IMPROVEMENTS III: MORE EMPHASIS ON LOW FREQUENCIES

The use of DSGEs 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 the high-frequency behavior of DSGEs. This is quite explicit in the Bank of England’s monograph rationalizing its recently developed BEQM model, and 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 “DSGEs that fit” is that to fit well, they need 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 (nonstochastically), but otherwise it mimics information from a modest-sized sample. Such a prior inherently is more informative about short-run than long-run behavior. This also could be fixed. We 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.

5. 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 further research into both of these topics as well as into improving the DS methodology itself.

ACKNOWLEDGMENTS

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

Comment

Jon FAUST

Economics Dept., Johns Hopkins University, Baltimore, MD 21218 (faustj@jhu.edu)

At one point more than 3 decades ago, there was something of a consensus among macroeconomists regarding a large-scale model of business cycles. One version of such a model, built in an impressive joint effort of academicians and the Federal Reserve, was beginning to play an important role in policy analysis at the Fed. In the 1970s, this consensus evaporated, as bad economic outcomes and apparent policy mistakes were associated with a breakdown in model performance.

While we continue to apportion blame for these events, three critiques of model failure are important. For concreteness, and

without abusing historical accuracy too much, I call these the Lucas (1976), Sims (1980), and Hendry (1985) critiques. The Lucas critique revealed deep difficulties in the very nature of policy analysis; the Sims and Hendry critiques were more about practical standards of good practice in model building. Sims

(1980) argued that we had no basis for believing the assumptions used to identify the models; arbitrary identifying assumptions lead to arbitrary answers and unreliable policy analysis. Sims took pains to demonstrate that this critique has full force even if the models seemed to “fit.” For example, he reminded us that false restrictions can improve forecasting performance. Hendry argued that the models simply did not fit. The models showed a glaring inability to account for arguably important features of their estimation samples.

Something of a new consensus on large-scale macromodeling is emerging; a wide range academics and central bank economists are once again jointly building large-scale macromodels, and central banks are beginning to use these models in policy analysis. This is a very good development. These models—like the models of the last consensus—embed great advances over what came before. I have argued more fully elsewhere (Faust 2005) that these models have an important positive role to play in policy analysis.

Still, as a central banker, I am concerned. Setting aside the vexatious Lucas critique, how can we be confident that we are not repeating the more mundane mistakes highlighted by Hendry and Sims?

The excellent article by Del Negro, Schorfheide, Smets, and Wouters provides a unified framework for answering this question. The marginal likelihood of the hyperparameter, λ , provides one answer to Hendry in the form of a metric for assessing the distance between the reduced form of the models and the data. The impulse response comparisons give a partial response to Sims, shedding light on the degree to which the deviation between model and data casts doubt on the model’s account of causal structure. Neither the desirability nor the possibility of handling these issues in a unified framework has ever been doubted, but practical implementation of such a framework has largely eluded model builders up to now. The demonstration in this article is a very positive contribution.

Still, even armed with these new tools, as a central banker, I am concerned. My Hendry-style reservations are easiest to describe. The DSGE models have implications for many more variables than are used in the empirical analysis. In particular, the models have predictions for the entire term structure of interest rates, yet only short-term interest rates are used in the empirical analysis. The expectations theory of the term structure holds (or almost holds) in the models, and this theory is known to be grossly inconsistent with the data—especially the U.S. data. To put it most contentiously, we have discovered one way to “fit” the dynamics of the quantity of investment: move the long-term interest rate in arbitrary, counterfactual ways. We may echo Hendry in stating that these models show a glaring inability to account for arguably important aspects of the data. The point is as mundane as it was in the 1970s—ignore inconvenient features of the data at your peril.

There is a simple lesson here from the methodological perspective of the article. All macromodels are simplifications and ultimately wrong. The article embeds our *wrong* model in a more general one, allowing us to ask “How wrong is it?” and “In what dimensions is it wrong?” In practice, such tools will be of little value unless we avoid arbitrary and unmotivated zero-one decisions about which empirical facts to allow into the analysis.

Of course, we cannot solve all problems at once. If we begin the analysis with too many empirical implications, then the approach can become computationally infeasible or the results can become messy and difficult to assess. The Bayesian framework provides a nice piecemeal alternative: Start with a “core” set of observable implications as in the article; obtain the posterior for θ ; then evaluate how the posterior changes, say, if we add various additional variables one-by-one. Some such exercise is clearly needed.

Protection against the Sims critique is more tricky. First, we must remember that, although “fit” might be necessary for a good policy analysis model, it is not sufficient. As a profession, we have groped about in the class of friction-laden DSGE models to prove the *existence* of a model that fits macro dynamics. The tools described in this article can help us evaluate existence claims. To answer Sims (and Koopmans), we need to prove *uniqueness*; we must rule out that there are other models in this class with similar fit but different causal structure. Sims (2001) and others (Faust 2005; Leeper 2005) have argued that uniqueness here is a wide-open question.

The tools described in this article play no direct role in proving uniqueness, but the impulse response comparisons shed some light on whether the inadequacies in fit call into question the substance of the causal mechanism of the model. This is the one area in which I disagree a bit with the authors. The goal of the article is to compare the impulse responses implied by some DSGE model parameter, θ , to those implied by a reduced-form VAR that cannot be exactly matched by *any* DSGE model parameter. Of course, we have an identification problem here—which identification of the VAR should we choose for the comparison? The authors choose the identification by taking the Ω implied by θ from the DSGE model and applying it to the reduced form VAR. The article argues that “this implies that we take the DSGE model literally in the directions of the VAR parameter space in which the data are uninformative” (p. 127).

This argument is incomplete; there are arbitrarily many ways to “take the DSGE model literally.” Because the VAR and θ are literally inconsistent, to take certain features literally, we must relax others. Which features to maintain and which to drop is a substantive choice. The authors take literally certain aspects of contemporaneous interactions under θ . Instead, they could have, for example, chosen to take θ ’s implications literally, say, for long-run neutrality of certain shocks. Once again, we confront Sims’ warning against arbitrary restrictions.

Although the choice of identification in the article is arbitrary, the framework of the article provides a nice starting point for a more substantive analysis. We take seriously the fact that the model is ultimately wrong and want to know whether it is wrong in dimensions that matter most to us. We should state a substantive prior over implications of θ that we expect to hold—even in the “true” model, which differs from that of θ . Such features might include, for example, certain long-run implications or Euler equation restrictions.

My suggestions regarding the Sims and Hendry critiques bear a close family resemblance. Regarding the Hendry critique, I said that we should avoid arbitrary zero-one choices about which data implication to expose the model to. Regarding the Sims critique, I am saying we should avoid arbitrary dogmatic choices about which implications of the model to take

literally in evaluating the causal interpretation of a more general model.

When we work with wrong models, we must make astute choices about which features of the data we hope to match, as well as about which theoretical implications of the model that we wish to take seriously (versus take to be uninteresting artifacts of our idealization). Perhaps the key practical problem in a large modeling effort is that these choices can become opaque and difficult to inspect. In one view, Sims and Hendry argued that the failures of the last consensus models were due to insufficient vigilance regarding these choices.

As a central banker, I find that this article makes me optimistic about the new consensus models. To be sure, we currently have at best a poor response to the Hendry and Sims critiques as applied to the new models. But the unified framework presented in this article illuminates a route to replacing formerly opaque and dogmatic choices with explicit and nuanced choices about which bits of data and theory we take seriously. To paraphrase Sims (1980), a long road remains to be traveled here. The opportunities for progress look to be immense.

Note. This work was completed while the author was an economist in the International Finance Division at the Federal Reserve Board. The views in this article are solely the responsibility of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff.

ADDITIONAL REFERENCES

- Faust, J. (2005), "Is Applied Monetary Policy Analysis Hard?" manuscript, Johns Hopkins University, Economics Dept., available at <http://e105.org/faustj/download/faustbdsgefit.pdf>.
- Hendry, D. (1985), "Monetary Economic Myth and Econometric Reality," *Oxford Review of Economic Policy*, 1, 72–84.
- Leeper, E. M. (2005), Discussion of "Price and Wage Inflation Targeting: Variations on a Theme by Erceg, Henderson and Levin," by M. B. Canzoneri, R. E. Cumby, and B. T. Diba, in *Models and Monetary Policy: Research in the Tradition of Dale Henderson, Richard Porter, and Peter Tinsley*, eds. J. Faust, A. Orphanides, and D. Reifschneider, Washington, DC: Federal Reserve Board, pp. 216–223.
- Lucas, R. (1976), "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.
- Sims, C. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- (2001), Comments on Papers by J. Gali and by S. Albanesi, V. V. Chari, and L. J. Christiano, manuscript, Princeton University, Economics Dept.

Comment

Lutz KILIAN

Department of Economics, University of Michigan, Ann Arbor, MI 48109 (lkilian@umich.edu)

1. INTRODUCTION

Empirical adaptations of the Keynesian model date back to the early days of econometrics. The traditional partial-equilibrium Keynesian model was devoid of dynamics. It took partial adjustment and adaptive expectations models to make these inherently static models suitable for econometric analysis. The resulting highly restrictive dynamics of this first generation of Keynesian empirical models were economically implausible, especially from a general equilibrium standpoint, contributing to the demise of these models in the 1980s (see Sims 1980).

The second generation of Keynesian empirical models, exemplified by Gali (1992), embedded the Keynesian model in a structural vector autoregression (VAR). Restrictions implied by the static partial equilibrium Keynesian model were imposed in modeling the contemporaneous interaction of the shocks. Together with long-run neutrality restrictions on the effect of demand shocks on the level of output, these restrictions served to identify the structural parameters, while the lag structure of the model was left unconstrained. Although this structural VAR approach dispensed with the strong restrictions on the dynamics implied by partial adjustment and adaptive expectations models, it remained suspect because of the built-in partial equilibrium assumptions and lack of microfoundations.

As empirical representations of traditional Keynesian models evolved in the 1980s and 1990s, dissatisfaction with the theoretical underpinnings of the traditional Keynesian model led to the

development of the class of New Keynesian models. The latter models are microfounded dynamic stochastic general equilibrium (DSGE) models. Apart from offering more credible identifying assumptions, these DSGE models imply cross-equation restrictions that may improve the efficiency of VAR model estimates compared with traditional Keynesian VAR models. The authors' chief contribution is to provide a coherent econometric framework for evaluating the fit of this third generation of Keynesian empirical models from a Bayesian standpoint.

2. ECONOMETRIC FRAMEWORK

The central idea of the article is to approximate the DSGE model by a VAR and to document how the model fit changes as we relax the cross-equation restrictions implied by the theoretical model. The premise is that we start with the modal belief that a given DSGE model is accurate, but allow for the possibility that our prior beliefs may be wrong. Prior beliefs about the accuracy of the DSGE model have implications for the values of the approximating VAR model parameters. Let the hyperparameter λ be the precision of a zero mean prior distribution over the deviations of the (nearly) unrestricted VAR

parameters from the restricted VAR parameters implied by the DSGE model. DSGE-VAR($\lambda|p$) defines a continuum of models indexed by λ conditional on the lag order p of the approximating VAR model. At one extreme, we have the unrestricted VAR(p) model, which coincides with DSGE-VAR($0|p$); at the other extreme is the (approximate) VAR representation of the DSGE model, denoted by DSGE-VAR($\infty|p$).

The marginal likelihood of λ is interpreted as a measure of in-sample fit with a built-in penalty for model complexity that depends on λ . The lower λ , the higher the complexity of the model (because there are more “degrees of freedom” in fitting the data). DSGE-VAR($\hat{\lambda}|p$) denotes the model evaluated at the peak of the marginal likelihood of λ , corresponding to the degree of model complexity favored by the data.

The log-linearized DSGE model has a state-space form that can be expressed as a VARMA model in the observables, which in turn can be approximated by a finite-order VAR model under weak conditions. The authors first verify that the impulse responses obtained from the exact DSGE-VARMA model are well approximated by the finite-order DSGE-VAR($\infty|4$) model approximation. They then compare the impulse responses of the approximate DSGE-VAR($\hat{\lambda}|4$) and DSGE-VAR($\infty|4$) models in an effort to generate insight into the sources of DSGE model misspecification. Although a lag order of $p = 4$ may be reasonable in this example, more discussion of the principles underlying the selection of p would be useful. Should we select the lag order to minimize the discrepancy between the DSGE-VARMA model and the DSGE-VAR($\hat{\lambda}|p$) model for given sample size or by other means? Does the subsequent analysis remain credible if for a given dataset, no good approximation can be achieved for any p ?

3. WHAT MAKES THE NEW KEYNESIAN MODEL FIT?

DSGE models contain endogenous propagation mechanisms that convert unobservable exogenous shocks into observable data for macroeconomic aggregates. Ideally, we would like these propagation mechanisms to explain all of the persistence and co-movement in the U.S. data. These mechanisms are typically weak, however. Unless the underlying exogenous driving processes are highly persistent, the dynamics of the simulated model data do not match the dynamics of the U.S. data very well. This problem is not specific to the New Keynesian model. The same problem arises in New Classical models. If we restrict the productivity shock in a standard RBC model to white noise, for example, the dynamics of the model are clearly at odds with the data.

The weakness of the propagation mechanism would not be a problem if economic theory were not quiet on the sources and nature of the exogenous shocks in the model. In practice, the number, type, and nature of the exogenous driving processes derive not from theory, but rather from the ingenuity of the empirical researcher. Just as the partial adjustment and adaptive expectation specifications in large-scale econometric models were not part of the original theoretical design, these shock processes are not part of the original theoretical model. Thus, in evaluating the fit of the New Keynesian model, the authors are not evaluating the model per se, but rather are evaluating the

model in conjunction with ad hoc assumptions about the shock processes.

The authors’ approach is to include six exogenous driving processes used in the previous literature: a technology shock, a shock to capital accumulation, government spending shocks, shocks to preference for leisure, an overall preference shock, and shocks to price markups. Some obvious questions are: (1) Were it not for the fact that we feel compelled to justify the use of a large-dimensional VAR model, would we ever have thought of adding all these shock processes?; (2) Holding fixed the dimensionality of the VAR model, could we have selected other shock processes instead that would have been equally reasonable?; and (3) How important is the AR(1) structure chosen for the shock processes in the article? Clearly, alternative specifications will alter the VARMA model structure that the VAR model is designed to approximate.

The consequences of the researcher getting the nature of the exogenous driving processes in the model wrong are of obvious concern. Specifically, the concern is that the apparent fit of the New Keynesian model may have more to do with the inclusion of suitable exogenous driving processes than with the realism of the model structure itself. What does the proposed methodology tell us, for example, when in reality there is no exogenous preference-for-leisure shock but the model allows for such a shock? What happens when we get the number of exogenous shocks right but the type of shocks wrong? What happens when the researcher allows for a highly persistent government spending shock process in a model with weak propagation mechanisms, but in reality that government spending shock process is not persistent? Will the measure of fit favor an incorrect model merely because it gives the model much more flexibility in fitting the data? What if technology follows a random walk, but the variability of technology innovations is much smaller in reality than the model allows for? Would a misspecified model simply inflate the variance of the technology shock to achieve a better fit? This is a concern, given the evidence reported by Basu (1996) that measured Solow residuals may reflect shifts in factor utilization and increasing returns to scale, along with exogenous changes in technology. It would be interesting to explore these issues.

In addition, one would want to be able to compare formally alternative models with the same economic structure, except for differences in the nature of the exogenous driving processes. There is a close link, however, between the number of exogenous driving processes and the dimensionality of the approximating VAR model. Moreover, the rank structure of the VAR model will depend on the nature of the underlying exogenous shocks. In this sense, alternative models are no longer nested, and how to compare them and it is not obvious how to compare them.

Another concern is that many of the model features as well as priors seem to have been obtained based on previous studies of essentially the same dataset. Clearly, we have learned a lot in recent years about what works (e.g., habit persistence) and what does not work when fitting the data. This introduces an element of data mining and blurs the distinction between priors and posteriors. Is it surprising that the New Keynesian model as specified in this article appears to fit the data, given that many of its features have been developed to fit this very dataset? Even

the real-time forecasting exercise does little to allay these concerns, because the model has been specified with knowledge of previous studies that used the entire dataset.

4. HOW REALISTIC IS THE NEW KEYNESIAN MODEL?

An important part of the appeal of the New Keynesian model studied by the authors is that the DSGE- $\text{VAR}(\hat{\lambda}|4)$ impulse responses are “realistic.” By this, the authors mean that relaxing the restrictions implied by the DSGE- $\text{VAR}(\infty|4)$ —which is essentially the DSGE-VARMA model in this example—in favor of the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model does not alter significantly the responses to an unanticipated monetary policy shock.

Realism here is judged only through the prism of a specific economic model structure, the premise being that the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model is realistic. The problem with this approach to measuring realism is that the underlying DSGE model may very well be false. Although the comparison between DSGE- $\text{VAR}(\infty|4)$ and DSGE- $\text{VAR}(\hat{\lambda}|4)$ model responses is informative about the realism of the prior mean of the values of the model parameters chosen for this specific DSGE model, it is not necessarily informative about the realism of the DSGE model structure itself.

The traditional approach to judging the realism of a DSGE model is to compare the time series properties of data generated from the DSGE model with those obtained from a reduced-form representation of the U.S. data. One method is to compare cross-autocorrelation matrices, as in the RBC literature; another is to focus on reduced-form impulse response analysis from unrestricted approximating VAR models, on multivariate spectral representations of the data (see Diebold, Ohanian, and Berkowitz 1998) or on statistical measures of co-movement (see Den Haan 2000). Such VAR (or VARMA) comparisons still implicitly invoke the assumption that there are at least as many shocks in the DSGE model as in the VAR model. If we are unsure of the number of exogenous driving processes, and are unwilling to make ad hoc assumptions about “stochastic noise” or measurement error (as in Rotemberg and Woodford 1997), the only remaining avenue is to focus on statistics that can be computed based on univariate representations of the data. An example of this approach is the analysis of univariate predictability measures presented by Diebold and Kilian (2001) and Inoue and Kilian (2002).

The problem with using reduced-form VAR evidence as the benchmark, as the authors point out, is that the unrestricted VAR model or DSGE- $\text{VAR}(0|4)$ model is dominated in terms of its time series fit by the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model. The authors present evidence that imposing some DSGE structure improves the in-sample fit of the VAR model, even when there is strong evidence that the DSGE- $\text{VAR}(\infty|4)$ model is misspecified. Unlike the evaluation of the realism of the DSGE- $\text{VAR}(\infty|4)$ model relative to the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model, which is limited to a comparison of specific identified impulse responses of interest to economists, the evidence that the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model is more realistic than the DSGE- $\text{VAR}(0|4)$ model is based on a comparison of their respective marginal likelihoods. The authors interpret this finding as evidence that the DSGE- $\text{VAR}(0|4)$ model should be discarded as a benchmark in favor of the $\text{VAR}(\hat{\lambda}|4)$ model.

As this discussion illustrates, there is a certain asymmetry in the way in which the authors judge the realism of alternative models. Judged by the marginal likelihood, the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model dominates not only the DSGE- $\text{VAR}(0|4)$ model, but also the DSGE- $\text{VAR}(\infty|4)$ model (see fig. 2 in the article). Rather than discard the DSGE- $\text{VAR}(\infty|4)$ model as they did with the DSGE- $\text{VAR}(0|4)$ model, the authors resurrect the DSGE- $\text{VAR}(\infty|4)$ for all practical purposes by appealing to the similarity of its responses to monetary policy shocks to those from the DSGE- $\text{VAR}(\hat{\lambda}|4)$. By the same token, they could have compared selected reduced-form statistics based on the DSGE- $\text{VAR}(\hat{\lambda}|4)$ and DSGE- $\text{VAR}(0|4)$ models to argue that the formal rejection of the DSGE- $\text{VAR}(0|4)$ model perhaps does not mean that the model does not provide a good benchmark for all practical purposes. At least they do not provide evidence to the contrary.

If we accept the marginal likelihood of the model as the metric for model comparisons, then the unrestricted VAR model ceases to be the relevant benchmark for model comparisons. What then is the relevant benchmark for comparing alternative models? The authors’ view that the reduced-form properties of the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model are a more suitable benchmark than those of the DSGE- $\text{VAR}(0|4)$ model seems compelling, as far as this article is concerned, but one would not expect this particular DSGE model to be the only DSGE model that helps improve the fit of the unrestricted VAR model or even the model that improves it the most. Improved fit may be implied by any number of potentially nonnested DSGE model structures relative to the approximating VAR model appropriate for that specific DSGE model. Thus the relevant comparison for finding a new benchmark for the literature seems to be across alternative DSGE- $\text{VAR}(\hat{\lambda}|p)$ models with different VAR dimensions, variables, and lag structures. It is not clear from the article how to conduct such a comparison of the realism of alternative models or what is the correct metric for evaluating alternative DSGE- $\text{VAR}(\hat{\lambda}|p)$ models. Clearly, the λ metric is specific to a given DSGE model and cannot be used to compare alternative DSGE models except when one model nests the other, as in the examples involving habit formation and price and wage indexing provided by the authors.

5. ON THE FORECASTING ABILITY OF THE NEW KEYNESIAN MODEL

The article also investigates the forecasting ability of the New Keynesian model at horizons of up to 2 years. The question of forecasting ability is of interest primarily to central bankers and those who seek to understand monetary policy decisions. The forecasting problem faced by central bankers differs from the forecasting problem typically studied by academics. The difference is that central bankers want not only accurate forecasting models (as measured by, e.g., the out-of-sample prediction mean squared error), but also forecasting models that lend themselves to economic interpretation or story telling. This fact helps explain why large-scale macroeconomic models have survived at central banks long after their credibility has been undermined in academic circles. The New Keynesian model studied by the authors is intended to provide an alternative to large-scale models derived from traditional Keynesian partial

equilibrium models. Its attractiveness is that it retains the economic interpretability of traditional econometric models, while being more microfounded, general equilibrium, and tractable. Of course, ease of interpretation may come at a price in terms of out-of-sample forecast accuracy.

The authors show that the DSGE- $\text{VAR}(\hat{\lambda}|4)$ model is about as accurate as or even more accurate than the unrestricted $\text{VAR}(4)$ model in forecasting from rolling regressions, whereas the DSGE- $\text{VAR}(\infty|4)$ model is distinctly less accurate in some cases. Although the unrestricted VAR model may seem to be a natural benchmark in this context, it is something of a straw man in that in practice no one would rely on unrestricted VAR models for forecasting macroeconomic aggregates. Within the class of linear reduced-form models, more realistic competitors would include Bayesian VAR models developed for forecasting, forecasts from single-equation models based on shrinkage estimators, factor model forecasts, and forecast combination methods for large cross-sections of time series data. It is less than obvious that the DSGE- $\text{VAR}(\infty|4)$ model would remain competitive with the latter forecasting models.

There is another key difference between forecasting exercises conducted by central bankers and forecasting exercises studied by most academics, however, suggesting that the latter comparison would not be the right comparison either. Central bankers invariably are interested in forecasting macroeconomic aggregates, such as inflation or output growth conditional on a pre-specified path of interest rates. Such conditional forecasts require a structural model (see Waggoner and Zha 1999). None of the standard methods of forecasting through forecast combination or shrinkage allows the imposition of an exogenous path

of the interest rate in prediction. Thus the most relevant comparison of the forecasting ability of the New Keynesian model would be against suitable alternative structural models, given the same path of the interest rate.

6. CONCLUSION

The improvement of structural time series models for macroeconomic policy analysis is a central task if time series analysis is to retain its importance for economic policy making. The authors are to be commended for having taken on this task. Although I have focused on potential weaknesses in the article, as befits a discussant, this criticism should not obscure the fact that the analysis in this article is impressive and likely to frame the discussion of structural model evaluation for years to come.

ADDITIONAL REFERENCES

- Basu, S. (1996), "Procyclical Productivity: Increasing Returns or Cyclical Utilization?" *Quarterly Journal of Economics*, 111, 719–751.
- Den Haan, W. J. (2000), "The Co-Movement Between Output and Prices," *Journal of Monetary Economics*, 46, 3–30.
- Diebold, F. X., and Kilian, L. (2001), "Measuring Predictability: Theory and Macroeconomic Applications," *Journal of Applied Econometrics*, 16, 657–669.
- Gali, J. (1992), "How Well Does the IS–LM Model Fit Postwar U.S. Data?" *Quarterly Journal of Economics*, 107, 709–738.
- Inoue, A., and Kilian, L. (2002), "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in $\text{VAR}(\infty)$ Models," *International Economic Review*, 43, 309–332.
- Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- Waggoner, D. F., and Zha, T. (1999), "Conditional Forecasts in Dynamic Multivariate Models," *Review of Economics and Statistics*, 81, 639–651.

Rejoinder

Marco Del NEGRO

Federal Reserve Bank of Atlanta, Atlanta, GA 30309 (marcodelnegro@frbatlanta.org)

Frank SCHORFHEIDE

Department of Economics, University of Pennsylvania, Philadelphia, PA 19104 (schorf@ssc.upenn.edu)

Frank SMETS

European Central Bank, D-60311 Frankfurt, Germany (Frank.Smets@ecb.int)

Rafael WOUTERS

National Bank of Belgium, B-1000 Bruxelles, Belgium (Rafael.Wouters@nbb.be)

We would like to thank all of the discussants for their stimulating comments. The comments contain many useful suggestions on how to extend and improve our framework for evaluation, forecasting, and policy analysis with dynamic stochastic general equilibrium (DSGE) models. Taken as a whole, the comments outline an entire research agenda, and as Jon Faust puts it: "The opportunities for progress look to be immense." In this rejoinder we briefly revisit some of the issues that were raised.

1. IDENTIFICATION

Several commentators either implicitly or explicitly raised the issue of identification. There are two dimensions to the identification problems. The first dimension has to do with

observational equivalence of different DSGE models, whereas the second dimension is related to the identification of structural shocks in the DSGE–vector autoregression (DSGE–VAR) model. We begin with the broader question of observational equivalence.

1.1 Observational Equivalence

Are there other DSGE models with similar fits but different causal structures (Faust)? This is a key question in the research agenda. Many central banks around the world are currently building and estimating DSGE models for use in policy analysis. These models often share the same features and—when estimated with Bayesian methods—similar priors for the deep parameters. Yet there may exist other DSGE models that are (almost) observationally equivalent (i.e., they fit the data just as well) but have very different policy implications, for instance, models in which nominal rigidities play a less important role. A related question is: Does the apparent fit of the DSGE model have more to do with the inclusion of suitable exogenous driving processes than with the realism of the model structure itself (Lutz Kilian)? The following simple example shows that the two questions are related. Consider the following two one-equation models:

$$\mathcal{M}_1: \quad y_t = \frac{1}{\alpha} \mathbb{E}_t[y_{t+1}] + u_t, \\ u_t = \rho_1 u_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$$

and

$$\mathcal{M}_2: \quad y_t = \frac{1}{\alpha} \mathbb{E}_t[y_{t+1}] + \rho_2 y_{t-1} + u_t, \\ u_t = \epsilon_t, \quad \epsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2).$$

Focusing on parameters that lead to a unique stable rational expectations solution, it can be verified that both models are observationally equivalent. The exogenous propagation mechanism in \mathcal{M}_1 and the endogenous propagation mechanism of \mathcal{M}_2 lead to identical AR(1) reduced-form dynamics. However, changes in α have different effects on the law of motion for y_t in the two specifications.

If we apply DSGE–VAR approach to model \mathcal{M}_1 only, then we are essentially assessing the fit of an (restricted) AR(1) model relative to a more general autoregressive specification. Such an analysis does not generate a warning flag that the data could have been generated from \mathcal{M}_2 instead of \mathcal{M}_1 . However, if we were to apply the DSGE–VAR analysis to both \mathcal{M}_1 and \mathcal{M}_2 , then (under commensurable prior distributions for the two models) the marginal likelihood functions $p(\lambda|Y, \mathcal{M}_i)$ for the two models would be the same, indicating observational equivalence.

The DSGE–VAR procedure in its current state allows for only local departures from the DSGE model under consideration. As such, it does not help uncover structural forms that are radically different. The important task of contemplating alternative models—and of setting up the estimation in such a way that these alternatives are given a fair chance—is left to the modeler. Again, this is a key issue for central banks, because these alternative may have very different policy implications. Once alternative DSGE models have been set up, one can think of generalizing the DSGE–VAR procedure to allow priors from these different models. We have not yet explored this route.

1.2 Identification in the DSGE–VAR

With respect to the second aspect of identification (i.e., identification of structural shocks in the DSGE–VAR), Faust, Kilian, and Chris Sims voice some well-taken concerns. Our approach essentially amounts to asking whether we can construct a structural VAR that fits better than the underlying DSGE model yet inherits most of the structural dynamic responses from the DSGE model. Again, it is important to keep in mind that an affirmative answer does not solve the *uniqueness* problem posed in Faust’s comment. However, at a minimum we now have a structural VAR with impulse responses that we can interpret in the context of a general equilibrium macro model. If the data are generated by the DSGE models, then our identification scheme ensures that the impulse responses of the DSGE–VAR($\hat{\lambda}$) and the DSGE–VAR(∞) are (in large samples) identical. Moreover, if the state-space representation of the DSGE model is well approximated by a finite-order VAR (as we document in fig. A.1), then our identification ensures that the DSGE–VAR($\hat{\lambda}$) essentially recovers true shocks driving the DSGE model. If, on the other hand, the responses of DSGE–VAR($\hat{\lambda}$) and DSGE–VAR(∞) do not line up, then we can potentially learn from the discrepancies how to modify the DSGE model to improve its fit.

Sims points out that our identification scheme for the DSGE–VAR is potentially sensitive to the ordering of variables. Note, however, that this sensitivity is quite different from the traditional problem of ordering variables in VARs that are identified based on an recursive identification scheme. Consider the following example. Suppose that

$$A_0(\theta) = \begin{bmatrix} a_{11}(\theta) & 0 \\ a_{21}(\theta) & a_{22}(\theta) \end{bmatrix}.$$

Our procedure implies that $\Omega^*(\theta) = \mathcal{I}$ for all θ . Thus the DSGE–VAR inherits the property that the first variable, say $y_{1,t}$, does not respond to the second shock, say $\epsilon_{2,t}$. By reversing the order of the endogenous variables, we obtain the following DSGE–VAR response:

$$\frac{\partial y_{1,t}}{\partial \epsilon_{2,t}} = -\Sigma_{21}^{tr} \sin[\tan^{-1}(-a_{22}/a_{21})] \\ + \Sigma_{22}^{tr} \cos[\tan^{-1}(-a_{22}/a_{21})],$$

which is 0 only if $-a_{22}/a_{11} = \Sigma_{22}^{tr}/\Sigma_{21}^{tr}$. Here Σ^{tr} is the lower-triangular matrix obtained from the Cholesky decomposition of Σ , where Σ is the innovation covariance matrix of the VAR after reordering.

Although we have not explored the sensitivity to reordering of the impulse responses reported in the article, we would expect the effects to be small, at least for moderate and large values of λ . Moreover, if the DSGE model implies zero restrictions on matrix $A_0(\theta)$ that one also would like to impose on the DSGE–VAR, when the endogenous variables should be sorted accordingly.

To the extent that $\Sigma^{tr}(\hat{\lambda})$ is close to $\Sigma^{tr}(\infty)$, our procedure implicitly matches short-run responses of DSGE model and DSGE–VAR, which is undesirable if one believes that the DSGE models miss some of the short-run adjustment dynamics. More generally, given the premise that the DSGE model is potentially misspecified, it might be desirable to consider random perturbations of the $\Omega^*(\theta)$ matrix as suggested by Sims and Faust. We plan to address these issues in future research.

2. IGNORING IMPORTANT ASPECTS OF THE DATA AND THE MODEL

Several comments fall into this category. As pointed out by Larry Christiano and Ron Gallant, nonnormality of error terms as well as heteroscedasticity are of concern *ex post*. Our distributional assumptions are driven mostly by computational convenience. A conjugate normal–inverse–Wishart prior combined with a Gaussian VAR likelihood function lets us easily integrate out the VAR parameters and construct a marginal posterior distribution for the DSGE model parameters. The downside of this simplification is that outliers might have a strong influence on parameter estimates and marginal likelihood values. In this respect, the simulations reported by Christiano are encouraging. Although heteroscedasticity is not as pronounced in quarterly macroeconomic data observations as it is in the financial datasets studied by Gallant, two features of the macroeconomic data certainly deserve more careful attention: (1) the inability of constant-parameter interest rate rules to fit the Volcker disinflation episode and (2) the overall reduced volatility in the macro aggregates in the mid-1980s, often termed the “great” moderation. Modeling these aspects of the data more carefully and reexamining the marginal likelihood functions is an extension certainly worth pursuing.

Faust points toward the more general problem of completely ignoring certain observations, such as the yield curve. He raises a key question for the current generation of DSGE model builders at central banks: Should central banks try to build a “catholic” model, that is, a model that tries to describe “all” of the macroeconomic data and thus is very large, or should they build different (smaller) models aimed at addressing different issues? Again, the DSGE–VAR offers no help when facing these broad questions.

Although our framework is ideal for controlling the relative weight placed on economic theory, in its current form, it is not suited for controlling the relative weight of particular observations in the estimation and would certainly require substantial modification to do so. One interpretation of the DSGE–VAR framework, stressed in earlier work (Del Negro and Schorfheide 2004), is that of mixed estimation; that is, the VAR is estimated based on the actual data and artificial data generated from the theoretical model. In the current implementation we use only a single hyperparameter, λ , that controls the relative number of actual and artificial observations. By introducing additional hyperparameters, one could essentially create more sophisticated weighting schemes for the artificial observations. Although we have not yet explored placing different weight on different series (e.g., weighting quantities differently than asset prices), ongoing research (by Del Negro, Diebold, and Schorfheide) is taking Fourier transformations of the artificial observations and reweightings them according to frequency, following up on Improvement III suggested by Sims.

Gallant stresses that we work with a linearized approximation of the DSGE model. We do this mostly for computational convenience. It would be relatively straightforward to replace the linear solution technique with, say, a higher-order perturbation method, retaining the linear structure of the VAR, or the auxiliary model in the language of indirect inference. However, a more interesting extension would be to make the VAR nonlinear as well, by, for instance, introducing time-varying parameters. Again, this extension is left for future research.

3. MARGINAL LIKELIHOOD OF λ

We are very grateful to Christiano for conducting extensive simulation exercises and illustrating the sampling properties of the marginal likelihood function of λ as well as its peak $\hat{\lambda}$. For the AR(1) example in Section 3.4 of our article, we can actually calculate the sampling distribution of $\hat{\lambda}$ analytically. For simplicity, we assume that the DSGE model implies that $\phi^* = 0$ and that the misspecification of the DSGE model is small in the sense that the data-generating process is given by

$$y_t = (\phi^* + \tilde{\phi}^\Delta T^{-1/2})y_{t-1} + u_t.$$

As $T \rightarrow \infty$, the sampling distribution of $\hat{\lambda}$ converges to

$$\hat{\lambda} \Rightarrow \begin{cases} \frac{1}{(\tilde{\phi}^\Delta + \mathcal{Z})^2 - 1} & \text{if } (\tilde{\phi}^\Delta + \mathcal{Z})^2 > 1 \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{Z} \sim \mathcal{N}(0, 1)$. Note that as the misspecification $\tilde{\phi}^\Delta$ approaches 0, the probability of obtaining $\lambda = \infty$ increases, although it does not converge to 1 even for $\tilde{\phi}^\Delta = 0$. In contrast, for large values of $\tilde{\phi}^\Delta$, the probability of $\hat{\lambda} = \infty$ is small in large samples. The qualitative implications of this simple example coincide with the simulation results obtained from the much more sophisticated and realistic setup studied by Christiano.

Interestingly, in the AR(1) example, $\hat{\lambda}$ is simply a monotone transformation of the likelihood ratio statistic, LR_T , for a test of $\tilde{\phi}^\Delta = 0$ that converges in distribution to

$$LR_T \Rightarrow (\tilde{\phi}^\Delta + \mathcal{Z})^2.$$

Thus a 5% critical value of 3.84 for the likelihood ratio statistic LR_T translates into a critical value of .35 for $\hat{\lambda}$. However, the goal of our article is not to develop a classical test of the hypothesis that the DSGE model restrictions are satisfied; instead, we stress the Bayesian interpretation of the marginal likelihood function $p(\lambda|Y)$, which does not require any cutoff or critical values.

We agree that it is important to study the entire shape of the marginal likelihood function $\ln p(Y|\lambda)$ (as we do in figs. 2 and 4) rather than simply report the peak $\hat{\lambda}$. Again, a look at the simple example can be instructive. It can be verified that

$$\begin{aligned} \ln \left[\frac{p(Y|\lambda = \hat{\lambda}, \phi^* = 0)}{p(Y|\lambda = \infty, \phi^* = 0)} \right] \\ \Rightarrow \begin{cases} \frac{1}{2} [((\tilde{\phi}^\Delta + \mathcal{Z})^2 - 1) - \ln(\tilde{\phi}^\Delta + \mathcal{Z})^2] & \text{if } (\tilde{\phi}^\Delta + \mathcal{Z})^2 > 1 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Suppose that $\tilde{\phi}^\Delta = 0$. Even if $\hat{\lambda}$ is as low as .35, the odds of $\lambda = \hat{\lambda}$ versus $\lambda = \infty$ are only $\exp(.75) \approx 2.1$, because $x^2 - 1 \approx \ln x^2$ for small values of x^2 . As the misspecification $\tilde{\phi}^\Delta$ increases, the average gap between the marginal likelihood function at its peak and at $\lambda = \infty$ also increases. These simple analytical calculations are qualitatively in line with the simulation evidence provided by Christiano.

Both Christiano and Kilian raise the issue of lag length selection, which we did not address formally in our article. There are two dimensions to the choice of lags. The first dimension is the

fit of the resulting empirical specification, that is, the DSGE–VAR($\hat{\lambda}$). A fairly natural procedure would be to determine the lag length based on the marginal likelihoods associated with DSGE–VAR($\hat{\lambda}|p$) for $p = 1, \dots, p_{max}$ (using Kilian's notation) and then condition on the resulting \hat{p} . Alternatively, the marginal likelihoods in figures 2 and 4 could be drawn as functions of both λ and p , or the lag length parameter p could be integrated out conditional on each λ . As Christiano's simulations confirm, the benefit gained from shrinking toward the theoretical model increases with the more free parameters included in VAR. The second dimension of the lag length choice is related to the accuracy of the VAR(p) approximation of the state-space representation of the DSGE model. We view this consideration as secondary. As in an indirect inference framework, there is no need for the auxiliary model to nest the underlying theoretical model. Having said this, a more accurate the VAR approximation certainly makes it easier to deduce from the gap between the DSGE–VAR($\hat{\lambda}$) and DSGE–VAR(∞) responses how best to modify the underlying DSGE model.

4. FORECAST EVALUATION

Christiano argues in his comments in favor of pseudo-out-of-sample forecast error statistics, such as root mean squared errors or the log determinant of the forecast error covariance matrix. While we certainly see value in these statistics we would like to clarify the relationship between marginal likelihoods and forecast error statistics. Define $Y^{t-1} = [y_1, \dots, y_{t-1}]$. We rewrite the marginal likelihood function as

$$\ln p(Y|\lambda) = \sum_{t=1}^T \ln \left[\int p(y_t|Y^{t-1}, \Phi, \Sigma) \times p(\Phi, \Sigma, \theta|Y^{t-1}, \lambda) d(\Phi, \Sigma, \theta) \right]. \quad (*)$$

Under a Gaussian likelihood

$$p(y_t|Y^{t-1}, \Phi, \Sigma) \propto |\Sigma| \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma^{-1} (y'_t - x'_t \Phi)' (y'_t - x'_t \Phi)] \right\}$$

and the marginal likelihood has indeed the flavor of a one-step-ahead pseudo-out-of-sample forecast error statistic. If outliers are a concern, then for the same reason that the Gaussian likelihood should be replaced by, say, a t likelihood, the squared

error statistics, say e^2 , should be replaced by something that is robust to outliers, such as $\ln(1 + e^2)$.

By default, the summation in (*) starts at $t = 1$, which makes the marginal likelihood sensitive to the choice of prior. Although for large values of t , the posterior distribution of the parameters $p(\Phi, \Sigma, \theta|y_1, \dots, y_{t-1}, \lambda)$ is heavily influenced by the sample information, this is not the case for small values of t . For instance, as $\lambda \rightarrow 0$, the density

$$\int p(y_1|\Phi, \Sigma) p(\Phi, \Sigma, \theta|\lambda) d(\Phi, \Sigma, \theta)$$

becomes increasingly smaller for any fixed value of y_1 , which heavily penalizes an unrestricted VAR. This penalty could be reduced by starting the summation in (*) for $t = T^*$ instead of $t = 1$, which is an important difference between marginal likelihood and pseudo-out-of-sample forecast error comparisons. A similar robustification could be attained either by replacing the marginal likelihood $p(Y|\lambda)$ with a predictive likelihood $p(Y|\lambda)/p(y_1, \dots, y_{T^*}|\lambda)$, or, as suggested by Sims, by adding a Minnesota-style prior to the DSGE model prior such that the overall prior remains proper even if $\lambda = 0$. In any case, our article reports both marginal likelihoods, as well as pseudo-out-of-sample forecast error statistics as a function of λ .

Kilian points out that the unrestricted VAR in our forecast error comparison is somewhat of a straw man and that we are excluding many well-established benchmark forecasting models that dominate an unrestricted VAR. In the interest of space, we decided not to include a more comprehensive forecast evaluation in the article. Some of the comparisons were provided (albeit for a smaller DSGE model) in earlier work (Del Negro and Schorfheide 2004), where we found that the DSGE–VAR forecasts are comparable to forecasts with a Bayesian VAR based on a Minnesota prior. As the impulse response analysis shows, the attractive feature of the DSGE–VAR is that its dynamics closely mimic those of a fully specified DSGE model, and thus the forecasts become interpretable in the context of modern macroeconomic theory.

5. CONCLUSION

In closing, we would again like to thank all of the commentators for their stimulating and constructive remarks. We hope to be able to address many of their suggestions in future research, and invite others to take part in this research agenda.