Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

On the Fit of New Keynesian Models

Marco Del Negro[a], Frank Schorfheide[b], Frank Smets[c] & Rafael Wouters[d]
[a] Federal Reserve Bank of Atlanta, Atlanta, GA 30309
[b] Department of Economics, University of Pennsylvania, Philadelphia, PA 19104
[c] European Central Bank, D-60311 Frankfurt, Germany
[d] National Bank of Belgium, B-1000 Bruxelles, Belgium
Published online: 01 Jan 2012.

Editor's Note: The following article was the *JBES* Invited Address presented at the Joint Statistical Meetings, Seattle, Washington, August 6–10, 2006.

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

Journal of Business & Economic Statistics
April 2007, Vol. 25, No. 2
DOI 10.1198/073500107000000016

(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 $i \in [0, 1]$,

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

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

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

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

$$Y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-(1+\lambda_{f,t})/\lambda_{f,t}} Y_t \qquad \text{and} \qquad P_t = \left[\int_0^1 P_t(i)^{-1/\lambda_{f,t}}\,di\right]^{-\lambda_{f,t}}. \tag{3}$$

2.2 Intermediate Goods Producers

Good i is made using the technology

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

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

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

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,

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

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

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

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

$$\max_{\tilde{P}_t(i)} \mathbb{E}_t\left[\sum_{s=0}^{\infty} \zeta_p^s \beta^s \Xi_{t+s}^p \left(\tilde{P}_t(i)\left(\prod_{l=1}^{s} \pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}\right) - MC_{t+s}\right) \times Y_{t+s}(i)\right]$$

$$\text{s.t.}\quad Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i)(\prod_{l=1}^{s}\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p})}{P_{t+s}}\right)^{-(1+\lambda_{f,t})/\lambda_{f,t}} Y_{t+s},$$

$$MC_{t+s} = \frac{\alpha^{-\alpha} W_{t+s}^{1-\alpha} R_{t+s}^{k\ \alpha}}{(1-\alpha)^{(1-\alpha)} Z_{t+s}^{1-\alpha}}, \tag{8}$$

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:

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

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,

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

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,

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

and

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

2.4 Households

The objective function for household j is given by

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

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

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

and

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

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

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

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

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

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

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

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, Hercovitz, and Krusell 1998), which follows the exogenous process

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

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

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

$$\text{s.t.} \quad \text{eq. (15) for } s = 0, \ldots, \infty, \text{(11a), and} \tag{19}$$

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

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

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

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,

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

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 output 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

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

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

$$G_t = (1 - 1/g_t) Y_t, \tag{23}$$

where g_t follows the exogenous process

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

2.6 Resource Constraint

The aggregate resource constraint

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

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

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

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 $x_t = [1, (\beta' y_{t-1})', \Delta y'_{t-1}, \ldots, \Delta y'_{t-p}]'$, and let $\Phi = [\Phi_0, \Phi_\beta, \Phi_1, \ldots, \Phi_p]'$.

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., Gourieroux 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,

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

where Σ_{tr} 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 $\Sigma_u = \Sigma_{tr}\Sigma'_{tr}$.

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

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

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,

$$\Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta) \qquad \text{and} \qquad \Sigma_u^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta). \tag{30}$$

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

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

where Y_* denotes the $\lambda T \times n$ matrix with rows $y_t^{*\prime}$ and X_* is the $\lambda T \times k$ matrix with rows $x_t^{*\prime}$. Taking expectations under the distribution generated by the DSGE model yields

$$\mathbb{E}_\theta^D\left[\ln\left[\frac{\mathcal{L}(\Phi^* + \Phi^\Delta, \Sigma_u^*|Y_*, X_*)}{\mathcal{L}(\Phi^*, \Sigma_u^*|Y_*, X_*)}\right]\right] = -\frac{1}{2}\operatorname{tr}\big[\Sigma_u^{*-1}(\Phi^{\Delta\prime}\lambda T\Gamma_{XX}\Phi^\Delta)\big]. \tag{32}$$

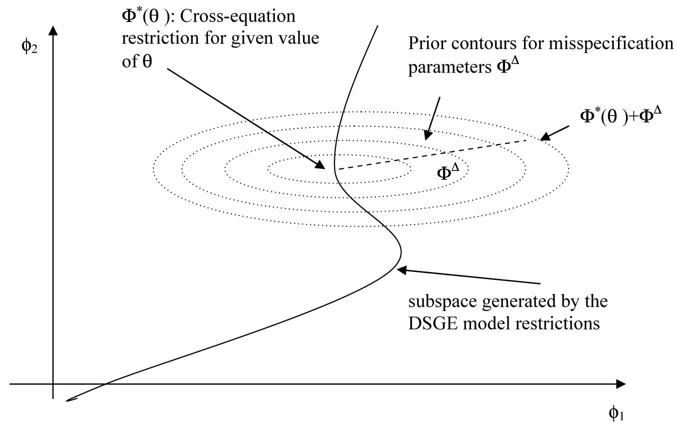

Figure 1. Stylized View of DSGE Model Misspecification. $\Phi = [\phi_1, \phi_2]'$ can be interpreted as the VAR parameters, and $\Phi'(\theta)$ is the restriction function implied by the DSGE model.

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

$$p(\Phi^\Delta|\Sigma_u^*) \propto \exp\left\{-\frac{1}{2}\operatorname{tr}\big[\lambda T\Sigma_u^{*-1}(\Phi^{\Delta\prime}\Gamma_{XX}\Phi^\Delta)\big]\right\}. \tag{33}$$

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

$$\begin{aligned}\Sigma_u|\theta, \lambda &\sim \mathcal{IW}(\lambda T\Sigma_u^*(\theta), \lambda T - k), \\ \Phi|\Sigma_u, \theta, \lambda &\sim \mathcal{N}\left(\Phi^*(\theta), \frac{1}{\lambda T}[\Sigma_u^{-1} \otimes \Gamma_{XX}(\theta)]^{-1}\right),\end{aligned} \tag{34}$$

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,

$$p(\Phi, \Sigma_u, \theta|Y, \lambda) = p(\Phi, \Sigma_u|Y, \theta, \lambda)p(\theta|Y, \lambda). \tag{35}$$

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,

$$\begin{aligned} &\Sigma_u|Y,\theta,\lambda \\ &\quad\sim \mathcal{IW}\big(T(\lambda+1)\hat{\Sigma}_{u,b}(\theta), T(\lambda+1)-k\big), \\ &\Phi|Y,\Sigma_u,\theta,\lambda \\ &\quad\sim \mathcal{N}\big(\hat{\Phi}_b(\theta), \Sigma_u\otimes[T(\lambda\Gamma_{XX}(\theta)+\hat{\Gamma}_{XX})]^{-1}\big), \end{aligned} \tag{36}$$

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

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

$$\lim_{\lambda\longrightarrow 0}\tilde{\sigma}^2(\lambda,\phi^*) = \frac{1}{T}\sum(y_t-\hat{\phi}y_{t-1})^2 \qquad \text{and}$$

$$\lim_{\lambda\longrightarrow\infty}\tilde{\sigma}^2(\lambda,\phi^*) = \frac{1}{T}\sum(y_t-\phi^* y_{t-1})^2,$$

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

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

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

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

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 can be uniquely decomposed into

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

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

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

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 P(1)	Prior P(2)	Prior Interval	DSGE–VECM($\hat{\lambda}$) posterior Mean	DSGE–VECM($\hat{\lambda}$) posterior Interval	DSGE posterior Mean	DSGE posterior 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]
ν_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 ν do so for the $\mathcal{IG}$ distribution, where $p_{\mathcal{IG}}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. 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 QII:1974–QI: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 ν_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 QI:1960–QI: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–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', ν_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–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–VAR(λ) 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–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–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–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 policymakers, 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–VAR(λ) 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,

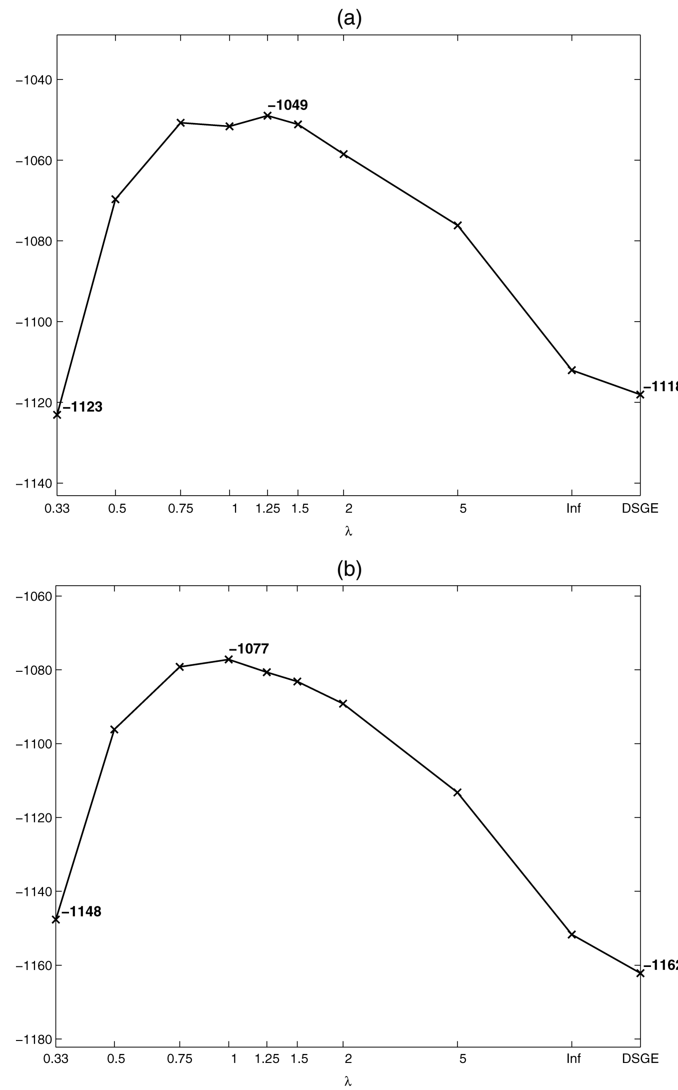

Figure 2. Marginal Likelihood as a Function of λ. (a) 30-year sample: QII:1974–QI:2004. (b) 30-year sample: QII:1970–QI:2000. The two panels depict the log-marginal likelihood function on the y-axis and the corresponding value of λ, rescaled between via the transformation λ/(1+λ), on the x-axis. The 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 particularly 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.

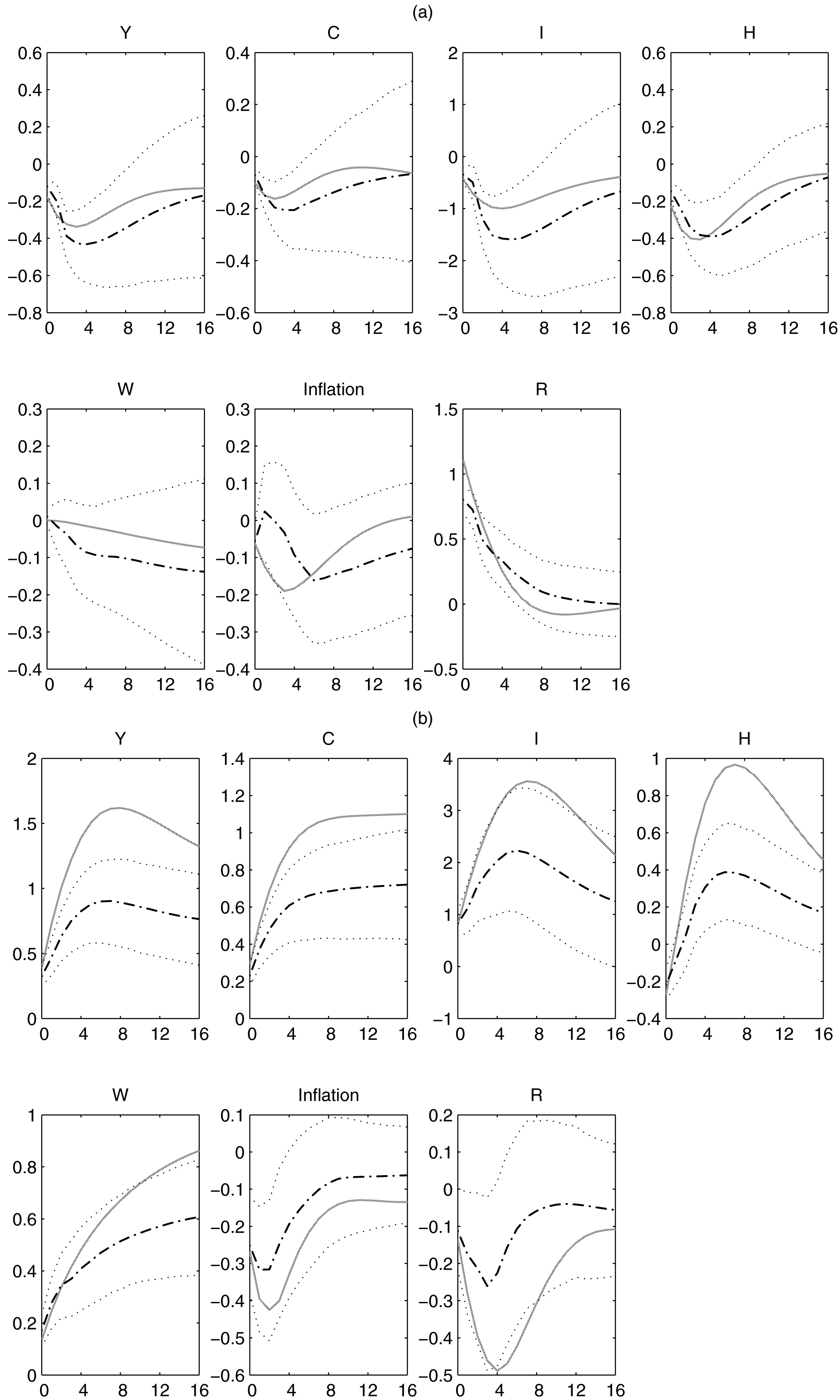

Figure 3. Impulse Response Functions: DSGE–VAR($\hat{\lambda}$) versus DSGE–VAR(∞). (a) Monetary policy shocks. (b) Technology shocks. This figure depicts posterior mean responses for the DSGE–VAR(∞) (gray solid lines) and the DSGE–VAR($\hat{\lambda}$) (dark dashed–dotted lines), and 90% bands (dark dotted lines) for DSGE–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–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–VAR(∞) and DSGE–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–VAR(∞) and DSGE–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

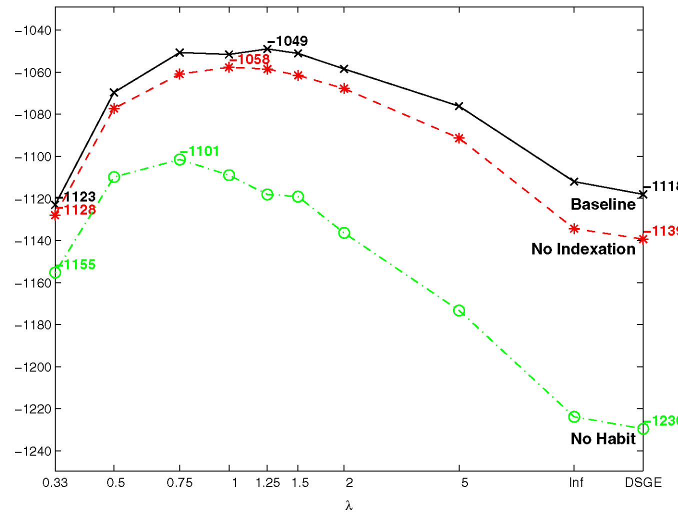

Figure 4. Marginal Likelihood as a Function of λ: Comparison Across Models. See Figure 2 for an explanation.

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–VAR(∞) and DSGE–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–VAR(∞) (gray solid line) and DSGE–VAR($\hat{\lambda}$) (dark dash-and-dotted lines), as well a the 90% bands (dark dotted lines) for DSGE–VAR($\hat{\lambda}$). Figure 5 is obtained based on the *no habit* model; therefore, the benchmark DSGE–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–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

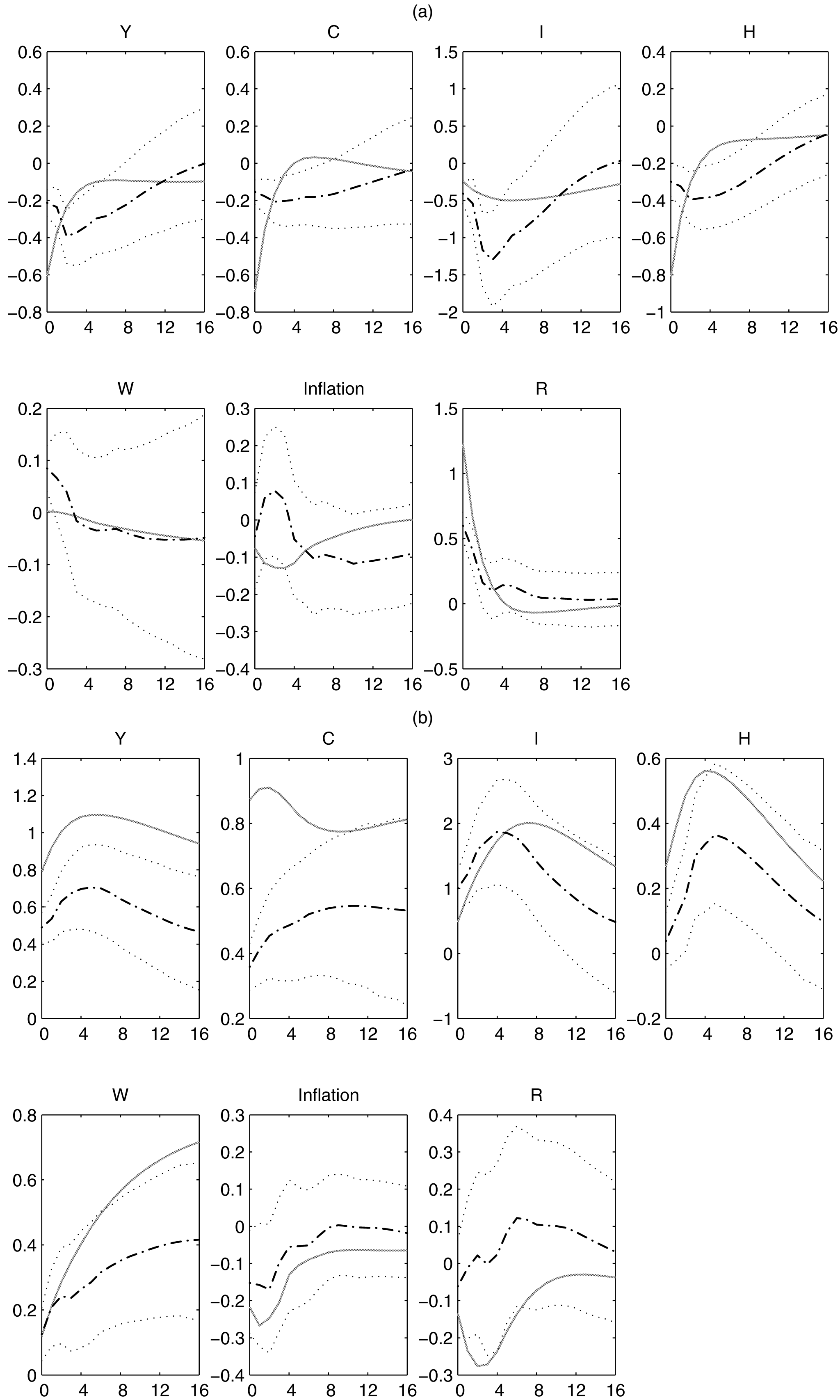

Figure 5. Impulse Response Functions for the no Habit Model: DSGE–VAR($\hat{\lambda}$) versus DSGE–VAR(∞). (a) Monetary policy shocks. (b) Technology shocks. See Figure 3 for an explanation.

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–VAR(∞), compared with the more gradual response in the DSGE–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–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–VAR(∞) and the benchmark, DSGE–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–VAR(∞) and compare it with that of the DSGE–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–VAR($\hat{\lambda}$) and DSGE–VAR(∞) 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–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–VAR(∞) 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–VAR($\hat{\lambda}$) is roughly as accurate as the unrestricted VAR in terms of hours per capita, whereas DSGE–VAR(∞) is far worse, especially in the long run. Conversely, DSGE–VAR(∞) performs well in terms of the nominal variables, inflation, and interest rate. For inflation, the forecasting accuracy of DSGE–VAR(∞) is inferior to that of DSGE–VAR($\hat{\lambda}$), but far better than that of the unrestricted VAR. For the nominal interest rate, DSGE–VAR(∞) outperforms DSGE–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

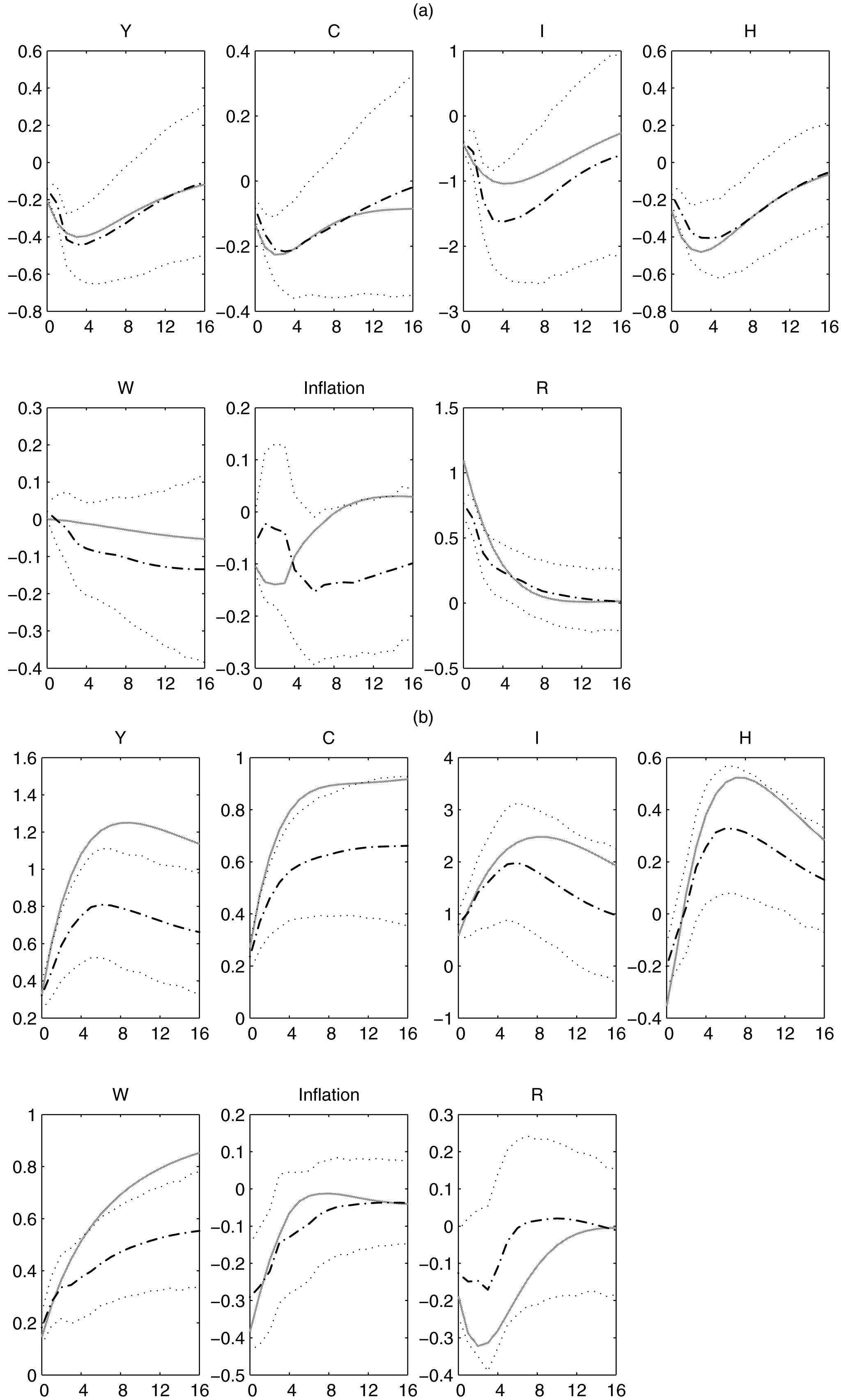

Figure 6. Impulse Response Functions for the no Indexation Model: DSGE–VAR($\hat{\lambda}$) versus DSGE–VAR(∞). (a) Monetary policy shocks. (b) Technology shocks. See Figure 3 for an explanation.

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.

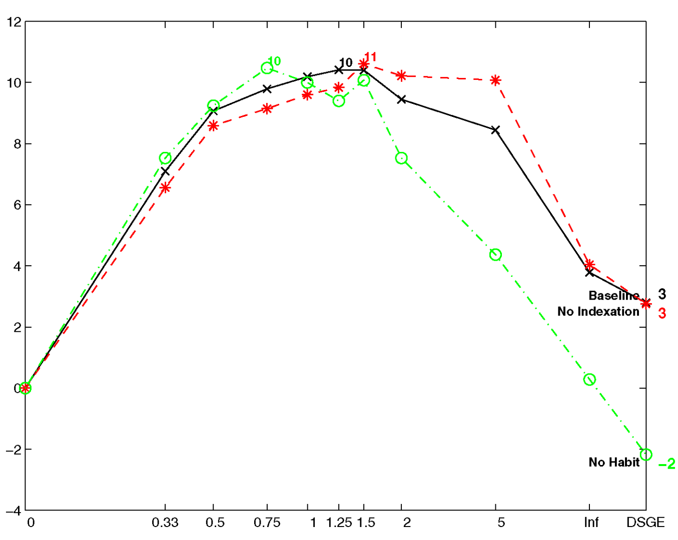

Figure 7. One-Period-Ahead RMSE Summary: Model Comparison. This figure depicts the improvement in the one-period-ahead multivariate statistic relative to an unrestricted VAR as a function of λ for three different models, the baseline model (solid line), the no indexation model (dashed line), and the 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

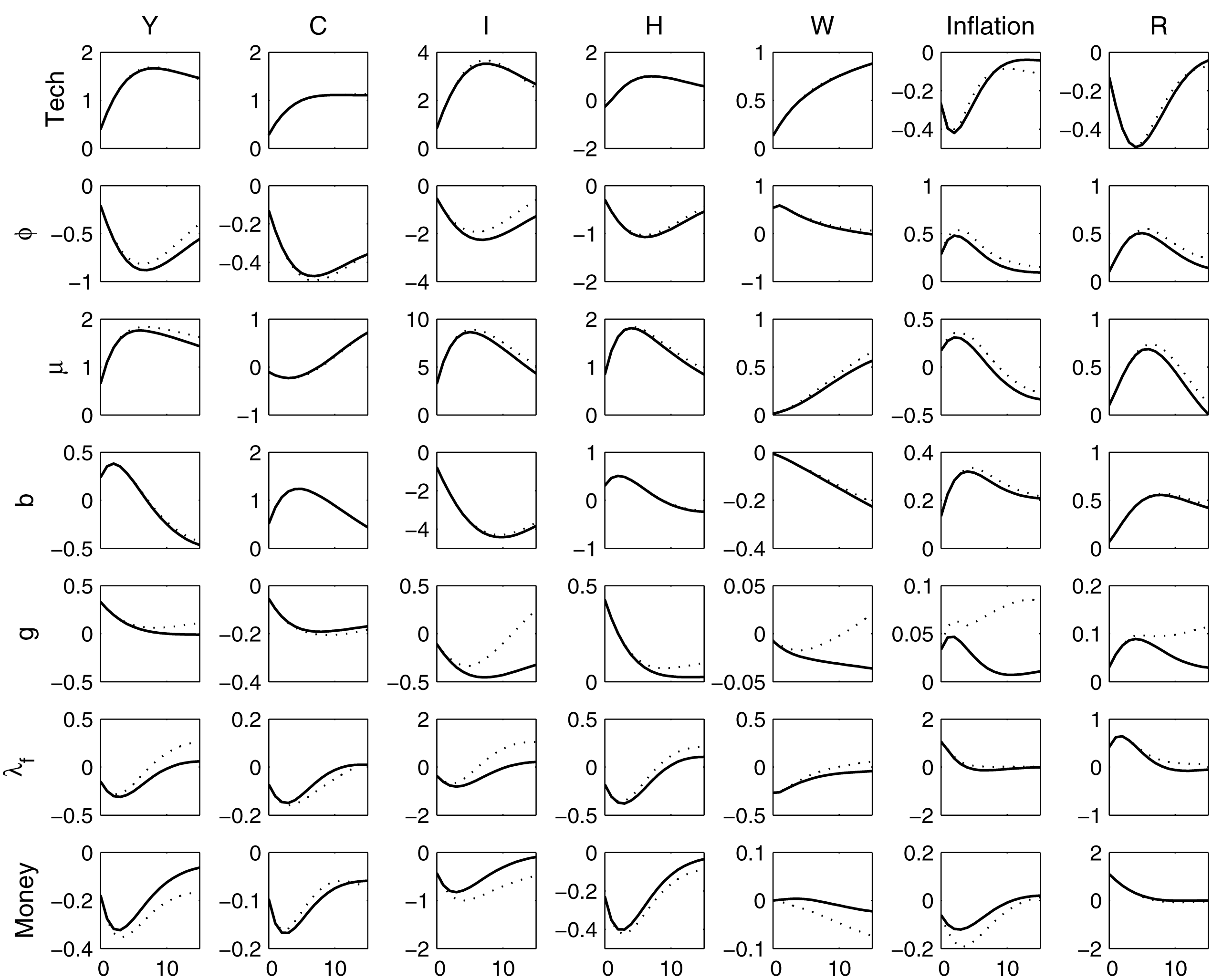

Figure A.1. Baseline Model Impulse Response Functions: DSGE Model versus DSGE–VAR(∞). This figure depicts the impulse responses of the endogenous variables to one standard deviation shocks for the DSGE–VAR(∞) (dotted lines) and for the state-space representation of the DSGE model (solid lines).

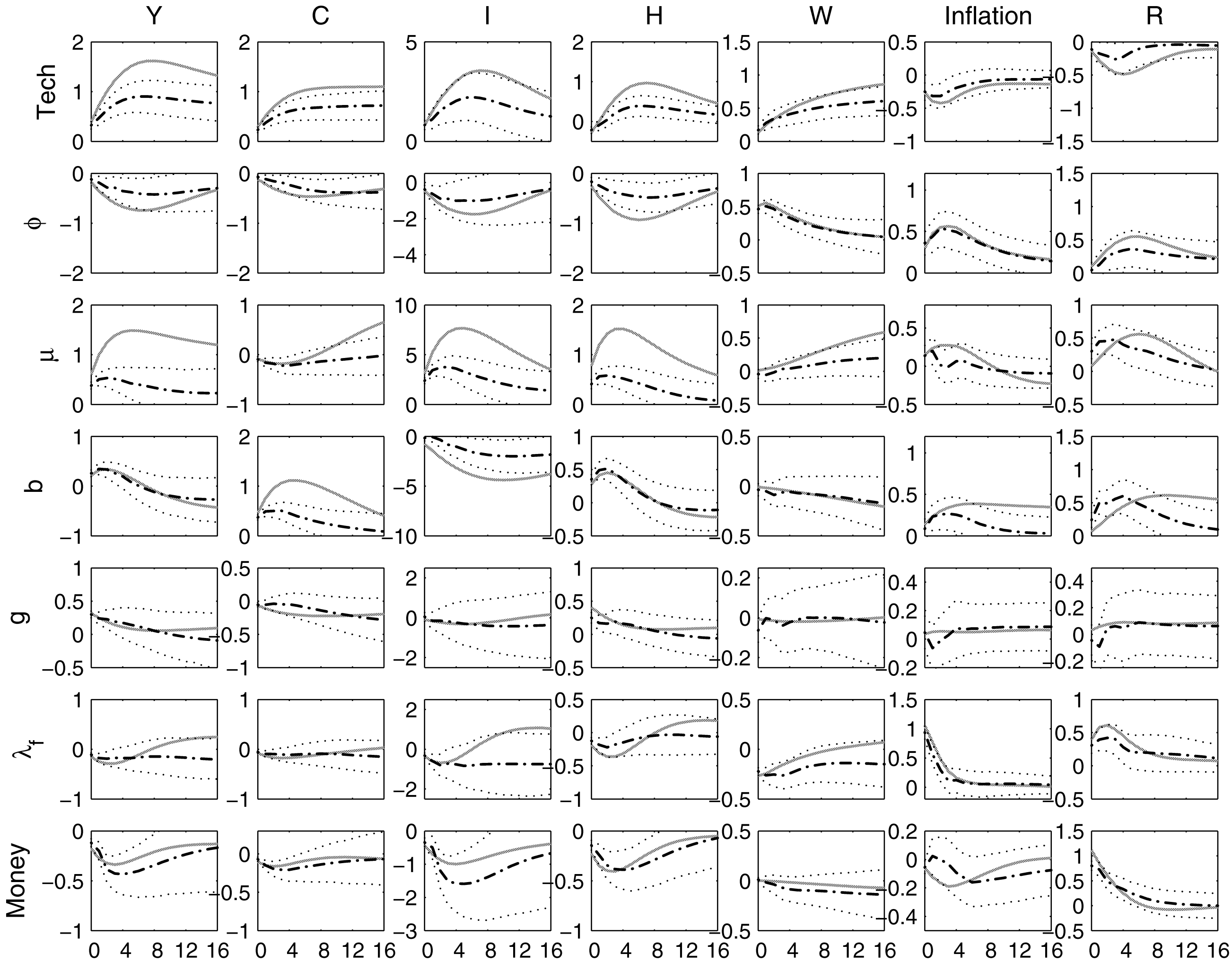

Figure A.2. Baseline Model Impulse Response Functions: DSGE–VAR($\hat{\lambda}$) versus DSGE–VAR(∞). This figure depicts mean responses of the endogenous variables to one standard deviation shocks for the DSGE–VAR(∞) (gray solid lines), the DSGE–VAR($\hat{\lambda}$) (dark dash-and-dotted lines), and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

Journal of Business & Economic Statistics
April 2007, Vol. 25, No. 2
DOI 10.1198/073500107000000061