On the Fit and Forecasting Performance of New-Keynesian Models

Marco Del Negro, Frank Schorfheide, Frank Smets, and Raf Wouters*

November, 2004

Abstract

The paper provides new tools for the evaluation of DSGE models, and applies it to a large-scale New Keynesian dynamic stochastic general equilibrium (DSGE) model with price and wage stickiness and capital accumulation. Specifically, we approximate the DSGE model by a vector autoregression (VAR), and then systematically relax the implied cross-equation restrictions. Let λ denote the extent to which the restrictions are being relaxed. We document how the in- and out-of-sample fit of the resulting specification (DSGE-VAR) changes as a function of λ. Furthermore, we learn about the precise nature of the misspecification by comparing the DSGE model's impulse responses to structural shocks with those of the best-fitting DSGE-VAR. We find that the degree of misspecification in large-scale DSGE models is no longer so large to prevent their use in day-to-day policy analysis, yet it is not small enough that it cannot be ignored. (JEL C11, C32, C53)

KEY WORDS: Bayesian Analysis, DSGE Models, Model Evaluation, Vector Autoregressions

---

*Marco Del Negro: Federal Reserve Bank of Atlanta, e-mail: *Marco.DelNegro@atl.frb.org*; Frank Schorfheide: University of Pennsylvania, Department of Economics, e-mail: *schorf@ssc.upenn.edu*; Frank Smets: European Central Bank and CEPR, e-mail: *Frank.Smets@ecb.int*; Raf Wouters: National Bank of Belgium, email: *Rafael.Wouters@nbb.be*. We thank seminar participants at the Atlanta Fed, New York University, Northwestern University, the Richmond Fed, Stanford University, the University of Virginia, Yale University, the workshop on "Empirical Methods and Applications to DSGE Models" at the Cleveland Fed, the 2^{nd} Euro Area Business Cycle Network (ECB, Fall 2003), 2004 SED, and 2004 SCE conferences for useful comments. The views expressed in this papers are solely our own 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.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are not just attractive from a theoretical perspective, but they are also 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, for instance, 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 Kydland and Prescott (1982). Subsequently, many authors have developed econometric frameworks that formalize aspects of the calibration approach, for instance, Canova (1994), DeJong, Ingram, and Whiteman (1996), Geweke (1999), Schorfheide (2000), and Dridi, Guay, Renault (2004). A common feature of many of these approaches is that DSGE model predictions are either implicitly or explicitly compared to those from a reference model. Much of the applied work related to monetary models has, for instance, proceeded by evaluating, and to some extend also estimating, 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). Examples include Nason and Cogley (1994), Rotemberg and Woodford (1997), Boivin and Giannoni (2003), and Christiano, Eichenbaum, and Evans (2004). As pointed out in Schorfheide (2000) such an evaluation is sensible as long as the VAR indeed dominates the DSGE model in terms of time series fit.

Smets and Wouters (2003a) lay out a large-scale monetary DSGE model in the New Keynesian tradition based on work by Christiano, Eichenbaum, and Evans (2004) and fit their DSGE model to Euro-area data. One of the remarkable empirical results is that the DSGE model outperforms vector autoregressions estimated with a fairly diffuse training sample prior in terms of its marginal likelihood. Loosely speaking, the log marginal likelihood can be interpreted as a measure of a one-step-ahead predictive score (Good, 1952). Previous studies using more stylized DSGE models, e.g., Schorfheide (2000), always found that even simple VARs dominate DSGE models. On the one hand, the Smets and Wouters (2003a) finding challenges the practice of assessing DSGE models on their ability to reproduce VAR impulse response functions without carefully documenting that the VAR indeed fits better than the DSGE model. On the other hand, it poses the question whether researchers from now on have to be less concerned about misspecification of DSGE models.

Moreover, the result suggests that it is worthwhile to carefully document the out-of-sample predictive performance of New-Keynesian DSGE models.

The contributions of our paper are twofold, one methodological and the other substantive. First, we develop a set of tools that is useful to assess the time series fit of a DSGE model. We construct a benchmark model that can assist to characterize and understand the degree of misspecification of the DSGE model. Second, we apply these tools to a variant of the Smets and Wouters (2003a) model and document its fit and forecasting performance based on post-war U.S. data.

Our approach to model evaluation is based on Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004). Both papers develop methods to tilt the coefficient estimates of a VAR toward the restrictions implied by a DSGE model in order to improve the time series fit of the estimated VAR. While the focus of this earlier work was to improve the empirical performance of a VAR, this paper emphasizes a different aspect. We approximate the state-space representation of a log-linear DSGE model by a vector autoregression with tight cross-coefficient restrictions. These restrictions are potentially misspecified and model fit can be improved by relaxing the restrictions. The weight that we place on the DSGE model restrictions is controlled by a hyperparameter λ. We refer to the resulting model as DSGE-VAR(λ).

Formally, we are using a Bayesian framework in which λ scales the inverse of a prior covariance matrix for parameters that capture deviations from the DSGE model restrictions. The posterior distribution of λ provides an overall assessment of the DSGE model restrictions. Posterior mass concentrated on large values of λ provides evidence in support of the DSGE model restrictions. The practice of assessing DSGE models based on their posterior odds relative to a VAR with diffuse prior can be viewed as a special case in which λ is restricted to be either ∞ or close to zero. Such a posterior odds comparison between the extremes, however, tends to be sensitive to the specification of the diffuse prior on the VAR. Sims (2003) noted that the posterior probabilities computed by Smets and Wouters do not give an accurate reflection of model uncertainty as they tend to switch between the extremes zero and one, depending on the choice of data set (Euro-area data in 2003a and U.S. data in 2003b) and the specification of the VAR prior (Minnesota prior versus training-sample prior). By considering an entire range of hyperparameter values between the extremes we are allowing for varying degrees of deviations from the DSGE model restrictions and our assessment misspecification becomes more refined and robust.

Second, in addition to studying the posterior distribution of λ we are computing a sequence of pseudo-out-of-sample forecasts for the state-space representation of the DSGE model, the DSGE-VAR with λ replaced by the hyperparameter value $\hat{\lambda}$ that has the highest posterior probability, and a VAR with a very diffuse prior. The resulting root-mean-squared forecast errors provide additional evidence on the fit of the DSGE model and how it changes as the model restrictions are being relaxed.

Third, if the posterior distribution of the hyperparameter suggests to relax the DSGE model restrictions, then the DSGE-VAR($\hat{\lambda}$) can be used as a benchmark for evaluating the dynamics of the DSGE model and to gain some insights on how to improve the structural model. Note that unlike a comparison of the DSGE model to a VAR estimated with simple least squares methods, our analysis guarantees that the DSGE model is not compared to a specification that fits worse, where fit is measured by the marginal likelihood.[1] We provide an identification scheme where the rotation matrix is such that in absence of misspecification the DSGE's and the DSGE-VAR's impulse responses to all shocks would coincide. To the extent that misspecification is mainly in the dynamics, as opposed to the covariance matrix of innovations, this identification implicitly matches the short-run responses of the DSGE-VAR to those of the underlying DSGE model. Hence, in constructing a benchmark for the evaluation of the DSGE model we are trying to stay as close to the original specification as possible without having to sacrifice time series fit.

The empirical findings are as follows. We document that the state space representation of the DSGE model is well approximated by a VAR with four lags in output growth, consumption growth, investment growth, real wage growth, hours worked, inflation, and nominal interest rates, provided the model-implied cointegration vectors are included as additional regressors. We refer to this specification as DSGE-VECM since the cointegration vectors are often called error correction terms in the time series literature. A preliminary estimation of the state space representation of the DSGE model confirms the well-known result that the exogenous driving processes of the model are highly persistent, pick up most of the serial correlation in the observed time series, and also have to offset some of the counterfactual co-trending implications of the DSGE model.

The posterior distribution of the hyperparameter λ has an inverse U-shape indicating that the fit of the autoregressive 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 as VARs with

[1] There is a long tradition in the forecasting literature to boost the predictive performance of VARs through the use of prior distributions dating back to Doan, Litterman, and Sims (1984).

very diffuse priors are clearly dominated by the DSGE-VECM($\hat{\lambda}$). 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 VECM with diffuse prior perform about equally well in terms of one-step ahead forecasts, but are clearly worse than the DSGE-VECM($\hat{\lambda}$). The forecast accuracy improvements obtained by optimally relaxing the DSGE model restrictions are largest in the medium run. We also document the forecast accuracy for individual series. While for most variables the forecasts improve as the restrictions are loosened there are two exceptions: real wage and inflation forecasts hardly improve.

When comparing impulse responses between the DSGE model and the DSGE-VECM($\hat{\lambda}$) we find that many responses are not only qualitatively, but also quantitatively in agreement. However, there are exceptions. For instance, the effects of the shock to the marginal rate of substitution between consumption and leisure are more persistent in the DSGE-VECM($\hat{\lambda}$) than in the DSGE model. According to the DSGE model, output and hours increase immediately in response to a government spending shock and quickly decay monotonically. The DSGE-VECM, on the other hand, predicts delayed, hump-shaped responses of both output and hours that are long-lasting. Moreover, the DSGE-VECM predicts a larger real effect of monetary policy shocks.

The paper is organized as follows. The DSGE model is presented in Section 2. Section 3 discusses how the state space representation of the DSGE model is approximated by a vector autoregressive specification and how a prior distribution for the DSGE model misspecification is generated. Moreover, a simple example is provided to illustrate the role of λ in assessing the overall fit of the DSGE model and in constructing a benchmark for forecast and impulse-response comparisons. Section 4 describes the data. Empirical results are presented in Section 5 and Section 6 concludes.

2 Model

This section describes the DSGE model, which is a slightly modified version of the DSGE model developed and estimated for the Euro area in Smets and Wouters (2003a). In particular, we introduce stochastic trends into the model, so that it can be fitted to unfiltered time series observations. The DSGE model is based on work of Christiano, Eichenbaum, and Evans (2004) and contains a large number of nominal and real frictions. To make this paper 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 $i \in [0,1]$:

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

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 mark-up over marginal costs. The final goods producers are perfectly competitive firms that buy intermediate goods, combine them to the final product 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:

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

2.2 Intermediate goods producers

Good i is made using the technology:

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

where the technology shock Z_t (common across all firms) follows a unit root process, and where $\mathcal{F}$ represent fixed costs faced by the firm. We define technology growth $z_t = \log(Z_t / Z_{t-1})$ and assume that z_t follows the autoregressive process:[2]

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

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

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

[2]Smets and Wouters (2003a) assume a stationary technology shock that follows an autoregressive process. Their estimate of the autocorrelation coefficient however are very close to the upper boundary of one. We therefore choose to assume a unit root process from the onset.

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 re-optimize 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 $\pi_t = P_t/P_{t-1}$ and π^* is the steady state inflation rate of the final good. In our empirical analysis we will restrict ι_p to be either zero or one. Those firms that are able to re-optimize prices choose the price level $\tilde{P}_t(i)$ that solves:

$$\begin{aligned}
&\max_{\tilde{P}_t(i)} I\!E_t \left[\sum_{s=0}^{\infty} \zeta_p^s \beta^s \Xi_{t+s}^p \left(\tilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p} \right) - MC_{t+s} \right) Y_{t+s}(i) \right] \\
&\text{s.t. } Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p} \right)}{P_{t+s}} \right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_{t+s}, \; MC_{t+s} = \frac{\alpha^{-\alpha} W_{t+s}^{1-\alpha} R_{t+s}^{k\ \alpha}}{(1-\alpha)^{(1-\alpha)} Z_{t+s}^{1-\alpha}},
\end{aligned} \tag{8}$$

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 where all firms will choose the same $\tilde{P}_t(i)$. Hence from (3) we obtain the following law of motion for the aggregate price level:

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

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:

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

where $\lambda_w \in (0,\infty)$ is a fixed parameter.[3] 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:

$$(a) \quad L_t(j) = \left(\frac{W_t(j)}{W_t} \right)^{-\frac{1+\lambda_w}{\lambda_w}} L_t \quad \text{and} \quad (b) \quad W_t = \left[\int_0^1 W_t(j)^{\frac{1}{\lambda_w}} di \right]^{\lambda_w}. \tag{11}$$

[3]Smets and Wouters (2003a) assume that i.i.d. shocks to the degree of labor substitutability are another source of disturbance in the economy.

2.4 Households

The objective function for household j is given by:

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

where $C_t(j)$ is consumption, $L_t(j)$ is labor supply, and $M_t(j)$ is money holdings. Household's preferences display habit-persistence. We depart from Smets and Wouters (2003b) in assuming separability in the utility function for a reason that will be discussed later. 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:

$$\ln \varphi_t = (1 - \rho_\varphi) \ln \varphi + \rho_\varphi \ln \varphi_{t-1} + \sigma_\varphi \epsilon_{\varphi,t}, \tag{13}$$

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

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

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

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

where $I_t(j)$ is investment, $B_t(j)$ is holdings of government bonds, 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 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 parenthesis 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 $R^k_t u_t(j)\bar{K}_{t-1}(j)$ in return. They however have to 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 $S'(\cdot) > 0$, $S''(\cdot) > 0$. 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:[4]

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

The households' wage setting is subject to nominal rigidities á la Calvo (1983). In each period a fraction ζ_w of households is unable to re-adjust 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 growth rate of the economy e^γ) and of last period's inflation times last period's productivity ($\pi_{t-1} e^{z_{t-1}}$). The weights are $1 - \iota_w$ and ι_w, respectively. Those households that are able to re-optimize their wage solve the problem:

$$\begin{array}{ll} \max_{\tilde{W}_t(j)} & I\!\!E_t \sum_{s=0}^{\infty} (\zeta_w \beta)^s b_{t+s} \left[-\dfrac{\varphi_{t+s}}{\nu_l + 1} L_{t+s}(j)^{\nu_l + 1} \right] \\ \text{s.t.} & (15) \text{ for } s = 0, \ldots, \infty, \text{ (11a), and} \\ & W_{t+s}(j) = \left(\Pi_{l=1}^{s} (\pi_* e^\gamma)^{1-\iota_w} (\pi_{t+l-1} e^{z_{t+l-1}})^{\iota_w} \right) \tilde{W}_t(j). \end{array} \tag{19}$$

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:

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

Finally, we assume 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. Since the amount of leisure will differ across households due to the wage rigidity, separability between labor and consumption in the utility function is key for this result.

[4] We have also experimented with the introduction of a deterministic trend Υ^t in equation (17), as in Greenwood, Hercovitz, and Krusell (1998). Since this added parameter does change the results or improve the fit for our empirical specification, we set it equal to 1.

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:

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

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 the one in Smets and Wouters (2003a,b), who introduce a time-varying inflation objective that varies stochastically according to a random walk. The random walk inflation target may help the model to fit the medium- and long-frequency fluctuations in inflation. In this paper, we are interested in assessing the model's fit of inflation without the extra help coming from the exogenous inflation target shocks. We consider two alternative specifications for the target level of output Y_t^* in (21). Under one specification the monetary authorities target the trend level of output: $Y_t^* = Y_t^s$. Under the alternative specification they target the level of output that would have prevailed in absence of nominal rigidities: $Y_t^* = Y_t^f$. The central bank supplies the money demanded by the household to support the desired nominal interest rate.

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 nominal lump-sum taxes (or subsidies) that also appear in household's budget constraint. Government spending is given by:

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

where g_t follows the 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 Altig, Christiano, Eichenbaum, and Linde (2002) 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 K_t and effective capital $\bar{K}_t$ all grow at the rate Z_t. Nominal interest rates, inflation, and hours worked 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 in Sims (2002) to construct a log-linear approximation of the model around the steady state. We collect all 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 and π_t is the inflation rate.

3 DSGE-VARs as Tools for Model Evaluation

The DSGE model generates a covariance-stationary distribution of the sequence $\{\Delta y_t\}$. We will now derive an (approximate) vector autoregressive representation for the DSGE model. As is well known, this representation imposes many cross-equation restrictions on the VAR parameters. We explicitly introduce model misspecification as deviations of the VAR parameters from these restrictions. We construct a joint prior distribution for the parameters of the DSGE model and the parameters that characterize the model misspecification. The prior for the misspecification parameters is centered at zero: the DSGE restrictions are the benchmark.

The prior covariance matrix of the misspecification parameters is scaled by a hyperparameter λ, which can be interpreted as the weight that we are placing on the DSGE model restrictions. Whenever λ is very high, the resulting model will be fairly close to the DSGE model itself as the prior on the misspecification parameters concentrates the mass around zero. Whenever λ is small the resulting model will be fairly close to an unrestricted VAR: The prior on the misspecification parameters is virtually flat. Hence, we have a family of models indexed by λ that essentially has an unrestricted VAR at one extreme and the DSGE model at the other extreme. We will call these models DSGE-VAR(λ). Here by model we mean a joint probability distribution for the data and the parameters. Using Markov-Chain-Monte-Carlo methods we can generate draws from the posterior distribution of the DSGE model parameters and the misspecification parameters. We are also able to make posterior inference with respect to the hyperparameter λ.

Our framework is useful for DSGE model evaluation in two respects. First, the posterior distribution of the hyperparameter λ provides an overall summary of fit. A substantial mass on large values of λ can be interpreted as evidence in favor of the restrictions imposed by the DSGE model. If there are misspecifications, they are likely to be small. Posterior estimates of λ near zero, on the other hand, indicate serious misspecification.

Second, the estimated DSGE-VAR provides a natural benchmark for comparing the dynamics of the DSGE model to those of a less restrictive specification. There is an extensive literature that evaluates DSGE models by comparing their impulse responses to those obtained from vector autoregressions, to name a few, Cogley and Nason (1994), Rotemberg and Woodford (1997), Schorfheide (2000), and, more recently, Boivin and Giannoni (2003) and Christiano, Eichenbaum, and Evans (2004). An important issue in such comparisons is the estimation and identification of the VAR that serves as a benchmark. Most authors use simple least squares techniques to estimate the VAR, which, unfortunately, leads to very noisy coefficient estimates. Moreover, it is often difficult to find identification schemes that are consistent with the DSGE model that is being estimated and that identify several structural shocks simultaneously.

Our benchmark for comparing impulse-responses is DSGE-VAR($\hat{\lambda}$), where $\hat{\lambda}$ maximizes the posterior density of λ, and, loosely speaking, selects the DSGE-VAR that yields the best fit according to one-step-ahead pseudo-out-of-sample forecasting performance. Thus, our benchmark links the magnitude of the deviation from the DSGE model restrictions to the degree of their misspecification, that is, it deviates from the restrictions only to the extent that the deviations improves fit. In our empirical analysis, we document that DSGE-VAR($\hat{\lambda}$) indeed yields more precise out-of-sample forecasts than an unrestricted VAR and therefore arguably represents a better benchmark. Importantly, DSGE-VAR is identified. We propose an identification scheme for the structural shocks that is implementable in high-dimensional systems and tries to keep the DSGE-VAR and DSGE model impulse responses similar. The shapes and magnitudes of the remaining discrepancies can provide valuable information about dynamic misspecifications of the DSGE model and how to overcome them by refining the structural model.

The econometric analysis in this paper is closely related to earlier work by Del Negro and Schorfheide (2004), who proposed a Bayesian procedure that tilts the VAR coefficient estimates toward the restrictions implied by the DSGE model. Loosely speaking, the procedure amounts to adding artificial observations generated from the DSGE model to the actual observations and then estimating the VAR based on this augmented data set. From a Bayesian perspective the artificial observations generate a prior distribution for the VAR coefficients that is centered around the DSGE model restrictions. Del Negro and Schorfheide (2004) focused on the question: how can one improve a VAR by using information from a DSGE model? The present paper asks the opposite question: How can one relax the restrictions of the DSGE model and evaluate the extent of their misspecification? Since these two questions can be viewed as opposite sides of the same coin, the priors and posteriors presented subsequently are almost identical to the ones used in Del Negro and Schorfheide (2004), but we present a different derivation and interpretation.

3.1 VAR and VECM Representations of the DSGE Model

The DSGE model generates a restricted and potentially misspecified moving average (MA) representation for the vector Δy_t. We approximate the MA representation by a VAR with p-lags:

$$\Delta y_t = \Phi_0^*(\theta) + \Phi_1^*(\theta)\Delta y_{t-1} + \ldots + \Phi_p^*(\theta)\Delta y_{t-p} + u_t. \tag{26}$$

We will assume that the vector of reduced-form innovations u_t is normally distributed conditional on past information with mean zero and covariance matrix $\Sigma_u^*(\theta)$. We denote the dimension of Δy_t by n and define the $k \times 1$ vector $x_t = [1, \Delta y'_{t-1}, \ldots, \Delta y'_{t-p}]$. Let $\Phi^*(\theta) = [\Phi_0^*(\theta), \ldots, \Phi_p^*(\theta)]'$. Then the VAR can be rewritten as

$$\Delta y'_t = x'_t \Phi^*(\theta) + u'_t. \tag{27}$$

In general, the VAR representation (26) is not exact if the number of lags p is finite. We define $\Gamma_{XX}(\theta) = I\!\!E_\theta^D[x_t x'_t]$ and $\Gamma_{XY}(\theta) = I\!\!E_\theta^D[x_t y'_t]$ and let

$$\Phi^*(\theta) = \Gamma_{XX}(\theta)^{-1}\Gamma_{XY}(\theta). \tag{28}$$

Here $I\!\!E_\theta^D[\cdot]$ refers to an expectation with respect to the distribution generated by the DSGE model. For $\Gamma_{XX}(\theta)$ to be well-defined it is important that x_t is stationary according to the DSGE model and that its covariance matrix is non-singular. Both conditions are satisfied for the model specified in Section 2. The model implied covariance matrix of u_t is defined as

$$\Sigma_u^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta), \tag{29}$$

where Γ_{YY} and Γ_{YX} are defined in the same way as Γ_{XX}.

The DSGE model presented in Section 2 implies that the set of variables that we consider for our empirical analysis has several common trends. For instance, output, consumption, and investment all grow that the rate Z_t. This suggests that we can obtain a better approximation of the DSGE model if we generate a moving-average representation from the following vector error correction (VECM) specification

$$\Delta y_t = \Phi_0^*(\theta) + \Phi_\beta^*(\theta)(\beta' y_{t-1}) + \Phi_1^*(\theta)\Delta y_{t-1} + \ldots + \Phi_p^*(\theta)\Delta y_{t-p} + u_t, \tag{30}$$

where the (stationary) error correction terms are defined as

$$\beta' y_{t-1} = \left[\begin{array}{l} \ln C_t - \ln Y_t \\ \ln I_t - \ln Y_t \\ \ln(W_t/P_t) - \ln Y_t \end{array} \right]. \tag{31}$$

We will refer to this specification as DSGE-VECM. We can easily encompass the DSGE-VECM in the notation developed above by redefining $x_t = [1, \beta' y_{t-1}, \Delta y'_{t-1}, \ldots, \Delta y'_{t-p}]'$ and $\Phi^*(\theta) = [\Phi_0^*(\theta), \Phi_\beta^*(\theta), \ldots, \Phi_p^*(\theta)]'$. For ease of exposition we will subsequently ignore the error made by approximating the state space representation of the log-linearized DSGE model with a finite-order vector autoregressive specification or, in other words, treat (26) or (30), respectively, as the structural model that imposes potentially misspecified restrictions on the matrices Φ and Σ_u. We will document the magnitude of the approximation error at the end of this section.

3.2 Misspecification and Bayesian Inference

We make the following assumptions about misspecification of the DSGE model. There is a vector θ and matrices Ψ^Δ and Σ_u^Δ such that the data are generated from the vector autoregressive model

$$y'_t = x'_t \Phi + u'_t \tag{32}$$

where

$$\Phi = \Phi^*(\theta) + \Phi^\Delta, \quad \Sigma_u = \Sigma_u^*(\theta) + \Sigma_u^\Delta. \tag{33}$$

and there does not exist a $\tilde{\theta} \in \Theta$ such that $\Phi = \Phi^*(\tilde{\theta})$ and $\Sigma_u = \Sigma_u^*(\tilde{\theta})$.

Our econometric analysis is casted in a Bayesian framework in which initial beliefs about the DSGE model parameter θ and the model misspecification matrices Ψ^Δ and Σ_u^Δ are summarized in a prior distribution. One can interpret the prior as describing how nature draws the misspecification. We will now motivate this prior distribution with a thought experiment. In this experiment, we assume that $\Sigma_u^\Delta = 0$ and condition on the DSGE model parameter vector θ.

We assume that the prior assigns low density to large values of the misspecification parameter Ψ^Δ. That is, we assume that nature is more likely to draw small than large misspecification matrices. This assumption reflects the belief that the DSGE model provides a good albeit not perfect approximation of reality. We measure the size of the misspecification Ψ^Δ by the ease with which it can be detected using likelihood ratios. Suppose that a sample of λT observations is generated from (26), where Φ is given by (33) and T is the size of the actual sample used in the estimation. We will construct a prior that has the property that its density is proportional to the expected likelihood ratio of Φ evaluated at its (misspecified) restricted value $\Phi^*(\theta)$ versus the true value $\Phi = \Phi^*(\theta) + \Phi^\Delta$. The log-likelihood ratio is

$$\begin{aligned} & \ln\left[\frac{\mathcal{L}(\Phi^* + \Phi^\Delta, \Sigma_u^*, \theta|Y,X)}{\mathcal{L}(\Phi^*, \Sigma_u^*, \theta|Y,X)}\right] \\ = & -\frac{1}{2}tr\Bigg[\Sigma_u^{*-1}\Bigg((\Phi^* + \Phi^\Delta)'X'X(\Phi^* + \Phi^\Delta) - 2(\Phi^* + \Phi^\Delta)'X'Y \\ & -\Phi^{*'}X'X\Phi^* + 2\Phi^{*'}X'Y\Bigg)\Bigg]. \end{aligned} \tag{34}$$

Y denotes the $\lambda T \times n$ matrix with rows y_t' and X_t is the $\lambda T \times k$ matrix with rows x_t'. After replacing Y by $X\Phi^* + U$ the log likelihood ratio simplifies to

$$\ln\left[\frac{\mathcal{L}(\Phi^* + \Phi^\Delta, \Sigma_u^*, \theta|Y,X)}{\mathcal{L}(\Phi^*, \Sigma_u^*, \theta|Y,X)}\right] = -\frac{1}{2}tr\left[\Sigma_u^{*-1}\left(\Phi^{\Delta'}X'X\Phi^\Delta - 2\Phi^{\Delta'}X'U\right)\right]. \tag{35}$$

Taking expectations over X and U using the distribution induced by the data generating process yields (minus) the Kullback-Leibler distance between the data generating process and the DSGE model:

$$I\!E^{VAR}_{\Phi^*,\Sigma_u^*}\left[\ln\left[\frac{\mathcal{L}(\Phi^* + \Phi^\Delta, \Sigma_u^*, \theta|Y,X)}{\mathcal{L}(\Phi^*, \Sigma_u^*, \theta|Y,X)}\right]\right] = -\frac{1}{2}tr\left[\Sigma_u^{*-1}\left(\Phi^{\Delta'}\lambda T\Gamma_{XX}\Phi^\Delta\right)\right], \tag{36}$$

where $I\!E^{VAR}_{\Phi^*,\Sigma_u^*}[\cdot]$ denotes the expectation under the probability distribution generated by (26).[5] We now choose a prior density that is proportional ($\propto$) to the Kullback-Leibler discrepancy:

$$p(\Phi^\Delta|\Sigma_u^*, \theta) \propto \exp\left\{-\frac{1}{2}tr\left[\lambda T\Sigma_u^{*-1}\left(\Phi^{\Delta'}\Gamma_{XX}\Phi^\Delta\right)\right]\right\}. \tag{37}$$

As the sample size λT increases the prior places more mass on misspecification matrices that are close to zero.

[5] It is straightforward to verify that $I\!E^{VAR}_{\Psi^*,\Sigma^*}[x_t x_t'] = I\!E^D_\theta[x_t x_t'] = \Gamma_{XX}(\theta)$.

For computational reasons it is convenient to transform this prior into a prior for Φ. Using standard arguments we deduce that this prior is multivariate normal

$$\Phi|\Sigma_u^*, \theta \sim \mathcal{N}\left(\Phi^*(\theta), \frac{1}{\lambda T}\left[\Sigma_u^{*-1} \otimes \Gamma_{XX}(\theta)\right]^{-1}\right). \tag{38}$$

The hyperparameter λ, which determines the length of the hypothetical sample as a multiple of the actual sample size T, scales the variance of the distribution that generates Φ^Δ and Φ. If λ is close to zero, the prior variance of the discrepancy Φ^Δ is large. Large values of λ, on the other hand, correspond to small model misspecification and for $\lambda = \infty$ beliefs about model misspecification degenerate to a point mass at zero.

In practice we also have to take potential misspecification of the covariance matrix $\Sigma_u^*(\theta)$ into account. Hence, we will use the following, slightly modified, prior distribution conditional on θ in the empirical analysis:

$$\Sigma_u|\theta \quad \sim \quad \mathcal{IW}\left(\lambda T \Sigma_u^*(\theta), \lambda T - k, n\right) \tag{39}$$

$$\Phi|\Sigma_u, \theta \quad \sim \quad \mathcal{N}\left(\Phi^*(\theta), \frac{1}{\lambda T}\left[\Sigma_u^{-1} \otimes \Gamma_{XX}(\theta)\right]^{-1}\right), \tag{40}$$

where $\mathcal{IW}$ denotes the inverted Wishart distribution. The latter induces a distribution for the discrepancy $\Sigma_u^\Delta = \Sigma_u - \Sigma_u^*$. The prior distribution is proper, i.e., has mass one, provided that $\lambda T \geq k + n$. Hence, we restrict the domain of λ to the interval $[(k+n)/T, \infty]$.

3.3 Posteriors

The posterior density is proportional to the prior density times 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:

$$p_\lambda(\Phi, \Sigma_u, \theta|Y) = p_\lambda(\Phi, \Sigma_u|Y, \theta) p_\lambda(\theta|Y). \tag{41}$$

The λ-subscript indicates the dependence of the posterior on the hyperparameter. 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:

$$\Sigma_u|Y, \theta \quad \sim \quad \mathcal{IW}\left((\lambda+1)T\hat{\Sigma}_{u,b}(\theta), (1+\lambda)T - k, n\right) \tag{42}$$

$$\Phi|Y, \Sigma_u, \theta \quad \sim \quad \mathcal{N}\left(\hat{\Phi}_b(\theta), \Sigma_u \otimes (\lambda T \Gamma_{XX}(\theta) + X'X)^{-1}\right), \tag{43}$$

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

$$\begin{aligned}
\hat{\Phi}_b(\theta) &= (\lambda T\Gamma_{XX}(\theta) + X'X)^{-1}(\lambda T\Gamma_{XY} + X'Y) \qquad (44)\\
&= \left(\frac{\lambda}{1+\lambda}\Gamma_{XX}(\theta) + \frac{1}{1+\lambda}\frac{X'X}{T}\right)^{-1}\left(\frac{\lambda}{1+\lambda}\Gamma_{XY} + \frac{1}{1+\lambda}\frac{X'Y}{T}\right)\\
\hat{\Sigma}_{u,b}(\theta) &= \frac{1}{(\lambda+1)T}\bigg[(\lambda T\Gamma_{YY}(\theta) + Y'Y) - (\lambda T\Gamma_{YX}(\theta) + Y'X)\\
&\quad \times(\lambda T\Gamma_{XX}(\theta) + X'X)^{-1}(\lambda T\Gamma_{XY}(\theta) + X'Y)\bigg]. \qquad (45)
\end{aligned}$$

Expressions (44) and (45) show that 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 $\lambda = (n+k)/T$ then the posterior mean is close to the OLS estimate $(X'X)^{-1}X'Y$. The formula for the marginal posterior density of θ and the description of a Markov-Chain-Monte-Carlo algorithm that generates draws from the joint posterior of Φ, Σ_u, and θ are provided in Del Negro and Schorfheide (2004). Note that the joint posterior of Φ, Σ, and θ implicitly defines a posterior distribution for the misspecification matrices Φ^Δ and Σ^Δ.

We will study the fit of the DSGE model by looking at the posterior distribution of the hyperparameter λ. For computational reasons, we only consider a finite set of values $\Lambda = \{l_1, \ldots, l_q\}$, where $l_1 = (n+k)/T$ and $l_q = \infty$. Moreover, we assign equal prior probabilities to the elements of Λ. According to Bayes Theorem, the posterior probabilities for the hyperparameter are proportional to the marginal data density

$$p_\lambda(Y) = \int p(Y|\theta, \Sigma, \Phi)p_\lambda(\theta, \Sigma, \Phi)d(\theta, \Sigma, \Phi). \qquad (46)$$

We denote the posterior mode of λ by

$$\hat{\lambda} = \text{argmax}_{\lambda\in\Lambda}\, p_\lambda(Y). \qquad (47)$$

It is common in the literature, e.g., Smets and Wouters (2002a,b) to use marginal data densities to document the fit of DSGE models relative to VARs. In our framework this corresponds to (approximately) to comparing $p_\lambda(Y)$ for the extreme values of λ, that is, $\lambda = \infty$ (DSGE model) and $\lambda = (k+n)/T$ (VAR with nearly flat prior). We are extending the analysis to intermediate values of λ as the posterior of the hyperparameter reveals information about the degree of the DSGE model misspecification.

Sims (2003) criticized the use of posterior odds between DSGE models and VARs with diffuse prior because they do not provide a realistic characterization of model uncertainty. The latter generate a rather flat marginal data density whereas the former have a data density that concentrates in a small subset of the observation space. Since the probability of observing data for which the marginal data densities of the two types of models are of similar magnitude is very small, the odds tend to decisively favor either the VAR or the DSGE model. Sims (2003) interprets this phenomenon as an indication that the model space is too sparse. Another criticism of posterior odds comparisons between tightly parameterized models such as DSGE models and more densely parameterized models such as VARs is that the odds are very sensitive to the choice of priors, in particular, for the more general model.[6] Our procedure "fills" the model space by considering a large set of intermediate models, indexed by λ, that lie between the VAR with diffuse prior and the DSGE model. Hence we are able to provide a more detailed characterization of model fit. Moreover, for practical purposes we found our procedure to be robust to minor modifications to the DSGE model and the priors.

3.4 The Role of λ in a Simple Example

This section illustrates the role of λ and the interpretation of the hyperparameter using a stylized example. Specifically, we first describe the relationship between the posterior of the hyperparameter λ and the degree of model misspecification. Second, we show that for intermediate values of the "true" misspecification DSGE-VAR($\hat{\lambda}$) will provide more precise parameter and impulse response function estimates (in a mean-squared error sense) as well as more accurate forecasts than either of the two extremes, the unrestricted model or the model where the restriction is dogmatically imposed. This is indeed the case we encounter in practice.

The example we consider is:

$$y_t = \phi y_{t-1} + u_t, \quad u_t \sim iid\mathcal{N}(0,1), \tag{48}$$

where y_t is a scalar, $\phi = \phi^* + \phi^\Delta$. The variance of the one-step ahead forecast errors is known to be one. We assume that according to the DSGE model $\phi^* = 0$ and abstract from the dependence of the DSGE model on an unknown parameter vector θ. In the notation developed previously $x_t = y_{t-1}$, $\Gamma_{XX} = 1$, and the prior is of the form

$$\phi = \phi^\Delta \sim \mathcal{N}\left(0, \frac{1}{\lambda T}\right). \tag{49}$$

[6] A discussion of this this issue can be found in most Bayesian textbooks, often under the heading "Lindley's Paradox," for instance, Robert (1994) and Gelman, Carlin, Stern, and Rubin (1995).

We restrict $\lambda \geq T^{-1}$, which means that when constructing the Kullback-Leibler distance between the DSGE model and the autoregressive data generating process we consider at least one hypothetical observation.

The joint density of $Y = [y_1, \dots, y_T]$ and ϕ conditional on λ and y_0 is (in our notation we are omitting the initial observation from the conditioning set):

$$p(Y, \phi|\lambda) = (2\pi)^{-(1+T)/2}(\lambda T)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2 + \lambda T\phi^2\right]\right\}. \tag{50}$$

Define

$$\hat{\phi}_b(\lambda) = (\lambda T + \sum y_{t-1}^2)^{-1}(\sum y_t y_{t-1}),$$

which is a simplified version of the expression in (44). It can be verified that the posterior distribution of ϕ is of the form:[7]

$$\phi|Y, \lambda \sim \mathcal{N}\left(\hat{\phi}_b(\lambda), \frac{1}{\lambda T + \sum y_{t-1}^2}\right) \tag{51}$$

and the log marginal data density is given by:

$$\ln p(Y|\lambda) = -\frac{T}{2}\ln(2\pi) + \frac{1}{2}\ln \lambda T - \frac{1}{2}\ln\left(\lambda T + \sum y_{t-1}^2\right) - \frac{1}{2}\left[\sum y_t^2 - \frac{(\sum y_t y_{t-1})^2}{\lambda T + \sum y_{t-1}^2}\right]. \tag{52}$$

Unlike in the empirical application where λ is restricted to take values on a finite grid, we now let λ take values in $I\!R^+$ subject to the restriction $\lambda > T^{-1}$ and use an (improper) prior that is uniform over the domain of λ. Hence, (52) can be interpreted as the log posterior density of the hyperparameter. As above, we denote its mode by $\hat{\lambda}$, and the resulting posterior mean estimator for ϕ by $\hat{\phi}_b(\hat{\lambda})$. Note that the maximum likelihood estimator for ϕ in this problem is given by $\hat{\phi}_{mle} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2}$.

The shape of the posterior of λ depends, of course, on the particular realization of Y. In order to provide a characterization of the posterior we assume that the observations have been generated from the following model:

$$y_t = (\phi^* + T^{-1/2}\widetilde{\phi^{\Delta}})y_{t-1} + u_t. \tag{53}$$

According to (53) the misspecification vanishes at rate $T^{-1/2}$ (local misspecification). The trade-off between the squared bias introduced by the potentially misspecified coefficient restriction $\phi = 0$ and the sampling variance due to the estimation of ϕ stays approximately constant, as both bias and variance decay at the same rate T^{-1} (see also Schorfheide (2004)). This setup formalizes the notion that DSGE models provide a fairly good albeit not perfect

[7] Detailed derivations are available from the authors upon request.

approximation of reality.[8] Letting the sample size T tend to infinity we are now able to characterize the posterior mode of λ via its asymptotic sampling distribution:

$$\hat{\lambda} \Longrightarrow \begin{cases} \frac{1}{(\widetilde{\phi^{\Delta}}+\mathcal{Z})^2-1} & \text{if } (\widetilde{\phi^{\Delta}}+\mathcal{Z})^2 > 1 \\ \infty & \text{otherwise} \end{cases}, \tag{54}$$

where $\mathcal{Z} \sim \mathcal{N}(0,1)$ and $\Longrightarrow$ signifies convergence in distribution. Thus, as the magnitude of the misspecification decreases, the probability that the posterior of λ peaks at ∞ increases. However, even if $\widetilde{\phi^{\Delta}} = 0$ this probability is typically not equal to one, as the local misspecification parameter $\widetilde{\phi^{\Delta}}$ cannot be estimated consistently. If the misspecification is large, $\hat{\lambda}$ is close to zero with high probability. Hence, also from a frequentist perspective, large values of $\hat{\lambda}$ can be interpreted as evidence in favor of small misspecifications.

We have shown that the value of $\hat{\lambda}$ reflects the amount of misspecification present in the data generating process. For the simple AR(1) model the true impulse response function is

$$\frac{\partial y_{T+h}}{\partial u_T} = \begin{cases} 1 & \text{if } h = 1 \\ T^{-h/2}\widetilde{\phi^{\Delta}}^{h} & \text{otherwise} \end{cases}. \tag{55}$$

Thus, both reliable impulse responses, that can serve as a benchmark for the evaluation of the DSGE model, as well as accurate forecasts from an autoregressive specification require a precise estimate of $\widetilde{\phi^{\Delta}}$. We now document that when the amount of actual misspecification is neither too small or too big, DSGE-VAR($\hat{\lambda}$) provides more accurate impulse response functions and forecasts than either of the two extremes, the VAR with diffuse prior or the VAR with DSGE model restriction dogmatically imposed.

The sampling distribution of the posterior mean $\hat{\phi}_b(\hat{\lambda})$ can be approximated in large samples by[9]

$$\sqrt{T}\hat{\phi}_b(\hat{\lambda}) \Longrightarrow \begin{cases} \widetilde{\phi^{\Delta}} + \mathcal{Z} - \frac{1}{\widetilde{\phi^{\Delta}}+\mathcal{Z}} & \text{if } (\widetilde{\phi^{\Delta}}+\mathcal{Z})^2 > 1 \\ 0 & \text{otherwise} \end{cases} \tag{56}$$

If the misspecification is large, then the posterior mean estimator corresponds, approximately, to the maximum likelihood estimator of ϕ, which has the limit distribution $\widetilde{\phi^{\Delta}} + \mathcal{Z}$ in our example. As the misspecification decreases, the probability that the DSGE-VAR($\hat{\lambda}$)

[8] If the data are generated under fixed misspecification $\phi = \phi^{\Delta}$ the posterior mode $\hat{\lambda}$ is driven to zero asymptotically. As the sample size increases the potential advantage from imposing the DSGE model restriction $\phi = 0$ vanishes as $\phi = \phi^{\Delta}$ can be consistently estimated. Such an analysis however does not capture the trade-offs that a researcher faces in finite samples.

[9] The estimator $\hat{\phi}_b(\hat{\lambda})$ is often called empirical Bayes estimator, see, for instance, Robert (1994). The most famous example of an empirical Bayes estimator is James' and Stein's celebrated estimator for the mean of a multivariate normal distribution.

will impose the DSGE model restrictions increases. Under a quadratic loss function the frequentist estimation risk for the misspecification parameter is given by

$$T \cdot I\!\!E\left[(\hat{\phi} - T^{-1/2}\widetilde{\phi}^{\Delta})^2\right]. \tag{57}$$

Moreover, for this simple model, the measure of expected forecast accuracy is proportional to (57). We plot this risk in Figure 1. For small values of $\widetilde{\phi^{\Delta}}$ the most precise estimate is obtained from the DSGE model ($\lambda = \infty$) itself. As the misspecification increases the autoregressive model with the diffuse prior $\lambda = T^{-1}$ eventually dominates the DSGE model. The DSGE-VAR($\hat{\lambda}$) has the property that it is preferable to the VAR with diffuse prior if the misspecification is small and that it dominates the DSGE model if the misspecification is large.

In our example there is a range of values for $|\widetilde{\phi^{\Delta}}|$, from 0.9 to 1.2, for which $\hat{\phi}_b(\hat{\lambda})$ delivers the best estimates of the misspecification and therefore is a desirable benchmark for the impulse-response function based evaluation of the DSGE model. We will present empirical evidence in Section 5 that this is indeed the relevant range of misspecification as pseudo out-of-sample forecasts for DSGE-VAR($\hat{\lambda}$) clearly dominate those from both the DSGE model and the VAR with diffuse prior.

3.5 Identification

According to the VAR approximation of the DSGE model the reduced-form innovations are functions of the structural shocks ϵ_t that generate the fluctuations in the DSGE model:

$$u_t = \Sigma_{tr}\Omega\epsilon_t, \tag{58}$$

where Σ_{tr} is the Cholesky decomposition of Σ and Ω is an orthonormal matrix with the property $\Omega\Omega' = I$. The matrix Ω is not identifiable since the likelihood function of the VAR depends only on the covariance matrix

$$\Sigma_u = \Sigma_{tr}\Omega\Omega'\Sigma_{tr}' = \Sigma_{tr}\Sigma_{tr}'.$$

For an impulse response function based evaluation of the DSGE model, a matrix Ω has to be chosen to compute responses to structural shocks from the benchmark model. While the literature contains many approaches to identify a small number of very specific shocks, such as a technology shock or a monetary policy shock, the identification of an entire vector of structural shocks in large dimensional VARs is still an open research question. One of the requirements is that such an identification scheme has the property that if the data are generated from the DSGE model, then the structural shocks are correctly identified.

Del Negro and Schorfheide (2004) proposed to construct Ω as follows. The DSGE model itself is identified in the sense that for each value of θ there is a unique matrix $A_0(\theta)$, obtained from the state space representation of the DSGE model, 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

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{DSGE} = A_0(\theta) = \Sigma^*_{tr}(\theta)\Omega^*(\theta), \tag{59}$$

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

$$\left(\frac{\partial \Delta y_t}{\partial \epsilon_t'}\right)_{VAR} = \Sigma_{tr}\Omega. \tag{60}$$

To identify the DSGE-VAR, we maintain the triangularization of its covariance matrix Σ_u and replace the rotation Ω in Equation (60) with the function $\Omega^*(\theta)$ that appears in (59). Loosely speaking, the rotation matrix is such that in absence of misspecification the DSGE's and the DSGE-VAR's impulse responses to all shocks would coincide. To the extent that misspecification is mainly in the dynamics, as opposed to the covariance matrix of innovations, the identification procedure can be interpreted as matching, at least qualitatively, the short-run responses of the VAR with those from the DSGE model. Since the DSGE model essentially determines the direction of the responses, the approach is similar in spirit to the sign-restriction identification schemes proposed by Canova and De Nicoló (2002) and Uhlig (2001), except that the sign restrictions are constructed directly from a fully-specified structural model.

The implementation of this identification procedure is straightforward in our framework. Since we are able to generate draws from the joint posterior distribution of Φ, Σ_u, and θ, we can for each draw (i) use Φ to construct a MA representation of Δy_t in terms of the reduced-form shocks u_t, (ii) compute a Cholesky decomposition of Σ_u, and (iii) calculate $\Omega = \Omega^*(\theta)$ to obtain a MA representation in terms of the structural shocks ϵ_t.

In Del Negro and Schorfheide (2004) the identification procedure was applied to a trivariate VAR in output growth, inflation, and nominal interest rates, driven by a technology shock, a government spending shock, and a monetary policy shock. In this paper we apply the approach to a seven-variable VAR. Once identification has been achieved a comparison with DSGE model impulse responses can generate important insights in the potential misspecification of the DSGE model. The spirit of this evaluation is to keep the autocovariance sequence associated with the benchmark model, that is the DSGE-VAR, as close to the DSGE model as possible without sacrificing the ability to track the historical time series.

3.6 How Well is the DSGE Model Approximated?

At the beginning of this section we described DSGE-VAR(λ) (or DSGE-VECM(λ)) as a continuum of specifications with an essentially unrestricted VAR at one extreme and the VAR approximation to the DSGE model at the other. The fact that for $\lambda = \infty$ we only obtain an approximation of the log-linearized DSGE model raises the question why we did not start out from a more general VARMA model that nests the moving average representation of the DSGE model. The answer is twofold. First, VARs have established themselves as popular and powerful tools for empirical research and forecasting in macroeconomics. Second, from a computational perspective the posterior of DSGE-VAR is much easier to analyze than the posterior of a DSGE-VARMA.[10]

The accuracy of the VAR approximation of the DSGE model depends on the invertibility of the DSGE model's moving average components and on the number of included autoregressive lags. Consider the following example. Suppose according to the log-linearized DSGE model

$$y_t = \theta\epsilon_t + \epsilon_{t-1} = (\theta + L)\epsilon_t, \quad \epsilon_t \sim iid\mathcal{N}(0,1), \tag{61}$$

where y_t is a scalar, L is the lag operator, and $0 \leq \theta < 1$. Thus, in response to ϵ_t, y increases between t and $t+1$ and subsequently drops to zero. Since the roots of the MA polynomial lie inside the unit circle, the lag polynomial is not invertible and y_t does not have an autoregressive representation in terms of the structural shocks ϵ_t.

Now consider the model

$$y_t = \eta_t + \theta\eta_{t-1} = (1 + \theta L)\eta_t, \quad \eta_t \sim iid\mathcal{N}(0,1), \tag{62}$$

which is observationally equivalent to (61) since it generates the same autocovariance sequence. However, the impulse response function looks quite different. In response to a

[10] For the VAR, we can calculate the marginal likelihood function conditional on the DSGE model parameters θ and the hyperparameter λ analytically. This marginal likelihood can be used for to generate draws from the marginal posterior of θ. For a VARMA model, this analytical calculation is not possible.

positive shock η_t, y falls between period t and $t+1$. Unlike (61), the model (62) is invertible and has the infinite-order autoregressive representation

$$y_t = -\sum_{j=1}^{\infty} (-\theta)^j y_{t-j} + \eta_t. \tag{63}$$

The analysis in this paper is based on a finite-order approximation of (63). Therefore, if the DSGE model had a non-invertible MA representation, then the impulse response functions generated from the VAR approximation of the DSGE model would be misleading. Alternatively, if the DSGE model corresponded to (62) with θ close to unity, then we would need many lags to obtain an accurate autoregressive approximation. In practice, the number of lags that can be used in an autoregressive approximation is typically constrained by the number of observations that are available to initialize lags and to estimate the coefficients.

Fernandez-Villaverde, Rubio-Ramirez, and Sargent (2004) provide necessary and sufficient conditions for the invertibility of the moving average components of linear state-space models. Chari, Kehoe, and McGrattan (2004) illustrate that a large number of autoregressive lags is needed to approximate the moving average representation of hours worked and output generated by a standard neoclassical growth model and to accurately recover the response of hours worked to a technology shock using long-run identification restrictions. To check whether the VAR approximation of our DSGE model is reliable, we compare in Figure 2 the impulse responses from the DSGE model with those from the DSGE-VAR(∞) with four lags. These latter impulse responses are obtained using the identification procedure described in the previous subsection. The impulse responses to output, consumption, investment, and the real wage are cumulative. We fixed the parameter vector θ at the posterior mean values reported in column 5 of Tables 2 and 3.

The solid and dash-and-dotted lines in Figure 2 represent the impulse responses of the DSGE model and the DSGE-VAR(∞), respectively. For many of the impulse responses, for example those of output, hours, inflation, and interest rates, the approximation is good. For instance, in terms of cumulative output the maximum difference between the DSGE model's and DSGE-VAR(∞)'s impulse responses over the horizon (sixteen quarters) is less than 10 basis points for both monetary policy and technology shocks. For other impulse responses however the approximation breaks down, most notably for the responses of consumption and investment.

The inclusion of the error-correction terms in the DSGE-VECM specification is able to reduce the approximation error substantially. For instance, the response of consumption to a discount rate shock reverts to zero after sixteen quarters according to both the

DSGE model and DSGE-VECM(∞), while it is well above one percent according to DSGE-VAR(∞).[11] Overall the responses of the DSGE-VECM(∞) (dotted lines in Figure 2) track the DSGE model's well. There are a few instances where the DSGE model's responses and DSGE-VECM(∞)'s are different, such as the responses of hours or the real wage to a g (government spending) shock, or the response of cumulative investment to a λ_f (mark-up) shock. Whenever this is the case, the magnitude of the difference in impulse-responses is still contained, however, relative to the overall variability of the series. The maximum differences between the DSGE model's and DSGE-VECM(∞)'s impulse responses are 160 basis points for cumulative investment, which is small relative to the overall variability of the series. To double check that even these minor differences eventually disappear, we also computed the responses of DSGE-VAR(∞) and DSGE-VECM(∞) with forty lags. Now, the impulse responses of the DSGE model are virtually indistinguishable from those of DSGE-VECM(∞), while the approximation of DSGE-VAR(∞) is about as good as that of DSGE-VECM(∞) with four lags only.

This finding suggests that DSGE-VECM(∞) is a parsimonious way to approximate the DSGE model in presence of cointegration restrictions, and appears to be fairly successful even with a moderate number of lags.[12] The remainder of the paper mainly focuses on results for DSGE-VECM, although we also discuss the results for the DSGE-VAR specification. In estimating and assessing the fit of the DSGE model we condition on the same information set used in DSGE-VECM. Specifically, we are conditioning on x_0, the p initial lags of the endogenous variables, as well as on the initial values of the cointegration vector (31).[13]

[11] It is well known in the context of non-stationary vector autoregressive systems that error-correction terms can eliminate unit roots from moving average polynomials that appear in representations for first differences. For instance, let $y_{1,t} = \theta y_{2,t} + \epsilon_{1,t}$ and $\Delta y_{2,t} = \epsilon_{2,t}$. If one expresses $\Delta y_{1,t}$ as a function of $\Delta y_{1,t-1}$ and $\Delta y_{2,t-1}$, then a unit root in the moving average polynomial arises $\Delta y_{1,t} = \theta \Delta y_{2,t-1} + \epsilon_{1,t} - \epsilon_{1,t-1}$ and an approximation of $\Delta y_{1,t}$ through a finite-order VAR in differences will be inaccurate. However, once the error correction term $y_{1,t} - \theta y_{2,t}$ is included, the unit root in the moving-average polynomial vanishes $\Delta y_{1,t} = -(y_{1,t-1} - \theta y_{2,t-1}) + \epsilon_{1,t} + \theta \epsilon_{2,t}$.

[12] To further investigate the issue of invertibility we randomly generated data from the DSGE model, and then checked whether DSGE-VECM(∞) was able to reproduce the original time series of structural shocks. We find that this is indeed the case, even with four lags.

[13] This is achieved by running the Kalman filter on the initial observations x_0, and then using the resulting mean and variance for the state as starting values in the estimation on $\Delta y_1 \ldots \Delta y_T$. Note that the initial values of the cointegration vector combined with the sample information $\Delta y_1 \ldots \Delta y_T$ implies that we are effectively giving the model information on the values of the cointegration vector for $t = 1 \ldots T$.

4 The Data

All data are obtained from Haver Analytics (Haver mnemonics are in italics). Real output, consumption, and investment are obtained by dividing the nominal series (*GDP*, *C*, and *I*, respectively) by population 16 years and older (*LF*+*LH*), and deflating using the chained-price GDP deflator (*JGDP*). The real wage is computed by dividing compensation per hour in the non-farm business sector (*LXNFC*) by 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. Labor supply is computed by dividing hours of all persons in the non-farm business sector (*LXNFH*) by population. Hours of all persons in the non-farm business sector is an index developed by the Bureau of Labor Statistics that includes the labor supply of both wage and salary workers and proprietors. This measure of labor supply best corresponds to our measure of real wages.[14] All growth rate are computed using log-differences from quarter to quarter, and are in percent. Inflation is computed using log-differences of the GDP deflator, in percent. The nominal rate corresponds to the effective Federal Funds Rate (*FFED*), also in percent. Data are available from QIII:1954 to QI:2004.

5 Empirical Results

The empirical analysis has five parts. First, in a preliminary analysis we estimate the state space representation of the log-linear DSGE model directly (not its VAR/VECM approximation) and use marginal data densities to choose a baseline specification. Second, we discuss parameter estimates for the baseline DSGE model and smoothed exogenous processes. Third, we estimate DSGE-VECM and DSGE-VAR models for various values of λ and document model fit in terms of marginal likelihoods and, equivalently, present the posterior distribution of the hyperparameter. Fourth, we compare the pseudo-out-of-sample forecasting performance of the DSGE model, the DSGE-VECM($\hat{\lambda}$), and the VECM with a diffuse prior. Finally, we document the discrepancy between the impulse response functions of the DSGE model and that of the DSGE-VECM($\hat{\lambda}$). All results reported in this paper that are based on Markov Chain Monte Carlo simulations are computed using 110,000 draws and discarding the first 10,000. We checked whether 110,000 draws were sufficient by repeating the estimation procedure several times and verifying that we obtain the same results for parameter estimates and log marginal likelihoods.

[14] Since we use an index as a measure of hours, which enter our specification in level, we need to pin down the average value of the index via an additional free parameter in the DSGE model.

5.1 Choosing a Baseline DSGE Model

Before relaxing the DSGE model restrictions we consider four different versions and determine which of the specifications attains the highest marginal data density. In the baseline specification, denoted by S0, prices and wages are indexed with respect to steady state price and wage inflation ($\iota_p = \iota_w = 0$), also known as static indexation. The output gap in the Taylor rule (21) is defined using the trend of output along the stochastic growth path $Y_t^* = Y_t^s$. Moreover, the fixed costs $\mathcal{F}$ in the production function for the intermediate goods producers in Eq. (4) are set to zero.

Alternative versions of the DSGE model are obtained by modifying one aspect of the baseline specification at a time. Specification S1 differs from the baseline version of the DSGE model in that prices and wages are indexed with respect to last period's price and wage inflation rates ($\iota_p = \iota_w = 1$), also known as dynamic indexation. In specification S2 the output gap in the Taylor rule is calculated based on the flexible price output $Y_t^* = Y_t^f$. Finally, in specification S3 the fixed costs are determined endogenously to erase steady state profits of the intermediate goods producers.

Log marginal likelihoods for the four specifications are reported in Table 1. The posterior odds are equal to the exponential of the log marginal likelihood differentials and summarize the the odds of specification Si, $i = 1, 2, 3$, versus the baseline specification S_0. S1 is clearly rejected by the data as the posterior odds are virtually zero. While Eichenbaum and Fisher (2004) find evidence in favor of dynamic indexation in a single-equation framework in which only the Phillips curve is estimated, their finding does not appear to hold in a multiple equation framework such as the one considered here. We also strongly reject specification S3 in which the steady state profits are forced to be zero. The odds against the flexible-price output target in the Taylor rule (S2) are less decisive but still favor the baseline version of the DSGE model. Hence, all subsequent results are based on specification S0.

5.2 In-Sample Fit of the DSGE Model and Parameter Estimates

Figure 3 provides a first visual diagnostic of the DSGE model. The figure plots the actual data (dark lines), as well as the one-period-ahead forecasts obtained from the Kalman filter (gray lines), computed using the posterior mean of θ reported in column 5 of Tables 2 and 3. The in-sample fit of the DSGE model is fairly satisfactory as there appear to be no big discrepancies between actual and fitted values. However, in terms of real activity and real wages, the model seem to have a hard time fitting the most volatile periods, such as the mid-seventies. Importantly, the model consistently over-predicts consumption growth in the first part of the sample, and under-predicts consumption growth in the second part, except during the 1990 recession. This is a first indication that the balanced growth path implications of the DSGE model are at odds with the data. In terms of nominal variables the model under-predicts inflation in the late seventies, and over-predicts inflation toward the end of the sample. Under(over)-predictions for inflation generally translate into under(over)-predictions for the nominal interest rate.

Tables 2 and 3 report on the estimates of the DSGE model parameters. The Table contains information on the prior distributions as well as the posterior means and the 90% probability intervals of the structural parameters based on the estimation of the state space representation of the DSGE model and the estimation of the DSGE-VECM($\hat{\lambda}$).[15] For now, we focus on the former. For most of the parameters the priors coincide with those used by Smets and Wouters (2003a,b). Onatsky and Williams (2004) estimate the Smets-Wouters model on Euro Area data with priors that are less informative than ours and those used in Smets and Wouters (2003a). While they obtain different parameter estimates, they find that the dynamics of their estimated DSGE model are similar to those obtained with the Smets and Wouters (2003a) parameter estimates. The parameter estimates are also generally in line with those of Smets and Wouters (2003b). 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 0.848 and 0.936, respectively. Smets and Wouters (2003b) also present high estimates of these parameters.

Of particular interest for the evaluation of the DSGE model are the parameters describing the autocorrelation of the underlying exogenous processes: ρ_z (technology), ρ_φ (preferences of leisure), ρ_μ (shocks to the capital accumulation equation), b (overall preference shifter), ρ_g (government spending), and ρ_{λ_f} (price markup shocks). Since we model the level of technology Z_t as a unit root root process, the estimate of ρ_z, which measures the serial correlation of technology growth z_t, is low. All other processes are strongly autocorrelated, particularly those for the government spending shock g_t. However, for most of these shocks the degree of persistence is not as high as that found in Smets and Wouters (2003b). The high persistence of many of the exogenous processes raises concerns about the ability of the DSGE model to generate endogenous propagation mechanisms. While this

[15] A few of the DSGE model's parameters were fixed at the onset: δ, λ_f and λ_w were set at 0.025, 0.3 and 0.3, respectively.

lack of internal propagation is well documented for small-scale DSGE model such as the benchmark three-equation model described, for instance, in Woodford (2003), it is also a concern for models with capital accumulation, variable capital utilization, adjustment costs, habit formation, as well as price and wage stickiness as the one estimated in this paper.

Figure 4 plots the Kalman-smoothed time series for the processes z_t (technology), φ_t (preferences of leisure), μ_t (shocks to the capital accumulation equation), b_t (overall preference shifter), g_t (government spending), and $\lambda_{f,t}$ (price markup shocks), computed using the posterior mean of θ. Not surprisingly, given the estimates of the autoregressive coefficients shown in Table 3, many of the exogenous processes are indeed persistent. For instance, leisure preference shocks are positive in the first part of the sample, where total hours are generally lower than average (see Figure 3), and mostly negative in the second part and particularly in the nineties, where hours are above average.

The government spending process g_t clearly has a downward trend which is deemed as very unlikely by the stationarity assumption stated in Eq. (24). The reason for this negative trend can be traced to the consistent under-prediction of consumption starting in the early eighties, documented in Figure 3. According to the model investment, output, and consumption all grow at the same rate, when measured in nominal terms (or in real terms when deflated by the same price deflator). In the data, consumption has been growing faster than either output or investment. The DSGE model's inability to account for this fact may explain the downward path of g_t in Figure 4. The impulse responses (Figure 2) show that government spending shocks have the largest – and opposite – effect on output and consumption, and generally a small impact on investment and the other series ("small" relative to the overall volatility of the series, as can be gauged from other impulse-responses in the same column). While latent government spending shocks can to some extent compensate for growth rate differentials and boost the in-sample fit of the DSGE model, they are less helpful in adjusting long horizon out-of-sample forecasts as we will document subsequently.

The above results are based on thirty years of data ($T = 120$), starting in QII:1974 and ending in QI:2004. We use thirty years because this the amount of data used in the estimation for the rolling sample forecasting exercise. The findings are qualitatively unchanged when we use the entire sample, from QIII:1955 to QI:2004. Also, we obtain similar results when we estimate the DSGE model without conditioning on the initial value of the cointegration vector (Eq. 31).

5.3 Relaxing the DSGE Model Restrictions

An important finding of Smets and Wouters (2003a, Table 2) is, that, at least for the Euro area data, large-scale new-Keynesian DSGE models can fit better than VARs. Their estimated DSGE model attains a higher marginal likelihood than VARs of lag order one to three with training-sample priors, and a VAR (1 lag) with a Minnesota-type prior. Only a VAR (3 lags) with Minnesota prior is able to outperform the DSGE model. The Smets and Wouters (2003a) finding is qualitatively different from earlier results for small-scale DSGE models, e.g., Schorfheide (2000), who finds that cash-in-advance models fitted to output growth and inflation data are unable to dominate VARs with up to four lags and training-sample priors.

The analysis in Smets and Wouters (2003a) has, however, some caveats as pointed out by Sims (2003). First, in- and out-of-sample comparisons are based on linearly detrended data instead of raw data. Second, the use of a training sample prior for some of the VAR specifications but not for the DSGE model potentially generates a disadvantage for the VAR if the training sample observations are qualitatively different from the observations in the estimation sample. Third, the set of models considered is sparse, as it only contains the DSGE model itself as well as VARs with fairly diffuse priors. In particular, the results seem to be quite sensitive to the specification of the VAR prior.

These shortcomings are addressed in the subsequent analysis through the use of DSGE-VECMs(λ) that are fitted to non-detrended data. Instead of simply considering extreme values for λ, that is, $\lambda = (k+n)/T$, which is the smallest value of λ for which our prior integrates to one, and $\lambda = \infty$, we consider a range of intermediate values that allow for various degrees of deviation from the DSGE model restrictions. While the DSGE-VECM(∞) provides a better approximation of the state-space representation of the DSGE model than the DSGE-VAR(∞), as documented in Section 3, we report in Table 4 log marginal likelihoods for both the DSGE-VECM(λ) and the DSGE-VAR(λ) specifications. The latter has the advantage that it relaxes some of the co-trending implications of the VECM that appear to be counterfactual according to the direct estimation of the DSGE model. For both the DSGE-VECM and the DSGE-VAR the number of lags is four.

We begin with a discussion of the DSGE-VECM results reported in columns 2 and 3 of Table 4. The second row of the Table contains the log marginal likelihood for the directly estimated DSGE model (state space representation), which is very similar to the value for the DSGE-VECM(∞) suggesting that the approximation error due to the lag truncation is indeed fairly small. As the weight on the DSGE model restrictions is reduced and λ is lowered to 1, the log marginal likelihood of the DSGE-VECM increases. Hence, taking a potential misspecification of the DSGE model restrictions into account improves the fit of the model. The posterior of λ peaks at 1 and the marginal likelihood falls as the hyperparameter is decreased to 0.33, which is the smallest value of λ that yields a proper prior and a well-defined marginal likelihood, in our application. Table 4 shows that the posterior distribution of λ has an inverse U-shape as one would expect if the DSGE model is to some extent misspecified. Unlike the analysis in Smets and Wouters (2003a) on detrended Euro-area data, our findings for non-detrended U.S. data are less favorable for the DSGE model. The fit of the DSGE model is a lot worse than that of the DSGE-VECM($\hat{\lambda}$). The differences in log marginal likelihoods are so large that the posterior odds of the DSGE model are practically zero.

Table 4 also presents log marginal likelihoods for the DSGE-VAR(λ). As for the DSGE-VECM, the posterior of λ has an inverse U-shape, but it peaks at $\lambda = 0.75$ instead of $\lambda = 1$. The likelihood discrepancy between the state space representation of the DSGE model and the VAR approximation is much larger than in the VECM case, which is qualitatively consistent with the impulse response comparison in Figure 2. Since the VAR specification does not impose the empirically inaccurate co-trending restrictions on consumption, investment, output, and real wages the DSGE-VAR($\hat{\lambda}$) fits better than the DSGE-VECM($\hat{\lambda}$). Nevertheless, we proceed with the analysis of the DSGE-VECM(λ) specifications as they come closer to nesting the DSGE model. While $\hat{\lambda}$ provides an overall measure of the degree of misspecification of the DSGE model we now explore the misspecification in more detail by considering the forecasting performance and a comparison of impulse responses.

Table 4 is based on 30 years of data ($T = 120$), starting in QII:1974 and ending in QI:2004. The results in Table 4 are remarkably robust: For each date of the rolling sample, from QIV:1985 to QI:2000, the shape of the posterior of λ is qualitatively the same, with the only difference that the peak of the posterior is $\lambda = .75$ for some dates and $\lambda = 1$ for others. We also varied the prior distribution for the structural parameters θ and did not find a significant effect on the shape of $p_\lambda(Y)$. In fact $\hat{\lambda}$ and the associated posterior densities appears to be much more robust than the odds ratio of the extremes that is DSGE-VECM(∞) versus DSGE-VECM($(n+k)/T$).

Finally, we discuss the posterior estimates of θ obtained from DSGE-VECM($\hat{\lambda}$) contained in Tables 2 and 3. Roughly speaking, these estimates are obtained by projecting the posterior estimates of Φ and Σ_u onto the restriction functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ (for details see Del Negro and Schorfheide, 2004). We find that the estimates obtained from DSGE-VECM($\hat{\lambda}$) are broadly in line with those obtained from the DSGE model. Interestingly, relaxing the cross-equation restrictions generally leads to a reduction in the estimated degree of persistence of the exogenous processes, as well as a reduction in the degree of stickiness of wages and prices.

5.4 Pseudo-Out-of-Sample Forecast Accuracy

We now discuss the pseudo-out-sample fit of the DSGE model (state-space representation)[16] and compare it to that of the DSGE-VECM($\hat{\lambda}$) and a VECM with diffuse prior. Unlike in the previous subsection, in which the diffuse prior was obtained by $\lambda = 0.33$ to guarantee that the corresponding marginal likelihood is well-defined, we now report forecasts from the DSGE-VECM(0). Since for $\lambda = 0$ the posterior mean of Φ is simply the OLS estimate of Φ we also refer to the DSGE-VECM(0) as unrestricted VECM.

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 fifty-eight periods. At each date of the rolling sample we use the previous 120 observations to re-estimate the models, and the following twelve quarters to assess forecasting accuracy, which is measured by the root mean squared error (RMSE) of the forecast. For the variables that enter the VECM in growth rates (output, consumption, investment, real wage) and inflation we forecast cumulative changes. For instance, the RMSE of inflation for twelve quarters ahead forecasts measures the error in forecasting cumulative inflation over the next three years (in essence, average inflation), as opposed to inflation exactly three years down the road. At each date, we also re-compute the posterior mode $\hat{\lambda}$ to construct forecasts from the DSGE-VECM($\hat{\lambda}$). As discussed above, the value of $\hat{\lambda}$ hovers between 0.75 and 1.00. When estimating the DSGE model we condition on the very same information that is used to initialize the lags that appear in the VECM specification. Table 5 documents for each series and for each forecast horizon the RMSE for DGSE-VECM(λ) as well as the percentage improvement (in parenthesis) in RMSE relative to both the DGSE model and the unrestricted VECM. 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.

[16] While the forecasts from the state-space representation of the DSGE model and the DSGE-VECM(∞) are very similar, we decided to report forecast errors for the former in Table 5.

In the context of the AR(1) example in Section 3 we illustrated that when the misspecification of the DSGE model is small the most precise estimate of the autoregressive coefficients is obtained by imposing the restrictions. On the other hand, if the misspecification is very large, it is best to ignore the DSGE model restrictions. However, according to Figure 1 there is an intermediate range for the misspecification values $\tilde{\Phi}$ in which the $\lambda = \infty$ are approximately as precise, in a mean squared error sense, as the estimates obtained under the diffuse prior. According to the multivariate forecast error statistic reported in Table 5 the improvement of the DSGE-VECM($\hat{\lambda}$) over the DSGE model and the unrestricted VECM are 13.8% and 15.0%, respectively. Hence, the the one-step-ahead forecasting performance for $\lambda = 0$ and $\lambda = \infty$ is approximately the same, which is consistent with the view that the misspecification of the DSGE model is modest and that it provides a good albeit not perfect approximation of reality. Relaxing, yet not ignoring the restrictions leads to an improvement in fit and forecasting performance.

In general, for one-step ahead forecasts DGSE-VECM($\hat{\lambda}$) appears to be more accurate than both the DSGE model and the VECM. Two exceptions are the real wage and the inflation forecasts, which hardly improve as the DSGE model restrictions are relaxed. While the DSGE model outperforms the unrestricted VECM in terms of interest rate forecasts, the VECM delivers more precise consumption and investment forecasts.[17]

According to the multivariate statistic the forecast accuracy improvements obtained by optimally relaxing the DSGE model restrictions are largest for medium-run (4 to 8 quarters) forecasts, and then tend to decline in the longer-run. For many of the individual series, such as output, consumption, investment, and hours, the improvements are substantial and increase steadily with the forecast horizons. An exception is again the real wage series. For forecasts beyond one quarter ahead, DGSE-VECM($\hat{\lambda}$) actually does worse than the DSGE model, and the discrepancy rises with the forecast horizon. In terms of medium and long-run forecasts the unrestricted VECM generally outperforms the DSGE model, which is another piece of evidence that the balanced-growth path restrictions embodied in the DSGE model are to some extent counterfactual.[18]

[17]Separately, we have also plotted the percentage increase in forecasting accuracy of DGSE-VECM(λ) relative to the unrestricted VECM, as measured by the RMSE. Consistently with the results in the previous section we find that for most series and forecasting horizons the increase in forecasting accuracy as a function of λ displays an inverse U-shape, first increasing a then declining as λ goes from zero to infinity.

[18]We reach by and large the same conclusions for the DSGE-VAR specification (results are available upon request). Consistently with the results in Table 4 the VECM specification does somewhat worse than the VAR specification in terms of RMSEs for most of the variables with the exception of consumption. For series like investment the reduction in long run forecast accuracy is quite large.

5.5 Comparing the Propagation of Shocks

We conclude the empirical analysis with an assessment of the DSGE model misspecification based on impulse response functions. A reliable benchmark model is needed in order to evaluate DSGE models based on impulse response functions. If the DSGE model were to fit better than the benchmark model, nothing could be learned about the DSGE model from a comparison of impulse response functions. The findings in Smets and Wouters (2003a) and the empirical results reported in the preceding subsections indicate that a VAR or a VECM specification estimated under a diffuse prior distribution is not always a useful benchmark. We found that the DSGE-VECM($\hat{\lambda}$) clearly dominates the DSGE-VECM(0) and DSGE-VECM($(n+k)/T$). The spirit of our impulse response function based evaluation is to relax the DSGE model restriction by reducing λ until the fit cannot be improved further. Thus, we are creating a benchmark that is favorable toward the DSGE model, in that we are trying to keep the deviation from the DSGE model restrictions as small as possible. In our application this is achieved by setting $\lambda = \hat{\lambda} = 1$.

Figure 5 shows the mean impulse-responses for DGSE-VECM($\hat{\lambda}$) (dash-and-dotted lines), the ninety percent bands (dotted lines), and the mean impulse-responses for the DGSE model. The impulse-responses for the DGSE model are computed using the same draws of DSGE model parameters θ that generate the DGSE-VECM($\hat{\lambda}$) impulse-responses. One important feature of the procedure developed in this paper is that it delivers identified DSGE-VECM impulse responses even when λ is less than infinity. Figure 5 shows that for $\lambda = 1$ the identification procedure is fairly successful also for relatively large systems with as many as seven shocks. By successful we mean that the impulse-responses to all seven shocks are economically interpretable, in that they agree with the DSGE model by and large in terms of the direction of the response.

We find that several of DGSE-VECM($\hat{\lambda}$)'s impulse-responses are not only qualitatively but also quantitatively in agreement with the DSGE model's. This is the case for the responses to capital adjustment shocks (μ), and mark-up shocks (λ_f). Other impulse responses exhibit discrepancies. The impulse-response to a technology growth shock (*Tech*) is more pronounced in the medium run for output, consumption, investment and hours under DGSE-VECM($\hat{\lambda}$) than under the DSGE model. Also, the response of inflation is more persistent. The effects of the preference shock (φ) on output, consumption, and hours are more persistent according to the VECM specification, which indicates a lack of internal propagation of labor supply shifts. The intertemporal preference shock (b) has a much larger effect on the nominal interest rate in the VECM than it has in the DSGE model. Since b

and R are related through the consumption Euler equation, the discrepancy suggests potential misspecification of the consumption-based pricing kernel. According to the DSGE model, output and hours increase immediately in response to a government spending shock and quickly decay monotonically. The VECM specification, on the other hand, predicts delayed, hump-shaped responses of both output and hours that are long-lasting. Moreover, the VECM implies that the government shock is accompanied by a fall in nominal interest rates whereas the DSGE model generates a small rise in R. While both DSGE model and VECM agree on the response of inflation and interest rates to a monetary policy shock, the VECM generates much stronger real effects.

6 Conclusions

Smets and Wouters (2003a) showed that large-scale New-Keynesian models with real and nominal rigidities can fit as well as VARs estimated under diffuse priors, and possibly better. This result implies that these models are becoming a tool usable for quantitative analysis by policy making institutions. Their finding suggests that it is now worthwhile to carefully document the out-of-sample performance of the DSGE model. In addition, it implies that more elaborate tools for model evaluation are necessary. It is not guaranteed that vector autoregressions estimated with simple least squares techniques, or from a Bayesian perspective, estimated under a very diffuse prior, provide a reliable benchmark. This paper has addressed both issues. We conducted a pseudo-out-of-sample forecasting experiment. Moreover, using techniques developed in Del Negro and Schorfheide (2004) we constructed a reliable benchmark by systematically relaxing the restrictions that the DSGE model poses on a vector autoregressive to optimize its fit measured by the marginal likelihood function. According to our empirical results, one of the biggest impediments to fitting a large vector of macroeconomic variables are the counterfactual co-trending implications of the DSGE model.

Thus, much work lies ahead both in terms of modeling and econometrics. We need to build models that can be successfully taken to non-detrended data – models that fulfill Kydland and Prescott (1982)'s original promise of integrating growth and business cycle theory, so they can at the same time match both growth and business cycle features of the data. On the econometrics side we need to develop approaches that use the DSGE model restrictions, but down-weight those frequencies where the DSGE model's implications are more at odds with the data, and emphasize those where the DSGE model may be most useful.

Progress in either direction may further enhance the use of DSGE models in quantitative policy analysis – the ultimate goal of our research agenda.

References

Altig, David, Lawrence J. Christiano, Martin Eichenbaum, and Jesper Linde (2002): "Technology Shocks and Aggregate Fluctuations," *Manuscript*, Northwestern University.

Boivin, Jean and Marc Giannoni (2003): "Has Monetary Policy Become More Effective," *NBER Working Paper*, **9459**.

Calvo, Guillermo (1983): " Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, **12**, 383-398.

Canova, Fabio (1994): "Statistical Inference in Calibrated Models". *Journal of Applied Econometrics*, **9**, S123-144.

Canova, F. and G. De Nicoló, "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7." *Journal of Monetary Economics* **49**(6) (2002), 1131-1159.

Chari, V.V., Patrick Kehoe, and Ellen McGrattan (2004): "Are Structural VARs Useful Guides for Developing Business Cycle Theories?" *Working Paper*, **631**, Federal Reserve Bank of Minneapolis, Research Department.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (2004): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, forthcoming.

DeJong, David N., Beth F. Ingram, and Charles H. Whiteman (1996): "A Bayesian Approach to Calibration". *Journal of Business Economics and Statistics*, **14**, 1-9.

DeJong, David N., Beth F. Ingram, and Charles H. Whiteman (2000): "A Bayesian Approach to Dynamic Macroeconomics". *Journal of Econometrics...*

Del Negro, Marco and Frank Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, **45**, 643-673.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984): "Forecasting and Conditional Projections Using Realistic Prior Distributions." *Econometric Reviews*, **3**, 1-100.

Dridi, Ramdan, Alain Guay, and Eric Renault (2003): "Calibration of DSGE Models: An Sequential Partial Indirect Inference Approach," *Manuscript*, University of North Carolina.

Eichenbaum, Martin and Jonas D.M. Fisher (2004): "Evaluating the Calvo Model of Sticky Prices," *NBER Working Paper*, **10617**.

Fernandez-Villaverde, Jesus, Juan Rubio-Ramirez, and Thomas J. Sargent (2004): "The A, B, C (and D's) of VARs," *Manuscript*, New York University.

Gelman, Andrew, John Carlin, Hal Stern and Donald Rubin (1995): "*Bayesian Data Analysis*". Chapman and Hall.

Geweke, John (1999): "Computational Experiments and Reality". *Manuscript*, University of Iowa.

Greenwood, Jeremy, Zvi Hercovitz, and Per Krusell (1998): "Long-Run Implications of Investment-Specific Technological Change." American Economic Review, **87**(3), 342-362.

Greenwood, Jeremy, Zvi Hercovitz, and Per Krusell (2000): "The Role of Investment-Specific Technical Change in the Business Cycle." European Economic Review, **44**, 91-115.

Good, I.J. (1952): "Rational Decisions". *Journal of the Royal Statistical Society B*, **14**, 107-114.

Ingram, Beth F. and Charles H. Whiteman (1994): "Supplanting the Minnesota prior – Forecasting macroeconomic time series using real business cycle model priors." *Journal of Monetary Economics*, **34**, 497-510.

Kydland Finn and Edward Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, **50**, 1345-70.

Onatski, Alexei and Noah Williams (2004): "Empirical and Policy Performance of a Forward-Looking Monetary Model," Manuscript, Princeton University, Department of Economics.

Nason, James M. and Timothy Cogley (1994): "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, **9**, S37-70.

Robert, Christian (1994): *"The Bayesian Choice,"* Springer-Verlag.

Rotemberg, Julio and Michael Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomics Annual*, **12**, 297-246.

Schorfheide, Frank (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, **15**, S645-670.

Schorfheide, Frank (2004): "VAR Forecasting Under Misspecification," *Journal of Econometrics*, forthcoming.

Sims, Christopher A. (2002): "Solving Rational Expectations Models." *Computational Economics*, **20**(1-2), 1-20.

Sims, Christopher A. (2003): "Probability Models for Monetary Policy Decisions," *Manuscript*, Princeton University.

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

Smets, Frank and Raf Wouters (2003b): "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach," *Manuscript*, European Central Bank.

Uhlig, Harald (2001) "What are the Effects of Monetary Policy? Evidence from an Agnostic Identification Procedure," *Manuscript*, Humboldt University.

Woodford, Michael (2003): *"Interest and Prices: Foundations of a Theory of Monetary Policy,"* Princeton University Press.

Zellner, Arnold (1971): *"Introduction to Bayesian Inference in Econometrics."* John Wiley & Sons, New York.

Table 1: Alternative DSGE Model Specifications: Log Marginal Likelihoods

	Log-Marginal Likelihood	(Difference wrt Baseline)	Post. Odds wrt Baseline
(S0) Baseline	-530.48	(0)	1
(S1) Dynamic Indexation ($\iota_p = \iota_w = 1$)	-561.97	(-31.49)	10^{-12} %
(S2) Flexible-price Output Target in Eq. (21)	-532.56	(-2.08)	12.49 %
(S3) $\mathcal{F}$ sets steady-state profits $= 0$	-559.35	(-28.87)	10^{-11} %

Notes: Baseline specification is: static price and wage indexation, output target is trend of output along stochastic growth path, fixed costs $\mathcal{F} = 0$. Posterior odds are computed as the exponential of the log-differences in marginal likelihood between two model specifications, and are expressed in percent. See Section 4 for a description of the data. Results are based on the sample period QII:1974 - QI:2004.

Table 2: DSGE Model's Parameter Estimates (Part I)

	Prior			DSGE Post.			DSGE-VECM($\hat{\lambda}$) Post.		
	Distr.	Mean	Stdd	Mean	Interval		Mean	Interval	
α	$\mathcal{B}$	0.250	0.100	0.172	0.149	0.195	0.153	0.125	0.181
ζ_p	$\mathcal{B}$	0.750	0.100	0.848	0.814	0.883	0.667	0.512	0.834
s'	$\mathcal{N}$	4.000	1.500	5.827	3.238	8.115	4.204	3.999	4.405
h	$\mathcal{B}$	0.800	0.100	0.793	0.725	0.857	0.691	0.585	0.791
$a'\,'$	$\mathcal{G}$	0.200	0.075	0.167	0.067	0.273	0.243	0.126	0.356
ν_l	$\mathcal{G}$	2.000	0.750	2.204	1.050	3.271	2.285	2.056	2.518
ζ_w	$\mathcal{B}$	0.750	0.100	0.936	0.913	0.959	0.812	0.720	0.905
r^*	$\mathcal{G}$	0.500	0.100	0.389	0.270	0.501	0.496	0.350	0.646
ψ_1	$\mathcal{G}$	1.700	0.100	1.799	1.640	1.944	1.768	1.609	1.993
ψ_2	$\mathcal{G}$	0.125	0.100	0.065	0.040	0.090	0.035	0.000	0.074
ρ_r	$\mathcal{B}$	0.800	0.100	0.815	0.775	0.855	0.799	0.748	0.850
π^*	$\mathcal{N}$	0.650	0.200	1.026	0.807	1.264	0.553	0.296	0.843
γ	$\mathcal{G}$	0.500	0.250	0.185	0.085	0.276	0.466	0.237	0.699
g^*	$\mathcal{B}$	0.150	0.050	0.224	0.199	0.253	0.157	0.095	0.217

Notes: See Section 2 for a definition of the DSGE model's parameters, and Section 4 for a description of the data. $\mathcal{B}$ is Beta-distribution, $\mathcal{G}$ is Gamma-distribution, $\mathcal{N}$ is Normal-distribution. Results are based on the sample period QII:1974 - QI:2004.

Table 3: DSGE Model's Parameter Estimates (Part II)

	Prior			DSGE Post.			DSGE-VECM($\hat{\lambda}$) Post.		
	Distr.	Mean	Stdd	Mean	Interval		Mean	Interval	
ρ_z	$\mathcal{B}$	0.200	0.050	0.218	0.146	0.294	0.210	0.134	0.283
ρ_ϕ	$\mathcal{B}$	0.800	0.050	0.705	0.625	0.791	0.856	0.766	0.951
ρ_{λ_f}	$\mathcal{B}$	0.800	0.050	0.518	0.449	0.589	0.690	0.471	0.895
ρ_μ	$\mathcal{B}$	0.800	0.050	0.884	0.834	0.937	0.706	0.608	0.797
ρ_b	$\mathcal{B}$	0.800	0.050	0.811	0.743	0.876	0.762	0.696	0.841
ρ_g	$\mathcal{B}$	0.800	0.050	0.951	0.928	0.975	0.900	0.846	0.955
σ_z	$\mathcal{IG}$	0.400	2.000	0.702	0.625	0.779	0.475	0.405	0.544
σ_ϕ	$\mathcal{IG}$	1.000	2.000	3.450	1.990	4.886	1.121	0.867	1.369
σ_{λ_f}	$\mathcal{IG}$	1.000	2.000	0.192	0.168	0.217	0.191	0.163	0.219
σ_μ	$\mathcal{IG}$	1.000	2.000	0.918	0.742	1.077	0.725	0.597	0.844
σ_b	$\mathcal{IG}$	0.200	2.000	0.538	0.439	0.630	0.302	0.231	0.370
σ_g	$\mathcal{IG}$	0.300	2.000	0.406	0.360	0.454	0.284	0.241	0.328
σ_r	$\mathcal{IG}$	0.200	2.000	0.271	0.242	0.300	0.169	0.143	0.197

Notes: See Section 2 for a definition of the DSGE model's parameters, and Section 4 for a description of the data. $\mathcal{B}$ is Beta-distribution, $\mathcal{IG}$ is Inverse-Gamma-distribution. The Inverse Gamma priors are of the form $p(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. We report s in the Mean-column and ν in the Stdd-column of the table. Results are based on the sample period QII:1974 - QI:2004.

Table 4: Relaxing DSGE Model Restrictions: Log Marginal Likelihoods

Weight on DSGE Restrictions	DSGE-VECM		DSGE-VAR	
λ	Log-Marginal Likelihood	(Difference wrt DSGE-VECM($\hat{\lambda}$))	Log-Marginal Likelihood	(Difference wrt DSGE-VAR($\hat{\lambda}$))
DSGE	-530.48	(-87.10)	-530.48	(-101.70)
∞	-529.00	(-85.62)	-501.56	(-72.78)
5	-488.65	(-45.27)	-481.04	(-52.26)
2	-461.13	(-17.75)	-451.23	(-22.45)
1.5	-454.14	(-10.76)	-444.20	(-15.42)
1.25	-450.50	(-7.12)	-437.62	(-8.84)
1	-443.38	(0)	-434.89	(-6.11)
0.75	-443.45	(-0.07)	-428.78	(0)
0.5	-456.41	(-13.03)	-436.07	(-7.29)
0.33	-506.41	(-63.03)	-473.32	(-44.54)

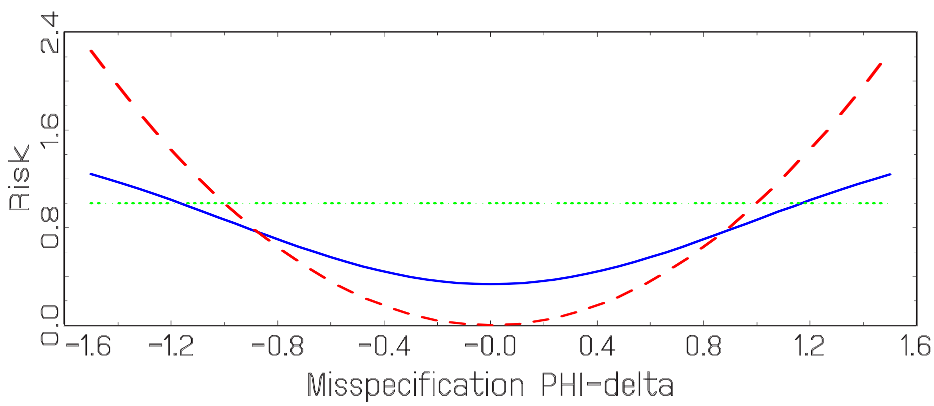
Notes: Column 1 shows the weight of the DSGE model prior λ. Columns 2 and 4 show the logarithm of the marginal likelihood for DSGE-VAR(λ) and DSGE-VECM(λ), respectively. Columns 3 and 5 show in parenthesis the difference between the logarithms of the marginal likelihood of DSGE-VAR(λ) and DSGE-VAR($\hat{\lambda}$), and of DSGE-VECM(λ) and DSGE-VECM($\hat{\lambda}$), respectively, where $\hat{\lambda}$ is the value of λ that maximizes the marginal likelihood. See Section 4 for a description of the data. Results are based on the sample period QII:1974 - QI:2004.

Table 5: Pseudo-Out-of-Sample Root Mean Squared Errors

Variable	Model	Forecast Horizon					
		1	2	4	6	8	12
Output	DSGE-VECM($\hat{\lambda}$)	0.577	0.909	1.753	2.505	3.141	3.888
	DSGE	(19.6)	(36.0)	(51.5)	(58.8)	(63.1)	(69.5)
	VECM	(21.3)	(24.4)	(25.0)	(21.5)	(17.0)	(12.7)
Consumption	DSGE-VECM($\hat{\lambda}$)	0.498	0.767	1.375	1.959	2.450	3.226
	DSGE	(22.9)	(33.9)	(42.1)	(46.1)	(49.4)	(54.7)
	VECM	(5.5)	(9.3)	(18.2)	(19.9)	(21.4)	(19.2)
Investment	DSGE-VECM($\hat{\lambda}$)	3.160	4.955	9.205	13.112	16.520	20.250
	DSGE	(29.3)	(41.0)	(53.1)	(59.7)	(63.7)	(69.9)
	VECM	(13.2)	(12.3)	(9.5)	(4.5)	(-2.1)	(-11.8)
Hours	DSGE-VECM($\hat{\lambda}$)	0.005	0.009	0.019	0.029	0.038	0.050
	DSGE	(16.3)	(29.9)	(43.7)	(48.5)	(50.2)	(52.5)
	VECM	(19.5)	(21.1)	(16.4)	(13.6)	(9.4)	(-0.2)
Real Wages	DSGE-VECM($\hat{\lambda}$)	0.611	1.022	1.875	2.563	3.162	4.256
	DSGE	(1.9)	(-0.3)	(-3.9)	(-8.0)	(-12.0)	(-13.2)
	VECM	(7.5)	(6.4)	(2.5)	(1.5)	(3.6)	(8.8)
Inflation	DSGE-VECM($\hat{\lambda}$)	0.233	0.450	0.833	1.305	1.803	2.820
	DSGE	(2.4)	(8.6)	(15.5)	(14.1)	(14.0)	(14.5)
	VECM	(5.2)	(5.6)	(9.0)	(13.4)	(15.3)	(16.5)
Fed Funds Rate	DSGE-VECM($\hat{\lambda}$)	0.465	0.780	1.288	1.712	2.180	2.596
	DSGE	(13.1)	(22.4)	(27.4)	(26.5)	(21.2)	(19.4)
	VECM	(28.3)	(28.8)	(29.2)	(31.4)	(30.1)	(29.1)
Multivariate	DSGE-VECM($\hat{\lambda}$)	1.368	0.939	0.523	0.281	0.117	-0.101
Statistic	DSGE	(13.8)	(18.2)	(21.7)	(20.7)	(17.8)	(5.0)
	VECM	(15.0)	(16.3)	(16.0)	(19.7)	(20.4)	(26.5)

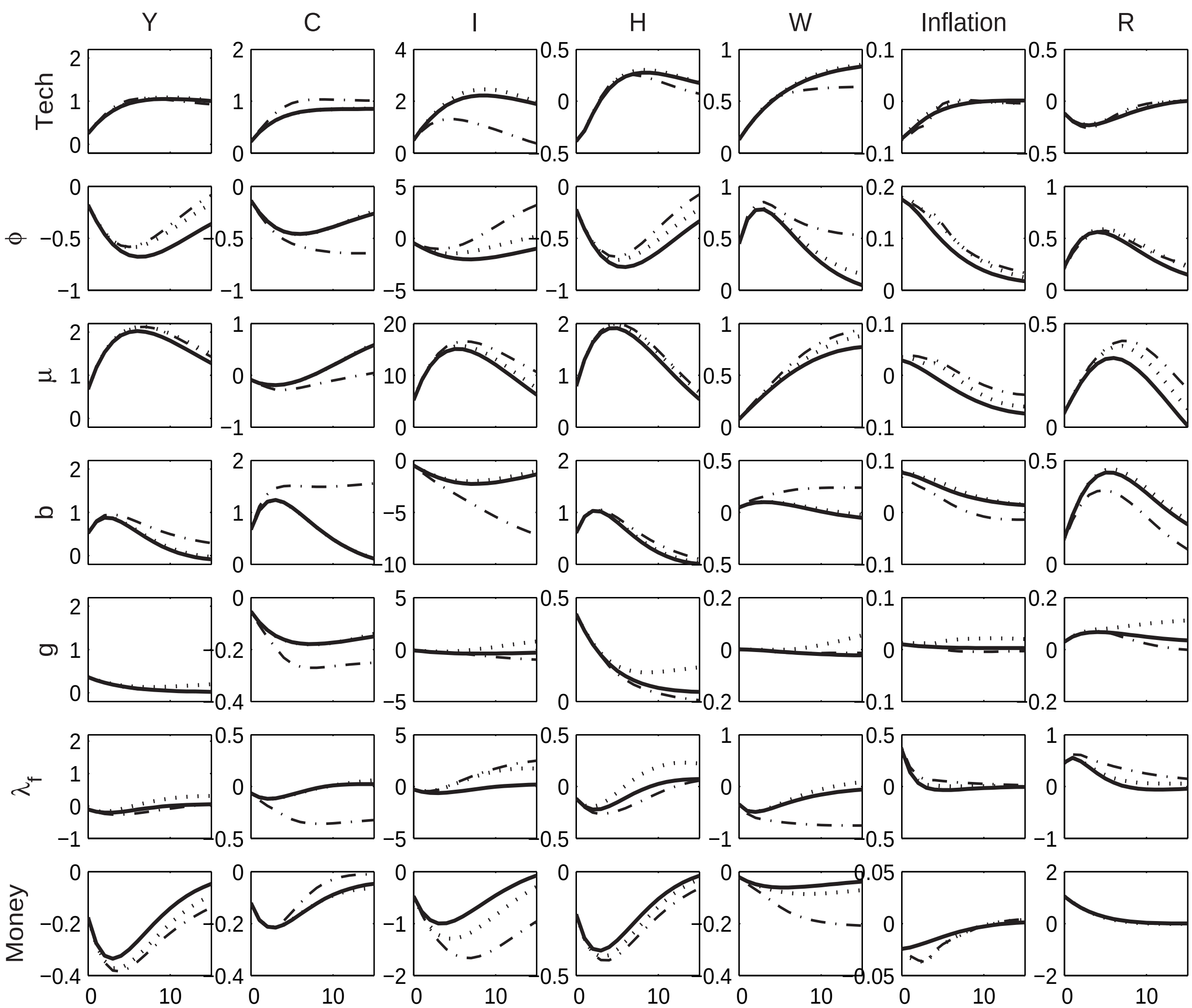
Notes: Results are based on the rolling sample QIV:1985 - QI:2000. At each date of the rolling sample we use the previous 120 observations to estimate the model, and the following twelve quarters to assess forecasting accuracy. For each date we also compute $\hat{\lambda}$, the value of λ that maximizes the marginal likelihood. For each variable, the table shows the root mean squared error (RMSE) of the forecast from DSGE-VECM($\hat{\lambda}$), and in parenthesis the improvement in forecast accuracy relative to the DSGE model and the unrestricted VECM, as measured by the percentage reduction (increase, if negative) in RMSE. The multivariate statistic is computed as the converse of the log-determinant of the variance-covariance matrix of forecast errors. The forecast horizon is measured in quarters. See Section 4 for a description of the data.

Figure 1: Estimation Risk: AR(1) Example



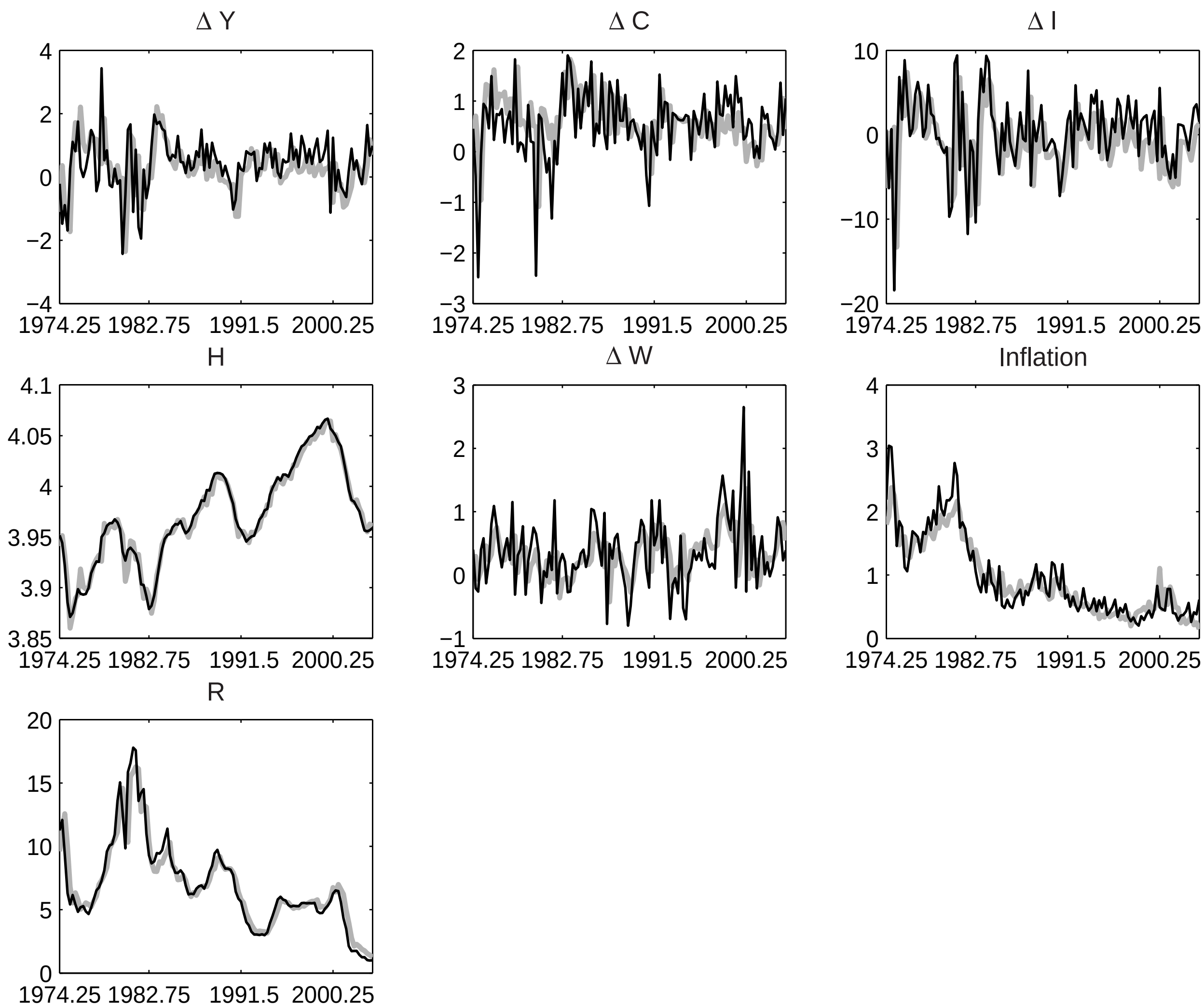
Notes: Figure depicts asymptotic risks as a function of local misspecification: *solid* is $\lambda = \hat{\lambda}$, dashed is $\lambda = \infty$, and dotted is $\lambda = T^{-1}$.

Figure 2: How Well Do DSGE-VAR/VECM Approximate the DSGE Model? Identified Impulse Responses



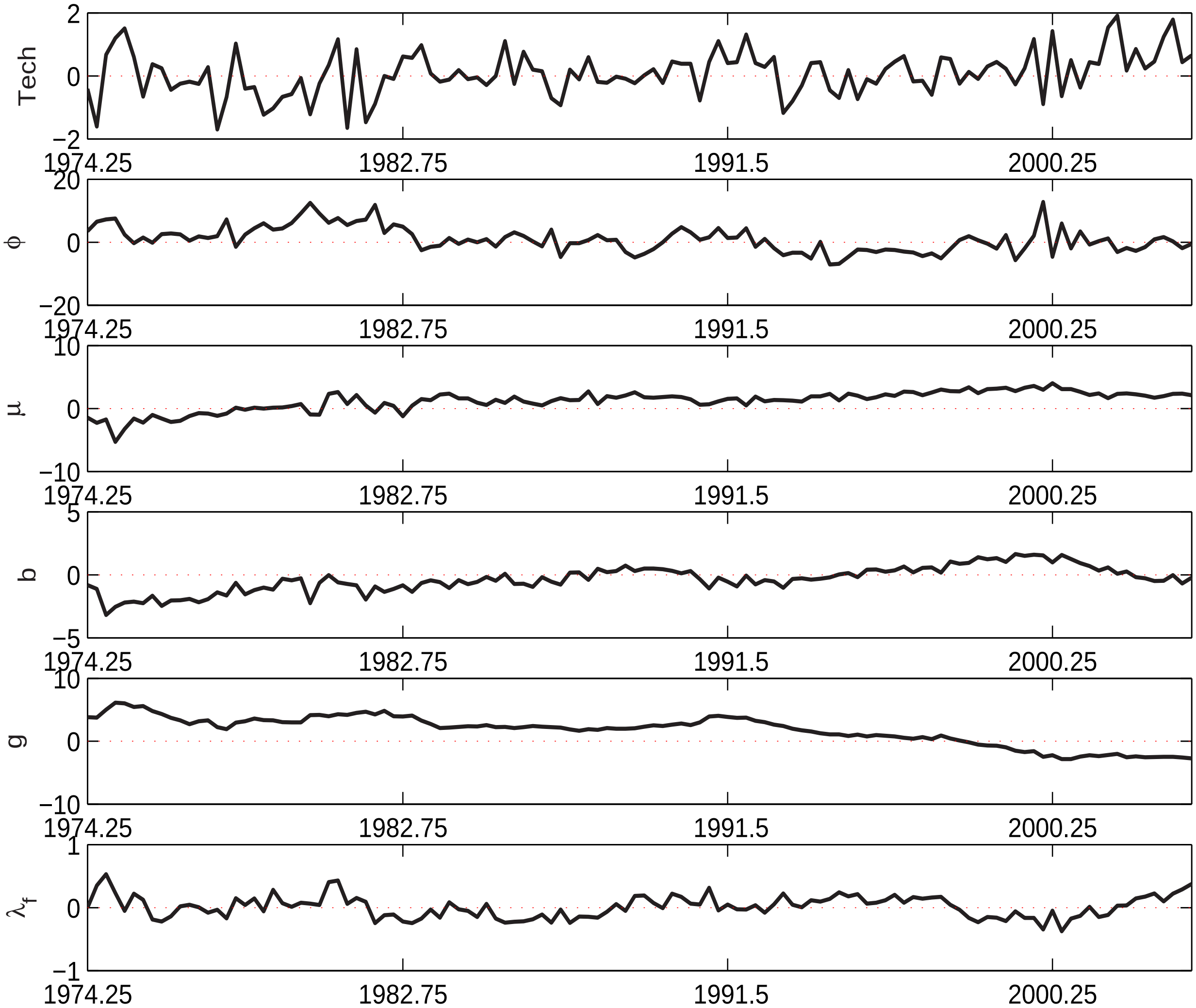
Notes: The solid, dash-and-dotted, and dotted lines represent the impulse responses from one to sixteen quarters ahead of the DSGE model, DSGE-VAR(∞), and DSGE-VECM(∞), respectively, with respect to the following shocks: Tech (technology), φ_t (preferences of leisure), μ_t (shocks to the capital accumulation equation), b_t (overall preference shifter), g_t (government spending), and $\lambda_{f,t}$ (price markup shocks), and Money (monetary policy). All impulse responses are computed setting the vector of DSGE model parameters θ at the posterior mean values reported in column 5 of Tables 2 and 3. These impulse responses for DSGE-VAR(∞) and DSGE-VECM(∞) are obtained using the identification procedure described in the section 3.5. All impulse responses are in percent. The impulse responses to output, consumption, investment, and the real wage are cumulative. Results are based on the sample period QII:1974 - QI:2004. See Section 4 for a description of the data.

Figure 3: In-Sample Fit of the DSGE Model



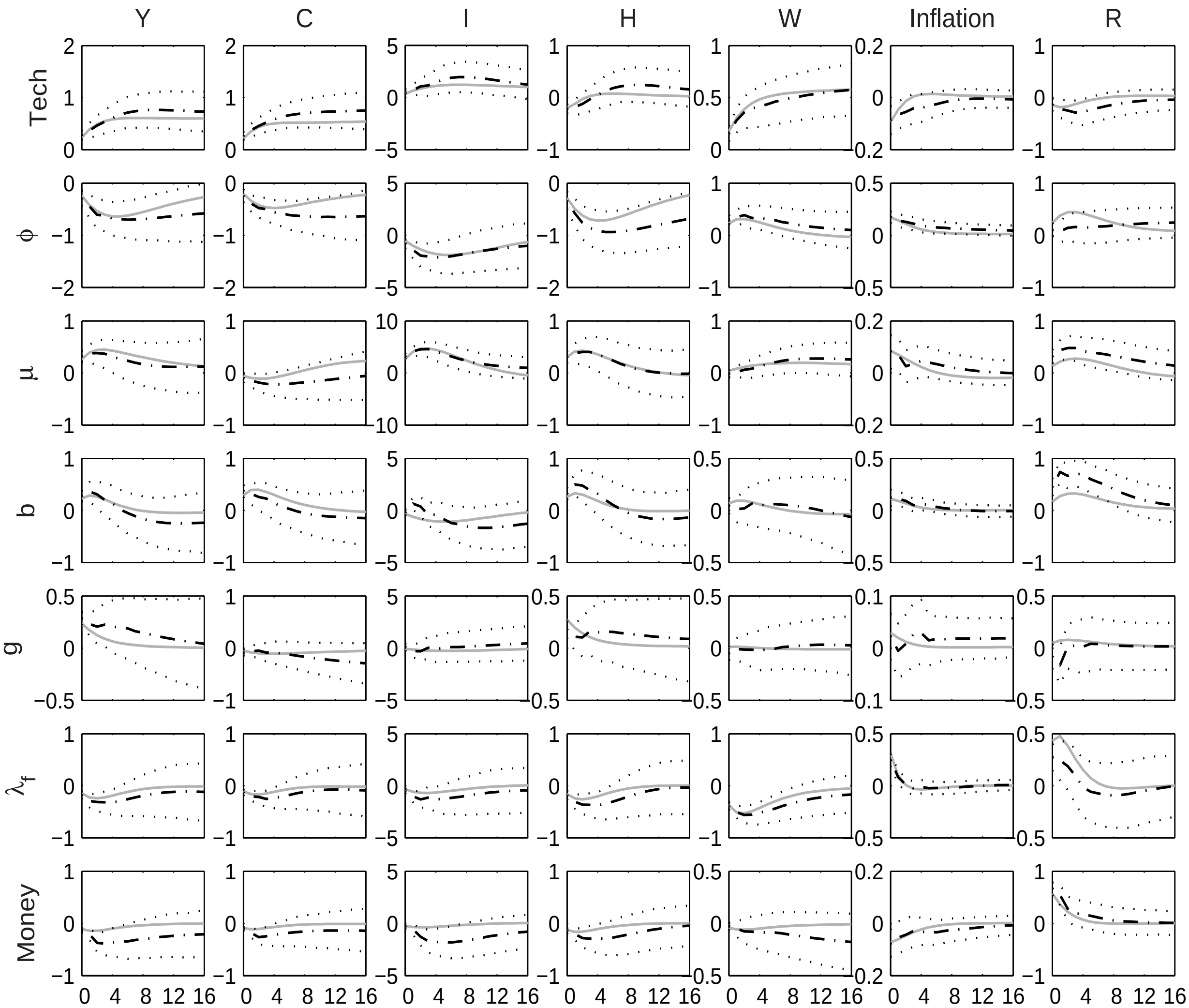
Notes: The figure plots the actual data (dark line), as well as the one-period-ahead forecasts obtained from the Kalman filter (gray line), computed using the vector θ of DGSE model parameters that maximizes the posterior. Results are based on the sample period QII:1974 - QI:2004. See Section 4 for a description of the data.

Figure 4: Exogenous Processes



Notes: The figure plots the Kalman-smoothed time series for the processes z_t (technology), φ_t (preferences of leisure), μ_t (shocks to the capital accumulation equation), b_t (overall preference shifter), g_t (government spending), and $\lambda_{f,t}$ (price markup shocks), computed using the vector θ of DGSE model parameters that maximizes the posterior. Results are based on the sample period QII:1974 - QI:2004. See Section 4 for a description of the data.

Figure 5: Impulse-Responses: DSGE-VECM($\hat{\lambda}$) versus the DSGE model



Notes: The black lines represent the mean impulse-responses (dash-and-dotted lines) of DSGE-VECM($\lambda = 1$) and the associated 90% bands (dotted lines). The gray lines represent the mean impulse-responses (solid lines) of the DSGE model and the associated 90% bands (dotted lines). The impulse-responses are computed with respect to the following shocks: Tech (technology), φ_t (preferences of leisure), μ_t (shocks to the capital accumulation equation), b_t (overall preference shifter), g_t (government spending), and $\lambda_{f,t}$ (price markup shocks), and Money (monetary policy). All impulse responses are in percent. The impulse responses to output, consumption, investment, and the real wage are cumulative. Results are based on the sample period QII:1974 - QI:2004. See Section 4 for a description of the data.