



BANK OF ENGLAND

# Staff Working Paper No. 716

## DSGE-based priors for BVARs and quasi-Bayesian DSGE estimation

Thomai Filippeli, Richard Harrison and  
Konstantinos Theodoridis

March 2018

Staff Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Bank of England or to state Bank of England policy. This paper should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee or Prudential Regulation Committee.



BANK OF ENGLAND

# Staff Working Paper No. 716

## DSGE-based priors for BVARs and quasi-Bayesian DSGE estimation

Thomai Filippeli,<sup>(1)</sup> Richard Harrison<sup>(2)</sup> and Konstantinos Theodoridis<sup>(3)</sup>

### Abstract

We present a new method for estimating Bayesian vector auto-regression (VAR) models using priors from a dynamic stochastic general equilibrium (DSGE) model. We use the DSGE model priors to determine the moments of an independent Normal-Wishart prior for the VAR parameters. Two hyper-parameters control the tightness of the DSGE-implied priors on the autoregressive coefficients and the residual covariance matrix respectively. Determining these hyper-parameters by selecting the values that maximize the marginal likelihood of the Bayesian VAR provides a method for isolating subsets of DSGE parameter priors that are at odds with the data. We illustrate the ability of our approach to correctly detect incorrect DSGE priors for the variance of structural shocks using a Monte Carlo experiment. We also demonstrate how posterior estimates of the DSGE parameter vector can be recovered from the BVAR posterior estimates: a new 'quasi-Bayesian' DSGE estimation. An empirical application on US data reveals economically meaningful differences in posterior parameter estimates when comparing our quasi-Bayesian estimator with Bayesian maximum likelihood. Our method also indicates that the DSGE prior implications for the residual covariance matrix are at odds with the data.

**Key words:** BVAR, DSGE, DSGE-VAR, Gibbs sampling, marginal likelihood evaluation, predictive likelihood evaluation, quasi-Bayesian DSGE estimation.

**JEL classification:** C11, C13, C32, C52.

---

(1) Queen Mary, University of London. Email: t.filippeli@qmul.ac.uk

(2) Bank of England. Email: richard.harrison@bankofengland.co.uk

(3) Cardiff Business School. Email: theodoridis1@cardiff.ac.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. We are grateful to Jana Eklund, Francesca Monti, Gary Koop, Tony Yates, Emilio Fernandez-Corugedo, Haroon Mumtaz and Martin Andreasen for helpful comments and insightful discussions.

The Bank's working paper series can be found at [www.bankofengland.co.uk/working-paper/Working-papers](http://www.bankofengland.co.uk/working-paper/Working-papers)

Publications and Design Team, Bank of England, Threadneedle Street, London, EC2R 8AH  
Telephone +44 (0)20 7601 4030 email [publications@bankofengland.co.uk](mailto:publications@bankofengland.co.uk)

# 1 Introduction

We present a new method for estimating Bayesian vector autoregression (VAR) models using priors from a dynamic stochastic general equilibrium (DSGE) model. In doing so, we contribute to a rich literature that seeks to combine the theoretical structure of DSGE models with the fit and forecasting performance of VAR models.

Dynamic stochastic general equilibrium (DSGE) models have become a workhorse for macroeconomic policy analysis. Part of the attraction of these models is the discipline imposed by a tight theoretical structure. The structure implies that the properties of the model are determined by a relatively small number of ‘deep’ parameters, describing the preferences and constraints of optimizing agents. While the small number of parameters sharpens policy analysis, it also places many restrictions on the implied time series properties of endogenous variables. So *a priori* we expect DSGE models to be at a disadvantage when used to fit the time series properties compared with more densely parameterized models such as VARs.

This observation prompted the development of a literature exploring the links between DSGE and VAR representations of the data and seeking to use information from DSGE models to apply some structure to VAR priors. As we discuss in Section 2, a key contribution to this literature is the DSGE-VAR methodology proposed by [del Negro and Schorfheide \(2004\)](#), henceforth ‘DNS’. They utilize a mapping from the linearized DSGE model solution to a VAR representation to build a hierarchical prior. The researcher specifies a prior for the vector of deep DSGE model parameters. The mapping can then be used to compute the coefficients of the implicit VAR representation, using the population moments of the data implied by the DSGE model. This is used as the prior for Bayesian VAR estimation.

The DNS approach can be interpreted in terms of augmenting the data sample with a sample of artificial observations from the DSGE model. A hyperparameter specifies the size of the sample of artificial observations relative to the actual data sample. DNS propose that the hyperparameter be selected to maximize the marginal data density of the VAR.

We also use the moments of the VAR representation of the DSGE model implied by the priors for the DSGE model parameters. However, in contrast to DNS, we use draws of the DSGE parameter vector from the prior distribution to characterize the moments of VAR representation. These moments are used to parameterize an independent Normal-Wishart distribution which forms the prior for the Bayesian VAR estimation. Using a non-conjugate prior increases the computational demands of our procedure. However, our approach uses more information from the DSGE model priors. Specifically, the second moments of the VAR priors capture the variance of the VAR coefficients implied by the priors for the deep DSGE parameters, rather than the variance of the *OLS estimators* of those parameters as in DNS.

Two hyper-parameters control the tightness of the DSGE-based priors on the VAR autoregressive coefficients and the residual covariance matrix respectively. We study cases in which the values of these hyper-parameters are selected to maximize the marginal data density. We show that this can be used as a method for isolating subsets of the DSGE parameter for which the priors are substantially at odds with the data. In particular, the parameters governing the variances of the structural shocks in the DSGE model have a strong influence on the implied covariance matrix of VAR residuals. If the relevant hyper-parameter indicates that DSGE prior information on the residual covariance matrix

should be down weighted, this may indicate that the priors for the parameters related to the structural shocks are out of line with the data.

We verify this property of our method using a Monte Carlo experiment with data generated by a variant of the medium-scale DSGE model of [Smets and Wouters \(2007\)](#). We study the case in which our procedure is applied using misspecified priors for the variance of the structural shocks. We find that our approach consistently down weights the DSGE-based prior for the VAR residual covariance matrix. Importantly, the hyper-parameter on the autoregressive VAR parameters is often chosen to place a relatively large weight on the DSGE-based prior, allowing our approach to selectively use information from the DSGE parameter priors. This property implies that, in our Monte Carlo study, our procedure fits better on average than the DNS method. That is because the DNS approach down-weights *all* information from the DSGE-based priors, even when priors for only a subset of the model parameters are substantially at odds with the data.

We also show how our method can be used to uncover estimates of the posterior distributions for the DSGE model parameters from the posterior distributions of the BVAR parameters. Posterior distributions from this ‘quasi-Bayesian’ procedure can be compared with the posterior estimates from Bayesian maximum likelihood estimation of the DSGE model. If the DSGE model estimated using Bayesian maximum likelihood fits poorly, then such a comparison can provide useful information about which particular parameters within the DSGE model parameter vector may be contributing to the poor fit. To illustrate this, we present an empirical application of our approach using US data and the [Smets and Wouters \(2007\)](#) model. We compare posterior DSGE parameter estimates from our approach to those obtained using Bayesian maximum likelihood. This comparison indicates some economically meaningful differences in the implied behavior of the model, including the parameters of the monetary policy reaction and the response of the model to a monetary policy shock: two topics of considerable interest to applied macroeconomists. As in our Monte Carlo study, our empirical application implies that the DSGE-based prior for the residual covariance matrix is at odds with the data.

The rest of the paper is organized as follows. The existing approaches for constructing priors from structural models are reviewed in Section 2. Our proposed methodology is described in Section 3. Section 4 presents a Monte Carlo exercise comparing our method to that of [del Negro and Schorfheide \(2004\)](#). An empirical application is presented in Section 5 and the final section concludes.

## 2 Existing literature and our contribution

As noted in the Introduction, there is a rich literature exploring the links between DSGE and VAR models. A seminal contribution is [Ingram and Whiteman \(1994\)](#), who construct priors for a BVAR using a simple Real Business Cycle (RBC, [King et al., 1988](#)) model. The authors exploit the simplicity of their model and the properties of the normal distribution to allow posterior estimation of the VAR implied by the DSGE model using single equation ‘mixed’ estimation procedures introduced by [Theil and Goldberger \(1961\)](#).

[De Jong et al. \(1993\)](#) built on this approach by assuming that the prior distribution for the VAR parameters has a conjugate Normal-Wishart distribution, estimating its moments using OLS on stochastically simulated data. Conjugate priors imply that the posterior moments of the VAR parameter

vector can be written as weighted averages of the prior moments and the OLS estimates. One implication of the De Jong et al. approach is that if the priors are strongly at odds with the data, then posterior inference would be damaged since the priors are dogmatically applied in their method.

The work of del Negro and Schorfheide (2004, 'DNS') provides both a theoretical underpinning for the De Jong et al. approach and a way to deal with issues arising from imposing priors that are strongly at odds with the data. As in De Jong et al., the actual data set is augmented with a number of artificial observations simulated from the DSGE model. The number of artificial observations is proportional to the size of the actual sample ( $T$ ) by a factor  $\lambda_{DS} \in (0, \infty)$ . DNS factorize the VAR likelihood of the augmented sample into the likelihood of the actual data and the likelihood of the artificial data, with the latter interpreted as the prior density of the VAR parameter vector  $\theta$ .<sup>1</sup> This representation of the likelihood can be viewed as the posterior kernel of the VAR parameter vector.

DNS show that the posterior distribution of the DSGE parameter vector can be obtained by combining the marginal likelihood of the VAR (which in their case has an analytic form) with the prior distribution of the DSGE parameter vector. DNS highlight that the empirical performance of the time-series model crucially depends on the choice of  $\lambda_{DS}$  and, therefore, they recommend that it is selected based on measures of fit such as the marginal likelihood. The resulting posterior density of the DSGE parameter vector consists of the parameter values that minimize the distance between the OLS estimated VAR parameter vector and the VAR parameters implied by the DSGE model. Loosely speaking, this can be viewed as the Bayesian version of the estimator proposed by Smith (1993).

In our view, the DNS approach suffers from two drawbacks. The first is that the DSGE-based prior moments are constructed by estimating a VAR( $p$ ) on data generated by the DSGE model and using the moments of the OLS estimators to characterise the priors. This is not in fact the prior distribution of the VAR parameters, vector but rather the distribution of the OLS *estimates* of the VAR (computed under the assumption that the DSGE model is the true data generating process). If the DSGE model, is the true data generating process then consistency of the OLS estimator ensures that the VAR estimates converge in probability to the VAR parameters implied by the DSGE model as the pseudo sample becomes infinitely large. However, the OLS covariance matrix does not measure the dispersion of the prior probability density function of the VAR parameters implied by the DSGE model: it measures the variance of the *estimation error*, which converges to zero as the pseudo sample becomes infinitely large.

These observations suggest that information captured by the second moments of the prior density of the DSGE model parameters is not fully utilised. The variance of the priors over the DSGE parameters reflects the researcher's confidence over the range of values that the parameter vector may take. It is likely that the researcher will have tighter priors over some parameters than others. In this case, some VAR parameters should also receive a larger prior weight than others. In our view, this is an important piece of information that helps researchers to improve the properties of DSGE models and so should not be ignored. For instance, the distribution of the VAR impulse responses under the prior depends on the variance of  $\theta$ , which under DSGE-based priors should also depend on the variance of  $\gamma$ .

The second drawback of the DNS approach is that the functional form of the prior covariance matrix

---

<sup>1</sup>To avoid the stochastic variation rising by the simulation of the model, the authors replace the non-standardized sample moments of the likelihood with their expected values.

strongly restricts the covariances among the VAR coefficients. For instance, it implies that the prior covariance matrix for each reduced form VAR equation has the same structure (up to a scaling factor) (see, De Jong et al., 1993; Sims and Zha, 1998). There is no reason to suggest that DSGE models would in general give rise to VAR representations with this property.

Our approach therefore allows a more complete and more flexible specification of the prior implied by the DSGE model. The inclusion of two hyperparameters allows our approach to detect possible sources of misspecification in the DSGE priors.

### 3 Our Method

Our approach uses the well-known mapping between the DSGE and VAR models together with the prior distribution of the DSGE parameter vector to derive the prior moments of the VAR parameter vector.<sup>2</sup> We introduce two hyperparameters to ensure well-behaved posterior inference. We also illustrate how the BVAR posterior moments can be used to obtain the posterior distribution of the DSGE parameter vector.

We start from the state space representation of the rational expectations solution of a (log) linearized DSGE model ( $\mathcal{M}$ ), which is given by

$$y_t = A(\gamma) \xi_t \tag{3.1}$$

$$\xi_t = B(\gamma) \xi_{t-1} + \Upsilon(\gamma) \omega_t \tag{3.2}$$

Equation (3.2) describes the evolution of the state vector of the model,  $\xi_t \in \mathbb{R}^{d_\xi}$ , where we use the notation  $da$  to indicate the dimension of the vector  $a$ . Equation (3.1) is the measurement equation mapping the unobserved state of the economy to the observable variables,  $y_t \in \mathbb{R}^{d_y}$ . We assume that the vector of the shocks,  $\omega_t \in \mathbb{R}^{d_\omega}$ , is normally distributed with mean zero and identity covariance matrix. The elements of the matrices  $A(\gamma)$ ,  $B(\gamma)$ , and  $\Upsilon(\gamma)$  are nonlinear functions of the DSGE parameter vector (the ‘structural parameter vector’),  $\gamma \in \Gamma$ .

The VAR( $p$ ) model ( $\mathcal{T}$ ) of the observable variables  $y$ , is:

$$y_t = \sum_{i=1}^p \Psi_i y_{t-i} + v_t \tag{3.3}$$

where the vector of the reduced form errors is normally distributed as  $v_t \sim N(0, \Sigma_v)$ . The standard regression representation of the VAR is:

$$Y = \Psi X + U \tag{3.4}$$

where  $\Psi = \begin{bmatrix} \Psi_1 & \cdots & \Psi_p \end{bmatrix}$  is the  $d_y \times (p \cdot d_y)$  matrix of the VAR coefficients,  $T$  is the sample size,  $Y$  is the  $d_y \times T$  data matrix of the observed variables,  $X$  is the  $(p \cdot d_y) \times T$  matrix of the lagged data and  $U$  is the  $d_y \times T$  matrix of the VAR innovations. In what follows, we will define  $\psi \equiv \text{vec}(\Psi)'$  and

---

<sup>2</sup>This mapping has been explored by many others, including Fernandez-Villaverde et al. (2007), Christiano et al. (2006) and Ravenna (2007).

$\sigma_v \equiv \text{vech}(\Sigma_v)'$  as the components of the VAR parameter vector  $\theta \equiv (\psi', \sigma_v')' \in \Theta$ .<sup>3</sup>

We depart from the existing literature by moving away from the assumption of conjugate priors. This is achieved by assuming that the prior distribution of the VAR parameter vector  $p(\theta)$  is independent Normal-Wishart. This slightly complicates the implementation of the approach, but the benefits repay the additional effort. In particular,  $p(\theta)$ , can be interpreted as the prior distribution of the reduced-form parameter vector  $\theta(\gamma)$ . As noted in Section 2, this contrasts with previous approaches (such as the method of del Negro and Schorfheide (2004)) that proxy this prior using the distribution of the OLS estimator  $\hat{\theta}$  under the assumption that the DSGE model is the true data generating process. As noted in Section 2, using the distribution of the OLS estimator  $\hat{\theta}$  produces some counterintuitive implications for the prior covariance matrix of  $\theta$ . Our approach circumvents these difficulties.

### 3.1 The Prior Moments of $\theta$

The starting point of our methodology is the DSGE model summarised by equations (3.1) and (3.2) 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<sup>4</sup>

$$M(\gamma) \equiv \left[ I_{da\xi} - \Upsilon(\gamma) [A(\gamma) \Upsilon(\gamma)]^{-1} A(\gamma) \right] B(\gamma) \quad (3.5)$$

are less than one in absolute terms, then there is an analytical mapping between the structural and VAR parameter vector:<sup>5</sup>

$$\phi : \gamma \rightarrow \theta \quad (3.6)$$

namely,

$$\psi(\gamma)_i = \text{vec}(\Psi(\gamma)_i) \quad (3.7)$$

$$\Psi(\gamma)_i \equiv A(\gamma) B(\gamma) M(\gamma)^{i-1} \Upsilon(\gamma) [A(\gamma) \Upsilon(\gamma)]^{-1} \quad (3.8)$$

$$\sigma_v(\gamma) = \text{vec}([A(\gamma) \Upsilon(\gamma)] [A(\gamma) \Upsilon(\gamma)]') \quad (3.9)$$

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’ or the ‘Change of Variable Theorem’ :  $\phi$  is not an injection (Koop, 2003, pp. 334).<sup>6</sup>

The analysis proceeds with the assumption that the prior for  $\theta(\gamma)$  is the *independent* Normal-Wishart distribution and its moments are approximated through stochastic simulation.<sup>7</sup> Specifically our prior

<sup>3</sup>The *vec* operators transforms a  $da \times da$  matrix to an  $da^2 \times 1$  vector by stacking the columns. The *vech* operator transforms a  $da \times da$  matrix to an  $(da(da+1)/2) \times 1$  vector by stacking the elements of and below the main diagonal.

<sup>4</sup>We define  $I_{da}$  as the  $(da \times da)$  identity matrix.

<sup>5</sup>This condition is known as the Poor Man’s Invertibility Condition (PMIC) (Fernandez-Villaverde et al., 2007). The mapping between the DSGE and VAR models that we use relies on the PMIC being satisfied. Though the requirement that the DSGE model satisfy the PMIC reduces the set of models to which our proposed procedure can be applied, this is also true of alternative approaches.

<sup>6</sup>See Ingram and Whiteman (1994) for a discussion of the mean value theorem.

<sup>7</sup>The terminology ‘Normal-Wishart’ to describe our prior comes from the time series econometrics literature which would often specify regression models in terms of the slope coefficients and ‘precision’ of the errors (the inverse of the variance of the errors). The Normal-Wishart specification applies to the slope and precision parameters. So in our case, the prior for the covariance matrix will be *inverse* Wishart.



for the VAR parameters are given by:

$$p(\psi(\gamma), \Sigma_v(\gamma)) = p(\psi(\gamma))p(\Sigma_v(\gamma)) \quad (3.10)$$

$$p(\psi(\gamma)) = N(\mu_\psi(\gamma), \Sigma_{\lambda\psi}(\gamma)) \quad (3.11)$$

$$p(\Sigma_v(\gamma)) = IW(\Pi(\gamma)^{-1}, \eta) \quad (3.12)$$

where  $p(a)$  denotes the prior distribution of vector  $a$ . We use the notation  $N(\mu_\alpha, \Sigma_\alpha)$  to represent the normal distribution, where  $\mu_\alpha$  and  $\Sigma_\alpha$  denote the mean and the covariance matrix of the vector  $\alpha$ , respectively. The Wishart distribution and its inverse are defined as  $W(\Pi, \eta)$  and  $IW(\Pi^{-1}, \eta)$ , respectively, where  $\eta$  is the degrees of freedom and  $\Pi$  is the scale matrix. The matrix  $\Sigma_{\lambda\psi}$  is described below.

Using equations (3.5)–(3.9) and 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,  $\{\theta_j\}_{j=1}^S$ .<sup>8</sup> Then from Theorem 3.1 and Proposition 3.2 of White (2001) it is known that the estimated moments converge to the true moments almost surely:

$$\hat{\mu}_\psi(\gamma) \equiv S^{-1} \sum_{j=1}^S \psi(\gamma)_j \xrightarrow{a.s.} \mu_\psi(\gamma) \quad (3.13)$$

$$\hat{\mu}_\sigma(\gamma) \equiv S^{-1} \sum_{j=1}^S \sigma(\gamma)_j \xrightarrow{a.s.} \mu_\sigma(\gamma) \quad (3.14)$$

$$\hat{\Sigma}_\psi(\gamma) \equiv S^{-1} \sum_{j=1}^S \left( \psi(\gamma)_j - \hat{\mu}_\psi \right) \left( \psi(\gamma)_j - \hat{\mu}_\psi \right)' \xrightarrow{a.s.} \Sigma_\psi(\gamma) \quad (3.15)$$

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

$$\mu_\sigma(\gamma) = \frac{1}{\eta - dy - 1} \text{vec}(\Pi) \text{ or } \Pi(\gamma) = (\eta - dy - 1) \Pi^*(\gamma) \quad (3.16)$$

where  $\text{vec}(\Pi^*(\gamma)) \equiv \mu_\sigma(\gamma)$ . In order to study the variance of the prior distribution of  $\Sigma_v$  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)<sup>9</sup>

$$\Sigma_{\sigma^{-1}}(\gamma) \equiv \frac{1}{\eta - dy - 1} \left( I_{dy^2} + K_{dy, dy} \right) \left( (\Pi^*(\gamma))^{-1} \otimes (\Pi^*(\gamma))^{-1} \right) \quad (3.17)$$

Expressions (3.5)–(3.17) establish the mapping between the moments of  $p(\gamma)$  and  $p(\theta(\gamma))$ .

<sup>8</sup>The steps are described in Appendix B.

<sup>9</sup>The symbol  $\otimes$  denotes the Kronecker product operator and  $K_{dm, dn}$  is a commutation matrix, such that for any  $dm \times dn$  matrix  $G$ ,  $K_{dm, dn} \text{vec}(G) = \text{vec}(G')$ .



### 3.2 Controlling the tightness of the DSGE priors

The economic assumptions underpinning DSGE models impose many restrictions on the behavior of the data because the parameter vector  $\gamma$  is typically of relatively low dimension, compared with the number of parameters in a comparable VAR. Here we explain how these constraints can be relaxed inside our framework.

#### Singularity of $\Sigma_\psi(\gamma)$

DSGE models are typically able to describe the behavior of a set of variables using a relative small number of so-called ‘deep’ parameters, relating to the optimization problems being solved by the agents in the model. In contrast, VAR models tend to have many more parameters since the lag order of a VAR is typically chosen to provide a good fit to the observed data.

This implies that the dimension of  $\gamma$  is expected to be (much) smaller than  $\psi$ . In turn, this means that  $\Sigma_\psi$ , the covariance matrix of the VAR parameters implied by the DSGE model, may not be positive definite. To avoid working with singular distributions, we impose more structure on  $\Sigma_\psi$ , by assuming that it is a block diagonal matrix.<sup>10</sup> Assuming that  $\Sigma_\psi$  is diagonal (similar to Minnesota priors) seems a natural choice when we are not very confident about the DSGE model’s predictions regarding the cross-moments between the autoregressive coefficients and so do not wish to impose these restrictions on the data. Our approach therefore allows us to set a more flexible prior for the covariance matrix of the autoregressive parameters,  $\psi$ , in contrast to the [del Negro and Schorfheide \(2004\)](#) approach which, as noted in Section 2, forces the prior covariance structure to be identical (up to scale) across VAR equations.

#### Hyper-Parameters

The work of [De Jong et al. \(1993\)](#) and [del Negro and Schorfheide \(2004\)](#) suggests that a device that relaxes the ‘strength’ of the DSGE model priors when they are at odds with the data is desirable to ensure well-behaved posterior inference.

Note that in the current framework  $\mu_\psi(\gamma)$  and  $\mu_\sigma(\gamma)$  represent the DSGE model’s predictions of the mean of the VAR coefficient and the residual variance-covariance vector, respectively. Since sample estimates from the data,  $\hat{\mu}_\psi$  and  $\hat{\mu}_\sigma$ , are easily computed, we need to ensure that these estimates lie within the support of the prior distribution implied by the DSGE model. This is possible by adjusting the variance around  $\mu_\psi(\gamma)$  and  $\mu_\sigma(\gamma)$ .

From equation (3.17) it is evident that a suitable choice of the degrees of freedom  $\eta$  allows us to either ‘shrink’ or ‘loosen’ the distribution around  $\mu_{\sigma-1}(\gamma)$ .<sup>11</sup> However, such a parameter does not exist for  $\Sigma_\psi(\gamma)$ . We therefore introduce the hyperparameter  $\lambda \in (0, \infty)$  to define:

$$\Sigma_{\lambda\psi}(\gamma) \equiv \lambda \Sigma_\psi(\gamma) \tag{3.18}$$

---

<sup>10</sup>The length and, consequently, the number of blocks depends on the number of the structural parameters. For instance, in the application in Section 5, 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 is seventy five ( $\Sigma_\psi$  is a diagonal matrix).

<sup>11</sup>Specifically,  $\Sigma_{\sigma-1}(\gamma) \rightarrow 0_{dy \times dy}$  as  $\eta \rightarrow \infty$  and  $\Sigma_{\sigma-1}(\gamma) \rightarrow \infty$  as  $\eta \rightarrow dy + 1$ .

Remark 3.1 below summarises the role of the hyper-parameter vector  $\bar{\lambda} \equiv (\lambda, 1/\eta)'$  in the posterior distribution of  $\theta$  and, consequently, illustrates why its selection should be based on measures of fit.

### 3.3 VAR Posterior Estimation

Our choice of independent Normal-Wishart priors means that the posterior distribution of  $\theta$  does not have an analytic form. However, Proposition 3.1 states that the posterior kernel of the VAR  $p(\theta|Y, \bar{\lambda})$  can be expressed as the product of two conditional distributions, meaning that it can be approximated using a Gibbs sampling scheme (Canova, 2005).

**Proposition 3.1** *If the prior for  $(\psi, \Sigma_v)$  is specified as in equations (3.10)–(3.12), then the posterior kernel of  $\theta$  can be written as the product of two conditional distributions*

$$p(\theta|Y, \bar{\lambda}) \propto N(\bar{\mu}_\psi, \bar{\Sigma}_\psi | \Sigma_v) IW(\bar{\Pi}, T + \eta | \Psi) \quad (3.19)$$

where

$$\bar{\Sigma}_\psi \equiv \left[ \hat{\Sigma}_\psi^{-1} + \Sigma_{\lambda\psi}(\gamma)^{-1} \right]^{-1} \quad (3.20)$$

$$\bar{\mu}_\psi \equiv \bar{\Sigma}_\psi \left[ \Sigma_{\lambda\psi}(\gamma)^{-1} \mu_\psi(\gamma) + \hat{\Sigma}_\psi^{-1} \hat{\psi} \right] \quad (3.21)$$

$$\bar{\Pi} \equiv \Pi(\gamma) + T\hat{\Sigma}_v + (\Psi - \hat{\Psi})' X' X (\Psi - \hat{\Psi}) \quad (3.22)$$

and  $\hat{\Sigma}_v$ ,  $\hat{\Psi}$ ,  $\hat{\psi}$  and  $\hat{\Sigma}_\psi$  are the OLS estimates of  $\Sigma_v$ ,  $\Psi$ ,  $\psi$  and  $\Sigma_\psi$ , respectively.  $\square$

The proof can be found in Appendix A.

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

**Remark 3.1** *From Proposition 3.1 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*

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

From the latter expression it can be concluded that

$$\bar{\mu}_\psi \rightarrow \hat{\psi} \quad \text{as } \lambda \rightarrow \infty \quad (3.23)$$

$$\bar{\mu}_\psi \rightarrow \mu_\psi(\gamma) \quad \text{as } \lambda \rightarrow 0 \quad (3.24)$$

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

$$\begin{aligned}\bar{\Sigma}_\psi &\equiv \left[ \hat{\Sigma}_\psi^{-1} + \Sigma_{\lambda\psi}(\gamma)^{-1} \right]^{-1} = \hat{\Sigma}_\psi - \hat{\Sigma}_\psi \left[ \hat{\Sigma}_\psi + \Sigma_{\lambda\psi}(\gamma) \right]^{-1} \hat{\Sigma}_\psi \\ &= \hat{\Sigma}_\psi - \hat{\Sigma}_\psi \left[ \hat{\Sigma}_\psi + \lambda \Sigma_\psi \right]^{-1} \hat{\Sigma}_\psi\end{aligned}\quad (3.25)$$

Meaning that

$$\bar{\Sigma}_\psi(\gamma) \rightarrow \hat{\Sigma}_\psi \quad \text{as } \lambda \rightarrow \infty \quad (3.26)$$

$$\bar{\Sigma}_\psi(\gamma) \rightarrow 0_{d_\psi \times d_\psi} \quad \text{as } \lambda \rightarrow 0 \quad (3.27)$$

3. From the properties of the inverse Wishart distribution is know that the posterior mean of  $\Sigma_v$  is (Poirier, 1995)

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

This implies that

$$\bar{\mu}_\sigma \rightarrow \text{vec} \left( \hat{\Sigma}_v \right) \quad \text{as } \eta - dy - 1 \rightarrow 0 \text{ and } \lambda \rightarrow \infty \quad (3.28)$$

$$\bar{\mu}_\sigma \rightarrow \mu_\sigma(\gamma) \quad \text{as } \eta \rightarrow \infty \quad (3.29)$$

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

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

It can be seen that

$$\text{vec}(\Sigma_{\sigma^{-1}}) \rightarrow \frac{1}{T} (I_{dy^2} + K_{dy,dy}) \left( \hat{\Sigma}_v^{-1} \otimes \hat{\Sigma}_v^{-1} \right) \quad \text{as } \eta \rightarrow dy + 1 \text{ and } \lambda \rightarrow \infty \quad (3.30)$$

$$\text{vec}(\Sigma_{\sigma^{-1}}) \rightarrow 0_{d_\sigma \times d_\sigma} \quad \text{as } \eta \rightarrow \infty \quad (3.31)$$

□

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

The previous section clearly shows that  $\bar{\lambda}$  affects the fit of the VAR. To guard against worsening empirical performance, the data can be used to guide our choice of this hyperparameter vector. This approach mirrors the suggestion of del Negro and Schorfheide (2004), to select the value of the hyperparameter in their DSGE-VAR approach based on model fit.

A natural measure for this purpose is the marginal likelihood of the VAR, defined as:

$$m_{\mathcal{T}}(Y) \equiv \int L_{\mathcal{T}}(Y|\theta) p(\theta) d\theta$$

where  $L_{\mathcal{T}}(Y|\theta)$  is the likelihood of the VAR. In this case  $\hat{\lambda}$  should be determined as:

$$\hat{\lambda} = \arg \max_{\bar{\lambda} \in (0, \infty) \times (dy+1, \infty)} m_{\mathcal{T}}(Y|\bar{\lambda}) \quad (3.32)$$

In our framework  $m_{\mathcal{T}}(Y|\bar{\lambda})$  does not have an analytic form. However, it can be approximated by the output of the Gibbs sampler using either the methodology developed by Chib (1995) or Geweke's modified harmonic mean estimator (Geweke, 1999).<sup>12</sup>

This estimation procedure implies that  $\hat{\lambda}$  can be interpreted as an indicator of model misspecification. If the DSGE model used to derive the prior of  $\theta$  is 'far away' from the true data generation process, then  $\hat{\lambda}$  is expected to be large, implying that  $\hat{\lambda}$  is an increasing function of model misspecification. Another useful by-product of estimation of  $\hat{\lambda}$  is that we are now able to identify which parts of the VAR parameter vector ( $\psi$  and/or  $\sigma$ ) are best (or worst) summarised by the DSGE model. The curvature of  $m_{\mathcal{T}}(Y|\bar{\lambda})$  surface with respect to  $\lambda$  and  $\eta$  identifies whether small changes in these arguments lead to large variations in the fit of the VAR, as measured by  $m_{\mathcal{T}}(Y|\bar{\lambda})$ .

### 3.4 Using quasi-Bayesian estimation to pinpoint misspecification

We argued in the previous section that  $\hat{\lambda}$  can be thought of as an indicator of model misspecification. In this section, we show how a form of quasi-Bayesian estimation can be useful in further examining the misspecification of the DSGE model. In the following subsections, we first present the estimation approach and then describe how this can be used for model evaluation exercises.

#### 3.4.1 Quasi-Bayesian estimation of $\gamma$

Our quasi-Bayesian estimator is a limited information estimator.<sup>13</sup> The studies of del Negro and Schorfheide (2004) and Christiano et al. (2010) can be viewed as attempts to incorporate an approach to selecting structural parameters based on limited information metrics within the Bayesian framework. Our approach takes a similar perspective and illustrates how the structural parameters can be selected to minimize the distance between the posterior moments of the BVAR parameters estimated using the approach described in Section 3.3 and the VAR parameters implied by the DSGE model.

From del Negro and Schorfheide (2004, Section 3.3.1), we know that the posterior distribution of  $\gamma$  can be obtained by combining the marginal likelihood of the VAR with the prior distribution of the structural parameter vector. However, in our case  $m_{\mathcal{T}}(Y|\gamma, \hat{\theta}, \hat{\lambda})$  does not have an analytic form. This difficulty is surmounted by using the Laplace approximation and the mapping (3.6).

<sup>12</sup>In practice, the researcher can create a two dimensional grid  $((0, \infty) \times (dy + 1, \infty))$  and evaluate  $m_{\mathcal{T}}(Y|\bar{\lambda})$  for all values of  $\bar{\lambda}$  inside this grid, setting  $\hat{\lambda}$  equal to the  $\bar{\lambda}$  that corresponds to  $m_{\mathcal{T}}(Y|\bar{\lambda})$  with the highest value.

<sup>13</sup>The theory regarding this type of estimator is developed in a series of papers by Kwan (1999), Kim (2002) and Chernozhukov and Hong (2003).

To see this, suppose that the likelihood is close to symmetric and highly peaked around the mode,  $\theta^*$ , given by

$$\theta^* = \arg \max L_{\mathcal{T}}(Y|\theta) p(\psi|\gamma, \hat{\lambda}) p(\sigma_v|\gamma, \hat{\lambda}) \quad (3.33)$$

Then  $m_{\mathcal{T}}(Y|\gamma, \hat{\theta}, \hat{\lambda})$  can be approximated by<sup>14</sup>

$$m_{\mathcal{T}}(Y|\theta^*, \hat{\lambda}) = (2\pi)^{d\theta} \left| -\nabla_{\theta}^2 \log p(\theta^*|Y, \hat{\lambda}) \right|^{-\frac{1}{2}} \exp \left[ \frac{1}{2} (\theta - \theta^*)' \nabla_{\theta}^2 \log p(\theta^*|Y, \hat{\lambda}) (\theta - \theta^*) \right] \quad (3.34)$$

(see, [Canova, 2005](#), Chapter 9). At this point we apply the same assumption adopted by [del Negro and Schorfheide](#) and we replace  $\theta$  by  $\theta(\gamma)$ , namely

$$m_{\mathcal{T}}(Y|\theta^*, \hat{\lambda}) = (2\pi)^{d\theta} \left| -\nabla_{\theta}^2 \log p(\theta^*|Y, \hat{\lambda}) \right|^{-\frac{1}{2}} \exp \left[ \frac{1}{2} (\theta(\gamma) - \theta^*)' \nabla_{\theta}^2 \log p(\theta^*|Y, \hat{\lambda}) (\theta(\gamma) - \theta^*) \right] \quad (3.35)$$

Consequently, in our approach, we have

$$p(\gamma|Y, \theta^*, \hat{\lambda}) \propto m_{\mathcal{T}}(Y|\theta^*, \hat{\lambda}) p(\gamma) \quad (3.36)$$

and the posterior distribution of  $\gamma$  can be constructed using the random walk Metropolis-Hastings Markov Chain Monte Carlo algorithm.

Intuitively, we can think of (3.36) as the set of  $\gamma$  values that minimize the distance between the posterior mode of the VAR parameter vector estimated in the data and the one implied by the model. The difference between this estimator and the one discussed in [del Negro and Schorfheide \(2004\)](#) is the set of instruments for the estimation of  $\gamma$ . In our case it is the posterior mode of  $\theta$ , while [del Negro and Schorfheide](#) use the OLS estimate.

If the DSGE model estimated using Bayesian maximum likelihood fits poorly, then it may be useful to compare the quasi-Bayesian estimate of  $\gamma$  with the posterior estimate of  $\gamma$  from Bayesian maximum likelihood. This can provide useful information about which particular parameters within the  $\gamma$  vector may be contributing to the poor fit of the DSGE model. We do this as part of our empirical application in Section 5.2.<sup>15</sup>

---

<sup>14</sup>We use  $\nabla_a f(a)$  and  $\nabla_a^2 f(a)$  to represent the matrices of the first and second derivatives of the vector function  $f(a)$  with respect to the vector  $a$ , respectively.

<sup>15</sup>Of course, by definition, limited information methods do not include all information relevant for estimation. For example, [Canova and Sala \(2009\)](#) argue that full information techniques deliver more accurate inference than limited information methods since the likelihood of the model conveys substantially more information than the limited information objective function, which aids the identification of the true parameter vector. However, [Iskrev \(2010\)](#) suggests that many DSGE models may suffer from identification problems, even when full information methods are used. [Theodoridis \(2011\)](#) illustrates that when a sufficient number of instruments, which fully summarize the likelihood of the model under normality, and the optimal weighting matrix are used then, in small samples, limited information estimation techniques outperform even Bayesian full information procedures with excessively tight priors around the true parameter vector.

### 3.4.2 DSGE Evaluation

We first note that the term inside the square brackets of (3.35)

$$\mathcal{W} \equiv (\theta(\gamma) - \theta^*)' \nabla_{\theta}^2 \log p(\theta^* | Y, \hat{\lambda}) (\theta(\gamma) - \theta^*) \quad (3.37)$$

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

$$\begin{aligned} \mathbb{E}_{\gamma}(\theta(\gamma) - \theta^*) &= \int (\theta(\gamma) - \theta^*) p(\gamma | Y, \theta^*, \hat{\lambda}) d\gamma \\ &= \int \theta(\gamma) p(\gamma | Y, \theta^*, \hat{\lambda}) d\gamma - \theta^* = 0_{d\theta \times 1} \end{aligned} \quad (3.38)$$

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

Under quadratic preferences, the following expressions

$$\mathbb{E}_{\gamma} \mathcal{W} = \int \mathcal{W} p(\gamma | Y, \theta^*, \hat{\lambda}) d\gamma \quad (3.39)$$

$$\mathbb{E}_{\gamma} (\mathcal{W} - \mathbb{E}_{\gamma} \mathcal{W})^2 = \int (\mathcal{W} - \mathbb{E}_{\gamma} \mathcal{W})^2 p(\gamma | Y, \theta^*, \hat{\lambda}) d\gamma \quad (3.40)$$

can be interpreted as the expected loss and risk of using  $\mathcal{M}$ , respectively (Canova, 2005; Schorfheide, 2000). Furthermore, the overlap between the posterior distribution of  $\mathcal{W}$  and the posterior distribution of  $\mathcal{W}^*$

$$\mathcal{W}^* \equiv (\theta - \theta^*)' \nabla_{\theta}^2 \log p(\theta^* | Y, \hat{\lambda}) (\theta - \theta^*) \quad (3.41)$$

which is obtained by the posterior estimation of the VAR, provides a more complete measure of the misspecification of the structural model.

## 4 Monte Carlo experiments

In this section, we present Monte Carlo experiments to illustrate two aspects of our approach. In Section 4.2, we conduct a Monte Carlo investigation of the quasi-Bayesian estimator described in Section 3.4.1. In Section 4.3 we demonstrate how our estimation methodology detects priors about DSGE model shock variances that are at odds with the data. Before doing so, we briefly describe the DSGE model in Section 4.1.

### 4.1 The DSGE model

We use a version of the model developed by Smets and Wouters (2007) with two simplifications: we use a smaller number of shocks (specifically we abstract from mark-up shocks); and we use a simpler specification of the reaction function for monetary policy.

As the model is well known, our description is brief.<sup>16</sup> All variables are expressed as log deviations from their steady-state values,  $\mathbb{E}_t$  denotes the rational expectation formed at time  $t$ , the steady-state value of a variable (e.g.,  $y_t$ ) is denoted with an over-bar ( $\bar{y}$ ) and all shocks ( $\omega_t^i$ ) are assumed to be normally distributed with zero mean and unit standard deviation.

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

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g \quad (4.1)$$

where  $y_t$  denotes total output. The consumption Euler equation is given by

$$c_t = \frac{\lambda}{1+\lambda} c_{t-1} + \frac{1}{1+\lambda} \mathbb{E}_t c_{t+1} + \frac{(1-\sigma_C)(\bar{W}^h \bar{L} | \bar{C})}{\sigma_C(1+\lambda)} (\mathbb{E}_t l_{t+1} - l_t) - \frac{1-\lambda}{\sigma_C(1+\lambda)} (r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b) \quad (4.2)$$

where  $l_t$  is the hours worked,  $r_t$  is the nominal interest rate,  $\pi_t$  is the rate of inflation and  $\varepsilon_t^b = \rho_g \varepsilon_{t-1}^b + \sigma_g \omega_t^b$  is a risk premium process. In the absence of habit formation ( $\lambda = 0$ ), equation (4.2) reduces to the standard forward looking consumption Euler equation. The linearised investment equation is given by

$$i_t = \frac{1}{1+\beta} i_{t-1} + \frac{\beta}{1+\beta} \mathbb{E}_t i_{t+1} + \frac{1}{(1+\beta)S''} q_t + \varepsilon_t^i \quad (4.3)$$

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

$$q_t = \beta(1-\delta) \mathbb{E}_t q_{t+1} + (1-\beta(1-\delta)) \mathbb{E}_t r_{t+1}^k - (r_t - \mathbb{E}_t \pi_{t+1} + \varepsilon_t^b) \quad (4.4)$$

where  $r_t^k = -(k_t - l_t) + w_t$  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 model, the aggregate production function is defined as

$$y_t = \phi_p (\alpha k_t^s + (1-\alpha) l_t + \varepsilon_t^a) \quad (4.5)$$

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,  $z_t$  (so  $k_t^s = k_{t-1} + z_t$ ). Cost minimization implies that capital utilization is proportional to the real rental rate of capital,  $z_t = \frac{1-\psi}{\psi} r_t^k$ . Total factor 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

$$k_t = (1-\delta) k_{t-1} + \delta i_t + (1+\beta) \delta S'' \varepsilon_t^i \quad (4.6)$$

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

<sup>16</sup>Readers interested in the microfoundations of the model are recommended to consult the original paper.



following New-Keynesian Phillips curve for inflation ( $\pi$ )

$$\pi_t = \frac{i_p}{1 + \beta i_p} \pi_{t-1} + \frac{\beta}{1 + \beta i_p} \mathbb{E}_t \pi_{t+1} + \frac{1}{(1 + \beta i_p)} \frac{(1 - \beta \xi_p)(1 - \xi_p)}{(\xi_p((\phi_p - 1)\varepsilon_p + 1))} mc_t \quad (4.7)$$

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 New-Keynesian Phillips curve for wages

$$\begin{aligned} w_t = & \frac{1}{1 + \beta} w_{t-1} + \frac{\beta}{1 + \beta} (\mathbb{E}_t w_{t+1} + \mathbb{E}_t \pi_{t+1}) - \frac{1 + \beta i_w}{1 + \beta} \pi_t + \frac{i_w}{1 + \beta} \pi_{t-1} \\ & + \frac{1}{1 + \beta} \frac{(1 - \beta \xi_w)(1 - \xi_w)}{(\xi_w((\phi_w - 1)\varepsilon_w + 1))} \mu_t^w \end{aligned} \quad (4.8)$$

where  $\mu_t^w = \left( \sigma_l l_t + \frac{1}{1 - \lambda} (c_t - \lambda c_{t-1}) \right) - w_t$  is the households' marginal benefit of supplying an extra unit of labour service. Lags of inflation appear in both (4.7) and (4.8) because price and wage setters that do not receive the Calvo signal to reset their price or wage may index it to inflation in the previous period. The extent of indexation for prices and wages respectively are captured by the parameters  $i_p$  and  $i_w$ .

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

$$r_t = \rho r_{t-1} + (1 - \rho) (r_\pi \pi_t + r_y y_t) + \varepsilon_t^r \quad (4.9)$$

where  $\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \sigma_r \omega_t^r$  is the monetary policy shock. Our version of the monetary policy rule specifies that the policymaker responds to detrended output rather than the output gap. This modification to [Smets and Wouters \(2007\)](#) means that we do not have to augment the model with a block of equations that determine the flexible price allocations in the model (and hence the output gap). This reduces the size of the model which reduces the computational burden somewhat.

## 4.2 The quasi-Bayesian estimator

In this section we present a small simulation exercise to investigate the performance of the DSGE Quasi-Bayesian (QB) estimator proposed in Section 3.4.1 against the standard full information Bayesian maximum likelihood approach (BML). The data generating process is the model described in Section 4.1 and it is assumed that only output, consumption, investment, inflation and nominal interest rates ( $y_t$ ,  $c_t$ ,  $i_t$ ,  $\pi_t$  and  $r_t$ , respectively) are observed.

The details of the Monte Carlo steps are described in Appendix C, though we highlight one particular aspect. One critique of our suggested approach could be that the minimization required to obtain  $\theta^*$  and  $\nabla_\theta^2 \log p(\theta^* | Y, \widehat{\lambda})$  will typically be very computationally expensive. For that reason, we replace this step by the estimation of the VAR model, and the use of the posterior mean and the variance-covariance matrix of  $p(\theta | Y, \widehat{\lambda})$  instead of  $\theta^*$  and  $\nabla_\theta^2 \log p(\theta^* | Y, \widehat{\lambda})$ , respectively. It is common practice for applied researchers to approximate these quantities from the output of the posterior simulation to reduce the computational burden ([Koop and Poirier, 1993](#)). We have verified that results using  $\theta^*$  and  $\nabla_\theta^2 \log p(\theta^* | Y, \widehat{\lambda})$  are almost identical, but we present results based on the approximation to demonstrate that a less computationally demanding variant of our approach performs well.

Before turning to the discussion of the results, it should be highlighted that the purpose of this

experiment is not to fully characterize the small-sample properties of the proposed estimator, for which a more substantial exercise would be required. Instead, we aim to assess whether the estimator described in Section 3.4.1 behaves reasonably in small samples and we choose the Bayesian maximum likelihood estimator as the benchmark case to assess this.

Table 1 provides description of the estimated parameters and their prior distributions used in the Monte Carlo exercise. The prior moments (mean and standard deviation) of the structural parameter vector are presented in the first two columns of Table 3. The next three columns display the median (calculated using the Mahalanobis metric<sup>17</sup>), the standard deviation and the bias for the BML estimator based on 1000 replications. The remaining columns capture the same information for the quasi-Bayesian estimator.

The results reveal that the quasi-Bayesian estimator performs better than the BML estimator not only in terms of bias but also in terms of efficiency. It is not surprising that our limited information estimator is less biased than the full information estimator in small samples (Ruge-Murcia, 2007; Theodoridis, 2011). However, it seems striking that the QB estimator also appears to be more efficient than the BML estimator. This result can be rationalized by appealing to the asymptotic theory of minimum distance estimators (Newey and McFadden, 1986). From (3.35) and (3.36) we see that the posterior distribution of the structural parameter vector consists of the values of  $\gamma$  that minimise the distance between  $\theta(\gamma)$  and  $\theta^*$ . In other words,  $\theta^*$  can be interpreted as the set of the instruments used in this estimation and  $-\nabla_{\theta}^2 \log p(\theta^*|Y, \hat{\lambda})^{-1}$  is its covariance matrix. Loosely speaking,  $p(\gamma|Y, \theta^*, \hat{\lambda})$  can be viewed as the distribution of an efficient estimator, a property exhibited by the limited information estimator but not the full information estimator. Clearly this property does not hold asymptotically where the BML is unambiguously more efficient than the QB estimator.

### 4.3 Identifying DSGE model misspecification

We construct a simple Monte Carlo experiment to demonstrate how our estimation methodology may provide information about the source of the DSGE model misspecification. As before, we simulate observations from the DSGE model described in Section 4.1. We then examine the performance of VARs based on *incorrect* DSGE priors.

In each replication, we draw DSGE model parameters  $\gamma$  from the prior distributions detailed in Table 1 and the first two columns of Table 3. We solve the model and generate a sample of 10000 observations, keeping only the final 200 observations as our synthetic sample. We then estimate VARs based on *incorrect* priors for the DSGE model parameters.<sup>18</sup> The incorrect priors are for the five elements of  $\gamma$  that correspond to the standard deviations of the structural shocks:  $\{\sigma_i, \sigma_g, \sigma_a, \sigma_b, \sigma_r\}$ . For these parameters we assume a prior that is double the true values of the parameters drawn in that iteration. The remaining twenty elements of  $\gamma$  are set equal to the values drawn in that iteration.

The incorrect DSGE priors are used to estimate VARs using the DSGE-VAR methodology of del Negro and Schorfheide (2004) and our proposed approach. The hyper-parameters controlling the weight on

<sup>17</sup>The Mahalanobis 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.

<sup>18</sup>This is in contrast to the Monte Carlo exercise in Section 4.2, which is intended to explore the performance of the estimation approach when the correct priors are used.

the DSGE model priors (respectively  $\bar{\lambda}$  and  $\lambda_{DS}$ ) are chosen to maximise the marginal likelihood.<sup>19</sup> Our goal is to examine how the misspecified DSGE priors affect the values of the hyperparameters and the consequent effects on inference.

We find that our methodology successfully identifies that the misspecification of the DSGE model priors is primarily manifest in the prior variance of the VAR residuals,  $\Sigma_\psi$ . This is evident in the fact that the estimation procedure selects the lowest value for the degrees of freedom  $\eta - dy - 1$  available in our grid in *all* replications. So our estimation procedure places very little weight on the DSGE prior for the VAR covariance matrix.

In contrast,  $\lambda$ , the hyperparameter controlling the tightness of the prior of the VAR coefficients  $\psi$  has a non-degenerate distribution. The mean value for  $\lambda$  is around 0.37. This implies that our procedure recognises the usefulness of the DSGE prior information for  $\psi$  and, on average, recommends that the econometrician shrink her estimate of the prior variance for  $\psi$  by around 60%.

When we repeat this exercise using the DSGE-VAR approach of [del Negro and Schorfheide \(2004\)](#), we find that the mean value for  $\lambda_{DS}$  is close to the mean value of  $\lambda$  for our approach. However, this similarity masks some differences in the distribution of the hyperparameters  $\lambda$  and  $\lambda_{DS}$ . This is illustrated in Table 5, which records the number of observations of  $\lambda$  and  $\lambda_{DS}$  for each of the grid points considered. We observe that the modal value for the hyperparameter  $\lambda$  is 0.2 (the lowest value in the grid) whereas the modal value for  $\lambda_{DS}$  is 0.35. However, the distribution of  $\lambda$  across the Monte Carlo replications has a heavy right tail, putting a weight of 1 or more on the DSGE model priors in 5% of the replications. In contrast, the  $\lambda_{DS}$  puts a weight of 1 or more on the DSGE priors in just 2 of the 1000 replications. This observation suggests that, by consistently placing a low weight on the implications of the misspecified DSGE priors for the VAR covariance matrix (by selecting a low  $\eta$  in all cases), our approach is free to place more weight on the VAR coefficient parameters  $\psi$  when it is appropriate to do so.

Figure 9 plots the estimated density of marginal likelihood values across Monte Carlo replications obtained from our approach (FHT) and the DSGE-VAR (DNS). We observe that the DSGE-VAR tends to fit less well (based on this measure of fit) than our proposed estimator. There are two possible explanations for this result.

First, it may be the case that our approach fits better precisely because it puts less weight on the DSGE priors when those priors are misspecified. That is, the improvement in fit may be driven by those instances in which  $\lambda = 0.2$ , so that little weight is being placed on the DSGE priors. Figure 10 illustrates that this is not in fact the case. The Figure shows the joint distribution of  $\{\lambda, \lambda_{DS}\}$  across the 1000 experiments. Each  $\{\lambda, \lambda_{DS}\}$  pair is represented by a circle, the area of which is proportional to the number of observations in the set of experiments. The color of the circle indicates the mean difference in marginal likelihood for those observations: a higher number indicates that our approach fits better than the DSGE-VAR. We observe that our approach produces better fit in almost all cases and that the extent of the improvement in average fit is not driven entirely by cases in which our approach heavily down-weights the DSGE priors (that is, for low values of  $\lambda$ ).

The second reason that our approach might be expected to dominate is that we have an additional

---

<sup>19</sup>We follow [del Negro and Schorfheide \(2004\)](#) and evaluate the marginal likelihood for a grid of hyperparameters, selecting the element of the grid with the highest marginal likelihood. The grid for  $\lambda_{DS}$  is the same used by [del Negro and Schorfheide \(2004\)](#):  $\{0.2, 0.35, 0.5, 0.7, 1, 1.25, 1.5, 2, 2.5, 5\}$ . We use the same grid for  $\lambda$ . The grid for  $\eta - dy - 1$  is  $\{0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$ .

hyperparameter ( $\eta$ ). Indeed, the purpose of the experiment in this section is to illustrate that the additional flexibility afforded by the additional hyperparameter could be particularly valuable when there is relatively localised misspecification of the DSGE model priors.

## 5 Empirical application

In this section we present an empirical application of our method and compare it to other approaches. Before doing so, we describe the data that we use and the priors for the DSGE model.

We use the data set for the United States constructed by [Smets and Wouters \(2007\)](#).<sup>20</sup> The model is estimated using data between 1966Q1 and 2004Q4. We fit the model to five time series: output, consumption, investment, the Fed funds rate and inflation.<sup>21</sup> The estimation sample is 1966Q1–1999Q4, with the period 2000Q1–2004Q4 used to evaluate out of sample forecasting performance.

We use the simplified version of the [Smets and Wouters \(2007\)](#) model, described in Section 4.1. The five shocks that drive the DSGE model are: a shock to consumer preferences, a shock to investment adjustment costs, a productivity shock, a monetary policy shock and a government spending shock. The shape of the prior distribution is given in Table 1, while the prior moments can be found in Table 4.

### 5.1 Selection of hyperparameters

In our approach, the posterior BVAR parameter estimates are weighted averages between the DSGE-based prior moments and the OLS estimates, with the weights controlled by the hyperparameter vector  $\bar{\lambda} = (\lambda, 1/\eta)'$ . The hyperparameter vector needs to be decided before the estimation and we use the marginal likelihood to select the value that maximizes the fit of the BVAR.<sup>22</sup>

The relationship between the hyperparameter vector  $\bar{\lambda}$  and the marginal likelihood of the VAR is shown in Figure 1. We observe an incompatibility between the DSGE-based priors and the data, since the marginal likelihood is maximized for a large value of  $\lambda$  and a small value of  $\eta$ . However, we are able to identify which part of the VAR parameter vector implied by the DSGE-based priors disagrees most with the data. The significant steepness of the surface with respect to small values of  $\eta$  indicates that small changes to the degrees of freedom parameter lead to substantial changes in the fit of the model, while the surface is relatively flat with respect to  $\lambda$  and large values of  $\eta$ . This suggests that the prior mean of the residual variance-covariance matrix is at odds with the data.

### 5.2 Assessing misspecification using quasi-Bayesian estimation

In this section, we use the quasi-Bayesian ('QB') estimator described in Section 3.4.1, to uncover posterior estimates for the DSGE parameter vector,  $\gamma$ , based on our BVAR posterior parameter

<sup>20</sup>This data set is publicly available from the website of the [American Economic Association](#).

<sup>21</sup>We use the Hodrick-Prescott filter to eliminate the zero frequency component of the non-stationary series (real output, real consumption and real investment).

<sup>22</sup>This process is computationally demanding since it requires the discretization of the interval where  $\bar{\lambda}$  is defined to specify a grid of  $\{\eta, \lambda\}$  pairs and posterior estimation of the BVAR model for each grid point.

estimates. We compare these estimates with those generated by full information Bayesian maximum likelihood (‘BML’) to investigate potential areas of misspecification in the DSGE model.

To provide an ‘eyeball’ assessment of fit, Figure 7 plots the Kalman filter estimates of the observed series of the two estimators against the data (black solid line), which can be regarded as the one-step-ahead within sample forecasts.<sup>23</sup> The in-sample fit of the structural model using the full information BML estimator (dashed red line) seems very good, though the QB estimator (dotted-dashed blue line) seems to fit marginally better.

Under the assumption that the BVAR model estimated using our approach is a good description of the data generating process, then the metric discussed in Section 3.4.2 can be used to assess plausibility the restrictions imposed by the BML and QB models on the data. Table 2 displays the  $\mathbb{E}_\gamma \mathcal{W}$  and  $\sqrt{\mathbb{E}_\gamma (\mathcal{W} - \mathbb{E}_\gamma \mathcal{W})^2}$  moments indicating that the expected loss and risk of using the BML instead of the QB model is substantial. Since the latter model is estimated by selecting the structural parameter vector to minimize the distance between the VAR parameter vector implied by the DSGE model and the VAR parameter estimated from the data, this is unsurprising.

The posterior distributions of BML- $\mathcal{W}$  (red dashed line) and QB- $\mathcal{W}$  (blue dashed-dotted line) are plotted against the posterior distribution of  $\mathcal{W}^*$  in Figure 4. While 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 QB model, though even in this case the overlap between the posterior distribution of QB- $\mathcal{W}$  and  $\mathcal{W}^*$  is almost zero. Our interpretation of this result is that model’s restrictions on the properties of the data are quite severe, even when  $\gamma$  has been selected to give the best chance to the DSGE model of reproducing the estimated VAR dynamics. Again this may be unsurprising since the VAR model has ninety well-estimated parameters and the DSGE only twenty five.

We can investigate which elements of the DSGE model structure may be placing the harshest restrictions on the behavior of the data, by examining the estimation results in more detail. The posterior distribution of the DSGE model parameters generated by the BML and QB estimators are plotted with the prior distributions in Figure 5. Table 4 provides summary statistics. The posterior variance for most of the QB estimates (eighteen out of twenty five) is smaller than the posterior variance of the BML estimates. This result is consistent with the Monte Carlo evidence presented in Section 4.2.

In terms of an economic interpretation, marginal cost variations have a smaller impact on current inflation in the QB case (0.079), compared to the BML estimate (0.243). Additionally, the labor supply curve responds less to wage and consumption movements in the QB case. The effect of the expected growth of hours worked on consumption in the Euler equation (4.2) is much smaller in the QB variant (0.079 versus 0.188 in the BML case). The response of monetary policy to inflation is larger in the QB case.

In general, the posterior estimates of the standard deviations of the structural shocks are smaller in the QB case (an exception being the monetary policy shock). The overlap of the posterior distributions of the persistence parameters of the risk premium, government spending and interest rate shocks is quite small. The fact that the BML and QB methods generate somewhat different estimates of the shock process parameters is consistent with the finding that the prior mean of the residual variance-

---

<sup>23</sup>The prediction errors are computed for models solved using parameter values equal to the estimated modes of the posterior densities in each case.

covariance matrix of the VAR is at odds with the data, as demonstrated in Section 5.1.<sup>24</sup>

The parameter differences we have documented generate different economic behavior, as can be seen from the impulse response function (IRF) comparison in Figure 6. For instance, in the QB case, a smaller and less persistent government spending shock crowds out less consumption and investment than in the BML case. The difference in consumption responses leads to different profiles for real wages (in the BML case, agents supply more labor pushing wages down) and, consequently, different profiles for marginal cost, inflation and interest rates.

Another example is the investment shock: a similar perturbation leads to a smaller rise in investment and, consequently, to a lower capital stock for the QB case. In the BML variant, the fall in the rental rate of capital exceeds the increases in the wages (due to higher labor demand) keeping the marginal cost below base from the fourth period after the shock. However, this does not happen in the QB variant, where marginal cost stays above base for the entire duration of the shock, with obvious implications for inflation and the interest rate.

The alternative estimation approaches deliver somewhat different impulse responses to a monetary policy shock, which is often the focus of interest in monetary DSGE models (see, Christiano et al., 2005). The higher estimated price stickiness in the QB variant gives rise to more persistent responses, despite the fact that the degree of interest rate smoothing in the Taylor rule is estimated to be smaller than the BML variant.

Figure 8 computes the contribution of each shock to the variance of the  $h$ -period-ahead forecast error of the observable vector ( $h = \infty$  captures the long-run effect). Although both decompositions look very similar, we note that the QB 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. This is consistent with the flatter QB Philips curve – inflation is mainly driven by productivity shocks – and the higher inflation weight in the QB Taylor rule. The contribution of the government spending shock to fluctuations in consumption and investment is small in the BML case, but its contribution is almost zero in the QB variant, in which consumption and investment are driven primarily by the risk premium and investment shock, respectively. Taken together with the comparison of impulse responses, we conclude that the alternative estimation approaches give rise to parameter estimates that have economically significant implications in some cases.

## 6 Conclusion

We have developed a new method for estimating BVAR models using priors from DSGE model. Our method uses the DSGE model priors to determine the moments of an independent Normal-Wishart prior for the VAR parameters. Two hyper-parameters control the tightness of the DSGE-implied priors on the autoregressive coefficients and the residual covariance matrix respectively. Compared with existing approaches, our method makes more use of the information about the second moments of the VAR parameters contained in the DSGE parameter priors. The hyper-parameters provide a device for the researcher to detect subsets of DSGE parameter priors that are particularly at odds

---

<sup>24</sup>Of course  $\Sigma_v$  is a function of all structural parameters, though the parameters governing the variability and persistence of the shock processes have a particularly important influence. From equation (3.2) we observe that the parameters controlling the variance of the structural shocks will enter the matrix  $\Upsilon$ . From equation (3.9) we see that  $\Sigma_v = [A\Upsilon][A\Upsilon]'$ .

with the data.

We have also shown how our method can be used to uncover the posterior density of the DSGE parameter vector from the posterior estimates of the VAR parameters. This new quasi-Bayesian estimator can also be used to assess instances in which the DSGE prior is at odds with the data. In particular, a comparison of the posterior estimates from our quasi-Bayesian approach with those from Bayesian maximum likelihood may provide useful information about specific parameter priors that are at odds with the data. We have illustrated the strengths of our approach using both Monte Carlo examples and an empirical exercise using US data.



## References

- BLANCHARD, O. J. AND C. M. KAHN (1980): “The Solution of Linear Difference Models under Rational Expectations,” *Econometrica*, 48, 1,305–11.
- CANOVA, F. (2005): *Methods for applied macroeconomic research*, Princeton: Princeton University Press.
- CANOVA, F. AND L. SALA (2009): “Back to square one: identification issues in DSGE models,” *Journal of Monetary Economics*, 56, 431–49.
- CHERNOZHUKOV, V. AND H. HONG (2003): “An MCMC approach to classical estimation,” *Journal of Econometrics*, 115, 293–346.
- CHIB, S. (1995): “Marginal Likelihood from the Gibbs Output,” *Journal of American Statistical Association*, 90, 1313–1321.
- CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): “Nominal Rigidities and the Dynamic Effects of a shock to Monetary Policy,” *Journal of Political Economy*, 113, 1–45.
- CHRISTIANO, L. J., M. EICHENBAUM, AND R. VIGFUSSON (2006): “Assessing Structural VARs,” NBER Working Paper 12353, National Bureau of Economic Research, Inc.
- CHRISTIANO, L. J., M. TRABANDT, AND K. WALENTIN (2010): “DSGE Models for Monetary Policy Analysis,” in *Handbook of Monetary Economics*, ed. by B. M. Friedman and M. Woodford, vol. 3, 285–367.
- 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.
- FERNANDEZ-VILLAVARDE, J., J. RUBIO-RAMIREZ, T. SARGENT, AND M. WATSON (2007): “ABCs (and Ds) of Understanding VARs,” *American Economic Review*, 97, 1,021–26.
- GEWEKE, J. (1999): “Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication,” *Econometric Reviews*, 18, 1–73.
- 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.
- 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.
- 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.
- NEWBY, W. K. AND D. MCFADDEN (1986): “Large sample estimation and hypothesis testing,” in *Handbook of Econometrics*, ed. by R. F. Engle and D. McFadden, vol. 4, 2,111–245.
- POIRIER, D. J. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach*, The MIT Press.
- RAVENNA, F. (2007): “Vector Autoregressions and Reduced Form Representations of DSGE Models,” *Journal of Monetary Economics*, 54, 2,048–64.
- 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. (2011): “An efficient minimum distance estimator for DSGE models,” Bank of England working papers 439, Bank of England.
- WHITE, H. (2001): *Asymptotic Theory for Econometricians*, San Diego: Academic Press.

## A Proofs

**Proof:** of Proposition 3.1

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

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

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

by setting

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

and

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

(A.2) reduces to

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

From (A.5), it can be seen that conditioning on  $\Sigma_v$ ,

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

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

$$\begin{aligned} p\left(\theta|Y, \hat{\theta}, \gamma, \bar{\lambda}\right) &\propto |\Sigma_{\lambda\psi}(\gamma)|^{-0.5d\psi} \exp\left\{-0.5\left[\Sigma_{\lambda\psi}(\gamma)^{-0.5}(\psi - \mu_\psi(\gamma))\right]'\left[\Sigma_{\lambda\psi}(\gamma)^{-0.5}(\psi - \mu_\psi(\gamma))\right]\right\} \\ &\quad \times |\Pi(\gamma)|^{0.5\eta} |\Sigma_v|^{-0.5(\eta+dy+1)} \exp\left\{-0.5\text{tr}\left(\Sigma_v^{-1}\Pi(\gamma)\right)\right\} \\ &\quad \times |\Sigma_v|^{-0.5T} \exp\left\{-0.5\left[\Sigma_v^{-1}\begin{bmatrix} (Y - X\hat{\Psi})'(Y - X\hat{\Psi}) \\ + (\Psi - \hat{\Psi})'X'X(\Psi - \hat{\Psi}) \end{bmatrix}\right]\right\} \\ &\propto |\Sigma_{\lambda\psi}(\gamma)|^{-0.5d\psi} \exp\left\{-0.5\left[\Sigma_{\lambda\psi}(\gamma)^{-0.5}(\psi - \mu_\psi(\gamma))\right]'\left[\Sigma_{\lambda\psi}(\gamma)^{-0.5}(\psi - \mu_\psi(\gamma))\right]\right\} \\ &\quad \times |\Pi(\gamma)|^{0.5\eta} |\Sigma_v|^{-0.5(T+\eta+dy+1)} \\ &\quad \exp\left\{-0.5\text{tr}\left(\Sigma_v^{-1}\begin{bmatrix} \Pi(\gamma) + T\hat{\Sigma}_v + \\ (\Psi - \hat{\Psi})'X'X(\Psi - \hat{\Psi}) \end{bmatrix}\right)\right\} \end{aligned} \quad (\text{A.7})$$

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

## B $\theta$ Draws

This Appendix explains how the draws of  $\theta$  are produced

1. Draw  $\gamma_j$  from  $p(\gamma)$  and solve the DSGE model
2. If  $\gamma_j$  satisfies Blanchard and Kahn's conditions (Blanchard and Kahn, 1980) calculate the  $M$  matrix (3.5) else go to step 1
3. If the eigenvalues of  $M$  are less than one in absolute terms calculate  $\psi_j$  and  $\sigma_j$  using equations (3.7)–(3.9) else go to step 1
4. Repeat steps 1 to 3  $S$  times

## C Implementation of Monte Carlo assessment of QB estimator

This Appendix describes the steps required to produce Table 3.

1. The draws described in Appendix B are used to construct the prior moments of  $\theta$  (3.5)–(3.9)
2. A sample of 200 observations is generated by the model described in Section 4.1 based on the prior moments of  $\gamma$  given by Table 3 for  $y_t$ ,  $c_t$ ,  $i_t$ ,  $\pi_t$  and  $r_t$
3. A BVAR with 3 lags is fitted to this data.
4. The posterior median and variance-covariance matrix of  $\theta$  are calculated. The former is obtained using the Mahanolobis metric, namely,

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

where  $\mu_{\bar{\theta}}$  and  $\Sigma_{\bar{\theta}}$  are the posterior mean and variance-covariance matrix of  $p(\theta|Y, \hat{\theta}, p(\gamma))$

5. Given the posterior moments of  $\theta$ , the posterior mode of  $\gamma$  is obtained by minimizing  $\tilde{m}(y_t|\theta^*, \lambda) p(\gamma)$ , 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  $\lambda = (\lambda_\delta, \eta)'$

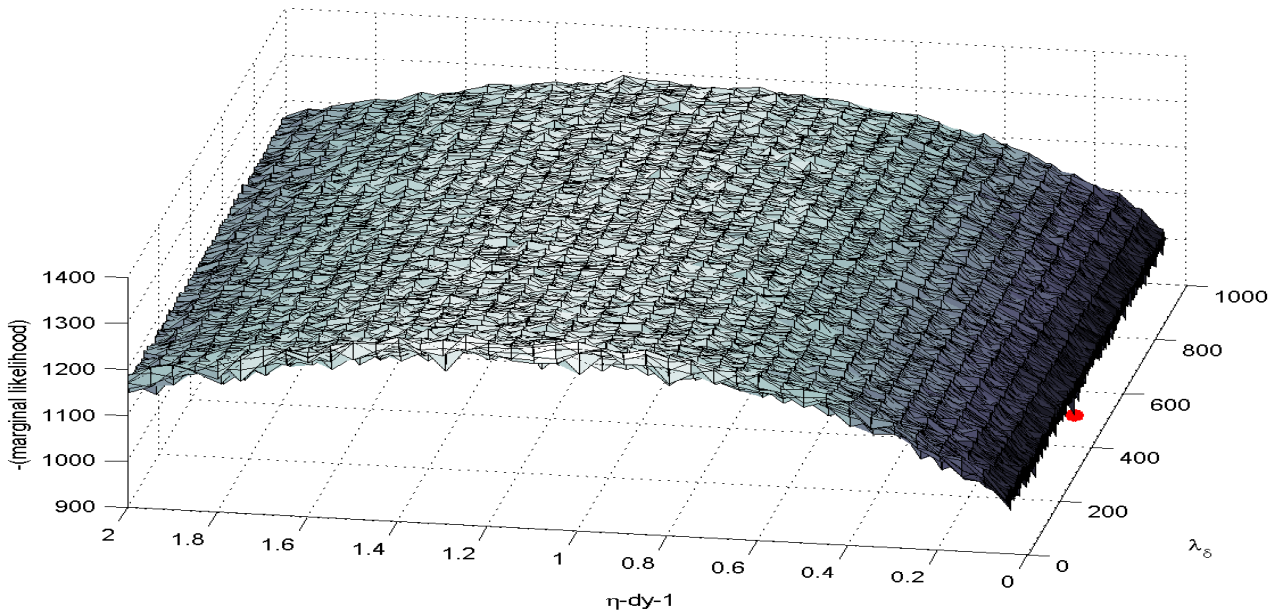


Figure 2: DS-VAR  $\lambda$

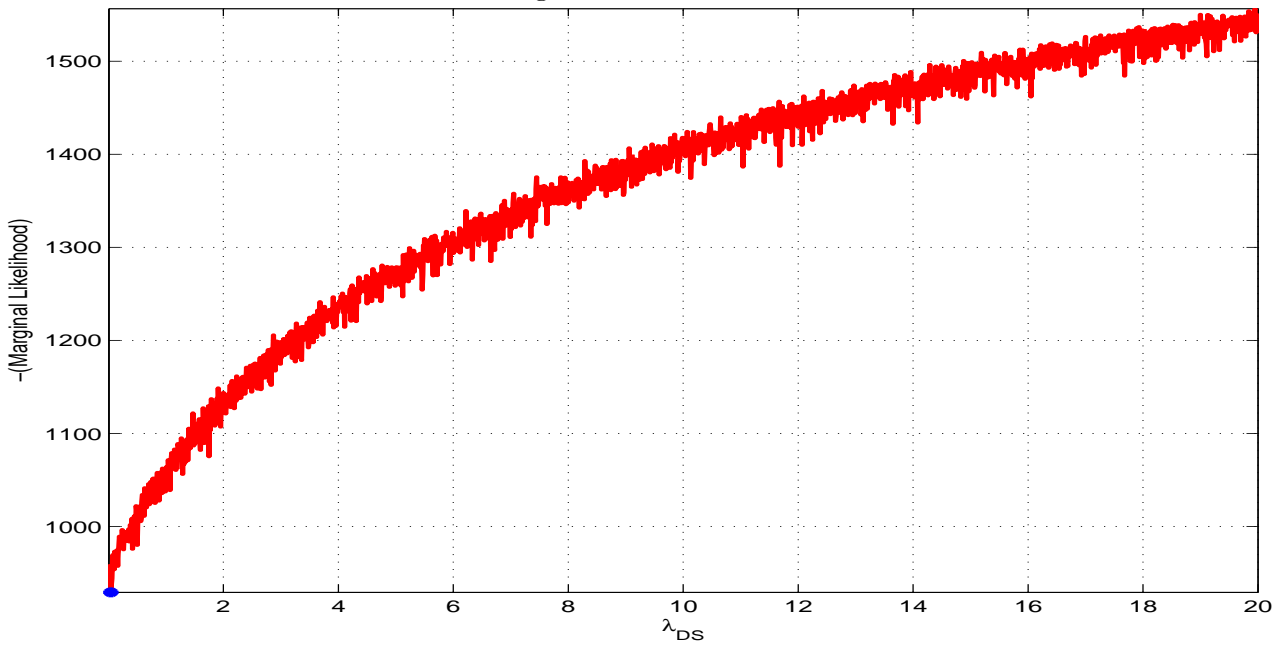


Figure 3: M-VAR  $\lambda$

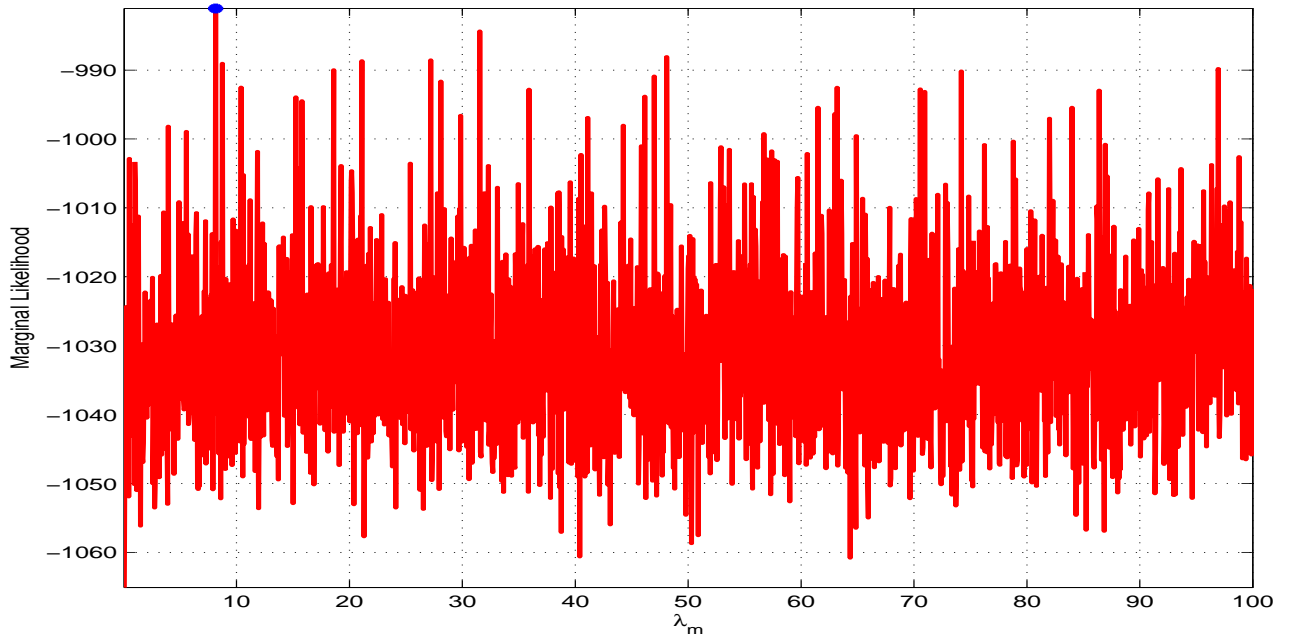


Figure 4: DSGE Evaluation

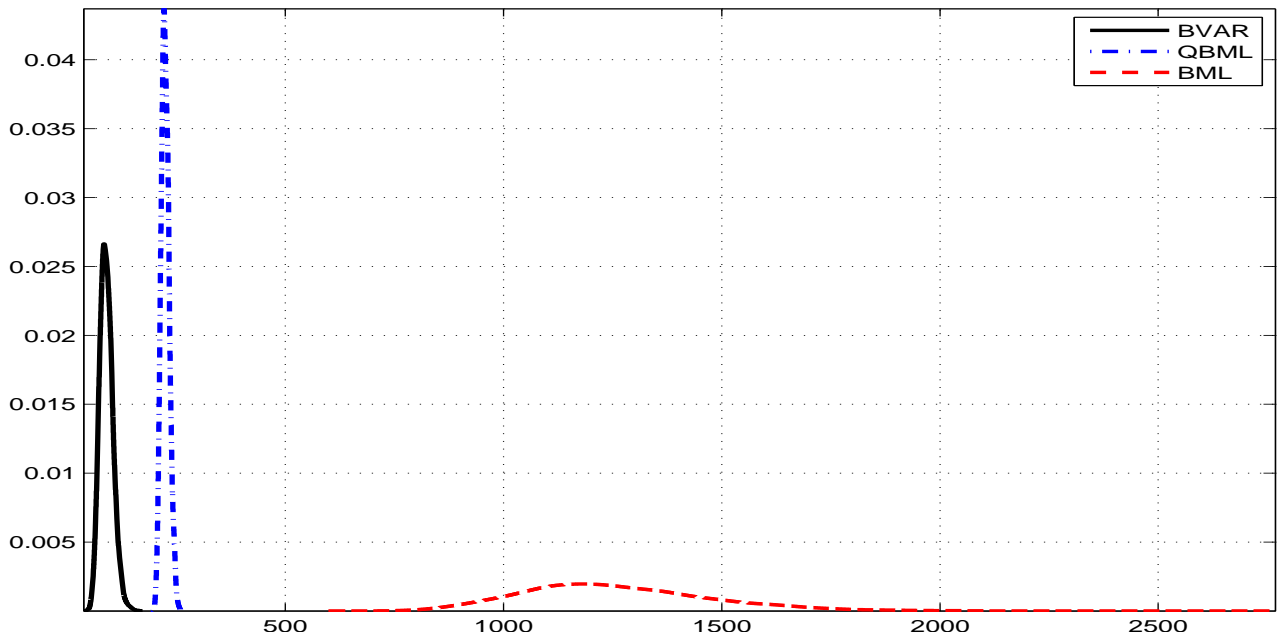




Figure 5: Prior versus Posterior DSGE Parameter Distribution

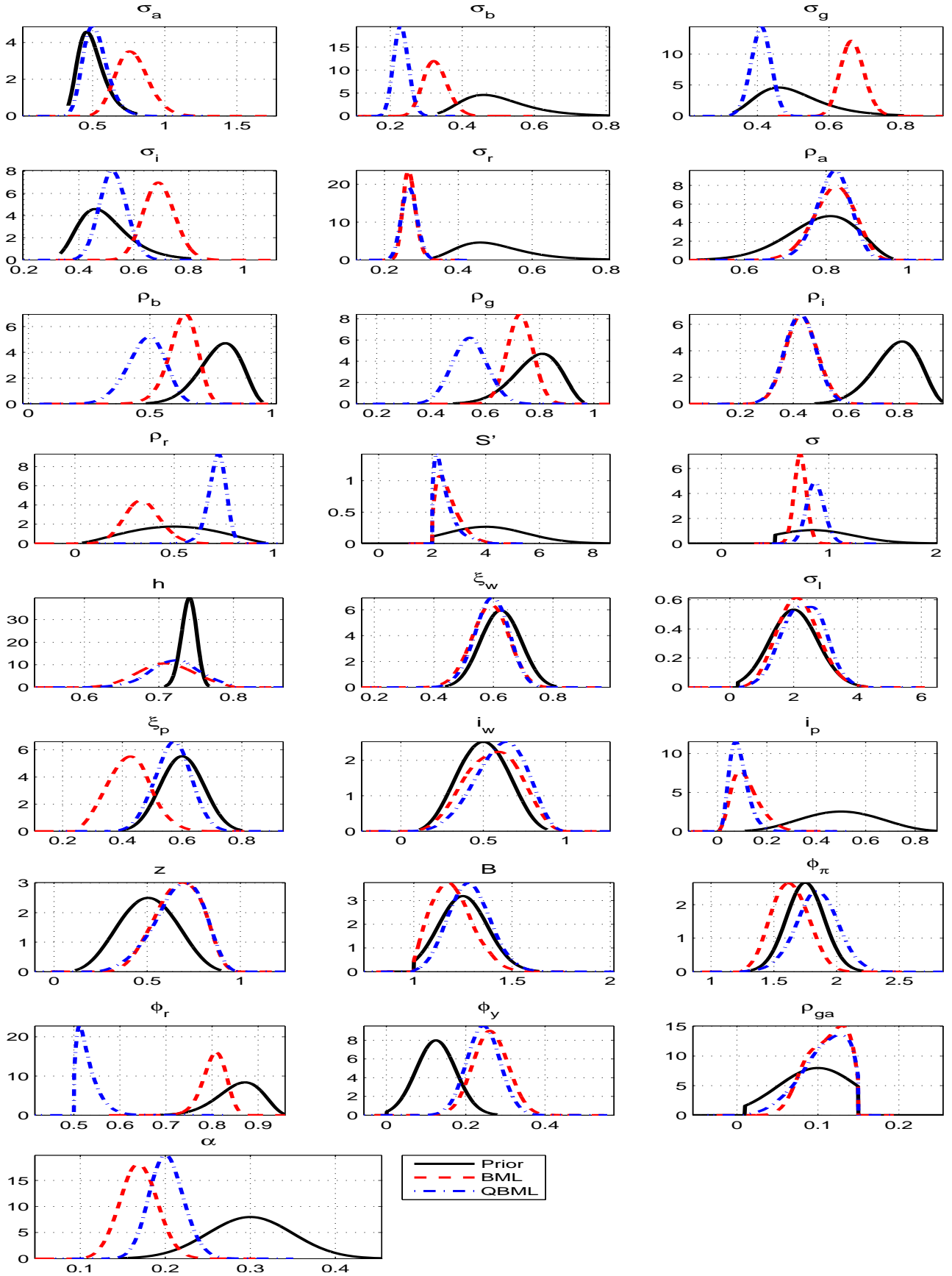


Figure 6: DSGE Posterior IRFs

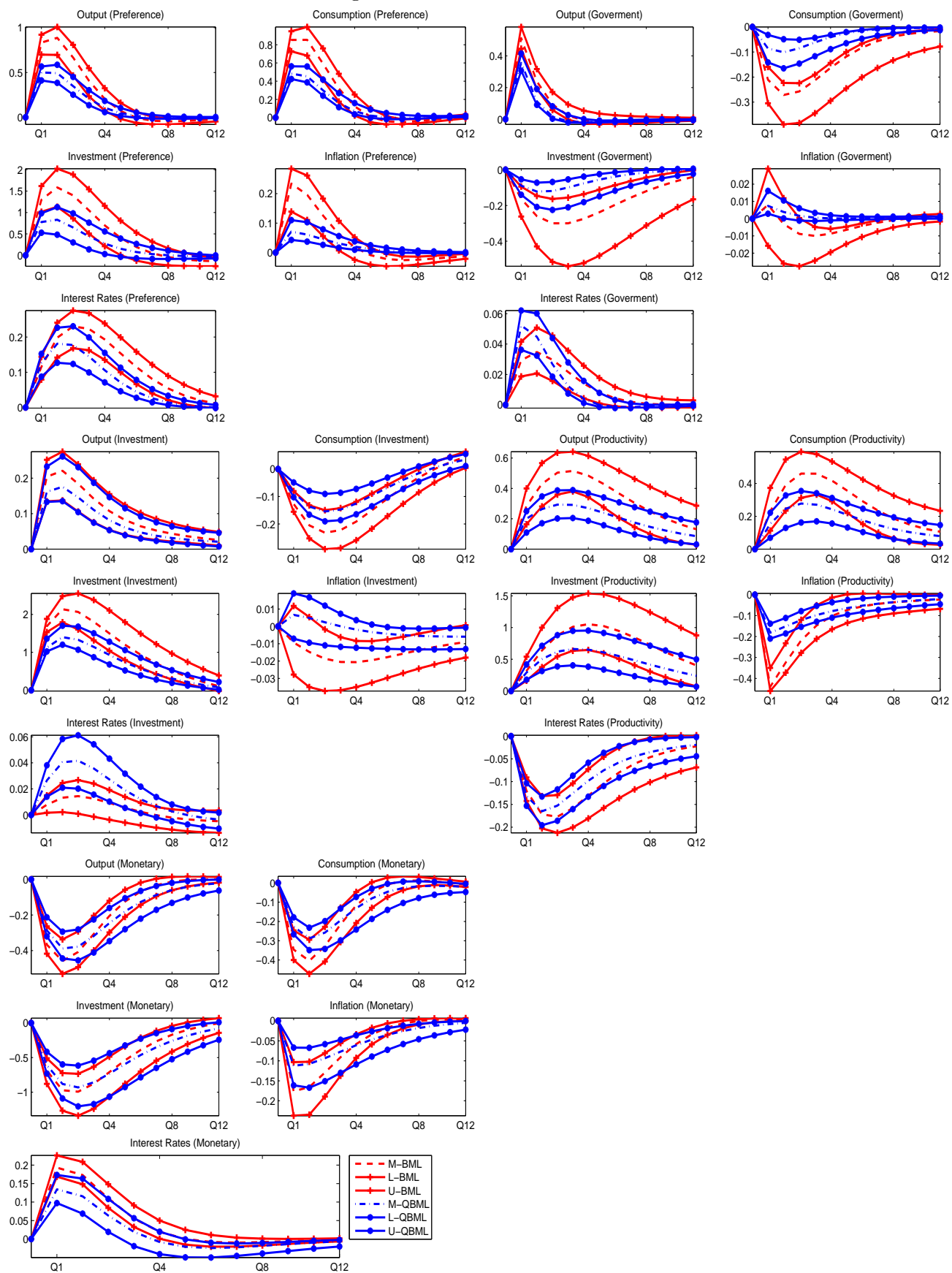


Figure 7: BML One-Step-Ahead Forecasts & FVD

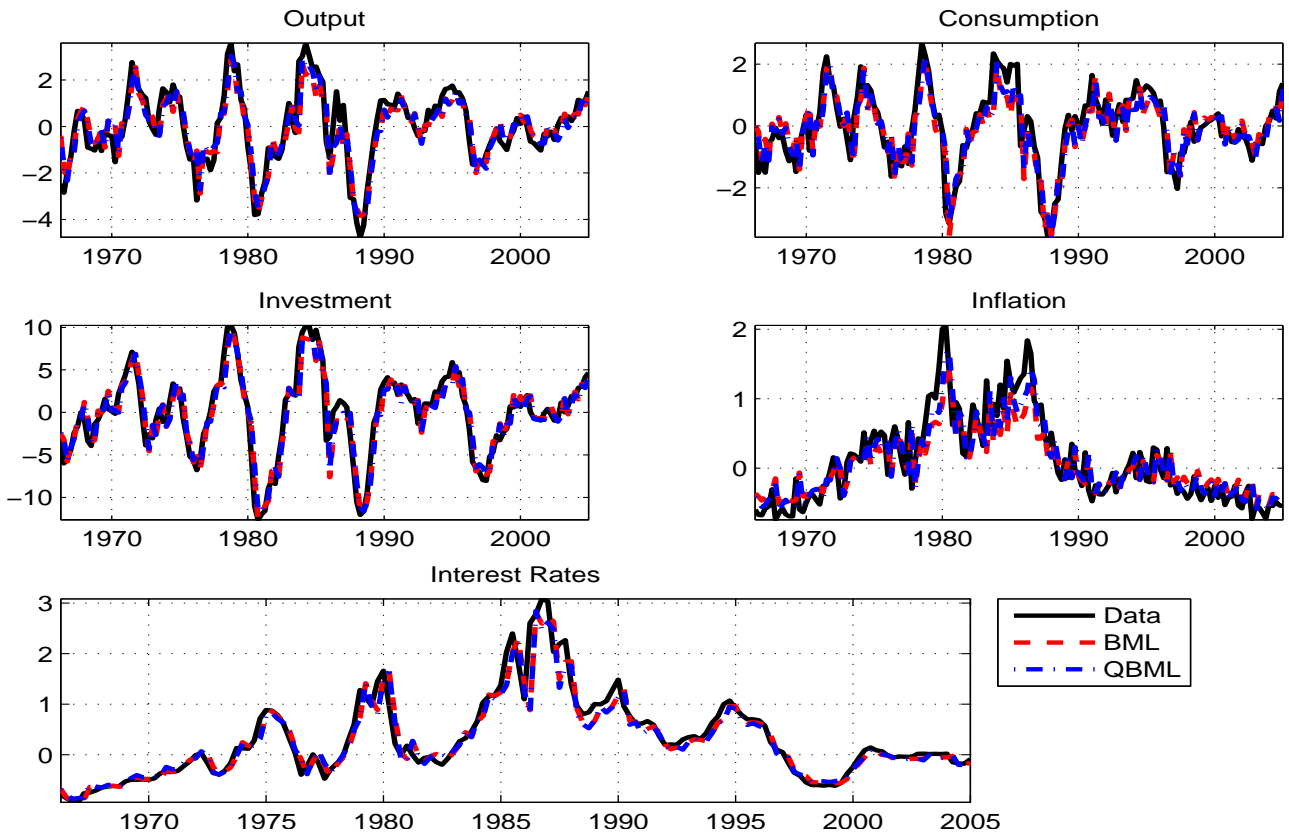
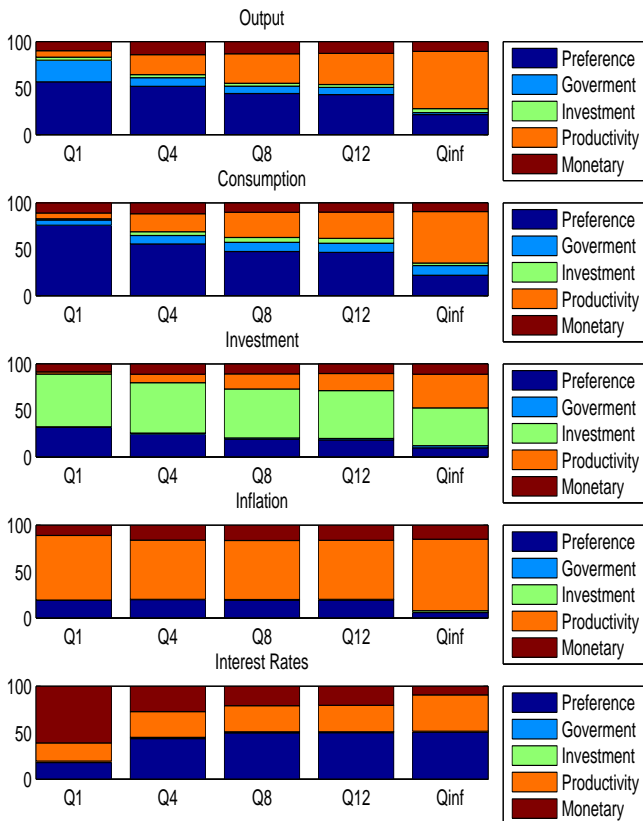


Figure 8: BML FVD



QB FVD

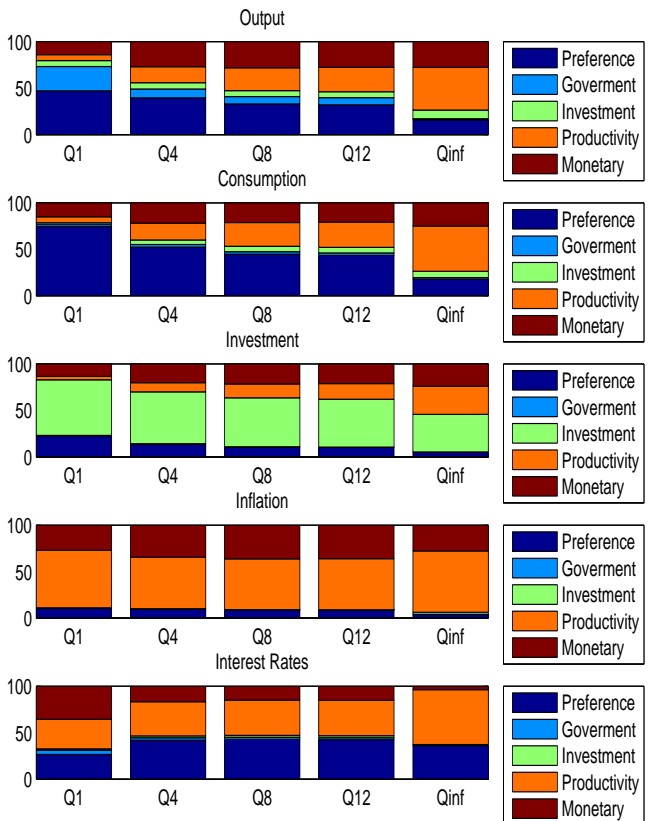


Figure 9: Estimated density of marginal likelihood values across Monte Carlo replications

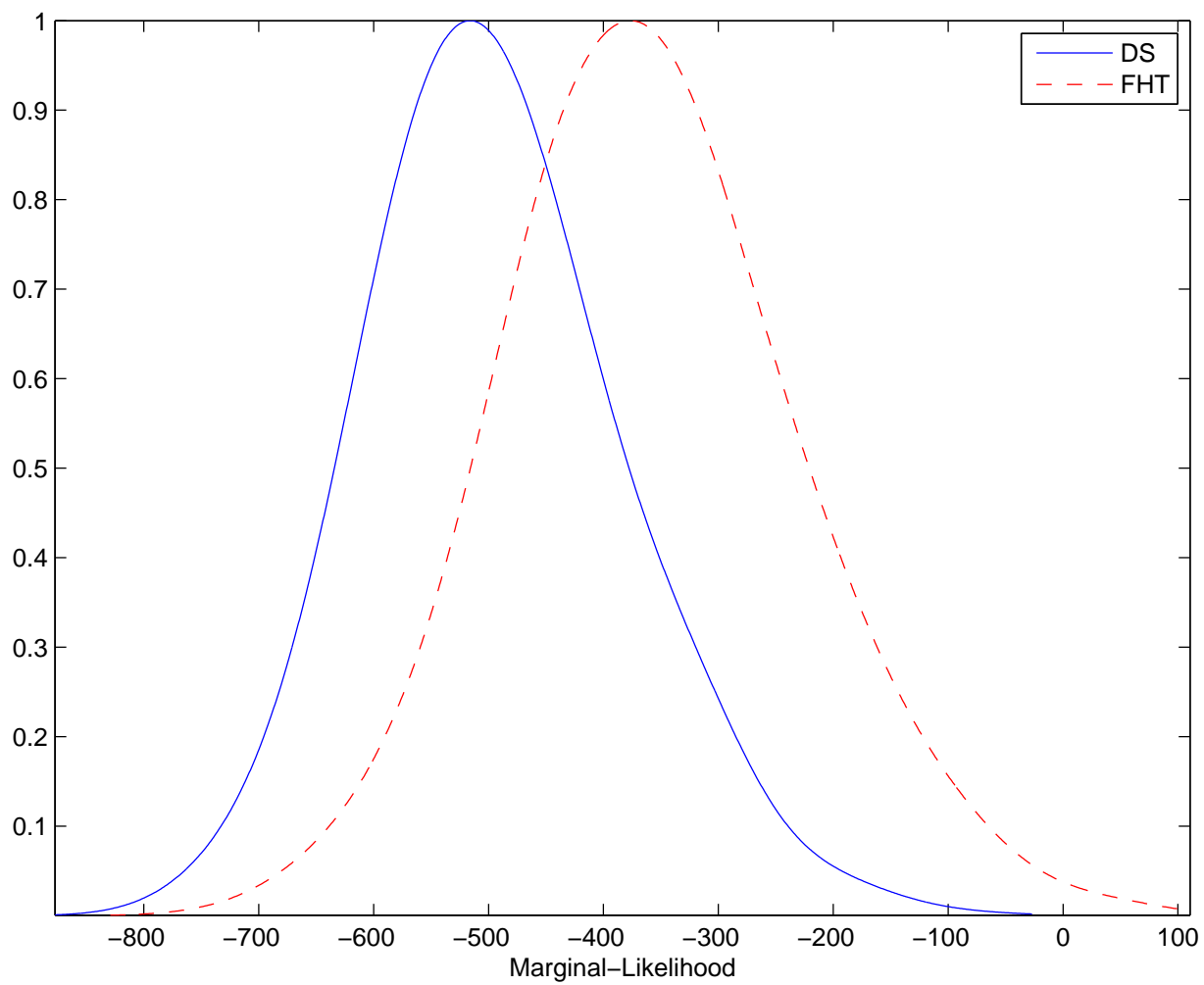
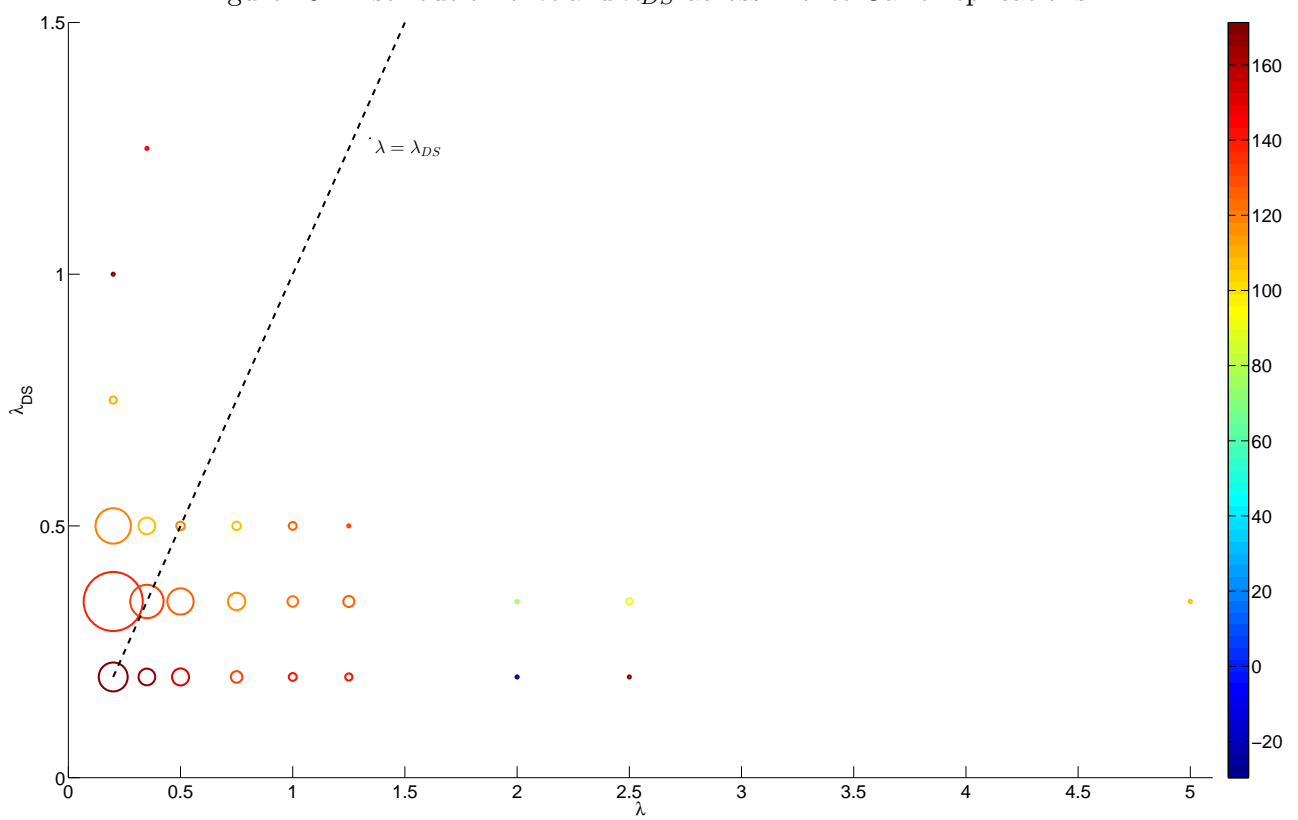


Figure 10: Distribution of  $\lambda$  and  $\lambda_{DS}$  across Monte Carlo replications



## E Tables

Table 1: DSGE Parameter Description & Prior Distribution

<i>Symbols</i>	<i>Description</i>	<i>Prior Distribution</i>
$B$	Fixed Cost	Normal
$S''$	Steady State Capital Adjustment Cost Elasticity	Normal
$\alpha$	Capital Production Share	Normal
$\sigma$	Intertemporal Substitution	Normal
$h$	Habit Persistence	Beta
$\xi_w$	Wages Calvo Parameter	Beta
$\sigma_l$	Labour Supply Elasticity	Normal
$\xi_p$	Prices Calvo Parameter	Beta
$i_w$	Wage Indexation	Beta
$i_p$	Price Indexation	Beta
$z$	Capital Utilisation Adjustment Cost	Beta
$\phi_\pi$	Taylor Inflation Parameter	Normal
$\phi_r$	Taylor Inertia Parameter	Beta
$\phi_y$	Taylor Output Gap Parameter	Normal
$\rho_i$	Investment Shock Persistence	Beta
$\rho_g$	Government Spending Shock Persistence	Beta
$\rho_a$	Productivity Shock Persistence	Beta
$\rho_b$	Premium Shock Shock Persistence	Beta
$\rho_r$	Monetary Policy Shock Shock Persistence	Beta
$\sigma_i$	Investment Shock Uncertainty	Inv. Gamma
$\sigma_g$	Government Spending Shock Uncertainty	Inv. Gamma
$\sigma_a$	Productivity Shock Uncertainty	Inv. Gamma
$\sigma_b$	Risk Premium Shock Uncertainty	Inv. Gamma
$\sigma_r$	Policy Shock Uncertainty	Inv. Gamma

Table 2: DSGE Evaluation

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

Table 3: Full versus Limited Information Bayesian Monte Carlo Evaluation

<i>Parameters</i>	Prior	Moments	<i>BML</i>			<i>QB</i>		
	<i>Mean</i>	<i>STD</i>	<i>Median</i>	<i>Bias</i>	<i>STD</i>	<i>Median</i>	<i>Bias</i>	<i>STD</i>
$\sigma_a$	0.462	0.100	0.461	0.002	0.069	0.451	0.023	0.013
$\sigma_b$	0.182	0.100	0.137	0.245	0.106	0.166	0.089	0.087
$\sigma_g$	0.609	0.100	0.560	0.080	0.095	0.613	0.006	0.018
$\sigma_r$	0.460	0.100	0.465	0.012	0.091	0.434	0.056	0.070
$\sigma_q$	0.240	0.100	0.180	0.248	0.097	0.268	0.119	0.070
$\rho_a$	0.950	0.100	0.948	0.002	0.131	0.985	0.036	0.018
$\rho_b$	0.180	0.100	0.113	0.371	0.085	0.169	0.061	0.046
$\rho_g$	0.970	0.100	0.861	0.112	0.189	0.983	0.013	0.009
$\rho_q$	0.710	0.100	0.813	0.145	0.097	0.698	0.017	0.059
$\rho_m$	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
$\sigma$	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
$\xi_w$	0.700	0.100	0.763	0.091	0.080	0.717	0.025	0.017
$\sigma_l$	1.830	0.750	1.841	0.006	0.568	1.818	0.007	0.057
$\xi_p$	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
$\phi_\pi$	2.040	0.150	2.059	0.009	0.086	2.071	0.015	0.033
$\phi_r$	0.810	0.100	0.749	0.075	0.091	0.814	0.005	0.052
$\phi_y$	0.300	0.050	0.291	0.030	0.038	0.292	0.026	0.014
$\rho_{ga}$	0.050	0.050	0.051	0.024	0.013	0.049	0.022	0.001
$\alpha$	0.190	0.050	0.176	0.072	0.045	0.193	0.014	0.018
<i>Total Bias</i>				2.275			1.762	
<i>Average Bias</i>				0.091			0.070	
<i>Total STD</i>					3.890			1.408



Table 4: Full versus Limited Information Bayesian Estimation

<i>Parameters</i>	<i>Prior</i>	<i>Moments</i>	<i>BML</i>			<i>QB</i>		
	<i>Mean</i>	<i>STD</i>	<i>Median</i>	<i>LwB</i>	<i>UpB</i>	<i>Median</i>	<i>LwB</i>	<i>UpB</i>
$\sigma_a$	0.500	0.100	0.832	0.576	1.017	0.533	0.384	0.710
$\sigma_b$	0.500	0.100	0.331	0.266	0.395	0.237	0.193	0.271
$\sigma_g$	0.500	0.100	0.676	0.605	0.732	0.413	0.355	0.462
$\sigma_r$	0.500	0.100	0.679	0.586	0.808	0.539	0.432	0.623
$\sigma_q$	0.500	0.100	0.273	0.238	0.305	0.265	0.229	0.309
$\rho_a$	0.750	0.100	0.797	0.713	0.905	0.811	0.729	0.892
$\rho_b$	0.750	0.100	0.627	0.504	0.731	0.514	0.325	0.615
$\rho_g$	0.750	0.100	0.718	0.636	0.818	0.517	0.42	0.666
$\rho_q$	0.750	0.100	0.433	0.318	0.547	0.402	0.314	0.539
$\rho_m$	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
$\sigma$	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
$\xi_w$	0.500	0.100	0.548	0.465	0.703	0.565	0.479	0.697
$\sigma_l$	2.000	0.750	1.870	0.988	3.427	2.425	1.096	3.486
$\xi_p$	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
$\phi_\pi$	1.750	0.150	1.558	1.367	1.905	1.801	1.510	2.159
$\phi_r$	0.750	0.100	0.790	0.748	0.847	0.517	0.501	0.589
$\phi_y$	0.125	0.050	0.247	0.178	0.347	0.254	0.165	0.324
$\rho_{ga}$	0.100	0.050	0.105	0.064	0.147	0.097	0.044	0.148
$\alpha$	0.300	0.050	0.192	0.127	0.211	0.194	0.164	0.242

Grid value	$\lambda$ frequency	$\lambda_{DS}$ frequency
0.20	544	165
0.35	162	575
0.50	101	169
0.75	49	5
1.00	23	1
1.25	18	1
2.00	2	0
2.50	5	0
5.00	1	0

Table 5: Distribution of hyper-parameters