

# School of Economics and Finance

## DSGE Priors for BVAR Models

Thomai Filippeli and Konstantinos Theodoridis

Working Paper No. 713

March 2014

ISSN 1473-0278



Queen Mary  
University of London

# DSGE Priors for BVAR Models\*

Thomai Filippeli<sup>†</sup>

Konstantinos Theodoridis<sup>‡</sup>

December 20, 2013

## Abstract

Similar to Ingram and Whiteman (1994), De Jong et al. (1993) and Del Negro and Schorfheide (2004) this study proposes a methodology of constructing Dynamic Stochastic General Equilibrium (DSGE) consistent prior distributions for Bayesian Vector Autoregressive (BVAR) models. The moments of the assumed Normal-Inverse Wishart (no conjugate) prior distribution of the VAR parameter vector are derived using the results developed by Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) regarding structural VAR (SVAR) models and the normal prior density of the DSGE parameter vector. In line with the results from previous studies, BVAR models with theoretical priors seem to achieve forecasting performance that is comparable – if not better – to the one obtained using theory free “Minnesota” priors (Doan et al., 1984). Additionally, the marginal-likelihood of the time-series model with theory founded priors – derived from the output of the Gibbs sampler – can be used to rank competing DSGE theories that aim to explain the same observed data (Geweke, 2005). Finally, motivated by the work of Christiano et al. (2010b,a) and Del Negro and Schorfheide (2004) we use the theoretical results developed by Chernozhukov and Hong (2003) and Theodoridis (2011) to derive the quasi Bayesian posterior distribution of the DSGE parameter vector.

*JEL Classification:* C11, C13, C32, C52

*Keywords:* BVAR, SVAR, DSGE, Gibbs Sampling, Marginal-Likelihood Evaluation, Predictive Density Evaluation, Quasi-Bayesian DSGE Estimation

---

\*The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. We would like to thank an anonymous referee whose valuable suggestions and comments helped us to improve the substance of the paper substantially

<sup>†</sup>Queen Mary University of London, Email: t.filippeli@qmul.ac.uk

<sup>‡</sup>Bank of England, Email: konstantinos.theodoridis@bankofengland.co.uk

# 1 Motivation

Bayesian inference relies on the properties of the posterior distribution of the parameter vector, which is proportional to the product of the likelihood times the prior distribution. The role of the latter is to integrate into the estimation the econometrician’s “knowledge” about the parameter vector. For example, let us say that we want to measure the persistence of the real consumption series by estimating an autoregressive of order one model and the prior knowledge is that real (detrended) macroeconomic series display high degree of sluggishness. These two features – stationarity and significant positive autocorrelation – can be incorporated into the estimation by using the beta distribution, which is defined between zero and one, as the prior density of the lag coefficient with a mean that reflects the high rigidity.

The importance of the prior distribution increases further as the length of parameter vector expands and the size of the sample remains relatively small, which is the case when a macro-variable VAR is estimated. For instance, a seven variable VAR with three lags – like the one considered in this study – consists of one hundred and seventy five parameters whose estimation poses serious difficulties even with fifty years of quarterly data, i.e. 1.14 observations per parameter. Under these circumstances, priors can be used to shrink the number of the estimated parameters by focusing on some of them and ignoring others.

Minnesota priors (Litterman, 1980, 1986; Doan et al., 1984) act exactly in this way and this probably explains why it is always econometricians’ first choice when the estimation of a macroeconomic BVAR model is considered (Kadiyala and Karlsson, 1997; Bandbura et al., 2010). They rely on the stylised fact that random walk models deliver superior forecasting performance, implying that the prior mean of all autoregressive coefficients, except those on the main diagonal of the first lag, is equal to zero. In terms of the previous example this means that only seven parameters – out of one hundred and seventy five – are non zero before the estimation.

The above assumption, however, seems inconsistent with the economic theory according to which macroeconomic variables are functions of common structural shocks and, consequently, they exhibit a high degree of comovement. For example, a real business cycle model with persistent productivity and labour supply shock processes (see, Hansen, 1985) predicts strong correlation among model’s endogenous series (i.e. consumption, investment, wages, output etc), meaning that the orthogonality assumption adopted by the Minnesota literature cannot be easily founded on the basis of economic theory. This lack of theoretical consistency limits researchers’ intuition and, therefore, complicates the posterior inference.

The question that naturally arises is whether we are able to construct prior distributions for BVAR models that do not conflict with economic theory. Similar to Ingram and Whiteman (1994), De Jong et al. (1993) and Del Negro and Schorfheide (2004), this study illustrates how the relation between VAR and DSGE models can be explored to deliver theory consistent priors. Why we would like to do this? DSGE models are devices where stylised facts can be decomposed into agents’ optimisation problems, implying that data features can be expressed as functions of the structural parameters. For example, the hump shaped real consumption response after a monetary policy shock observed in VAR studies (Christiano et al., 1998) could be reproduced by a DSGE model where households form consumption habits (Smets and Wouters, 2007), and by varying the consumption smoothing parameter in this model the researcher controls the peak – or the trough – of the shock and, consequently, the time-series properties of the entire state vector. This mapping between the DSGE parameters and stylised facts is what guides model developers to decide about the prior mean of the structural parameter vector,

while its standard deviation aims to capture most of the values used in the literature. Answering the previous question, it seems that the prior distribution of the structural parameter vector is well motivated and, given the close association between DSGE and VAR models, this information can be used to develop prior distributions for the time-series model, which are clearly not subject to the orthogonality critique discussed in the previous paragraph.

Also we illustrate how the mapping between the DSGE and SVAR models – identified by Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) – the normal prior density of the structural parameter vector and the “Mean Value Theorem” (see, White, 2001) are combined to derive the moments of the Normal-Wishart prior distribution of the time-series model. Non conjugate priors implies that the posterior distribution of the VAR parameter vector does not have an analytic form. However, the kernel of the latter distribution can be re-written as the product of two conditional distributions – Normal-Inverse Wishart – and the posterior distribution of the reduced-form parameter vector can be approximated using the Monte Carlo Markov Chain (MCMC) Gibbs sampler.

Arguing in favour of DSGE consistent priors does not make much sense if the fit or/and the forecasting performance of the VAR model deteriorates relative to theory free priors. A set of metrics – marginal-likelihood assessment, mean squared forecast error comparison and predictive density evaluation – is employed here to assess the performance of the proposed priors. Similar to the previous studies, all results indicate that theory motivated priors deliver time-series properties that are comparable – if not better – to the one obtained using theory free priors.

The marginal-likelihood of the time series model with theoretical priors – approximated from the output of the Gibbs sampler using either the methodology developed by Chib (1995) or Geweke’s modified harmonic mean estimator (Geweke, 1999) – can be viewed as a measure of fit that assesses the quality of the restrictions imposed by the structural model or/and the prior distribution of the DSGE parameter vector on the data. This information can be used to rank competing DSGE theories or/and structural parameter prior moments thus helping the researcher to increase the fit of his DSGE model. An increasing number of studies – for instance, Iskrev (2010) and Caglar et al. (2012) – suggests that DSGE models are weakly identified, making the marginal-likelihood of the estimated structural model a spurious measure of fit that it cannot be used with confidence, as the posterior estimates tend to be dominated by their prior moments (Kleibergen and Mavroeidis, 2011). In other words, we believe that the “indirect” likelihood is the only reliable way of assessing competing DSGE models.

Finally, we use the impulse-response matching estimator proposed by Theodoridis (2011) and the theoretical results developed by Chernozhukov and Hong (2003) to derive the quasi Bayesian posterior distribution of the DSGE parameter vector. This estimator has many similarities with the one proposed recently by Christiano et al. (2010b,a), however, ours utilises all moments/conditions – for all shocks and observable variables – implied by the structural model. The estimator proposed by Theodoridis under certain conditions mimics the maximum-likelihood one – see the discussion in Lewis and Reinsel (1985) – and this perhaps explains why Canova and Sala (2009) use the term “full-information” for such estimators.<sup>1</sup>

The paper is organised as follows. The notation needed for this study is developed next, the structural model is reviewed in Section 3, the existing methodologies of constructing theoretical priors are discussed in Section 4, the proposed methodology is described in Section 5, an application is considered in Section 6 and the final section concludes.

---

<sup>1</sup>The term “full-information” is loosely used here.

## 2 Notation

The required notation is developed in this section.  $\mathbb{R}$  denotes the real line,  $da$  indicates the dimension of the vector  $a$ ,  $\mathbb{R}^{da} \equiv \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$  is the  $da$ -cartesian power of the real line,  $I_{da}$  stands for the  $(da \times da)$  identity matrix, the  $vec$  operators transforms a matrix with dimensions  $da \times da$  to an  $da^2 \times 1$  vector by stacking the columns. The symbol  $\otimes$  denotes the Kronecker product operator, while,  $\nabla_a f(a)$  represents the matrix of the first derivatives of the vector function  $f(a)$  with respect to the vector  $a$ . The state space representation of a solved (log) linear approximated DSGE model is given by

$$y_t = A(\gamma) \xi_t \quad (1)$$

$$\xi_t = B(\gamma) \xi_{t-1} + \Upsilon(\gamma) \omega_t \quad (2)$$

where the equation (2) describes the evolution of the state vector ( $\xi_t \in \mathbb{R}^{d\xi}$ ) of the model, expression (1) illustrates the relation between the unobserved state of the economy and the observable variables ( $y_t \in \mathbb{R}^{dy}$ ), the vector of the structural shocks ( $\omega_t \in \mathbb{R}^{d\omega}$ ) is normally distributed with mean zero and  $I_{d\omega}$  covariance matrix ( $N \sim (0, I_{d\omega})$ ) and the elements of the matrices  $A(\gamma)$ ,  $B(\gamma)$ , and  $\Upsilon(\gamma)$  are nonlinear functions of the DSGE parameter vector, which is also called the structural parameter vector,  $\gamma \in \Gamma$ . The VAR( $p$ ) model is described by

$$y_t = \sum_{i=1}^p \Delta_i y_{t-i} + v_t \quad (3)$$

where  $v_t$ , the vector of the reduced form error, is normally distributed with zero mean and  $\Sigma_v$  variance-covariance matrix and its standard regression representation is

$$Y = \Delta X + V \quad (4)$$

where  $\Delta = \begin{bmatrix} \Delta_1 & \dots & \Delta_p \end{bmatrix}$  is the  $dy \times pdy$  matrix of the VAR coefficients,  $T$  is the sample size,  $Y$  is the  $dy \times T$  data matrix of the observed variables,  $X$  is the  $pdy \times T$  matrix of the lagged data and  $V$  is the  $dy \times T$  matrix of the VAR innovations.  $\delta \equiv vec(\Delta)'$  and  $\sigma_v \equiv vec(\Sigma_v)'$  are the components of the VAR parameter vector  $\theta \equiv (\delta', \sigma_v')' \in \Theta$ , which is called the reduced-form parameter vector;  $\Gamma$  and  $\Theta$  are compact subsets of  $\mathbb{R}^{d\gamma}$  and  $\mathbb{R}^{d\theta}$ , respectively. The OLS estimates of  $\Sigma_v$ ,  $\Delta$ ,  $\delta$ ,  $\sigma_v$  and  $\theta$  are defined as  $\widehat{\Sigma}_v$ ,  $\widehat{\Delta}$ ,  $\widehat{\delta}$ ,  $\widehat{\sigma}_v$  and  $\widehat{\theta}$ , respectively.  $C_{yy} = \mathbb{E}(Y_t Y_t')$ ,  $C_{xx} = \mathbb{E}(X_t X_t')$ ,  $C_{yx} = \mathbb{E}(Y_t X_t')$ ,  $\widehat{C}_{yy} = T^{-1} Y Y'$ ,  $\widehat{C}_{xx} = T^{-1} X X'$ ,  $\widehat{C}_{yx} = T^{-1} Y X'$  are the population moments of the data and their estimates.  $N(\mu_\alpha, \Sigma_\alpha)$  stands for the normal distribution, where  $\mu_\alpha$  and  $\Sigma_\alpha$  denote the mean and the covariance matrix of the vector  $\alpha$ , respectively. The Wishart and its inverse distributions with  $\eta$  degrees of freedom and  $\Pi$  scale matrix are defined as  $W(\Pi, \eta)$  and  $IW(\Pi^{-1}, \eta)$ , respectively.  $p(a)$  denotes the prior distribution of  $a$ ,  $L(Y|\theta)$  and  $m(Y) \equiv \int L(Y|\theta) p(\theta) d\theta$  stand for the likelihood and marginal-likelihood of the VAR. Finally, the companion matrix  $\bar{\Delta}$  is defined as

$$\bar{\Delta} \equiv \begin{bmatrix} \Delta_1 & \Delta_2 & \dots & \dots & \Delta_p \\ I_{dy} & 0_{dy \times dy} & \dots & \vdots & 0_{dy \times dy} \\ 0_{dy \times dy} & I_{dy} & \dots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{dy \times dy} & 0_{dy \times dy} & \dots & I_{dy} & 0_{dy \times dy} \end{bmatrix} \quad (5)$$

### 3 Structural Model

This section briefly discusses some of the key linearised equilibrium conditions of the model developed by Smets and Wouters (2007).<sup>2</sup> Readers who are interested in agents' decision problems are recommended to consult the references mentioned above directly. All the variables are expressed as log deviations from their steady-state values,  $\mathbb{E}_t$  denotes expectation formed at time  $t$ ,  $'-'$  denotes the steady state values and all the shocks ( $\omega_t^i$ ) are assumed to be normally distributed with zero mean and unit standard deviation.

The demand side of the economy consists of consumption ( $c_t$ ), investment ( $i_t$ ), capital utilisation ( $z_t$ ) and government spending  $\varepsilon_t^g = \rho_g \varepsilon_{t-1}^g + \sigma_g \omega_t^g$ , which is assumed to be exogenous. The market clearing condition is given by

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g \quad (6)$$

where  $y_t$  denotes the total output, Table 1 provides a full description of the model's parameters and their prior moments.

The consumption Euler equation is given by

$$c_t = \frac{h}{1+h} c_{t-1} + \frac{1}{1+h} \mathbb{E}_t c_{t+1} + \frac{(1-\sigma_C)(\bar{W}^h \bar{L} | \bar{C})}{\sigma_C(1+h)} (\mathbb{E}_t l_{t+1} - l_t) - \frac{1-h}{\sigma_C(1+h)} (r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b) \quad (7)$$

where  $l_t$  is the hours worked,  $r_t$  is the nominal interest rate,  $\pi_t$  is the rate of inflation and  $\varepsilon_t^b = \rho_g \varepsilon_{t-1}^b + \sigma_g \omega_t^b$  is the risk premium shock. If the degree of habits is zero ( $h = 0$ ), equation (7) reduces to the standard forward looking consumption Euler equation. The linearised investment equation is given by

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} \mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta)S''} q_t + \varepsilon_t^i \quad (8)$$

where  $i_t$  denotes the investment and  $q_t$  is the real value of existing capital stock (Tobin's Q). The sensitivity of investment to real value of the existing capital stock depends on the parameter  $S''$  (see, Christiano et al., 2005). The corresponding arbitrage equation for the value of capital is given by

$$q_t = \beta(1-\delta) \mathbb{E}_t q_{t+1} + (1-\beta(1-\delta)) \mathbb{E}_t r_{t+1}^k - (r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b) \quad (9)$$

where  $r_t^k = -(k_t^s - l_t) + w_t$  denotes the real rental rate of capital which is negatively related to the capital-labour ratio and positively to the real wage.

On the supply side of the economy, the aggregate production function is defined as

$$y_t = \phi_p (\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a) \quad (10)$$

where  $k_t^s$  represents capital services which is a linear function of lagged installed capital ( $k_{t-1}$ ) and the degree of capital utilisation,  $k_t^s = k_{t-1} + z_t$ . Capital utilization, on the other hand, is proportional to the real rental rate of capital,  $z_t = \frac{1-\psi}{\psi} r_t^k$ . The total factor of productivity follows an AR(1) process,

---

<sup>2</sup>The only difference between the model described in this Section and that developed by Smets and Wouters (2007) is that we have switched off the linear trend.

Table 1: DSGE Parameter Description &amp; Prior Moments

<i>Symbols</i>	<i>Description</i>	<i>Prior Mean</i>	<i>Prior Standard Deviation</i>	<i>Probability Density Function</i>
$B$	Fixed Cost	1.250	0.125	Normal
$S''$	Steady State Capital Adjustment Cost Elasticity	4.000	1.500	Normal
$\alpha$	Capital Production Share	0.300	0.050	Normal
$\sigma$	Intertemporal Substitution	0.850	0.375	Normal
$h$	Habit Persistence	0.700	0.050	Beta
$\xi_w$	Wages Calvo Parameter	0.500	0.100	Beta
$\sigma_l$	Labour Supply Elasticity	2.000	0.750	Normal
$\xi_p$	Prices Calvo Parameter	0.500	0.100	Beta
$i_w$	Wage Indexation	0.500	0.150	Beta
$i_p$	Price Indexation	0.500	0.150	Beta
$z$	Capital Utilisation Adjustment Cost	0.500	0.150	Beta
$\phi_\pi$	Taylor Inflation Parameter	1.750	0.150	Normal
$\phi_r$	Taylor Inertia Parameter	0.750	0.100	Beta
$\phi_y$	Taylor Output Gap Parameter	0.125	0.05	Normal
$\rho_i$	Investment Shock Persistence	0.750	0.100	Beta
$\rho_g$	Government Spending Shock Persistence	0.750	0.100	Beta
$\rho_a$	Productivity Shock Persistence	0.750	0.100	Beta
$\rho_b$	Premium Shock Shock Persistence	0.750	0.100	Beta
$\rho_p$	Prices Markup Shock Persistence	0.750	0.100	Beta
$\rho_w$	Wages Markup Shock Persistence	0.750	0.100	Beta
$\theta_p$	Prices Markup Shock Moving Average	0.750	0.100	Beta
$\theta_w$	Wages Markup Shock Moving Average	0.750	0.100	Beta
$\rho_r$	Monetary Policy Shock Shock Persistence	0.750	0.100	Beta
$\sigma_i$	Investment Shock Uncertainty	0.500	0.100	Inv. Gamma
$\sigma_g$	Government Spending Shock Uncertainty	0.500	0.100	Inv. Gamma
$\sigma_a$	Productivity Shock Uncertainty	0.500	0.100	Inv. Gamma
$\sigma_b$	Risk Premium Shock Uncertainty	0.500	0.100	Inv. Gamma
$\sigma_p$	Prices Markup Shock Uncertainty	0.500	0.100	Inv. Gamma
$\sigma_w$	Wages Markup Shock Uncertainty	0.500	0.100	Inv. Gamma
$\sigma_r$	Policy Shock Uncertainty	0.500	0.100	Inv. Gamma

$\varepsilon_t^a = \rho_g \varepsilon_{t-1}^a + \sigma_g \omega_t^a$ . The accumulation process of installed capital is simply described as

$$k_t = (1 - \delta) k_{t-1} + \delta i_t + (1 + \beta) \delta S'' \varepsilon_t^i \quad (11)$$

where the investment shock,  $\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \sigma_i \omega_t^i$ , increases the stock of capital in the economy exogenously. Monopolistic competition within the production sector and Calvo-pricing constraints gives the following New-Keynesian Phillips curve for inflation

$$\pi_t = \frac{i_p}{1 + \beta i_p} \pi_{t-1} + \frac{\beta}{1 + \beta i_p} \mathbb{E}_t \pi_{t+1} + \frac{1}{(1 + \beta i_p)} \frac{(1 - \beta \xi_p)(1 - \xi_p)}{(\xi_p((\phi_p - 1)\varepsilon_p + 1))} mc_t + \varepsilon_t^p \quad (12)$$

where  $mc_t = \alpha r_t^k + (1 - \alpha) w_t - \varepsilon_t^a$  is the marginal cost of production and  $\varepsilon_t^p = \rho_p \varepsilon_{t-1}^p + \sigma_p \omega_t^p - \theta_p \sigma_p \omega_{t-1}^p$  denotes the price mark-up shock. Monopolistic competition and Calvo wage-setting in the labour

market also gives rise to a similar wage New-Keynesian Phillips curve

$$w_t = \frac{1}{1+\beta}w_{t-1} + \frac{\beta}{1+\beta}(\mathbb{E}_t w_{t+1} + \mathbb{E}_t \pi_{t+1}) - \frac{1+\beta i_w}{1+\beta}\pi_t + \frac{i_w}{1+\beta}\pi_{t-1} + \frac{1}{1+\beta} \frac{(1-\beta\xi_w)(1-\xi_w)}{(\xi_w((\phi_w-1)\varepsilon_w+1))} \mu_t^w + \varepsilon_t^w \quad (13)$$

where  $\mu_t^w = \left(\sigma_l l_t + \frac{1}{1-h}(c_t - hc_{t-1})\right) - w_t$  is the households' marginal benefit of supplying an extra unit of labour service and  $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \sigma_w \omega_t^w - \theta_w \sigma_w \omega_{t-1}^w$  is the wage mark-up shock.

Finally, the monetary policy maker is assumed to set the nominal interest rate according to the following Taylor-type rule

$$r_t = \rho r_{t-1} + (1-\rho)(r_\pi \pi_t + r_y y_t) + \varepsilon_t^r \quad (14)$$

where  $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \sigma_r \omega_t^r$  is the monetary policy shock.

Similarly to Smets and Wouters (2007) we assume that real consumption, real investment, real output, hours, inflation, real wages and nominal interest rates are observed.

## 4 Existing Methodologies

### Ingram and Whiteman (1994)

Ingram and Whiteman were the first to construct theoretical priors for BVAR models based on a simple Real Business Cycle (King et al., 1988) model, with two state variables (capital and technology process) and one stochastic disturbance (the innovation of the technology process). The simplicity of the model allows the state vector to be expressed as a linear function of the observed series – this is achieved through the use of the Generalized Inverse<sup>3</sup> –

$$\left[ (A(\gamma)' A(\gamma))^{-1} A(\gamma)' \right] y_t = \xi_t \quad (15)$$

meaning that the DSGE system – equations (1)-(2) – can be rewritten as a VAR(1)

$$\begin{aligned} y_t &= A(\gamma) B(\gamma) \left[ (A(\gamma)' A(\gamma))^{-1} A(\gamma)' \right] y_{t-1} + A(\gamma) \Upsilon(\gamma) \omega_t \\ &= \Delta_1(\gamma) y_{t-1} + C(\gamma) \omega_t \end{aligned} \quad (16)$$

with the last expression illustrating the mapping between the VAR and DSGE parameter vector

$$\theta(\gamma) \equiv \left( \text{vec}(\Delta_1(\gamma))', \text{vec}(C(\gamma) C(\gamma)')' \right)' \quad (17)$$

The existence of this mapping and the properties of the normal distribution – the distribution of a non-linear function of normal variables can be approximated by the normal distribution – are, possibly, the reasons behind Ingram and Whitemans' selection of the normal probability density function (pdf) as the prior distribution of the DSGE parameter vector. To be precise, the assumption that  $\gamma$  is normally distributed with mean and variance-covariance matrix equal to  $\mu_\gamma$  and  $\Sigma_\gamma$ , respectively, and the use of the ‘‘Mean Value Theorem’’ (White, 2001) imply that  $\theta(\gamma)$  is also normally distributed with mean and variance-covariance matrix given by  $\mu_\theta(\gamma) \equiv \theta(\mu_\gamma)$  and  $\Sigma_\theta(\gamma) \equiv \nabla_\gamma \theta(\mu_\gamma) \Sigma_\gamma \nabla_\gamma \theta(\mu_\gamma)'$ , respectively.

---

<sup>3</sup>See, Magnus and Neudecker (2002)



Having derived the prior distribution of the time-series parameter vector, the posterior estimation of the VAR is carried out by adopting single equation “mixed” estimation procedures introduced by Theil and Goldberger (1961).

### De Jong et al. (1993) & Del Negro and Schorfheide (2004)

The work of Del Negro and Schorfheide can be viewed as an attempt to provide the theoretical validation behind the study of De Jong et al. (1993) and, therefore, we review only the former procedure. In this case the actual data set is augmented with a number of artificial observations simulated by the structural model –  $Y(\gamma) \equiv \{Y(\gamma)_t\}_{t=1}^{\lambda T}$  and  $X(\gamma) \equiv \{X(\gamma)_t\}_{t=1}^{\lambda T}$  –, which is proportional to the size of the actual sample –  $\lambda T$  where  $\lambda \in (0, \infty)$ . The VAR likelihood of this augmented sample is factorised into the likelihood of the true data and the likelihood of the artificial one

$$L(Y(\gamma), Y|\theta) = L(Y|\theta) L(Y(\gamma)|\theta) \quad (18)$$

with the latter to be interpreted as the prior density of  $\theta$ . To avoid the stochastic variation rising by the simulation of the model, the authors replace the non-standardised sample moments of the likelihood –  $Y(\gamma)Y(\gamma)'$ ,  $Y(\gamma)X(\gamma)'$  and  $X(\gamma)X(\gamma)'$  – with their expected values, which are written as function of the structural parameters and the hyperparameter that controls the size of the simulated data –  $\lambda TC_{yy}(\gamma)$ ,  $\lambda TC_{xx}(\gamma)$  and  $\lambda TC_{yx}(\gamma)$ . Since the likelihood can be decomposed into the product of the conditional Normal times the Wishart distribution (see, Canova, 2005), expression (18) can be viewed as the kernel of the posterior distribution of the VAR parameter vector

$$p(\theta|Y) \propto L(Y|\theta) p(\delta|\sigma_v, \gamma) p(\sigma_v|\gamma) \quad (19)$$

with the prior moments written as function of  $\gamma$

$$p(\sigma_v|\gamma) \equiv IW(\lambda T \Sigma_v(\gamma), \lambda T - (pdy + 1)) \quad (20)$$

$$p(\delta|\sigma_v, \gamma) \equiv N(\Delta(\gamma), \Sigma_v \otimes (\lambda TC_{xx}(\gamma))^{-1}) \quad (21)$$

$$\Delta(\gamma) \equiv C_{xx}(\gamma)^{-1} C_{xy}(\gamma) \quad (22)$$

$$\Sigma_v(\gamma) \equiv C_{yy}(\gamma) - C_{yx}(\gamma) \Delta(\gamma) \quad (23)$$

Conjugate priors imply an analytic posterior distribution for  $\theta$  – conditional Normal-Inverse Wishart – and its moments can be written as weighted average between the DSGE implied moments and the OLS estimates

$$p(\sigma_v|Y, \hat{\sigma}_v, \gamma) \equiv IW((\lambda + 1) T \bar{\Sigma}_v(\gamma), (\lambda + 1) T - (dyp + 1)) \quad (24)$$

$$p(\delta|Y, \hat{\delta}, \sigma_v, \gamma) \equiv N(\bar{\Delta}(\gamma), \Sigma_v \otimes (\lambda TC_{xx}(\gamma))^{-1}) \quad (25)$$

$$\bar{\Delta}(\gamma) \equiv (\lambda TC_{xx}(\gamma) + X'X)^{-1} (\lambda TC_{xx}(\gamma) \Delta(\gamma) + X'X \hat{\Delta}) \quad (26)$$

$$\bar{\Sigma}_v(\gamma) \equiv \frac{1}{(\lambda + 1) T} [(\lambda TC_{yy}(\gamma) + Y'Y) - (\lambda TC_{yx}(\gamma) + Y'X) \bar{\Delta}(\gamma)] \quad (27)$$

It is not hard to see that this is not exactly the prior distribution of the reduced-form parameter vector  $\theta(\gamma)$  but the distribution of the OLS VAR estimate  $\hat{\theta}(\gamma)$  under the assumption that the DSGE model is the true data generation process. The consistency property of the OLS estimator ensures that  $\hat{\theta}(\gamma)$

converges in probability to  $\theta(\gamma)$  as the size of the pseudo sample tends to infinite –  $\widehat{\theta}(\gamma) - \theta(\gamma) \xrightarrow{\text{Pr}} 0_{d\theta}$  as  $\lambda \rightarrow \infty$  –, however, the variance-covariance matrix  $\Sigma_v \otimes (\lambda TC_{xx}(\gamma))^{-1}$  does not measure the dispersion of the prior probability density function of  $\theta(\gamma) - p(\theta(\gamma))$  – but it reflects the variance of the estimation error –  $\mathbb{E} \left( \widehat{\theta}(\gamma) - \theta(\gamma) \right) \left( \widehat{\theta}(\gamma) - \theta(\gamma) \right)'$ , which converges to zero as  $\lambda \rightarrow \infty$ . This implies that information captured by the second moments of  $p(\gamma)$  is not properly explored. The variance of  $\gamma$  reflects econometrician's confidence about  $\mu_\gamma$  and researchers may be more certain about some parameters than others, meaning that some VAR parameters should get more prior weight than others. Additionally, the functional form of the variance-covariance matrix –  $\Sigma_v \otimes (\lambda TC_{xx}(\gamma))^{-1}$  – strongly restricts the covariances among the VAR coefficients. For instance, it implies that ratios of variances of coefficients on the same variable in different equations are identical (see, De Jong et al., 1993; Sims and Zha, 1998), a feature clearly not in line with the DSGE model.

## 5 The Proposed Method

This section describes the proposed methodology, which can be seen as an extension of Ingram and Whitemans' work. The mapping between DSGE and SVAR models – identified by the studies of Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) – and the assumption that the DSGE parameter vector is normally distributed are used to derive the moments of the Normal-Inverse Wishart (no conjugate) prior distribution of the SVAR parameter vector.

The starting point of our methodology is the DSGE model described by equations (1) and (2). It is known from the work of Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) that when the number of shocks coincides with the number of the observable variables and the eigenvalues of the matrix

$$M(\gamma) \equiv \left[ I_{d\xi} - \Upsilon(\gamma) [A(\gamma) \Upsilon(\gamma)]^{-1} A(\gamma) \right] B(\gamma) \quad (28)$$

are less than one in absolute terms then there is an analytical mapping between the structural and VAR parameter vector

$$\phi : \gamma \rightarrow \theta \quad (29)$$

namely,

$$\delta(\gamma)_i = \text{vec}(\Delta(\gamma)_i) \quad (30)$$

$$\Delta(\gamma)_i \equiv A(\gamma) B(\gamma) M(\gamma)^{i-1} \Upsilon(\gamma) [A(\gamma) \Upsilon(\gamma)]^{-1} \quad (31)$$

$$\Sigma_u(\gamma) = [A(\gamma) \Upsilon(\gamma)] [A(\gamma) \Upsilon(\gamma)]' \quad (32)$$

where  $i = 1, \dots, \infty$ . In this framework, it is not hard to see that the VAR residuals are linked with the structural disturbances via

$$v_t = A(\gamma) \Upsilon(\gamma) \omega_t \quad (33)$$

### Prior distribution of $\delta$

The assumption that the VAR coefficient parameter vector  $\delta$  is normally distributed implies that we need only two moments – mean and variance – to define its prior distribution. In the current setup these can be obtained by combining (i) the assumption that the structural parameter vector  $\gamma$  is

normally distributed, (ii) the ‘‘Mean Value Theorem’’ and (iii) equations (30) and (31),

$$\mu_\delta(\gamma) \equiv \phi(\mu_\gamma) \quad (34)$$

$$\tilde{\Sigma}_\delta(\gamma) = \nabla_\gamma \phi(\mu_\gamma) \Sigma_\gamma \nabla_\gamma \phi(\mu_\gamma)' \quad (35)$$

From the discussion in the previous paragraph we know that the selection of the prior mean of the structural parameter vector –  $\mu_\gamma$  – must satisfy the condition that the eigenvalues of the matrix  $M(\mu_\gamma)$  are less than one in absolute terms in order the mapping (29) to exist.

### Singularity of $\tilde{\Sigma}_\delta(\gamma)$

Structural parameters have an economic interpretation and they are not arbitrary added like the VAR ones aiming to capture all the structure observed in the data – for example, we usually increase the number of VAR lags up to the point where the estimated residuals behave like identically independently distributed (i.i.d.) processes. This implies that the dimension of the DSGE parameter vector  $\gamma$  is expected to be (much) smaller than the reduced-form parameter vector  $\delta(\gamma)$ , meaning that the variance-covariance matrix of the latter –  $\tilde{\Sigma}_\delta(\gamma)$  – is positive semi-definite. This singularity – the rank of  $\tilde{\Sigma}_\delta(\gamma)$  is equal to the dimension of  $\gamma$  and not equal to the number of elements of  $\delta$  – is another restriction imposed by the theory on the data. We take this constraint into account when we construct the prior distribution of the VAR parameter vector and this is in sharp contrast with De Jong et al. (1993) and Del Negro and Schorfheide (2004) who implicitly ignore this feature (see the discussion in Section 4).

To be precise, we relax this constraint by using only the diagonal elements of  $\tilde{\Sigma}_\delta(\gamma)$  and setting all the off diagonal elements equal to zero

$$\Sigma_\delta(\gamma) = \text{diag}(\tilde{\Sigma}_\delta(\gamma)) \quad (36)$$

This choice is motivated by the Minnesota priors literature and, in the current framework, it means that we discard cross-correlation moments restrictions implied by the DSGE model regarding VAR autoregressive parameters. By taking this route, we avoid working with singular distributions since the variance-covariance matrix  $\Sigma_\delta(\gamma)$  is now – by construction – a positive-definite matrix.

We would like also to emphasise that for those special cases where the number of the structural parameters exceeds the number of the VAR coefficients the above action is not required

$$\Sigma_\delta(\gamma) = \tilde{\Sigma}_\delta(\gamma)$$

and the rest of the analysis remains exactly the same.

### Prior distribution of $\Sigma_v$

In this case only two parameters – the degrees of freedom ( $\eta$ ) and the scale matrix ( $\Pi$ ) – are required to define the prior distribution of the reduced-form error variance-covariance matrix. We use the first moment of the Inverse-Wishart distribution (see, Poirier, 1995) to decide about them. To be precise, the mean of a random variable that follows the Inverse-Wishart distributed with  $\eta$  degrees of freedom and  $\Pi$  scale is given by

$$\mu_{\Sigma_v} = \frac{1}{\eta - dy - 1} \Pi \quad (37)$$

We set  $\mu_{\Sigma_v}$  equal to  $\Sigma_v(\mu_\gamma)$  – expression (31) – and  $\eta$  equal to  $dy + 2$ , meaning that

$$\Pi(\gamma) = [A(\mu_\gamma) \Upsilon(\mu_\gamma)] [A(\mu_\gamma) \Upsilon(\mu_\gamma)]' \quad (38)$$

In other words, we select the parameters  $\eta$  and  $\Pi(\gamma)$  in such way that the prior mean of  $\Sigma_v$  coincides with the reduced-form error variance-covariance matrix implied by the DSGE model when the structural parameter vector is set equal to its prior mean.<sup>4</sup>

### Normality of $\gamma$

The normal pdf does not fully describe the stochastic properties of the DSGE parameter vector, for instance, Calvo probability, indexation and shock persistence parameters are defined between 0 and 1, while the variances of the structural disturbances cannot be nonnegative. However, under the assumption that the VAR coefficient vector  $\delta$  is normally distributed we need only two moments from the pdf of the structural parameter vector to define the prior distribution of  $\delta$ , and this is also true for the prior distribution of  $\Sigma_v$ . In other words, ignoring information about the higher moments of  $\gamma$  does not seem to have a substantial effect on the specification of prior distribution of the VAR parameter vector.

To illustrate this point we use the model discussed in Section 3 to conduct the following simulation exercise.<sup>5</sup> We draw  $\gamma_N$  from the normal pdf – where the proposed methodology relies on – and  $\gamma_{GD}$  from distribution functions typically used in the DSGE estimation literature – they can be found in the last column of Table 1. Both  $\gamma_N$  and  $\gamma_{GD}$  have the same first and second moments and they are given by the third and fourth column of Table 1. We then use equations (28), (31) and (32) to calculate the impulse response function for both  $\theta(\gamma_N)$  and  $\theta(\gamma_{GD})$

$$\begin{aligned} IRF(\theta(\gamma_N)) &= (vec(B_1(\gamma_N))', \dots, vec(B_p(\gamma_N))')' \\ IRF(\theta(\gamma_{GD})) &= (vec(B_1(\gamma_{GD}))', \dots, vec(B_p(\gamma_{GD}))')' \end{aligned}$$

where  $B_j(\gamma_N) = J\bar{\Delta}(\gamma_N)^{j-1} J' A(\gamma_N) \Upsilon(\gamma_N)$ ,  $B_j(\gamma_{GD}) = J\bar{\Delta}(\gamma_{GD})^{j-1} J' A(\gamma_{GD}) \Upsilon(\gamma_{GD})$  for  $j = 1, \dots, p$ ,  $J = \begin{bmatrix} I_{dy} & 0_{dy \times (p-1)dy} \end{bmatrix}$ , finally, the matrices  $\bar{\Delta}(\gamma_N)$  and  $\bar{\Delta}(\gamma_{GD})$  are obtained after replacing in the matrix  $\bar{\Delta}$  – Expression (5) – the matrices  $\Delta_j$  with  $\Delta_j(\gamma_N)$  and  $\Delta_j(\gamma_{GD})$ , respectively. These steps are repeated 10000 times and Figures 1 and 2 plot the prior mean and the 16%-84% confidence interval of  $IRF(\theta(\gamma_N))$  – black solid line and shaded area, respectively – against the same moments of  $IRF(\theta(\gamma_{GD}))$  – red dashed and red dashed-cross lines, respectively.

From these charts it seems fair to conclude that the prior distribution of  $IRF(\theta(\gamma_N))$  almost coincides with the prior distribution  $IRF(\theta(\gamma_{GD}))$  and our simplifying assumption does not ‘harm’ the prior distribution of  $\theta(\gamma_N)$  and, consequently, the prior distribution of  $IRF(\theta(\gamma_N))$ .<sup>6</sup>

<sup>4</sup>Similar to Ingram and Whiteman (1994) and De Jong et al. (1993), however, in contrast to Del Negro and Schorfheide (2004), our methodology does not offer a device of controlling the tightness of the prior such as the  $\lambda$  hyperparameter in Del Negro and Schorfheide (2004). An anonymous referee suggested us a “way” of extending our analysis to allow for such hyperparameter that we are going to investigate in a future work.

<sup>5</sup>The exact steps of this Monte Carlo exercise are described in Appendix A.

<sup>6</sup>Instead of presenting 175 histograms – the total number of VAR parameters,  $3 \times 7^2 + 0.5 \times 7 \times (7 + 1)$  – that plot the prior distribution of the individual components  $\theta(\gamma_N)$  against the elements of  $\theta(\gamma_{GD})$  we thought that it would be more constructive for the reader to present the impulse response function as it best summarises all the VAR parameters. However, all the graphs are available from the authors upon request.

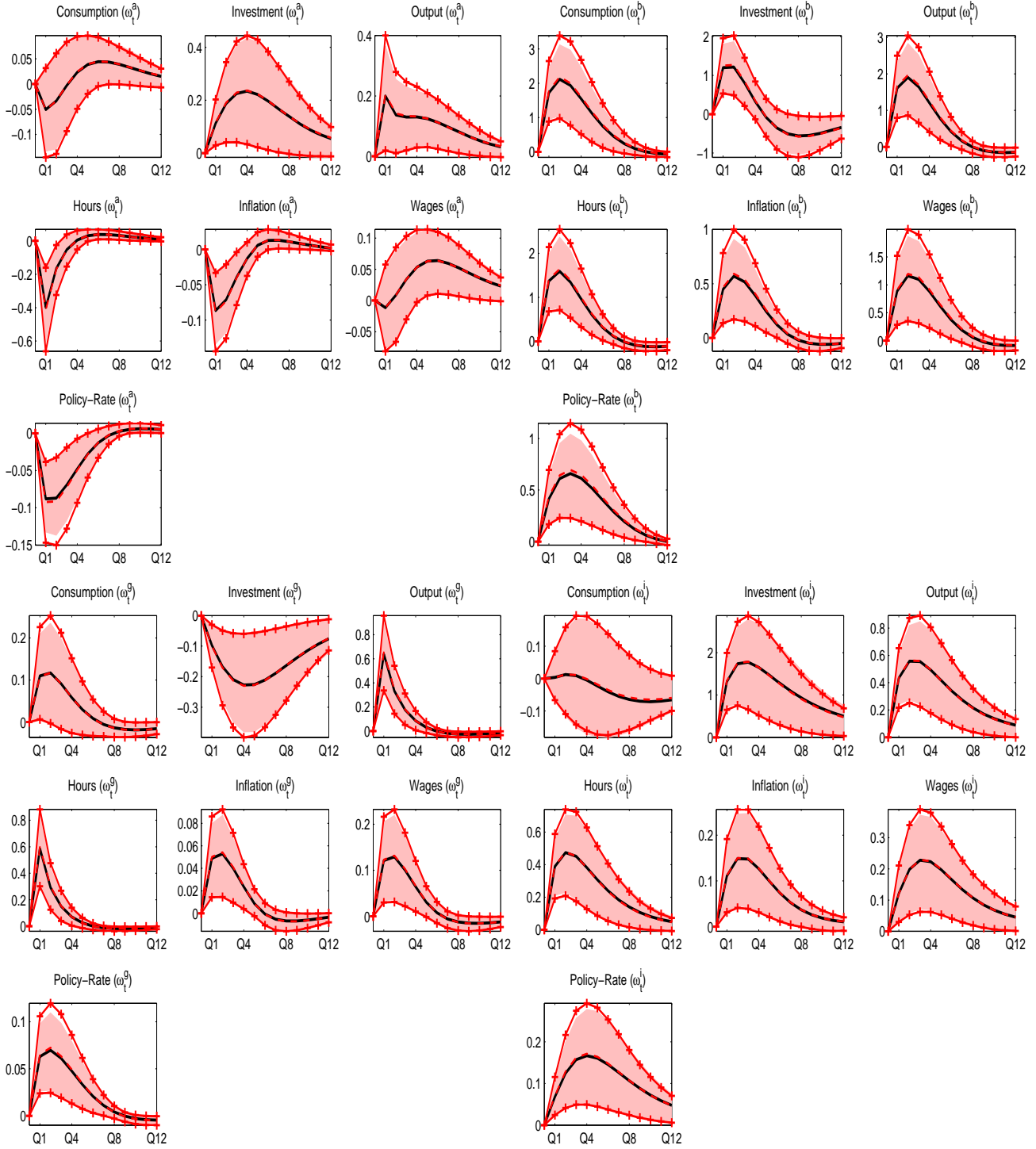


Figure 1: Simulated Impulse Responses Prior Distribution I:  $(\omega_t^a)$  denotes the responses to a productivity,  $(\omega_t^b)$  to a preference,  $(\omega_t^g)$  to a government spending and  $(\omega_t^i)$  to an investment specific shock. The solid black line represents the pointwise mean impulse response function, and the shaded area is the corresponding 16th and 84th percentiles drawing  $\gamma_N$  from the prior normal pdf. The red dashed and red dashed-cross lines represents the same moments of the impulse responses function drawing  $\gamma_{GD}$  from more “standard” probability density functions.

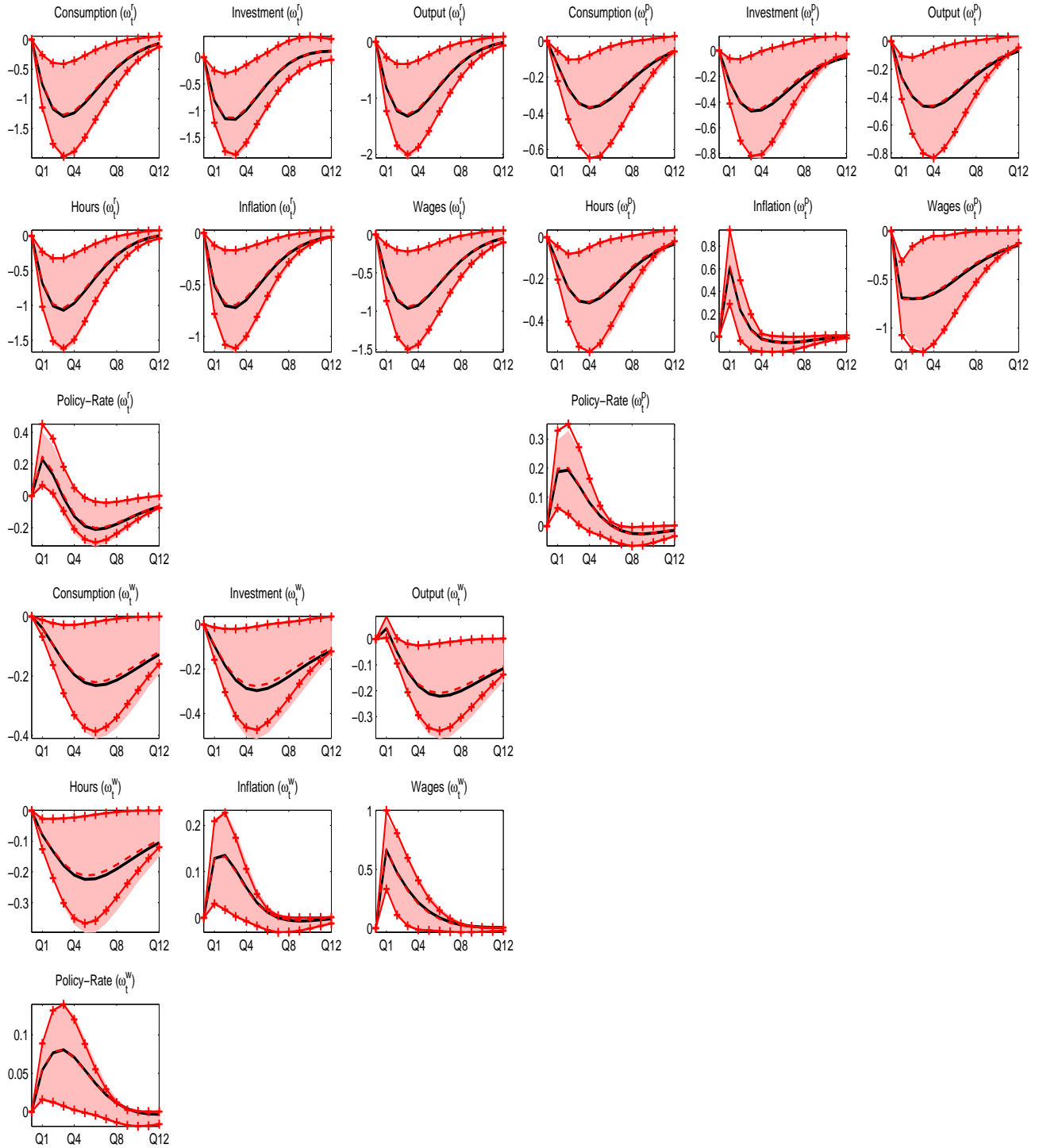


Figure 2: Simulated Impulse Responses Prior Distribution II :  $(\omega_t^r)$  denotes the responses to a monetary policy,  $(\omega_t^p)$  to a prices markup and  $(\omega_t^w)$  to a wages markup shock. The solid black line represents the pointwise mean impulse response function, and the shaded area is the corresponding 16th and 84th percentiles drawing  $\gamma_N$  from the prior normal pdf. The red dashed and red dashed-cross lines represents the same moments of the impulse responses function drawing  $\gamma_{GD}$  from more “standard” probability density functions.

## VAR Posterior Estimation

The posterior distribution of the VAR parameter vector  $\theta$  does not have an analytic form since no conjugate priors are assumed. However, it is well known that the posterior kernel of the VAR parameter vector can be reexpressed as the product of two conditional distributions – see, for instance, Canova (2005) and Koop and Korobilis (2010)

$$p(\theta|Y, \gamma) \propto N(\bar{\mu}_\delta, \bar{\Sigma}_\delta | \Sigma_v) IW(\bar{\Pi}, T + \eta | \Delta) \quad (39)$$

where

$$\bar{\Sigma}_\delta \equiv \left[ \hat{\Sigma}_\delta^{-1} + \Sigma_\delta(\gamma)^{-1} \right]^{-1} \quad (40)$$

$$\bar{\mu}_\delta \equiv \bar{\Sigma}_\delta \left[ \Sigma_\delta(\gamma)^{-1} \mu_\delta(\gamma) + \hat{\Sigma}_\delta^{-1} \hat{\delta} \right] \quad (41)$$

$$\bar{\Pi} \equiv \Pi(\gamma) + T \hat{\Sigma}_v + (\Delta - \hat{\Delta})' X' X (\Delta - \hat{\Delta}) \quad (42)$$

From expressions (40)-(42) it is apparent that in our case the conditional moments are expressed as weighted averages between the DSGE implied VAR prior moments and the OLS estimates. Finally, writing  $p(\theta|Y, \gamma)$  as the product of two conditional distributions allows us to employ the MCMC Gibbs sampler to approximate the posterior distribution of the VAR parameter vector (see, Canova (2005); Koop (2003)).

## DSGE Posterior Estimation

Del Negro and Schorfheide (2004) derive the quasi Bayesian posterior distribution of the DSGE parameter vector –  $\gamma$  – by combining the marginal likelihood function of the VAR model with the prior distribution of the structural parameter vector. Motivated by the work of Christiano et al. (2010b,a) and using the theoretical results developed by Chernozhukov and Hong (2003) and Theodoridis (2011), we are also able to derive the posterior distribution of the DSGE parameter vector. Theodoridis (2011) proposes an impulse-response matching estimator that:

- A)** matches the  $k$ -period responses of the whole vector of the observable variables described by the structural model – caused after a small perturbation to the entire vector of the structural errors – with those observed in the historical data, which have been recovered through the use of a structurally identified vector autoregressive model, and
- B)** minimises the distance between the reduced-form error covariance matrix implied by the structural model and the one estimated in the data<sup>7</sup>

$$\hat{\gamma} = \arg \min m(\gamma | \bar{\theta}) = \arg \min \left\| \mathcal{W}^{1/2} \left[ IRF(\theta(\gamma)) - \widehat{IRF}(\bar{\theta}) \right] \right\| \quad (43)$$

where  $\widehat{IRF}(\bar{\theta})$  is the mean of the posterior distribution of the VAR impulse-response function – discussed earlier,  $\mathcal{W}$  is the inverse of the variance-covariance matrix of the posterior distribution of

<sup>7</sup>Canova and Sala (2009) call this type of estimators ‘full-information’ as it utilises all the available impulse-response information.

VAR impulse-response function and  $\|\cdot\|$  denotes the Euclidean norm.<sup>8</sup>

Theodoridis (2011) shows that this estimator is efficient and from the work of Chernozhukov and Hong (2003) it is known that when a minimum distance estimator is efficient then its quasi Bayesian version

$$p(\gamma|\bar{\theta}) \propto m(\gamma|\bar{\theta})p(\gamma) \quad (44)$$

can be approximated through the random walk Metropolis-Hastings MCMC resampling scheme. The implementation steps are described in Appendix B.

### SVAR Identification

As it is explained in Theodoridis (2011) the identification of the VAR and the estimation of the structural model take place, simultaneously. This idea relies heavily on the fast growing methodology where external instruments are used to identify the VAR structural shocks known as “proxy SVAR” (see Mertens and Ravn (2013a), Mertens and Ravn (2013b), Stock and Watson (2012) and Carriero et al. (2013)).

The starting point of this approach is that the fact that although structural shocks ( $\omega_t$ ) are unobserved, there are available instruments ( $z_t$ ) correlated with the structural disturbances that can be used to identify them. An obvious example of such an instrument is the Romer and Romer (2004) measure of U.S. monetary policy shocks. These instruments can be used to identify the structural shocks through a set of moment conditions. These necessary identification moments/restrictions are given by

$$\begin{aligned} E(\omega_t^i, z_t^j) &= v \neq 0 && \text{if } i = j \\ E(\omega_t^i, z_t^j) &= 0 && \text{otherwise} \end{aligned} \quad (45)$$

In Theodoridis (2011) these instruments are the DSGE disturbances<sup>9</sup>

$$z_t = \omega_t \quad (46)$$

where by using expression (33) we derive the identification matrix implied by the DSGE model. For instance

$$\begin{aligned} E(v_t, z_t') &= E([A(\gamma) \Upsilon(\gamma)] \omega_t, z_t') \\ &= E([A(\gamma) \Upsilon(\gamma)] \omega_t, \omega_t') \\ &= [A(\gamma) \Upsilon(\gamma)] E(\omega_t, \omega_t') \\ &= A(\gamma) \Upsilon(\gamma) \end{aligned}$$

were the second equality is obtained using expression (46). However, matrices  $A(\gamma)$  and  $\Upsilon(\gamma)$  are unknown quantities since  $\gamma$  is not known, meaning that the estimation of the structural model and the identification of the VAR take place simultaneously. Loosely speaking, the identification matrix is obtained by regressing the VAR residuals on the structural errors (Mertens and Ravn, 2013a).

<sup>8</sup>In Theodoridis (2011) the set of instruments employed for the estimation of the structural parameter vector correspond to the impulse response function calculated using the OLS estimate of  $\theta - \widehat{IRF}(\hat{\theta})$  – and  $\mathcal{W}$  is the inverse of the asymptotic variance-covariance matrix of  $\widehat{IRF}(\hat{\theta})$ .

<sup>9</sup>Similarly, Stock and Watson (2012) use the monetary policy shock from Smets and Wouters (2007) model to identify the policy shock in the VAR.



## 6 Applications

### 6.1 DSGE Posterior Estimation

In this subsection we estimate the structural model discussed in Section 3. We use the macroeconomic data set compiled by Smets and Wouters (2007).<sup>10</sup> Due to the absence of any kind of model trend – either deterministic or stochastic – the elimination of the zero frequency component of the non-stationary real output, real consumption and real investment series is achieved by linearly detrending them.<sup>11</sup> The estimation sample is 1966Q1–1999Q4 and the period 2000Q1–2004Q4 is used for the evaluation of the out of sample forecasting performance of the BVARs discussed below.

Figure 3 plots the posterior distribution of the structural parameters – red solid line – against their prior distributions – dashed black line. The first striking observation is that the posterior uncertainty is substantially smaller than the prior one – the peak of  $p(\gamma|\bar{\theta})$  is significantly higher than the peak of  $p(\gamma)$  – and this is true for all parameters. The second interesting feature is that in most cases the posterior mode has shifted away from the prior mode; Table 2 summarises the same information. These two observations seem to suggest that the posterior inference is not driven by prior assumptions and this is a supporting evidence for the proposed methodology.<sup>12</sup>

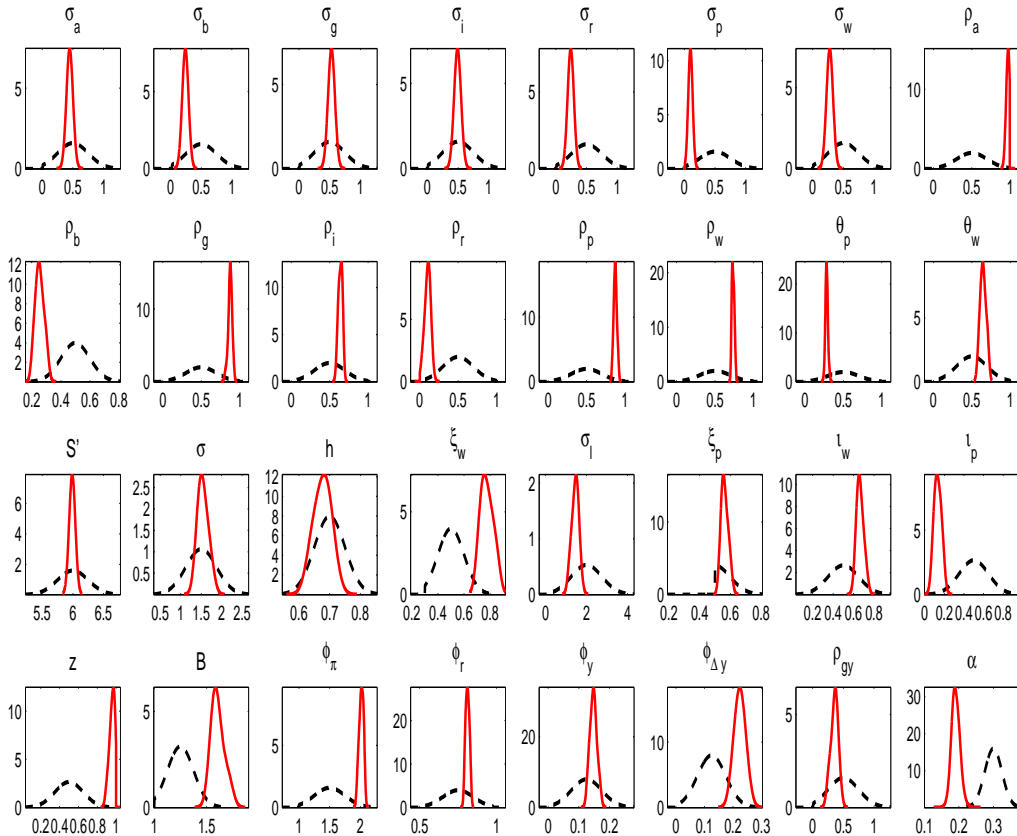


Figure 3: Posterior (red solid line) versus Prior (black dashed line) Structural Parameter Distribution

<sup>10</sup>This data set is publicly available from the website of the American Economic Association.

<sup>11</sup>This choice is consistent with the work of Smets and Wouters (2007).

<sup>12</sup>Canova and Ferroni (2012) employ similar checks to assess the contribution of the prior moments to the posterior DSGE estimates.

Table 2: Structural Parameter Posterior Moments

	Mean	Mean	STD	5%	95%
$\sigma_a$	0.500	0.449	0.019	0.419	0.482
$\sigma_b$	0.500	0.242	0.011	0.223	0.260
$\sigma_g$	0.500	0.530	0.025	0.489	0.572
$\sigma_i$	0.500	0.492	0.026	0.451	0.535
$\sigma_r$	0.500	0.243	0.010	0.227	0.261
$\sigma_p$	0.500	0.100	0.008	0.089	0.114
$\sigma_w$	0.500	0.279	0.020	0.247	0.313
$\rho_a$	0.500	0.973	0.008	0.957	0.985
$\rho_b$	0.500	0.254	0.028	0.210	0.302
$\rho_g$	0.500	0.877	0.024	0.829	0.914
$\rho_i$	0.500	0.640	0.024	0.598	0.673
$\rho_r$	0.500	0.103	0.030	0.045	0.138
$\rho_p$	0.500	0.874	0.020	0.841	0.909
$\rho_w$	0.500	0.736	0.015	0.715	0.764
$\theta_p$	0.500	0.282	0.015	0.258	0.310
$\theta_w$	0.500	0.647	0.035	0.589	0.706
$S'$	6.000	5.990	0.007	5.981	6.001
$\sigma$	1.500	1.544	0.101	1.419	1.728
$h$	0.700	0.676	0.024	0.636	0.712
$\xi_w$	0.500	0.775	0.043	0.713	0.851
$\sigma_l$	2.000	1.449	0.138	1.219	1.676
$\xi_p$	0.500	0.563	0.019	0.539	0.598
$\iota_w$	0.500	0.664	0.028	0.627	0.718
$\iota_p$	0.500	0.133	0.031	0.088	0.181
$z$	0.500	0.958	0.028	0.903	0.996
$B$	1.250	1.606	0.055	1.534	1.722
$\phi_\pi$	1.500	2.015	0.034	1.956	2.063
$\phi_r$	0.750	0.811	0.012	0.793	0.830
$\phi_y$	0.125	0.147	0.011	0.130	0.167
$\phi_{\Delta y}$	0.125	0.224	0.015	0.199	0.250
$\rho_{gy}$	0.500	0.356	0.054	0.259	0.443
$\alpha$	0.300	0.191	0.011	0.173	0.211

We believe that the most prominent element of our approach is the similarity of our posterior estimates – second column of Table 2 – with those from the study of Smets and Wouters (2007) – Tables 1A and 1B. Although we use different type of priors – normal pdf – and limited information estimation techniques we obtain very similar estimates with the latter study. This seems very encouraging for the following reasons:

- it is another evidence – along with simulated prior distribution of  $IRF(\theta(\gamma_N))$ , Figures 1 and 2 – that our assumption about normality regarding the prior distribution of the DSGE parameter vector  $\gamma$  does not distort the posterior inference
- our limited information quasi Bayesian estimator mimics the full-information one and this is important as there are several – both econometric and economic – advantages of using impulse-response matching estimators – see the discussion in Theodoridis (2011)
- our methodology offers a naturally way to assess how well the structural model replicates the dynamics observed in the real world by comparing the impulse responses of the DSGE model

against the VAR. This informal metric has been extensively used in DSGE applied studies— see, for instance, Rotemberg and Woodford (1998) and Christiano et al. (2005)<sup>13</sup>

Figures 4 and 5 display this measure of fit for our exercise, where we plot the mean – solid line – and the 16%-84% confidence interval – shadowed area – of the posterior distribution of the VAR identified impulse response function against the DSGE posterior mean – dashed line. The following discussion illustrates how these charts can be used to improve the fit of the model.

**Productivity Shock:** The model underestimates the persistence of the wealth effect (see, Gali, 1999).

In the data, hours fall persistently below the steady state due to the shock, while the model predicts that they are back to their equilibrium after 2 years. This leads to DSGE responses of higher output and higher consumption provided that the effect on investment is correctly estimated. Moreover, higher supply in the theoretical economy causes larger fall in prices than it actually happens in the real world. The model policymaker responds to weak inflation by cutting interest rates more than what we observe in the real economy as he aims to stimulate the demand and bring inflation back to its target.

**Time Preference Shock:** In the data this shock increases consumption and lowers investment. This is hard to be replicated by the model as this shock enters positively in both the Euler consumption and investment equations. As it is explained by Smets and Wouters (2007) this shock captures – in a reduced-form manner – the financial accelerator mechanism proposed by Bernanke et al. (1999). In other words, in the model this shock by construction co-moves consumption and investment.

**Government Spending Shock:** In the data this shock crowds out consumption and investment by more than it is predicted by theoretical economy. The fiscal authorities in the model run a balanced budget and this is achieved through lump-sum taxes. However, in the real economy the government uses instruments that distort heavily consumption and investment decisions such as VAT and capital taxes, respectively.

**Investment Efficiency Shock:** This shock in the data causes investment to rise but reduces consumption. However, it behaves differently in the model where it raises both consumption and investment and, consequently, leads to higher output than the VAR output response. Inflation in the DSGE model does not increase as much as it is predicted by the VAR model and as a result the theoretical monetary authorities do not need to ‘tight’ policy as much as in the real world.

**Monetary Policy Shock:** Interestingly this shock has significantly larger effect on the real side of the economy on the data than it is actually predicted by the model. However, this is not true for inflation, where VAR evidence suggests that inflation overshoots the steady-state in the second period, while according to the model it returns back to the target monotonically. This probably signals that the structural model misses a working capital channel (see, Christiano et al., 2005), as it would imply that interest rate changes directly affect firms’ marginal cost and puts an upward pressure on inflation.

Finally, the wage and price mark-up shocks are well replicated by the model.

---

<sup>13</sup>Data dynamics are summarised by the impulse responses and they are influenced by the choice of the identification matrix.

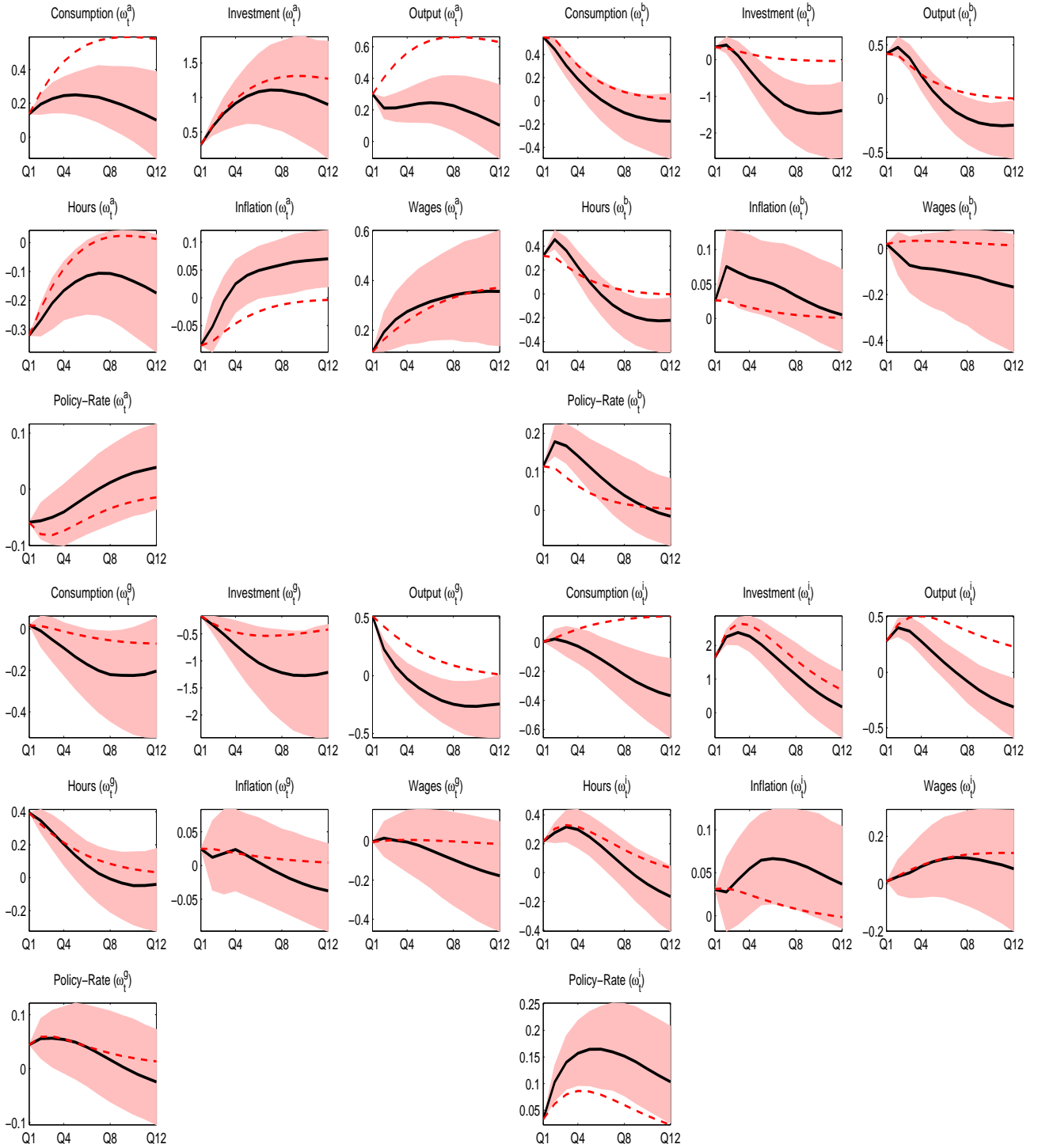


Figure 4: Impulse Responses Posterior Distribution I:  $(\omega_t^a)$  denotes the responses to a productivity,  $(\omega_t^b)$  to a preference,  $(\omega_t^g)$  to a government spending and  $(\omega_t^i)$  to an investment specific shock. The solid black line represents the pointwise mean impulse response function, and the shaded area is the corresponding 16th and 84th percentiles of the BVAR model. The red dashed denotes the posterior mean DSGE impulse response function.

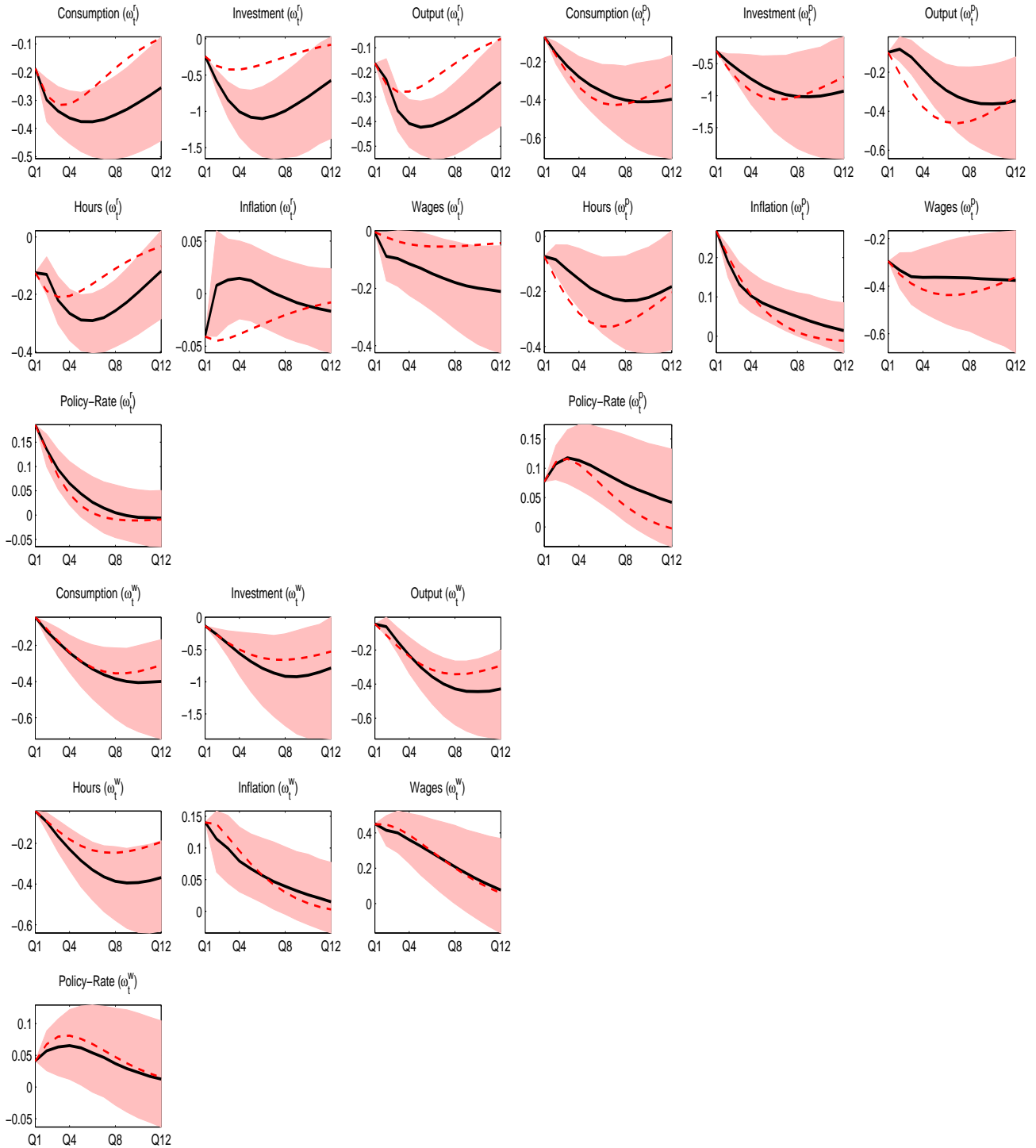


Figure 5: Impulse Responses Posterior Distribution II: ( $\omega_t^r$ ) denotes the responses to a monetary policy, ( $\omega_t^p$ ) to a prices markup and ( $\omega_t^w$ ) to a wages markup shock. The solid black line represents the pointwise mean impulse response function, and the shaded area is the corresponding 16th and 84th percentiles of the BVAR model. The red dashed denotes the posterior mean DSGE impulse response function.

## 6.2 BVAR Analysis

It is time to examine the empirical performance of the priors proposed in this study. We use the structural model considered in Section 3 to derive eight BVAR models with theory driven priors:

1.  $\mathcal{M}0$ : Benchmark DSGE Model
2.  $\mathcal{M}1$ : Flexible wages,  $i_w = 0$  and  $\xi_w = 0.05$
3.  $\mathcal{M}2$ : Flexible prices,  $i_p = 0$  and  $\xi_p = 0.05$
4.  $\mathcal{M}3$ : Flexible wages and prices,  $i_w = 0$ ,  $\xi_w = 0.05$ ,  $i_p = 0$  and  $\xi_p = 0.05$
5.  $\mathcal{M}4$ : No consumption habits,  $h = 0$
6.  $\mathcal{M}5$ : No consumption habits and flexible wages and prices,  $h = 0$ ,  $i_w = 0$ ,  $\xi_w = 0.05$ ,  $i_p = 0$  and  $\xi_p = 0.05$
7.  $\mathcal{M}6$ : No capital utilisation,  $z = 0$
8.  $\mathcal{M}7$ : No capital utilisation and flexible wages and prices,  $z = 0$ ,  $i_w = 0$ ,  $\xi_w = 0.05$ ,  $i_p = 0$  and  $\xi_p = 0.05$

Finally, we augment this set with Minnesota priors –  $\mathcal{M}$ .<sup>14</sup>

In this exercise we want to investigate, i) which BVAR model fits better the past data and ii) which time-series model delivers more accurate out-of-sample forecasts.

### Marginal-Likelihood Evaluation

The first row of Table 3 illustrates the log marginal-likelihood of all estimated BVAR models, the second line reports their posterior probabilities –  $\Pr_j \equiv \frac{m_j(Y)}{\sum_j m_j(Y)}$ , where  $j = \mathcal{M}0, \mathcal{M}1, \mathcal{M}2, \mathcal{M}3, \mathcal{M}4, \mathcal{M}5, \mathcal{M}6, \mathcal{M}7$  and  $\mathcal{M}$  (see, Koop, 2003) and the Bayes factors  $\frac{m_j(Y)}{m_{\mathcal{M}}(Y)}$  and  $\frac{m_{\mathcal{M}0}(Y)}{m_j(Y)}$  are presented in lines three and four, respectively.<sup>15</sup> Two messages can be drawn from Table 3:

Table 3: Log Marginal-Likelihood & Posterior Probabilities

	$\mathcal{M}0$	$\mathcal{M}1$	$\mathcal{M}2$	$\mathcal{M}3$	$\mathcal{M}4$	$\mathcal{M}5$	$\mathcal{M}6$	$\mathcal{M}7$	$\mathcal{M}$
$\log m(Y)$	-801.795	-885.019	-850.567	-947.942	-806.082	-891.875	-806.081	-955.7244	-958.589
$Pr$	0.973	0.000	0.000	0.000	0.013	0.000	0.013	0.000	0.000
$\frac{m_j(Y)}{m_{\mathcal{M}}(Y)}$	1.25E+68	8.94E+31	8.20E+46	4.21E+04	1.71E+66	9.41E+28	1.71E+66	1.76E+01	1
$\frac{m_{\mathcal{M}0}(Y)}{m_j(Y)}$	1	1.39E+36	1.52E+21	2.96E+63	72.835	1.32E+39	72.786	7.10E+66	1.25E+68

### I. DSGE priors do not harm the fit of the time-series model

All statistics seem to suggest that theory based priors do not deteriorate the empirical performance of the VAR model. Posterior probabilities and Bayes factor quantities – in-sample fit statistics – do not seem to support the “orthogonality” assumption imposed by Minnesota priors on macroeconomic data. For example, a truly bayesian econometrician would never use a unique model to draw his inference but this would be based on the weighted average of all elements

<sup>14</sup>The details about Minnesota priors can be found in Appendix C.

<sup>15</sup>The marginal-likelihood in this study has been approximated using Geweke’s modified harmonic mean estimator (Geweke, 1999). The calculation step can be found in either An and Schorfheide (2007) or Schorfheide (2000)

of the model set – nine BVAR models in our case – and the weights used for this aggregation would be the posterior probabilities. The second line of Table 3 seems to indicate that – in this example – the contribution of the BVAR model with Minnesota priors to the total outcome would be almost zero. Additionally, the third line, which reports the Bayes factor that compares the in-sample fit of the BVAR models with theoretical priors against the one with Minnesota priors, seems again to suggest that time-series models with DSGE driven priors achieve higher in-sample empirical performance.

## II. Wage and price rigidities do increase the fit of the time-series model

We know from the work of Christiano et al. (2005) and Smets and Wouters (2007) that wage and price frictions are important DSGE modelling devices that help structural models to match the properties of the aggregated macroeconomic data. The time-series evidence presented in this section point again towards this direction, as labour and price rigidities seem to help the VAR model to achieve higher in-sample fit.

The analysis so far examines the in-sample performance of the time-series models considered in this study and – sometimes – this is not representative for their out-of-sample forecasting performance.

### Univariate Out-of-Sample Forecast Evaluation

Chart 6 plots the mean-square-forecast-error for one, four, eight and twelve-quarters-ahead forecasts calculated using the posterior mean of the VAR parameter vector. Just to remind the reader that all models are estimated over the period between 1966Q1 and 1999Q4 and the period between 2000Q1 and 2004Q4 is used to calculate the out of sample measures of fit. The univariate measure considered in this section suggests that no model delivers superior forecasts. This is important because the use of Minnesota priors is usually motivated on the basis that these theory free priors deliver better out-of-sample forecasting fit.

Focusing on the labour market variables – hours and real wages – we see two very interesting features:

- Minnesota priors deliver the less accurate forecasts for these series
- Flexible price-wage priors provide significantly better forecasts for hours and wages<sup>16</sup>

### Multivariate Out-of-Sample Forecast Evaluation

The same results hold even when multivariate measures of forecasting performance are used. Under the assumption that one, four, eight and twelve quarters ahead forecasts –  $\mathbb{E}_T(y_{T+h}|\mathcal{T})$ , where  $h = 1, 4, 8$  and  $12$  and  $\mathcal{T} = \mathcal{M}0, \mathcal{M}1, \mathcal{M}2, \mathcal{M}3, \mathcal{M}4, \mathcal{M}5, \mathcal{M}6, \mathcal{M}7$  and  $\mathcal{M}-$  are normally distributed we are able to use the predictive density function

$$pdf(y_{T+h}|\mathcal{T}) \equiv (2\pi)^{-dy/2} |\Sigma_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}|^{-1/2} \exp\left(-\frac{1}{2}\left(y_{T+h} - \mu_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}\right)' \Sigma_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}^{-1} \left(y_{T+h} - \mu_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}\right)\right) \quad (47)$$

---

<sup>16</sup>This is an interesting evidence that deserves further research that goes beyond the scope of this study.

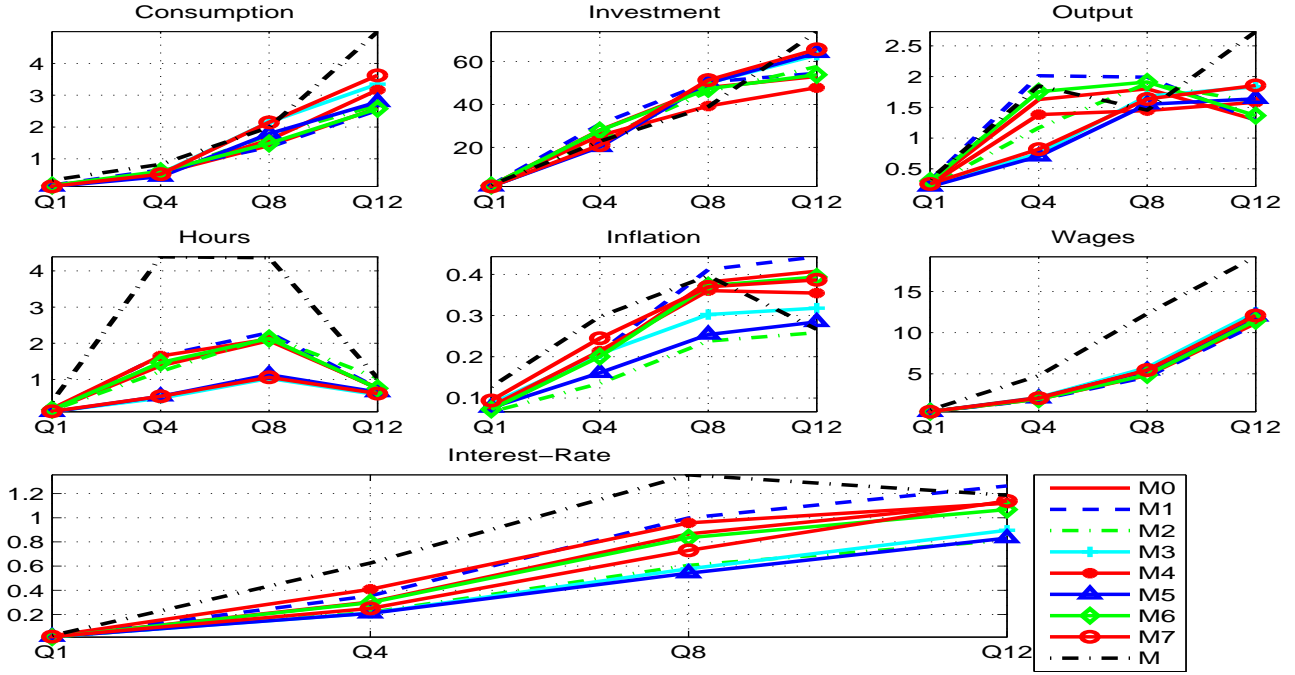


Figure 6: Mean Square Forecast Error

to construct forecast probability weights

$$\Pr(y_{T+h}|\mathcal{T}) \equiv \frac{pdf(y_{T+h}|\mathcal{T})}{\sum_{\mathcal{T}} pdf(y_{T+h}|\mathcal{T})} \quad (48)$$

which illustrate how likely the out-run is, conditional on model  $\mathcal{T}$ . If we put ourselves into truly Bayesian econometrician's shoes, then we need some weights to construct the average forecast of all these seven BVAR models. These weights can be the posterior probabilities calculated using the marginal-likelihood – Table 3 – or the one implied by (48). In the second case, we use model's forecasting performance in the current period to construct our next period forecasts.<sup>17</sup>

The first four subplots of Chart 7 illustrate these probabilities for one, four, eight and twelve-quarters-ahead forecasts, respectively, and for the period 2000Q1-2004Q4. Again, all models for all periods – with few exceptions – appear to get the same support from the data and there are no apparent reasons why Minnesota priors should be preferred.

The final measure considered in this study aims to assess the forecasting performance of each model for the entire period – 2000Q1-2004Q4

$$\Pr(Y_h|\mathcal{T}) \equiv \frac{\sum_{\mathcal{T}} pdf(y_{T+h}|\mathcal{T})}{\sum_{\mathcal{T}} \sum_{\mathcal{T}} pdf(y_{T+h}|\mathcal{T})} \quad (49)$$

The fifth subplot of Chart 7 illustrates this metric, where it can be seen that the BVAR model with theoretical priors, which assign zero rigidities to wages and prices, seems to perform significantly better than any other model over this period for forecast horizons greater than four quarters.

Summarising the outcome of this exercise we can conclude that theory motivated priors do not damage the empirical – both in-sample and out-of-sample – performance of the time series model. In fact, some evidence suggests that BVAR models with DSGE priors attain better in and out-of-sample fit.

<sup>17</sup>The steps required to calculate these weights are described in Appendix C.



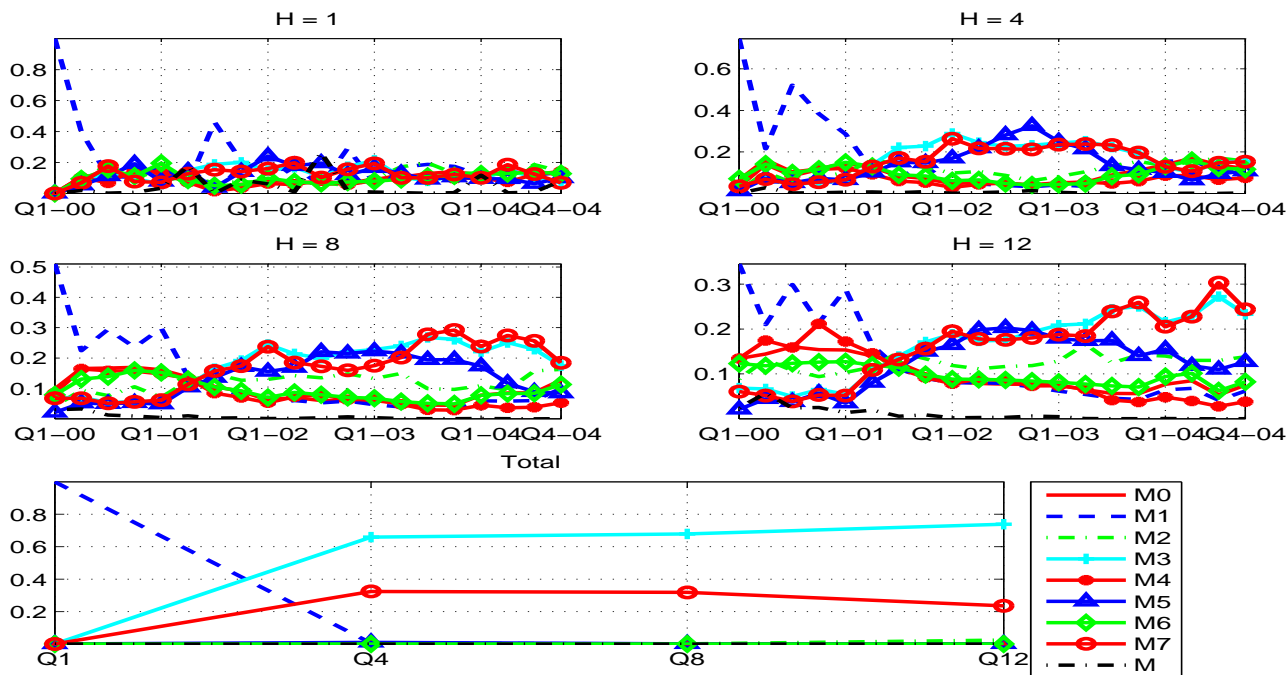


Figure 7: Forecast Time-Varying Probability Weights

## 7 Conclusion

This paper proposes a methodology of constructing DSGE theory consistent prior distributions for BVAR models eliminating theoretical difficulties rising from the use of theory free Minnesota prior distribution. Similar to previous studies, the empirical exercises presented here illustrate that theoretical consistency does not “harm” the forecasting-fit cost of the VAR. Additionally, the marginal-likelihood of the time-series model with theoretical prior can be used to assess the fit of competing DSGE theories that aim to explain the same observed data.

The second contribution of this study is the quasi Bayesian impulse response matching estimator that is used for the posterior estimation of the structural parameter vector. The proposed DSGE estimator shares some similarities with the one introduced by Christiano et al. (2010b,a), however, the one discussed in this study is a “full-information” one – using Canova and Sala (2009) terminology – as it matches all the responses for all observable variables and for all structural shocks. Our estimator is inline with the finding of the work of Canova and Sala (2009) and the analysis of Iskrev (2010) that strongly recommend to “use as many implications of the model as possible to avoid the presence of estimation identification failures”.

## References

- AN, S. AND F. SCHORFHEIDE (2007): “Bayesian Analysis of DSGE Models,” *Econometric Reviews*, 26, 113–172.
- BANDBURA, M., D. GIANNONE, AND L. REICHLIN (2010): “Large Bayesian vector auto regressions,” *Journal of Applied Econometrics*, 25, 71–92.
- BERNANKE, B. S., M. GERTLER, AND S. GILCHRIST (1999): “The financial accelerator in a quantitative business cycle framework,” in *Handbook of Macroeconomics*, ed. by J. B. Taylor and M. Woodford, vol. 1 of *Handbook of Macroeconomics*, chap. 21, 1341–1393.
- CAGLAR, E., J. S. CHADHA, AND K. SHIBAYAMA (2012): “Bayesian Estimation of DSGE Models: Is the Workhorse Model Identified?” TUSYAD-Koc University Economic Research Forum Working Papers 1205, TUSIAD-Koc University Economic Research Forum.
- CANOVA, F. (2005): *Methods for applied macroeconomic research*, Princeton: Princeton University Press.
- CANOVA, F. AND F. FERRONI (2012): “The dynamics of US inflation: Can monetary policy explain the changes?” *Journal of Econometrics*, 167, 47–60.
- CANOVA, F. AND L. SALA (2009): “Back to square one: identification issues in DSGE models,” *Journal of Monetary Economics*, 56, 431–49.
- CARRIERO, A., H. MUMTAZ, K. THEODORIDIS, AND A. THEOPHILOPOULOU (2013): “The Impact of Uncertainty Shocks under Measurement Error. A Proxy SVAR Approach,” Working Papers 707, Queen Mary, University of London, School of Economics and Finance.
- CHERNOZHUKOV, V. AND H. HONG (2003): “An MCMC approach to classical estimation,” *Journal of Econometrics*, 115, 293–346.
- CHIB, S. (1995): “Marginal Likelihood from the Gibbs Output,” *Journal of American Statistical Association*, 90, 1313–1321.
- CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): “Nominal Rigidities and the Dynamic Effects of a shock to Monetary Policy,” *Journal of Political Economy*, 113, 1–45.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1998): “Monetary Policy Shocks: what Have We Learned and to What End?” NBER Working Paper 6400, National Bureau of Economic Research, Inc.
- CHRISTIANO, L. J., M. EICHENBAUM, AND R. VIGFUSSON (2006): “Assessing Structural VARs,” NBER Working Paper 12353, National Bureau of Economic Research, Inc.
- CHRISTIANO, L. J., M. TRABANDT, AND K. VALENTIN (2010a): “DSGE Models for Monetary Policy Analysis,” in *Handbook of Monetary Economics*, ed. by B. M. Friedman and M. Woodford, vol. 3, 285–367.
- (2010b): “Involuntary Unemployment and the Business Cycle,” NBER Working Paper 15801, National Bureau of Economic Research, Inc.

- DE JONG, D., B. INGRAM, AND C. WHITEMAN (1993): “Analyzing VARs with Monetary Business Cycle Model Priors,” *Proceedings of the American Statistical Association, Bayesian Statistics Section*, 160–9.
- DEL NEGRO, M. AND F. SCHORFHEIDE (2004): “Priors from General Equilibrium Models for VARs,” *International Economic Review*, 45, 643–673.
- DOAN, T., R. LITTERMAN, AND C. SIMS (1984): “Forecasting and conditional projection using realistic prior distributions,” *Econometric Reviews*, 3, 1–100.
- FERNANDEZ-VILLAVARDE, J., J. RUBIO-RAMIREZ, T. SARGENT, AND M. WATSON (2007): “ABCs (and Ds) of Understanding VARs,” *American Economic Review*, 97, 1,021–26.
- GALI, J. (1999): “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?” *American Economic Review*, 89, 249–271.
- GEWEKE, J. (1999): “Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication,” *Econometric Reviews*, 18, 1–73.
- GEWEKE, J. F. (2005): *Contemporary Bayesian Econometrics and Statistics*, Wiley Series in Probability and Statistics.
- HANSEN, G. (1985): “Indivisible Labor and the Business Cycle,” *Journal of Monetary Economics*, 16, 281–308.
- INGRAM, B. AND C. WHITEMAN (1994): “Supplanting the ‘Minnesota’ Prior Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors,” *Journal of Monetary Economics*, 34, 497–510.
- ISKREV, N. (2010): “Local identification in DSGE models,” *Journal of Monetary Economics*, 57, 189–202.
- KADIYALA, K. R. AND S. KARLSSON (1997): “Numerical Methods for Estimation and Inference in Bayesian VAR-Models,” *Journal of Applied Econometrics*, 12, 99–132.
- KING, R., C. PLOSSER, AND S. REBELO (1988): “Production, Growth, and Business Cycles: I. The Basic Neoclassical Model,” *Journal of Monetary Economics*, 21, 195–232.
- KLEIBERGEN, F. AND S. MAVROEIDIS (2011): “Identification issues in Bayesian analysis of structural macroeconomic models with an application to the Phillips curve,” mimeo.
- KOOP, G. (2003): *Bayesian econometrics*, Chichester, England: Wiley & Sons.
- KOOP, G. AND D. KOROBILIS (2010): “Bayesian Multivariate Time Series Methods for Empirical Macroeconomics,” *Foundations and Trends in Econometrics*, 3, 267–358.
- LEWIS, R. AND G. REINSEL (1985): “Prediction of Multivariate Time Series by Autoregressive Model Fitting,” *Journal of Multivariate Analysis*, 16, 393–411.
- LITTERMAN, R. (1980): “Techniques for Forecasting with Vector Autoregressions,” Phd, University of Minnesota.

- (1986): “Forecasting with Bayesian Vector Autoregressions - Five Years of Experience,” *Journal of Business and Economic Statistics*, 4, 25–38.
- MAGNUS, J. AND H. NEUDECKER (2002): *Matrix Differential Calculus with Application in Statistics and Econometrics*, New York: WILEY.
- MERTENS, K. AND M. O. RAVN (2013a): “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States,” *American Economic Review*, forthcoming.
- (2013b): “A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers,” *Journal of Monetary Economics*, forthcoming.
- POIRIER, D. J. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach*, The MIT Press.
- RAVENNA, F. (2007): “Vector Autoregressions and Reduced Form Representations of DSGE Models,” *Journal of Monetary Economics*, 54, 2,048–64.
- ROMER, C. D. AND D. H. ROMER (2004): “A New Measure of Monetary Shocks: Derivation and Implications,” *American Economic Review*, 94, 1055–1084.
- ROTEMBERG, J. J. AND M. WOODFORD (1998): “An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: expanded Version,” NBER Technical Working Paper 233, National Bureau of Economic Research, Inc.
- SCHORFHEIDE, F. (2000): “Loss function based evaluation of DSGE models,” *Journal of Applied Econometrics*, 15, 645–670.
- SIMS, C. A. AND T. ZHA (1998): “Bayesian Methods for Dynamic Multivariate Models,” *International Economic Review*, 39, 949–68.
- SMETS, F. AND R. WOUTERS (2007): “Shocks and Frictions in US Business Cycles: a Bayesian DSGE Approach,” *American Economic Review*, 97, 586–606.
- STOCK, J. H. AND M. W. WATSON (2012): “Disentangling the Channels of the 2007-2009 Recession,” NBER Working Papers 18094, National Bureau of Economic Research, Inc.
- THEIL, H. AND A. S. GOLDBERGER (1961): “On Pure and Mixed Statistical Estimation in Economics,” *International Economic Review*, 2, 65–78.
- THEODORIDIS, K. (2011): “An efficient minimum distance estimator for DSGE models,” Bank of England working papers 439, Bank of England.
- WHITE, H. (2001): *Asymptotic Theory for Econometricians*, San Diego: Academic Press.

## A Simulation

### STEPS:

1. We draw  $\gamma_N$  from the normal distribution and its first and second moments are given by the third and fourth column of Table 1, respectively.

2. We draw  $\gamma_G$  from more “DSGE standard” probability density distributions, which are reported in the last column of Table 1. The first two moments  $\gamma_G$  are the same with those of  $\gamma_N$ .
3. We use  $\gamma_N$  and equations (28), (31) - (32) to construct  $\Delta(\gamma_N)_i$ ,  $A(\gamma_N)$   $\Upsilon(\gamma_N)$
4. We do the same with  $\gamma_G$ ,  $\Delta(\gamma_G)_i$  and  $A(\gamma_G)$   $\Upsilon(\gamma_N)$

## B Random-Walk Implementation

This implementation steps follow closely An and Schorfheide (2007)

### STEPS:

1. Use a numerical optimization routine to maximize  $\ln m(\gamma|\bar{\theta}) + \ln p(\gamma)$ . Denote the posterior mode by  $\bar{\gamma}$ .
2. Let  $\tilde{\Sigma}$  be the inverse of the Hessian computed at the posterior mode  $\bar{\gamma}$ .
3. Draw  $\gamma^{(0)}$  from  $N(\bar{\theta}, c_0^2 \tilde{\Sigma})$  or directly specify a starting value.
4. For  $s = 1, \dots, n_{sim}$ , draw  $\gamma$  from the proposal distribution  $N(\gamma^{(s-1)}, c^2 \tilde{\Sigma})$ . The jump from  $\gamma^{(s-1)}$  is accepted ( $\gamma^{(s)} = \gamma$ ) with with probability  $\min \left\{ 1, \frac{m(\gamma|\bar{\theta})p(\gamma)}{m(\gamma^{(s-1)}|\bar{\theta})p(\gamma^{(s-1)})} \right\}$

The constant  $c$  is calibrated to deliver an acceptance rate about 25%.

## C Forecast Probability Weights

For each  $T = 2000Q1, \dots, 2004Q4$  and for each  $\mathcal{T} = \mathcal{M}0, \dots, \mathcal{M}5$

1. Draw  $\Delta$  and  $\Sigma_v$  from the posterior distribution of  $p(\theta|Y, \mathcal{T})$
2. Draw  $v_{T+h}$  from  $N(0, \Sigma_v)$ , were  $h = 1, \dots, 12$
3. Use  $\Delta$ ,  $v_{T+h}$  and  $y_T$  to simulate  $\tilde{y}_{T+h}^j$  from the VAR model
4. Repeat Steps 1-3  $J$  times
5. Use  $\left\{ \tilde{y}_{T+h}^j \right\}_{j=1}^J$  to calculate  $\mu_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}$  and  $\Sigma_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}$
6. Finally, use the actual data realisation -  $y_{T+h}$ ,  $\mu_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}$ ,  $\Sigma_{\mathbb{E}_T(y_{T+h}|\mathcal{T})}$  and expression (48) to calculate  $pdf(y_{T+h}|\mathcal{T})$

These steps deliver a time series of  $pdf(y_{T+h}|\mathcal{T})$ .

## D Minnesota Priors

The posterior inference is obtained as follows. It is assumed that the prior distribution of the VAR parameter vector has a Normal-Wishart conjugate form

$$\delta|\Sigma_v \sim N(\delta_0, \Sigma_v \otimes \Omega_0), \Sigma_v \sim IW(v_0, S_0). \quad (50)$$

where  $\delta$  is obtained by stacking the columns of  $\Delta$ . The prior moments of  $\delta$  are given by

$$E[(\Delta_k) i, j] = \begin{cases} \delta_i & i = j, k = 1 \\ 0 & \text{otherwise} \end{cases}, \quad Var[(\Delta_k) i, j] = \lambda \sigma_i^2 / \sigma_j^2,$$

and as it is explained by Bandbura et al. (2010) they can be constructed using the following dummy observations

$$Y_D = \begin{pmatrix} \frac{diag(\delta_1 \sigma_1 \dots \delta_N \sigma_N)}{\lambda} \\ 0_{N \times (K-1)N} \\ \dots \\ diag(\sigma_1 \dots \sigma_N) \\ \dots \\ 0_{1 \times N} \end{pmatrix} \quad \text{and} \quad X_D = \begin{pmatrix} \frac{J_K \otimes diag(\sigma_1 \dots \sigma_N)}{\lambda} \\ 0_{N \times NK} \\ \dots \\ 0_{1 \times NK} \end{pmatrix} \quad (51)$$

where  $J_K = diag(1, 2, \dots, K)$  and  $diag$  denotes the diagonal matrix. The prior moments of (50) are just functions of  $Y_D$  and  $X_D$ ,  $\Delta_0 = Y_D X_D' (X_D X_D')^{-1}$ ,  $\Omega_0 = (X_D X_D')^{-1}$ ,  $S_0 = (Y_D - \Delta_0 X_D) (Y_D - \Delta_0 X_D)'$  and  $v_0 = T_D - NK$ . Finally, the hyper-parameter  $\lambda$  controls the tightness of the prior. Since the normal-inverted Wishart prior is conjugate, the conditional posterior distribution of this model is also normal-inverted Wishart (Kadiyala and Karlsson, 1997)

$$\delta | \Sigma_v, Y \sim N(\bar{\delta}, \Sigma_v \otimes \bar{\Omega}), \quad \Sigma_v | Y \sim IW(\bar{v}, \bar{S}), \quad (52)$$

where the bar denotes that the parameters are those of the posterior distribution. Defining  $\hat{\Delta}$  and  $\hat{V}$  as the OLS estimates, we have that  $\bar{\Delta} = (\Omega_0^{-1} \Psi_0 + Y X') (\Omega_0^{-1} + X' X)^{-1}$ ,  $\bar{\Omega} = (\Omega_0^{-1} + X' X)^{-1}$ ,  $\bar{v} = v_0 + T$ , and  $\bar{S} = \hat{\Delta} X X' \hat{\Delta}' + \Delta_0 \Omega_0^{-1} \Delta_0' + S_0 + \hat{V} \hat{V}' - \bar{\Delta} \bar{\Omega}^{-1} \bar{\Delta}'$ .

The values of the persistence –  $\delta_i$  – and the error standard deviation –  $\sigma_i$  – parameters of the AR(1) model are obtained from its OLS estimation. Finally,  $\lambda$  has been set equal to 8.15 and this value maximises the marginal-likelihood over the grid  $\lambda \in (0, 100)$ .

**This working paper has been produced by  
the School of Economics and Finance at  
Queen Mary, University of London**

**Copyright © 2014 Thomai Filippeli and Konstantinos Theodoridis  
All rights reserved**

**School of Economics and Finance  
Queen Mary, University of London  
Mile End Road  
London E1 4NS  
Tel: +44 (0)20 7882 7356  
Fax: +44 (0)20 8983 3580  
Web: [www.econ.qmul.ac.uk/papers/wp.htm](http://www.econ.qmul.ac.uk/papers/wp.htm)**