Theoretical Priors for BVAR Models & Quasi-Bayesian DSGE Model Estimation

Preliminary*

Thomai Filippeli[†]
University of Macedonia

Richard Harrison[‡]
Bank of England

Konstantinos Theodoridis[§]
Bank of England

February 12, 2011

Abstract

We build upon the work of Ingram and Whiteman (1994), De Jong et al. (1993) and Del Negro and Schorfheide (2004) to propose a methodology of constructing Dynamic Stochastic General Equilibrium (DSGE) consistent prior distributions for Bayesian Vector Autoregressive (BVAR) models. Further, motivated by the studies of Del Negro and Schorfheide (2004), Christiano et al. (2010) and the theoretical results developed by Chernozhukov and Hong (2003) we illustrate how the posterior moments of the BVAR parameter vector can be used to obtain the posterior inference with respect to the DSGE model. The moments of the assumed Normal-Inverse Wishart 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 prior density of the DSGE parameter vector. Two data driven hyper-parameters unwind the "intensity" of these theoretical priors avoiding bimodality problems that could possibly arise from the strong disagreement between "tight" priors and the data (see, De Jong et al., 1993). The combination of the VAR marginal likelihood function – approximated using the "Laplace" transform – with the prior distribution of the DSGE parameter vector delivers the posterior distribution of the latter. In line with the results from previous studies, BVAR models with theoretical priors seem to achieve forecasting performance that is comparable – if not better – with the one obtained using theory free "Minnesota" priors (Doan et al., 1984). Finally, a small monte carlo experiment and an empirical exercise reveals very supportive results for the quasi bayesian estimator proposed in this study relatively to the standard full information bayesian maximum likelihood estimator.

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

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

*The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. The authors wish to thank Jana Eklund, Francesca Monti, Tony Yates, Emilio Fernandez-Corugedo, Mumtaz Harron, and Martin Andreasen for helpful comments and insightful discussions.

[†]Email: filtho@uom.gr

[‡]Email: Richard.Harrison@bankofengland.co.uk

[§]Email: Konstantinos.Theodoridis@bankofengland.co.uk

Contents

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 scientist's "knowledge" about the parameter vector. For example, we want to assess the persistence of a macroeconomic variable by estimating an autoregressive model of order one – AR(1) – and it is widely accepted that (detrended) macroeconomic series – such as real consumption, real investment, real wages etc – display high degree of persistence. This information – stationarity and significant positive autocorrelation – could be summarised using the beta distribution – defined between zero and one – for the lag coefficient with a mean that reflects this high rigidity.

The importance of the prior distribution further increases 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 five variable VAR with three lags – like the one used in this study – contains ninety parameters whose estimation poses serious difficulties even with fifty years of quarterly data, i.e. 2.2 observations per parameter. In 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) work exactly in this way and this probably explains why it is always scientists' first choice when the estimation of a macroeconomic bayesian VAR 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 parameters, except those on the main diagonal of the first lag, is equal to zero. In terms of the previous example this means that only five from seventy five autoregressive parameters 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, therefore, are characterised by high degree of comovement. For example, a real business cycle model with a unit root technological process and a persistent labour supply shock (see, Hansen, 1985) predicts strong correlation between model's real observable series (i.e. consumption, investment, wages, output etc), meaning that orthogonality among macro variables cannot easily founded on the basis of the economic theory. This lack of theoretical consistency limits researchers' intuition and, consequently, complicates the posterior inference.

The question that naturally arises is if we can construct prior distributions for BVAR models that do not conflict with the theory. Similar to Ingram and Whiteman (1994) 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 and, consequently, data features are expressed as functions of structural parameters. For example, the hump shaped real consumption response after a monetary policy shock observed in VAR studies (Christiano et al., 1998) can be reproduced by a DSGE model where households form consumption habits (Smets and Wouters, 2007) and by varying the consumption smoothing parameter the researcher controls the peak – or the through – of the shock and, consequently, affects the time series properties of the entire vector of model variables. This mapping between the DSGE parameters and the stylised facts guides model

developers to decide about the prior mean of the structural parameter vector, while the standard deviation aims to capture most of the values used in the literature. Answering to 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.

Certainly, arguing in favour of DSGE driven priors it does not seem rational if we have to sacrifice the fit or/and the forecasting performance of the VAR in order to get some theoretical consistency. A set of metrics – marginal likelihood assessment, mean squared forecast error comparison and predictive density evaluation – is used here to measure the performance of the proposed priors. Similar to the previous studies, all results indicate that theory consistent priors deliver time series properties that are comparable – if not better – to the one obtained using theory free priors.

In Section 4 it is illustrated how the mapping between the DSGE and SVAR models – identified by Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) – and random draws from the prior distribution of the DSGE parameter vector are used to construct the moments of the Normal-Wishart prior distribution of the time series model. Since no conjugate priors are used, the posterior distribution of the VAR parameter vector is approximated by using the monte carlo markov chain (MCMC) Gibbs sampling sampler, where the conditional posterior moments are expressed as weighted averages between the ordinary least square (OLS) VAR estimates and the prior moments (Proposition 4.1). Two hyperparameters – the prior degrees of freedom of the Wishart distribution and the tightness factor of the prior variance-covariance matrix of the VAR coefficient vector – control the weights attached to these DSGE derived prior moments (Remark 4.1). This hyperparameter vector represents scientist's confidence regarding the data plausibility of the DSGE model, which can be decided on the basis of "rules of thumb" or data driven techniques such as the marginal likelihood criterion. Motivated by the work of Del Negro and Schorfheide (2004), the present study adopts the second strategy, since this allows the norm of this vector to be interpreted as a measure of fit of the DSGE model.

Del Negro and Schorfheide (2004) illustrate that the hierarchical structure of these theory driven prior distributions enables the estimation of the posterior distribution of the DSGE parameter vector. This is achieved by combining the marginal likelihood function of the VAR model with the prior distribution of the structural parameter vector. In our case the marginal likelihood does not have an analytic form and it is approximated by using the "Laplace" transform[1] and replacing the unrestricted VAR parameters with the one implied by the DSGE model. The small monte carlo experiment and the empirical exercise undertaken in Section 6, using a version of the model developed by Smets and Wouters (2007), reveal very encouraging results for the proposed limited information estimation relatively to the full information maximum likelihood one.

The paper is organised as follows. The notation needed for this paper is developed next, the existing methodologies of constructing theoretical priors are reviewed in Section 3, the proposed one is described in Sections 4 and 5.1, an application is considered in Section 6 and the final section concludes.

[1] See Koop (2003).

2 Notation

The required notation is developed in this section. To be precise, $\mathbb{R}$ denotes the real line, da indicates the dimension of the vector a, $\mathbb{R}^{da} \equiv \mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}$ is the da-cartesian power of the real line, I_{da} stands for the $(da \times da)$ identity matrix, the vec and $vech$ operators transform a matrix with dimensions $da \times da$ to an $da^2 \times 1$ vector by stacking the columns, and to an $(da(da+1)/2) \times 1$ vector by stacking the elements of and below the main diagonal, respectively. The symbol $\otimes$ denotes the Kronecker product operator, while, $\nabla_a f(a)$ and $\nabla_a^2 f(a)$ represent the matrices of the first and second derivatives of the vector function $f(a)$ with respect to the vector a, respectively. $K_{dm,dn}$ is a commutation matrix, such that for any $dm \times dn$ matrix G, $K_{dm,dn} vec(G) = vec(G')$. The state space representation of a solved (log) linear approximated DSGE model ($\mathcal{M}$) is given by

$$y_t = A(\gamma)\,\xi_t \tag{2.1}$$

$$\xi_t = B(\gamma)\,\xi_{t-1} + \Upsilon(\gamma)\,\omega_t \tag{2.2}$$

where the equation (2.2) describes the evolution of the state vector $\left(\xi_t \in \mathbb{R}^{d\xi}\right)$ of the model, expression (2.1) illustrates the relation between the unobserved state of the economy and the observable variables $\left(y_t \in \mathbb{R}^{dy}\right)$, the vector of the structural shocks $\left(\omega_t \in \mathbb{R}^{d\omega}\right)$ is normally distributed with mean zero and I_{dy} covariance matrix $(N \sim (0, I_{dy}))$ and the elements of the matrices $A(\gamma)$, $B(\gamma)$, and $\Upsilon(\gamma)$ are nonlinear functions of the DSGE paymaster vector, which is also called the structural parameter vector, $\gamma \in \Gamma$. $\Psi(a)$ implies that the quantity Ψ is expressed as a function of a. The VAR(p) model ($\mathcal{T}$) is described by

$$y_t = \sum_{i=1}^{p} \Delta_i y_{t-i} + v_t \tag{2.3}$$

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

$$Y = \Delta X + U \tag{2.4}$$

where $\Delta = \left[\begin{array}{ccc} \Delta_1 & \cdots & \Delta_p \end{array}\right]$ 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 U is the $dy \times T$ matrix of the VAR innovations. $\delta \equiv vec(\Delta)'$ and $\sigma_v \equiv vech(\Sigma_v)'$ are the components of the VAR parameter vector $\theta \equiv \left(\delta', \sigma_v'\right)' \in \Theta$, which is called the reduced-form parameter vector; Γ and Θ are compact subsets of $\mathbb{R}^{d\gamma}$ and $\mathbb{R}^{d\theta}$, respectively. The OLS estimates of Σ_v, Δ, δ, σ_v and θ 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 data's population moments and their estimates. $N(\mu_\alpha, \Sigma_\alpha)$ stands for the normal distribution, where μ_α and Σ_α denote the mean and the covariance matrix of the vector α, respectively. The Wishart and its inverse distributions with η degrees of freedom and Π scale matrix are defined as $W(\Pi, \eta)$ and $IW\left(\Pi^{-1}, \eta\right)$, respectively. $p(a)$ denotes the prior distribution of a, $L_{\mathcal{T}}(Y|\theta)$ and $m_{\mathcal{T}}(Y) \equiv \int L_{\mathcal{T}}(Y|\theta)\, p(\theta)\, d\theta$ stand for the likelihood and marginal likelihood of the VAR, respectively, while the same quantities for the DSGE model are given by $L_{\mathcal{M}}(Y|\gamma)$ and $m_{\mathcal{M}}(Y) \equiv \int L_{\mathcal{M}}(Y|\gamma)\, p(\gamma)\, d\gamma$.

3 Existing Methodologies

The methodologies reviewed in this section are grouped into two categories, the analytic type priors where the mapping between the structural and reduced-form parameter vector has a closed-form and the regression type priors where this relation is approximated by the OLS estimator.

3.1 Analytic Type Priors

Ingram and Whiteman (1994)

Ingram and Whiteman (1994) were the first to construct theoretical priors for BVAR models based on a simple Real Business Cycle (RBC, 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 them to express the state vector as a linear function of the observed series – this is acheived through the use of the Generalized Inverse[2] –

$$\left[\left(A\left(\gamma\right)'A\left(\gamma\right)\right)^{-1}A\left(\gamma\right)'\right]y_t=\xi_t \tag{3.1}$$

meaning that the DSGE system – equations (2.1)-(2.2) – can be rewritten as a VAR(1)

$$\begin{aligned} y_t &= A\left(\gamma\right)B\left(\gamma\right)\left[\left(A\left(\gamma\right)'A\left(\gamma\right)\right)^{-1}A\left(\gamma\right)'\right]y_{t-1}+A\left(\gamma\right)\Upsilon\left(\gamma\right)\varepsilon_t \\ &= \Delta_1\left(\gamma\right)y_{t-1}+C\left(\gamma\right)\varepsilon_t \end{aligned} \tag{3.2}$$

with the last expression to illustrate the mapping between the VAR and DSGE parameter vector

$$\theta(\gamma)\equiv\left(vec\left(\Delta_1\left(\gamma\right)\right)',vec\left(C\left(\gamma\right)C\left(\gamma\right)'\right)'\right)' \tag{3.3}$$

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 γ is normally distributed with mean and variance-covariance matrix equal to μ_γ and Σ_γ, respectively, and the use of the "Mean Value Theorem"(see, White, 2001) imply that $\theta\left(\gamma\right)$ is also normally distributed with mean and variance-covariance matrix given by $\mu_\theta\left(\gamma\right)\equiv\theta\left(\mu_\gamma\right)$ and $\Sigma_\theta\left(\gamma\right)\equiv\nabla_\gamma\theta\left(\mu_\gamma\right)\Sigma_\gamma\nabla_\gamma\theta\left(\mu_\gamma\right)'$, respectively.

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).

Clearly, this procedure is not fully Bayesian and it is not suitable for multivariate analysis, meaning that measures for, either the fit of the time series model, or/and the forecasting uncertainty of the entire vector cannot be easily or/and consistently derived (De Jong et al., 1993).

[2] See, Magnus and Neudecker (2002)

3.2 Regression Type Priors

De Jong et al. (1993)

De Jong et al. identify the shortcuts of the previous methodology – γ is normally distributed and single equation non-bayesian estimation inference – and the problems that arise from them. They proceed by making an assumption about the functional form of $p(\theta)$ – conditional Normal-Inverse Wishart (conjugate) distribution– and derive its moments through stochastic simulation. To be precise, pseudo data –$Y(\gamma)$ and $X(\gamma)$ – of length equal to 1350 is simulated by the DSGE model, a VAR(1) is fitted to this artificial data set and the OLS estimates of the VAR coefficient vector, its variance-covariance matrix and the residual variance-covariance matrix are stored. This exercise is repeated 1000 times and the averages of these three quantities are calculated. Conjugate priors imply an analytic posterior distribution, namely

$$
\begin{aligned}
p(\theta|Y,) &= p\left(\delta|\sigma_v,\widehat{\delta},\gamma\right)p\left(\sigma_v|\widehat{\sigma}_v,\gamma\right) && (3.4)\\
p\left(\delta|Y,\sigma_v,\widehat{\delta},\gamma\right) &= N\left(\bar{\mu}_\delta,\bar{\Sigma}_\delta\right) && (3.5)\\
p\left(\sigma_v^{-1}|Y,\widehat{\sigma}_v,\gamma\right) &= IW\left(\bar{\Pi},T+\eta\right) && (3.6)\\
\bar{\Sigma}_\delta &= \Sigma_v\otimes\left(XX'+X(\gamma)X(\gamma)'\right)^{-1} && (3.7)\\
\bar{\mu}_\delta &= \bar{\Sigma}_\delta\left[\left(\Sigma_v^{-1}\otimes XX'\right)\hat{\delta}+\left(\Sigma_v^{-1}\otimes X(\gamma)X(\gamma)'\right)\delta(\gamma)\right] && (3.8)\\
\bar{\Pi} &= \eta\Sigma_v(\gamma)^{-1}+T\widehat{\Sigma}_v \\
&\quad +\left(\widehat{\Delta}-\Delta\right)X(\gamma)X(\gamma)'\left(XX'+X(\gamma)X(\gamma)'\right)^{-1}XX'\left(\widehat{\Delta}-\Delta\right)' && (3.9)
\end{aligned}
$$

where the posterior moments of the VAR parameter vector – expressions (3.7)-(3.9) – are written as weighted averages between the prior moments and the OLS estimates. De Jong et al. do not account for the fact that priors might strongly disagree with the data and, consequently, should not be "dogmatically"[3] imposed since this would damage the posterior inference. Actually, their empirical exercise reveals this problem where most of the posterior autocorrelation distributions are bimodal.

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 to deal with pathologies arising from imposing priors that are not compatible with the data. Further, Del Negro and Schorfheide illustrate that the posterior distribution of the structural parameter vector can be obtained by combining the marginal likelihood of the VAR model with the prior distribution of γ.

To be precise, similar to De Jong et al. the actual data set is augmented with a number of artificial observations simulated by the structural model – $Y(\gamma)\equiv\{Y(\gamma)_t\}_{t=1}^{\lambda_{DS}T}$ and $X(\gamma)\equiv\{X(\gamma)_t\}_{t=1}^{\lambda_{DS}T}$ –, which is proportional to the size of the actual sample – $\lambda_{DS}T$ where $\lambda_{DS}\in(0,\infty)$ – and the VAR likelihood of this augmented sample is factorised into the likelihood of the true data and the likelihood of the artificial one

$$
L_{\mathcal{T}}(Y(\gamma),Y|\theta)=L_{\mathcal{T}}(Y|\theta)\,L_{\mathcal{T}}(Y(\gamma)|\theta) \qquad (3.10)
$$

[3] Terminology used by Del Negro and Schorfheide (2004).

with the latter to be interpreted as the prior density of θ. The authors in order to avoid the stochastic variation, arising by the simulation of the model, replace the non-standarised 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 – $\lambda_{DS}TC_{yy}(\gamma)$, $\lambda_{DS}TC_{xx}(\gamma)$ and $\lambda_{DS}TC_{yx}(\gamma)$. Since the likelihood can be decomposed into the product of the conditional Normal times the Wishart distribution (see, Canova, 2005), then (3.10) can be viewed as the posterior kernel of the VAR parameter vector

$$p(\theta|Y) \propto L_{\mathcal{T}}(Y|\theta)\, p(\delta|\sigma_v,\gamma)\, p(\sigma_v|\gamma) \tag{3.11}$$

with the prior moments be expressed as function of γ

$$\begin{aligned}
p(\sigma_v|\gamma) &\equiv IW(\lambda_{DS}T\Sigma_v(\gamma), \lambda_{DS}T-(pdy+1)) && (3.12)\\
p(\delta|\sigma_v,\gamma) &\equiv N\left(\Delta(\gamma), \Sigma_v\otimes(\lambda_{DS}TC_{xx}(\gamma))^{-1}\right) && (3.13)\\
\Delta(\gamma) &\equiv C_{xx}(\gamma)^{-1}C_{xy}(\gamma) && (3.14)\\
\Sigma_v(\gamma) &\equiv C_{yy}(\gamma)-C_{yx}(\gamma)\Delta(\gamma) && (3.15)
\end{aligned}$$

As it has already been mentioned, conjugate priors imply an analytic posterior distribution for θ – conditional Normal-Inverse Wishart – and its moments are written as weighted average between the DSGE implied moments and the OLS estimates

$$\begin{aligned}
p(\sigma_v|Y,\widehat{\sigma}_v,\gamma) &\equiv IW\left((\lambda_{DS}+1)T\bar{\Sigma}_v(\gamma), (\lambda_{DS}+1)T-(dyp+1)\right) && (3.16)\\
p\left(\delta|Y,\widehat{\delta},\sigma_v,\gamma\right) &\equiv N\left(\bar{\Delta}(\gamma), \Sigma_v\otimes(\lambda_{DS}TC_{xx}(\gamma))^{-1}\right) && (3.17)\\
\bar{\Delta}(\gamma) &\equiv \left(\lambda_{DS}TC_{xx}(\gamma)+X'X\right)^{-1}\left(\lambda_{DS}TC_{xx}(\gamma)\Delta(\gamma)+X'X\widehat{\Delta}\right) && (3.18)\\
\bar{\Sigma}_v(\gamma) &\equiv \frac{1}{(\lambda_{DS}+1)T}\left[\left(\lambda_{DS}TC_{yy}(\gamma)+Y'Y\right)-\left(\lambda_{DS}TC_{yx}(\gamma)+Y'X\right)\bar{\Delta}(\gamma)\right] && (3.19)
\end{aligned}$$

Del Negro and Schorfheide do alert the readers that the empirical performance of the time series model crucially depends on the choice of λ_{DS} and, therefore, they recommend its selection to be based on measurers of fit such as the marginal likelihood.

Del Negro and Schorfheide also illustrate that the posterior distribution of the DSGE parameter vector can be obtained by combining the marginal likelihood of the VAR – in their case it has an analytic form – with the prior distribution of γ

$$p\left(\gamma|Y,\widehat{\theta},\lambda_{DS}\right) \propto m_{\mathcal{T}}\left(Y|\gamma,\widehat{\theta},\lambda_{DS}\right)p(\gamma) \tag{3.20}$$

This is clearly a major improvement over all the earlier methodologies since it enables the posterior estimation of γ to be based on limited information techniques. To be precise, $p\left(\gamma|Y,\widehat{\theta},\lambda_{DS}\right)$ consists of values of γ that minimise the distance between the OLS estimated VAR parameter vector and the one implied by the DSGE model; loosely speaking, this could be viewed as the Bayesian version of the estimator proposed by Smith (1993).

The theoretical prior moments in the last two methodologies are constructed by fitting a VAR(p) model in the data simulated by the structural model. This is not exactly the prior distribution of the reduced-form parameter vector $\theta(\gamma)$ but the distribution of the OLS VAR estimate $\widehat{\theta}_{\lambda_{DS}}(\gamma)$

under the assumption that the DSGE model is the true data generation process. Certainly, the consistency property of the OLS estimator ensures that $\widehat{\theta}_{\lambda_{DS}}(\gamma)$ converges in probability to $\theta(\gamma)$ as the size of the pseudo sample tends to infinite – $\widehat{\theta}_{\lambda_{DS}}(\gamma) - \theta(\gamma) \xrightarrow{\Pr} 0_{d\theta}$ as $\lambda_{DS} \to \infty$ –, however, the variance-covariance matrix $\Sigma_v \otimes (\lambda_{DS} TC_{xx}(\gamma))^{-1}$ 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 – $\mathbb{E}\left(\widehat{\theta}_{\lambda_{DS}}(\gamma) - \theta(\gamma)\right)\left(\widehat{\theta}_{\lambda_{DS}}(\gamma) - \theta(\gamma)\right)'$, which converges to zero as $\lambda_{DS} \to \infty$. This implies that information captured by the second moments of $p(\gamma)$ is not fully utilised by these approaches. The variance of γ reflects scientist's confidence about μ_γ and, very likely, researchers might be more certain about some parameters than others, meaning that some VAR parameters should get more prior weight than others. We believe this is a crucial piece of information that helps researchers to improve the properties of the DSGE model and, therefore, it should not be ignored. For instance, the prior distribution of the VAR responses depends on the variance of θ, which under theory driven priors should also depend on the variance of γ.

Additionally, the functional form of the variance-covariance matrix – $\Sigma_v \otimes (\lambda_{DS} TC_{xx}(\gamma))^{-1}$ – 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), this feature does not seem in line with the DSGE model.

4 The Proposed Method

This section describes the proposed methodology, this falls into the analytic type priors category and it can be seen as a generalisation of Ingram and Whitemans' work. Briefly, it relies on the mapping between the DSGE and SVAR models – Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) – and the prior distribution of the structural parameter vector to derive the prior moments of the VAR parameter vector. Two hyperparameters ensure well behaved posterior inference and, finally, it is illustrated how the BVAR posterior moments can be used to obtain the posterior distribution of the DSGE parameter vector.

4.1 The Prior Moments of θ

The starting point of our methodology is the DSGE model summarised by equations (2.1) and (2.2) and the prior distribution of the structural parameter vector. From the work of Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007) it is known that when the number of shocks coincides with the number of the observable variables and the eigenvalues of the matrix

$$M(\gamma) \equiv \left[I_{d\xi} - \Upsilon(\gamma)\left[A(\gamma)\Upsilon(\gamma)\right]^{-1} A(\gamma)\right] B(\gamma) \tag{4.1}$$

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

$$\phi : \gamma \to \theta \tag{4.2}$$

namely,

$$\begin{aligned}
\delta(\gamma)_i &= vec(\Delta(\gamma)_i) && (4.3)\\
\Delta(\gamma)_i &\equiv A(\gamma) B(\gamma) M(\gamma)^{i-1} \Upsilon(\gamma) [A(\gamma) \Upsilon(\gamma)]^{-1} && (4.4)\\
\sigma(\gamma) &= vec\left([A(\gamma) \Upsilon(\gamma)] [A(\gamma) \Upsilon(\gamma)]'\right) && (4.5)
\end{aligned}$$

The non-linearity of the mapping and the non-normality of $p(\gamma)$ implies that the functional form of $p(\theta(\gamma))$ does not have a closed-form and it cannot be approximated using either the "Mean Value Theorem" – discussed in Ingram and Whitemans' case – or the "Change of Variable Theorem" – ϕ is not an injection (Koop, 2003, pp. 334).

The analysis proceeds with the assumption that the pdf of $\theta(\gamma)$ is the Normal-Inverse Wishart (not conjugate) distribution and its moments are approximated through stochastic simulation. To be precise, using equations (4.1)-(4.5) and random draws from $p(\gamma)$ we are able construct a pseudo set of identically independently distributed (i.i.d) draws of the reduced-form VAR parameter vector, $\{\theta_j\}_{j=1}^{S}$. Then form Theorem 3.1 and Proposition 3.2 of White (2001) it is known that the estimated moments converge to the true one "Almost Sure"

$$\begin{aligned}
\widehat{\mu}_\delta(\gamma) &\equiv S^{-1} \sum_{j=1}^{S} \delta(\gamma)_j \overset{a.s.}{\to} \mu_\delta(\gamma) && (4.6)\\
\widehat{\mu}_\sigma(\gamma) &\equiv S^{-1} \sum_{j=1}^{S} \sigma(\gamma)_j \overset{a.s.}{\to} \mu_\sigma(\gamma) && (4.7)\\
\widehat{\Sigma}_\delta(\gamma) &\equiv S^{-1} \sum_{j=1}^{S} \left(\delta(\gamma)_j - \widehat{\mu}_\delta\right)\left(\delta(\gamma)_j - \widehat{\mu}_\delta\right)' \overset{a.s.}{\to} \Sigma_\delta(\gamma) && (4.8)
\end{aligned}$$

The steps are fully described in the Appendix B.

From the properties of the Inverse Wishart distribution (Poirier, 1995) it is known that the scale matrix Π is related to $\mu_\sigma(\gamma)$ through the following relationship

$$\mu_\sigma(\gamma) = \frac{1}{\eta - dy - 1} vec(\Pi) \text{ or } \Pi(\gamma) = (\eta - dy - 1)\Pi^*(\gamma) \quad (4.9)$$

where $vec(\Pi^*(\gamma)) \equiv \mu_\sigma(\gamma)$. In order to study the variance of the posterior distribution of Σ_v we need work with the properties of the Wishart distribution since Magnus and Neudecker (1979) provide an analytic expression for the second moment of the Wishart distribution. The mapping between the Inverse-Wishart and the Wishart distribution is given by Theorem 3.5.5 of Poirier (1995)

$$\Sigma_{\sigma^{-1}}(\gamma) \equiv \frac{1}{\eta - dy - 1}\left(I_{dy^2} + K_{dy,dy}\right)\left((\Pi^*(\gamma))^{-1} \otimes (\Pi^*(\gamma))^{-1}\right) \quad (4.10)$$

Expressions (4.1)-(4.10) establish the mapping between the moments of $p(\gamma)$ and $p(\theta(\gamma))$.

4.2 No "Dogmatic" Priors

It was explained in the introduction that DSGE models should be viewed as devices where stylised facts can be decomposed to agents' optimisation problems, meaning that it should not be expected to

explain all the features of the observed data. DSGE models, therefore, impose a large – and some times very severe – amount of restrictions on the data and this subsection explains how these constraints can be relaxed inside this framework.

Singularity of $\Sigma_\delta(\gamma)$

Structural parameters have – by construction – an economic interpretation and they are not arbitrary added like the VAR ones aiming to fully capture all the structure observed in the real world – we usually increase the number of VAR lags up to the point where the estimated residuals are i.i.d.. This implies that the dimension of γ is expected to be (much) smaller than $\delta(\gamma)$, meaning that Σ_δ is a positive semi-definite matrix. In order to avoid working with singular distributions we need to impose more structure on Σ_δ and this has to be a block diagonal matrix.

The length and, consequently, the number of blocks depends on the number of the structural parameter, for instance, in the example presented below there are twenty five structural parameters and seventy five autoregressive coefficients, meaning that the minimum number of blocks is three – twenty five parameters length, one for each lag – and the maximum one is seventy five – Σ_δ is a diagonal matrix.

Assuming that Σ_δ is diagonal – similar to Minnesota priors – seems a natural choice when we are not very confident about model's predictions regarding the cross-moments between the autoregressive coefficients and we do not wish to impose these restrictions on the data.

Hyper-Parameters

It is clear from the work of De Jong et al. (1993) and Del Negro and Schorfheide (2004) that we need a device that either unwinds the "strength" of the theoretical priors when they are not compatible with the data – ensuring well behave posterior inference – or attaches more weight to them when model's predictions are inline with the true data generation process.

Notice that in the current framework $\mu_\delta(\gamma)$ and $\mu_\sigma(\gamma)$ represent model's predictions about the mean of the VAR coefficient and residual variance-covariance vector, respectively. Since these quantities can be easily estimated from the data – $\widehat{\mu}_\delta$ and $\widehat{\mu}_\sigma$ – we need to ensure that the latter lie inside the support of the distribution implied by the DSGE model, which can be succeeded by varying the variance around $\mu_\delta(\gamma)$ and $\mu_\sigma(\gamma)$.

From equation (4.10) it is clear that by changing the degrees of freedom parameter we can either "shrink" – $\Sigma_{\sigma^{-1}}(\gamma) \to 0_{dy \times dy}$ as $\eta \to \infty$ – or "loose" – $\Sigma_{\sigma^{-1}}(\gamma) \to \infty$ as $\eta \to dy + 1$ – the distribution around $\mu_{\sigma^{-1}}(\gamma)$. Unfortunately, such parameter does not exist for $\Sigma_\delta(\gamma)$, meaning that we need to "invent" one

$$\Sigma_{\lambda\delta}(\gamma) \equiv \lambda \Sigma_\delta(\gamma) \tag{4.11}$$

where $\lambda \in (0, \infty)$.

Remark 4.1 below summarises the role of the hyper-parameter vector $\bar{\lambda} \equiv (\lambda, 1/\eta)'$ in the posterior distribution of θ and, consequently, illustrates why its selection should be based on measures of fit.

However, before this we need study how the posterior distribution of the VAR parameter vector is constructed.

4.3 VAR Posterior Estimation

The posterior distribution of θ does not have an analytic form since no conjugate prior are assumed, however, Proposition 4.1 states that the posterior kernel of the VAR can be rewritten as the product of two conditional distributions, meaning that $p\left(\theta|Y,\widehat{\theta},\gamma,\bar{\lambda}\right)$ can be approximated using the MCMC Gibbs sampling scheme (Canova, 2005).

Proposition 4.1 *If the prior distribution of δ is the normal one with mean and variance-covariance matrix equal to $\mu_\delta(\gamma)$ and $\Sigma_{\lambda\delta}(\gamma)$, respectively, and the prior distribution of Σ_v is the Inverse Wishart with η degrees of freedom and $\Pi(\gamma)$ scale matrix. Then the posterior kernel of θ can be written as the product of two condition distributions*

$$p\left(\theta|Y,\widehat{\theta},\gamma,\bar{\lambda}\right) \propto N\left(\bar{\mu}_\delta,\bar{\Sigma}_\delta|\Sigma_v\right) IW\left(\bar{\Pi},T+\eta|\Delta\right) \tag{4.12}$$

where

$$\begin{aligned}
\bar{\Sigma}_\delta &\equiv \left[\widehat{\Sigma}_\delta^{-1}+\Sigma_{\lambda\delta}(\gamma)^{-1}\right]^{-1} && (4.13)\\
\bar{\mu}_\delta &\equiv \bar{\Sigma}_\delta\left[\Sigma_{\lambda\delta}(\gamma)^{-1}\mu_\delta(\gamma)+\widehat{\Sigma}_\delta^{-1}\widehat{\delta}\right] && (4.14)\\
\bar{\Pi} &\equiv \Pi(\gamma)+T\widehat{\Sigma}_v+\left(\Delta-\widehat{\Delta}\right)' X'X\left(\Delta-\widehat{\Delta}\right) && (4.15)
\end{aligned}$$

□

The proof can be found in Appendix A.

The following Remark highlights the role of $\bar{\lambda}$ in $p\left(\theta|Y,\widehat{\theta},\gamma,\bar{\lambda}\right)$

Remark 4.1 *From Proposition 4.1 it can be seen that:*

1. *The posterior mean of the conditional Normal distribution of the VAR coefficient vector is a weighted average between the prior mean and the OLS estimate*

$$\begin{aligned}
\bar{\mu}_\delta &\equiv \left[\widehat{\Sigma}_\delta^{-1}+\Sigma_{\lambda\delta}(\gamma)^{-1}\right]^{-1}\left[\Sigma_{\lambda\delta}(\gamma)^{-1}\mu_\delta(\gamma)+\widehat{\Sigma}_\delta^{-1}\widehat{\delta}\right]\\
&= \left[\Sigma_{\lambda\delta}(\gamma)^{-1}+\widehat{\Sigma}_\delta^{-1}\right]^{-1}\widehat{\Sigma}_\delta^{-1}\widehat{\delta}+\left[\Sigma_{\lambda\delta}(\gamma)^{-1}+\widehat{\Sigma}_\delta^{-1}\right]^{-1}\Sigma_{\lambda\delta}(\gamma)^{-1}\mu_\delta(\gamma)\\
&= \left[\frac{1}{\lambda}\Sigma_\delta(\gamma)^{-1}+\widehat{\Sigma}_\delta^{-1}\right]^{-1}\widehat{\Sigma}_\delta^{-1}\widehat{\delta}+\lambda\left[\Sigma_\delta(\gamma)^{-1}+\lambda\widehat{\Sigma}_\delta^{-1}\right]^{-1}\frac{1}{\lambda}\Sigma_\delta(\gamma)^{-1}\mu_\delta(\gamma)\\
&= \left[\frac{1}{\lambda}\Sigma_\delta(\gamma)^{-1}+\widehat{\Sigma}_\delta^{-1}\right]^{-1}\widehat{\Sigma}_\delta^{-1}\widehat{\delta}+\left[\Sigma_\delta(\gamma)^{-1}+\lambda\widehat{\Sigma}_\delta^{-1}\right]^{-1}\Sigma_\delta(\gamma)^{-1}\mu_\delta(\gamma)
\end{aligned}$$

   From the latter expression it can be concluded that

$$\begin{aligned}
\bar{\mu}_\delta &\to \widehat{\delta} \quad as\ \ \lambda\to\infty && (4.16)\\
\bar{\mu}_\delta &\to \mu_\delta(\gamma) \quad as\ \ \lambda\to 0 && (4.17)
\end{aligned}$$

2. *Form Theorem A.4.3(b) of Poirier (1995) it is known that*

$$\begin{aligned}\bar{\Sigma}_\delta &\equiv \left[\widehat{\Sigma}_\delta^{-1} + \Sigma_{\lambda\delta}(\gamma)^{-1}\right]^{-1} = \widehat{\Sigma}_\delta - \widehat{\Sigma}_\delta\left[\widehat{\Sigma}_\delta + \Sigma_{\lambda\delta}(\gamma)\right]^{-1}\widehat{\Sigma}_\delta \\ &= \widehat{\Sigma}_\delta - \widehat{\Sigma}_\delta\left[\widehat{\Sigma}_\delta + \lambda\Sigma_\delta\right]^{-1}\widehat{\Sigma}_\delta \qquad (4.18)\end{aligned}$$

Meaning that

$$\bar{\Sigma}_\delta(\gamma) \to \widehat{\Sigma}_\delta \quad as \quad \lambda \to \infty \qquad (4.19)$$

$$\bar{\Sigma}_\delta(\gamma) \to 0_{d\delta\times d\delta} \quad as \quad \lambda \to 0 \qquad (4.20)$$

3. *From the properties of the inverse Wishart distribution is know that the posterior mean of Σ_v is (Poirier, 1995)*

$$\begin{aligned}\bar{\mu}_\sigma &= \frac{\eta - dy - 1}{T + \eta - dy - 1}\mu_\sigma(\gamma) + \frac{T}{T + \eta - dy - 1}vec\left(\widehat{\Sigma}_v\right) \\ &\quad + \frac{1}{T + \eta - dy - 1}vec\left[\left(\Delta - \widehat{\Delta}\right)' X'X\left(\Delta - \widehat{\Delta}\right)\right]\end{aligned}$$

This implies that

$$\bar{\mu}_\sigma \to vec\left(\widehat{\Sigma}_v\right) \ as\ \eta - dy - 1 \to 0 \ and\ \lambda \to \infty \qquad (4.21)$$

$$\bar{\mu}_\sigma \to \mu_\sigma(\gamma) \quad as\ \eta \to \infty \qquad (4.22)$$

4. *From Magnus and Neudecker (1979) and Theorem 3.5.5 of Poirier (1995) it is known that*

$$\begin{aligned}vec\left(\Sigma_{\sigma^{-1}}\right) &= (T + \eta - dy - 1)\left(I_{dy^2} + K_{dy,dy}\right)\left(\bar{\Pi}^{-1}\otimes\bar{\Pi}^{-1}\right) \\ &= (T + \eta - dy - 1)\left(I_{dy^2} + K_{dy,dy}\right) \\ &\quad \begin{pmatrix} \left[(\eta - dy - 1)\Pi^* + T\widehat{\Sigma}_v + \left(\Delta - \widehat{\Delta}\right)' X'X\left(\Delta - \widehat{\Delta}\right)\right]^{-1} \\ \otimes \left[(\eta - dy - 1)\Pi^* + T\widehat{\Sigma}_v + \left(\Delta - \widehat{\Delta}\right)' X'X\left(\Delta - \widehat{\Delta}\right)\right]^{-1} \end{pmatrix}\end{aligned}$$

It can be seen that

$$vec\left(\Sigma_{\sigma^{-1}}\right) \to \frac{1}{T}\left(I_{dy^2} + K_{dy,dy}\right)\left(\widehat{\Sigma}_v^{-1}\otimes\widehat{\Sigma}_v^{-1}\right) \ as\ \eta \to dy + 1 \ and\ \lambda \to \infty \qquad (4.23)$$

$$vec\left(\Sigma_{\sigma^{-1}}\right) \to 0_{d\sigma\times d\sigma} \quad as\ \eta \to \infty \qquad (4.24)$$

□

From Remark 4.1 it seems clear that $\bar{\lambda}$ can be interpreted as a "measure of confidence" regarding the DSGE model. To be precise, when the researcher is highly uncertain about the plausibility of the structural model – $\eta \to dy + 1$ and $\lambda \to \infty$ – then his best strategy is to discount any structural information and to proceed with the OLS estimates. However, the other extreme scenario, if he strongly believes that the DSGE model is the true data generation process – $\eta \to \infty$ and $\lambda \to 0$ – then his analysis should be based on $\theta(\gamma)$ ignoring any data information.

Estimation & Interpretation of $\bar{\lambda}$

The previous section clearly states the necessity of rationally selecting $\bar{\lambda}$ to avoid damaging the empirical fit of the time series model. Since worsening the performance of the VAR model is highly undesirable, data can be used to guide our choice regarding this hyperparameter vector.

The most natural measure for this purpose seems the marginal likelihood of time series model, meaning that $\widehat{\bar{\lambda}}$ should be the one that delivers the highest value

$$\widehat{\bar{\lambda}} = \arg \max_{\bar{\lambda} \in (0,\infty)\times(dy+1,\infty)} m_{\mathcal{T}}\left(Y|\gamma, \widehat{\theta}, \bar{\lambda}\right) \tag{4.25}$$

In our framework $m_{\mathcal{T}}\left(Y|\gamma, \widehat{\theta}, \bar{\lambda}\right)$ does not have an analytic form, however, it can be approximated by the output of the Gibbs smapler using either the methodology developed by Chib (1995) or Geweke's modified harmonic mean estimator (Geweke, 1999), and this increases the complexity of calculating $\widehat{\bar{\lambda}}$. To see this, notice that we need to create a two dimensional grid – $(0,\infty)\times(dy+1,\infty)$ – and to evaluate $m_{\mathcal{T}}\left(Y|\gamma, \widehat{\theta}, \bar{\lambda}\right)$ for all values of $\bar{\lambda}$ inside this grid, having done this we set $\widehat{\bar{\lambda}}$ equal to the $\bar{\lambda}$ that corresponds to $m_{\mathcal{T}}\left(Y|\gamma, \widehat{\theta}, \bar{\lambda}\right)$ with the highest value; see Chart 1.

This estimation procedure implies that $\widehat{\bar{\lambda}}$ can be interpreted as a "measure of fit" that assesses both model's transmission mechanisms and the prior specification of γ. For example, say that the true data generation process is a model where households form consumption habits, a fraction of them set their wages optimally while others use rules of thumb, some of the firms that produce interemediate goods in this economy form their pricing desicions optimally and the others index their prices on lagged inflation. Given this assumption, we can think three reasons why $\widehat{\bar{\lambda}}$ could possibly be large if another model is used to derive prior distribution of the VAR parameter vector. The first one is that one or more of the rigidities are absent, this would imply low order VAR dynamics meaning that some of the elements of $\mu_{\delta}(\gamma)$ would be zero and this does not reconcile with the data $\widehat{\mu}_{\delta}$. Another reason is that the values attached to γ are not the right one, for instance, the slope of the Phillips curve is not the right one, meaning that the impact of monetary policy is not well measured or the slope of labour supply curve is not right misquoting the importance of the substitution effect relative to the wealth one, clearly, this would imply that $\mu_{\theta}(\gamma)$ is very different from $\widehat{\mu}_{\theta}$. Finally, a combination of the two – model's mispecification and wrong calibration – seems the most likely reason for large values of $\widehat{\bar{\lambda}}$.

Another interesting feature of the methodology estimating $\widehat{\bar{\lambda}}$ is that we are able to identify which part of the VAR parameter vector – δ or/and σ – failed to be well summarised by the DSGE model. The curvature of the surface – obtained by evaluating $m_{\mathcal{T}}\left(Y|\gamma, \widehat{\theta}, \bar{\lambda}\right)$ for all values of the $\bar{\lambda}$ grid – indicates whose hyperparameter's small changes can lead to big fit improvements. In the exercise presented below it seems that this the case for Σ_v – extreme curvature for small values of $\eta - dy - 1$ and relative flat with respect to λ_{δ} and large values of η, Chart 1 – and this seems consistent with the Quasi-Bayesian estimation of the DSGE model, where the posterior estimates of standard deviations and most of the persistent parameters of the structural disturbances seem smaller than the one obtained using full information Bayesian maximum likelihood techniques.

5 DSGE Quasi-Bayesian Estimation

5.1 DSGE Estimation

Before the evolution of the MCMC methodology DSGE models were mainly estimated using classical limited information (LMI) methods (Smith, 1993; Rotemberg and Woodford, 1998; Christiano et al., 2005) and this was due to their good small samples properties (Ruge-Murcia, 2007; Theodoridis, 2010), the lack of the requirement the structural model to be viewed as the true data generation process and the non-smoothness of the DSGE likelihood.

As it is explained in Theodoridis (2010), the most attractive feature of the LMI methodology is that you actually observe the targets you are trying to "hit". This allows you to identify the strengths and the weaknesses of the model, once the structural parameters have been estimated, and to derive some useful economic conclusions about the model. This does not seem so obvious in the full information (FLI) case where the estimated vector minimises the distance between the model and the true data generation process and, since that the latter is unknow, this means that we have no idea how big this distance is or/and what this implies about model's economics.

Canova and Sala (2009) argue that FLI techniques deliver more accurate inference than LMI one. Their point is that the likelihood of the model conveys substantially more information than the LMI objective function and this helps the identification of the true parameter vector. However, Iskrev (2010) illustrates that DSGE models are inherently weakly identified and this symptom applies to both FLI and LMI – when enough instruments are used so the estimated parameter to be identified – estimation techniques. Theodoridis (2010) shows that when a sufficient number of instruments, which fully summarise the likelihood of the model under normality, and the optimal weighted matrix are used then, in small samples, LMI estimation techniques outperform even Bayesian FLI procedures with unrealistic tight priors around the true parameter vector.

The studies of Del Negro and Schorfheide (2004) and Christiano et al. (2010) can be viewed as attempts to incorporate inside the Bayesian framework this long lasting knowledge about selecting structural parameters based on LMI metrics. The technique presented in this section walks along this path and illustrates how the structural parameters can be selected in order to minimise the distance between the posterior moments of the VAR parameter vector and the one implied by the DSGE model.

To be precise, we know from the work of Del Negro and Schorfheide (2004, Section 3.3.1) that the posterior distribution of γ can be obtained by combining the marginal likelihood of the VAR with the prior distribution of the structural parameter vector – expression (3.20) –, however, in our case $m_{\mathcal{T}}\left(Y|\gamma,\widehat{\theta},\widehat{\widehat{\lambda}}\right)$ does not have an analytic form. This difficulty is bypassed using the "Laplace" approximation and the mapping (4.2). To see this, notice that under the assumption that the likelihood is close to symmetric and highly picked around the mode

$$\theta^{*}=\arg\max L_{\mathcal{T}}\left(Y|\theta\right)p\left(\delta|\gamma,\widehat{\widehat{\lambda}}\right)p\left(\sigma_{v}|\gamma,\widehat{\widehat{\lambda}}\right) \tag{5.1}$$

$m_{\mathcal{T}}\left(Y|\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right)$ can be approximated by

$$\begin{aligned} m_{\mathcal{T}}\left(Y|\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right) &= (2\pi)^{d\theta}\left|-\nabla_\theta^2 \log p\left(\theta^*|Y,\widehat{\theta},\gamma,\widehat{\overline{\lambda}}\right)\right|^{-\frac{1}{2}} \\ &\quad \exp\left[\frac{1}{2}\left(\theta-\theta^*\right)'\nabla_\theta^2 \log p\left(\theta^*|Y,\widehat{\theta},\gamma,\widehat{\overline{\lambda}}\right)\left(\theta-\theta^*\right)\right] \end{aligned} \tag{5.2}$$

(see, Canova, 2005, Chapter 9), at this point we apply the same assumption adopted by Del Negro and Schorfheide and we replace θ by $\theta(\gamma)$ (4.2), namely

$$\begin{aligned} m_{\mathcal{T}}\left(Y|\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right) &= (2\pi)^{d\theta}\left|-\nabla_\theta^2 \log p\left(\theta^*|Y,\widehat{\theta},\gamma,\widehat{\overline{\lambda}}\right)\right|^{-\frac{1}{2}} \\ &\quad \exp\left[\frac{1}{2}\left(\theta(\gamma)-\theta^*\right)'\nabla_\theta^2 \log p\left(\theta^*|Y,\widehat{\theta},\gamma,\widehat{\overline{\lambda}}\right)\left(\theta(\gamma)-\theta^*\right)\right] \end{aligned} \tag{5.3}$$

Consequently, (3.20) in the present framework can be expressed as

$$p\left(\gamma|Y,\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right) \propto m_{\mathcal{T}}\left(Y|\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right)p(\gamma) \tag{5.4}$$

and the posterior distribution of γ can be constructed through the random walk Metropolis Hasting MCMC resampling scheme.

The theory regarding this type of limited information or Quasi Bayesian estimators is developed in a series of papers by Kwan (1999), Kim (2002) and Chernozhukov and Hong (2003).

5.2 DSGE Evaluation

The term inside the square brackets of (5.3)

$$\mathcal{W} \equiv \left(\theta(\gamma)-\theta^*\right)'\nabla_\theta^2 \log p\left(\theta^*|Y,\widehat{\theta},\gamma,\widehat{\overline{\lambda}}\right)\left(\theta(\gamma)-\theta^*\right) \tag{5.5}$$

is a norm that assesses the plausibility of the DSGE model relatively to the estimated VAR. To see this notice that under the extreme assumption that the structural model is the true data generation process and – for simplicity – the posterior mode is equal to the posterior mean then the expected difference between $\theta(\gamma)$ and θ^* should be equal to zero

$$\begin{aligned} \mathbb{E}_\gamma\left(\theta(\gamma)-\theta^*\right) &= \int\left(\theta(\gamma)-\theta^*\right)p\left(\gamma|Y,\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right)d\gamma \\ &= \int\theta(\gamma)\,p\left(\gamma|Y,\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right)d\gamma-\theta^* = 0_{d\theta\times 1} \end{aligned} \tag{5.6}$$

Alternatively, if the model is heavily misspecified then $\mathbb{E}_\gamma\left(\theta(\gamma)-\theta^*\right)$ is expected to be large.

Under quadratic preferences, the following expressions

$$\mathbb{E}_\gamma\mathcal{W} = \int\mathcal{W}p\left(\gamma|Y,\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right)d\gamma \tag{5.7}$$

$$\mathbb{E}_\gamma\left(\mathcal{W}-\mathbb{E}_\gamma\mathcal{W}\right)^2 = \int\left(\mathcal{W}-\mathbb{E}_\gamma\mathcal{W}\right)^2 p\left(\gamma|Y,\theta^*,\gamma,\widehat{\theta},\widehat{\overline{\lambda}}\right)d\gamma \tag{5.8}$$

can be interpreted as the expected loss and risk of using $\mathcal{M}$, respectively (Canova, 2005; Schorfheide,

2000). Additionally, the overlap between the posterior distribution of $\mathcal{W}$ and the postior distribution of $\mathcal{W}^*$

$$\mathcal{W}^* \equiv (\theta - \theta^*)' \nabla^2_\theta \log p\left(\theta^* | Y, \widehat{\theta}, \gamma, \widehat{\bar{\lambda}}\right)(\theta - \theta^*) \tag{5.9}$$

provides a more complete measure of the misspecifications of the DSGE model.

6 Application

6.1 Structural Model

To make the paper self-contained, this subsection briefly discusses some of the key linearised equilibrium conditions, this is a version of the model developed by Smets and Wouters (2007). 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 (ω^i_t) 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 $\varepsilon^g_t = \rho_g \varepsilon^g_{t-1} + \sigma_g \omega^g_t$, which is assumed to be exogenous. The market clearing condition is given by

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon^g_t \tag{6.1}$$

where y_t denotes the total output and Table (1) provides a full description of the model's parameters and the shape of their distribution. The consumption Euler equation is given by

$$\begin{aligned} c_t = {} & \frac{\lambda}{1+\lambda} c_{t-1} + \frac{1}{1+\lambda}\mathbb{E}_t c_{t+1} + \frac{(1-\sigma_C)\left(\bar{W}^h \bar{L} | \bar{C}\right)}{\sigma_C (1+\lambda)}\left(\mathbb{E}_t l_{t+1} - l_t\right) \\ & - \frac{1-\lambda}{\sigma_C (1+\lambda)}\left(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon^b_t\right) \end{aligned} \tag{6.2}$$

where l_t is the hours worked, r_t is the nominal interest rate, π_t is the rate of inflation and $\varepsilon^b_t = \rho_g \varepsilon^b_{t-1} + \sigma_g \omega^b_t$ is the risk premium shock. If the degree of habits is zero $(\lambda = 0)$, equation (6.2) reduces to the standard forward looking consumption Euler equation. The linearised investment equation is given by

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta}\mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta) S''} q_t + \varepsilon^i_t \tag{6.3}$$

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

$$q_t = \beta(1-\delta)\mathbb{E}_t q_{t+1} + (1 - \beta(1-\delta))\mathbb{E}_t r^k_{t+1} - \left(r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon^b_t\right) \tag{6.4}$$

where $r^k_t = -(k_t - l_t) + w_t$ 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 define as

$$y_t = \phi_p\left(\alpha k^s_t + (1-\alpha) l_t + \varepsilon^a_t\right) \tag{6.5}$$

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

$$k_t = (1-\delta)\, k_{t-1} + \delta i_t + (1+\beta)\, \delta S'' \varepsilon_t^i \tag{6.6}$$

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

$$\pi_t = \frac{i_p}{1+\beta i_p}\pi_{t-1} + \frac{\beta}{1+\beta i_p}\mathbb{E}_t\pi_{t+1} + \frac{1}{(1+\beta i_p)}\frac{\left(1-\beta\xi_p\right)\left(1-\xi_p\right)}{\left(\xi_p\left(\left(\phi_p-1\right)\varepsilon_p+1\right)\right)} mc_t \tag{6.7}$$

where $mc_t = \alpha r_t^k + (1-a)\, w_t - \varepsilon_t^a$ is the marginal cost of production. Monopolistic competition in the labour market also gives rise to a similar wage New-Keynesian Phillips curve

$$\begin{aligned} w_t \;=\; & \frac{1}{1+\beta} w_{t-1} + \frac{\beta}{1+\beta}\left(\mathbb{E}_t w_{t+1} + \mathbb{E}_t \pi_{t+1}\right) - \frac{1+\beta i_w}{1+\beta}\pi_t + \frac{i_w}{1+\beta}\pi_{t-1} \\ & + \frac{1}{1+\beta}\frac{\left(1-\beta\xi_w\right)\left(1-\xi_w\right)}{\left(\xi_w\left(\left(\phi_w-1\right)\varepsilon_w+1\right)\right)}\mu_t^w \end{aligned} \tag{6.8}$$

where $\mu_t^w = \left(\sigma_l l_t + \frac{1}{1-\lambda}\left(c_t - \lambda c_{t-1}\right)\right) - w_t$ is the households' marginal benefit of supplying an extra unit of labour service.

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

$$r_t = \rho r_{t-1} + (1-\rho)\left(r_\pi \pi_t + r_y y_t\right) + \varepsilon_t^r \tag{6.9}$$

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

6.2 Monte Carlo Evaluation

This section undertakes a small simulation exercise to investigate the performance of the DSGE Quasi-Bayesian estimator – QBML – proposed in Section 5.1 against the standard full information Bayesian maximum likelihood – BML – one. The data generation process is the model described in Section 6.1 and it is assumed that only output, consumption, investment, inflation and nominal interest rates (y_t, c_t, i_t, π_t and r_t, respectively) are observed in this economy.

The details of this exercise are described in Appendix C, however, we would like to highlight that the first minimisation required to obtain θ^* and $\nabla_\theta^2 \log p\left(\theta^* | Y, \widehat{\theta}, \gamma, \widehat{\overline{\lambda}}\right)$ has been replaced by the estimation of the VAR model, and the use of the posterior mean and the variance-covariance matrix of $p\left(\theta | Y, \widehat{\theta}, \gamma, \widehat{\overline{\lambda}}\right)$ instead of θ^* and $\nabla_\theta^2 \log p\left(\theta^* | Y, \widehat{\theta}, \gamma, \widehat{\overline{\lambda}}\right)$, respectively. It is common practice for applied researchers to approximate the latter quantities from the output of the posterior simulation (Koop and Poirier, 1993), since this reduces the computation effort. We also run the same experiment using θ^* and $\nabla_\theta^2 \log p\left(\theta^* | Y, \widehat{\theta}, \gamma, \widehat{\overline{\lambda}}\right)$, the results are almost identically – marginally better for the proposed estimator – and, therefore, not presented here.

Before turning into the discussion of the results, it should be highlighted that the purpose of this experiment is not to establish the small properties of the proposed estimator – where a more serious exercise would be required – but to illustrate that estimator described in Section 5.1 behaves decently in small samples and we choose the full information maximum likelihood estimator as the benchmark case.

The prior moments – mean and standard deviation – of the structural parameter vector are presented in the first two columns of Table 4. The next three display the median – calculated using the Mahanalobis metric[4] –, the standard deviation and the bias for the BML estimator after 1000 simulations. Finally, the remaining columns capture the same information regarding the proposed estimator.

As it can be seen the Quasi-Bayesian estimator performs better than the full information one not only in terms of bias but also in terms of efficiency. Clearly, these differences are not huge and it would not be wise to argue in favor of the proposed estimator and against the standard maximum likelihood one, however, this simulation does illustrate that the QBML estimator performs relatively well even in small samples.

It is not surprising the fact that in small samples a limited information estimator is less biased than the full information one (Ruge-Murcia, 2007; Theodoridis, 2010), however, it seems striking that the former does better in terms of the efficiency as well. Our candidate explanation relies on the asymptotic theory of minimum distance estimators (Newey and Mcfadden, 1994), to be precise, from (5.3) and (5.4) it is clear that the posterior distribution of the structural parameter vector consists of the values of γ that minimise the distance between $\theta(\gamma)$ and θ^*. In other words, θ^* can be interpreted as the set of the instruments used in this estimation and $-\nabla^2_\theta \log p\left(\theta^*|Y,\widehat{\theta},\gamma,\widehat{\widehat{\lambda}}\right)^{-1}$ is its covariance matrix. Loosely speaking, $p\left(\gamma|Y,\theta^*,\gamma,\widehat{\theta},\widehat{\widehat{\lambda}}\right)$ can be viewed as the distribution of an efficient estimator and we believe that this property shows up in this simulation for the limited information one and not for the full one – certainly, this property does not hold asymptotically where the BML is more efficient than the QBML estimator.

6.3 Empirical Exercise

We also investigate the performance of the two estimators using actual macroeconomic data and we choose for this purpose the set constructed by Smets and Wouters (2007).[5] The use of the Hodrick-Prescott filter to eliminate the zero frequency component of the non-stationary series – real output, real consumption and real investment – is implied by the absence of any kind of model trend – either deterministic or stochastic. The model is estimated using data between 1966Q1 and 2004Q4, the shape of the prior distribution is given by the Table 1, while the prior moments can be found in Table 5.

Fit of $\mathcal{M}$

Chart 7 plots the Kalman filter estimates of the observed series of the two estimators against the data (black solid line), loosely speaking, these lines represent the one-step-ahead in sample forecasts and,

[4]The Mahanalobis distance has also been used by Jorda et al. (2010) to construct simultaneous confidence regions for forecast paths and by Minford et al. (2009) for DSGE model validation.

[5]This data set is publicly available from the website of the American Economic Association.

consequently, the above graph illustrates the estimation fit of the two models. It is clear that the in sample fit of the structural model using the full information Bayesian maximum likelihood estimator (dashed red line) is really good, however, the one obtained by the Quasi-Bayesian estimator (dotted-dashed blue line) seems marginally better. Having establish that both models are well, we turn into the discussion of the results.

Posterior Distribution of γ

The posterior distribution of the BML and QBML estimators along with the prior distribution of the structural parameter vector are presented in Chart 5; Table 5 summarises the same information. It seems that the posterior variance for most of the QBML estimates (eighteen out of twenty five) is smaller – higher peaked mode – than the BML one, this is also reflected in the posterior distribution of the impulse response function (IRF) exercise – Chart 6 – and, it coincides consistent with the monte carlo results.

In terms of economics, marginal cost variations have smaller impact on the current inflation in the QBML (0.079) than in the BML (0.243) economy, additionally, the labour supply curve responds less to wage and consumption changes under the limited than the full information estimates. The effect of the current real interest rate to current consumption seems the same in both economies – BMl (0.216) and QBML (0.1982) –, however, the effect of the expected growth of hours worked is substantially reduced in the QBML economy (0.079) – BML (0.188). Indexation and Calvo probability parameters do not look very different, however, the QBML monetary policy maker reacts more aggressively towards inflation than the BML monetary authorities. The QBML standard deviations estimates are much smaller than the BML ones – except from the interest rate shock –, while the posterior distribution overlap between the two estimates regarding the persistence parameter of the risk premium, government spending and interest rate shocks is quite small.

Impulse Responses

These parameter differences imply different expectations for the agents of the two economies, as it can been seen form the IRF comparison exercise – Chart 6. For instance, in the QBML economy, a smaller and less persistent government spending shock crowds out less consumption and investment than the one hitting the BML economy. The difference in consumption responses leads to different profiles for real wages – BML agents supply more labour than it is actually demanded pushing wages down, the QBML labour supply does not exceed demand – and, consequently, different profiles for marginal cost, inflation and interest rates. Another example is the investment shock, a similar perturbation leads to a smaller investment increase and, consequently, to lower level of capital stock in the QBML than BML economy. The fall of the BML rental rate of capital exceeds the increases in the wages – due to higher labour demand – pushing and keeping the marginal cost below zero from the fourth period after the shock. However, this does not happen in the QBML economy where the marginal cost stays positive for all the duration of the shock, implying different inflation and interest rate profiles from those observed in the BML economy.

Generally speaking, except from the magnitude differences – mainly due standard deviation discrepancies, Chart 5 – both type of impulse responses look very similar – in terms of the shape and the sign.

This seems to be justified by the large overlap between the posterior distribution of the behavioural structural parameters of the two estimators.

Forecast Variance Decomposition

Chart 8 reveals the contribution of each shock to the variance of the h-periods-ahead forecast error of the observable vector – $h = \infty$ captures the long-run effect. Although both decompositions look very similar, it is clear that the QBML estimator assigns more importance to the monetary policy shock and reduces the effects of the productivity disturbance to all series except the nominal interest rate, where the opposite is observed. This is possibly due to the flatter QBML Philips curve – inflation is mainly driven by productivity shocks – and the higher inflation weight in the QBML Taylor rule. Although the effect of the government spending shock to consumption and investment is tiny in the BML economy, its contribution reduces almost to zero in the QBML one and the latter macroeconomic series are mainly driven by the risk premium and investment shock, respectively.

DSGE Evaluation

Under the assumption that the BVAR model proposed in this study is a good description of the data generation process – Table 3 provides some support to this hypothesis – then the metric discussed in Section 5.2 can be used to assess plausibility the restrictions imposed by the BML and QBML models on the data.

Table 2 displays the $\mathbb{E}_\gamma \mathcal{W}$ and $\sqrt{\mathbb{E}_\gamma \left(\mathcal{W} - \mathbb{E}_\gamma \mathcal{W}\right)^2}$ moments indicating that the expected loss and risk of using the BML instead of the QBML model is enormous. Given the way the latter model is estimated – the structural parameter vector is selected in order to minimise the distance between the VAR parameter vector implied by the DSGE and the one estimated in the data – and the differences between the posterior estimates, this results should not come as surprise.

The posterior distributions of the BML-$\mathcal{W}$ (red dashed line) and QBML-$\mathcal{W}$ (blue dashed-dotted line) are plotted against the posterior distribution of $\mathcal{W}^*$ in Chart 4. Although, the BML model appears to be well estimated its VAR predictions seem to be very different from those observed in the data. The picture looks better for the QBML model, however, the overlap between the posterior distribution of QBML-$\mathcal{W}$ and $\mathcal{W}^*$ is almost zero. This probably indicates that model's restrictions are quite severe even when γ has been selected aiming to give the best chance to the DSGE model to reproduce the estimated VAR dynamics. Again this should not surprise us since the VAR model has ninety well estimated parameters and the DSGE only twenty five.

6.4 BVAR Analysis

It is time to assess the empirical performance of the analytic type priors proposed in this study. This is achieved by using two more BVAR models – one with theory free Minnesota priors and one with Del Negro and Schorfheide (2004)s' regression type priors – and three metrics – marginal likelihood, mean square forecast error and predictive density. It should be emphasised that the estimation sample this time is 1966Q1–1999Q4 and the period 2000Q1–2004Q4 is used for the evaluation of the out of sample forecasting performance of the BVARs.

Estimation of λ

As it has been illustrated above the posterior estimates are weighted averages between the prior moments and the OLS estimates and these weights are controlled by the hyperparameter λ – this is also true for the Minnesota prior (see, Bandbura et al., 2010). The degree of the tightness parameter needs to be decided before the estimation and we rely on the marginal likelihood measure to identify the value that maximises the fit of the time series model. This process is computationally demanding since it requires the discretisation of the interval where λ is defined and the posterior estimation of the BVAR model for all these grid values.

Charts 3 and 2 provides the graphical representation of the mapping between λ and the marginal likelihood of the Minnesota (M-VAR) and the Del Negro and Schorfheide (2004)s' (DS-VAR) BVARs, respectively. Clearly, as λ_{DS} increases the fit of the MS-VAR model deteriorates, indicating that DSGE prior moments are not inline with the data and their contribution should be kept at the minimum. On the other hand, Minnesota posterior estimates appear robust to λ_m choices.

The relation between the hyperparameter vector $\bar{\lambda}$ and the marginal likelihood of the VAR is presented by Chart 1. We can see again that the compatibility between the DSGE based priors and the data is not great and, therefore, their role to the posterior inference should be limited. However, we are now able to identify which part of the VAR parameter vector implied by the DSGE model disagrees more with the data. To be precise, the significant steepness of the surface with respect to small values of η indicates that tiny changes to the degrees of freedom parameter lead to enormous changes to the fit of the model, while the surface is relatively flat with respect to λ_δ and large values of η. This seems to suggest that the prior mean of the residual variance-covariance matrix does not get much support by the data and – although, $\Sigma_v(\gamma)$ is a function of all structural parameters – this seems consistent with the fact we have got very different limited information estimates regarding the shock processes parameters from the full information ones.

Marginal Likelihood Evaluation

The marginal likelihood is probably the most traditional measure of model's fit, Table 3 presents these values for the BML and all three BVAR models. Given the discussion so far it is not hard to see why the DSGE model has the worst fit. On the other hand, the DS-VAR seems to perform better than any other model and the explanation is the tiny weight given to theoretical priors – $\lambda_{DS} = 0.04$ –, meaning that the economic theory is discounted heavily and the posterior estimates almost coincide with the OLS one. The other time series model which again uses theory consistent priors – FHT-VAR – achieves better fit than the VAR model with Minnesota priors – M-VAR.

Impulse Responses & Forecast Variance Decompositions

We use the "sign restriction" methodology (Canova and De Nicolo, 2002; Uhlig, 2005) to identify the structural shocks from the set of the VAR residuals. The sign restrictions – twenty five in total – are those implied by the BML estimated DSGE model and Charts 9, 10 and 11 display the posterior distribution of the impulse responses and the forecast variance decomposition of the M-VAR, DS-VAR and FHT-VAR model, respectively.

Qualitatively, the impulse responses posterior distribution looks very similar for all time series models, even quantitatively, the differences do not look enormous. To highlight this point and to compare these VAR identified IRFs with those obtained by the DSGE models we plot all the IRF posterior means in Chart 12. It can be seen there that in most cases the QBML model produces impulse responses that are closer to the VAR-IRFs than those implied by the BML model, however, there are also some important differences that are worthy to be mentioned.

For instance, the DSGE models – especially the QBML – seems to heavily underestimate the investment crowding out effect after a government spending shock, this leads to different profiles for output – the VAR models seem to suggest that output gets negative after the second period of the shock –, marginal cost – the rental rate of capital and wages probably rise further in the economies described by the VARs – and, consequently, inflation. Another interesting feature is that the QBML model seems to replicate the VAR-IRFs for the demand components and inflation after a productivity shock, however, it enormously overestimates the interest rate fall and this is probably due to a stronger estimate about the persistence of the productivity effect and the inflation reaction coefficient in the Taylor rule.

Again the VAR forecast error variance decomposition looks very similar for all VAR models, however, there are some important differences with those obtained by the DSGE models – Chart 8. Unlike with the structural models the investment forecast is not mainly explained by an investment shock – at least in the short run – and the contribution of the monetary policy shock to ivestment has increased. This is also consistent with the IRF picture – Chart 12 – where the VAR investment does not rise – after an investment shock – as much as the DSGE responses and a much smaller interest rate increase – after a monetary policy shock – leads to a larger the ivestment fall than the those predicted by the structural models. Inflation in the VAR world is more demand driven, this squares with the IRFs where government spending and investment demand shocks seem to have larger effects on inflation. Again form Chart 12 it can be seen that the interest rate hardly responds to the productivity and government spending shocks – this is mainly due to the fact that inflation barely reacts to these perturbations – and this what shows up in its forecast variance forecast decomposition where the risk premium shock is the main driver.

Forecast Evaluation

Chart 13 plots the mean-square-forecast-error for one, four, eight and twelve-quarters-ahead forecasts. This univariate measure of forecasting performance seems to suggest that no model delivers superior forecasts and this is important because Minnesota priors are usually considered as the one that improve the forecasting fit of the VAR model.

This is also seems to be the case when multivariate measures of forecasting performance are used. To be precise, under the assumption that the functional form of the multivariate predictive density function is the normal pdf we are able to construct forecast probability weights that measure the possibility that the actual vector data outturn is produced by the specific model. The first four subplots of Chart 14 illustrate these probability for the period 2000Q1-2004Q4 for one, four, eight and twelve-quarters-ahead forecasts, respectively. These time varying weights can be also used to constructed an average forecast that it is robust to the worst case outcome. The final subplot illustrates the same probability weight but for the entire period this time. It seems again from Chart 14 that Minnesota priors do not

deliver superior forecasting performance and all BVAR models do get considerable data support.

7 Conclusion

This paper proposes a methodology of constructing DSGE theory driven prior distribution for BVAR models eliminating conceptual difficulties that arise from the use of theory free Minnesota prior distribution. Similar to the previous studies, the empirical exercises presented here illustrate that theoretical consistency does not induce any forecasting-fit cost. We next show how the posterior moments of the VAR parameter vector can be used to obtain the posterior inference with respect to the DSGE parameter vector. Monte carlo simulations and evidences from an empirical exercise seem to suggest some important gains of using this Quasi-Bayesian estimator. A measure of fit that assesses the plausibility of the restrictions imposed by the DSGE model on the data relatively to the BVAR model, which can also be used to rank candidate structural models, naturally arises in the present framework. A step for future research is to use similar type of methodology for structural VAR model (**?**).

References

Bandbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92.

Blanchard, J. and M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectation," *Econometrica*, 48, 1305–1311.

Canova, F. (2005): *Methods for Applied Macroeconomic Research*, Princeton: Princeton University Press.

Canova, F. and G. De Nicolo (2002): "Monetary disturbances matter for business fluctuations in the G-7," *Journal of Monetary Economics*, 49, 1131–1159.

Canova, F. and L. Sala (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56, 431–449.

Chernozhukov, V. and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics*, 115, 293–346.

Chib, S. (1995): "Marginal Likelihood form 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 Papers 6400.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2006): "Assessing Structural VARs," NBER Working Papers 12353, National Bureau of Economic Research, Inc.

Christiano, L. J., M. Trabandt, and K. Walentin (2010): "DSGE Models for Monetary Policy Analysis," Working Paper 16074, National Bureau of Economic Research.

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-Villaverde, J., J. Rubio-Ramirez, T. Sargent, and M. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021–1026.

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," *Econometric Reviews*, 18, 1–73.

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.

Jorda, O., M. Knuppel, and M. Marcellino (2010): "Empirical Simultaneous Confidence Regions for Path-Forecasts," CEPR Discussion Papers 7797, C.E.P.R. Discussion Papers.

Kadiyala, K. R. and S. Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, 12, 99–132.

Kim, J.-Y. (2002): "Limited information likelihood and Bayesian analysis," *Journal of Econometrics*, 107, 175–193.

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.

Koop, G. (2003): *Bayesian Econometrics*, Chichester, England: Wiley & Sons.

Koop, G. and D. J. Poirier (1993): "Bayesian analysis of logit models using natural conjugate priors," *Journal of Econometrics*, 56, 323–40.

Kwan, Y. K. (1999): "Asymptotic Bayesian analysis based on a limited information estimator," *Journal of Econometrics*, 88, 99–121.

Litterman, R. (1980): "Techniques for Forecasting with Vector Autoregresions," 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.

Magnus, J. R. and H. Neudecker (1979): "The Commutation Matrix: Some Properties and Applications," *The Annals of Statistics*, 7, pp. 381–394.

Minford, P., K. Theodoridis, and D. Meenagh (2009): "Testing a Model of the UK by the Method of Indirect Inference," *Open Economies Review*, 20, 265–291.

Newey, W. and D. Mcfadden (1994): "Large Samples Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. Engel and D. Mcfadden.

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, 2048–2064.

Rotemberg, J. J. and M. Woodford (1998): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy: Expanded Version," Working Paper 233, National Bureau of Economic Research.

Ruge-Murcia, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31, 2599–36.

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.

Smith, A. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

Theil, H. and A. S. Goldberger (1961): "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, 2, 65–78.

Theodoridis, K. (2010): "DSGE Efficient IRF Minimum Distance Estimation," mimeo.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.

White, H. (2001): *Asymptotic Theory for Econometricians*, San Diego: Academic Press.

A Proofs

Proof: *of Proposition 4.1*

$$\begin{aligned} p\left(\theta|Y,\widehat{\theta},\gamma,\bar{\lambda}\right) &\propto \left|\Sigma_{\lambda\delta}(\gamma)\right|^{-0.5d\delta}\exp\left\{-0.5\left[\Sigma_{\lambda\delta}(\gamma)^{-0.5}(\delta-\mu_{\delta}(\gamma))\right]'\left[\Sigma_{\lambda\delta}(\gamma)^{-0.5}(\delta-\mu_{\delta}(\gamma))\right]\right\} \\ &\quad\times\left|\Pi(\gamma)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad\times\left|\Sigma_v\right|^{-0.5T}\exp\left\{\begin{array}{c}-0.5\left[\left(\Sigma_v^{-0.5}\otimes I_T\right)y-\left(\Sigma_v^{-0.5}\otimes X\right)\delta\right]' \\ \left[\left(\Sigma_v^{-0.5}\otimes I_T\right)y-\left(\Sigma_v^{-0.5}\otimes X\right)\delta\right]\end{array}\right\} \\ &= \left|\Sigma_{\lambda\delta}(\gamma)\right|^{-0.5d\delta}\left|\Pi(\gamma)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(T+\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad\exp\left\{-0.5\left(\begin{array}{c}\left[\left(\Sigma_v^{-0.5}\otimes I_T\right)y-\left(\Sigma_v^{-0.5}\otimes X\right)\delta\right]' \\ \left[\left(\Sigma_v^{-0.5}\otimes I_T\right)y-\left(\Sigma_v^{-0.5}\otimes X\right)\delta\right] \\ +\left[\Sigma_{\lambda\delta}(\gamma)^{-0.5}(\delta-\mu_{\delta}(\gamma))\right]'\left[\Sigma_{\lambda\delta}(\gamma)^{-0.5}(\delta-\mu_{\delta}(\gamma))\right]\end{array}\right)\right\} \end{aligned} \tag{A.1}$$

Similar to Canova (2005, Chapter 10, page: 354) we let $\mathcal{Y}\equiv\left[\begin{array}{cc}\Sigma_{\lambda\delta}(\gamma)^{-0.5}\mu_{\delta}(\gamma) & \left(\Sigma_v^{-0.5}\otimes I_T\right)y\end{array}\right]'$ and $\mathcal{X}\equiv\left[\begin{array}{cc}\Sigma_{\lambda\delta}(\gamma)^{-0.5} & \left(\Sigma_v^{-0.5}\otimes X\right)\end{array}\right]'$, then (A.1) becomes

$$\begin{aligned} p\left(\theta|Y,\widehat{\theta},\gamma,\bar{\lambda}\right) &\propto \left|\Sigma_{\lambda\delta}(\gamma)\right|^{-0.5d\delta}\left|\Pi(\gamma)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad\times\left|\Sigma_v\right|^{-0.5T}\exp\left\{-0.5(\mathcal{Y}-\mathcal{X}\delta)'(\mathcal{Y}-\mathcal{X}\delta)\right\} \\ &= \left|\Sigma_{\lambda\delta}(\gamma)\right|^{-0.5d\delta}\left|\Pi(\gamma)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad\times\left|\Sigma_v\right|^{-0.5T}\exp\left\{\begin{array}{c}-0.5\left[(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta})-\mathcal{X}(\delta-\bar{\mu}_{\delta})\right]' \\ \left[(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta})-\mathcal{X}(\delta-\bar{\mu}_{\delta})\right]\end{array}\right\} \\ &= \left|\Sigma_{\lambda\delta}(\gamma)\right|^{-0.5d\delta}\left|\Pi(\gamma)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad\times\left|\Sigma_v\right|^{-0.5T}\exp\left\{-0.5\left[\begin{array}{c}(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta})'(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta}) \\ -(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta})'\mathcal{X}(\delta-\bar{\mu}_{\delta}) \\ -(\delta-\bar{\mu}_{\delta})'\mathcal{X}'(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta}) \\ +(\delta-\bar{\mu}_{\delta})'\mathcal{X}'\mathcal{X}(\delta-\bar{\mu}_{\delta})\end{array}\right]\right\} \end{aligned} \tag{A.2}$$

by setting

$$\begin{aligned} \bar{\mu}_{\delta} &= \left(\mathcal{X}'\mathcal{X}\right)^{-1}\mathcal{X}'\mathcal{Y}=\left[\Sigma_{\lambda\delta}(\gamma)^{-1}+\left(\Sigma_v^{-1}\otimes X'X\right)\right]^{-1}\left[\Sigma_{\lambda\delta}(\gamma)^{-1}\mu_{\delta}(\gamma)+\left(\Sigma_v^{-1}\otimes X\right)'y\right] \\ &= \left[\Sigma_{\lambda\delta}(\gamma)^{-1}+\widehat{\Sigma}_{\delta}^{-1}\right]^{-1}\left[\Sigma_{\lambda\delta}(\gamma)^{-1}\mu_{\delta}(\gamma)++\widehat{\Sigma}_{\delta}^{-1}\widehat{\delta}\right] \end{aligned} \tag{A.3}$$

and

$$\bar{\Sigma}_{\delta}\equiv\left(\mathcal{X}'\mathcal{X}\right)^{-1}=\left[\Sigma_{\lambda\delta}(\gamma)^{-1}+\widehat{\Sigma}_{\delta}^{-1}\right]^{-1} \tag{A.4}$$

(A.2) reduces to

$$\begin{aligned} p\left(\theta|Y_t,\lambda,p(\gamma)\right) &\propto \left|\Sigma_{\lambda\delta}(\gamma)\right|^{-0.5d\delta}\left|\Pi(\gamma)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad\left|\Sigma_v\right|^{-0.5T}\exp\left\{-0.5\left[\begin{array}{c}(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta})'(\mathcal{Y}-\mathcal{X}\bar{\mu}_{\delta}) \\ +(\delta-\bar{\mu}_{\delta})'\bar{\Sigma}_{\delta}^{-1}(\delta-\bar{\mu}_{\delta})\end{array}\right]\right\} \end{aligned} \tag{A.5}$$

From (A.5), it can be seen that conditioning on Σ_v,

$$p\left(\delta|\Sigma_v^{-1},Y,\widehat{\theta},\gamma,\bar{\lambda}\right)\propto\left|\bar{\Sigma}_\delta\right|^{-0.5d\delta}\exp\left\{-0.5\left(\delta-\bar{\mu}_\delta\right)'\bar{\Sigma}_\delta^{-1}\left(\delta-\bar{\mu}_\delta\right)\right\} \tag{A.6}$$

meaning that δ is normally distributed with mean and variance equal to $\bar{\mu}_\delta$ and $\bar{\Sigma}_\delta$, respectively. Alternatively, equation (A.1) can also be written as

$$\begin{aligned}
p\left(\theta|Y,\widehat{\theta},\gamma,\bar{\lambda}\right) \propto\ & \left|\Sigma_{\lambda\delta}\left(\gamma\right)\right|^{-0.5d\delta}\exp\left\{-0.5\left[\Sigma_{\lambda\delta}\left(\gamma\right)^{-0.5}\left(\delta-\mu_\delta\left(\gamma\right)\right)\right]'\left[\Sigma_{\lambda\delta}\left(\gamma\right)^{-0.5}\left(\delta-\mu_\delta\left(\gamma\right)\right)\right]\right\} \\
& \times\left|\Pi\left(\gamma\right)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(\eta+dy+1)}\exp\left\{-0.5tr\left(\Sigma_v^{-1}\Pi\left(\gamma\right)\right)\right\} \\
& \times\left|\Sigma_v\right|^{-0.5T}\exp\left\{-0.5\left[\Sigma_v^{-1}\left[\begin{array}{c}\left(Y-X\widehat{\Delta}\right)'\left(Y-X\widehat{\Delta}\right)\\+\left(\Delta-\widehat{\Delta}\right)'X'X\left(\Delta-\widehat{\Delta}\right)\end{array}\right]\right]\right\} \\
\propto\ & \left|\Sigma_{\lambda\delta}\left(\gamma\right)\right|^{-0.5d\delta}\exp\left\{-0.5\left[\Sigma_{\lambda\delta}\left(\gamma\right)^{-0.5}\left(\delta-\mu_\delta\left(\gamma\right)\right)\right]'\left[\Sigma_{\lambda\delta}\left(\gamma\right)^{-0.5}\left(\delta-\mu_\delta\left(\gamma\right)\right)\right]\right\} \\
& \times\left|\Pi\left(\gamma\right)\right|^{0.5\eta}\left|\Sigma_v\right|^{-0.5(T+\eta+dy+1)} \\
& \exp\left\{-0.5tr\left(\Sigma_v^{-1}\left[\begin{array}{c}\Pi\left(\gamma\right)+T\widehat{\Sigma}_v+\\\left(\Delta-\widehat{\Delta}\right)'X'X\left(\Delta-\widehat{\Delta}\right)\end{array}\right]\right)\right\}
\end{aligned} \tag{A.7}$$

which, conditional on δ, is the the Wishart distribution with $\Pi\left(\gamma\right)+T\widehat{\Sigma}_v+\left(\Delta-\widehat{\Delta}\right)'X'X\left(\Delta-\widehat{\Delta}\right)$ scale matrix and $T+\eta$ degrees of freedom. ■

B θ Draws

This section explains how the draws of θ are produced

1. Draw γ_j from $p(\gamma)$ and solve the DSGE model
2. If γ_j satisfies Blanchard and Kahn's conditions (Blanchard and Kahn, 1980) calculate the M matrix (4.1) else go to step 1
3. If the eigenvalues of M are less than one in absolute terms calculate δ_j and σ_j using equations (4.3)–(4.5) else go to step 1
4. Repeat steps 1 to 3 S times

C Monte Carlo Steps

This section describes the step required to produce Table 4

1. The draws described in Section B are used to construct the prior moments of θ (4.1 - 4.5)
2. A sample of 200 observations is generated by the model described in Section 6.1 based on the prior moments of γ given by Table 4 for y_t, c_t, i_t, π_t and r_t
3. A BVAR with 3 lags is fitted to this data
4. The posterior median and variance-covariance matrix of θ are calculated. The former is obtained using the Mahanolobis metric, namely,

$$\bar{\theta}^{median} = \arg\max \left(\bar{\theta}_j - \mu_{\bar{\theta}}\right)' \Sigma_{\bar{\theta}}^{-1} \left(\bar{\theta}_j - \mu_{\bar{\theta}}\right)$$

   where $\mu_{\bar{\theta}}$ and $\Sigma_{\bar{\theta}}$ are the posterior mean and variance-covariance matrix of $p\left(\theta|Y, \widehat{\theta}, p(\gamma)\right)$
5. Given the posterior moments of θ, the posterior mode of γ is obtained by minimising $\tilde{m}(y_t|\theta^*, \lambda)\, p(\gamma)$, namely

$$\gamma^* = \arg\max \tilde{m}(y_t|\theta^*, \lambda)\, p(\gamma)$$

6. Steps 2-5 are repeated 1000 times

D Figures

Figure 1: FT-VAR $\lambda = (\lambda_\delta, \eta)'$

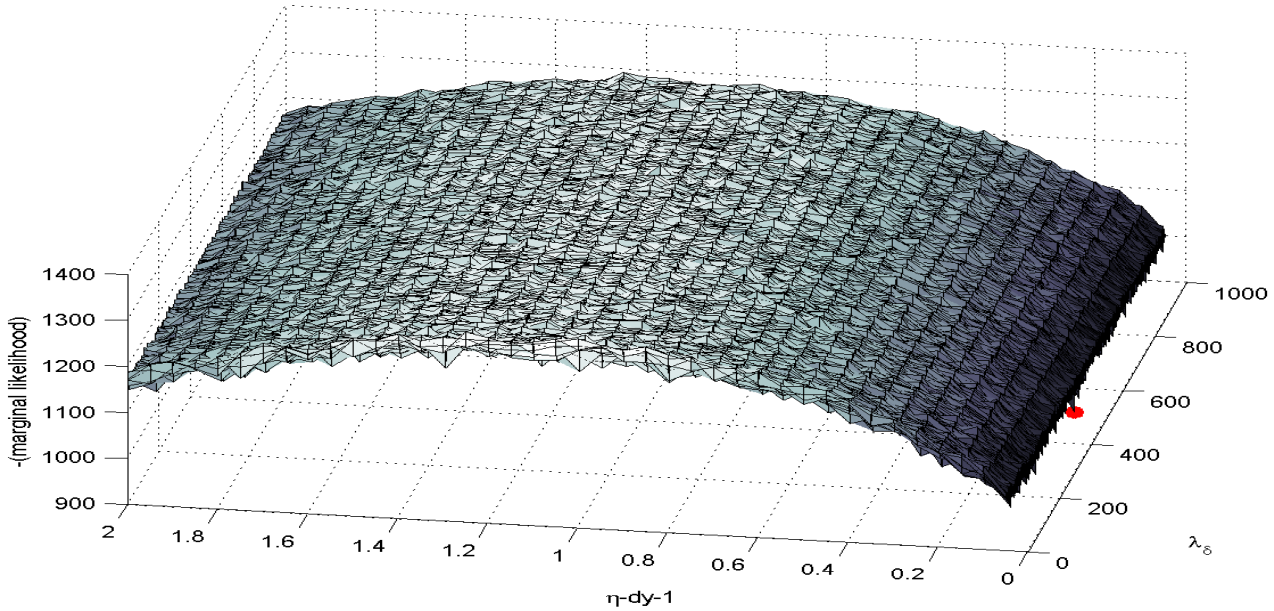


Figure 2: DS-VAR λ

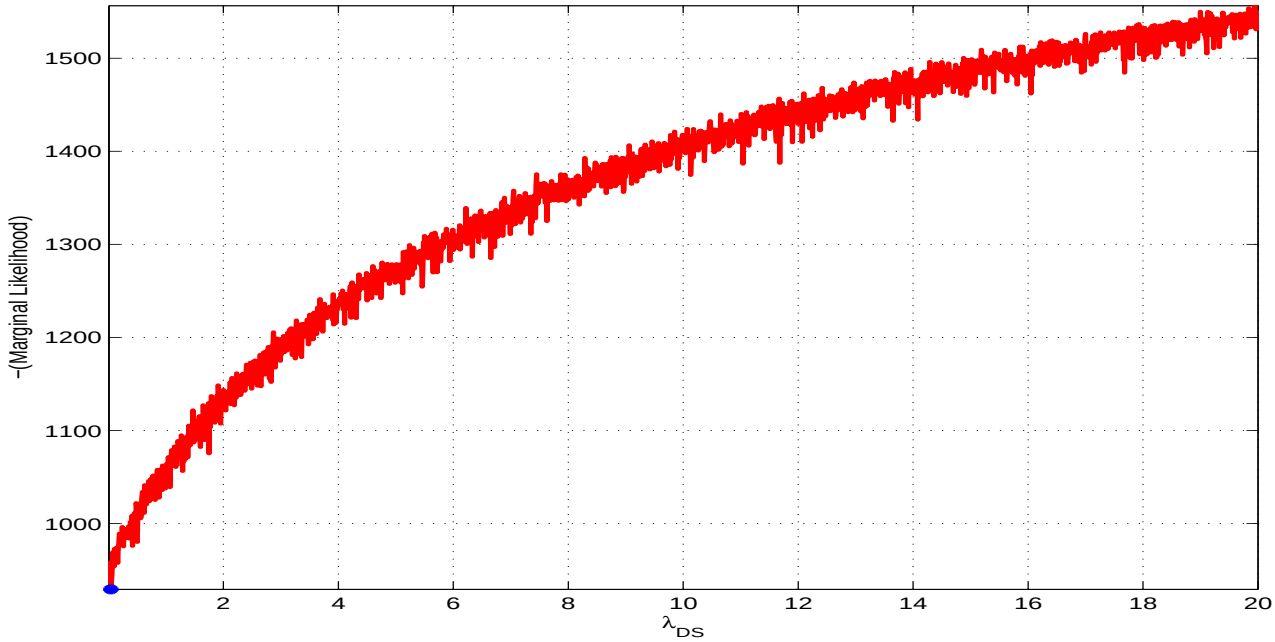


Figure 3: M-VAR λ

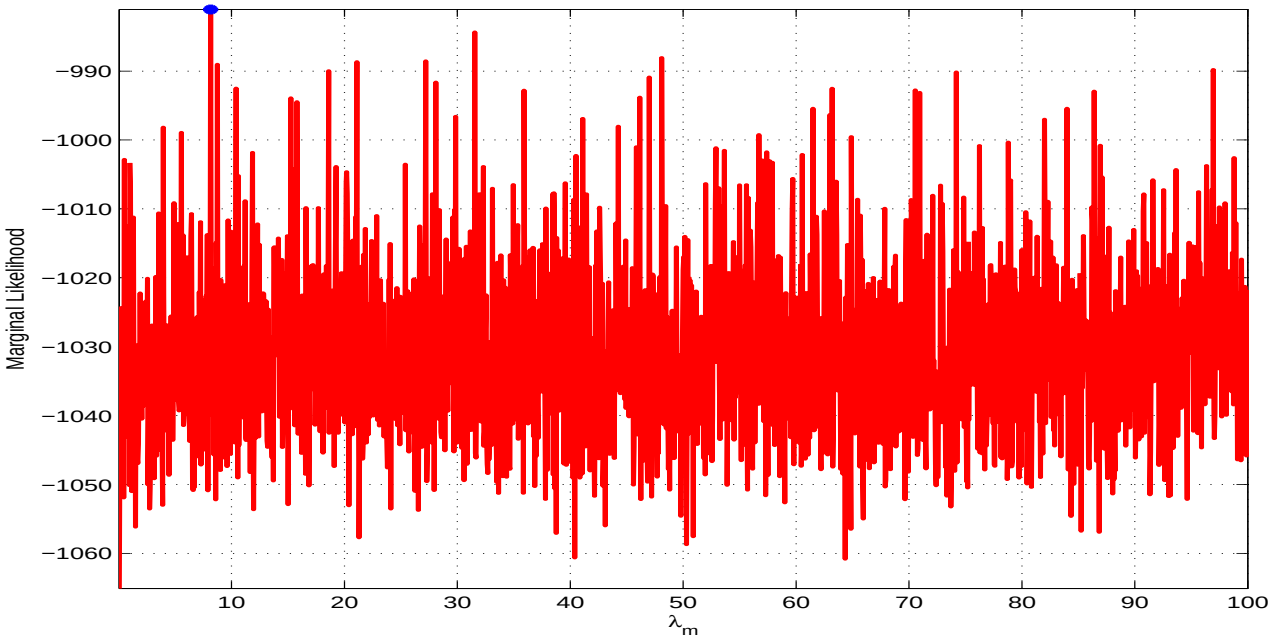


Figure 4: DSGE Evaluation

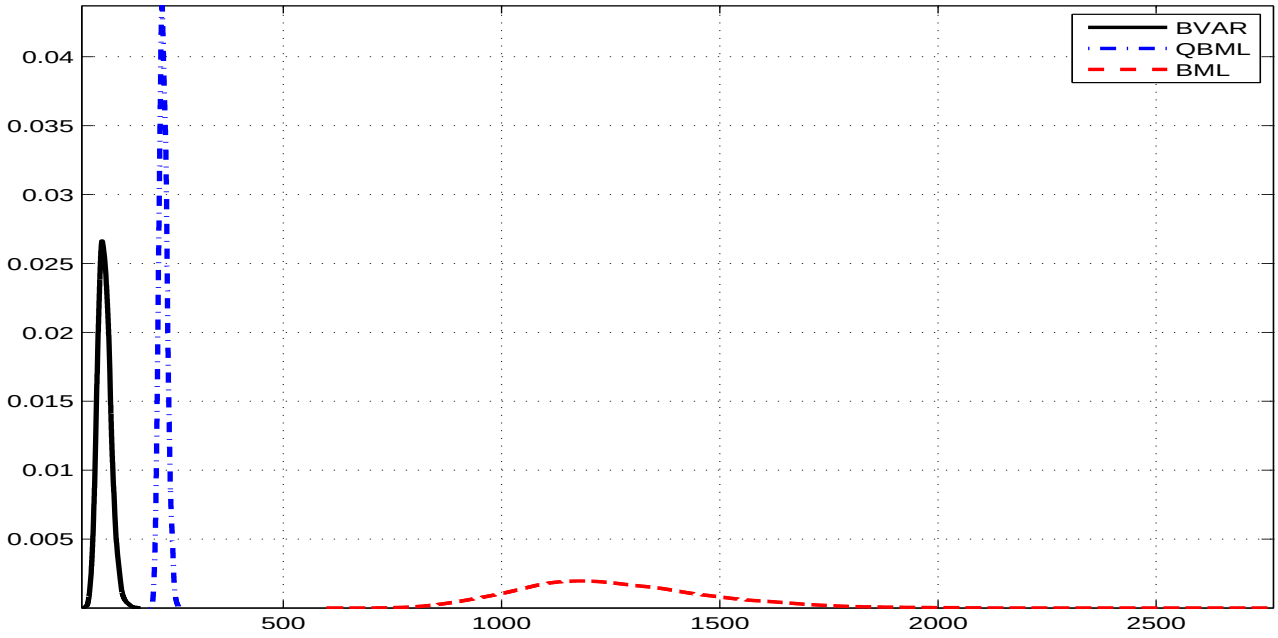


Figure 5: Prior versus Posterior DSGE Parameter Distribution

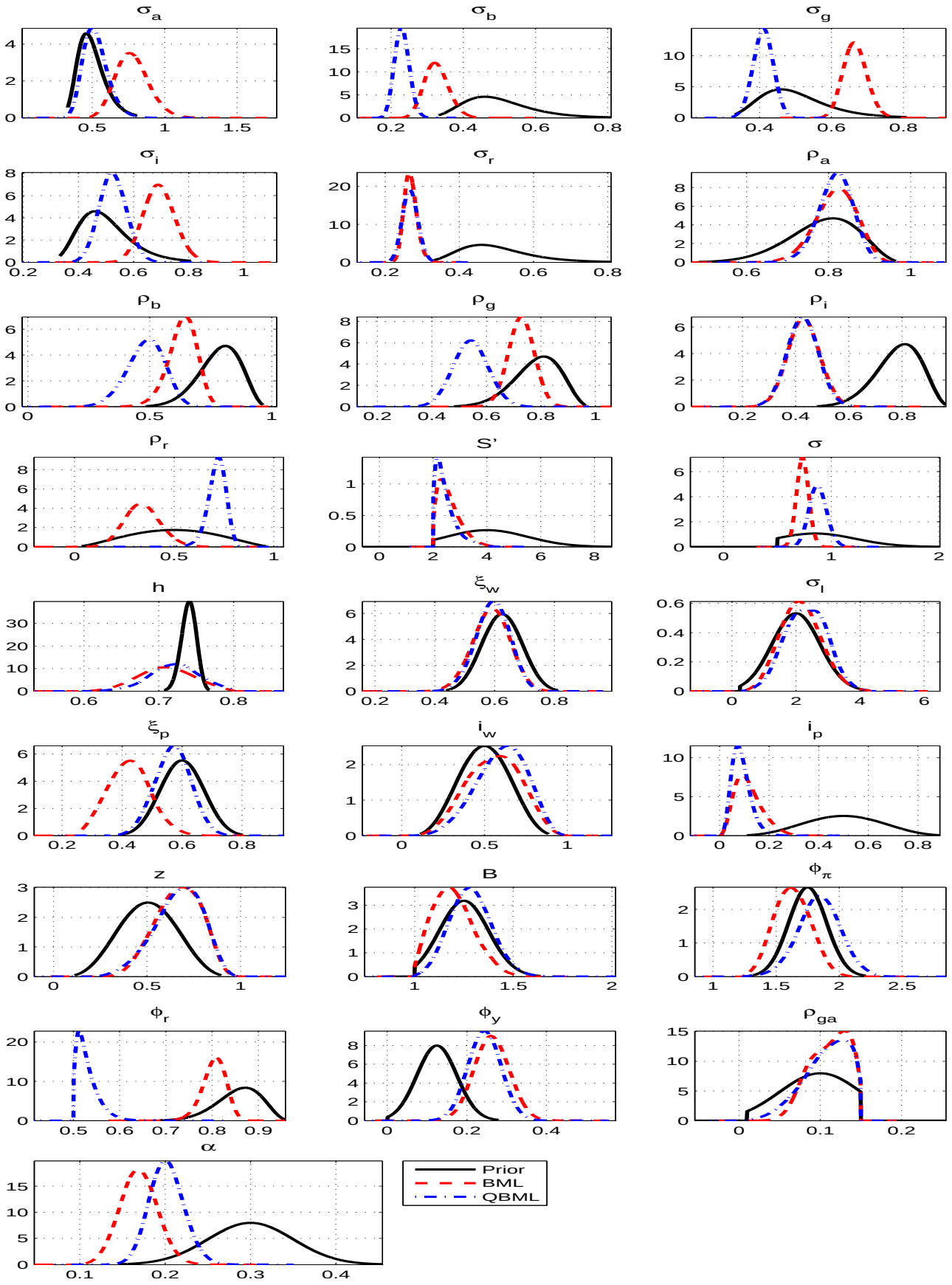


Figure 6: DSGE Posterior IRFs

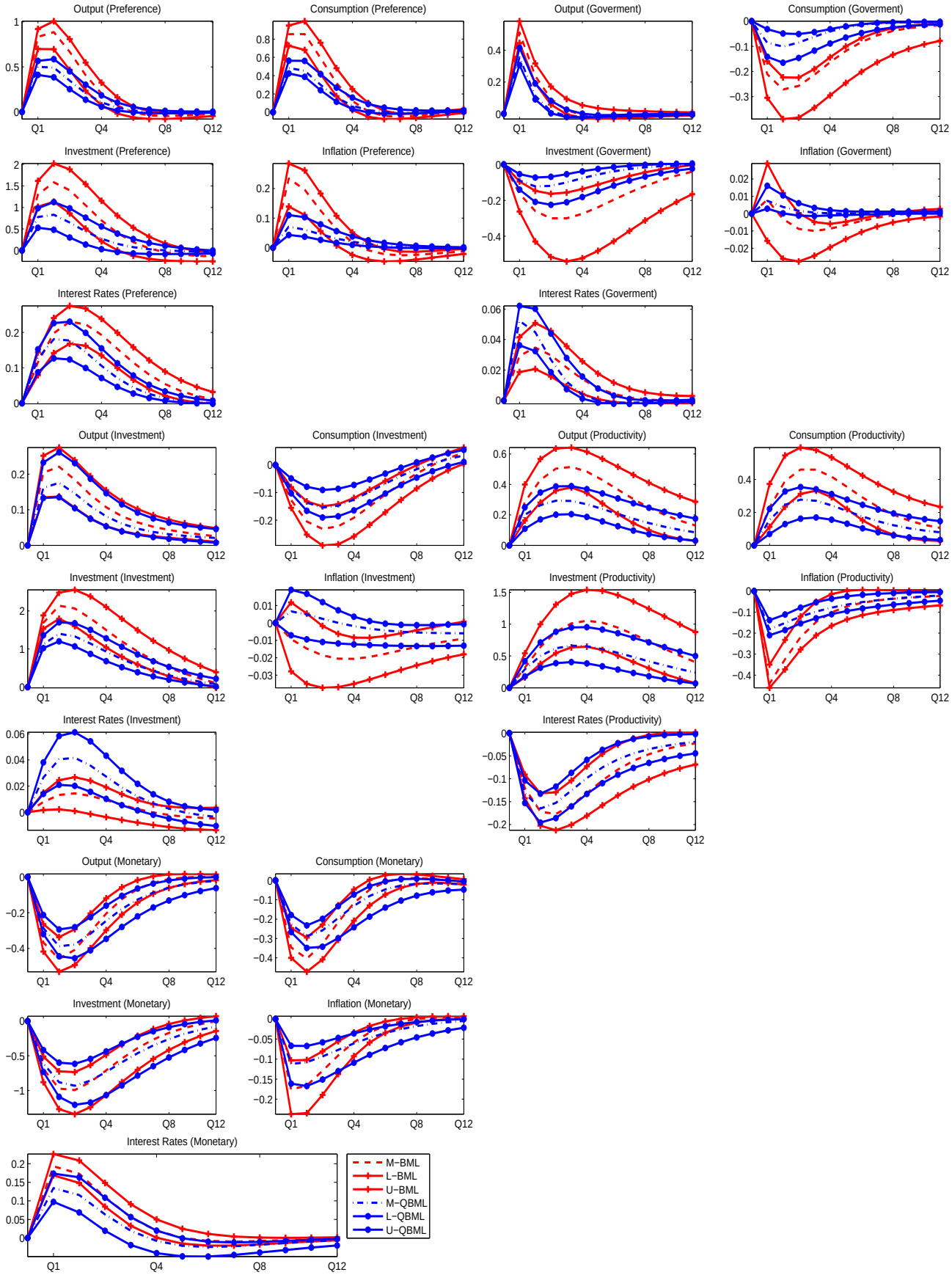


Figure 7: BML One-Step-Ahead Forecasts & FVD

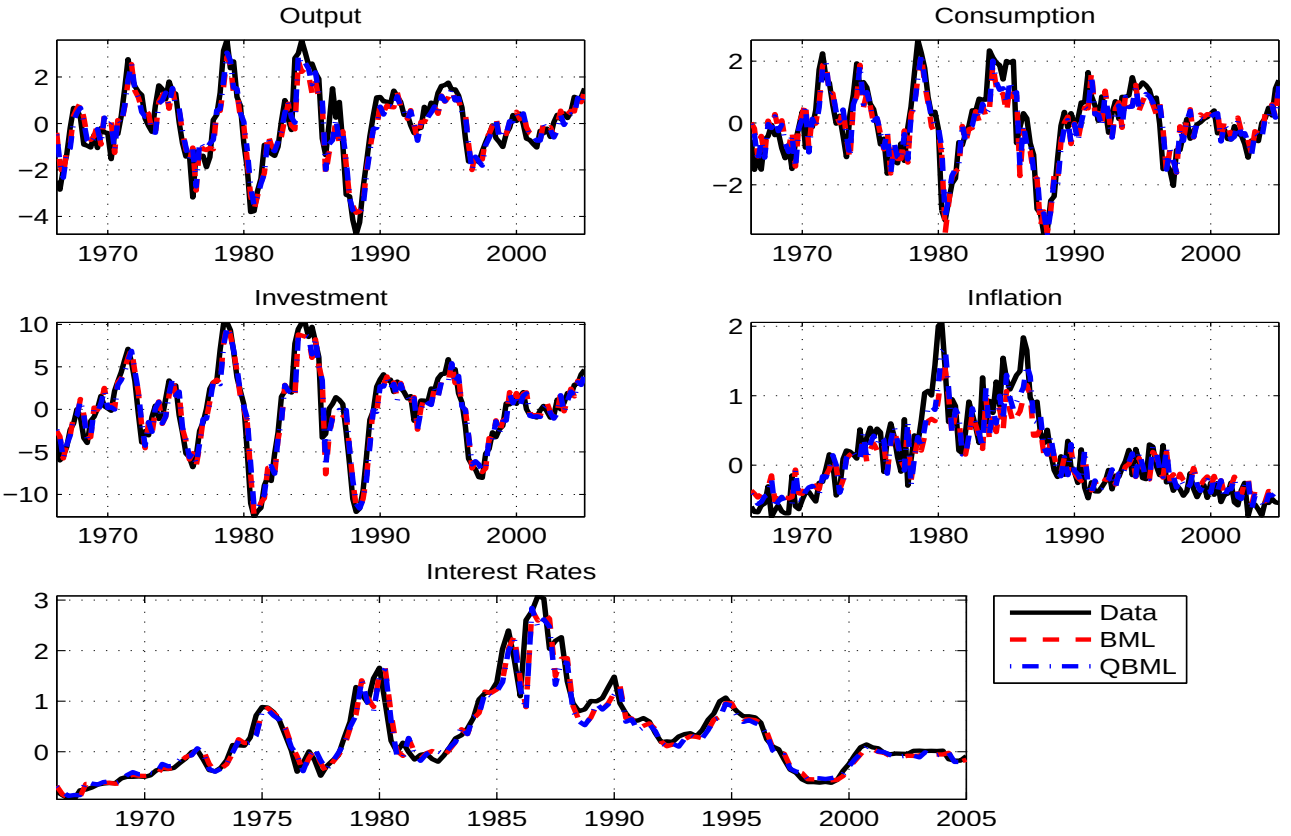


Figure 8: BML FVD QBML FVD

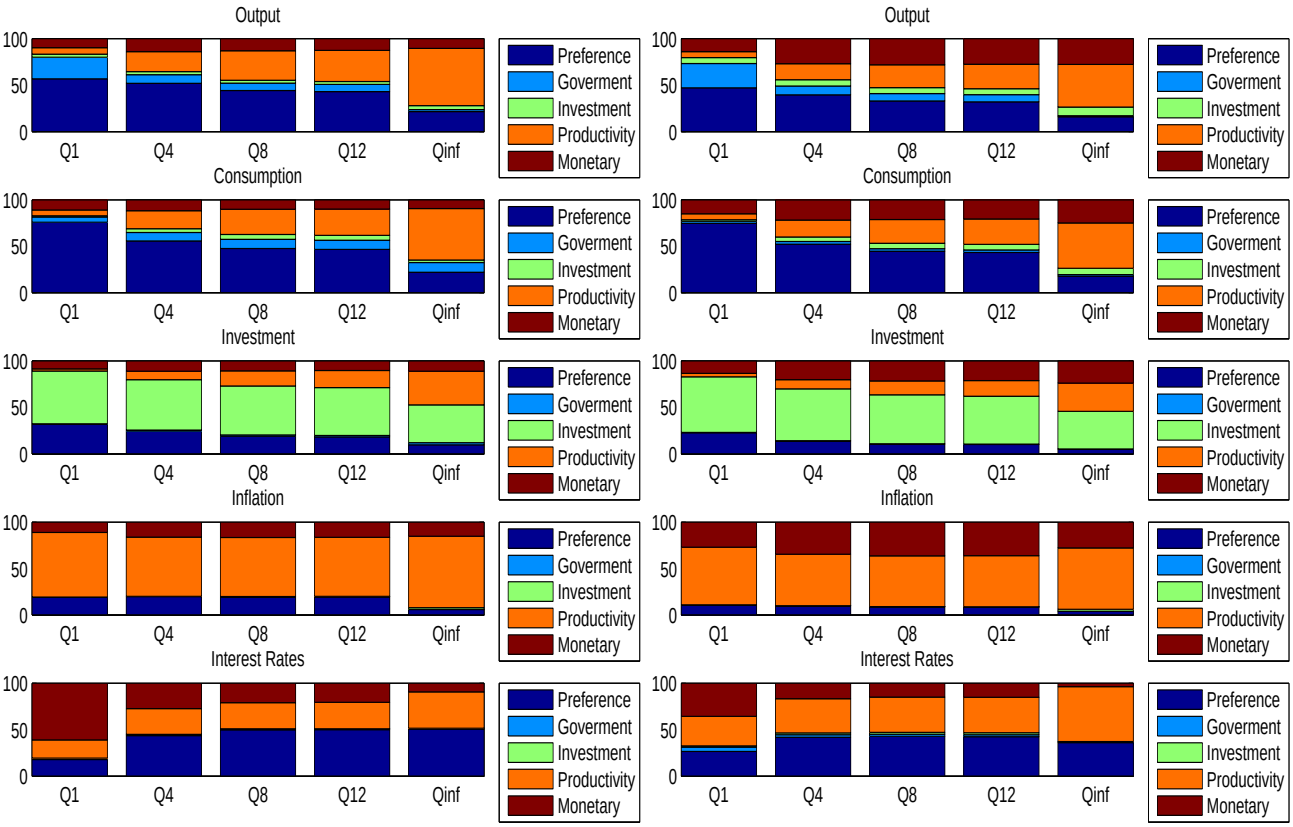


Figure 9: M-BVAR Posterior IRFs & FVD

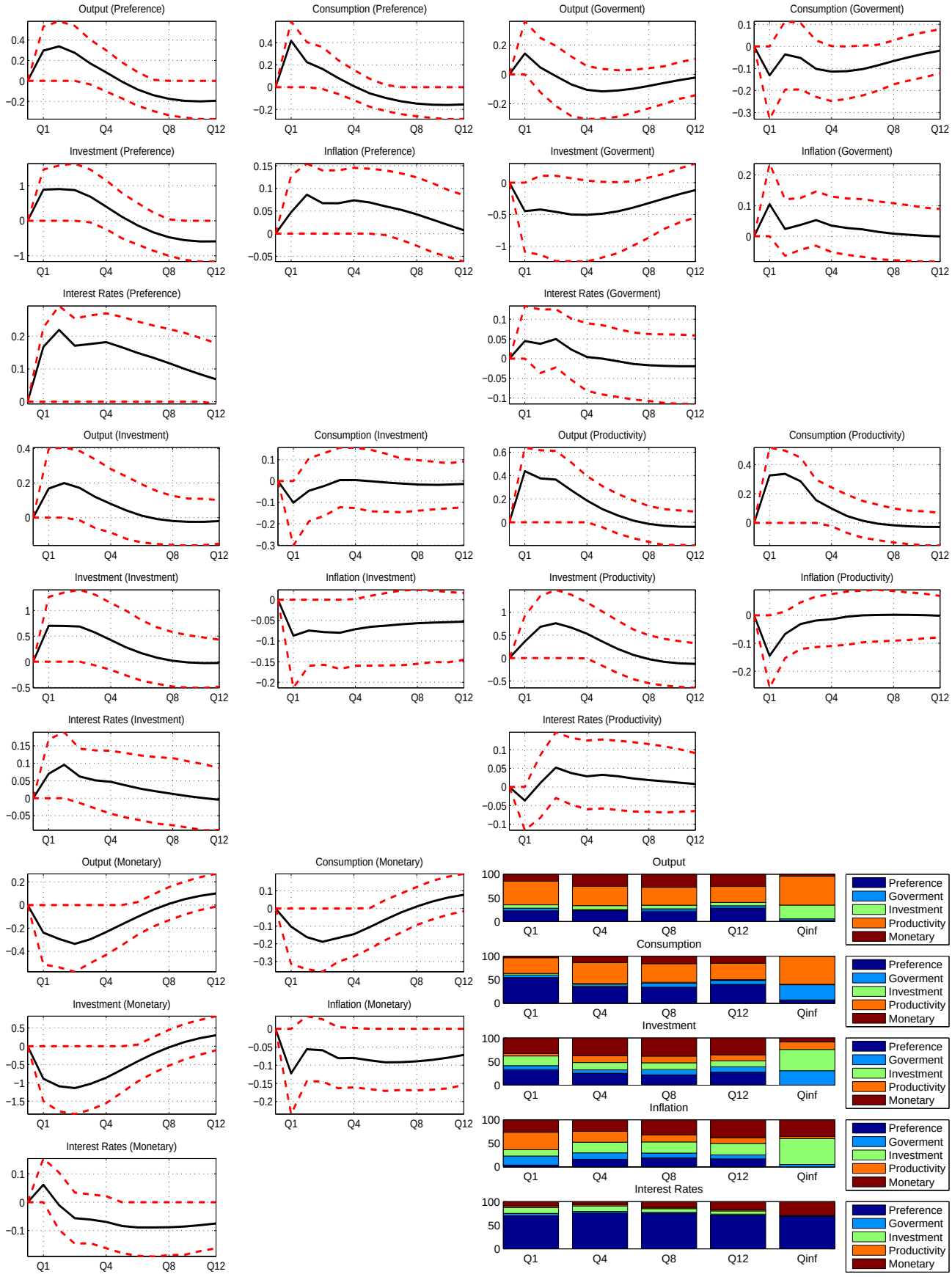


Figure 10: DS-BVAR Posterior IRFs & FVD

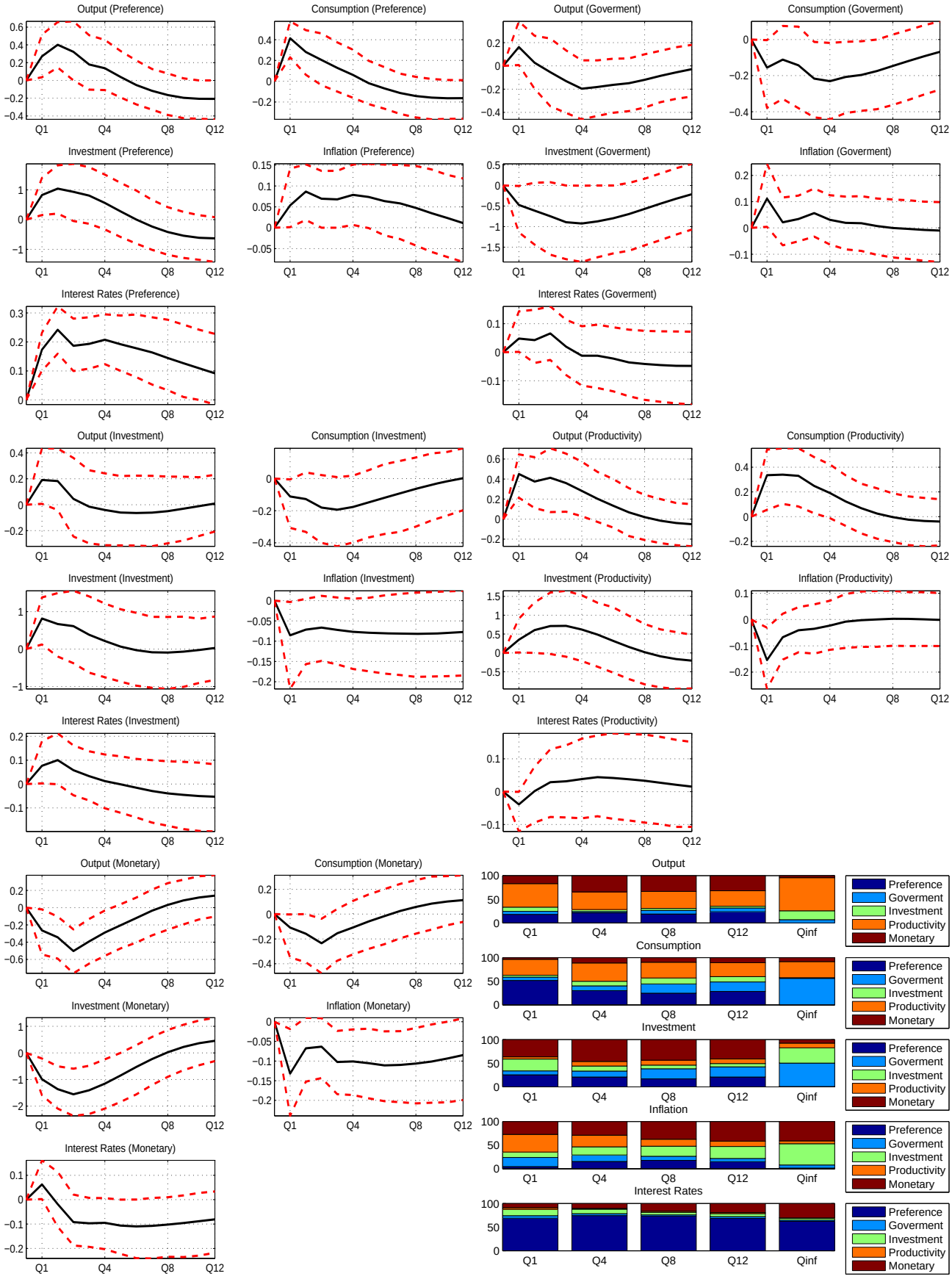


Figure 11: FHT-BVAR Posterior IRFs & FVD

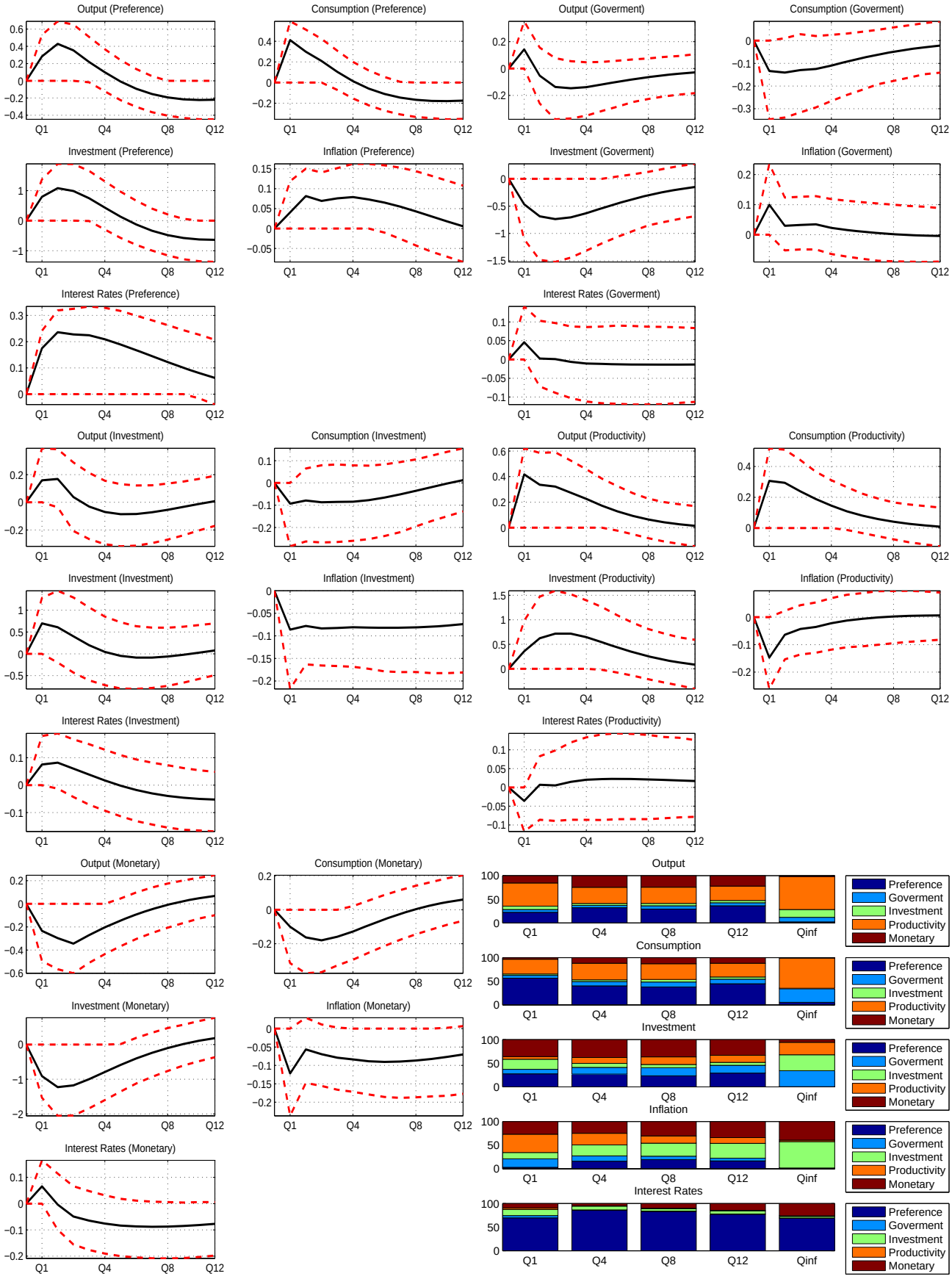


Figure 12: IRF Comparison

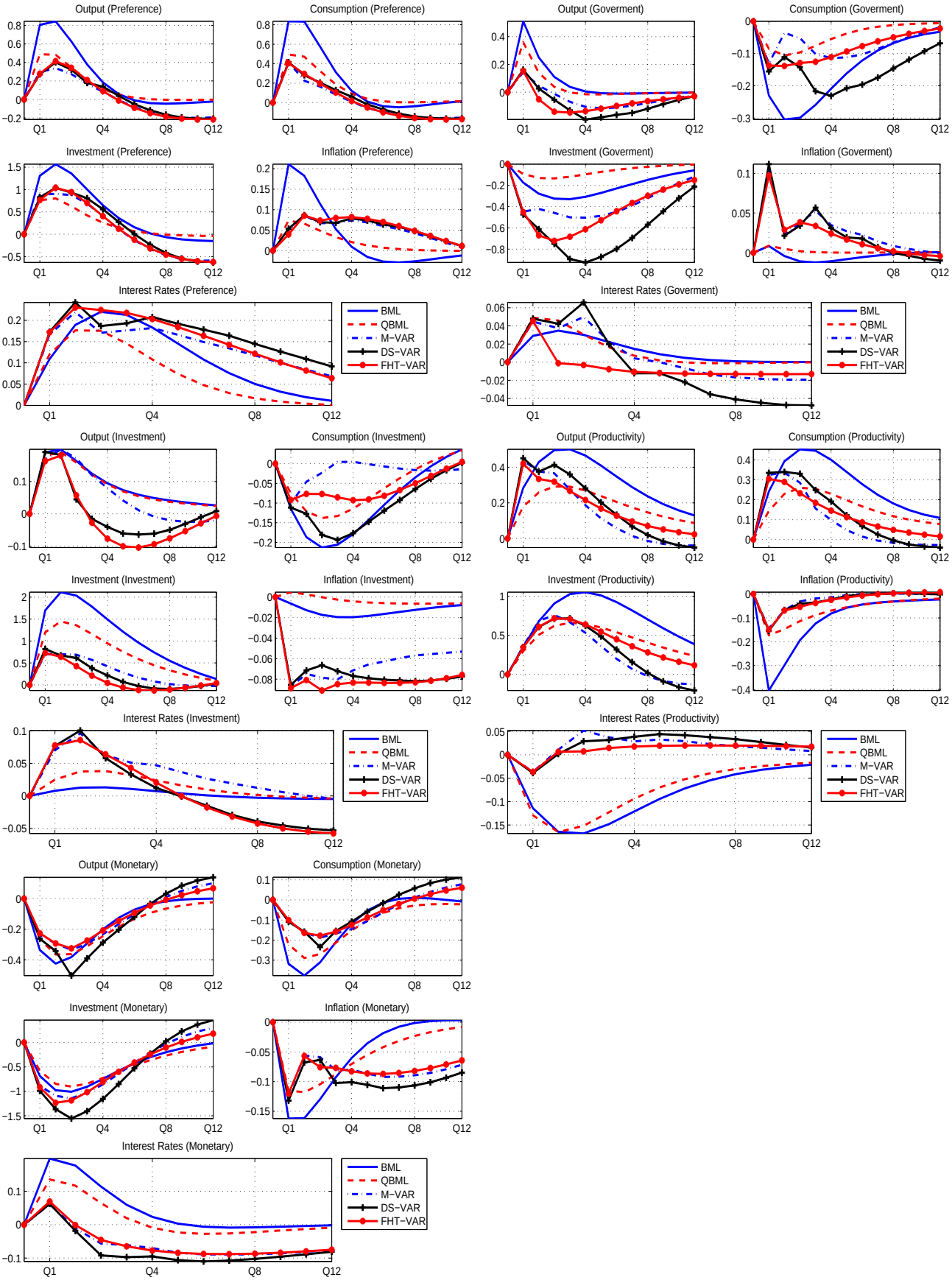


Figure 13: Mean Square Forecast Error

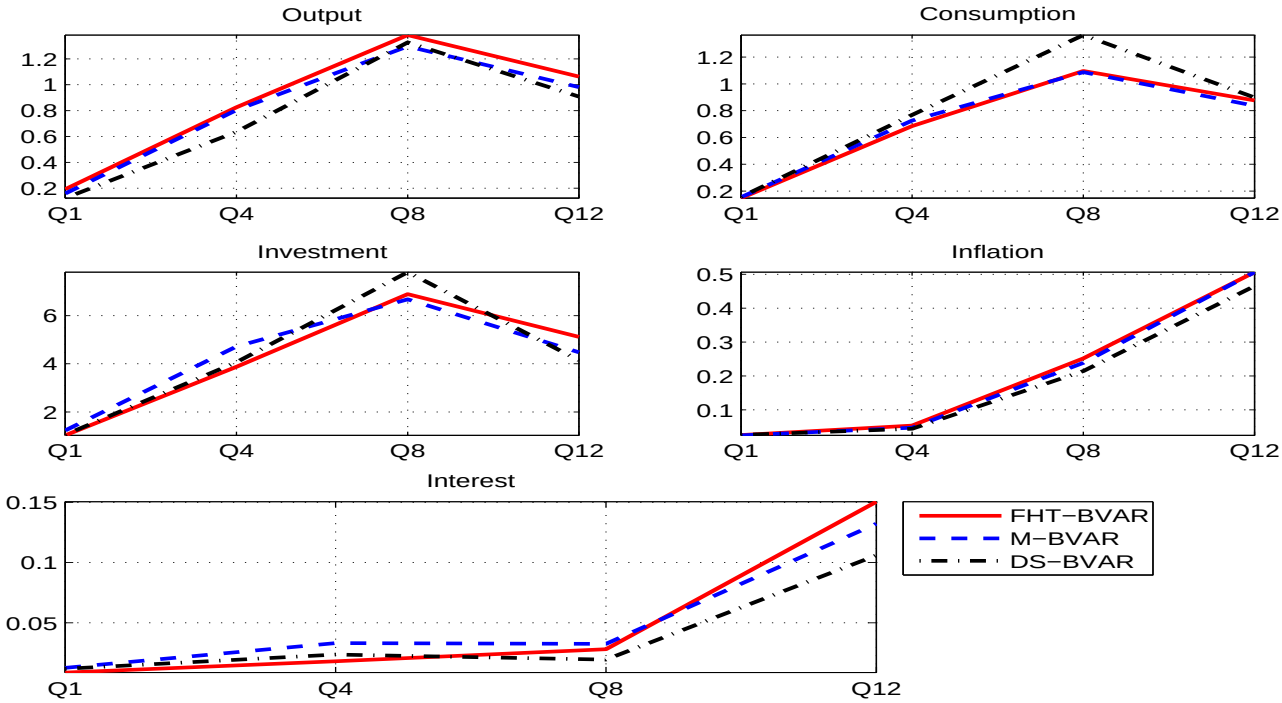
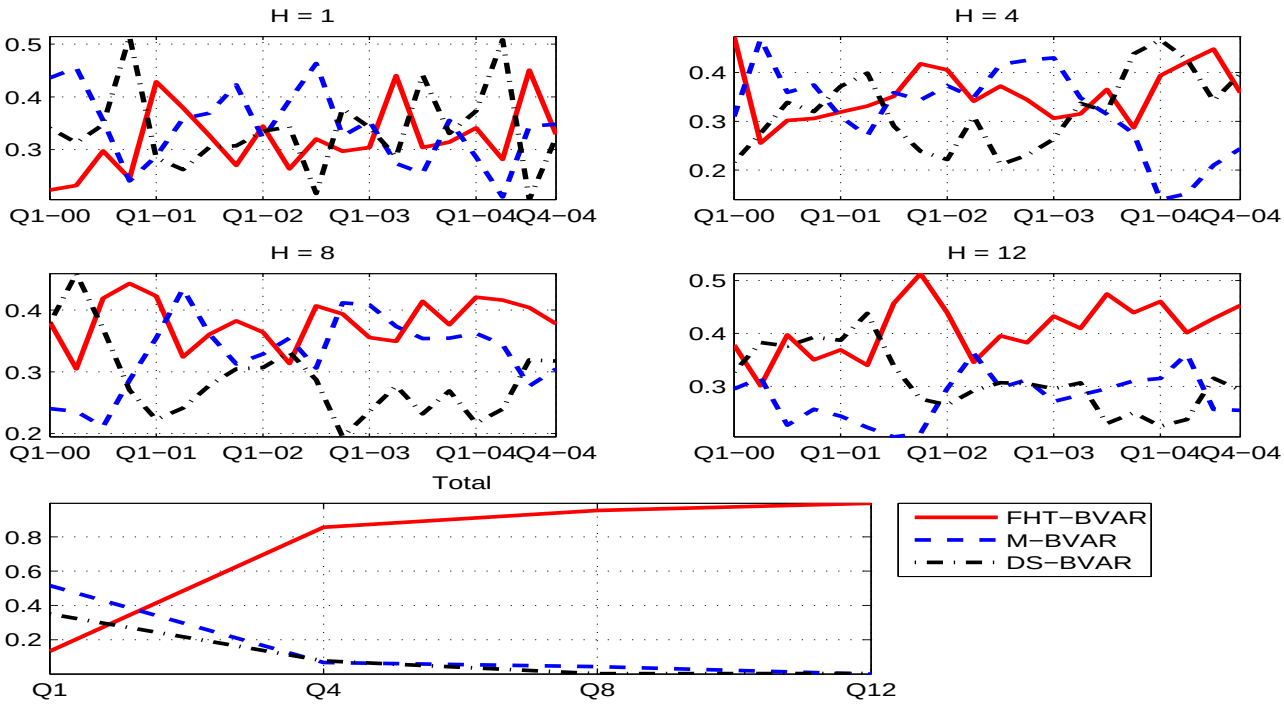


Figure 14: Forecast Probability Weights



E Tables

Table 1: DSGE Parameter Description & Prior Distribution

Symbols	*Description*	*Prior Distribution*
B	Fixed Cost	Normal
S"	Steady State Capital Adjustment Cost Elasticity	Normal
α	Capital Production Share	Normal
σ	Intertemporal Substitution	Normal
h	Habit Persistence	Beta
ξ_w	Wages Calvo Parameter	Beta
σ_l	Labour Supply Elasticity	Normal
ξ_p	Prices Calvo Parameter	Beta
i_w	Wage Indexation	Beta
i_p	Price Indexation	Beta
z	Capital Utilisation Adjustment Cost	Beta
ϕ_π	Taylor Inflation Parameter	Normal
ϕ_r	Taylor Inertia Parameter	Beta
ϕ_y	Taylor Output Gap Parameter	Normal
ρ_i	Investment Shock Persistence	Beta
ρ_g	Government Spending Shock Persistence	Beta
ρ_a	Productivity Shock Persistence	Beta
ρ_b	Premium Shock Shock Persistence	Beta
ρ_r	Monetary Policy Shock Shock Persistence	Beta
σ_i	Investment Shock Uncertainty	Inv. Gammma
σ_g	Government Spending Shock Uncertainty	Inv. Gammma
σ_a	Productivity Shock Uncertainty	Inv. Gammma
σ_b	Risk Premium Shock Uncertainty	Inv. Gammma
σ_r	Policy Shock Uncertainty	Inv. Gammma

Table 2: DSGE Evaluation

Model	$\mathbb{E}_\gamma \mathcal{W}$	$\sqrt{\mathbb{E}_\gamma \left(\mathcal{W} - \mathbb{E}_\gamma \mathcal{W}\right)^2}$
BML	1257.300	215.452
QBML	224.727	9.178

Table 3: Marginal Likelihood

Model	Value
BML	-1100.5
FHT-BVAR	-949.679
DS-BVAR	-893.533
M-BVAR	-995.104

Table 4: Full versus Limited Information Bayesian Monte Carlo Evaluation

	Prior	Moments	*BML*			*QBML*		
Parameters	*Mean*	*STD*	*Median*	*Bias*	*STD*	*Median*	*Bias*	*STD*
σ_a	0.462	0.100	0.461	0.002	0.069	0.451	0.023	0.013
σ_b	0.182	0.100	0.137	0.245	0.106	0.166	0.089	0.087
σ_g	0.609	0.100	0.560	0.080	0.095	0.613	0.006	0.018
σ_r	0.460	0.100	0.465	0.012	0.091	0.434	0.056	0.070
σ_q	0.240	0.100	0.180	0.248	0.097	0.268	0.119	0.070
ρ_a	0.950	0.100	0.948	0.002	0.131	0.985	0.036	0.018
ρ_b	0.180	0.100	0.113	0.371	0.085	0.169	0.061	0.046
ρ_g	0.970	0.100	0.861	0.112	0.189	0.983	0.013	0.009
ρ_q	0.710	0.100	0.813	0.145	0.097	0.698	0.017	0.059
ρ_m	0.120	0.100	0.084	0.300	0.094	0.026	0.782	0.021
S''	5.740	1.500	6.356	0.107	1.036	5.583	0.027	0.392
σ	0.900	0.375	0.888	0.013	0.353	0.966	0.074	0.242
h	0.710	0.050	0.684	0.037	0.045	0.725	0.021	0.013
ξ_w	0.700	0.100	0.763	0.091	0.080	0.717	0.025	0.017
σ_l	1.830	0.750	1.841	0.006	0.568	1.818	0.007	0.057
ξ_p	0.660	0.100	0.695	0.052	0.089	0.712	0.078	0.038
i_w	0.580	0.150	0.591	0.019	0.063	0.595	0.025	0.012
i_p	0.240	0.150	0.191	0.205	0.144	0.190	0.207	0.065
z	0.540	0.150	0.547	0.014	0.130	0.536	0.008	0.029
B	1.600	0.125	1.592	0.005	0.054	1.607	0.004	0.013
ϕ_π	2.040	0.150	2.059	0.009	0.086	2.071	0.015	0.033
ϕ_r	0.810	0.100	0.749	0.075	0.091	0.814	0.005	0.052
ϕ_y	0.300	0.050	0.291	0.030	0.038	0.292	0.026	0.014
ρ_{ga}	0.050	0.050	0.051	0.024	0.013	0.049	0.022	0.001
α	0.190	0.050	0.176	0.072	0.045	0.193	0.014	0.018
Total Bias				2.275			1.762	
Average Bias				0.091			0.070	
Total STD					3.890			1.408

Table 5: Full versus Limited Information Bayesian Estimation

	Prior	*Moments*	*BML*			*QBML*		
Parameters	*Mean*	*STD*	*Median*	*LwB*	*UpB*	*Median*	*LwB*	*UpB*
σ_a	0.500	0.100	0.832	0.576	1.017	0.533	0.384	0.710
σ_b	0.500	0.100	0.331	0.266	0.395	0.237	0.193	0.271
σ_g	0.500	0.100	0.676	0.605	0.732	0.413	0.355	0.462
σ_r	0.500	0.100	0.679	0.586	0.808	0.539	0.432	0.623
σ_q	0.500	0.100	0.273	0.238	0.305	0.265	0.229	0.309
ρ_a	0.750	0.100	0.797	0.713	0.905	0.811	0.729	0.892
ρ_b	0.750	0.100	0.627	0.504	0.731	0.514	0.325	0.615
ρ_g	0.750	0.100	0.718	0.636	0.818	0.517	0.42	0.666
ρ_q	0.750	0.100	0.433	0.318	0.547	0.402	0.314	0.539
ρ_m	0.500	0.200	0.373	0.177	0.524	0.718	0.611	0.789
S''	4.000	1.500	2.241	2.040	3.624	2.245	2.019	3.513
σ	0.850	0.375	0.719	0.627	0.847	0.840	0.725	1.039
h	0.700	0.050	0.731	0.640	0.777	0.715	0.653	0.784
ξ_w	0.500	0.100	0.548	0.465	0.703	0.565	0.479	0.697
σ_l	2.000	0.750	1.870	0.988	3.427	2.425	1.096	3.486
ξ_p	0.500	0.100	0.427	0.294	0.563	0.541	0.452	0.681
i_w	0.500	0.150	0.593	0.247	0.839	0.520	0.288	0.857
i_p	0.500	0.150	0.115	0.036	0.241	0.060	0.032	0.172
z	0.500	0.150	0.678	0.428	0.869	0.633	0.390	0.875
B	1.250	0.125	1.184	1.030	1.409	1.367	1.090	1.495
ϕ_π	1.750	0.150	1.558	1.367	1.905	1.801	1.510	2.159
ϕ_r	0.750	0.100	0.790	0.748	0.847	0.517	0.501	0.589
ϕ_y	0.125	0.050	0.247	0.178	0.347	0.254	0.165	0.324
ρ_{ga}	0.100	0.050	0.105	0.064	0.147	0.097	0.044	0.148
α	0.300	0.050	0.192	0.127	0.211	0.194	0.164	0.242