A DSGE-VAR for the Euro Area

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

December, 2003

Abstract

This paper uses a modified version of the DSGE model estimated in Smets and Wouters (2003) to generate a prior distribution for a vector autoregression, following the approach in Del Negro and Schorfheide (2003). This DSGE-VAR is fitted to Euro area data on GDP, consumption, investment, nominal wages, hours worked, inflation, M2, and a short-term interest rate. We document the fit of the DSGE-VAR.

Preliminary and Incomplete

JEL CLASSIFICATION: C11, C32, C53

KEY WORDS: Bayesian Analysis, DSGE Models, Forecasting, 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*. The views expressed in this papers are solely our own and do not necessarily reflect those of the Federal Reserve Bank of Atlanta, 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, they are emerging as s useful tool for empirical research in macroeconomics, forecasting, and quantitative policy analysis.

- The simple models that have been used in the past have a poor forecasting record. For instance, Schorfheide (2000) computes posterior probabilities, which can be interpreted as measures of one-step-ahead predictive performance, of cash-in-advance models and a vector autoregression (VAR). He finds that output growth and inflation data strongly favor the VAR.

- More recently, Smets and Wouters (2003a), henceforth SW, have developed an elaborate DSGE model with capital accumulation as well as various nominal and real frictions. They evaluate forecasting performance and the posterior odds of their model versus a VAR based on (detrended) Euro-area data and U.S. data (Smets and Wouters, 2003c). While the Euro-area data prefer the VAR the U.S. data lead to posterior probabilities that strongly favor the DSGE model.

- Del Negro and Schorfheide (2003), henceforth DS, show that forecasts with a simple three equation New Keynesian DSGE model can be improved by systematically relaxing the DSGE model restrictions. In their framework the DSGE model is used to generate a prior distribution for the coefficients of a VAR. The prior concentrates most of its probability mass near the restrictions that the DSGE model imposes on the VAR representation and pulls the likelihood estimate of the VAR parameters toward the DSGE model restrictions, without dogmatically imposing them. DS document that the resulting specification, which we will label as DSGE-VAR, outperforms both the DSGE model itself as well as a VAR in terms of forecasting performance.

- In this paper we use a variant of the SW model to generate a prior distribution for an eight-variable VAR in output, consumption, investment, hours worked, nominal wages, money stock, prices, and interest rates. We consider various specifications of the VAR: in differences, in vector error correction form, and in levels, extending the DS framework to non-stationary endogenous variables. Our approach is particularly attractive for Euro area applications in which the the sample size is fairly small compared to the dimensionality of the autoregressive model that is being estimated. The Bayesian estimation procedure can be interpreted as augmenting the sample of actual observations with artificial data generated from the DSGE model.

- Sims (2003) points out that the posterior probabilities computed by SW tend to switch between the extremes zero and one, depending on the choice of the data set (U.S. versus Euro-area) and the specification of the VAR prior (Minnesota prior versus training sample prior). In his view these probabilities do not give an accurate reflection of model uncertainty and are largely an artifact of a model space that is too sparse. Following arguments in Gelman, Carlin, Stern, and Rubin (1994), Sims advocates filling the model space by connecting distinct model specifications with continuous parameters and characterize the model uncertainty through the posterior probability distribution of these additional parameters. This posterior will be less sensitive to the choice of prior distribution than the posterior odds for the original models.

- The DS framework can be viewed as an attempt to connect VAR and DSGE model using a continuous hyperparameter. This hyperparameter controls the variance and therefore the weight of the DSGE model prior relative to the sample. For extreme values of this parameter (zero or infinity) either an unrestricted VAR or the DSGE model is estimated. Allowing for intermediate values of the hyperparameter is similar in spirit to Sims' (2003) notion of completing the model space with one caveat: we do not have a strict structural interpretation of the specifications that are being estimated with intermediate values of this parameter. However, we are able to construct DSGE model-based identification schemes and carry out a structural VAR analysis.

- Empirical findings.

The paper is organized as follows. The DSGE model is presented in Section 2. Section 3 reviews the DS approach of generating a prior distribution from the DSGE model for a VAR. We discuss extensions to vector autoregressive models with non-stationary endogenous variables. Empirical results are presented in Section 4 and Section 5 concludes.

2 Model

To generate a prior distribution for the coefficients of a vector autoregression we use 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, largely based on the work of Christiano, Eichenbaum, and Evans (2003), contains a large number of nominal and real frictions and various structural shocks. To make this paper self-contained we provide a brief description of the model.

2.1 Final and Intermediate Goods' Production

The final good Y_t produced in the model economy is a composite made of a continuum of intermediate goods i:

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

The final goods producers buy the intermediate goods on the market, package Y_t, and resell it to the households. The parameter λ_f in the aggregation function is time varying. The final goods firms maximize profits in a perfectly competitive environment:

$$\begin{array}{ll} \max_{Y_t, Y_t(i)} P_t Y_t - \int_0^1 P_t(i) Y_t(i) di & \\ \text{s.t. } Y_t = \left[\int_0^1 Y_t(i)^{\frac{1}{1+\lambda_{f,t}}} di \right]^{1+\lambda_{f,t}} & (\mu_{f,t}) \end{array} \tag{1}$$

Combining the first-order conditions of this maximization problem with the zero profit condition one obtains the price of the composite good:

$$P_t = \left[\int_0^1 P_t(i)^{\frac{1}{\lambda_{f,t}}} di \right]^{\lambda_{f,t}}. \tag{2}$$

Intermediate goods producers i uses the following technology:

$$Y_t(i) = \max \{ K_t(i)^\alpha (Z_t L_t(i))^{1-\alpha} - Z_t^* \Phi, 0 \}, \tag{3}$$

where Z_t is a labor-augmenting stochastic trend and Φ is a fixed cost. We denote the growth rate of technology by $z_t = \log(Z_t/Z_{t-1})$ and assume that z_t follows an autoregressive process with mean γ:

$$(z_t - \gamma) = \rho_z (z_{t-1} - \gamma) + \epsilon_{z,t}, \tag{4}$$

It will turn out that the overall growth rate of the economy is

$$Z_t^* = Z_t \Upsilon^{\left(\frac{\alpha}{1-\alpha} t\right)}, \quad \Upsilon > 1$$

because we will assume subsequently that investment goods are becoming more efficient over time. The intermediate goods' firms period t nominal profits are given by

$$P_t(i) Y_t(i) - W_t L_t(i) - R_t^k K_t(i), \tag{5}$$

where W_t is the nominal wage and R_t^k is the rental cost of capital. Profit maximization implies that the capital-labor ratio is proportional to the ratio of the factor prices:

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

Total variable costs are given by

$$VC = \left(W_t + R_t^k \frac{K_t(i)}{L_t(i)}\right) L_t(i) = \left(W_t + R_t^k \frac{K_t(i)}{L_t(i)}\right) \tilde{Y}_t(i) Z_t^{-(1-\alpha)} \left(\frac{K_t(i)}{L_t(i)}\right)^{-\alpha}, \quad (6)$$

where $\tilde{Y}_t(i) = K_t(i)^\alpha (Z_t L_t(i))^{1-\alpha}$ is the variable component of output. Hence, the marginal cost MC_t is the same for all firms and equal to:

$$MC_t = \left(W_t + R_t^k \frac{K_t(i)}{L_t(i)}\right) Z_t^{-(1-\alpha)} \left(\frac{K_t(i)}{L_t(i)}\right)^{-\alpha}. \quad (7)$$

Profits can then be expressed as $[P_t(i) - MC_t]Y_t(i) - MC_t Z_t^* \Phi$. Note that fixed cost component does not depend on the firms' decisions, it can be safely ignored. As in Calvo (1983), we assume that each firm can readjust prices with probability $1-\zeta_p$ in each period. Those firms that cannot adjust their price $P_t(i)$ will increase it at the steady state rate of inflation π^*. For those firms that can adjust prices, the problem is to choose a price level $\tilde{P}_t(i)$ that maximizes the expected present discounted value of profits in all states where the firm is stuck with that price in the future:

$$\max_{\tilde{P}_t(i)} \quad I\!E_t \sum_{s=0}^{\infty} \left[\zeta_p^s \beta^s Q_{t+s} \left(\tilde{P}_t(i)\pi^{*s} - MC_{t+s}\right) Y_{t+s}(i)\right] \quad (8)$$

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

where Q_{t+s} is today's value of a future dollar for the consumers in a particular state of nature. Under the assumption that households have access to a complete set of state-contingent claims $Q_t = \Xi_t^p$ in equilibrium, where Ξ_t^p is the Lagrange multiplier associated with the consumer's nominal budget constraint. The first-order condition for intermediate firm i is:

$$I\!E_t \sum_{s=0}^{\infty} \zeta_p^s \beta^s Q_{t+s} \left(\tilde{P}_t(i)\pi^{*s} - (1+\lambda_{f,t+s})MC_{t+s}\right) Y_{t+s}(i) = 0 \quad (9)$$

Since expected marginal costs are the same across firms, all firms that can readjust prices will choose the same $\tilde{P}_t(i)$, so we can drop the i index from now on. From 2 it follows that the aggregate price level evolves according to:

$$P_t = [(1-\zeta_p)\tilde{P}_t^{\frac{1}{\lambda_{f,t}}} + \zeta_p(\pi^* P_{t-1})^{\frac{1}{\lambda_{f,t}}}]^{\lambda_{f,t}}. \quad (10)$$

2.2 The Household Sector

There is a continuum of households j in the economy. Households differ in that they supply a differentiated type of labor. Each household maximizes an intertemporal utility function:

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s b_{t+s} \left[\ln(C_{t+s}(j) - hC_{t+s-1}(j)) - \frac{\varphi_{t+s}}{\nu+1} L_{t+s}(j)^{\nu+1} + \chi_{t+s} \log\left(\frac{M_{t+s}(j)}{P_{t+s}}\right) \right] \quad (11)$$

where $C_t(j)$ is consumption, $L_t(j)$ is labor supply (total available hours are normalized to one), and $M_t(j)$ are money holdings. Consumption C_{t+s} enters the utility function relative to the habit stock hC_{t+s-1}. ν represents the inverse of the elasticity of work effort with respect to the real wage. Equation (11) contains three time-varying preference parameters: b_t affects the intertemporal substitution of households, φ_t represents a labor supply shift, and χ_t captures fluctuations of the preference for money holdings.

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

$$\begin{aligned} & P_{t+s}C_{t+s}(j) + P_{t+s}I_{t+s}(j) + B_t(j) + M_{t+s}(j) + T_{t+s}(j) \\ \leq \; & R_{t+s}B_{t+s-1}(j) + M_{t+s-1}(j) + \Pi_{t+s} + W_{t+s}(j)L_{t+s}(j) \\ & + \left(R^k_{t+s}u_{t+s}(j)\bar{K}_{t+s-1}(j) - P_{t+s}a(u_{t+s}(j))\Upsilon^{-t}\bar{K}_{t+s-1}(j)\right), \end{aligned} \quad (12)$$

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, Π_t is the per-capita profit the household gets from owning firms (assume household pool their firm shares, so that they all receive the same profit), $W_t(j)$ is the 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, $u_t(j)$, and end up renting to firms in period t an amount of "effective" capital equal to:

$$K_t(j) = u_t(j)\bar{K}_{t-1}(j), \quad (13)$$

and getting $R^k_t u_t(j)\bar{K}_{t-1}(j)$ in return. However, households have to pay a cost of utilization in terms of the consumption good which is equal to $a(u_t(j))\Upsilon^{-t}\bar{K}_{t-1}(j)$. Households accumulate capital according to the equation:

$$\bar{K}_t(j) = (1-\delta)\bar{K}_{t-1}(j) + \Upsilon^t \mu^\Upsilon_t \left(1 - S(\frac{I_t(j)}{I_{t-1}(j)})\right) I_t(j), \quad (14)$$

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^Υ_t is a stochastic disturbance to the price of investment relative to consumption. Due to the trend Υ^t investment becomes more efficient over time, in the sense that the same amount of investment goods leads to a larger increase in the effective capital stock as time progresses.

Let $\Xi_t^p(j)$ be the Lagrange multiplier associated with the budget constraint 12. We assume there is a complete set of state contingent securities in nominal terms, although we do not explicitly write them in the household's budget constraint. This assumption implies that $\Xi_t^p(j)$ must be the same for all households in all periods and across all states of nature: $\Xi_t^p(j) = \Xi_t^p$ for all j and t. Although we so far kept the j index for all the appropriate variables, it turns out that the assumption of complete markets lets us drop the index. In equilibrium households will make the same choice of consumption, money demand, investment and capital utilization.

2.3 Labor Market

Labor used by the intermediate goods producers L_t is a composite:

$$L_t = \left[\int_0^1 L_t(j)^{\frac{1}{1+\lambda_w}} di\right]^{1+\lambda_w}. \tag{15}$$

There are labor packers who buy the labor from the households, package L_t, and resell it to the intermediate goods producers. Labor packers maximize profits in a perfectly competitive environment. From the first-order conditions of the labor packers one obtains:

$$L_t(j) = \left(\frac{W_t(j)}{W_t}\right)^{-\frac{1+\lambda_w}{\lambda_w}} L_t. \tag{16}$$

Combining this condition with the zero profit condition one obtains an expression for the wage:

$$W_t = \left[\int_0^1 W_t(j)^{\frac{1}{\lambda_w}} di\right]^{\lambda_w}, \tag{17}$$

where λ_w is a parameter. Given the structure of the labor market, the household has market power: she can choose her wage subject to 16. However, she is also subject to Calvo-type nominal rigidities. Households can readjust wages with probability $1 - \zeta_w$ in each period. For those that cannot adjust wages, $W_t(j)$ will increase at the steady state rate of inflation π_* multiplied by the growth rate of the economy $e^\gamma \Upsilon^{\frac{\alpha}{1-\alpha}}$. For those that can adjust, the problem is to choose a wage $\tilde{W}_t(j)$ that maximizes utility in all states of nature where the household is stuck with that wage in the future:

$$\begin{aligned} \max_{\tilde{W}_t(j)} \quad & \mathbb{E}_t \sum_{s=0}^{\infty} (\zeta_w \beta)^s b_{t+s} \left[\dots - \frac{\varphi_{t+s}}{\nu+1} L_{t+s}(j)^{\nu+1} + \dots\right] \\ \text{s.t.} \quad & 12 \text{ for } s = 0, \dots, \infty, \text{ 16, and} \\ & W_{t+s}(j) = (\pi_* e^\gamma \Upsilon^{\alpha/(1-\alpha)})^s \tilde{W}_t(j) \end{aligned} \tag{18}$$

where the ... indicate the terms in the utility function that are irrelevant for this problem. The first-order conditions for this problem are:

$$
\begin{aligned}
0 &= I\!\!E_t \sum_{s=0}^{\infty} (\zeta_w \beta)^s \Xi_{t+s} \Bigg[- \frac{(\pi_* e^{\gamma} \Upsilon^{\alpha/(1-\alpha)})^s \tilde{W}_t(j)}{P_{t+s}} \\
&\quad + (1+\lambda_w) \frac{b_{t+s} \varphi_{t+s} L^{\nu}_{t+s}(j)}{\Xi_{t+s}} \Bigg] L_{t+s}(j).
\end{aligned} \tag{19}
$$

In absence of nominal rigidities this condition would amount to setting the real wage equal to ratio of the marginal utility of leisure over the marginal utility of consumption times the markup $(1+\lambda_w)$. Under the complete market assumption $\tilde{W}_t(j) = \tilde{W}_t$, all j. Then from (17) it follows that:

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

2.4 Completing the Model

The market clearing condition for the final goods market is

$$
Y_t = C_t + I_t + G_t, \tag{21}
$$

where G_t is exogenous government spending because it has to relate C_t, I_t, G_t to aggregate capital and employment. We define the exogenous process g_t such that government spending can be expressed as a fraction of output:

$$
G_t = (1 - 1/g_t) Y_t. \tag{22}
$$

The government adjusts the nominal lump-sum taxes (or subsidies) T_t to ensure that its budget constraint is satisfied in every period:

$$
P_t G_t + R_{t-1} B_{t-1} + M_{t-1} = T_t + M_t + B_t. \tag{23}
$$

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:

$$
\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*} \right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi^*} \right)^{\psi_1} \left(\frac{Y_t}{Y_t^*} \right)^{\psi_2} \right]^{1-\rho_R} \tag{24}
$$

where R^* is the steady state nominal rate and Y_t^* is nominal output. The parameter ρ_R determines the degree of interest rate smoothing. The central bank supplies the money demanded by the household to support the desired nominal interest rate. The laws of motion for the exogenous processes are summarized in Table 1.

2.5 Model Solution

As in Altig, Christiano, Eichenbaum, and Linde (2002) our model economy evolves along stochastic growth path. Output Y_t, consumption C_t, and investment I_t grow at the rate Z_t^*. Physical capital K_t and effective capital $\bar{K}_t$ both grow at the rate $\Upsilon^t Z_t^*$. Money stock M_t and nominal wages W_t grow proportionally to $P_t Z_t^*$, whereas the growth rate of the nominal rental costs of capital is $P_t \Upsilon^{-t}$. 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 and derive a state-space representation for

$$\Delta y_t = [\Delta \ln Y_t, \Delta \ln C_t, \Delta \ln I_t, \Delta \ln W_t, \ln L_t, \pi_t, \Delta \ln M_t, R_t]',$$

where Δ denotes the temporal difference operator and π_t is the inflation rate. From the state-space representation we construct the VAR prior.

3 DSGE Model Priors

A less restrictive moving-average representation for the $n \times 1$ vector y_t than the one implied by the DSGE model of the previous section can be obtained from a vector autoregressive model:

$$\Delta y_t = \Phi_0 + \Phi_1 \Delta y_{t-1} + \ldots + \Phi_p \Delta y_{t-p} + u_t, \tag{25}$$

where u_t is a vector of one-step-ahead forecast errors. VARs have a long tradition in applied macroeconomics as a tool for forecasting, policy analysis, and business cycle analysis. One drawback of VARs is that they are not very parsimonious: in many applications, data availability poses a serious constraint on the number of endogenous variables and the number of lags that can effectively be included in a VAR without overfitting the data. A solution to this problem of too many parameters is to use a prior distribution that essentially adds information to the estimation problem. This prior distribution will be obtained from the DSGE model presented in the previous section. We use the method developed in Del Negro and Schorfheide (2003). Subsequently, we sketch the main ideas of the procedure. Loosely speaking, our prior adds artificial observations from the DSGE model to the actual data set and leads to an estimation of the VAR based on a mixed sample of artificial and actual observations.

3.1 Baseline Specification

Our baseline specification is the VAR in differences given in Equation (25). We assume that the innovations u_t have a multivariate normal distribution $\mathcal{N}(0, \Sigma_u)$ conditional on past observations of Δy_t. Let Y be the $T \times n$ matrix with rows $\Delta y_t'$. Let $k = 1 + np$, X be the $T \times k$ matrix with rows $x_t' = [1, \Delta y_{t-1}', \ldots, \Delta y_{t-p}']$, U be the $T \times n$ matrix with rows u_t', and $\Phi = [\Phi_0, \Phi_1, \ldots, \Phi_p]'$. The VAR can be expressed as $Y = X\Phi + U$ with likelihood function

$$\begin{aligned} & p(Y|\Phi, \Sigma_u) \propto \\ & \quad |\Sigma_u|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma_u^{-1}(Y'Y - \Phi'X'Y - Y'X\Phi + \Phi'X'X\Phi)] \right\} \end{aligned} \tag{26}$$

conditional on observations $\Delta y_{1-p}, \ldots, \Delta y_0$. Although the DSGE model presented in Section 2 does not have a finite-order vector autoregressive representation in terms of Δy_t, the VAR can be interpreted as an approximation to the moving-average representation of the DSGE model. The magnitude of the discrepancy becomes smaller the more lags are included in the VAR. Since θ is of much lower dimension than the VAR parameter vector, the DSGE model imposes a restrictions on the (approximate) vector autoregressive representation of Δy_t.

According to our DSGE model, the vector of endogenous variables Δy_t is covariance stationary and the expected values of sample moments of artificial data $\sum \Delta y_t^* \Delta y_t^{*'}$, $\sum \Delta y_t^* x_t^{*'}$, and $\sum x_t^* x_t^{*'}$ are given by the (scaled) population moments $\lambda T \Gamma_{dydy}^*(\theta)$, $\lambda T \Gamma_{dyx}^*(\theta)$, and $\lambda T \Gamma_{xx}^*(\theta)$, where, for instance, $\Gamma_{dydy}^*(\theta) = I\!E_\theta[\Delta y_t \Delta y_t']$. These expected values depends on the DSGE model parameter θ. Define the functions

$$\Phi^*(\theta) = \Gamma_{xx}^{*^{-1}}(\theta)\Gamma_{xdy}^*(\theta) \tag{27}$$

$$\Sigma_u^*(\theta) = \Gamma_{dydy}^*(\theta) - \Gamma_{dyx}^*(\theta)\Gamma_{xx}^{*^{-1}}(\theta)\Gamma_{xdy}^*(\theta). \tag{28}$$

The functions $\Phi^*(\theta)$ and $\Sigma_u^*(\theta)$ trace out a subspace of the VAR parameter space and can be interpreted as follows. Suppose that data are generated from a DSGE model with parameters θ. Among the p'th order VARs the one with the coefficient matrix $\Phi^*(\theta)$ minimizes the one-step-ahead quadratic forecast error loss. The corresponding forecast error covariance matrix is given by $\Sigma_u^*(\theta)$.

Conditional on θ our prior distribution of the VAR parameters is of the Inverted-Wishart ($\mathcal{IW}$) – Normal ($\mathcal{N}$) form. It belongs to the same family of probability distributions as the posterior characterized in Equations (33) and (34) below. This distribution can be interpreted as follows. It is the posterior distribution of someone who updates the non-informative prior $p(\Phi, \Sigma_u) \propto |\Sigma_u|^{-(n+1)/2}$ with the sample of artificial observations generated from the DSGE model. Provided that $\lambda T \geq k+n$ and $\Gamma_{xx}(\theta)$ is invertible, the resulting probability density is proper (it integrates to one) and non-degenerate (its support is not restricted to a subspace of the VAR parameter space).

The specification of the prior is completed with a distribution of the DSGE model parameters, details of which we discuss in Section 4. Overall our prior has the hierarchical structure

$$p(\Phi, \Sigma_u, \theta) = p(\Phi, \Sigma_u|\theta)p(\theta). \tag{29}$$

The posterior density is proportional to the prior density and the likelihood function. In order to study the posterior distribution we factorize it into the posterior density of the VAR parameters given the DSGE model parameters and the marginal posterior density of the DSGE model parameters:

$$p(\Phi, \Sigma_u, \theta|Y) = p(\Phi, \Sigma_u|Y, \theta)p(\theta|Y). \tag{30}$$

Let $\tilde{\Phi}(\theta)$ and $\tilde{\Sigma}_u(\theta)$ be the maximum-likelihood estimates of Φ and Σ_u, respectively, based on artificial sample and actual sample

$$\tilde{\Phi}(\theta) = (\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T \Gamma^*_{xdy} + X'Y) \tag{31}$$

$$\begin{aligned} \tilde{\Sigma}_u(\theta) = \frac{1}{(\lambda+1)T}\Big[&(\lambda T \Gamma^*_{yy}(\theta) + Y'Y) \\ &-(\lambda T \Gamma^*_{yx}(\theta) + Y'X)(\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}(\lambda T \Gamma^*_{xdy}(\theta) + X'Y)\Big]. \end{aligned} \tag{32}$$

Since conditional on θ the DSGE model prior and the likelihood function are conjugate, 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:

$$\Sigma_u|Y, \theta \sim \mathcal{IW}\Big((\lambda+1)T\tilde{\Sigma}_u(\theta), (1+\lambda)T - k, n\Big) \tag{33}$$

$$\Phi|Y, \Sigma_u, \theta \sim \mathcal{N}\Big(\tilde{\Phi}(\theta), \Sigma_u \otimes (\lambda T \Gamma^*_{xx}(\theta) + X'X)^{-1}\Big). \tag{34}$$

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 (2003). The ability to compute the population moments $\Gamma^*_{dydy}(\theta)$, $\Gamma^*_{xdy}(\theta)$, and $\Gamma^*_{xx}(\theta)$ analytically from the log-linearized solution to the DSGE model and the use of conjugate priors for the VAR parameters makes the approach very efficient from a computational point of view.

The hyperparameter λ determines the effective sample size for the artificial observations, which is λT. If λ is small the prior is diffuse, and the actual observations dominate the artificial observations in the posterior. Not surprisingly, the empirical performance of a VAR with DSGE model prior will crucially depend on the choice of λ. We will choose a grid $\Lambda = \{l_1, \ldots, l_q\}$ for the hyperparameter and assign prior probabilities to these grid points. Based on the marginal data density

$$p_\lambda(Y) = \int p_\lambda(Y|\theta)p(\theta)d\theta \tag{35}$$

We can compute posterior probabilities for the grid points and either average over different the values of λ or condition on the one that has the highest posterior probability. This marginal data density can also be used to determine an appropriate lag length for the VAR.

3.2 VEC Specification

The DSGE model 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^*. The common trend structure suggests to include vector error correction terms in the specification (25) and to consider the following model:

$$\Delta y_t = \Phi_0 + \Phi_\beta(\beta' y_{t-1}) + \Phi_1 \Delta y_{t-1} + \ldots + \Phi_p \Delta y_{t-p} + u_t, \tag{36}$$

According to our DSGE model the error correction terms are

$$\beta' y_{t-1} = \begin{bmatrix} \ln C_t - \ln Y_t \\ \ln I_t - \ln Y_t \\ \ln W_t - \ln Y_t - \ln P_t \\ \ln M_t - \ln Y_t - \ln P_t \end{bmatrix}$$

The elements of the vector $\beta' y_{t-1}$ are stationary according to our model and the prior can be constructed as above. However, we now let $k = 2 + np$, X be the $T \times k$ matrix with rows $x_t' = [1, (\beta' y_{t-1})', \Delta y_{t-1}', \ldots, \Delta y_{t-p}']$, and $\Phi = [\Phi_0, \Phi_\beta, \Phi_1, \ldots, \Phi_p]'$.

3.3 VAR in Levels

As a second alternative to the VAR in differences (25) we consider a VAR in levels. Let

$$y_t = [\ln Y_t, \ln C_t, \ln I_t, \ln W_t, L_t, P_t, \ln M_t, R_t]',$$

and

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_{p+1} y_{t-p-1} + u_t. \tag{37}$$

This level-VAR is consistent with various trend patterns of the endogenous variables. However, the construction of the DSGE model prior is slightly more difficult since most of the endogenous variables are non-stationary and do not have time-invariant moments. To transform the endogenous variables we define the following rotation matrix

$$D_{\lambda T} = \begin{bmatrix} (\lambda T)^{-\frac{3}{2}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -(\lambda T)^{-\frac{1}{2}} & (\lambda T)^{-\frac{1}{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\ -(\lambda T)^{-\frac{1}{2}} & 0 & (\lambda T)^{-\frac{1}{2}} & 0 & 0 & 0 & 0 & 0 \\ -(\lambda T)^{-\frac{1}{2}} & 0 & 0 & (\lambda T)^{-\frac{1}{2}} & 0 & -(\lambda T)^{-\frac{1}{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & -(\lambda T)^{-\frac{1}{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -(\lambda T)^{-\frac{3}{2}} & 0 & 0 \\ -(\lambda T)^{-\frac{1}{2}} & 0 & 0 & 0 & 0 & -(\lambda T)^{-\frac{1}{2}} & (\lambda T)^{-\frac{1}{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & (\lambda T)^{-\frac{1}{2}} \end{bmatrix}$$

According to the DSGE model

$$\sum_{t=1}^{\lambda T} D_{\lambda T}(y^*_{t-1} y^{*'}_{t-1}) D'_{\lambda T} \xrightarrow{p} \Gamma_{DyyD}(\theta) \tag{38}$$

converges in probability. In order to incorporate the model population moments of the rotated variables into the mixed estimation of the DSGE model we have to undo the rotation:

$$\Gamma_{yy}(\theta) = D^{-1}_{\lambda T} \Gamma_{DyyD} (D'_{\lambda T})^{-1} \tag{39}$$

This procedure is valid as long as the growth rate of technology and the steady state inflation rate are strictly positive and can be used to derive the matrices $\Gamma_{xx}(\theta)$ and $\Gamma_{xy}(\theta)$ that are needed for the prior.

3.4 Identification

DS propose an identification scheme for the DSGE-VAR, which is described in the remainder of this section. To achieve identification we need to construct a mapping between the structural shocks ϵ_t and the one-step-ahead forecast errors u_t. Let Σ_{tr} be the Cholesky decomposition of Σ_u. It is well known that in any exactly identified structural VAR the relationship between u_t and ϵ_t can be characterized as follows:

$$u_t = \Sigma_{tr} \Omega \epsilon_t, \tag{40}$$

where Ω is an orthonormal matrix and the structural shocks are from now on standardized to have unit variance, that is $\mathbb{E}[\epsilon_t \epsilon_t'] = \mathcal{I}$. According to Equation (25) the initial impact of ϵ_t on the endogenous variables y_t in the VAR is given by

$$\left(\frac{\partial \Delta y_t}{\partial \epsilon_t'} \right)_{VAR} = \Sigma_{tr} \Omega. \tag{41}$$

The identification problem arises from the fact that the data are silent about the choice of the rotation matrix Ω. In our framework, it is quite natural in our framework to use the structural model also to identify the VAR. Thus, we will now construct a rotation matrix Ω based on the dynamic equilibrium model.

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 y_t. Using a QR factorization of $A_0(\theta)$, the initial response of y_t to the structural shocks can be can be uniquely decomposed into

$$\left(\frac{\partial y_t}{\partial \epsilon_t'} \right)_{DSGE} = A_0(\theta) = \Sigma^*_{tr}(\theta) \Omega^*(\theta), \tag{42}$$

where $\Sigma^*_{tr}(\theta)$ is lower triangular and $\Omega^*(\theta)$ is orthonormal. To identify the VAR, we maintain the triangularization of its covariance matrix Σ_u and replace the rotation Ω in Equation (41) with the function $\Omega^*(\theta)$ that appears in (42).

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 y_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.

4 Empirical Results

(So far only the baseline version has been implemented. Results are very preliminary.)

- Data set: we are using time series from the database for the (euro) Area-wide model, maintained by the European Central Bank. The database has been constructed from euro area Monthly Bulletin data and Eurostat data where available. It has then been backdated with aggregated country data from various sources. The database covers a wide range of quarterly euro area macroeconomic time-series. We use data starting in the first quarter of 1986, as at that time inflation had come down to a relatively low level. A description can be be found in Fagan, Henry, and Mestre (2001). We are using the following series in our empirical analysis: log real GDP per capita (lgdprpc), log nominal consumption deflated by GDP deflator per capita (lconrpc), log nominal investment deflated by GDP deflator per capita (linvrpc), log hours worked per capita (lhoursuppc) log hourly nominal wage (lhupwin), log GDP deflator (lyed), log nominal M2 per capita (lm2pc) nominal short-term interest rate (3 months) (stn).

- Figure 1 presents time series plots of various ratios: consumption-output ratio appears fairly stable over time. Fluctuations of investment-output ratio are fairly persistent, slight downward trend. Very persistent movement of hours, peaks in the 1991 and decreases throughout the 90's, rises again in 1997. Velocity is falling from 1986 to 1992, and rising afterwards. Real wage as fraction of output is falling throughout the sample period.

- We fit our DSGE-VAR to unfiltered data. DSGE model implies that consumption-output ratio, investment-output ratio, hours worked, velocity, and the ratio of ratio of real wage to GDP are stationary. These "long-run" implications are to some extend at odds with the data. If DSGE is fitted to the data directly we expect the autocorrelation estimates for some of the exogenous shocks to be close to unity. In order to fit the data well there is need to relax DSGE model restrictions.

- We consider a VAR with 3 lags, specified in growth rates. Since we have 8 endogenous variables, our VAR has $8 + 3 * 64 + 8 * 9/2 = 236$ parameters. Estimation period: 1986:I to 2002:IV. We are using 68 observations per equation. Each equation has 25 parameters (plus variances and covariances). Need informative prior distribution to estimate a VAR of this size with a fairly short sample of observation. The DSGE model is more tightly parameterized. It has 47 parameters.

- Tables 3 and 4, columns 2 and 3 contain information on prior distribution for structural parameters. Some of the parameters are fixed. Overall priors are tight, mainly for numerical reasons at this point.

- Table 2: we use a modified harmonic mean estimator (Geweke, 1999) to approximate the log marginal data density defined in 35. The marginal data density can be used to calculate posterior odds. Let $\pi_{i,0}$ and $\pi_{i,T}$ denote prior and posterior probabilities of specification $\mathcal{M}_i$. Then

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}}{\pi_{2,0}} \exp[\ln p(Y|\mathcal{M}_1) - \ln p(Y|\mathcal{M}_2)].$$

If we are assigning equal prior probability to the λ grid points

$$\Lambda = \{0.6, 0.75, 1, 2, 5, 10\}$$

then we find that $\lambda = 0.75$ has the highest posterior probability. Interpretation: the mixed sample that is used to estimate VAR consists of 51 artificial and 68 actual observations.

We also report marginal data densities for a VAR(4). These are generally lower than for a VAR(3) (except for $\lambda = 5$). Conditional on four lags the optimal choice of λ is 1. Since a VAR(4) has more coefficients, the data prefer more artificial observations from the DSGE model to pin down the additional parameters. Overall, VAR(3) with $\lambda = 0.75$ has highest posterior probability.

- Parameter estimates are reported in Tables 3 and 4. Due to numerical difficulties we are fixing a number of parameters at this stage: α, δ, λ_w, L^*, χ, λ_f, and g^*. We also impose the absence of serial correlation for a number of exogenous processes: z_t, $\lambda_{f,t}$, b_t. Since priors are fairly tight, posterior estimates stay close to prior mean. However, there are a few exceptions: for $\lambda = 5$ the posterior mean of the wage stickiness parameter ζ_w is substantially smaller than the prior mean, indicating less wage rigidity. The posterior distribution for the habit parameter h indicates a smaller role for habit formation than implied by the prior. The data also shift our beliefs about the policy parameter ψ_2. The prior mean is 0.13, whereas the posterior means are 0.28, and 0.32, respectively.

- How important is price stickiness? We increase the probability that the firm is able to adjust its price, that is we choose a prior for ζ_p that concentrates near zero. The log marginal data densities for this specification are $-466.71(\lambda = 0.6)$, $-473.89(\lambda = 0.75)$, and $-486.35(\lambda = 1)$. A comparison with the results reported in Table 2 suggests that: the fit of the DSGE-VAR deteriorates if we remove the price stickiness from the model, conditional on the flexible price version our criterion implies that we should add fewer artificial observations to the mixed sample than under the sticky-price specification.

References

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

Chang, Yongsung, Joao Gomes, and Frank Schorfheide (2002): "Learning-by-Doing as Propagation Mechanism," *American Economic Review*, **92**(5), 1498-1520.

Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans (2003): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, forthcoming.

Del Negro, Marco and Frank Schorfheide (2003): "Priors from General Equilibrium Models for VARs," *International Economic Review*, forthcoming.

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

Fagan, Gabriel, Jerome Henry, and Ricardo Mestre (2001): "An Area-wide Model (AWM) for the Euro Area," ECB Working Paper, **42**, January 2001.

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.

Schorfheide, Frank (2000): "Loss Function-Based Evaluation of DSGE Models." *Journal of Applied Econometrics*, **15**, 645-670.

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*, forthcoming.

Smets, Frank and Raf Wouters (2003b): "Forecasting with a Bayesian DSGE Model: An Application to the Euro Area," *Manuscript*, European Central Bank.

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

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

Table 1: EXOGENOUS PROCESSES

Interpretation	Law of Motion
Technology growth	$z_t = \gamma(1-\rho_z) + \rho_z z_{t-1} + \epsilon_{z,t}$
Price mark-up shift	$\ln(\lambda_{f,t}/\lambda_f) = \rho_\lambda \ln(\lambda_{f,t-1}/\lambda_f) + \epsilon_{\lambda,t}$
Capital adjustment costs	$\ln \mu_t = \rho_\mu \ln \mu_{t-1} + \epsilon_{\mu,t}$
Intertemporal preference shift	$\ln b_t = \rho_b \ln b_{t-1} + \epsilon_{b,t}$
Labor supply shift	$\ln(\varphi_t/\varphi) = \rho_\varphi \ln(\varphi_{t-1}/\varphi) + \epsilon_{\varphi,t}$
Money demand shift	$\ln(\chi_t/\chi) = \rho_\chi \ln(\chi_{t-1}/\chi) + \epsilon_{\chi,t}$
Government spending	$\ln(g_t/g) = \rho_g \ln(g_{t-1}/g_t) + \epsilon_{g,t}$
Monetary policy	$\epsilon_{R,t}$

Table 2: CHOICE OF HYPERPARAMETER

λ	Log Marg. Data Density	
	VAR(3)	VAR(4)
0.60	-334.61	-392.42
0.75	-332.94	-366.41
1.00	-351.05	-364.99
2.00	-420.08	-426.23
5.00	-597.45	- 589.21
10.00	-726.24	- 737.64

Notes: The ratio $\exp[\ln p_{\lambda_1}(Y) - \ln p_{\lambda_2}(Y)]$ can be interpreted as posterior odds of λ_1 versus λ_2 if the prior odds are equal to one. Filenames: m21100682002753mh and m21100682002754mh. Sample range: 1986:I to 2002:IV.

Table 3: PRIOR AND POSTERIOR OF DSGE MODEL PARAMETERS

Parameter	Prior		Posterior, $\lambda = 0.75$			Posterior, $\lambda = 5$		
	Mean	Stdd	Mean	90 % Intv.		Mean	90 % Intv.	
α	0.25							
ζ_p	0.75	0.10	0.75	0.72	0.77	0.70	0.68	0.72
δ	0.03							
Υ	0.10	0.05	0.12	0.07	0.17	0.16	0.09	0.22
Φ	0.50	0.25	1.36	1.31	1.43	1.14	1.01	1.24
s'	4.00	1.50	3.73	3.66	3.77	3.93	3.83	4.03
h	0.80	0.10	0.47	0.41	0.55	0.44	0.41	0.47
a''	0.20	0.08	0.12	0.04	0.20	0.05	0.03	0.08
ν	2.00	0.75	2.54	2.48	2.58	2.49	2.43	2.61
ζ_w	0.75	0.10	0.74	0.70	0.77	0.59	0.54	0.64
λ_w	0.30							
r^*	0.90	0.10	0.96	0.90	1.03	1.06	0.93	1.16
ψ_1	1.70	0.10	1.59	1.53	1.63	1.61	1.53	1.68
ψ_2	0.13	0.10	0.28	0.20	0.36	0.32	0.25	0.38
ρ_r	0.80	0.10	0.91	0.89	0.93	0.88	0.86	0.90
π^*	0.65	0.05	0.67	0.65	0.71	0.78	0.67	0.89
γ	0.50	0.25	0.56	0.51	0.61	0.50	0.46	0.54
L^*	0.00							
χ	0.10							
λ_f	0.30							
g^*	0.15							
L^{adj}	5.00	1.00	5.70	5.64	5.76	5.56	5.47	5.64

Notes: We report posterior means and 90% probability intervals based on the output of the Metropolis-Hastings Algorithm. Sample range: 1986:I to 2002:IV. Filename: m21100682002753mom.

Table 4: Prior and Posterior of DSGE Model Parameters

Parameter	Prior		Posterior, $\lambda = 0.75$			Posterior, $\lambda = 5$		
	Mean	Stdd	Mean	90 % Interval		Mean	90 % Interval	
ρ_z	0.00							
ρ_ϕ	0.85	0.10	0.62	0.57	0.67	0.39	0.31	0.46
ρ_χ	0.85	0.10	0.90	0.86	0.94	0.74	0.67	0.80
ρ_{λ_f}	0.00							
ρ_μ	0.85	0.10	0.70	0.60	0.76	0.81	0.75	0.86
ρ_b	0.00							
ρ_g	0.90							
σ_z	0.40	2.00	0.29	0.26	0.33	0.51	0.44	0.56
σ_ϕ	1.00	2.00	1.77	1.71	1.85	2.04	1.98	2.12
σ_χ	1.00	2.00	1.88	1.85	1.92	1.92	1.83	1.96
σ_{λ_f}	1.00	2.00	6.84	6.75	6.93	6.76	6.72	6.80
σ_μ	1.00	2.00	1.38	1.32	1.42	1.35	1.31	1.41
σ_b	0.20	2.00	0.27	0.22	0.31	0.48	0.43	0.55
σ_g	0.30	2.00	0.17	0.15	0.19	0.29	0.27	0.32
σ_r	0.10	2.00	0.04	0.03	0.04	0.07	0.06	0.09

Notes: We report posterior means and 90% probability intervals based on the output of the Metropolis-Hastings Algorithm. Sample range: 1986:I to 2002:IV. Filename: m21100682002753mom.

Figure 1: TIME SERIES PLOTS OF SELECTED RATIOS