

# Bayesian prior elicitation in DSGE models: macro vs micro priors

Giulio Nicoletti\*and Marco J.Lombardi†

January 5, 2009

PRELIMINARY AND INCOMPLETE PLEASE DO NOT  
QUOTE WITHOUT PERMISSION

## Abstract

DSGE models are commonly estimated by bayesian methods. Prior knowledge can either be information concerning deep parameters values or it can be information concerning some statistical moments of macroeconomic data. Del Negro and Schorfheide (2008) show that using these two sources of information can lead to non clear-cut conclusions concerning posterior estimates. We dig into the details of their result showing that model misspecification and nuisance parameters are to blame for that. The problem we highlight is general when using macroeconomic information as prior and it is not confined to the DS specific application.

## 1 Introduction

Since the seminal contributions by Sims and Zha (1999), Schorfheide (2000) and Smets and Wouters(2003), Bayesian estimation methods have gained ground as a very attractive alternative over classical methods in the field of dynamic stochastic general equilibrium models (DSGE), both among academicians and practitioners.

Within the class of DSGE models, prior knowledge can either be information concerning deep parameters values or it can be information concerning some statistical moments of the macroeconomic data. The former prior has to be elicited from unrelated microeconomic studies and therefore deep

---

\*Bank of Italy and UCL

†ECB

parameters are usually modeled as independent random variables. The latter prior has to be elicited from a pre-sample of data and it concerns some functions of the deep parameters such as some moments of simulated data or impulse response functions. Due to the mapping from the deep parameters to that function, a macro prior generally implies that parameters are correlated. A macro prior is then characterized by two key ingredients, a correlation structure of parameters and an informational content of the pre-sample.

Del Negro and Schorfheide -DS heretofore- (2008) apply a macro prior to the subset of parameters concerning the exogenous states in a Smets and Wouters type of model. They use information from moments of a presample and find that posteriors distributions change with respect to the case when micro priors only are used. Moreover their estimates are sensitive to the prior setting of those parameters which are *not* informed by the macro prior. Depending upon how those latter are set, two posterior distribution emerge, either one with both high nominal and wage rigidities or one where both rigidities are low. They also found that computing the marginal likelihood turns out not to be useful in assessing which mode is more worth of trust.

This paper contributes to the literature on prior elicitation and DSGE models, by further investigating why the macro and the micro elicitation approach can deliver such different results. The DS result can be due to the informational content of the presample, say a structural break in the moments of the data, for example after and before the Great Moderation, or, more plainly, it can be related to the prior correlation structure. In turn this latter might either relate to a weak identification problem or to misspecification of the DSGE model. In this paper we aim at disentangling the effects of each of those issues above mentioned. We first discuss macro priors within DSGE models and we show that they are in general characterized by the presence of nuisance parameters which cannot be integrated out of the prior and therefore they must be fixed at some predetermined value. We then introduce a simple DSGE model, one of those used in Rabanal and Rubio-Ramirez (2005), and we estimate it by a benchmark micro-prior and by the DS prior. We reproduce the gist of Del Negro and Schorfheide's findings: deep parameter posterior estimates dramatically differ when either macro or micro priors are employed. Moreover results hinge upon how nuisance parameters are set, due to the correlation structure of parameters in the likelihood. We show that the informational content of the presample can be relevant in shaping posterior distributions when the DSGE model is misspecified, although it is the main driver for the DS result.

In order to be able to fully disentangle the effects of the presample information from that of prior correlation, we introduce priors which embed a

meaningful correlation structure but their mode can be elicited in the same way as for micro priors, i.e. without referring to a pre-sample. We obtain this feature by the means of a prior based on the Generalized Likelihood Uncertainty Estimation approach discussed by Saltelli et al. (2008) within the framework of Global Sensitivity Analysis. We use this prior for different subsets of parameters and we show how posterior estimates change in each case, due to correlation between deep parameters. Since it is not clear whether the different posterior estimates are due to a weak identification problem or it is due to misspecification we assess the two hypothesis we recur to simulated data. This controls for misspecification, since a weak identification problems related to the tranformation from structural to reduced form of the model should be detected on both simulated and real data. We simulate our sample by calibrating the parameters of the model at the mode which was found by benchmark priors and running MCMC estimation by adopting in turn correlated and uncorrelated priors. We conclude that model misspecification is mostly to blame for the instability of estimates to the change of the priors.

The paper is structured as follows, in section 2 we define macro priors more properly; in section 3 we introduce our model and in 3.1 we describe our benchmark and the DS prior; results follow. We then introduce our generalized prior in section 6 with results. A discussion concerning misspecification vs weak identification follows in section 7 and some conclusions follow.

## 2 Macro Priors

In this section we provide the reader with a more precise definition of what we have called so far as macro priors. We define macro priors as distributions over a set of deep parameters for which the following holds:

1. The prior mode is elicited from information concerning some macroeconomic indicator such as moments of the data or impulse response functions.
2. The prior over the deep parameters embeds a correlation structure which is related to the inverse mapping from the macroeconomic indicator to the deep parameters.
3. There is no uncertainty concerning the target value of the indicator.

While the first point is rather trivial and it corresponds to what it should be expected from the use of a presample in order to tune the prior, the second and third are more substantial and they deserve some further comments.

The second point is what characterizes the macro approach: the prior distribution is formed over a joint set of parameters and it entails a correlation structure induced by the mapping between deep parameters and the macro indicator which is chosen. The correlation can be further decomposed in two parts, the mapping between structural and reduced form of the model, i.e. a macro prior includes the set of cross equation restrictions over the data generating process implied by the DSGE. That is common to all macroeconomic indicators, and a second transformation from the reduced form to the specific indicator at hand.

Take a structural DSGE model of the general form:

$$A_1(\psi)E_t X_{t+1} + A_2(\psi)X_t + A_3(\psi)X_{t-1} + B_4(\psi)\epsilon_t = 0, \quad (1)$$

$X_t$  gathers both endogenous and exogenous state variables.  $A$ s and  $B$ s are matrices of coefficients,  $E_t$  is the conditional expectations operator and  $\epsilon_t$  is the vector of innovations, which we take as normally distributed and with a unit variance since the variance-covariance matrix is taken up by the  $B_4$  term. The model is then solved to get its reduced form:

$$x_t = F(\psi)x_{t-1} + G(\psi)\epsilon_t, x_t \equiv [y_t; z_t],$$

where  $x$  is the vector of states, including for simplicity observables  $y$  and unobservable exogenous states  $z_t$ . All indicators such as moments or impulse response functions are constructed from the reduced form, therefore all priors constructed by targeting one of these indicators are likely to display some common correlation structure.

In general not all parameters are informed with a macro prior for two main reasons. First, for some of them there is available some microeconomic information which should be taken into account. Second, the correlation structure in the likelihood function is known to be somewhat ill-behaved when all parameters are considered (Iskrev (2008)). Nevertheless, when only a subvector  $\psi_1$  is informed with a macro prior, the other  $\psi_2$  must be treated as a nuisance parameter, i.e. it should be integrated out of the prior distribution. This procedure is not viable in a DSGE where an analytical form of the likelihood is not available; therefore nuisances are to be kept fixed at some predefined level. When nuisance parameters are kept fixed it is not guaranteed that the correlation structure of the joint prior replicates the structure of correlation in the likelihood. In particular when they are fixed at some prior mode and this is distant from the likelihood mode, the researcher is likely to provide estimates of the  $\psi_2$  vector with an additional bias with respect to what she might want to take into the analysis. Building on this intuition we contrast the case when nuisance parameters are fixed as close as possible to

the mode found by a numerical optimizer to the case where they are fixed at some prior mode.

The third point makes it clear that a macro prior is different from a prior over moments or over some macroeconomic indicators. In this latter case the researcher should start from specifying a distribution over the moments and then transform it, by an inverse function theorem, into a distribution over the deep parameters. In order to do that the mapping from the deep parameters to the indicators must be injective, i.e. there must be a one-to-one mapping from the deep parameters to the set of indicators. This approach has been undertaken for example by Kociecki (2005) in the context of VAR analysis but, to our knowledge, it has not been undertaken for DSGE models, we leave it for future research.

In our macro prior there is no a priori uncertainty over the moments; moreover the number of indicators can be lower, equal or higher than the number of deep parameters. From this point of view macro prior is similar to a dummy prior approach where the sample is augmented by dummy observations simulated from a model which incorporates restrictions over macroeconomic indicators. The Del Negro and Schorfheide's approach is exactly a dummy prior in the sense that since it computes the likelihood of a sample which is augmented by simulated data and uses that as posterior distribution for the parameters.

### 3 Model application and data

We use the same model as one of those estimated in Rabanal and Rubio-Ramirez (2005). The reason for choosing a small scale model rather than a larger one is twofold. On the one hand DS(2008) already document the problem at hand by using a medium scale model à la Smets and Wouters. Taking that latter model or even a larger one would therefore not add value to our analysis, while it would make it more complicated and less transparent, as already seen in DS(2008). On the other hand we already know from Del Negro and Schorfheide (2004) that even larger models are misspecified, albeit probably less than smaller ones, but on top of that they have more parameters and this makes identification more difficult<sup>1</sup>. Overall we believe that the model we picked represents a good trade-off between transparency

---

<sup>1</sup>A different route could have been taken, by simulating data samples and allowing the model which generates the data to differ from the one used in the estimation. While this route is suitable in classical-gradient method optimization, it can be very time consuming in MCMC. Moreover in order to undertake meaningful simulations we should have a too strong view on how the misspecification is produced in reality

and realism of our analysis.

The (linearized) model is described by the following set of equations:

$$\begin{aligned}\frac{1}{\sigma}y_t &= \frac{1}{\sigma}E_t y_{t+1} - (r - E_t \pi_{t+1} + E_t g_{t+1} - g_t); \\ \pi_t &= \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa_{pp}(mc_t + e_p); \\ r &= \rho_r r_{t-1} + (1 - \rho_r)(\gamma_\pi \pi_t + \gamma_y y_t) + e_{ms}; \\ y_t &= a_t + (1 - \delta)n_t; \\ mc_t &= r w_t + n_t - y_t; \\ r w_t &= y_t \frac{1}{\sigma} + \gamma n_t - g_t;\end{aligned}$$

The first three equations are the Euler equation, the (backward looking) Phillip's curve and a standard Taylor rule. The backward looking component in the Phillip's curve is derived from the assumption that non updating producers partially index their prices as a function of past inflation. The remaining equations are more standard: the production function, the definition of marginal costs and the supply for labour. As common in the literature, some constants in the equations above are given by a non linear combinations of deep parameters, as follows:

$$\begin{aligned}\gamma_b &= \omega / (1 + \omega \beta) \\ \gamma_f &= (\beta / (1 + \omega \beta)); \\ \kappa_{pp} &= \frac{(1 - \delta)(1 - \theta_p \beta)(1 - \theta_p)}{(\theta_p(1 + \delta(\epsilon - 1)))(1 + \omega \beta)},\end{aligned}$$

$\omega$  is the degree of backward looking indexation in the Phillips curve,  $\beta$  is the discount factor,  $\theta_p$  is the probability of price adjustment in the Calvo model,  $\delta$  is the share of capital in the Cobb Douglas production function (0.36) and  $\epsilon$  is the elasticity of substitution among varieties in the bundle of commodities.

There are four shocks in this economy, technology  $a$ , intertemporal preferences  $g$ , monetary  $e_{ms}$  and mark up shocks  $e_p$ ; the first two autoregressive processes, while the latter are i.i.d.

$$\begin{aligned}a_t &= \rho_a a_{t-1} + e_a, \\ g_t &= \rho_g g_{t-1} + e_g.\end{aligned}$$

Some parameters which are known to be difficult to pin down, i.e. either because not separately identifiable from other model parameters or are hard

to estimate with detrended data, are calibrated at values which are standard in the literature (see R-R 2005). The calibrated terms are given by  $\beta = 0.99, \delta = 0.36, \epsilon = 6$ . The vector of estimated parameters  $\psi$  is then given by:

$$\psi \equiv (\sigma, \theta_p, \gamma, \omega, \rho_r, \gamma_y, \gamma_\pi, \rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_\lambda),$$

where the exogenous states parameters are the last six terms.

As measurements we use US data on real wages, inflation, GDP growth rate and the fed funds rates (sample 1960:01-2001:04). Series on output, prices and wages come from the Bureau of Labor Statistics<sup>2</sup>. Since we do not include a common trend (stochastic or deterministic) in the estimation process, data have to be made stationary prior to estimation. We use inflation and output growth rates and we demean them, while we have used an Hodrick-Prescott filter on real wages.

Measurements, on the right hand side, and their model counterparts, on the left hand side, are given by:

$$\begin{aligned} \text{Real Wage}\% &= 100 r w_t \\ \text{Real Output Growth}\% &= 100(y - y(-1)), \\ \text{Annualized inflation rate} &= 400(\ln P_t - \ln P_{t-1}), \\ \text{Annualized interest rate} &= 400 r_t \end{aligned}$$

on the left hand side the observables and on the right their model counterpart.

### 3.1 Benchmark priors

As a benchmark we choose a prior in which all parameters are independent random variables. We inform the prior modes and the distributions by following Rabanal and Rubio-Ramirez (2005)<sup>3</sup>.

For numerical reasons, we adopt some transformations of the parameter space. Since both the elasticity of substitution  $\sigma$  and the Calvo probability  $\theta_p$  lie in a bounded space (i.e. between zero and one), we transform them in order for our algorithms to search in a larger space. We define:

$$\sigma^{-1} = 1/\sigma \quad ; \quad \Theta_p = 1 - 1/\theta_p,$$

this greatly improves the performance of the algorithm, by lowering the condition number of many of the matrices involved in the kalman filter. The

---

<sup>2</sup>These are: output for the non farm business sector and its deflator and hourly compensation for the nonfarm business sector

<sup>3</sup>Results do not change when more or less diffuse priors are adopted, provided they are independent random distributions

transformation at hand is convenient from an economic point of view since the transformed variables can be interpreted as, respectively, the risk aversion of agents and the average amount of time that lasts until a price is revised. As it is common practice in the literature we formalize our priors as function of the transformed parameters<sup>4</sup>

The prior mode on the elasticity of intertemporal substitution is set to one, the logarithmic utility case when substitution and income effect compensate, while the Calvo parameter is set in such a way that prices adjust on average every quarter, which is above the lower bound of the commonly accepted microeconomic estimates (see ). The elasticity of labour is also set equal to one. We choose a uniform distribution between zero and one for the degree of price indexation in the economy. The coefficients on the Taylor rule are informed on the basis of previous studies, the  $\gamma_\pi = 1.5$  being the reference value since the original work by John Taylor (Taylor 1993). More details concerning the shape and some statistics of the benchmark priors are reported in the appendix.

For the exogenous parameters block we use uniform distributions.<sup>5</sup> We use a slightly larger support for what concerns the cost push shock, which is known to contribute to inflation dynamics with quite large movements. Overall, the chosen shapes of the prior distributions are consistent with those in the paper by Rabanal and Rubio-Ramirez (2005).

### 3.2 The DS prior

The kernel of the Del Negro and Schorfheide's prior corresponds to the likelihood of the VAR which (approximately) embeds the restrictions implied by the DSGE model<sup>6</sup>. The VAR approximation of the DSGE is denoted as follows:

$$Y_t = \Phi_1(\psi)Y_{t-1} + \Phi_2(\psi)Y_{t-2} + \dots + \Phi_j(\psi)Y_{t-j} + \epsilon_t,$$

or in matrix notation

$$Y_t = \Phi(\psi)X_t + \epsilon_t,$$

---

<sup>4</sup>About the average duration of prices, there is also microeconomic evidence readily available, therefore it makes much sense to form a prior on the transformed variable.

<sup>5</sup>Switching to an inverted gamma prior for the variances of shocks does not alter results in a significant sense

<sup>6</sup>DSGEs do not always have a finite VAR representation; a simple RBC model estimated with capital as unobserved variable has a finite VARMA structure, but no exact VAR one. To check the quality of the approximation we simulated datasets from an RBC model and we estimated both the RBC and the approximated VAR by likelihood methods. The difference in the sampling distributions seems to be rather small. Results are available upon request.

with  $\Phi(\psi) \equiv \Phi_1(\psi) \dots \Phi_j(\psi)$  and  $X_t \equiv [Y_{t-1} Y_{t-j}]$ .

The  $\Phi$ s are numerically constructed using the OLS regression coefficients formulas from DSGE-simulated data moments. The  $\Phi$ s depend upon the whole vector of structural parameters  $\psi$ . The likelihood of the VAR approximation of the DSGE is easily expressed as a function of few data moments, conditional on the values of deep parameters:

$$L(Y, X | \psi) = |\Sigma(\psi)|^{-(T^*+n-1)} \exp\left(\frac{T^*}{2} [\Sigma(\psi)]^{-1} (\Gamma_{YY} - 2\Phi(\psi)\Gamma_{XY} + \Phi(\psi)'\Gamma_{XX}\Phi(\psi)')\right), \quad (2)$$

where  $\Sigma(\psi)$  is the theoretical variance covariance matrix of the data implied by the DSGE,  $T^*$  is a scale factor for the strength of the prior.  $\Gamma$ s are sample data moments (a pre-sample is used) with  $Y$  as the dependent variables (in our case wages, output growth, inflation and interest rates) and  $X$  are lagged data.

The rationale underlying the DS prior stems from a dummy prior approach where the researcher constructs an augmented sample which combines the original data with some ‘dummy’ observations which have been simulated from the model. The likelihood of the augmented sample corresponds then to the posterior distribution of parameters. By approximating the DSGE by a VAR, DS are able to tune their dummy prior directly on moments of the data without the need of calibrating deep parameters.

Following DS, we partition the vector of parameters  $\psi$  in two,

$$\psi \equiv [\psi_{endo}, \psi_{exo}],$$

which are respectively the parameters related to exogenous states and the other structural parameters. The DS prior is applied to the subvector  $\psi_{exo}$  only, while we assume that sufficient micro information is available for  $\psi_{endo}$ , which is then informed by the benchmark prior  $p^{micro}$ . Assuming independence across the two subvectors, the prior over the whole set of parameter is then given by:

$$p(\psi) = p^{micro}(\psi_{endo})L(Y, X | \psi_{exo}, \psi_{endo}^-).$$

When nuisance, the  $\psi_{endo}$  is fixed at some level  $\psi_{endo}^-$ , since the VAR likelihood cannot be integrated out of the prior, at least not before an MCMC has run and converged.

The table below reports some quantiles and the mean of the DS prior. They were estimated by sampling parameters from a uniform distribution and computing the likelihood of the VAR as described in DS(2008). Those weights are then used in a weighted kernel density estimation to describe the

distribution of the prior<sup>7</sup>. The prior is computed for two different values of what turn out to be the main nuisance parameters in our analysis.

| Parameters ( $\sigma^{-1}, \Theta_p$ at 2) | 0.01 perc. | Mean   | 0.99 perc. |
|--|------------|--------|------------|
| $\rho_a$                                   | 0.6464     | 0.8284 | 0.9856     |
| $\rho_g$                                   | 0.6051     | 0.7954 | 0.9691     |
| $\sigma_a$                                 | 0          | 0.0515 | 0.2176     |
| $\sigma_m$                                 | 0          | 0.1411 | 0.2944     |
| $\sigma_p$                                 | 0.0408     | 0.2326 | 0.3860     |
| $\sigma_g$                                 | 0          | 0.0504 | 0.2633     |
| Parameters ( $\sigma^{-1}, \Theta_p$ at 5) | 0.01 perc. | Mean   | 0.99 perc. |
| $\rho_a$                                   | 0.5640     | 0.7503 | 0.9223     |
| $\rho_g$                                   | 0.5927     | 0.7933 | 0.9367     |
| $\sigma_a$                                 | 0          | 0.0606 | 0.2030     |
| $\sigma_m$                                 | 0          | 0.0606 | 0.2172     |
| $\sigma_p$                                 | 0          | 0.1197 | 0.2554     |
| $\sigma_g$                                 | 0.0321     | 0.2172 | 0.3739     |

It is interesting to note that variances change widely when different values of the nuisance parameters are selected, moreover the persistence of the technology shock changes, while the autoregressive parameter of the preference shock is almost not affected. This is due to the fact that our set of measurements do not provide enough information to well identify the technology shock, which turns out to be very sensitive to the assumed parameters. While the preference shock is not sensitive to changes in its persistence parameter we found that the smoothed technology shock turns out to be heavily influenced by changes in both its own parameter and changes in the persistence parameter of the other shock. We compared the smoothed preference and technology shock obtained by calibrating all parameters and just changing the persistence parameter of the exogenous state. As a simple example, clearly not meant to be conclusive evidence, we contrast below the smoothed states when  $(\rho_a, \rho_g) = (0.8, 0.8)$  against the case  $(\rho_a, \rho_g) = (0.833, 0.9)$ , all other parameters being the same. The second set of parameter values is closer to the mode which we discuss at a later stage, in section 4.1.

This exercise was undertaken on our data sample and cross checked on simulated data, so that the misspecification of the DSGE model is not an issue.

---

<sup>7</sup>Due to the kernel estimation numerical method (normal kernel) in some cases we obtained negative values for the lower quantile of some variances; those are approximated as zeros.

## 4 Estimation

In a Bayesian framework the prior information is combined with the likelihood. To form that latter we write a state space by using the reduced form of the model as transition equation. To form the likelihood of the model we use the Kalman filter innovations. Its precise implementation follows the description in Durbin and Koopman (2003). The filtering equations are iterated until convergence of the variance covariance matrix of the states is reached and then the steady state kalman gain and variance are used. We form projections of the states  $\hat{x}_{t|t}$  contingent to observables available at time  $t$ . The projection of the states at time  $t$ , after information is updated, is

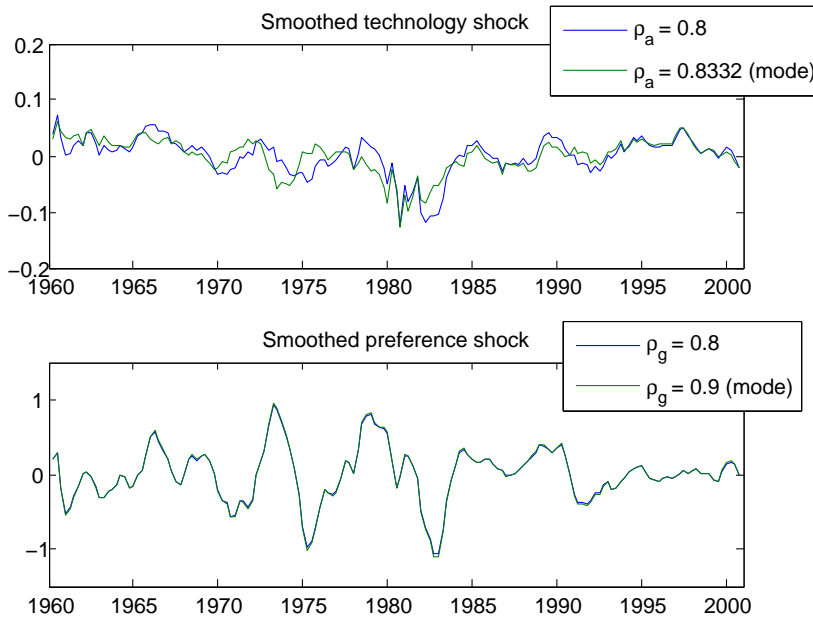


Figure 1: Shocks to technology and preference

given by:

$$\hat{x}_{t|t} = Fx_{t-1|t-1} + K\nu_t, \quad \nu_t = y_t - y_{t|t-1}.$$

The kalman gain  $K$  is a matrix of weights such as to provide the best linear projection of the unobservable states (i.e. the shocks) over the space of innovations  $\nu_t$ .

Since an analytic form of the posterior distribution cannot be obtained, we use Monte Carlo Markov Chains methods to draw from it<sup>8</sup>. MCMC enable the researcher to draw from an unknown target distribution which can be evaluated numerically for single draws only up to a normalizing constant. As it is common in the literature, see An and Schorfheide (2007), we use a random walk Metropolis Hastings algorithm. The estimation procedure can then be summarized as follows:

1. A numerical optimizer is used in order to find the posterior mode. We used the `csminwell.m` program by Chris Sims. To make estimations comparable this step is run only once with benchmark priors and the estimates are used to run estimation for all different priors.
2. The inverted hessian of the posterior  $H$  is computed at the mode and it is decomposed via eigenvalue-eigenvector decomposition to form a matrix  $D$ :  $H = D'\Lambda D$ . In DSGE models there are off-diagonal terms in the information matrix and this procedure essentially rotates the parameter space in order to decorrelate the parameters. This way of doing, described in Roberts and ..., greatly improves the efficiency of MCMC.
3. A candidate vector of parameters is drawn as according to a random walk with normally distributed innovations:

$$\theta^* = \theta_i + \mu\sqrt{\Lambda}D\epsilon_t,$$

the  $D$  matrix provides a direction of search for the MH algorithm and  $\lambda$  is a general scale parameter, which we use to tune the acceptance rate of the chain. To make estimation with different priors numerically comparable, each chain is started at the same  $\theta_0$ , which is the one numerically found in step one by benchmark priors.

---

<sup>8</sup>An alternative would be to use importance sampling, drawing from some distribution  $f$  and using the weights  $P/f$ , where  $P$  is the posterior kernel, to adapt the distribution. This method has been explored in An and Schorfheide, and it has been shown to be not superior to MCMC

4. For each draw the (log) prior is computed and the (log)likelihood is computed by the means of the Kalman Filter. They are combined to get a posterior draw.
5. A draw is accepted with a probability given by posterior odds:

$$\theta_{i+1} = \theta^* \text{ if ratio} = \frac{P(\theta^*)}{P(\theta_i)} > \text{rand}(U(0,1)),$$

where  $P$  is the posterior kernel (log-likelihood plus log-prior) and  $\text{rand}$  is a uniformly distributed random variable. The MH always accepts draws which provide an higher posterior kernel with respect to the actual one, but it can accept draws that do not improve upon the previous ones with a probability lower than one. By iterating the chain over time it approximates the posterior marginal distribution of the parameters. The posterior appears only as a ratio, the researcher might know it only up to a constant, as in our case.

6. The algorithm is run in one chain for 200.000 draws, the  $\lambda$  is calibrated in order to keep the acceptance rate between 20 and 40 % as common in the literature.

We assessed convergence by inspecting CUMSUM statistics which suggested us to select the last 50.000 draws of the chains to draw inference upon the parameters (see the appendix).

## 4.1 Results from the numerical mode finder

As first step we get the following numerical mode by using benchmark priors. To make the experiments comparable we initialized all the algorithms at this same set of values and we applied to it a small random perturbation, proportional to the scale parameter  $\mu$ .

| Parameters    | Numerical posterior mode |
|---------------|--------------------------|
| $\sigma^{-1}$ | 8.5361                   |
| $\Theta_p$    | 7.2722                   |
| $\gamma$      | 0.3735                   |
| $\omega$      | 0.7689                   |
| $\rho_r$      | 0.3943                   |
| $\gamma_y$    | 0.0415                   |
| $\gamma_\pi$  | 1.0901                   |
| $\rho_a$      | 0.8332                   |
| $\rho_g$      | 0.9001                   |
| $\sigma_a$    | 0.0148                   |
| $\sigma_{ms}$ | 0.0047                   |
| $\sigma_p$    | 0.7501                   |
| $\sigma_g$    | 0.1563                   |

## 5 Results from Metropolis-Hastings: Benchmark vs DS

Adopting the benchmark priors we get the following posterior distributions for the parameters:

| Parameters    | 0.01 perc. | 0.99 perc. | Posterior mean |
|---------------|------------|------------|----------------|
| $\sigma^{-1}$ | 4.6972     | 17.7151    | 9.8850         |
| $\Theta_p$ )  | 5.5936     | 10.1753    | 7.9071         |
| $\gamma$      | 0.2077     | 0.6842     | 0.3630         |
| $\omega$      | 0.6456     | 0.9082     | 0.7701         |
| $\rho_r$      | 0.2777     | 0.4926     | 0.3950         |
| $\gamma_y$    | 0.0115     | 0.0724     | 0.0408         |
| $\gamma_\pi$  | 1.0029     | 1.2097     | 1.0890         |
| $\rho_a$      | 0.7543     | 0.9036     | 0.8329         |
| $\rho_g$      | 0.8409     | 0.9424     | 0.8961         |
| $\sigma_a$    | 0.0085     | 0.0258     | 0.0165         |
| $\sigma_m$    | 0.0040     | 0.0057     | 0.0047         |
| $\sigma_p$    | 0.4454     | 1.4693     | 0.9167         |
| $\sigma_g$    | 0.0874     | 0.3300     | 0.1818         |

With the exception of the intertemporal elasticity of substitution estimates are quite close to the numerical mode. A low parameter value makes marginal costs less sensitive to variations in output and therefore it can be seen as a substitute mechanism to price rigidity. The degree of inflation

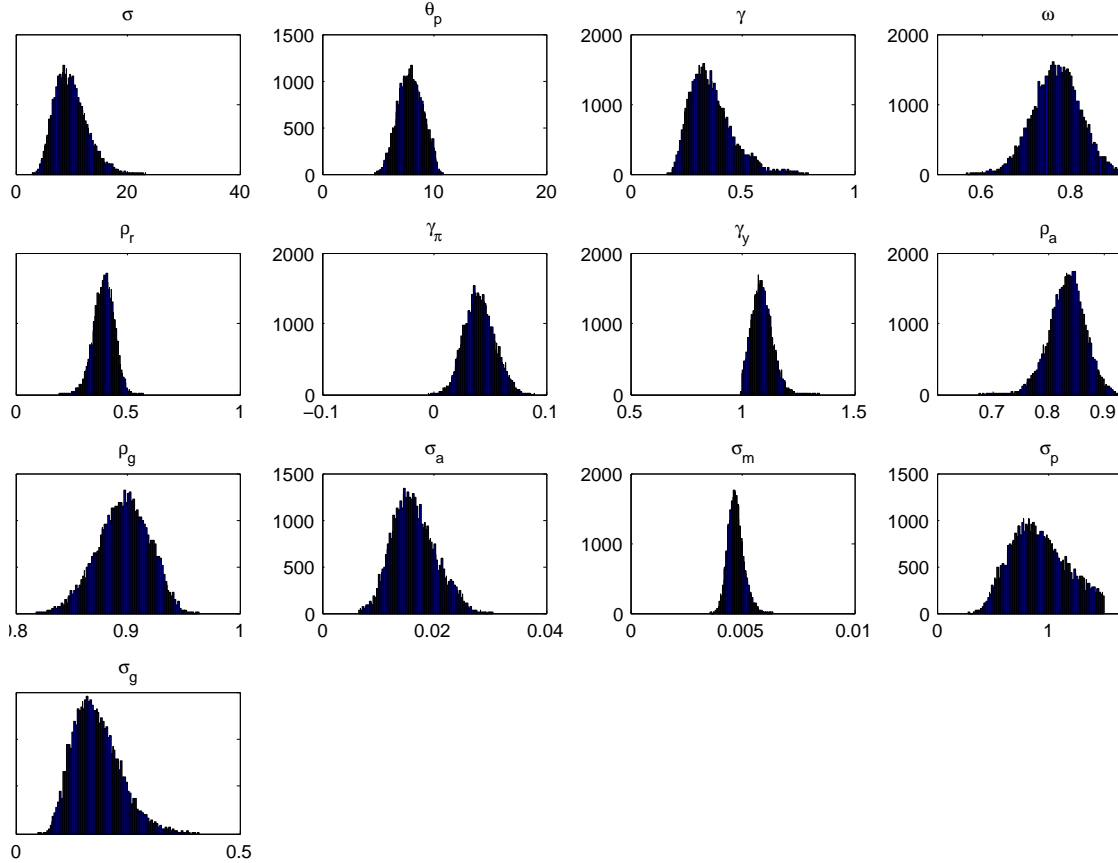


Figure 2: Posterior distribution for benchmark prior, histograms.

indexation turns out to be quite high;  $\omega = 0.76$ , which might be a figure overstated by the lack of rigid wages in the model.

In the subsequent part we contrast the results obtained with the benchmark prior with the ones obtained by using the DS (2008) prior. We apply the DS prior to the subvector of parameters  $\psi_{exo}$  concerning the exogenous states. Concerning all the other parameters we used our benchmark priors.

We implemented the DS prior by cutting the sample in two parts, a pre-sample of around forty observations, leaving the remaining for estimation<sup>9</sup>. We chose to fix the scale parameter  $T^*$  at 6, following DS(2008). Since they use a different number of observations in the presample we checked that for  $T^*$  in the range between 4 and 8 results do not change.

In exposing our results, we will restrict the attention of the reader to the two nuisance parameters which turn out to be the most relevant in twisting the estimates, even if several checks over all the vector of nuisance parameters were undertaken. We consider two different settings of the nuisance

<sup>9</sup>Results are not very sensitive to this choice, we also tried different splits without seeing much qualitative difference, albeit a little quantitative one

parameters:  $\sigma^{-1} = 2, \Theta_p = 2$  against the case  $\sigma^{-1} = 5, \Theta_p = 5$ . In the first case nuisance parameters are set closer to their prior mode, as implicitly suggested by the paper of DS, while in the second case they are set closer to the numerical mode<sup>10</sup>.

|  |            |            |                |
|--|------------|------------|----------------|
| Parameters ( $\sigma^{-1}, \Theta_p$ at 2) | 0.01 perc. | 0.99 perc. | Posterior Mean |
| $\sigma^{-1}$                              | 10.3426    | 23.6727    | 15.7359        |
| $\Theta_p$                                 | 3.1205     | 4.9284     | 3.9258         |
| $\gamma$                                   | 0.4376     | 0.8156     | 0.6270         |
| $\omega$                                   | 0.7361     | 0.9786     | 0.8935         |
| $\rho_r$                                   | 0.3380     | 0.5534     | 0.4539         |
| $\gamma_y$                                 | 0.0594     | 0.1315     | 0.0930         |
| $\gamma_\pi$                               | 1.0176     | 1.2788     | 1.1297         |
| $\rho_a$                                   | 0.6856     | 0.8270     | 0.7601         |
| $\rho_g$                                   | 0.8435     | 0.8928     | 0.8724         |
| $\sigma_a$                                 | 0.0080     | 0.0149     | 0.0109         |
| $\sigma_m$                                 | 0.0040     | 0.0057     | 0.0048         |
| $\sigma_p$                                 | 0.1438     | 0.3621     | 0.2282         |
| $\sigma_g$                                 | 0.2015     | 0.4277     | 0.2993         |
| Parameters ( $\sigma^{-1}, \Theta_p$ at 5) | 0.01 perc. | 0.99 perc. | Posterior mean |
| $\sigma^{-1}$                              | 4.7907     | 11.6364    | 7.3624         |
| $\Theta_p$                                 | 8.1756     | 10.1361    | 9.1578         |
| $\gamma$                                   | 0.2786     | 0.5678     | 0.4165         |
| $\omega$                                   | 0.6984     | 0.9504     | 0.8188         |
| $\rho_r$                                   | 0.0862     | 0.3692     | 0.2263         |
| $\gamma_y$                                 | 0.0264     | 0.1202     | 0.0740         |
| $\gamma_\pi$                               | 1.0019     | 1.1973     | 1.0908         |
| $\rho_a$                                   | 0.6520     | 0.9066     | 0.8142         |
| $\rho_g$                                   | 0.8087     | 0.9217     | 0.8646         |
| $\sigma_a$                                 | 0.0085     | 0.0173     | 0.0119         |
| $\sigma_m$                                 | 0.0060     | 0.0082     | 0.0073         |
| $\sigma_p$                                 | 1.0877     | 1.4887     | 1.3153         |
| $\sigma_g$                                 | 0.0962     | 0.2134     | 0.1401         |

The parameter estimates which are mainly affected by the change in nuisance parameters are the intertemporal elasticity of substitution, the elasticity of labour supply, the duration of the Calvo pricing scheme, the indexation to past inflation and the size of cost push shocks. Going from the first to the second case the size of the cost push shock increases and it reaches values

<sup>10</sup>Even setting them at the prior mode did not change results by much

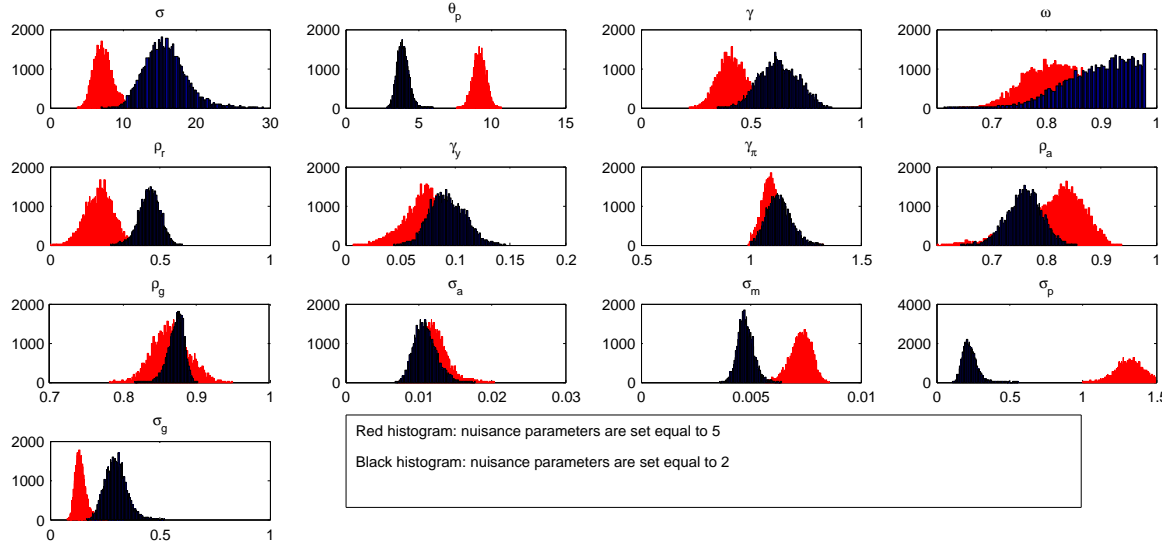


Figure 3: Posterior distribution for DS prior.

which are closer to the benchmark, while all other estimates change accordingly.

This can be better understood by looking at how the Phillips curve is specified:

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \frac{(1 - \delta)(1 - \theta_p \beta)(1 - \theta_p)}{(\theta_p(1 + \delta(\epsilon - 1)))(1 + \omega \beta)} (m c_t + e_p),$$

the impact of the cost push shock  $e_p$  is loaded into inflation  $\pi_t$  by a term which is a function of the Calvo probability of adjustment ( $\theta_p$ ); the higher the probability, the lower the impact of the mark-up shock, this needs a higher variance of the exogenous shock. Having a lower rigidity on the nominal side, the role of indexation should rise, while the sensitivity of real costs rise in order to compensate.<sup>11</sup>

## 5.1 Checking the model

A different but related issue is to which extent a model with dummy priors on moments is more effective than a model with benchmark priors in reproducing the moments of the data. Even when a model is misspecified a QML

<sup>11</sup>To check about the amount of complementarity in the parameters we computed principal components of the hessian of the likelihood; three factors explain most of information matrix.

procedure will consistently estimate those moments which are directly targeted (see Mittelhammer et al 2003 page []). We check that issue by drawing 10,000 draws from the posterior distributions obtained under benchmark and under the two different DS(2008) priors and compute the average variance-covariance matrices and we compare them with those of the data.

[TO BE ADDED]

## 6 Generalized Likelihood Priors

In this section we introduce a non parametric prior which draws upon a method known in the Global Sensitivity Analysis literature as Generalized Likelihood Uncertainty Approach (GLUE), described in Saltelli et al (2004)<sup>12</sup>. GLUE is an approximated method which allows the researcher to quickly define probability measures over events starting from a measure of fit or a loss function. Start from a function  $g$  of the deep parameter vector  $g(\psi)$ , i.e. some moments or impulse response function, define some target value of the parameters  $\psi^*$  which will correspond to the mode of the prior distribution. Define a loss  $L(g(\psi), g(\psi^*))$ . Then the a measure over  $\psi$  can be described by the following weights for a sample  $\psi^i$ :

$$Pr(\psi^i) = \frac{1}{L(g(\psi^i), g(\psi^*)^K)},$$

where  $K$  measures the strength of the prior. This can be normalized such as it sums to one.

In the case of normal shocks the likelihood corresponds to a particular GLUE where the penalty function is given by the standard deviation of the residuals errors. With this respect the approach is called Generalized Likelihood.

For our application we choose impulse response functions as an indicator. The loss is given by a distance between a target and the impulse response. While this approach can be further generalized by taking impulse response function by previous VAR studies, in this paper we select target deep parameters in the same way as it is done in a micro prior approach and leave the latter for further research. This fulfils two of the requirements concerning macro priors: it is a correlated prior, since it embeds a correlation structure which is due to the mapping from the reduced to the structural form, i.e. this is the same as in the likelihood (see Iskrev (2008)); moreover there is no uncertainty over the target, our point 3.

---

<sup>12</sup>See page 184 and following ones

Given a vector of parameters  $\psi$ , the corresponding impulse response from the model is denoted  $g_j(\psi)$ , which is a  $n \times s$  dimension matrix,  $n$  being the number of variables of the impulse,  $s$  the number of shocks. Subscript  $j$  denotes the number of steps ahead of the IRF. In the remainder of our analysis we concentrate on  $j = 0$  and  $j = 1$  in order to reduce the non linear nature of our problem; this avoids the criticism by Canova and Sala (2008) that minimizing the distance of a very non linear object such as impulse responses can be problematic. Still, from a numerical perspective allowing  $j < 5$  did not change our results.

We then set a vector of ‘target’ deep parameters which will correspond to the mode of our prior: those are set exactly as in the micro prior framework. The distance between the target impulse and the current impulse is defined by:

$$d_j(\psi^1, \bar{\psi}^2) \equiv g_j(\psi^1, \bar{\psi}^2) - g_j(\psi^{1*}, \bar{\psi}^2),$$

as only  $\psi_1$  is informed with the non-parametric prior, the subvector  $\psi^2$  is kept fixed at some level  $\bar{\psi}^2$ , to be discussed in the results. We define the kernel of our prior distribution and the proper prior as weights given by:

$$w = \frac{1}{L(d_j(\psi))}, w^{norm} = \frac{w}{(w * e^T)},$$

where  $L$  is a function over  $j$ , for example a sup function or a quadratic one.  $w^{norm}$  is the normalized (probability) vector, as it sums to one.

Once the weights are computed they describe a joint distribution of the vector  $\psi_1$ , for a fixed value of  $\psi_2$ .

We compared estimation results obtained by using priors on three different blocks of parameters each time and informing the remaining parameters with the benchmark prior: parameters in monetary policy rule,  $\psi_1 \equiv [\gamma_\pi, \gamma_y]$ . Parameters non related to the exogenous shocks, with the exception of the intertemporal elasticity of substitution which we find it hard to identify when a joint prior is used<sup>13</sup>  $\psi_1 \equiv [\Theta_p, \gamma, \omega, \rho_r]$ . The exogenous variables block, as in DS:  $\psi_1 \equiv [\rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_\lambda]$ .

We set nuisance parameters at the mode found numerically in section 4.1: we do so in order to escape the extra distortion shown in previous sections. As distance function between the IRF  $d$  we use a quadratic function of the same type that is used in impulse response matching:

$$d = vec(g(\psi)) W vec(g(\psi)),$$

---

<sup>13</sup>In linear models this parameter can be difficult to identify, see the paper by An and Schorfheide (2007)

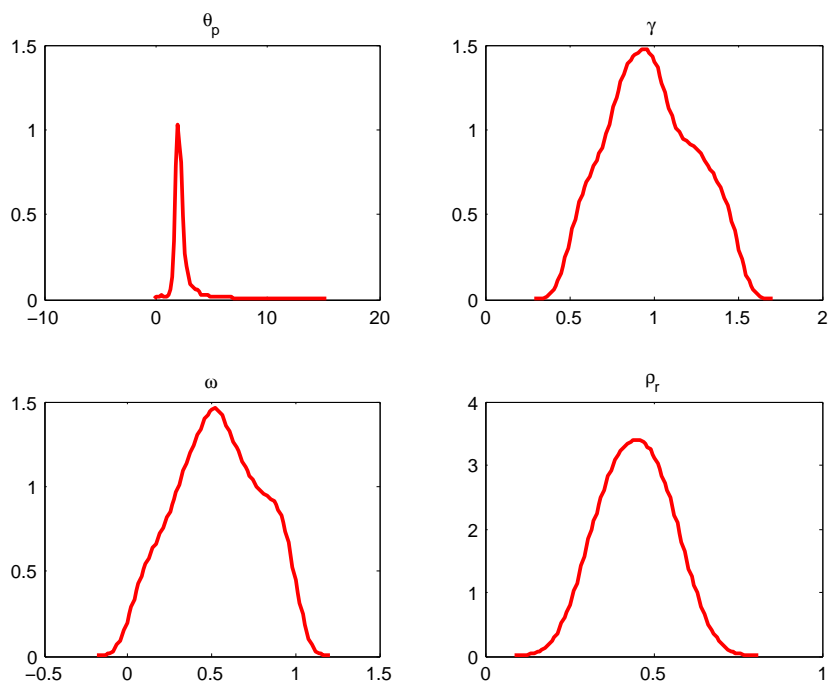


Figure 4: Block II. Univariate kernel density

where  $W$  is an  $8 \times 8$  identity matrix. We checked that a sup type of function delivers no different results. We used impulse on all the four observed variables and we matched two impulse points, the impact and the first impulse; no big difference emerges when more impulse points are taken into account.

With the first and third block of priors we found no substantial difference with respect to the benchmark case. To save space, results are therefore not shown, while a description of the prior is left to the appendix. For the second block of parameters we have the same prior modes as in the benchmark case:  $\Theta_p = 2, \gamma = 1, \omega = 0.5, \rho_r = 0.45$ . Means and quantiles are reported below:

| Parameters | 0.01 perc. | Prior Mean | 0.99 perc. |
|------------|------------|------------|------------|
| $\Theta_p$ | 0.41       | 2.07       | 13.03      |
| $\gamma$   | 0.49       | 0.96       | 1.51       |
| $\omega$   | 0.00       | 0.53       | 1.03       |
| $\rho_r$   | 0.20       | 0.46       | 0.68       |

with the kernel density plots:

## 6.1 Results from GLUE priors

When the second block of IRF priors is chosen estimates change somehow with respect to the benchmark case, mainly for those parameters regarding the standard deviation of exogenous shocks and the amount of rigidities needed to fit the data. The posterior of the mark-up shock hit hard on the prior boundary, forcing us to double the prior interval<sup>14</sup>.

| Parameters    | 0.01 perc. | 0.99 perc. | Posterior mean |
|---------------|------------|------------|----------------|
| $\sigma^{-1}$ | 4.5478     | 18.3013    | 9.7274         |
| $\Theta_p$ )  | 6.7766     | 12.0558    | 9.6645         |
| $\gamma$      | 0.1690     | 0.5210     | 0.2834         |
| $\omega$      | 0.6444     | 0.8792     | 0.7601         |
| $\rho_r$      | 0.2720     | 0.4933     | 0.3904         |
| $\gamma_y$    | 0.0052     | 0.0623     | 0.0311         |
| $\gamma_\pi$  | 1.0036     | 1.2049     | 1.0892         |
| $\rho_a$      | 0.7664     | 0.9087     | 0.8484         |
| $\rho_g$      | 0.8266     | 0.9476     | 0.8945         |
| $\sigma_a$    | 0.0107     | 0.0324     | 0.0212         |
| $\sigma_m$    | 0.0040     | 0.0056     | 0.0047         |
| $\sigma_p$    | 0.6507     | 1.9763     | 1.3461         |
| $\sigma_g$    | 0.0846     | 0.3333     | 0.1767         |

The result lies upon correlation among three key parameters: the elasticity of labour supply, the standard deviation of the mark-up shock and the Calvo parameter. There is a simple intuition for that. More rigid prices need stronger mark-up shocks in order to match inflation moments while a weaker elasticity of labour supply helps in keeping marginal costs low and it partially substitutes out the need for rigid wages. The result is similar to what it was found by DS(2008) in a more complex model; it is hard to disentangle between labor market features affecting the marginal costs from price rigidity behaviour.

## 7 Weak identification vs misspecification

In the preceding section we have shown that even after nuisance parameters are set close to the posterior mode rather than to the prior, estimates can be quite sensitive to a correlated prior. This depends upon which specific block of parameters is chosen to be informed with the prior. Moreover there

<sup>14</sup>As a check we rerun part of the previous estimation enlarging the prior, but previous result did not change substantially

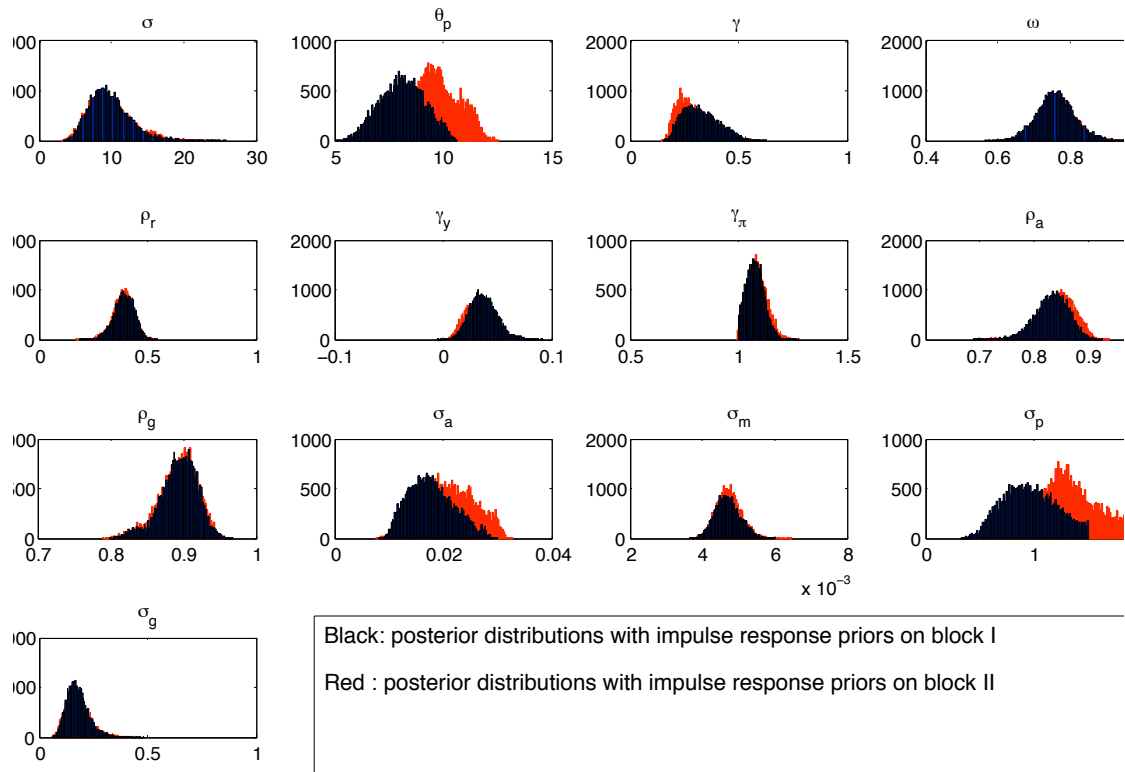


Figure 5: Block II. Posteriors: univariate kernel density

seems to be a meaningful pattern of change in the estimates, which is related to the fact that some higher parameter values such as the elasticity of labor supply can make up for lower nominal rigidities. This might suggest that deep parameters are only weakly identified in the model and then a correlated prior can effectively twist the estimates. A different interpretation or that the model is misspecified and then the structure of prior correlation interacts with the misspecification at hand. In order to disentangle which of the two explanations above can be more correct we undertook experiments on simulated data. In fact it can be assessed whether the model was weakly identified at the benchmark mode by simulating a dataset using that mode and by undertaking estimation.

We used our DGSE model, as described in section 3, to simulate a dataset of size comparable to our real dataset. Parameters were calibrated at the mode numerically found in the real dataset. We performed the estimation using both standard priors and our correlated prior on the second block. We contrast the results below:

Once misspecification is controlled for, we find no basic difference between the two estimates.<sup>15</sup> This suggests that misspecification is to blame for the instability of the estimates, although a further assessment of this issue is left for further research. As a check we ran a kalman smoother in order to recover the supposedly i.i.d. innovations of the shocks and to compute their contemporaneous correlation structure, shown in the table below:

| Shocks     | Technology | Monetary | Cost Push | Preference |
|------------|------------|----------|-----------|------------|
| Technology | 1.0000     | -0.5482  | 0.3077    | 0.0132     |
| Monetary   | -0.5482    | 1.0000   | -0.5675   | 0.1551     |
| Cost Push  | 0.3077     | -0.5675  | 1.0000    | 0.0138     |
| Preference | 0.0132     | 0.1551   | 0.0138    | 1.0000     |

It is easy to see that all shocks are highly correlated, which is a sign of misspecification<sup>16</sup>, with the exception of the preference shock.

<sup>15</sup>Some caution should also be used since we did not undertake a proper Monte Carlo experiment. With MCMC estimation this can take more than 300 hours of computing time.

<sup>16</sup>We assessed confidence bands for the correlation matrix by undertaking a little Monte Carlo experiment with 1000 simulated datasets. This provided us with the insight that in this model a reasonable confidence interval for correlations is (-0.2-0.2) with a mode of zero. Results are available upon request.

## 8 Conclusions

In this paper we compared posterior estimates results obtained using macro vs micro priors. We first highlighted the fact that macro priors are prone to problems concerning nuisance parameters, second we checked how much a presample based on data moments such as the DS prior can twist estimates when the model is misspecified. The intuition for that to happen is as follows: a VAR directly targets some moments of the data without the need of constructing unobservable variables, therefore it would correspond to a QML estimation procedure. This can be shown to yield a consistent estimate of the moments which are directly targeted even when the original model is

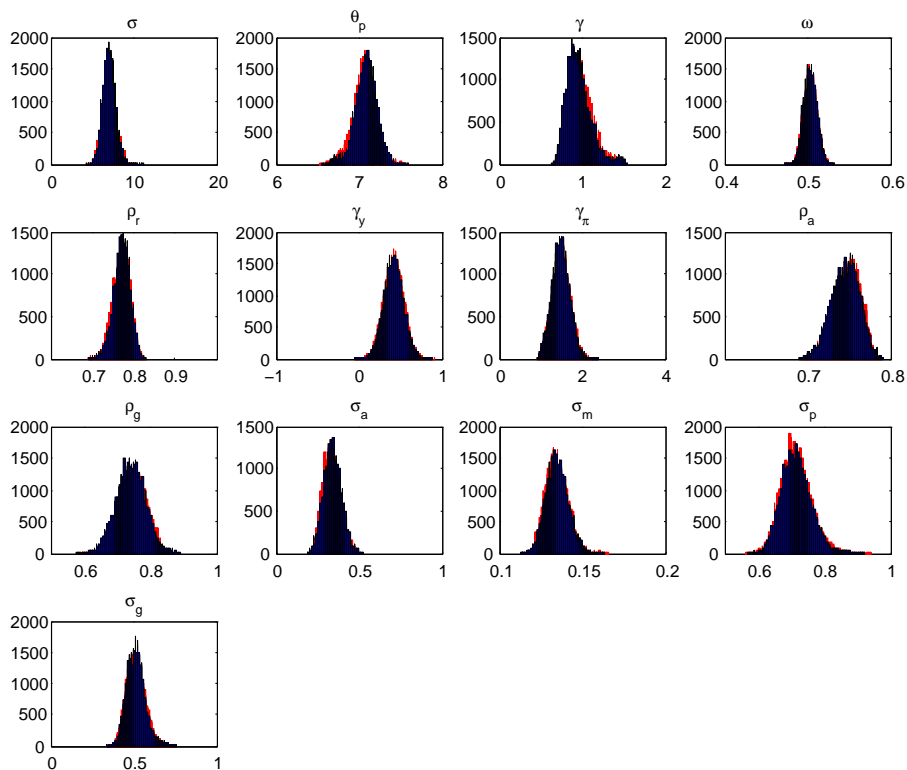


Figure 6: II block and independent prior benchmark

misspecified (see Mittelhammer et al 2003 page).

We then showed a simple way to form a prior which informs parameter correlation on the basis of the mapping between the structural and the reduced form of DSGE models via impulse response functions. The idea of forming priors as a matching of IRF is interesting on its own when the target impulse response is chosen from previous VAR studies. This would be the counterpart of a presample, with the clear advantage of having much stronger and wide knowledge on impulse responses rather than on data moments. Although matching impulse responses is not very different from matching data and theoretical moments as DS(2008) do, the advantage of our approach is the ability to match features that might not be very relevant from the variance decomposition of the data, but could be important for policymakers, such as monetary policy shocks. In this paper we do not pursue this way here and we leave it for future research.

The paper has two main implications concerning the use of correlated priors. First, estimated parameters heavily depend upon nuisance parameters which cannot be integrated out. We show this is the key to understand the Del Negro and Schorfheide (2008) result that posterior estimates change when a correlated prior on exogenous states parameters is adopted. Once the effect of nuisance parameters is offset, a joint prior on exogenous state parameters has a weaker influence on posterior estimates both using the DS prior. Still presample information can be helpful in making the moments of the simulated data more close to the real ones.

Second, we verify that applying our correlated prior on some key sets of parameters can twist the posterior estimates even after nuisance parameters are taken into account. By the means of a simulated dataset we check that this result is due to misspecification of the DSGE model. The change in the posterior estimates mostly concerns the Calvo probability of price adjustment and the variance of mark-up shocks. While in this paper we do not take a stand concerning the origin of the misspecification, our result is in line with the argument in Bills and Klenow (2004) which showed that the Calvo model prediction about an inverse correlation between inflation persistence and its volatility as regulated by the Calvo probability of price adjustment is fundamentally at odds with micro data. Rudd and Whelan (2006) take the argument of misspecification further by considering a generalized Phillips curve with indexation such as the one we have here and using macro data.

## 9 References

1. Altissimo, F. Mojon, B. and Zaffaroni, P. (2007) ‘Fast Micro and Slow Macro: can aggregation explain the persistence of inflation ?’ ECB working papers, 729.
2. An, S. and Schorfheide, F. (2007): ‘Bayesian Analysis of DSGE Models’, *Econometric Reviews*, pp. 187-192.
3. Berger, J.O., Liseo, B. and Wolpert, R. (1999), ‘Integrated likelihood methods for eliminating nuisance parameters’, *Statistical Science*, 14, 1,28
4. Bills, M. and Klenow, P. (2004), ‘Some Evidence on the Importance of Sticky Prices’, *Journal of Political Economy*, 112, October 2004, 947-985.
5. Christiano, L., Eichenbaum, M. and Evans, C. (2005), ‘Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy’, *Journal of Political Economy*, 113, 1-45.
6. Del Negro, M. and Schorfheide, F. 2004, ‘Priors from General Equilibrium Models for VARs’, *45 International Economic Review*,
7. Del Negro, M. and Schorfheide, F. 2008, ‘Forming Priors for DSGE Models and How it Matters for Nominal Rigidities’, *Manuscript*
8. Durbin, J. and Koopman, S.J. ‘Filtering and Smoothing of State Vector for Diffuse State Space Models’, *Journal of Time Series Analysis*, 24 (1), pp. 85-98.
9. Hansen, L.P. and Sargent, T. 1980, ‘ Forming and Estimating Linear Rational Expectation Models’, *Journal of Economic Dynamics and Control*
10. Iskrev, N. 2008, ‘How much do we learn from the estimation of DSGE models? A case study of identification issues in a New Keynesian business cycle model’, *Manuscript*.
11. Jondeau E. and Le Bihan H. 2008, ‘Examining Bias in Estimators of Linear Rational Expectations Models under Misspecification’, *Journal of Econometrics* 143, pp.375 - 395.

12. Klein, P. (2000): 'Using the Generalized Schur form to Solve Multivariate Linear Rational Expectations Models', *Journal of Economic Dynamics and Control*, 24(10), 1405-1423.
13. Kleinberger, F. and Zivot, E. (2003), 'Classical and Bayesian Approaches to instrumental Variables Regression', *Journal of Econometrics*, Vol 114(1), pp 29-72.
14. Kociecki, A. (2005) 'Priors for Impulse Responses in Bayesian Structural VAR Models'.
15. Rabanal, P and Rubio-Ramirez, J.F., 2005 'Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach', *Journal of Monetary Economics*, Vol. 52 pp 1151-1166.
16. Rudd, J. and Whelan, K. 2006. "Can Rational Expectations Sticky-Price Models Explain Inflation Dynamics?," *American Economic Review*, vol. 96(1), pages 303-320, March.
17. Saltelli, A., Tarantola, S., Campolongo F. and Ratto, M. (2004) 'Sensitivity Analysis in Practice' Wiley-VCH.
18. Sims, C., (2002): 'Solving Linear Rational Expectations Models', *Computational Economics*, 20 (1-2).
19. Sims, C. and Zha, T. 1999, 'Error Bands for Impulse Responses' *Econometrica*, Vol. 67, No. 5 (September 1999), pp 1113-1155.
20. S. James Press, 'Subjective and Objective Bayesian Statistics', Wiley Interscience 2004
21. Schorfheide, F. (2000), 'Loss Function-Based Evaluation of DSGE Models' *Journal of Applied Econometrics*, 15:6, 645-70.
22. Smets, F. and Wouters R. (2003), 'An estimated stochastic general equilibrium model of the euro area', *Journal of Economic Association*, Vol 1 (5), pp. 1123-1175.
23. Taylor, J. (1993) 'Discretion versus Policy Rules in Practice, Carnegie-Rochester Conference Series on Public Policy' 39, 195-214

| Params        | Benchmark        | DS      | DS      |
|---------------|------------------|---------|---------|
|               |                  | $\wp_1$ | $\wp_2$ |
| $\rho_a$      | 0.5 (0.01-0.99)  |         |         |
| $\rho_g$      | 0.5 (0.01-0.99)  |         |         |
| $\sigma_a$    | 0.5 (0.01-0.99)  |         |         |
| $\sigma_{ms}$ | 0.5 (0.01-0.99)  |         |         |
| $\sigma_p$    | 0.75 (0.01-1.49) |         |         |
| $\sigma_g$    | 0.5 (0.01-0.99)  |         |         |

Table 1: Prior distributions and 0.01-0.99 % quantiles

## 10 Appendix

We report some details concerning and shapes of the benchmark priors for the endogenous parameters:

| Parameters    | Benchmark prior mode | Prior shape      |
|---------------|----------------------|------------------|
| $\sigma^{-1}$ | 2                    | Gamma(2,1)       |
| $\Theta_p$    | 2                    | Gamma(2,1)       |
| $\gamma$      | 1                    | Normal(1,0.5)    |
| $\omega$      | 0.5                  | U(0,1)           |
| $\rho_r$      | 0.5                  | Normal(0.5,0.2)  |
| $\gamma_y$    | 0.25                 | Normal(0.25,0.1) |
| $\gamma_\pi$  | 1.5                  | Normal(1.5,0.5)  |

For the exogenous parameters benchmark distributions are uniform. We compare them with the DS priors for two different set of values of the nuisance parameters, either  $\wp_1 \equiv [\sigma^{-1} = 2, \Theta_p = 2]$ , or  $\wp_2 \equiv [\sigma^{-1} = 5, \Theta_p = 5]$ .

## 11 GLUE priors

Concerning GLUE priors, we do not truncate them with ‘ad hoc’ bounds on the support of priors. This means that there is a non-zero albeit small prior probability over the whole support. The only exception for this practice is on the Calvo pricing parameters where we imposed that on average prices cannot take more than 15 quarters to adjust. This constraint is a sensible choice given microeconomic evidence. Below we report some relevant quantiles (median, first and last centile) of the priors; the modes of the priors can easily be inferred from the kernel density plots which are also attached below.

## 11.1 The first block

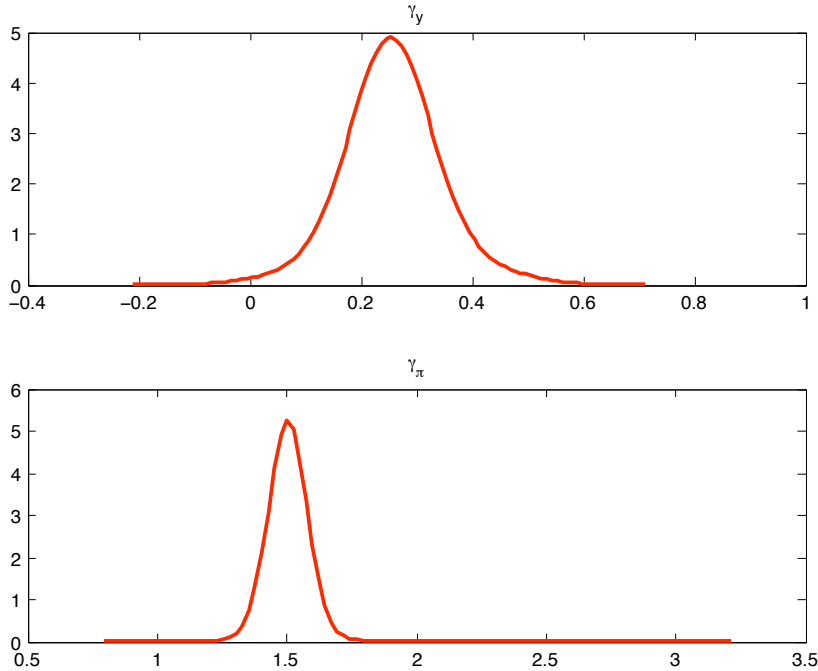


Figure 7: Block I. Univariate kernel density of  $\gamma_\pi$  and  $\gamma_y$ .

The figure 3 shows slices of the prior. Modes were set up as in the benchmark case:  $\gamma_\pi = 1.5$  and  $\gamma_y = 0.25$ . Due to the model structure, the two parameters turn out to be correlated, whereas a standard OLS regression of the Taylor rule would make them uncorrelated by construction. In figure 3 we report the kernel density estimates for single parameters<sup>17</sup>. Quantiles for our prior are shown below:

| Parameters   | 0.01 perc. | Median | 0.99 perc. |
|--------------|------------|--------|------------|
| $\gamma_y$   | 0.0130     | 0.2546 | 0.4963     |
| $\gamma_\pi$ | 1.2797     | 1.4996 | 1.7195     |

A three dimensional kernel density provides a more appealing representation of the correlation structure of the prior (figure 4):

The contour plot is also illustrative of the properties of our joint prior.

<sup>17</sup>We sampled coefficients from a uniform distribution, we computed the normalized weights, and we used a (weighted) kernel density procedure to plot their distributions and in order to compute quantiles

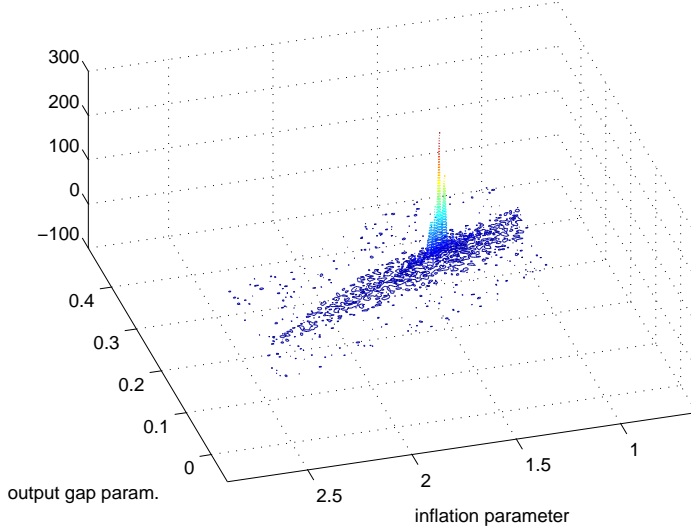


Figure 8: Block I. Bivariate kernel density of  $\gamma_\pi$  and  $\gamma_y$ .

On the X-axis of figure 5, we have the support of  $\gamma_\pi$ , on the Y-axis its probability, as according to our weights. The prior is plotted for different values of the second parameter  $\gamma_y$ . Since there is correlation between the parameters, the peak of the distribution w.r.t.  $\gamma_\pi$  changes when the second parameter changes; the prior tends then to evaluate not a single distribution centered in  $\gamma_\pi = 1.5$ , but an envelope of the distribution for the different values of  $\gamma_y$ ; this tends to penalize heavily draws which are very unlikely from a joint point of view.

## 11.2 Third block

With respect to the third block of priors we inform their modes on the basis of the means of the uniform distributions chosen as standard benchmark priors. Then we have the following :

| Parameters | 0.01 perc. | Median | 0.99 perc. |
|------------|------------|--------|------------|
| $\rho_a$   | -0.04      | 0.45   | 0.95       |
| $\rho_g$   | -0.06      | 0.50   | 1.04       |
| $\sigma_a$ | 0.00       | 0.48   | 0.98       |
| $\sigma_m$ | 0.00       | 0.50   | 0.96       |
| $\sigma_p$ | 0.00       | 0.71   | 1.50       |
| $\sigma_g$ | 0.00       | 0.49   | 1.03       |

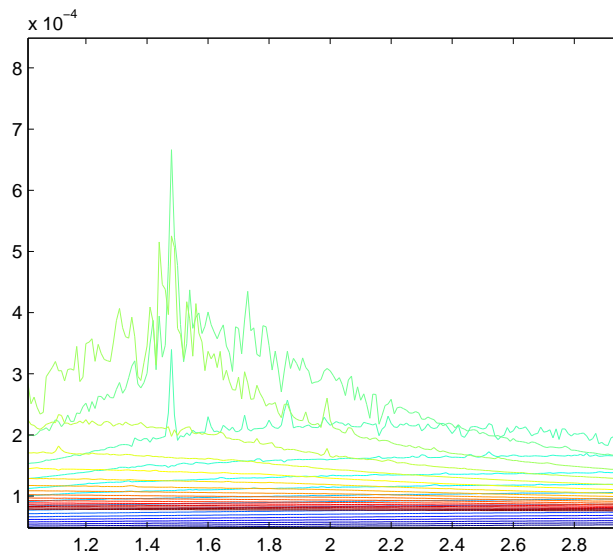


Figure 9: Block I. Joint prior contour plots

and kernel density plots follow:

### 11.3 Cumsum plots

We show cumsum (CS) plots of the posterior estimates, where cumsum is the difference between a rolling mean of posterior draws and the overall mean, scaled by the standard deviation of the chain:

$$CS_t = \left( \frac{1}{t} \sum_{n=1}^t \theta^n - \mu_\theta \right) \sigma_\theta,$$

in order to have the percentage oscillation of the CS statistic around the mean value. In order to compute posterior estimates we retain draws after a  $t$  such that the CS statistic oscillates by no more than 5%. In general this is achieved by selecting the last 50.000 draws of our chains, which is what we do. Below we report in the graphs for the last 100.000 draws of the chains.

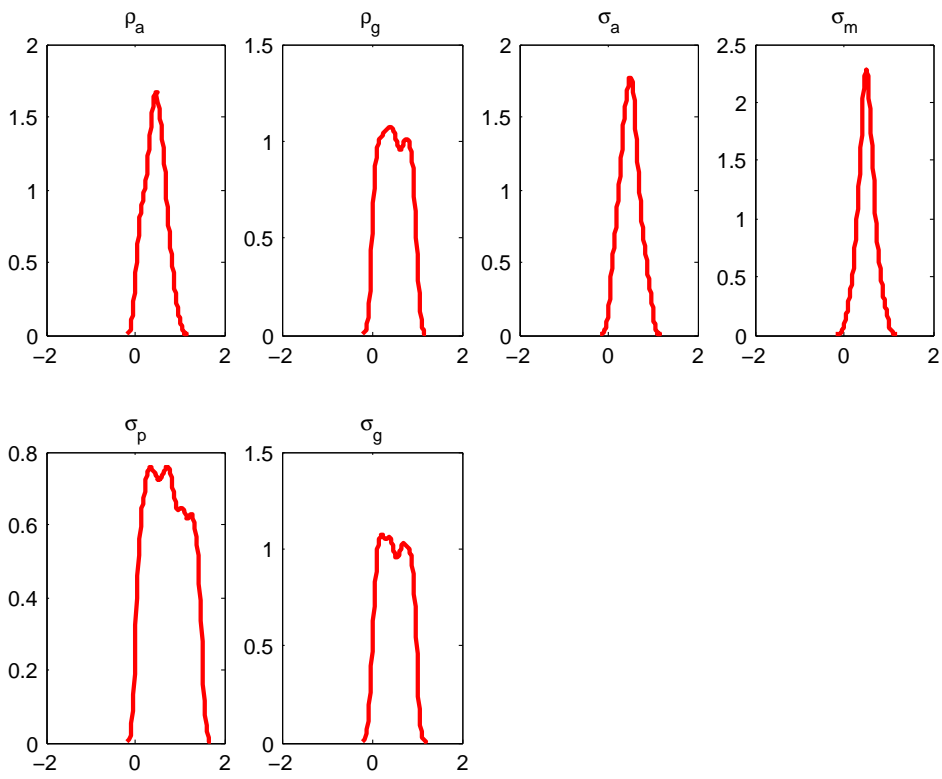


Figure 10: III block: univariate kernel density

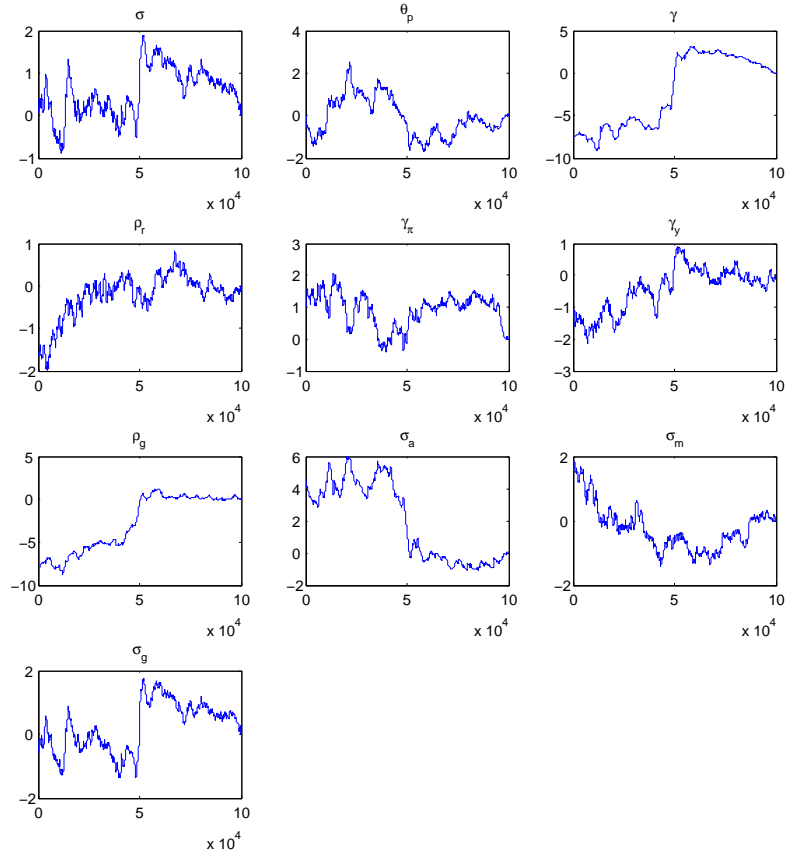


Figure 11: Cumsum plots : benchmark priors

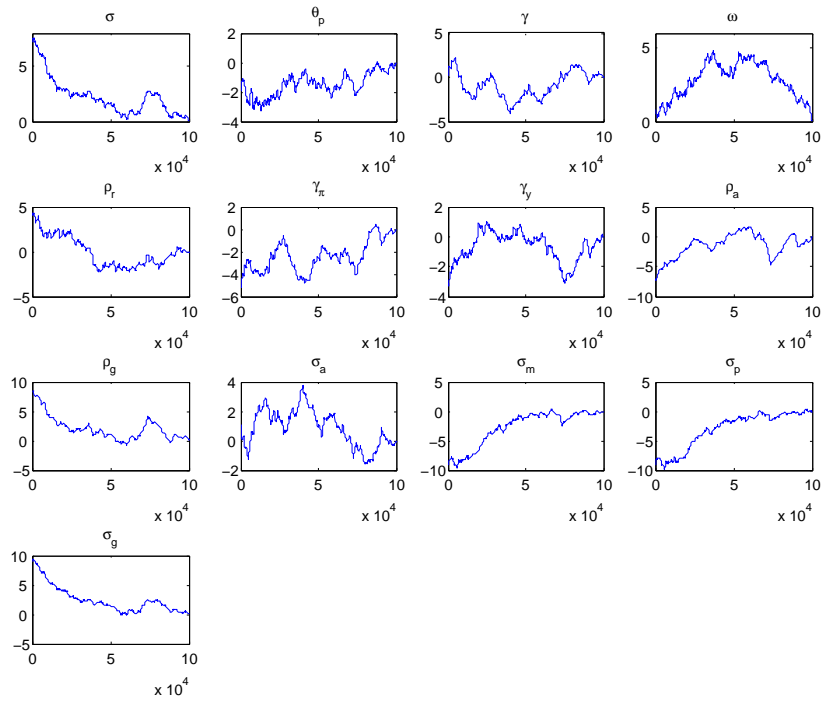


Figure 12: Cumsum plots : DS prior,  $\varphi_1$

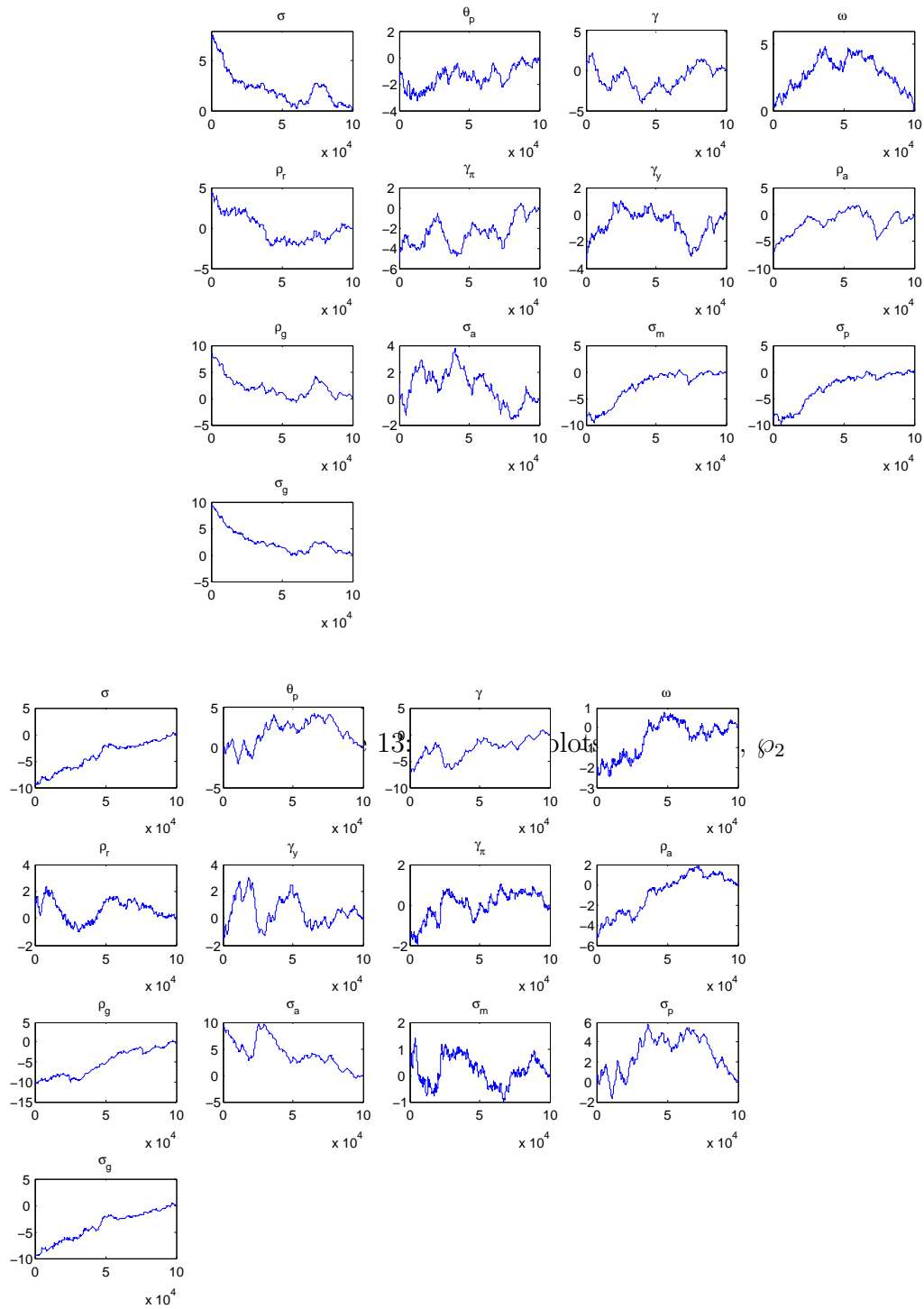


Figure 14: Cumsum plots: GLUE priors, block 2