Online Appendix for
Selecting Structural Innovation in DSGE models*

Filippo Ferroni
Chicago FED

Stefano Grassi
Department of Economics and Finance,
University of Rome 'Tor Vergata' and CREATES

Miguel A. León-Ledesma
School of Economics and MaGHiC
University of Kent

March 8, 2018

A Appendix

A.1 A basic RBC with analytical solution

The representative agent maximizes the following stream of future utility

$$\max \mathrm{E}_t \sum_{t=1}^{\infty} \beta^t \log c_t$$

subject to the following constraints:

$$\begin{aligned} y_t &= c_t + k_t \\ y_t &= z_t k_{t-1}^{\alpha}, \end{aligned}$$

where y_t is output, c_t consumption, k_t the stock of capital. β is the time discount factor and α is the capital share in production. The system is perturbed by one exogenous disturbance, technology z_t, which follows an AR process:

$$\log z_t = \rho_z \log z_{t-1} + \epsilon_t \quad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2).$$

The lagrangian is:

$$\mathbf{L} = \mathrm{E}_0 \sum_{t=0}^{\infty} \beta_t \left[\log c_t - \lambda_t \left(c_t + k_t - z_t k_{t-1}^{\alpha} \right) \right].$$

The first order conditions are:

$$\begin{aligned} 1/c_t &= \lambda_t \\ 1/c_t &= \beta \mathrm{E}_t \left(1/c_{t+1} \alpha z_{t+1} k_t^{\alpha-1} \right). \end{aligned}$$

*The views in this appendix are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Federal Reserve Bank of Chicago or any other person associated with the Federal Reserve System.

Permanent income model. Guess a solution of the form $c_t = \gamma y_t$, constant saving rate and substitute into the Euler equation.

$$\begin{aligned} 1 &= \beta \mathrm{E}_t \left(\gamma y_t / \gamma y_{t+1} \alpha z_{t+1} k_t^{\alpha-1} \right), \\ &= \beta \mathrm{E}_t \left(y_t / y_{t+1} \alpha y_{t+1} / k_t \right), \\ &= \beta \mathrm{E}_t \left(\alpha y_t / (y_t - c_t) \right), \\ &= \beta \mathrm{E}_t \left(\alpha y_t / (y_t - \gamma y_t) \right), \\ 1 &= \frac{\alpha\beta}{1-\gamma}. \end{aligned}$$

Hence, $\gamma = 1 - \alpha\beta$. This implies that:

$$k_t = (1-\gamma) y_t = \alpha\beta z_t k_{t-1}^{\alpha}.$$

In logs, we can specify a linear state space model in three equations, a law of motion for the exogenous state (z), a law of motion for the endogenous state (k), and the measurement equation (y) as follows:

$$\begin{aligned} \log y_t &= \log z_t + \alpha \log k_{t-1} + e_t, \qquad e_t \sim \mathrm{N}(0, \sigma_e^2), \\ \log k_t &= \log \alpha\beta + \alpha \log k_{t-1} + \log z_t, \\ \log z_{t+1} &= \rho_z \log z_t + \epsilon_{t+1}, \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2). \end{aligned}$$

At the non stochastic steady state, we have $\log k = 1/(1-\alpha) \log \alpha\beta$ and $\log y = \alpha \log k$

$$\begin{aligned} y_t &= z_t + \alpha k_{t-1} + e_t, \qquad e_t \sim \mathrm{N}(0, \sigma_e^2), \\ k_t &= \alpha k_{t-1} + z_t, \\ z_{t+1} &= \rho_z z_t + \epsilon_{t+1}, \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2), \end{aligned}$$

where small case variables indicate now the log deviation from steady state.

A.2 Stochastic variable selection in state space models

Stochastic variable selection has a long tradition in Bayesian analysis (see, among others, George and McCulloch, 1993, 1997 and the references therein). Recently, this methodology has been extended to state space models and, in particular, to the selection of the unobserved components (level, slope and seasonal cycles) that are the key ingredients in state space modeling (see Frühwirth-Schnatter, 2004, Frühwirth-Schnatter and Wagner, 2010, Grassi and Proietti, 2014 and Proietti and Grassi, 2015). This approached, called stochastic model selection search (SMSS), hinges on two basic ingredients: the non-centered representation of the unobserved components model and the consequent reparameterization of the variance hyperparameters as regression parameters with unrestricted support.
Consider, for example, modeling a time series $\mathbf{y} = \{y_1, \ldots, y_t\}$ using a local level model, see Harvey (1989) for an introduction:

$$\begin{aligned} y_t &= z_t + e_t, \qquad e_t \sim \mathrm{N}(0, \sigma_e^2), \\ z_t &= z_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2), \end{aligned} \tag{1}$$

where the latent process z_t follows a random walk starting from unknown initial value μ_0. A typical specification problem arising for this model is to decide if the random walk z_t is time-varying rather than a simple constant. It is well know that testing $\sigma_\epsilon^2 = 0$ versus $\sigma_\epsilon^2 > 0$ results in a non-regular testing problem,

because the null hypothesis lies on the boundary of the parameter space, see Harvey (1989) and Harvey (2001).
A similar specification problem is deciding which components are present in this time series model. For instance, is it necessary to include z_t, that follows a random walk, or should it be removed because z_t is simply a constant? This is another non-regular problem because, again, the null hypothesis can be rephrased as testing $\sigma_\epsilon^2 = 0$ versus $\sigma_\epsilon^2 > 0$.

The stochastic model specification search methodology proposed by Frühwirth-Schnatter and Wagner (2010) (FS-W) is based on a reparameterisation of (1) with respect to location and scale, known as the non-centred representation. See also Gelfand et al. (1995) and Frühwirth-Schnatter (2004). To give a simple example, the model in equation (1) has the following non-centered representation:

$$\begin{aligned} y_t &= \mu_0 + \sqrt{\theta_1}\tilde{z}_t + e_t, \qquad e_t \sim \mathrm{N}(0, \sigma_e^2), \\ \tilde{z}_t &= \tilde{z}_{t-1} + \tilde{\epsilon}_t, \qquad \tilde{\epsilon}_t \sim \mathrm{N}(0,1), \\ \tilde{z}_0 &= 0, \qquad \theta_1 = \sigma_\epsilon^2, \end{aligned} \tag{2}$$

where the latent states has been rewritten as follows:

$$\begin{aligned} z_t &= \mu_0 + \sqrt{\theta_1}\tilde{z}_t, \qquad t = 1, \ldots, T, \\ \tilde{z}_t &= \tilde{z}_{t-1} + \tilde{\epsilon}_t, \qquad \tilde{\epsilon}_t \sim \mathrm{N}(0,1), \end{aligned} \tag{3}$$

where $\tilde{z}_0$ is the starting value of the random walk and $\tilde{z}_t \sim \mathrm{N}(0, t)$.
Between the non-centered and centered representation, there exists a one to one relation that can be easily shown using (2) and (3):

$$\begin{aligned} y_t &= \mu_0 + \sqrt{\theta_1}\tilde{z}_t + e_t, \qquad e_t \sim \mathrm{N}(0, \sigma_e^2), \\ \tilde{z}_t &= \tilde{z}_{t-1} + \tilde{\epsilon}_t, \qquad \tilde{\epsilon}_t \sim \mathrm{N}(0,1), \end{aligned} \tag{4}$$

and rewriting:

$$\begin{aligned} z_t - z_{t-1} &= \sqrt{\theta_1}(\tilde{z}_t - \tilde{z}_{t-1}), \\ &= \sqrt{\theta_1}\tilde{\epsilon}_t = \epsilon_t, \qquad \epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2). \end{aligned} \tag{5}$$

FS-W's key idea is that the non-centered representation is not identified since the model in equation (2) with $(-\sqrt{\theta_1})(-\tilde{z}_t)$ is observationally equivalent to the same model with $(\sqrt{\theta_1})(\tilde{z}_t)$. As a consequence, the likelihood function is symmetric around zero along the $\sqrt{\theta_1}$ dimension and bimodal if the true $\sqrt{\theta_1}$ is larger than zero. This fact can be exploited to quantify how far the posterior of $\sqrt{\theta_1}$ is removed from zero and, in turn, the value of the variance. We stress that the posterior density can also be 0 allowing for boundary conditions. To estimate the model in equation (2) that is equivalent to the model in (1) a standard RW-MH algorithm can be used, see Gamerman and Lopes (2006) and Geweke (2005). Finally, we have to underline that this methodology can easily be extended to more complex state space models as shown in FS-W and in Grassi and Proietti (2014) and Proietti and Grassi (2015).

We extend this methodology to linear DSGE models. To show the workings of this extension, consider the basic RBC model in section 3 and A.1. In logs, we can specify a linear state space model in three equations, a law of motion for the exogenous state (z), a law of motion for the endogenous state (k), and the measurement equation (y) as follows:

$$\begin{aligned} \log y_t &= \log z_t + \alpha \log k_{t-1} + e_t, \qquad e_t \sim \mathrm{N}(0, \sigma_e^2), \\ \log k_t &= \log \alpha\beta + \alpha \log k_{t-1} + \log z_t, \\ \log z_{t+1} &= \rho_z \ln z_t + \epsilon_{t+1}, \qquad \epsilon_{t+1} \sim \mathrm{N}(0, \sigma_\epsilon^2). \end{aligned} \tag{6}$$

At the non stochastic steady state, we have $\log k = 1/(1-\alpha)\log\alpha\beta$ and $\log y = \alpha \log k$ and dropping the log and transforming the model to eliminate the forward looking term z_{t+1} see Hall et al. (2014), we get:

$$\begin{aligned} y_t &= z_t + \alpha k_{t-1} + e_t, & e_t &\sim \mathrm{N}(0, \sigma_e^2), \\ k_{t-1} &= \alpha k_{t-2} + z_{t-1}, & & \\ z_t &= \rho_z z_{t-1} + \epsilon_t, & \epsilon_t &\sim \mathrm{N}(0, \sigma_\epsilon^2), \end{aligned} \tag{7}$$

where, abusing notation, lower case variables now indicate the log deviation from steady state. Given this, the non-centred state space representation of the model is:

$$\begin{aligned} y_t &= \mu_0 + \sqrt{\theta_1}\tilde{z}_t + \alpha\tilde{k}_{t-1} + e_t, & e_t &\sim \mathrm{N}(0, \sigma_e^2), \\ \tilde{z}_t &= \rho_z \tilde{z}_{t-1} + \tilde{\epsilon}_t, & \tilde{\epsilon}_t &\sim \mathrm{N}(0, 1), \\ \tilde{k}_{t-1} &= \alpha\tilde{k}_{t-2} + \sqrt{\theta_1}\tilde{z}_{t-1}. & & \end{aligned} \tag{8}$$

This formulation is identified as the following steps show. Define the process:

$$z_t = \mu_0 + \sqrt{\theta}_1 \tilde{z}_t.$$

Then we have the following formulation:

$$\begin{aligned} z_t - z_{t-1} &= \sqrt{\theta}_1(\tilde{z}_t - \tilde{z}_{t-1}) \\ &= \sqrt{\theta}_1 \tilde{\epsilon}_t, \end{aligned}$$

$$\begin{aligned} \tilde{k}_{t-1} - \tilde{k}_{t-2} &= \alpha\tilde{k}_{t-2} + \sqrt{\theta}_1\tilde{z}_{t-2} - \alpha\tilde{k}_{t-3} + \sqrt{\theta}_1\tilde{z}_{t-3} \\ &= \alpha(\tilde{k}_{t-2} - \tilde{k}_{t-3}) + \sqrt{\theta}_1(\tilde{z}_{t-1} - \tilde{z}_{t-2}) \\ &= \alpha(\tilde{k}_{t-2} - \tilde{k}_{t-3}) + \sqrt{\theta}_1\tilde{\epsilon}_{t-1}. \end{aligned}$$

It is clear that $\tilde{k}_{t-1}$ is related to the error term $\sqrt{\theta}\tilde{\epsilon}_{t-1}$ as can also be shown in equation (6). The state space formulation of the model then becomes:

$$\begin{aligned} y_t &= \mu_0 + \begin{pmatrix} \sqrt{\theta}_1 & \alpha \end{pmatrix} \begin{pmatrix} \tilde{z}_t \\ \tilde{k}_{t-1} \end{pmatrix} + e_t, \\ \begin{pmatrix} \tilde{z}_t \\ \tilde{k}_{t-1} \end{pmatrix} &= \begin{pmatrix} \rho & 0 \\ \sqrt{\theta}_1 & \alpha \end{pmatrix} \begin{pmatrix} \tilde{z}_{t-1} \\ \tilde{k}_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tilde{\epsilon}_t. \end{aligned} \tag{9}$$

The noncenter DSGE model given in equation (9) can be extended to more complicated model, if we apply the same formulation to a more complicated DSGE model the general non-centered representation becomes:

$$\begin{aligned} y_t &= \Lambda_1 s_{1,t} + \Lambda_2 \Sigma^{1/2} \tilde{s}_{2,t} + e_t, & e_t &\sim \mathrm{N}(0, \sigma_e^2), \\ \tilde{\mathbf{s}}_{t+1} &= \mathrm{A}(\boldsymbol{\Theta})\tilde{\mathbf{s}}_t + \mathrm{B}(\boldsymbol{\Theta})\tilde{\epsilon}_{t+1}, & \tilde{\epsilon}_{t+1} &\sim \mathrm{N}(0, \mathrm{I}_n), \end{aligned}$$

where Λ_1 and Λ_2 are selection matrices, $\mathbf{s}_t = [s_{1,t}, s_{2,t}]$ is a stack of the latent states and $\mathrm{N}(0, \mathrm{I}_n)$ is the multivariate normal distribution with unitary variance. While the standard deviation of $\tilde{\epsilon}_{t+1}$ is fixed and normalized to one in estimation, the diagonal elements of $\Sigma^{1/2}$ are estimated with a normal prior.

A.3 Metropolis Hastings MCMC adjusted for the sign switch

Here, we explain the steps to adjust the RW Metropolis-Hastings MCMC for a random sign switch. Partition the vector of parameters $\mathbf{\Theta}$ as composed of a column vector of structural standard deviation parameters (σ) and a column vector of the remaining parameters ($\mathbf{\Gamma}$), i.e. $\mathbf{\Gamma} = [\Omega, \mathbf{\Theta}]$. Given an initial value, $\mathbf{\Theta}_0$, and the information matrix, Ω, from the maximization step, given the size of the jump c, and given a positive sign for the standard deviation, i.e. $\mathrm{W} = 1$, for $\ell = 1, \ldots, L$

1. Draw a candidate draw from $\mathbf{\Theta}^* \sim \mathrm{N}(\mathbf{\Theta}_{\ell-1}, c\Omega)$.
2. Plug it in the DSGE model, $\mathrm{E}_t\mathrm{F}(x_{t+1}, x_t, x_{t-1}, \epsilon_t; \mathbf{\Theta}^*) = 0$.
3. Solve the DSGE, and obtain the state space representation

$$\begin{aligned} y_t &= \Lambda\ s_t + e_t, \\ s_{t+1} &= \mathrm{A}(\mathbf{\Theta}^*)s_t + \mathrm{B}(\mathbf{\Theta}^*)\ \Sigma(\sigma^*)\ \mathrm{W}\ \epsilon_{t+1}. \end{aligned}$$

4. Compute the likelihood using the Kalman filter, i.e. $\mathcal{L}(y|\mathbf{\Theta}^*)$.
5. Contrast the kernels of the candidate draw and previous accepted draws:

$$\mathrm{R} = \frac{p(\mathbf{\Theta}^*)\mathcal{L}(\mathbf{\Theta}^*|y)}{p(\mathbf{\Theta}_{\ell-1})\mathcal{L}(\mathbf{\Theta}_{\ell-1}|y)}.$$

6. Keep the draw with certain probability. Draw $u \sim \mathrm{U}(0,1)$ and:

$$\begin{aligned} if\ \ \mathrm{R} > u, &\quad \mathbf{\Theta}_\ell = \mathbf{\Theta}^*, \\ if\ \ \mathrm{R} \leq u, &\quad \mathbf{\Theta}_\ell = \mathbf{\Theta}_{\ell-1}. \end{aligned}$$

7. Switch the sign of the standard deviation with 0.5 probability. Draw form a binomial distribution, $X \sim b(1/2)$, and set the sign of the standard deviation with $\mathrm{W} = -1 + 2X$. Multiply the standard deviation of the structural shocks times W,

$$\sigma_\ell = \mathrm{W}\sigma_\ell.$$

8. Go back to 1.

A.4 Gibbs -Metropolis Hastings MCMC for non-diagonal and rank-deficient matrix

One might be interested in estimating a non-diagonal covariance matrix, Σ_ϵ, with rank $r = rank(\Sigma_\epsilon) < n$. Such practice might be motivated by Cúrdia and Reis (2010) who offer reasons for why arbitrary restrictions on the correlation structure of DSGE model disturbances may be incorrect. It is possible to design an estimation procedure that accounts for both a non-diagonal covariance structure of the data and a rank-deficient covariance matrix.

The estimation procedure combines the ideas of the conjugate-conditional algorithm of Cúrdia and Reis (2010) and the singular generalized IW (see Uhlig, 1994 and Díaz-García and Gutiérrez-Jáimez, 1997). In particular, the sampling of the parameters needs to be partitioned in two blocks: the covariance matrix of the structural shocks (Σ_ϵ) and $\mathbf{\Gamma} = [\Omega, \mathbf{\Theta}]$, where Ω is the covariance matrix of the measurement errors and $\mathbf{\Theta}$ are all other parameters.

Conditional on a value of $\mathbf{\Theta}$ and on a sequence of states $s_{1:T}$, we can derive a sequence of i.i.d. structural shocks as follows:

$$\mathrm{B}(\mathbf{\Theta})^{+}\ (s_{t+1} - \mathrm{A}(\mathbf{\Theta})s_t) = z_{t+1} = \epsilon_{t+1} \sim \mathrm{N}_r(0, \Sigma_\epsilon),$$

where $\mathrm{B}(\mathbf{\Theta})^{+}$ is the left Moore-Penrose generalized inverse of $\mathrm{B}(\mathbf{\Theta})$. Conditional on a sequence of states $s_{1:T}$, this model can be cast in matrix form as:

$$\mathrm{Z} = \mathrm{E},$$

where $\mathrm{Z} = (z_1, \ldots, z_T)'$ and $p(\mathrm{E}|\Sigma_\epsilon) \equiv \mathrm{N}_{T,r}(0, \Sigma_\epsilon \otimes \mathrm{I}_T)$. As in Díaz-García and Gutiérrez-Jáimez (2006), we denote by $\mathrm{N}_{T,r}(0, \Sigma_\epsilon \otimes \mathrm{I}_T)$ the $T \times n$ multivariate singular Normal distribution with rank r. The likelihood of the normal singular population can be written as:

$$\mathcal{L}(\Sigma_\epsilon|\mathrm{Z}) \propto \left(\prod_{k=1}^{r} \lambda_k\right)^{-\frac{T}{2}} \exp\left(-1/2\ \mathrm{trace}(\Sigma_\epsilon^{+}\mathrm{Z}'\mathrm{Z})\right),$$

where λ_k are the non-null eigenvalues of Σ_ϵ. We consider the following non informative prior density for Σ_ϵ:

$$\mathrm{P}(\Sigma_\epsilon) \propto \left(\prod_{k=1}^{r} \lambda_k\right)^{-\frac{2n-r+1}{2}},$$

where λ_k are the non null eigenvalues of Σ_ϵ. Combining prior and posterior, we obtain the posterior distribution of Σ_ϵ:

$$p(\Sigma_\epsilon|\mathrm{Z}) \propto \left(\prod_{k=1}^{r} \lambda_k\right)^{-\frac{T+2n-r+1}{2}} \exp\left(-1/2\ \mathrm{trace}(\Sigma_\epsilon^{+}\mathrm{Z}'\mathrm{Z})\right). \tag{10}$$

This is an n-dimension singular generalized IW of rank r with $\nu = T - n + 1$ degrees of freedom and scale matrix $\mathrm{G} = \mathrm{Z}'\mathrm{Z}$, denoted by $\mathrm{W}^{+}(r, \nu, \mathrm{G})$.

We are now in a position to propose the following algorithms. Given r, $\Sigma_\epsilon^{(0)}, \mathbf{\Theta}^{(0)}$:

Algorithm 1

1. Draw $s_{1:T}^{(j)}$ *from* $p(s_{1:T}^{(j)} \mid y_{1:T},\ \mathbf{\Theta}^{(j-1)},\ \Sigma_\epsilon^{(j-1)})$.
This distribution is derived from the state space and, given linearity assumptions, it is a gaussian normal distribution.

2. Draw $\Sigma_\epsilon^{(j)}$ *from* $p(\Sigma_\epsilon^{(j)} \mid y_{1:T},\ \mathbf{\Theta}^{(j-1)},\ s_{1:T}^{(j)})$.
This distribution is an n-dimension singular generalized IW, $W^{+}(r, \nu, G^{(j)})$*, where* $G^{(j)} = Z^{(j)'}Z^{(j)}$*,* $Z^{(j)} = (z_1^{(j)}, \ldots, z_T^{(j)})'$ *and* $z^{(j)} = B(\mathbf{\Theta}^{(j-1)(+)}\ (s_{t+1}^{(j)} - A(\mathbf{\Theta}^{(j-1)})s_t^{(j)})$*, and degrees of freedom* $\nu = T - n + 1$

3. Draw $\mathbf{\Theta}^{*}$ *form a normal centered in* $\mathbf{\Theta}^{(j-1)}$ *and accept the draw with a Metropolis-Hastings probability, i.e.*

$$\min\left\{\frac{\mathcal{L}(y_{1:T}|\mathbf{\Theta}^{*}, \Sigma_\epsilon^{(j)})p(\mathbf{\Theta}^{*})}{\mathcal{L}(y_{1:T}|\mathbf{\Theta}^{(j)}, \Sigma_\epsilon^{(j)})p(\mathbf{\Theta}^{(j-1)})}, 1\right\}.$$

In step (3) the likelihood is computed using the Kalman filter recursions. Since the state space is augmented with n_y measurement errors, the covariance matrix of the observables is full rank, hence invertible. Therefore, the Kalman gain, defined as the product of the covariance between states and observables times the inverse of the variance of the observables, can be computed and all the remaining recursions are unaffected.

In order to obtain draws at step (2) use the following algorithm:

Algorithm 2 *Singular IW.*
If any U is an n-dimensional Wishart singular with degrees of freedom ν and scale matrix C, where both U and C are $n \times n$ symmetric positive semi-definite singular matrices of rank r, then we can draw U as follows:

1. *For $C = PLP'$, where PLP' is the non-singular part of the spectral decomposition of C, calculate $PL^{1/2}$.*
2. *Generate $x_1, \ldots, x_\nu$ independently from $N(0, I_r)$.*
3. *$U = \sum_{i=1}^{\nu} W_i W_i'$, where $W_i = Bx_i$ and $B = PL^{1/2}$.*

*Then U is drawn from the n-dimensional Wishart singular with ν degrees of freedom, scale matrix C, and rank r, and U^+ is drawn from the n-dimensional singular generalized **Inverse** Wishart with ν degrees of freedom, scale matrix C, and rank r. $+$ stands for the Moore-Penrose generalized inverse.*

For step (1), we know that $s_{t|T}$ is normally distributed as a singular multivariate normal distribution, i.e. $s_{t|T} \sim \mathrm{N}_r(s_{t-1|T}, Q_{t-1|T})$ where $\mathrm{Q}_{t-1|T}$ is the covariance of the states. Drawing from this distribution is easy: take the non singular part of the spectral decomposition of $\mathrm{Q}_{t-1|T}$, i.e. $\mathrm{Q}_{t-1|T} = PLP'$, draw x from a normal $\mathrm{N}(0, I_r)$ and $s_{t|T}^{(j)} = PL^{1/2}x$.

This algorithm relies on the assumption that the rank of the covariance matrix of structural shocks is known. In applied work, of course, this is not the case. Two approaches can be used to tackle this problem. The first is to run a preliminary test on the data to select the number of common factors that explain a pre-specified portion of the volatility of the observed data. The second is to estimate different specifications with increasing rank dimension from 1 to the number of shocks and select the one that maximizes the marginal likelihood. Once the rank of the covariance matrix is established, redundant or non-fundamental shocks can be obtained by looking a the null space of the posterior distribution of the covariance matrix.

A.5 Marginal Likelihood of RBC. Small sample, large sample and Montecarlo.

We simulate artificial data from the RBC model of section A.1 with $\alpha = 0.33$, $\rho = 0.95$ and $\sigma_e = 0.08$. We generate 50 data set of 1500 data points from the RBC model with $\sigma = 0, 0.05, 0.1, 0.3$, and retain the last 100 (1000) for computing the marginal likelihood with Laplace approximation. We assumed the following priors where m stands for the mean, and SD for the standard deviations,

- IG with $m = 0.05, 0.1, 0.3, 0.4$ and $SD = 5$;
- Normal $m = 0.00, 0.1, 0.3, 0.4$ and $SD = 5$;

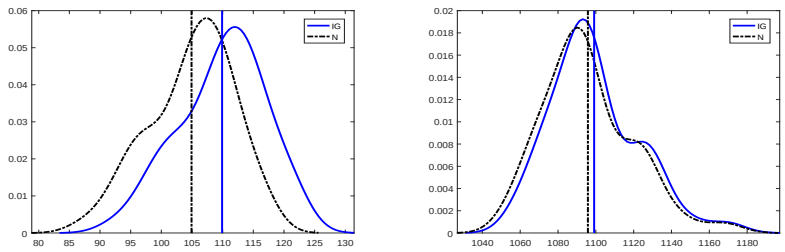
Table 1 reports the log marginal likelihood under different prior settings and DGP with different values for σ, where we consider both short and large sample. The inverse gamma prior with prior mean close to the true value of the DGP has a marginal likelihood larger than the corresponding values assuming the normal prior, even when the inverse gamma prior is unable to select the correct model when $\sigma = 0$. This occurs both in short and large sample.

Figure 1 reports the average and the kernel distribution of log marginal likelihood across 50 different DGP generated with $\sigma = 0$, in blue inverse gamma priors with mean 0.05 and standard deviation 5 and dash dotted balck normal priors with mean 0.05 and standard deviation 5 (left panel short sample and right panel large sample). We observe that the log marginal likelihood computed assuming the inverse gamma prior is on average larger than the marginal likelihood with normal prior, even when the inverse gamma prior is unable to select the correct model. The log marginal likelihood difference beween the normal and inverse gamma prior is of about 5 points in favor of the inverse gamma prior.

Table 1: *Log marginal likelihood (laplace approximation) of different DGP with different prior locations for the normal (N) and inverse gamma (IG) case. The standard deviation is fixed at 5.*

	$\sigma = 0$		$\sigma = 0.05$		$\sigma = 0.1$		$\sigma = 0.3$	
	IG	N	IG	N	IG	N	IG	N
prior mean				$T = 100$				
0.05	88.6	84.5	64.3	59.1	42.0	37.9	-29.1	-30.3
0.1	87.4	84.5	64.1	59.1	42.7	37.9	-27.8	-30.3
0.2	84.6	84.5	62.3	59.1	42.0	37.9	-26.8	-30.3
0.3	81.5	84.5	60.1	59.1	40.4	37.9	-26.5	-30.3
0.4	78.2	84.5	57.8	59.1	38.4	37.9	-26.5	-30.3
prior mean				$T = 100$				
0.05	1102.9	1100.8	870.7	865.2	563.5	559.7	-552.6	-551.7
0.1	1098.3	1100.8	870.1	865.2	564.4	559.7	-551.2	-551.7
0.2	1089.8	1100.8	865.1	865.2	564	559.7	-549.9	-551.7
0.3	1082	1100.8	858.2	865.2	562	559.7	-549.2	-551.7
0.4	1074.8	1100.8	850.7	865.2	558.9	559.7	-548.8	-551.7

Figure 1: *Montecarlo. Distribution of log marginal likelihood estimates (laplace approximation) of normal and inverse gamma prior when DGP has $\sigma = 0$. Small and large samples. Vertical lines means across dataset.*



(a) ML Distribution $T = 100$ (b) ML Distribution $T = 1000$

A.6 Inference Distorisions with SW(2007) model

In fully fledged DSGE models, the persistence of model dynamics is controlled not only by the autoregressive parameters, but also by the deep parameters capturing real and nominal frictions in the economy. To quantify these distortions, we consider a baseline DSGE model as presented in Smets and Wouters (2007) (henceforth SW). This model is selected because of its widespread use for policy analysis among academics and policymakers, and because it is frequently adopted to study cyclical dynamics and their sources of fluctuations in developed economies. We retain the nominal and real frictions originally present in the model, but we make a number of simplifications which reduce the computational burden of the experiment but bear no consequences on the conclusions we reach. First, we assume that all exogenous processes are stationary. Second, we assume that all the shocks are uncorrelated and follow an autoregressive process of order one. Third, since we do not want to have our results driven by identification issues (see Komunjer and Ng, 2011 or Iskrev, 2010), we fix a number of parameters and estimate only a subset of them. We estimate the standard errors, autoregressive parameters, and the parameters driving price and wage indexation and stickiness, habit in consumption, intertemporal elasticity of substitution and the inverse of the elasticity of investment (relative to an increase in the price of installed capital).

Table 2: *Montecarlo experiment with 100 artificial datasets. The table reports the bias measured as the difference between the average posterior mean (MCMC posterior estimates of the parameter) and the true value. The priors considered are Normal, IG and Exp. The table also reports the true simulated value. ϕ is the inverse of the elasticity of substitution of installed capital, λ habit in consumption, ζ_w wage stickiness, i_w wage indexation, i_p price indexation, ζ_p price stickiness, r_p MP rule response to inflation, r_{dy} MP response to output, r_y MP response to output, ρ interest smoothing, σ_c intertemporal substitution.*

	Structural parameters $\boldsymbol{\Theta}$ Bias										
	ϕ	λ	ζ_w	i_w	i_p	ζ_p	r_p	r_{dy}	r_y	ρ	σ_c
IG Prior	-1.32	-0.06	-0.07	-0.19	-0.17	-0.11	0.19	-0.04	0.02	-0.04	0.10
Normal Prior	0.00	0.00	0.00	0.01	0.01	0.00	-0.06	0.00	0.01	0.00	-0.01
Exp prior	0.01	0.01	0.01	0.02	-0.01	0.00	0.01	0.01	0.02	0.01	0.01
True value	5.74	0.71	0.70	0.59	0.24	0.65	2.05	0.22	0.09	0.81	1.38

To study the effect of estimating non-existent shocks, we switch off the price markup, the wage markup, and the investment specific shocks and add seven measurement i.i.d. errors (one for each observable) with standard deviation equal 0.08 which we estimate along with the other parameters.[1] With this calibration, measurement errors explain on average less then 3% of the volatility of observables. We simulate 1,000 data points and use the last 200 for inference. We consider seven observable variables: output y_t, consumption c_t, investment i_t, wages w_t, inflation π_t, interest rates r_t, and hours worked h_t. We estimate the model assuming IG, Exp, and Normal priors for the standard deviations and the same priors as in SW for the remaining parameters. For each of these three specifications, we estimated the standard deviations of measurement errors for which we assumed inverse gamma distributions centered on the true value with a loose precision. We run a 300,000 draws MCMC routine starting from the posterior kernel mode and burn-in the first 200,000 of the chain and keep randomly 1,000 for inference. Converge is checked by means of Brook and Gelman (1998) diagnostics for a subset of estimates [2]

We simulated various samples and estimated the posterior distributions of the parameters using our three different prior distributions, i.e. Normal, Exp and IG. In all samples, normal and exponential priors were able to retrieve primal shocks from non-existent ones. We measured the bias as the difference between the average posterior mean of different samples and the true value, and we report the results for deep structural parameters in table 2 (see table 6 in the paper for the description of the parameters). A positive value means that we are *overestimating* a parameter, and a negative value that we are *underestimating* it. Bar few exceptions, the bias using Normal or Exp priors is negligible as the order of magnitude is small. In all cases, the bias using IG priors is larger than the one using Normal or Exp priors. With IG priors, we obtain sizable distortions in parameters capturing persistence. In particular, price and wage indexation parameters are systematically underestimated.

Incorrect assumptions about the existence of structural shocks do not only distort parameter estimates, but they have deep consequences for the implications of the model regarding the sources of business cycle fluctuations or the dynamic transmission of structural shocks which are important for policy analysis. Table 3 reports the variance decomposition of output, inflation, wages, and the interest rate in terms of primal shocks under various prior assumptions about their standard deviations. Price and wage markup shocks

[1] We also reduced this value to 0.05 and increased it to 0.35 and the main conclusions are unaffected. We discuss a more complete set of robustness checks in section 3.1.

[2] In particular, we computed the interval range of the pooled chains and the average interval range within chains and verified when these two lines stabilize and lie one on top of each other. In most of the cases, convergence is achieved within 150,000 draws.

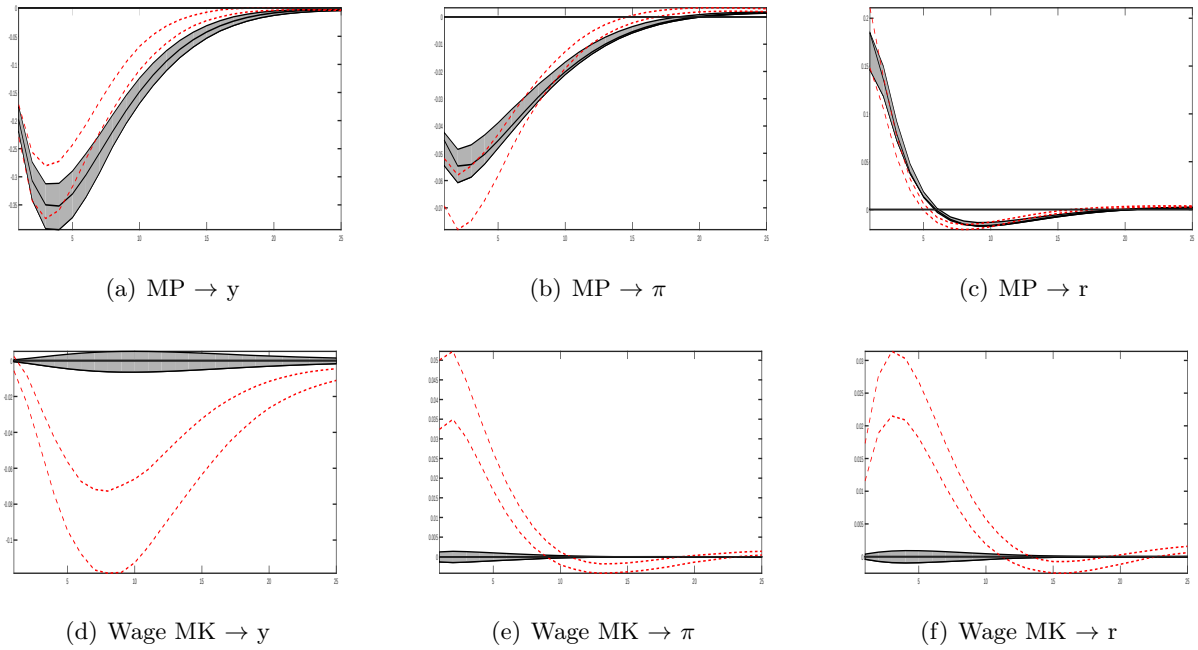
should not explain fluctuations in any of these variables. This is the case for Normal and Exp priors. It is not the case, however, for IG priors where wage and price markup shocks together explain 16% of the volatility of inflation and 8% of the volatility of wages. Moreover, the transmission of shocks is altered in a

Table 3: *Fraction of the variance of output (y), inflation (π), interest rate (r) and wages (w) explained by the non-existing shocks. In brackets, the 5%-95% confidence sets. The non-existing shocks are: price markup (σ_p), wage markup (σ_w) and investment specific (σ_i) shocks. In the true DGP the explained variances are zero.*

Priors	y	π	r	w
IG prior on σ_i	0 [0,0]	0 [0,0]	0[0,0]	0[0,0]
IG prior on σ_w	1 [0,1]	9[6,14]	2 [1,3]	8[5,13]
IG prior on σ_p	0 [0,0]	7 [5,8]	0 [0,1]	0[0,0]
Normal prior on σ_i	0 [0,0]	0 [0,0]	0 [0,0]	0[0,0]
Normal prior on σ_w	0 [0,0]	0 [0,0]	0 [0,0]	0[0,0]
Normal prior on σ_p	0 [0,0]	0 [0,0]	0 [0,0]	0[0,0]
Exp prior on σ_i	0 [0,0]	0 [0,0]	0 [0,0]	0[0,0]
Exp prior on σ_w	0 [0,0]	0 [0,0]	0 [0, 0]	0[0,0]
Exp prior on σ_p	0 [0,0]	0 [0,1]	0 [0,0]	0[0,0]

substantial way. Figure 2 reports the transmission of monetary policy shocks (top row) and wage markup shocks (bottom row) to output, inflation, and the interest rate. Gray areas (red dashed lines) represent the 90% confidence sets of the response assuming Normal (IG) priors on standard deviations and the black solid line the true response. The responses to a monetary policy shock are qualitatively different under the two settings. IG priors tend to produce less persistent dynamic responses. Moreover, on impact, we *overestimate* the reaction of inflation to an interest rate hike and *underestimate* the reaction of output. For an interest rate hike of 15-20 basis points, inflation declines more than than it should (black line) and output recovers much faster. In this context, disinflation trajectories might result to be less costly in terms of output loss relative to what they truly are. With normal priors (gray shaded areas) the disinflation trajectories are correctly picked up in terms of size and in terms of speed of adjustment. Even more striking, as would be expected, are the responses to "non-existing" shocks. In the normal prior setup, we obtain statistically insignificant dynamics for all the variables of interest to an increase in the wage markup. Conversely, with IG priors, output and inflation react strongly and their responses are statistically and economically significant.

Figure 2: *Impulse response function of: output (y), inflation (π) and interest rate (r) to a monetary policy shock (MP, top row) and to a wage markup shock (Wage MK bottom row) with IG prior (red dashed lines) and Normal prior (gray shaded area). Black line true impulse response function. Results for the Exp prior are not reported for readability reason. They are not different from the Normal prior.*



(a) MP → y (b) MP → π (c) MP → r

(d) Wage MK → y (e) Wage MK → π (f) Wage MK → r

A.7 The sensitivity of the posterior to the standard deviation prior location

In this section we study the sensitivity of the posterior to the standard deviation prior location using US data and the SW model.

A noteworthy drawback of using inverse gamma priors on the STD of shocks concerns the sensitivity of the posterior estimates to the prior location. On the contrary, the posterior analysis is invariant to the prior location when assuming normal priors. In other words, even if the parameter estimates driving some impulse response of interest were not too different between inverse gamma and normal priors, posterior estimates are affected by the inverse gamma prior location. This is an added advantage of using normal priors for STDs.

When we consider the SW model with the same the number of shocks and of observables (i.e. 7), the prior distribution of the shocks STD is largely uninfluential for the posterior parameter estimates[3]. All the shocks are needed to explain the stochastic dimension of the observables. And, regardless of the prior assumptions on the STD of the shocks, we obtain statistically indistinguishable posterior estimates. However, in the inverse gamma prior setting, the marginal likelihood varies substantially depending on the prior location, even if the posterior parameter estimates do not. In this context, marginal likelihood comparisons with inverse gamma priors are difficult to interpret.

When we have more shocks than observables (i.e. with measurement errors), the likelihood has something to say about the most likely combination of shocks, i.e. the model specification. As a consequence, it is not longer true that the STD prior distribution does not matter for the posterior analysis. The prior density could favor or penalize the likelihood of a specific model shock configuration, depending on the prior dispersion and location. As a result, it might influence posterior analysis. While this not wrong *per-se*, one should be aware of the extent to which the priors on STDs influence posterior analysis.

[3] Assuming 'reasonable' locations and dispersions.

Table 4: *Log marginal likelihood (Laplace approximation) of the SW model (no measurement errors) and its components, under different prior locations for the normal (N) and inverse gamma (IG) case.*

	Prior Mean							
	0.08	0.1	0.2	0.4	0.8	1	1.2	1.5
				Normal prior				
Laplace Approximation	-916.7	-916.7	-916.7	-916.7	-916.7	-916.7	-916.7	-916.7
				Inverse Gamma prior				
Laplace Approximation	-907.6	-904.6	-896.1	-891.1	-898.1	-906.4	-916.9	-935.8
(log) likelihood	-801.1	-801.1	-801.2	-801.4	-802.8	-804.2	-806.4	-811.1
(log) prior	-28.1	-25.1	-16.6	-11.3	-17.0	-23.8	-32.2	-46.6
(log) constant	32.2	32.2	32.2	32.2	32.2	32.2	32.2	32.2
(log) det inverse Hessian	-110.5	-110.5	-110.5	-110.5	-110.6	-110.6	-110.5	-110.3

Contrary to the inverse gamma case, the normal prior on STD is largely uninfluential for the computation of the posterior parameter distributions and of the marginal likelihood (as long as we have a sufficiently loose precision). It does not really matter if we postulate a priori that the STD of the measurement error on - say - interest rate is close to zero or not when we assume normal priors. The NN setting generates estimates of the posterior distributions and of the marginal data density that are invariant to the prior mean. This appealing property of the normal prior on shocks STD does not carry over for the inverse gamma distribution, where the posterior parameter and the marginal likelihood estimates are very sensitive to the inverse gamma location parameter. In such a case, different configurations of prior means lead to different estimates of the parameters and of the marginal likelihood. This brings us back to square one, as we need a device to select among the prior hyper-parameter.

To show these two points empirically, we run the following exercises. We first considered the case with 7 primal shocks and seven observable variables and no measurement errors. We specified various locations for the prior mean of the STD under the normal and inverse gamma settings. In particular, we assumed that the prior mean for the STD lies in this discrete range of values $[0.08, 0.1, 0.2, 0.4, 0.8, 1, 1.2, 1.5]$ for both the normal prior and the inverse gamma case, and in both setups, we assumed a loose standard deviation equal to 10. For each of these prior specifications (eight for each prior distribution), we estimated the mode of the posterior kernel and computed the log marginal likelihood with the Laplace approximation. Regardless of the prior location, the estimated posterior mode is the same in the normal and inverse gamma setup[4]. Table 4 reports the log marginal likelihood of this exercise and its components. While there are no variations in the marginal likelihood in the normal prior case, there are significant differences in terms of marginal likelihood when inverse gamma priors are used. In other words, since the likelihood for the case with, say, $IG(0.1, 10)$ is the same as the one with $IG(0.4, 10)$, the 13 Log ML difference reported is entirely coming form the prior density rather than from an improvement in the Kalman filter one-step ahead prediction error. The same argument applies when contrasting the marginal likelihood of the N and IG setup. In this sense, marginal likelihood comparisons do not appear to be a useful tool to compare model fit.

When we also introduce measurement errors, the likelihood can express a preference between model specifications (i.e. shocks configuration) which can be favored or penalized by the prior. To assess the extent to which this occurs, we considered different prior locations for measurement errors and for primal shocks, varying on the same discrete range of values. This generated 64 different variants of priors for each prior distribution assumption, the NN and the $IGIG$. For each of them, we estimated the mode of the posterior kernel and computed the log marginal likelihood with the Laplace approximation.

[4]To save space, we do not show the results here. They can be consulted on the online appendix in section B, Figure 3-5.

Table 5: *Log marginal likelihood (Laplace approximation) of the SW model with measurement errors, under different prior locations for the normal (NN) and inverse gamma (IGIG) case.*

	$IGIG$ Prior mean							
STR/ME	0.08	0.1	0.2	0.4	0.8	1	1.2	1.5
0.08	-863	-861	-854	-851	-860	-868	-879	-898
0.1	-862	-869	-852	-850	-859	-868	-879	-897
0.2	-861	-858	-849	-848	-860	-870	-881	-900
0.4	-859	-857	-853	-854	-869	-879	-892	-912
0.8	-877	-876	-877	-877	-894	-907	-921	-943
1	-888	-887	-887	-888	-909	-922	-937	-960
1.2	-900	-900	-900	-904	-925	-939	-953	-976
1.5	-921	-920	-922	-927	-952	-966	-981	-1003
	NN							
	Prior location $me, str = 0, 0.1, ..., 1.5$							
	-893							

Table 5 reports the log marginal likelihood of the two specifications, $IGIG$ and NN, under different prior locations. As before, in the NN case with a sufficiently loose prior, the log marginal likelihood (ML) is invariant to the prior mean, meaning that normal priors do not favor nor penalize the likelihood. In the $IGIG$ case there are large swings in the ML, which are the result of the extent to which priors and likelihood are in accordance. Under this prior parametrization, we find a maximum difference of 140 ML points. Even if we focus on more 'reasonable' prior locations, say between 0.1 and 0.2, we still find large changes in the ML. For instance, a variant with $IG(0.1, 10)$ on both measurement and primal shocks has a ML of -869 and a variant with $IG(0.2, 10)$ on both measurement and primal shocks has a ML of -849. The difference is still substantial and, from a Bayesian model selection perspective, it would justify selecting the latter prior location over the former.

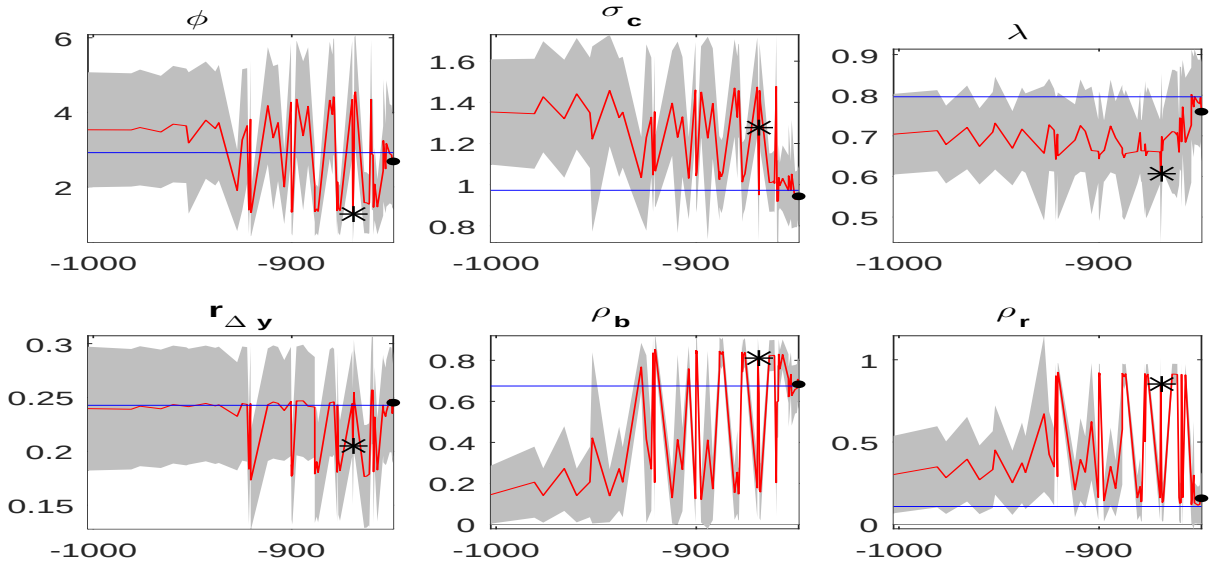
In this case, the large swings in marginal likelihood are associated with large variations in parameter estimates. While they turn out to be invariant for the NN case, they are unstable in the $IGIG$ one. Figure 8 reports the mode estimate (red) and the uncertainty (gray) around it for the $IGIG$ variants against the marginal likelihoods so that the mode estimates are ordered from the lowest to the highest marginal likelihood values. We also plot the mode estimates (blue) of the NN variants. Interestingly, even at the left end of the plots (where marginal likelihoods are substantially similar), some parameters exhibit large swings in the mode point estimate, e.g. ϕ, ρ_b and ρ_r. Even if one believes that the point estimate analysis does not portray the full picture and wishes to simulate the full posterior distribution, some issues about using the inverse gamma prior remain. This is particularly the case for the autoregressive parameter of the monetary policy shock.[5] And as highlighted in the previous section, this parameter is crucial in order to obtain sensible estimates of the transmission of monetary policy shocks.[6]

One could argue that a finer grid of prior mean values tailored to each individual (primal or measurement error) shock might lead to select the vector of shock-specific hyper-parameters that maximizes the

[5] In the online appendix (section A.6), we report a detailed posterior analysis for this parameter, e.g. convergence of the chains and shape of the posterior distribution. The main finding is that with inverse gamma prior on both measurement errors and structural shocks, the posterior distribution of ρ_r is bimodal.

[6] Note that in figure 8 we report, with a black dot, the mode of the estimate of parameters for the $IGIG$ variant that maximizes the marginal likelihood. The mode for ρ_r differs substantially from the mean reported in table **??** because, as mentioned, the posterior is bimodal. As can be seen in figure 2 in the online appendix, the chain flips between 0.2 and 0.8.

Figure 3: *Estimated mode selected parameters for NN (blue) and $IGIG$ (red) under different prior locations. Gray bands report the 90% confidence around the mode of specification $IGIG$ (red). On the horizontal axis the marginal likelihood of the $IGIG$ reported in table 5. Black dots represent the estimated variant that maximizes the ML. The asterisk the $IGIG(0.1, 10)$*



marginal likelihood and generates meaningful transmission mechanisms. This could, in principle, be the case. However, we can see two major drawbacks with this approach. First, practical. It would blow up the computing time for estimation. Second, methodological. As Bayesian econometricians, we want to put a reasonable 'distance' between our prior assumptions and the information contained in the data. Such an approach would inevitably blur this distinction.

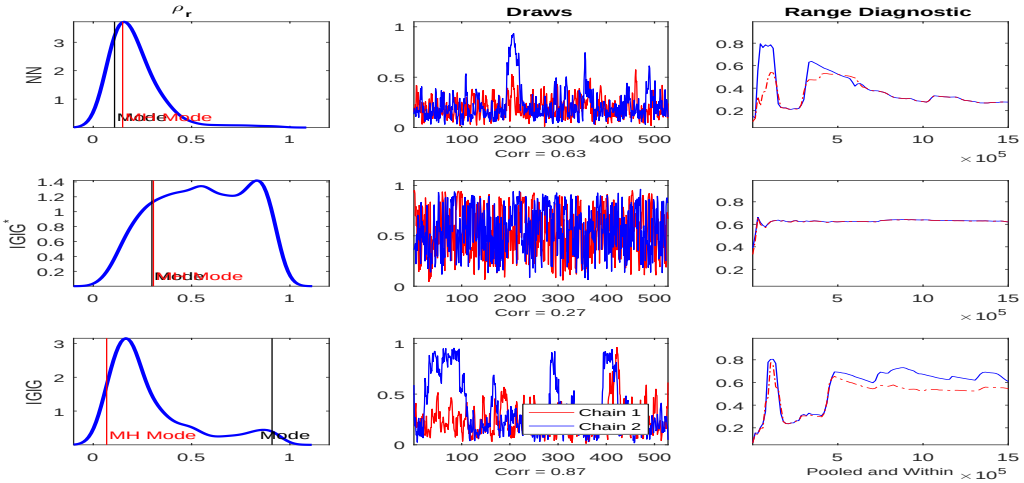
In all, when we have the same number of shocks as observables, normal and inverse gamma priors deliver the same posterior estimates of the parameters regardless of the prior location. However, marginal likelihood comparisons are difficult to interpret as the marginal likelihood estimates using inverse gamma priors are not invariant to the prior location. When there are more shocks than observables, inverse gamma priors offer a poor platform to generate posterior estimates that are independent of the prior locations. Moreover, it is not straightforward how to select amongst them. Normal priors overcome all these concerns. They offer a direct way of selecting the fundamental drivers of economic fluctuations and are insensitive to the choice of prior locations.

A.7.1 Posterior distribution of ρ_r in the SW model with different priors on STD

As pointed out in the paper, the autoregressive parameter of the monetary policy surprise, ρ_r, is crucial for capturing the monetary policy transmission dynamics. Figure 4 reports the posterior distributions and the draws from the MH algorithm used to construct the posterior, the mode at the optimization step and the mode of the MH algorithm in the three setups, NN, $IGIG$ where $IG^{str}(0.1, 10)$ and $IG^{meas}(0.1, 10)$, and $IGIG^*$ where $IG^{str}(0.2, 10)$ and $IG^{meas}(0.4, 10)$. For each specification, we run 2 chains of 1.5 million draws and target for all chains and specifications an acceptance rate of 25-35 percent; convergence is checked by means of Brook and Gelman (1998) diagnostics as reported in the the last column of figure 4. In particular, we report the interval range of the pooled chains and the average interval range within chains. We have convergence when these two lines stabilize and lie one on top of each other. To inspect the posterior

distribution of ρ_r, we discard the first 750,000 draws of each chain and keep randomly one every thousand draws. When looking at the convergence diagnostics of ρ_r in the $IGIG^*$ setup (second row third column),

Figure 4: *Posterior distributions of ρ_r and the draws from the MH algorithm used to construct the posterior, the mode at the optimization step (black) and the mode of the MH algorithm (red) in the NN (top), the $IGIG^*$ (middle) and the $IGIG_{0.1}$ (bottom) specification*



we would conclude that we achieved convergence for this parameter. However, a closer look at the MCMC chain and the posterior distribution, we see that the posterior distributions of ρ_r is clearly bimodal in the $IGIG$ settings; in both chains, the parameter keeps on flipping between small (0.2) and large values (0.8). Similar are the results for the $IGIG$ specification in which the second chain is characterized by a two regimes sequence (third row second column). Moreover, the mode at the maximization step (black vertical line) is very different from the mode at the end of the Metropolis-Hastings algorithm (red vertical line). As a consequence, the pooled and within chain interval range lines do not stabilize and do not lie one on top of each other, signaling that the MCMC has not converged. The bi-modality and non-convergence in the NN specification are only marginal concerns.

B Additional Tables and Graphs

Table 6: *Decomposition of the Laplace Approximation of the Marginal Likelihood (ML) for the RBC model in Section 3 with different locations for the priors when the true DGP features $\sigma = 0$. Decomposes the ML into a constant, the determinant of the inverse Hessian, and the Kernel. The Kernel is decomposed into the prior and the likelihood.*

	Inverse Gamma prior location				
	0.05	0.1	0.2	0.3	0.4
LaplaceApproximation	88.55	87.36	84.58	81.50	78.20
LaplaceConstant	3.68	3.68	3.68	3.68	3.68
LaplaceHessian	-14.46	-14.35	-14.46	-14.62	-14.78
LaplaceKernel	99.33	98.03	95.36	92.44	89.30
LaplacePrior	7.88	6.83	4.83	2.67	0.35
LaplaceLikelihood	91.45	91.20	90.53	89.77	88.95
	Normal prior location				
	0.05	0.1	0.2	0.3	0.4
LaplaceApproximation	84.47	84.47	84.47	84.47	84.47
LaplaceConstant	3.68	3.68	3.68	3.68	3.68
LaplaceHessian	-12.89	-12.89	-12.89	-12.89	-12.89
LaplaceKernel	93.68	93.68	93.68	93.68	93.68
LaplacePrior	2.08	2.08	2.08	2.08	2.07
LaplaceLikelihood	91.61	91.61	91.61	91.61	91.61

Table 7: *Full MCMC estimates of the RBC model. The model is estimated the with Normal, IG and Exp priors for the standard deviations of structural shocks. The table reports for all the structural parameters* $\mathbf{\Theta}$*: the posterior median, the lower (5%) and upper (95%) quantile(credible set). The parameters are set as follows:*$\alpha = 0.30$*,* $\rho = 0.70$*,* $\sigma_e = 0.08$*. Four values of the structural shock* $\sigma = \{0, 0.05, 0.10, .020\}$ *are considered. The table reports in italics the parameters that are estimated to be zero and with a ——– line parameters that are not identified.*

$\mathbf{\Theta}$	$\sigma = 0$ Median [Lower,Upper]	$\sigma = 0.05$ Median [Lower,Upper]	$\sigma = 0.1$ Median [Lower,Upper]	$\sigma = 0.2$ Median [Lower,Upper]
		IG Prior		
α	0.302 [0.218 , 0.382]	0.325 [0.245 , 0.407]	0.328 [0.247 , 0.409]	0.335 [0.256 , 0.416]
ρ	0.219 [0.076 , 0.447]	0.774 [0.584 , 0.905]	0.760 [0.580 , 0.889]	0.889 [0.794 , 0.958]
σ	0.053 [0.039 , 0.071]	0.072 [0.053 , 0.098]	0.085 [0.063 , 0.112]	0.129 [0.097 , 0.172]
σ_e	0.079 [0.064 , 0.093]	0.087 [0.070 , 0.105]	0.081 [0.061 , 0.101]	0.145 [0.118 , 0.176]
		Normal Prior		
α	0.322 [0.240 , 0.404]	0.327 [0.238 , 0.407]	0.332 [0.243 , 0.408]	0.329 [0.241 , 0.414]
ρ	——– [——-,——-]	0.841 [0.684 , 0.954]	0.735 [0.539 , 0.894]	0.902 [0.811 , 0.966]
σ	*0.000* [-0.051 , 0.054]	0.060 [0.040 , 0.088]	0.083 [0.059 , 0.110]	0.134 [0.100 , 0.176]
σ_e	0.085 [0.061 , 0.101]	0.093 [0.076 , 0.112]	0.082 [0.063 , 0.106]	0.145 [0.117 , 0.178]
		Exp Prior		
α	0.330 [0.245 , 0.408]	0.330 [0.248 , 0.412]	0.331 [0.247 , 0.412]	0.332 [0.249 , 0.412]
ρ	——– [——-,——-]	0.815 [0.637 , 0.928]	0.765 [0.579 , 0.900]	0.882 [0.780 , 0.955]
σ	*0.003* [0.001 , 0.025]	0.062 [0.041 , 0.091]	0.083 [0.058 , 0.111]	0.136 [0.101 , 0.181]
σ_e	0.091 [0.081 , 0.102]	0.091 [0.074 , 0.110]	0.082 [0.062 , 0.102]	0.143 [0.114 , 0.174]

Table 8: *Full MCMC estimates of model the RBC model. The model is estimated using the Normal, the IG and the Exp prior for the standard deviations of structural shock, under different prior settings. The table reports for all the structural parameters* $\boldsymbol{\Theta}$*: the posterior mode, the lower (5%) and upper (95%) quantile(credible set) and the marginal likelihood (ML) computed with the Laplace approximation. The parameters are set as follows:*$\alpha = 0.30$*,* $\rho = 0.95$*,* $\sigma_e = 0.08$*, and the structural shock* $\sigma = 0.05$*.*

	$\sigma = 0.05$		
$\boldsymbol{\Theta}$	Mode [Lower,Upper]	Mode [Lower,Upper]	Mode [Lower,Upper]
	IG(**0.05**,5)	Normal(**0.05**,5)	Exp(**5**,5)
α	0.31 [0.23 , 0.38]	0.30 [0.22 , 0.38]	0.30 [0.22 , 0.38]
ρ	0.88 [0.77 , 0.98]	0.86 [0.78 , 0.98]	0.88 [0.76 ,0.98]
σ	0.04 [0.02 , 0.05]	0.05 [0.02 , 0.08]	0.04 [0.01, 0.06]
σ_e	0.10 [0.09 , 0.12]	0.10 [0.08 , 0.11]	0.10 [0.08 , 0.12]
ML	64.31	59.15	61.36
	IG(**0.1**,5)	Normal(**0.1**,5)	
α	0.30 [0.21 , 0.38]	0.30 [0.22 , 0.39]	
ρ	0.82 [0.67 , 0.95]	0.85 [0.70 , 0.99]	
σ	0.05 [0.03 , 0.07]	0.05 [0.02, 0.07]	—
σ_e	0.10 [0.08 , 0.11]	0.10 [0.08 , 0.12]	
ML	64.06	59.15	
	IG(**0.2**,5)	Normal(**0.2**,5)	
α	0.30 [0.22 , 0.38]	0.30 [0.22 , 0.39]	
ρ	0.73 [0.56 , 0.90]	0.85 [0.72 , 0.97]	
σ	0.07 [0.05 , 0.09]	0.05 [0.02 , 0.07]	—
σ_e	0.09 [0.07 , 0.11]	0.10 [0.08 , 0.12]	
ML	62.30	59.15	

Table 9: *Full MCMC estimates of model the RBC model. The model is estimated using the Normal, the IG and the Exp prior for the standard deviations of structural shock, under different prior settings. The table reports for all the structural parameters* $\boldsymbol{\Theta}$*: the posterior mode, the lower (5%) and upper (95%) quantile(credible set) and the marginal likelihood (ML) computed with the Laplace approximation. The parameters are set as follows:*$\alpha = 0.30$*,* $\rho = 0.95$*,* $\sigma_e = 0.08$*, and the structural shock* $\sigma = 0.10$*.*

	$\sigma = 0.1$		
$\boldsymbol{\Theta}$	Mode [Lower,Upper]	Mode [Lower,Upper]	Mode [Lower,Upper]
	IG(**0.05**,5)	Normal(**0.05**,5)	Exp(**5**,5)
α	0.31 [0.23 , 0.40]	0.30 [0.22 , 0.38]	0.31 [0.23 , 0.40]
ρ	0.91 [0.85 , 0.98]	0.89 [0.83 , 0.98]	0.89 [0.82 , 0.98]
σ	0.06 [0.04 , 0.09]	0.07 [0.04 , 0.10]	0.07 [0.04 , 0.10]
σ_e	0.11 [0.09 , 0.13]	0.10 [0.08 , 0.12]	0.11 [0.09 , 0.13]
ML	42.03	37.89	39.70
	IG(**0.1**,5)	Normal(**0.1**,5)	
α	0.30 [0.22 , 0.39]	0.30 [0.21 , 0.38]	
ρ	0.90 [0.83 , 0.98]	0.90 [0.84 , 0.98]	
σ	0.07 [0.04 , 0.09]	0.07 [0.04 , 0.10]	—
σ_e	0.11 [0.09 , 0.13]	0.11 [0.09 , 0.13]	
ML	42.71	37.89	
	IG(**0.2**,5)	Normal(**0.2**,5)	
α	0.30 [0.22 , 0.38]	0.30 [0.22 , 0.38]	
ρ	0.88 [0.80 , 0.96]	0.89 [0.82 , 0.97]	
σ	0.08 [0.05 , 0.10]	0.07 [0.05 , 0.10]	—
σ_e	0.10 [0.08 , 0.12]	0.11 [0.09 , 0.13]	
ML	42.00	37.89	

Table 10: *Full MCMC estimates of model the RBC model. The model is estimated using the Normal, the IG and the Exp prior for the standard deviations of structural shock, under different prior settings. The table reports for all the structural parameters* $\boldsymbol{\Theta}$*: the posterior mode, the lower (5%) and upper (95%) quantile(credible set) and the marginal likelihood (ML) computed with the Laplace approximation. The parameters are set as follows:*$\alpha = 0.30$*,* $\rho = 0.95$*,* $\sigma_e = 0.08$*, and the structural shock* $\sigma = 0.20$*.*

	$\sigma = 0.20$		
$\boldsymbol{\Theta}$	Mode [Lower,Upper]	Mode [Lower,Upper]	Mode [Lower,Upper]
	IG(**0.05**,5)	Normal(**0.05**,5)	Exp(**5**,5)
α	0.31 [0.22 , 0.39]	0.30 [0.22 , 0.39]	0.31 [0.23, 0.40]
ρ	0.92 [0.88 , 0.97]	0.92 [0.86 , 0.98]	0.92 [0.88 , 0.98]
σ	0.19 [0.14 , 0.24]	0.20 [0.14 , 0.25]	0.19 [0.14 , 0.24]
σ_e	0.16 [0.20 , 0.12]	0.16 [0.11 , 0.20]	0.16 [0.12 , 0.20]
ML	-29.15	-30.29	-29.50
	IG(**0.1**,5)	Normal(**0.1**,5)	
α	0.30 [0.22 , 0.39]	0.31 [0.23 , 0.39]	
ρ	0.92 [0.87 , 0.98]	0.93 [0.88 , 0.98]	
σ	0.19 [0.14 , 0.24]	0.18 [0.13 , 0.23]	—
σ_e	0.16 [0.11 , 0.20]	0.16 [0.13 , 0.20]	
ML	-27.84	-30.29	
	IG(**0.2**,5)	Normal(**0.2**,5)	
α	0.31 [0.23 , 0.40]	0.30 [0.24 , 0.39]	
ρ	0.93 [0.88 , 0.98]	0.93 [0.88 , 0.98]	
σ	0.18 [0.13 , 0.24]	0.19 [0.14 , 0.23]	—
σ_e	0.16 [0.12 , 0.20]	0.16 [0.13 , 0.20]	
ML	-26.77	-30.29	

Figure 5: *Estimated mode of N (blue) and IGIG (red) under different prior locations without ME. Gray bands report the 90% confidence around the mode of specification IGIG (red).*

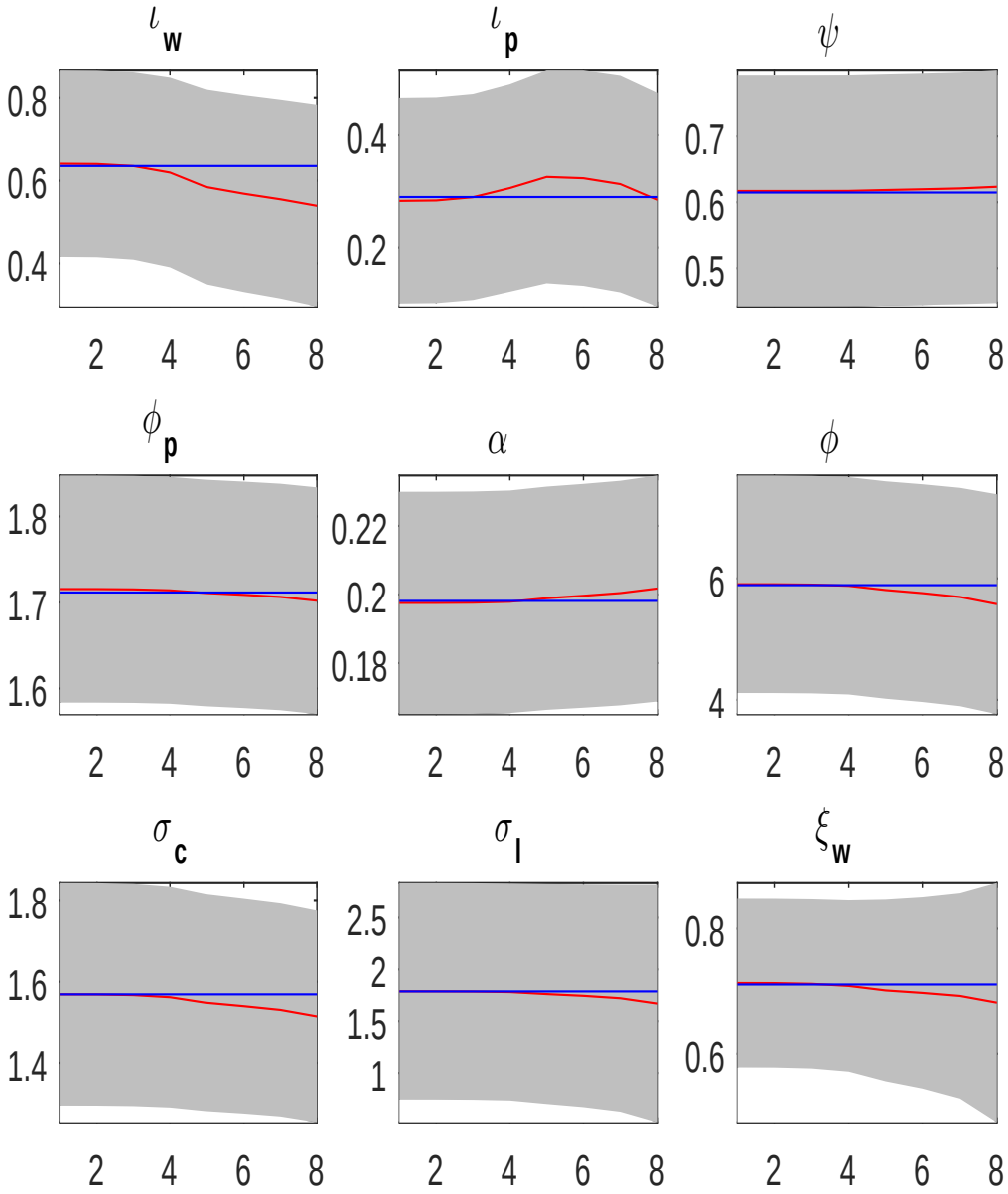


Figure 6: *Estimated mode of N (blue) and IG (red) under different prior locations without ME. Gray bands report the 90% confidence around the mode of specification IGIG (red).*

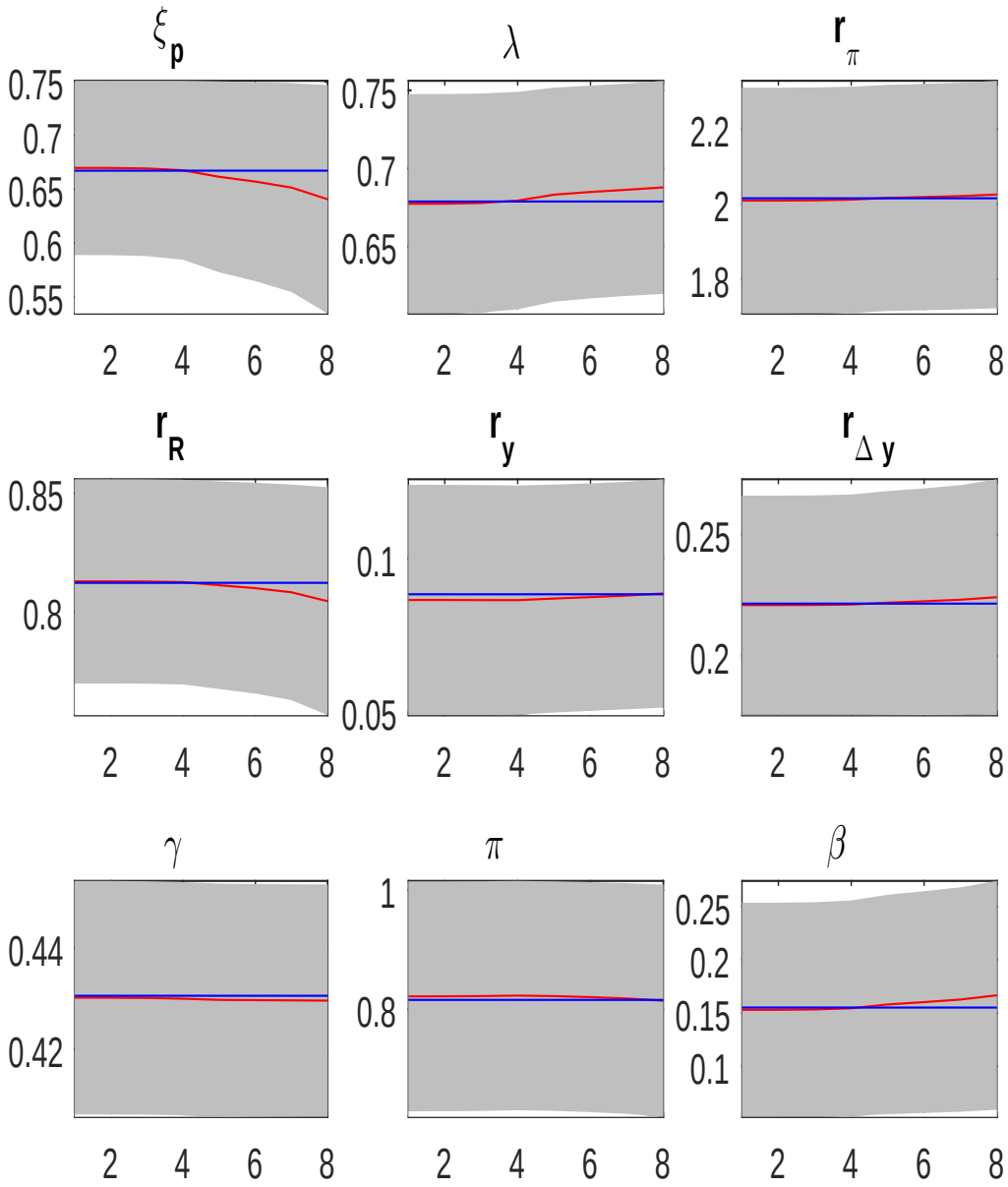


Figure 7: *Estimated mode of N (blue) and IG (red) under different prior locations without ME. Gray bands report the 90% confidence around the mode of specification IGIG (red).*

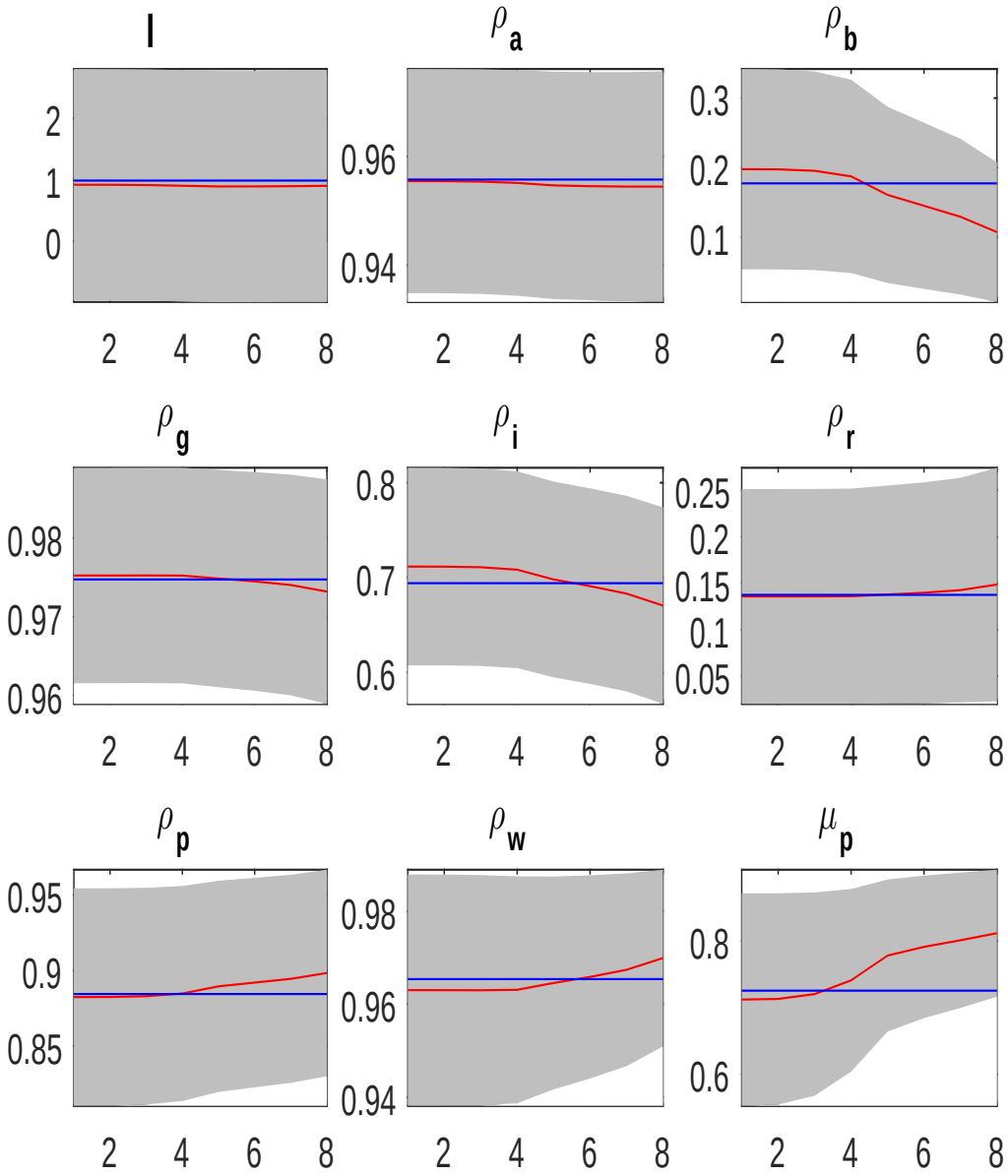


Figure 8: *Estimated mode of NN (blue) and IGIG (red) under different prior locations with measurement errors. Gray bands report the 90% confidence around the mode of specification IGIG (red). On the x axis the marginal likelihood of the IGIG specification. Black dot represents the estimated variant that maximizes the ML. The asterisk the* $IGIG(0.1, 10)$

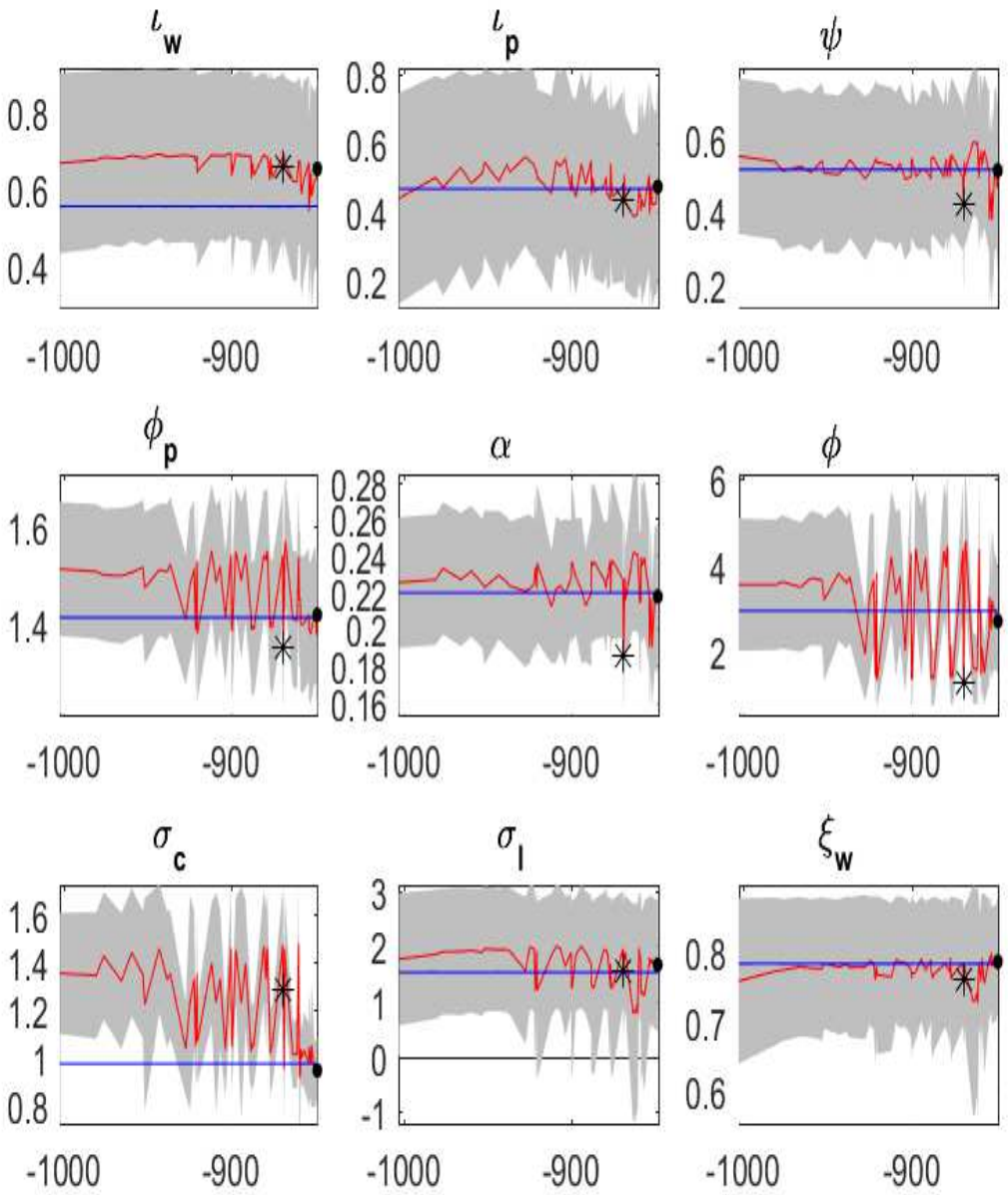


Figure 9: *Estimated mode of NN (blue) and IGIG (red) under different prior locations with measurement errors. Gray bands report the 90% confidence around the mode of specification IGIG (red). On the x axis the marginal likelihood of the IGIG specification. Black dot represents the estimated variant that maximizes the ML. The asterisk the* $IGIG(0.1, 10)$

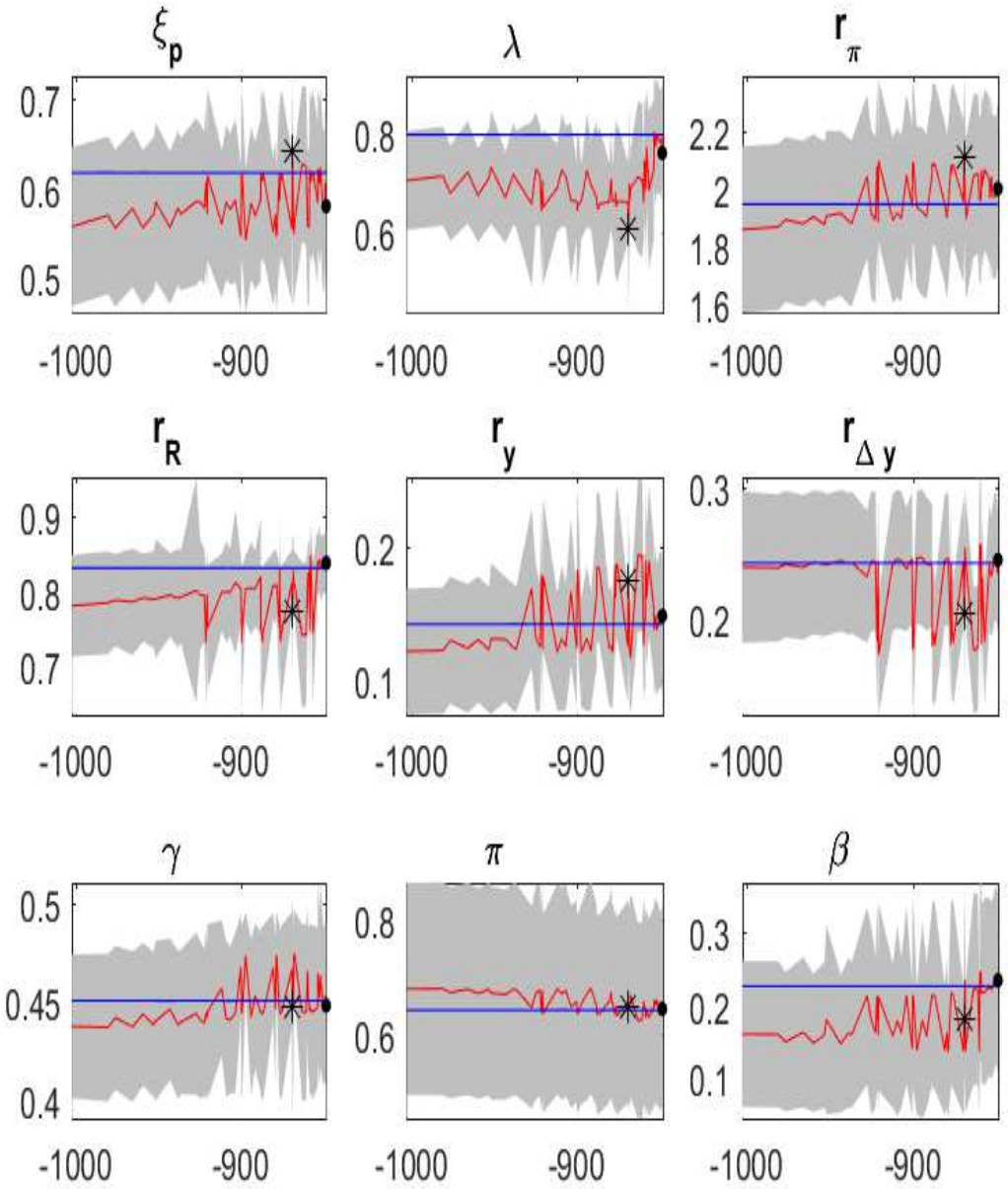


Figure 10: *Estimated mode of NN (blue) and IGIG (red) under different prior locations with measurement errors. Gray bands report the 90% confidence around the mode of specification IGIG (red). On the x axis the marginal likelihood of the IGIG specification. Black dot represents the estimated variant that maximizes the ML. The asterisk the* $IGIG(0.1, 10)$

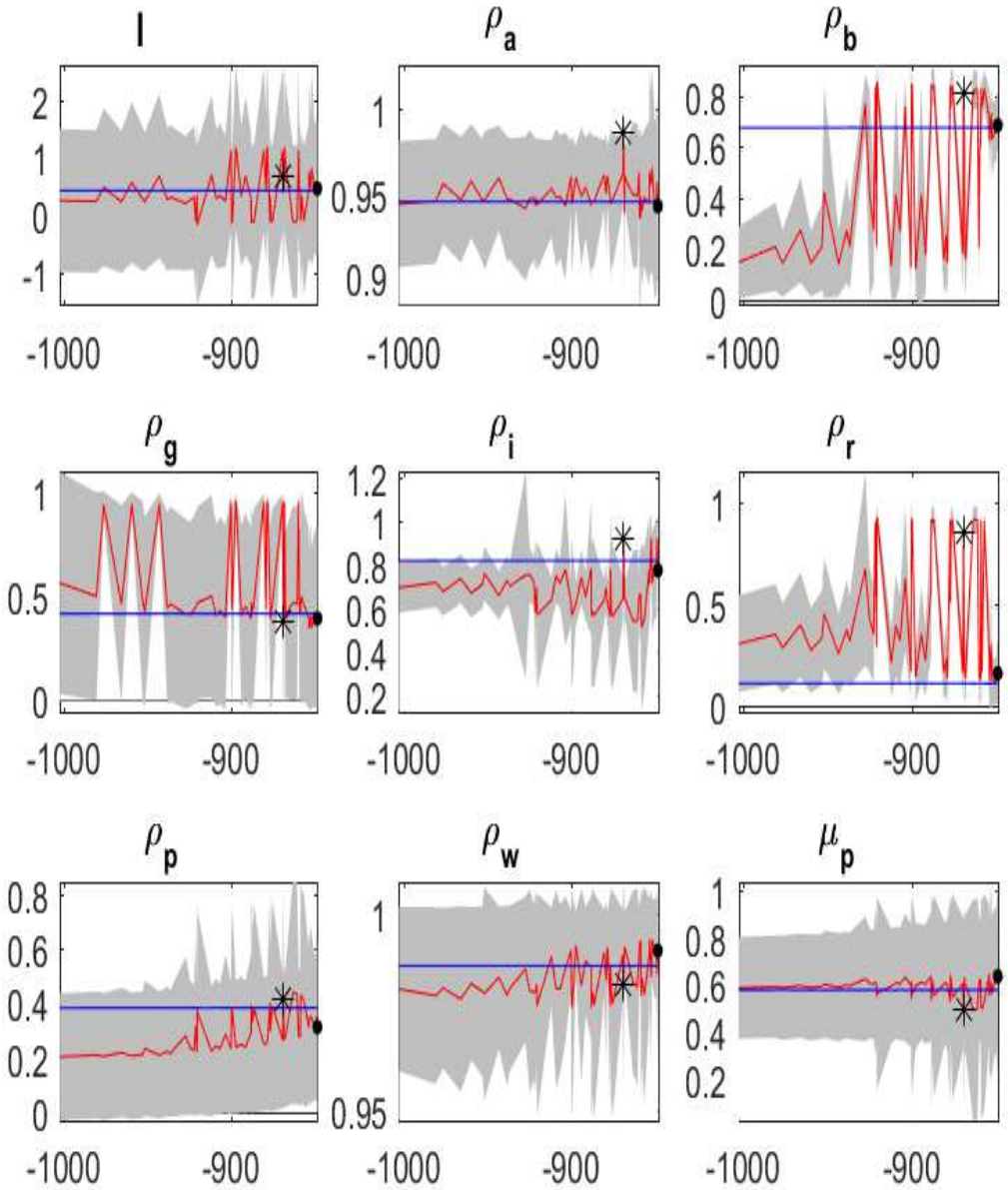


Figure 11: *Historical decomposition of output growth (dy) in terms of structural shocks in the IG case.*

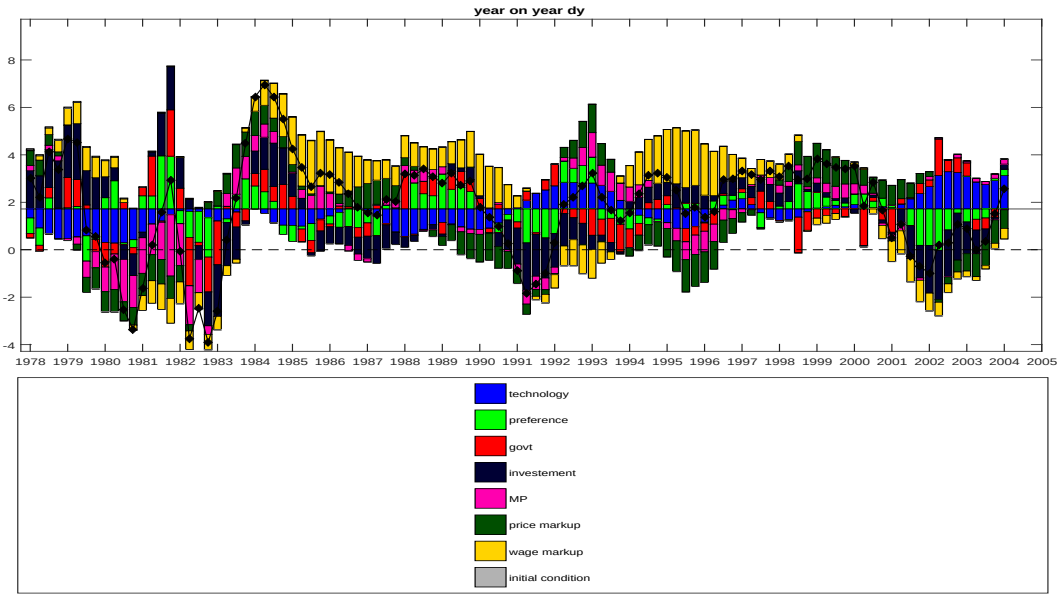
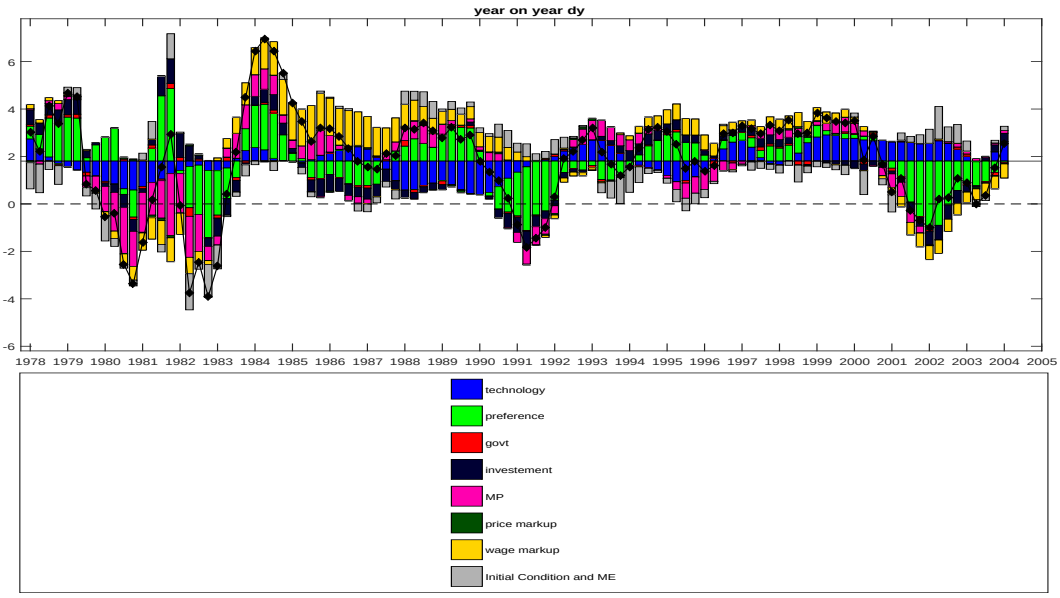


Figure 12: *Historical decomposition of output growth (dy) in terms of structural shocks in the NN case.*



References

Cúrdia, V. and Reis, R. (2010). Correlated Disturbances and U.S. Business Cycles. NBER Working Papers 15774, National Bureau of Economic Research, Inc.

Díaz-García, J. A. and Gutiérrez-Jáimez, R. (1997). Proof of the conjectures of H. Uhlig on the singular multivaraite beta and the Jacobian of a certain tranformation. *Annals of Statistics*, 25:2018–2033.

Díaz-García, J. A. and Gutiérrez-Jáimez, R. (2006). Distribution of the generalized inverse of a random matrix and its applications. *Journal of Statistical Planning and Inference*, 136:183–192.

Frühwirth-Schnatter, S. (2004). *Efficient Bayesian Parameter Estimation.* State Space and Unobserved Component Models: Theory and Applications. Cambridge University.

Frühwirth-Schnatter, S. and Wagner, H. (2010). Stochastic Model Specification Search for Gaussian and Partial non-Gaussian State Space Models. *Journal of Econometrics*, 154:85–100.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* Second Edition (Chapman and Hall/CRC Texts in Statistical Science), New York, USA.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient Parameterizations for Normal Linear Mixed Models. *Biometrika*, 82:479–488.

George, E. I. and McCulloch, R. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, 88:881–889.

George, E. I. and McCulloch, R. (1997). Approaches for Bayesian Variable Selection. *Statistica Sinica*, 7:339–373.

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics.* Wiley, New York, USA.

Grassi, S. and Proietti, T. (2014). Characterizing Economic Trends by Bayesian Stochastic Model Specification Search. *Computational Statistics and Data Analysis*, 71:359–374.

Hall, J., Pitt, M. K., and Kohn, R. (2014). Bayesian inference for nonlinear structural time series models. *Journal of Econometrics*, 179:99–111.

Harvey, A. C. (1989). *Forecasting, Structural Time Series and the Kalman Filter.* Cambridge University Press, Cambridge, UK.

Harvey, A. C. (2001). Testing in Unobserved Components Models. *Journal of Forecasting*, 20:1–19.

Iskrev, N. (2010). Local Identification in DSGE Models. *Journal of Monetary Economics*, 57:189–202.

Komunjer, I. and Ng, S. (2011). Dynamic Identification of Dynamic Stochastic General Equilibrium Models. *Econometrica*, 79:1995–2032.

Proietti, T. and Grassi, S. (2015). Stochastic Trends and Seasonality in Economic Time Series: New Evidence from Bayesian Stochastic Model Specification Search. *Empirical Economics*, 48:983–1011.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97:586–606.

Uhlig, H. (1994). On Singular Wishart and Singular Multivariate Beta Distributions. *Annals of Statistics*, 22:395–405.