EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1289 / JANUARY 2011

BAYESIAN PRIOR ELICITATION IN DSGE MODELS

MACRO- VS MICRO-PRIORS

by Marco J. Lombardi
and Giulio Nicoletti

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1289 / JANUARY 2011

BAYESIAN PRIOR ELICITATION IN DSGE MODELS

MACRO- VS MICRO-PRIORS [1]

by Marco J. Lombardi [2]

and Giulio Nicoletti [3]

In 2011 all ECB publications feature a motif taken from the €100 banknote.

1 Preliminary versions of this paper were presented at the 2009 ICEEE conference in Ancona and at the 2009 FESAMES congress in Tokyo. Without implicating, we would like to thank Luc Bauwens, Alessandro Rossi, Chris Sims, Frank Smets, Raf Wouters and an anonymous referee at the ECB WP series for their useful feedback and suggestions.

2 European Central Bank, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany; email: marco.lombardi@ecb.europa.eu

3 Corresponding author: Banca d'Italia, Via Nazionale, 91, 00184 Roma, Italy; e-mail: giulio.nicoletti@bancaditalia.it

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract

Bayesian approaches to the estimation of DSGE models are becoming increasingly popular. Prior knowledge is normally formalized either be information concerning deep parameters' values ('microprior') or some macroeconomic indicator, e.g. moments of observable variables ('macroprior'). In this paper we introduce a non parametric prior which is elicited from impulse response functions. Results show that using either a microprior or a macroprior can lead to different posterior estimates. We probe into the details of our result, showing that model misspecification is to blame for that.

JEL Classification Numbers: C11, C51, E30

Keywords: DSGE Models; Bayesian Estimation; Prior Distribution; Impulse Response Function

Non Technical Summary

Bayesian estimation methods have gained ground as a very attractive alternative over classical methods in the field of dynamic stochastic general equilibrium models (DSGE), among both academicians and practitioners. Unlike in the frequentist approach, a Bayesian researcher uses both information from the available data and prior knowledge, in order to provide so-called posterior estimates. Hence, an important aspect of the Bayesian approach relates to how the researcher elicits prior information and the way the latter influences posterior estimates, as results to be used for policy analysis should in principle be reasonably robust to a different prior specification.

For DSGE models, prior knowledge is normally expressed in the form of independent probability distributions placed on each of the structural parameters. Hence, the resulting joint prior distribution has independent components. Often such independent distributions are informed by using previous microeconometric studies.

This prior selection scheme is very simple and ready to implement, but suffers from several shortcomings. First, a priori independence of the deep parameters entails implicit assumptions for the a priori beliefs on data moments and the responses of macroeconomic variables after some shock (e.g. impulse response functions) which may indeed be at odds with prior knowledge drawing from microeconometric studies.

Second, for some parameters, microeconometric information can be abundant, while for others it can be scant At the same time there might be abundant evidence concerning the behaviour of macroeconomic variables, such as its response after some shock occurs (i.e. IRFs), that researchers might want to exploit.

Based on such considerations, Del Negro and Schorfheide (2008) pioneered a different approach in which prior distributions are formed on the basis of a priori beliefs on macroeconomic indicators (for example simulated moments) rather than on the deep parameters. This implies that the a priori belief on deep parameters is described by a joint distribution whose dependence structure hinges on the mapping between the deep parameters and the macroeconomic indicator at hand.

The contribution of our paper is twofold. First, we aim to investigate if and how the dependence structure of macropriors shape posterior estimates, and in particular how estimates of nominal rigidities are affected by the prior. Second, we find it convenient to do that by introducing a new type of prior, elicited in a non parametric way from impulse response functions. A clear advantage of a prior based on impulse response functions is that, differently from the deep parameters and, although to a lesser extent, from the sample moments used by Del Negro and Schorfheide (2008), researchers and policy makers often have a clear prior view of how the response of the economic system to certain shocks should look like.

Our main finding is that, when an IRF-prior is applied to exogenous state parameters, posterior estimates can differ with respect to the case of the benchmark independent case only by a little amount. Our second finding is that when applying the IRF-prior to a different set of deep parameters posterior estimates change more widely.

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), among both academicians and practitioners. Unlike in the frequentist approach, a Bayesian researcher uses both information from the available data and prior knowledge, in order to provide so-called posterior estimates. Hence, an important aspect of the Bayesian methodology relates to how the researcher elicits prior information and the way the latter influences posterior estimates.

For DSGE models, prior knowledge is normally expressed in the form of independent probability distributions placed on each of the structural parameters. Hence, the resulting joint prior distribution has independent components. Often such independent distributions are informed by using previous microeconometric studies. In the following, we will refer to this approach as 'microprior'.

The microprior selection scheme is very simple and ready to implement, but suffers from several shortcomings. First, a priori independence of the deep parameters entails implicit assumptions for the a priori beliefs on data moments and the responses of macroeconomic variables after some shock (e.g. impulse response functions, IRFs heretofore) which may indeed be at odds with prior knowledge drawing from macroeconometric studies. Second, for some parameters, microeconometric information can be abundant (e.g. the frequency at which firms adjust their prices, a direct measure of nominal rigidities in the economy), while for others it can be scant (e.g. concerning the average size of some shock or its persistence). At the same time there might be abundant evidence concerning the behaviour of macroeconomic variables, such as its response after some shock occurs (i.e. IRFs), that researchers might want to exploit.

Based on such considerations, Del Negro and Schorfheide (2008) – DS heretofore – pioneered a different approach in which prior distributions are formed on the basis of a priori beliefs on macroeconomic indicators (for example simulated moments or IRFs) rather than on the deep parameters; we will label this approach as 'macroprior'. A macroprior implies that the a priori belief on deep parameters is described by a joint distribution whose dependence structure hinges on the mapping between the deep parameters and the macroeconomic indicator at hand. More specifically, DS propose an hybrid approach in which a macroprior is used for the subset of parameters concerning the exogenous states, e.g. the persistence of exogenous shocks and their standard deviations, and a microprior for the other parameters such as those related to the amount of nominal rigidities in the economy. In the context of a Smets and Wouters (2003) type of model, DS find that posterior estimates, in particular those related to the nominal rigidities in the economy, which are *not* informed with the macroprior, change with respect to the case when only micropriors are used.

The contribution of our paper is twofold. First, we aim to investigate if and how the dependence structure of macropriors shape posterior estimates, as hinted by the results by DS, and in particular how estimates of nominal rigidities are affected by the prior. Second, we find it convenient to do that by introducing a new type of macroprior, elicited in a non parametric way from impulse response functions (IRF-prior heretofore). Concerning our IRF-prior, we build a controlled experiment, where we constrain the prior to have the same location (e.g. the mean/mode) as in the microprior. By doing this, any difference in posterior estimates can then be attributed to the second order properties of the IRF-prior, e.g. the dependence structure of parameters and its spread. A clear advantage of the IRF-prior is that, differently from the deep parameters and, although to a lesser extent, from the sample moments used in DS, researchers and policy makers often have a clear prior view of how the response of the economic system to certain shocks should look like, even without explicitly referring to a data-based presample.[1] We compare the way our prior's dependence structure shapes posterior estimates with what is obtained with both a microprior and the DS prior.

Our main finding is that, when an IRF-prior is applied to exogenous state parameters we find that posterior estimates can differ with respect to the case

[1]For example, in the setting of structural VARs, identifying schemes based on placing restrictions on the signs of the impulse responses Uhlig (2005) are becoming increasingly popular.

of the benchmark independent case by a little amount. This happens when the joint prior is set to be rather tight, while the effects are negligible when the overall variance of the prior is larger. This differs from the findings of DS, which report large differences in the posterior estimates: we apply their method to our case and we are able to explain how this difference is produced. Our second finding is that when applying the IRF-prior to a different set of deep parameters posterior estimates change more widely. As an additional result, by recurring to an exercise on simulated data, we highlight the fact that this last feature is likely to be due to model misspecification.

The paper is structured as follows: in section 2 we describe the general formulation of our IRF-prior. In section 3 we introduce the DSGE model used for our experiments. Section 3.1 details the estimation strategy we follow and 3.2 shows posterior estimates with our benchmark prior. In section 4 the IRF-prior is applied to the block of parameters which pertain to exogenous states. In subsection 4.2 we contrast our results with those obtained with the DS prior. In section 4.3 we present changes in posterior estimates obtained when a IRF-prior is applied to a different block of deep parameter and we discuss whether this might be due to misspecification or to weak identification. Some conclusions follow. A discussion of more technical parts, such as the 'intractable normalizing constants', the DSGE model specification and a presentation of DS method are presented in the appendix.

2 Impulse Response Priors

The reduced form of a (linearized) DSGE model can be written in state space form An and Schorfheide (2007):

$$X_t = F(\psi)X_{t-1} + G(\psi)\epsilon_t, \quad (1)$$

$$\epsilon_t \sim N(0, \Sigma_\epsilon), \quad (2)$$

$$Y_t = HX_t, \quad (3)$$

the transition equation corresponds to the solution (e.g. the reduced form) of the model. The n-dimensional vector of states X_t collects all variables which are known to economic agents (households, firms and government) at time t Sims (2002). X_t includes so called exogenous states (e.g. total factor productivity), which follow, individually, a scalar autoregressive process. ϵ_t denotes the l-dimension vector of i.i.d. normally distributed (and uncorrelated) structural shocks, which appear either as innovations of exogenous state processes and/or directly in one or more other equations. Their standard deviation is a diagonal matrix Σ_ϵ, or, as vector, σ_ϵ. Both the $(n \times n)$ F matrix and the $(n \times l)$ G matrix in the transition equation are functions of ψ, the vector of so called deep (or structural) parameters (e.g. parameters in the utility functions of the agents or parameters related to technology). The vector ξ gathers both σ_ϵ and ψ:

$$\xi \equiv [\psi, \ \ \sigma_\epsilon].$$

For simplicity we will refer to ξ as the vector of 'deep parameters'.

The $l \times n$ selection matrix H links states X_t to the variables Y_t which are observed by the econometrician. As common in the literature, we also assume that the number of observed variables is equal to the number of shocks. Y_t is then a vector of dimension l.[2]

Referring to variable Y, the j-period-ahead impulse response to a shock ϵ_t at time t, is defined as the difference between the j-steps ahead projection of Y_t, conditional on both the $t-1$ information (I_{t-1}) and the knowledge of ϵ_t, and the same j-steps-ahead projection without the knowledge of ϵ_t Koop, Pesaran, and Potter (1996):

$$\gamma_{t,j} \equiv E[Y_{t+j} \mid \epsilon_t = \sigma_\epsilon, I_{t-1}] - E[Y_{t+j} \mid I_{t-1}], j = 0, \ldots, m \tag{4}$$

where, as convenient, we conditioned the shocks on the their standard deviations. In our linear state space model (1-3), following definition (4), the impulse

[2]This assumption is not important to define impulse responses while it plays a relevant role in the estimation process, as it is not possible to run the Kalman Filter on a stochastically singular system.

$\gamma_j(\xi)$ at horizon $j = 0, \ldots m$ reduces to:

$$\gamma_j(\xi) = HF^j(\psi)G(\psi)\Sigma_\epsilon,\ j = 0, 1, \ldots m, \tag{5}$$

$\gamma_j(\xi)$ is a $l \times l$ matrix. Stacking all horizons we denote:

$$\gamma(\xi) \equiv [\gamma_0(\xi) \ldots \gamma_j(\xi) \ldots \gamma_m(\xi)],\ 0 < \forall j < m,$$

which is a $l \times l \times (m+1)$ matrix. We follow this notation in non-ambiguous cases.

Our nonparametric prior is constructed as follows. For a given ξ, compute a distance function between the model impulse $\gamma(\xi)$ and some target impulse γ^*, which is supplied by the analyst, as it contains his prior ideas on the response. The distance is denoted with $d(\gamma(\xi), \gamma^*)$ and it maps the $l \times l \times (m+1)$ space (measurement, shock and impulse horizon) into the positive real line. Our prior kernel w is given by a logistic[3] transformation of the distance: [4]

$$w(\xi \mid \gamma^*, K) = \frac{exp(-d/K)}{K(1 + exp(-d/K))^2}, \tag{6}$$

where K is the variance of the logistic and it controls for the degree of belief the analyst has on his prior. A proper prior probability distribution can be obtained by normalizing the kernel:

$$\bar{w}(\xi \mid \gamma^*, K) = \frac{w(\xi \mid \gamma^*, K)}{\int w(\xi \mid \gamma^*, K)d\xi}, \tag{7}$$

but the knowledge of the normalization constant (e.g. the denominator of 7) is

[3]Results are robust to the choice of the penalty function; we experimented with a normal distribution-like function (consistently with the steady state priors explored by Del Negro and Schorfheide (2008)), reporting similar results as for the logistic case. For the model used in section 3 the logistic function reported faster convergence properties for the MCMC. Results on the normal penalty function are available upon request.

[4]Different distance functions can also be used; for example, we experimented the inverse of the distance, with broadly similar results. However, the inverse function is not continuous on a measure zero set, e.g. when the distance is exactly zero. In practice we found it to be not a problem unless the parameter space of the prior is of dimension one. Yet, in this case the MCMC sampler can get stuck at the point of discontinuity.

not required since the model is estimated by MCMC techniques. As the prior kernel is directly used in the estimation process in the rest of the presentation we omit the proper prior $\bar{w}$ and we directly express formulas in terms of the kernel w.

Using a penalty function draws upon the idea of a distance between impulses, already in use for estimating DSGE models Christiano, Eichenbaum, and Evans (2005). The distance we use is a quadratic function of the type used in impulse response matching:[5]

$$d(\xi \mid \gamma^*) = vec(\gamma(\xi) - \gamma^*)' \, W \, vec((\gamma(\xi) - \gamma^*)), \tag{8}$$

where W is an $(l^2 * (m+1)) \times (l^2 * (m+1))$ identity matrix and vec denotes the vec matrix operator.[6] A branch of research which is closely connected to our approach is the literature known as Generalized Likelihood Uncertainty Estimation, within the framework of global sensitivity analysis, where some measure of distance is also used to inform distribution measures (see Saltelli, Tarantola, Campolongo, and Ratto (2004) pages 182 and following).

Formula (6) defines a joint prior kernel over the whole vector of parameters ξ, but, as explained in the introduction, the best use of prior knowledge of researchers might entail using a macroprior only on a subset of parameters, $\xi_\dagger$. Partitioning the vector of parameters ξ into two parts,

$$\xi \equiv [\xi_\dagger, \, \xi_b],$$

only $\xi_\dagger$ is informed by IRF-priors and ξ_b is the subvector informed by independent distributions $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$. Hyperparameters (location and spread) are now denoted ξ_b^*, Σ_b^*; were the benchmark prior applied also to the † parameters, its hyperparameters would be denoted as $\xi_\dagger^*, \Sigma_\dagger^*$. The 'hybrid' prior over the

[5] We experiment also different functions: a sup distance; and the function described by Uhlig (2005) in the context of sign restrictions for VAR models. Since we have not found strikingly different results using different distance measures, we present the quadratic function case, except when stated differently.

[6] In order to avoid the criticism of Canova and Sala (2009) that the surface described by the distance between impulse responses can be very ill-behaved due to the non linearity of IRFs, we set $m = 1$, that is we use only the impact of the impulse response and the first lead; results are in general robust to using $m < 5$.

whole vector of parameters, $p(\xi|\xi_b^*, \Sigma_b^*, \gamma^*, K)$ can now be defined as (proportional to) the product of the kernel of the IRF-prior conditional to ξ_b, times an independent prior on ξ_b:

$$p(\xi|\xi_b^*, \Sigma_b^*, \gamma^*, K) \propto w(\xi_\dagger \mid \xi_b, \gamma^*, K)p(\xi_b|\xi_b^*, \Sigma_b^*), \tag{9}$$

where the conditional distribution is computed from the ratio between joint and marginal IRF prior:

$$w(\xi_\dagger \mid \xi_b, \gamma^*, K) = \frac{w(\xi_\dagger, \xi_b \mid \gamma^*, K)}{w(\xi_b \mid \gamma^*, K)}. \tag{10}$$

Expression (9) is of little use for estimation: while the numerator in (10) is computable, the denominator cannot be analytically computed and, by the same token, it varies over draws of ξ_b, making it unsuitable for MCMC estimation. This is a (well) known issue in the MCMC literature as the 'intractable normalizing constant problem' Moeller, Pettitt, Berthelsen, and Reeves (2006). DS propose to approximate the conditional distribution at hand by the joint IRF-distribution (i.e. the numerator of 10, which is known), where the conditioning parameters ξ_b are kept fixed at some value ($\xi_b^?$):

$$w(\xi_\dagger \mid \xi_b, \gamma^*, K) \approx w(\xi_\dagger, \xi_b = \xi_b^? \mid \gamma^*, K). \tag{11}$$

Introducing the calibrated set of parameters $\xi_b^?$, equation (9) is replaced by the following one:

$$p(\xi_\dagger, \xi_b = \xi_b^? \mid \xi_b^*, \Sigma_b^*, \gamma^*, K) \propto w(\xi_\dagger, \xi_b = \xi_b^? \mid \gamma^*, K)p(\xi_b \mid \xi_b^*, \Sigma_b^*), \tag{12}$$

where now the full expression for the kernel is written as:

$$w(\xi_\dagger, \xi_b = \xi_b^? \mid \gamma^*, K) = \frac{exp(-d(\xi_\dagger, \xi_b = \xi_b^? \mid \gamma^*, K)/K)}{K(1 + exp(-d(\xi_\dagger, \xi_b = \xi_b^? \mid \gamma^*, K)/K))^2}. \tag{13}$$

Equations (12) and (13) together with the definition of the distance (8), as modified to take into account the partitioning of the set of deep parameters, define the IRF-prior we will be using in the remainder of our analysis. The subvector

ξ_b plays now a double role:

1. ξ_b is informed with an independent prior (micropriori) $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$.
2. It is calibrated in the IRF-prior (and in the DS prior) at some value $\xi_b^?$.

In the analysis by DS, which is based on sample moments rather than IRFs, the same type of calibrated parameters appear. From their analysis we have little guidance on how much point 1 vs. 2 above drive posterior estimates. Also it is not clear how much having a non-independent prior matters for results. To shed light on those issues, in the following we run a set of experiments with our prior and we cross-check some of our results by using the DS methodology. In the first set of experiments we apply the IRF-prior on the exogenous state parameters (section 4) as in DS, in the second one (section 4.3) we apply it to a different set of deep parameters more related to the structure of the economy, such as the amount of nominal rigidities; more details of the selected parameters are offered to the reader at the end of section 3. In order to control for the effects of point 2 in both experiments we use a target impulse γ^* derived from the DSGE model itself

$$\gamma^* = \gamma(\xi_{IRF}^*, \xi_b = \xi_b^?),$$

as ξ_{IRF}^* is the same location hyperparameter as in the microprior, we also focus on the role of second order properties (spread and correlation structure) of the macroprior in shaping the posterior distribution. As we check, once calibrated parameters are controlled for, posterior estimates do not vary much as calibrated parameters are considered at different $\xi_b^?$.[7] In section 4.2 we show how the DS prior is sensitive to calibrated $\xi_b^?$. In particular, when comparing results for the benchmark and the DS prior, the two posteriors come closer when also $\xi_b^?$ is closer to the posterior mode obtained under the benchmark prior. Our results show that the introduction of calibrated parameters is not at all innocuous and it is the main determinant of the large impact of prior knowledge on posterior estimates as found by DS.[8]

[7] In a more general case, when γ^* is taken from VAR studies or when pre-sample information is considered, as it is the case of the DS prior, this is not likely to be the case. Further research is in progress by the authors on this topic.

[8] Our prior could be computed for a given set of points, integrated with respect to the sub-

3 The DSGE Model

Our workhorse is a New Keynesian model which features Calvo pricing with indexation, no capital, no habits in consumption and no wage rigidities; the model follows the description in Rabanal and Rubio-Ramirez (2005) – RR heretofore.[9]. The set of its structural parameters is given by:

$$\xi \equiv (\beta, \eta, \delta, \sigma, \theta_p, \gamma, \omega, \rho_r, \gamma_y, \gamma_\pi, \rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p),$$

β is the discount factor, η is the elasticity of substitution among varieties in the bundle of commodities, δ is the share of capital in the Cobb Douglas production function, σ is the elasticity of intertemporal substitution and θ_p is the (Calvo) probability of price adjustment, which lies in a bounded space (0,1). γ is the elasticity of labor supply, ω is the degree of backward looking indexation in the Phillips curve, $\rho_r, \gamma_y, \gamma_\pi$ are the parameters concerning the Taylor rule (interest rate inertia, output term and inflation term). The rest of parameters pertain to exogenous state variables, technology (a) and preference shocks (g) which are autocorrelated (ρ_a, ρ_g), while the monetary shock (m) and the mark-up shocks (p) are modeled as i.i.d. Gaussian.

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 also RR); the calibrated elements are set at $\beta = 0.99, \delta = 0.36, \epsilon = 6$. Since both the elasticity of substitution σ and the Calvo probability θ_p lie in a bounded space (i.e. between zero and one), we transform them as:

$$\sigma^{-1} \equiv 1/\sigma \;\; ; \;\; \Theta_p \equiv \frac{1}{1-\theta_p}.$$

The transformation at hand is also convenient from an economic point of view: transformed variables are, respectively, the risk aversion of agents and the aver-

vector ξ_b and then an interpolation scheme might be used to fill missing points during the estimation procedure. Due to interpolation, this procedure turned out to be very unstable from a numerical point of view and we abandoned it.

[9]Some further details concerning the model and the motivation for choosing a small scale model are provided in appendix E.

age amount of time that it takes until a price is revised. As it is common practice in literature we formalize our priors as functions of the transformed parameters.

The vector of estimated parameters ξ hence reduces to:

$$\xi \equiv (\sigma^{-1}, \Theta_p, \gamma, \omega, \rho_r, \gamma_y, \gamma_\pi, \rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p),$$

and exogenous state parameters are the last six terms.

In the next section we sketch the general estimation procedure. As we run two different experiments with IRF-priors, in each one we inform a different block of parameters by our macroprior. These blocks are respectively given by: [10]

1. The exogenous states parameters block, as in Del Negro and Schorfheide (2008):

$$\xi_\dagger \equiv [\rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p].$$

   For this case the application of the DS prior is also discussed.

2. Parameters non related to the exogenous shocks, with the exception of the intertemporal elasticity of substitution which we found hard to identify when a joint prior is used:[11]

$$\xi_\dagger \equiv [\Theta_p, \gamma, \omega, \rho_r].$$

3.1 Estimation

From the model solution, we write a state-space model as $(1-3)$, by using the reduced form of the model as transition equation. The state vector X_t is given by real wages (rw_t), inflation (π_t), GDP (y_t), the FED funds rates (r_t); two autoregressive processes of order one: a preference shock process (g_t) and a technology shock process (a_t). The vector of i.i.d. shocks ϵ_t is given by the innovations of the two the AR(1) processes (technology and preference shock)

[10] We also ran a third experiment covering only parameters in the Taylor rule, but we omit it since it yields no different results with respect to the benchmark prior.

[11] In linear models this parameter can be difficult to identify, see the paper by An and Schorfheide (2007).

plus monetary and mark up shock (see the appendix for more details). The measurement equation is given by:[12]

$$\text{Real Wage} = 100\, rw_t,$$

$$\text{Real Output Growth\%} = 100(y_t - y_{t-1}),$$

$$\text{Annualized inflation rate} = 400\pi_t \equiv 400(\ln P_t - \ln P_{t-1}),$$

$$\text{Annualized interest rate} = 400\, r_t,$$

where measurements are on the left hand side, their model counterparts on the right hand side. To form the likelihood of the model we use the Kalman Filter innovations: its implementation follows the description in Durbin and Koopman (2001). Since an analytical form of the posterior distribution cannot be obtained, we use an MCMC Random Walk Metropolis-Hastings to draw samples from it.

A detailed description of how the estimation procedure works with the different priors is summarized below; we denote differently only those steps which differ across different priors (b, IRF, DS are respectively for benchmark priors, IRF prior, DS prior):

1b Set benchmark priors: $p(\xi \mid \xi^*, \Sigma^*)$. Combine with the Likelihood function to form the posterior kernel.

1IRF Set benchmark $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$ and compute a target $\gamma^* = \gamma(\xi_\dagger^*, \xi_b^? = \xi_b^*)$ (compute posterior kernel).

1DS Set $p(\xi_b \mid \xi_b^*, \Sigma_b^*)$ and compute sample moments (Γ^*) on a pre-sample (compute posterior kernel).

2 Use a numerical optimizer to find the mode of the posterior defined above for benchmark priors (1b), call the mode $\bar{\xi}$.[13] To make the different estima-

[12]The sample is 1960:01-2001:04. Series on output, prices and wages come from the Bureau of Labor Statistics, output for the non farm business sector and its deflator and hourly compensation for the nonfarm business sector. Fed Funds rates from the FRED database of the FED of St. Louis. Data have been made stationary prior to estimation: inflation and output are in growth rates and we demean them, whereas a linear trend is subtracted from the real wages series.

[13]We used the csminwell.m program by Chris Sims.

tions comparable this step is run only once with benchmark priors and the estimates are used to run estimation for all different priors.[14]

3 Compute the inverse Hessian of the posterior H at $\bar{\xi}$ and decompose via eigenvalue-eigenvector decomposition to form a matrix D: $H = D'\Lambda D$.

4 Draw a candidate vector of parameters (ξ^c) according to a random walk with normally distributed innovations $\eta \sim N(0,1)$:

$$\xi^c = \xi_i + \mu \Lambda^{-\frac{1}{2}} D\eta,$$

where $\xi^c = [\xi^c_\dagger, \xi^c_b]$ and μ 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 ξ_0, which is numerically found in step one by benchmark priors.

5b Combine Likelihood and benchmark prior to get a posterior kernel for draw ξ^c.

5IRF Compute the impulse response function $\gamma(\xi^c_\dagger, \xi^?_b = \xi^*_b,)$ and form the prior kernel $w(\xi^c_\dagger, \xi^?_b = \xi^*_b \mid \gamma^*, K)$, combine with $p(\xi^c_b \mid \xi^*_b, \Sigma^*_b)$ and the Likelihood to get a posterior kernel for draw ξ^c.

5DS Compute the simulated data moments implied by draw $\xi^c_\dagger, \xi^?_b$ where $\xi^?_b$ is fixed (we discuss this in section 4.2) and get the DS prior kernel (described in 23) combine with the benchmark $p(\xi^c_b \mid \xi^*_b, \Sigma^*_b)$ and with the Likelihood to get the posterior kernel of ξ^c.

6 Accept the draw ξ^c with probability depending on posterior odds ratio:

$$\xi_{i+1} = \xi^c \text{ if ratio} \equiv \frac{P(\xi^c)}{P(\xi_i)} > \text{rand(U(0,1))}, \quad (14)$$

$$\xi_{i+1} = \xi_i, \text{otherwise}, \quad (15)$$

[14]It is clearly possible to form a posterior, to be used for numerical maximization, applying points 1IRF and 1DS; MCMC posterior results are robust to this.

where P is the posterior kernel (likelihood times prior) and $rand$ is a uniformly-distributed random variable.

7 The algorithm is run in one chain for 200.000 draws, the μ is calibrated in order to keep the acceptance rate between 20 and 40 % as in common literature concerning Bayesian DSGE estimation.

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; further details on convergence diagnostics are reported in the appendix.

3.2 Posterior results with the benchmark prior

In our benchmark prior all parameters are independent random variables which follow from the description in Rabanal and Rubio-Ramirez (2005).[15] The prior mode of 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 two quarters; both distributions are modeled as gamma. The elasticity of labour is also set equal to one with a normal distribution. We chose 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 Taylor (1993). For the exogenous parameters block we use uniform distributions.[16] 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. The chosen shapes and parameters of the prior distributions, are reported in table 1.

The posterior distributions generated by employing the benchmark prior are shown in figure 1.

---

[15]Results do not change when more or less diffuse priors are adopted, provided they are independent distributions.

[16]Switching to an inverted gamma prior for the variances of shocks does not alter results significantly.

Parameters	Mean	Distribution
σ^{-1}	2	Gamma(2, 1)
Θ_p	2	Gamma(2, 1)
γ	1	Normal(1, 0.25)
ω	0.5	Uniform(0, 1)
ρ_r	0.5	Uniform(0, 1)
γ_y	0.25	Normal(0.25, 0.125)
γ_π	1.5	Normal(1.5, 0.25)
ρ_a	0.5	Uniform(0, 1)
ρ_g	0.5	Uniform(0, 1)
σ_a	0.5	Uniform(0, 1)
σ_m	0.5	Uniform(0, 1)
σ_g	0.5	Uniform(0, 1)
σ_p	0.75	Uniform(0, 1.5)

Table 1: Sufficient statistics for the benchmark prior (microprior)

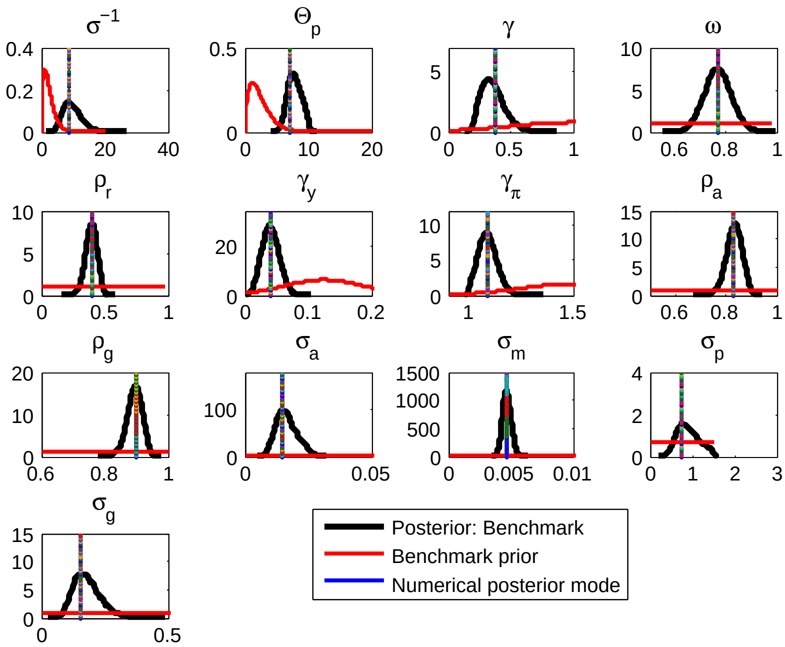


Figure 1: Benchmark prior, posterior distribution, numerical posterior mode

Posterior estimates are quite close to the numerical posterior mode. The elasticity of labor supply turns out quite low compared to its prior mean value; this makes marginal costs less sensitive to variations in output and therefore it can be seen as a substitute mechanism to wage rigidity. The degree of inflation indexation turns out to be quite high ($\omega = 0.77$), which might also be a consequence of the lack of rigid wages in the model. Estimated price rigidities (Θ_p) also turn out to be quite high compared to microeconometric evidence implied by our prior, which is a common feature in this class of models. Parameter estimates are similar to those obtained by Rabanal and Rubio-Ramirez (2005) with a similar dataset.

4 The impact of IRF priors

4.1 IRF priors: first block

In this section we apply our IRF-prior to the exogenous states block of parameters:

$$\xi_{\dagger} \equiv [\rho_a, \rho_g, \sigma_a, \sigma_m, \sigma_g, \sigma_p].$$

Kernel density plots (figure 4.1) show both the case $K = 0.5$ and the more informative prior $K = 0.05$, together with the benchmark prior.

We report means and quantiles of the IRF-prior below, for the case $K = 0.5$:[17]

[17] Due to the nonlinear mapping between the structural and the reduced-form parameter spaces, analytical expressions for IRF-priors cannot be obtained, therefore we draw the parameters we are interested in from a uniform distribution and we compute our normalized prior. This is then used to re-weight the original draws as in an importance sampling scheme. This approach is used to compute (weighted) kernel densities, quantiles and the linear correlation measures in the tables.

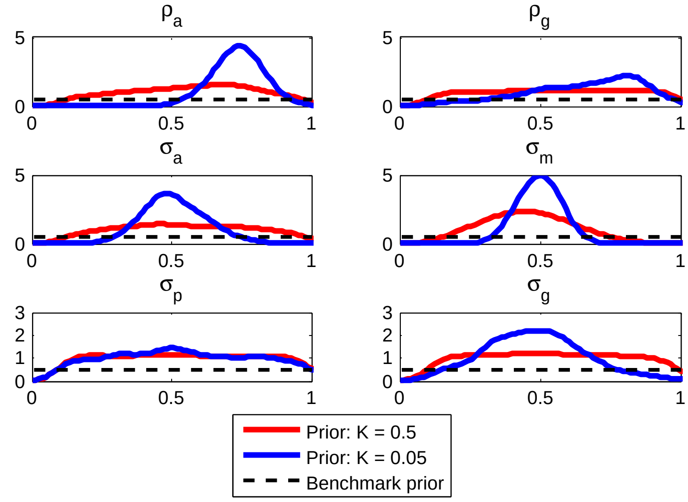


Figure 2: I block of parameters: univariate kernel density

Parameters	0.01 perc.	Mean	0.99 perc.
ρ_a	0.06	0.64	1.00
ρ_g	0.06	0.55	1.01
σ_a	0.07	0.49	1.00
σ_m	0.09	0.46	0.96
σ_p	0.06	0.53	1.01
σ_g	0.06	0.54	1.01

The amount of information conveyed by the prior differs across parameters; it is quite high for the standard deviation of technology and monetary policy shocks, while it is lower for the remaining parameters.

By using importance sampling we also compute the correlation structure implicit in our prior; parameters turn out to be highly correlated.

Parameters	ρ_a	ρ_g	σ_a	σ_m	σ_p	σ_g
ρ_a	1.00	0.97	0.61	0.96	0.69	0.85
ρ_g	0.97	1.00	0.67	0.92	0.72	0.92
σ_a	0.61	0.67	1.00	0.65	0.98	0.87
σ_m	0.96	0.92	0.65	1.00	0.74	0.82
σ_p	0.69	0.72	0.98	0.74	1.00	0.88
σ_g	0.85	0.92	0.87	0.82	0.88	1.00

From the MCMC posterior estimates we found that there is little difference with respect to the benchmark prior case. Figure 3 shows the case $K = 0.05$, as for $K = 0.5$ the difference turned out to be almost negligible. Regarding σ_p, in spite of the fact that the IRF-prior appears to be almost as informative as the benchmark in figure 4.1, this is the only posterior for which there was some noticeable shrinkage using $K = 0.05$.

4.2 DS reassessed

In this section we compare benchmark estimates with posterior results obtained by applying a DS-prior to the set of exogenous state parameters. Some description and technical details concerning the DS-prior are in appendix; estimation follows the steps outlined in section (3.1). We use a pre-sample of around forty observations on which we compute moments, leaving the remaining for estimation[18]. Following DS (2008), the moments computed for the observable variables are restricted to the variance covariance matrix and the order one (cross) autocorrelation. We fix the scale parameter of the strength of the prior, T^*, at 6, also following DS.[19]

For what concerns the calibrated parameters, experiments show that two parameters matter the most in shaping results: the parameter for risk aversion and the average amount of time after which a firm is able to change its price. We present two experiments, respectively DS1 and DS2 below, which differ only

[18] Results are not very sensitive to this choice, we also tried different splits without seeing much qualitative difference, albeit a little quantitative one.

[19] We checked that for T^* in the range between 4 and 8 results do not change.

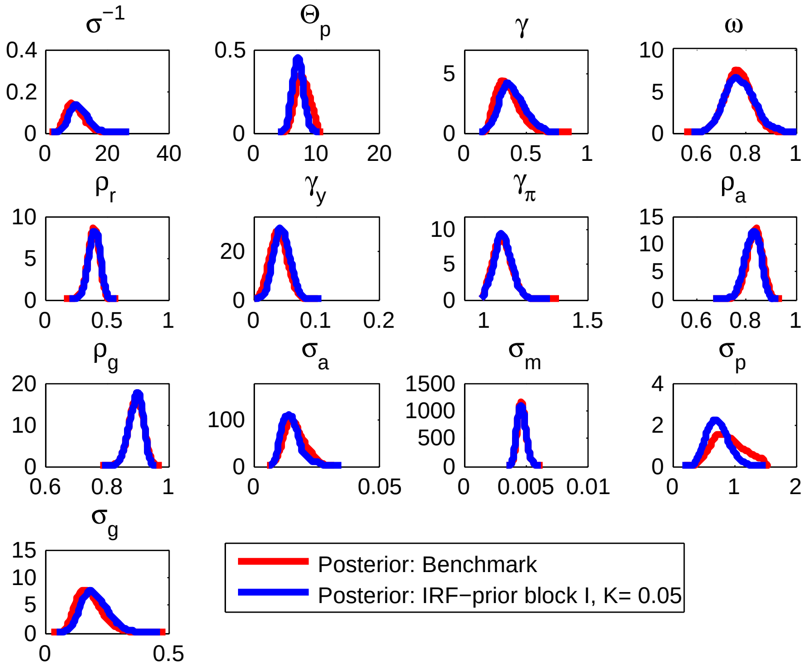


Figure 3: Posteriors: univariate kernel density, using benchmark prior and IRF-prior block I.

for the calibration of those two parameters. The following points summarize the experiments:

- Calibrated parameters are:

$$\xi_b^? = [\sigma^{?-1}, \Theta_p^?, \gamma^?, \omega^?, \rho_r^?, \gamma_y^?, \gamma_\pi^?],$$

  all but two are set at the benchmark prior mean.

DS1 ($\sigma^{?-1}$, $\Theta_p^?$) are both set equal to 2, the benchmark prior mean.

DS2 ($\sigma^{?-1}$, $\Theta_p^?$) are both set equal to 5, closer to the posterior mean found with benchmark priors.

In both experiments the benchmark prior applied to subvector ξ_b is the same, as described in table 1.

Such an (apparently) small difference in calibration between the two experiments is sufficient to produce dramatic differences in posterior estimates, as shown in figure 4. In particular we find that posterior nominal rigidities (Θ_p) are very low when $\Theta_p^? = 2$, which is in line with the findings of DS (2008). MCMC posterior estimates for DS1 and DS2 with respect to benchmark are shown in figure 4.

When the two parameters are set closer to the posterior mode obtained under benchmark prior (DS1 experiment, reported in blue), estimates for nominal rigidities become much closer to those obtained by using the benchmark prior (red line). The DS prior seems to be quite informative, posterior distributions being narrower with respect to the benchmark prior case. The DS2 experiment $(\sigma^{?-1}, \Theta_p^? = 2)$ is reported in green; in this case the DS method produces posterior estimates for nominal rigidities which are far away and much lower than the benchmark results. DS found that, in general, deep parameters estimates can be sensitive to priors. We conjecture that not only do priors matter, but (mostly) matters how parameters which need to be calibrated are fixed by the researcher; luckily it seems to be possible to find calibrations for which benchmark and macroprior results do coincide, while maintaining an high degree of informativeness by using the DS prior.

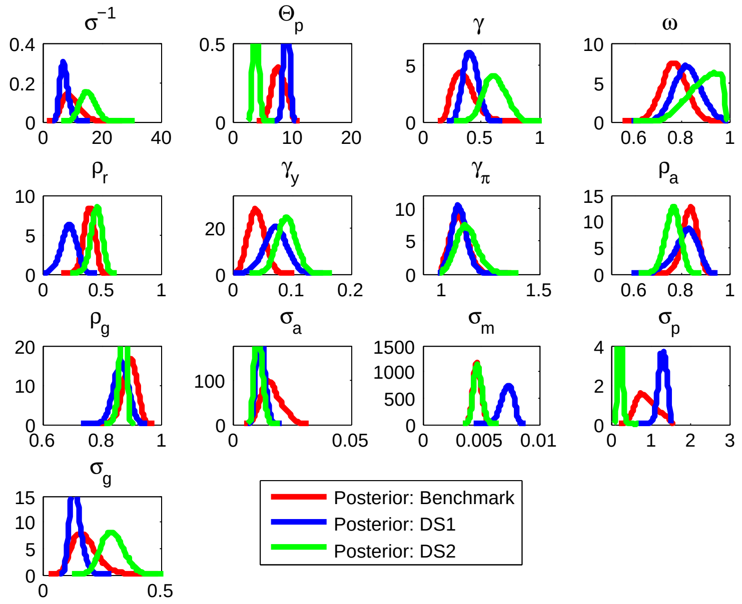


Figure 4: MCMC posterior distribution using the benchmark prior and DS prior. Two different settings of the calibrated rigidities, DS1, 'high' values of calibrated parameters, DS2, 'low' values of calibrated parameters.

4.3 IRF-priors: Second block

In our second experiment IRF-priors are used for a different block of parameters, namely the average length of price adjustment, the elasticity of labor supply, the degree of inflation indexation and the smoothing parameter in the Taylor rule: $\xi_\dagger \equiv [\Theta_p, \gamma, \omega, \rho_r]$.

For the IRF prior we impose (roughly) the same prior means as in the benchmark: means and quantiles are reported below.

Parameters	0.01 perc.	Prior Mean	0.99 perc.
Θ_p	1.01	2.11	7.03
γ	0.01	1.06	2.03
ω	0.06	0.49	0.92
ρ_r	0.01	0.43	0.78

with the kernel density plots shown in figure 5:[20]

Also in this case the strenght of the prior differs across parameters; differently from the previous block, the difference between setting $K = 0.5$ or a tighter $K = 0.05$ appears to be rather limited. The correlation structure of this prior turns out to be as follows:

Parameters	Θ_p	γ	ω	ρ_r
Θ_p	1.00	0.78	0.91	0.73
γ	0.78	1.00	0.96	0.98
ω	0.91	0.96	1.00	0.94
ρ_r	0.72	0.98	0.94	1.00

For what concerns the estimation, we follow the steps described in section 3.1; calibrated parameters were set according to $\xi_b^? = \xi_b^*$, i.e. according to the hyperparameters of the benchmark prior; no relevant difference was found by changing calibrated parameters both in the target and IRF implied by the metropolis draws.

From our second experiment we report changes in posterior estimates with respect to the benchmark prior case, for some parameters regarding the stan-

[20]To show how overall prior tightness operates we show both $K = 0.5$ and $K = 0.05$

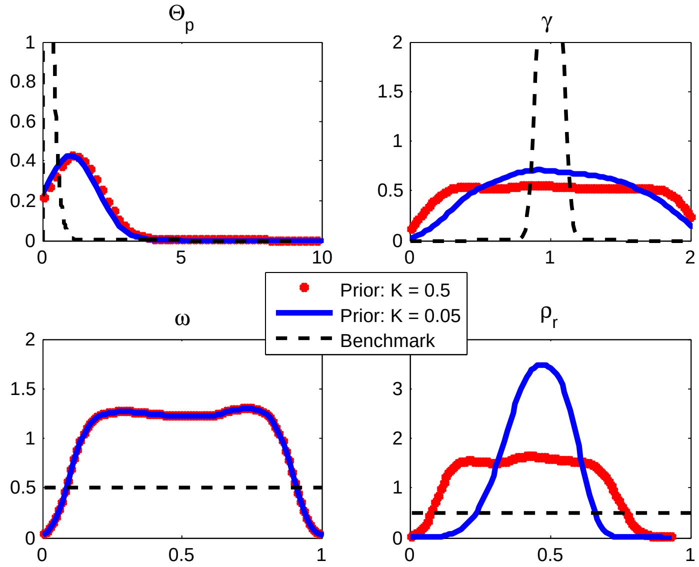


Figure 5: Block II. Univariate kernel density (Note that Θ_p is cut at one, kernel estimates do not take censoring into account)

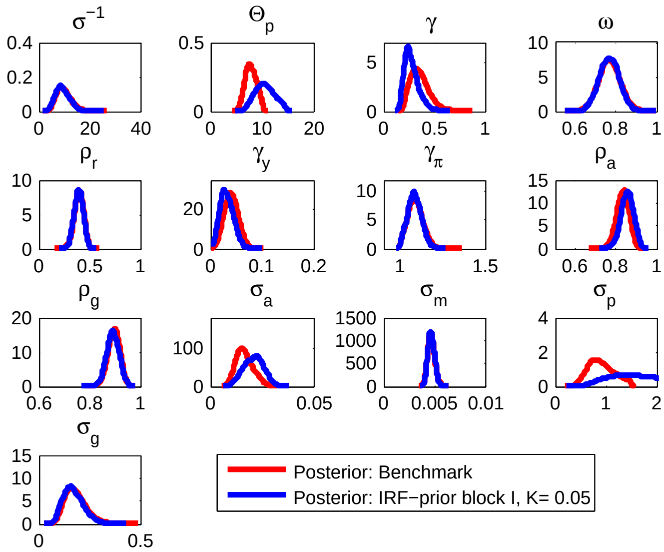


Figure 6: IRF prior on block II. Posteriors: univariate kernel density

dard deviation of the exogenous states: technology and more dramatically, the mark-up shock.

Among the structural parameter the most relevant changes concerned the amount of nominal rigidities Θ_p and the elasticity of labor supply γ. For this block no sizeable difference was found between the case $K = 0.5$ and $K = 0.05$

The change in the posterior estimates produced by a change in the priors seems to follow some meaningful pattern: higher nominal rigidities (Θ_p) are met by a lower elasticity of labor supply γ and by larger mark-up shocks σ_p. This can be better understood by looking at how the Phillips curve and the marginal costs are specified in the model. After some trivial algebraic manipulation we see that,

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa_{pp}(\theta_p)\left(mc_t + e_p\right), \quad (16)$$

$$mc_t = \left(\frac{1}{\sigma} - 1 + \frac{\gamma + 1}{1-\delta}\right) y_t - \frac{\gamma+1}{1-\delta} a_t - g_t \quad (17)$$

the impact of the cost push shock e_p is loaded into inflation π_t by a term κ_{pp} which is a negative function of the expected duration of price rigidity (Θ_p). For given expectations of future inflation, more rigid prices will need a larger variance of e_p in order to match the same variance of inflation present in the data. The elasticity of labor supply enters instead in the definition of marginal costs; a lower value of the elasticity γ helps in keeping marginal costs lower after shocks to output. A lower estimate of γ would then imply an higher variance of technology shocks (σ_a), since technology a_t is multiplied by the parameter γ in equation (17). The specific changes in posterior estimates under the two priors might suggest a problem of identification: deep parameters are only jointly identified in the model, but individually they are not well identified.

A possible interpretation of the difference in posterior estimates would then be that a correlated prior would effectively twist estimates by imposing a stronger correlation structure among parameters, when those are not well identified individually. A different interpretation would be that the model is misspecified and the structure of dependence in the prior exacerbates the problem by imposing the cross equation restrictions of the DSGE model as prior information. In order to get some insights about which of the two explanations above appears more plausible, we simulated artificial data from the model and we estimated it by recurring to both our IRF-prior and the benchmark independent prior. Data is simulated by setting parameters consistently with the numerical posterior mode on real data.

We contrast results in figure 7: once misspecification is controlled for, the difference between the two posterior estimates is greatly reduced with respect to what was apparent from figure 4.3, and appears to affect almost only the parameter σ_p.[21] This can suggest that misspecification is to blame for the instability of the estimates. As a check for misspecification, 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.

[21] Some caution should 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.

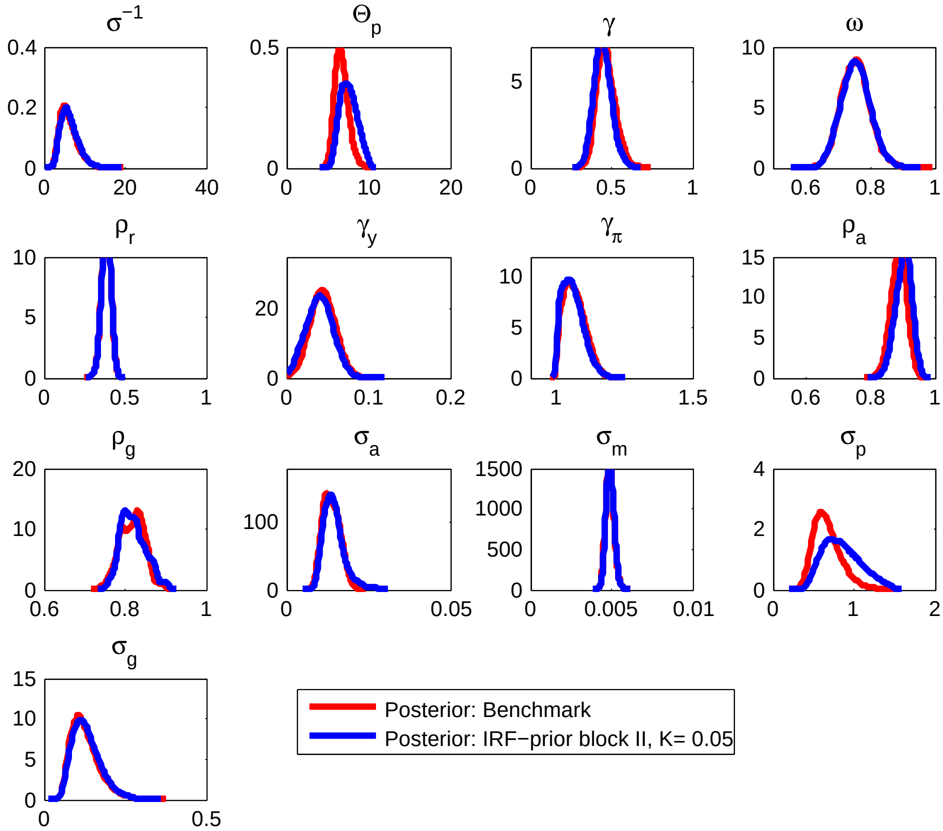


Figure 7: II block and independent prior benchmark: simulated data

Shocks	Technology	Monetary	Cost Push	Preference
Technology	1.00	-0.55	0.31	0.01
Monetary	-0.55	1.00	-0.57	0.16
Cost Push	0.31	-0.57	1.00	0.014
Preference	0.01	0.16	0.01	1.00

It is easy to see that all shocks are highly cross-correlated with the only exception of the preference shock, which is a sign of misspecification.[22]

[22]We assessed confidence bands for the correlation matrix by undertaking a little Monte Carlo

5 Conclusions

In this paper we studied the effect of joint priors on posterior estimates for DSGE models. We introduced a new macro prior based on impulse response functions ('IRF-prior') and we compared posterior estimates obtained using ours against a benchmark microprior.

The paper has two main implications concerning the use of macropriors. First, when a macroprior is adopted, posterior distributions depend upon parameters which need to be calibrated. The way researchers calibrate those parameters is not neutral for results. We show this to be the key to understand the Del Negro and Schorfheide (2008) result that posterior estimates change when a joint prior on exogenous states parameters is adopted. Once we offset the role of calibrated parameters, a joint prior on exogenous state parameters seems to have a weak or negligible influence on the posterior estimates of the parameters which regulate the amount of nominal rigidities in the economy. We confirm this finding by using both our IRF-prior and the DS prior.

Second, we verify that applying our IRF-prior on a key set of parameters can twist the posterior estimates even after calibrated parameters are taken into account. The change in the posterior estimates mostly concerns the Calvo probability of price adjustment and the variance of mark-up shocks. By the means of a simulated dataset we argue that this result could be attributed to misspecification of the DSGE model.

Appendices

A. The problem of intractable normalizing constants

Consistently with our definitions of the IRF-prior, the Metropolis-Hastings Ratio which defines the probability of accepting a candidate draw ξ^c over the previous ξ^i is given by:

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.

$$H(\xi^c \mid \xi^i) = \frac{L(X \mid \xi^c) g(\xi^i \mid \xi^c) w(\xi^c_b, \xi^c_{irf})}{L(X \mid \xi^i) g(\xi^c \mid \xi^i) w(\xi^i_b, \xi^i_{irf})} \frac{w(\xi^i_b)}{w(\xi^c_b)}, \tag{18}$$

where L denotes the likelihood function from the DSGE model (X are the data) and g is the proposal or candidate distribution; when using a random walk as a proposal, this term cancels out. The kernel w rather than its proper version $\bar{w}$ is adopted, as its normalizing constant would be eliminated in the ratio, as standard in an MCMC approach. To simplify notation we omit in the expression the term related to the benchmark prior. The last term in the expression is the ratio between 'normalizing constants'; it is not analytically available and it varies over the draws c and i so the ratio cannot be computed.

Moeller, Pettitt, Berthelsen, and Reeves (2006) show that this type of problem can be tackled by introducing an auxiliary variable x which is defined over the same space as $\xi_\dagger$, distributed with a conditional density $f(x \mid \xi_\dagger)$ chosen by the researcher. Introduce then a proposal density $\pi(x \mid \xi_b)$ (indepedent to g) to be used in the Metropolis sampler, as follows:

$$\pi(x \mid \xi_b) = \frac{w(x, \xi_b)}{w(\xi_b)}, \tag{19}$$

where $w(x, \xi_b)$ is in our case the joint distribution of x, ξ_b as induced by our IRF-prior. The marginal density $w(\xi_b)$ was defined before. Now consider estimating the joint set ξ, x by a Metropolis sampler. Its ratio would be given by:

$$H(\xi^c, x^c \mid \xi^i, x^i) = \frac{L(X \mid \xi^c) g(\xi^i \mid \xi^c) w(\xi^c_b, \xi^c_{irf})}{L(X \mid \xi^i) g(\xi^c \mid \xi^i) w(\xi^i_b, \xi^i_{irf})} \frac{w(\xi^i_b)}{w(\xi^c_b)} \frac{f(x^c \mid \xi^c_b) \pi(x^i \mid \xi^i_b)}{f(x^i \mid \xi^i_b) \pi(x^c \mid \xi^c_b)}, \tag{20}$$

now plug (19) in (20) to remove the normalizing constants. Assuming also a random walk for the g we obtain:

$$H(\xi^c, x^c \mid \xi^i, x^i) = \frac{L(X \mid \xi^c) w(\xi^c_b, \xi^c_{irf})}{L(X \mid \xi^i) w(\xi^i_b, \xi^i_{irf})} \frac{f(x^c \mid \xi^c_b)}{f(x^i \mid \xi^i_b)} \frac{w(x^i, \xi^i_b)}{w(x^c, \xi^c_b)}, \tag{21}$$

the solution comes as one assumes that the distribution followed by the auxiliary variable x is given by:

$$f(x^c \mid \xi_b) = \frac{w(x^c|\xi_b = \xi_b^?)}{w(\xi_b^?)},$$

where, as in the main text, $w(x^c, \xi_b = \xi_b^?)$ denotes the IRF joint prior kernel when ξ_b is at some fixed value $\xi_b^?$. The final expression for the metropolis ratio is given by:

$$H(\xi^c, x^c \mid \xi^i, x^i) = \frac{L(X \mid \xi^c)w(\xi_b^c, \xi_{irf}^c)}{L(X \mid \xi^i)w(\xi_b^i, \xi_{irf}^i)} \frac{w(x^c \mid \xi_b = \xi_b^?)}{w(x^i \mid \xi_b = \xi_b^?)} \frac{w(x^i, \xi_b^i)}{w(x^c, \xi_b^c)}, \tag{22}$$

(22) reduces to the shortcut outlined by DS(2008) only if $x^i = \xi_{irf}^i$ for all i.

B. The Model

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 that it yields more simple and transparent results. Larger models would still suffer from misspecification problems, albeit probably less than smaller ones, but on top of that, they have more parameters and this can make identification more difficult. Overall, we believe that the model we chose represents a good trade-off between transparency and realism for our analysis.

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

$$\frac{1}{\sigma}y_t = \frac{1}{\sigma}E_t y_{t+1} - (r_t - E_t\pi_{t+1} + E_t g_{t+1} - g_t);$$

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa_{pp}(mc_t + e_p);$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\gamma_\pi \pi_t + \gamma_y y_t) + e_m;$$

$$y_t = a_t + (1-\delta)n_t;$$

$$mc_t = rw_t + n_t - y_t;$$

$$rw_t = y_t \frac{1}{\sigma} + \gamma n_t - g_t;$$

The complete set of variables is given by: $\{y_t, mc_t, rw_t, \pi_t, r_t, n_t\}$, respectively the GDP, marginal costs, real wages, inflation rates, interest rates, the amount of hours worked. There are four shocks in this economy, technology a, intertemporal preferences g, monetary e_m and mark up shocks e_p; the first two autoregressive processes, while the latter are i.i.d.

$$a_t = \rho_a a_{t-1} + e_a,$$

$$g_t = \rho_g g_{t-1} + e_g.$$

The first three equations are the Euler equation, the (backward looking) Phillips curve and a standard Taylor rule. The backward looking component in the Phillips curve is derived from the assumption that non updating producers partially index their prices as a function of past inflation. The remaining equations are 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:

$$\gamma_b = \omega/(1+\omega\beta)$$
$$\gamma_f = (\beta/(1+\omega\beta));$$

$$\kappa_{pp} = \frac{(1-\delta)(1-\theta_p\beta)(1-\theta_p)}{(\theta_p(1+\delta(\epsilon-1)))(1+\omega\beta)},$$

ω is the degree of backward looking indexation in the Phillips curve, β is the discount factor (calibrated at 0.99), θ_p is the probability of price adjustment in the Calvo model, δ is the share of capital in the Cobb Douglas production function (0.36) and ϵ is the elasticity of substitution among varieties in the bundle of commodities (6).

C. Some further details concerning the DS prior

The DS prior uses information about sample moments, elicited from a presample of data. IRFs and moments are closely related concepts: they coincide when impulse responses from all shocks are considered. Differently from simulated moments, the advantage of IRFs is that they allow the researcher to draw prior information only from the shocks whose behaviour is most known to researchers, even if they contribute to the overall data moments by little. A good example of that is the monetary policy shock. In general, IRFs are the most studied objects in macroeconomics and potentially allow to better draw information from prior knowledge.

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.[23] The VAR approximation of the DSGE is denoted as follows:

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

or in matrix notation

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

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

The Φs are numerically constructed using the OLS regression coefficients formulas from DSGE-simulated data moments and they depend upon the whole vector of structural parameters ξ. 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:

$$L(Y, X \mid \xi) =\mid \Sigma(\xi) \mid^{-(T^*+n-1)/2} \exp\left(\frac{T^*}{2} tr[\Sigma(\xi)]^{-1}(\Gamma_{YY} - 2\Phi(\xi)\Gamma_{XY} + \Phi(\xi)'\Gamma_{XX}\Phi(\xi \tag{23}$$

where $\Sigma(\xi)$ is the theoretical variance covariance matrix of the data implied by the DSGE, T^* is a scale factor for the strength of the prior. Γs 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. Expression (23) provides the kernel of the DS prior concerning parameters which

[23]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.

Table 2: DS Prior statistics: $\sigma^{?-1}, \Theta_p^?$ at 2

Parameters	0.01 perc.	Prior mean	0.99 perc.
ρ_a	0.6464	0.8284	0.9856
ρ_g	0.6051	0.7954	0.9691
σ_a	0	0.0515	0.2176
σ_m	0	0.1411	0.2944
σ_g	0	0.0504	0.2633
σ_p	0.0408	0.2326	0.3860

Table 3: DS Prior statistics: $\sigma^{?-1}, \Theta_p^?$ at 5

Parameters	0.01 perc.	Prior Mean	0.99 perc.
ρ_a	0.5640	0.7503	0.9223
ρ_g	0.5927	0.7933	0.9367
σ_a	0	0.0606	0.2030
σ_m	0	0.0606	0.2172
σ_g	0.0321	0.2172	0.3739
σ_p	0	0.1197	0.2554

are informed by the prior. When only a subset of the deep parameters ξ_{ds} is informed by the DS prior, the kernel above should be multiplied with a microprior on ξ_b. The same issue of calibrated parameters as in IRF-prior also arises when computing (23). The rationale underlying the DS prior stems from a dummy prior approach, see Sims (2008), where the researcher constructs an augmented sample which combines the original data with some 'dummy' observations which have been simulated from the model. 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 directly deep parameters.

The table below reports some quantiles and the mean of the DS prior as applied to our pre-sample. We sampled parameters from a uniform distribution and computed (23) at each draw: those are then weights for a weighted kernel density estimation. The prior is computed for two different values of calibrated parameters.

It is interesting to note that variances change widely when different values of the calibrated 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.

D. Detailed estimation results

Table 4: Numerical mode results

Parameters	Numerical posterior mode
σ^{-1}	8.5361
Θ_p	7.2722
γ	0.3735
ω	0.7689
ρ_r	0.3943
γ_y	0.0415
γ_π	1.0901
ρ_a	0.8332
ρ_g	0.9001
σ_a	0.0148
σ_m	0.0047
σ_g	0.1563
σ_p	0.7501

Table 5: Posterior: benchmark priors

Parameters	0.01 perc.	0.99 perc.	Posterior mean
σ^{-1}	4.6972	17.7151	9.8850
Θ_p	5.5936	10.1753	7.9071
γ	0.2077	0.6842	0.3630
ω	0.6456	0.9082	0.7701
ρ_r	0.2777	0.4926	0.3950
γ_y	0.0115	0.0724	0.0408
γ_π	1.0029	1.2097	1.0890
ρ_a	0.7543	0.9036	0.8329
ρ_g	0.8409	0.9424	0.8961
σ_a	0.0085	0.0258	0.0165
σ_m	0.0040	0.0057	0.0047
σ_g	0.0874	0.3300	0.1818
σ_p	0.4454	1.4693	0.9167

Table 6: Posterior: IRF-prior (First block)

Parameters	0.01 perc.	0.99 perc.	Posterior mean
$\sigma^{-}1$	4.8299	18.6660	10.3330
Θ_p	5.5119	10.0485	7.6757
γ	0.1758	0.6198	0.3650
ω	0.6466	0.8961	0.7657
ρ_r	0.2814	0.4959	0.3948
γ_y	0.0142	0.0735	0.0425
γ_π	1.0036	1.2060	1.0918
ρ_a	0.7504	0.8955	0.8314
ρ_g	0.8161	0.9416	0.8921
σ_a	0.0092	0.0315	0.0165
σ_m	0.0040	0.0056	0.0047
σ_p	0.4285	1.4680	0.8665
σ_g	0.0909	0.3332	0.1885

Table 7: Posterior: IRF-prior (Second block)

Parameters	0.01 perc.	0.99 perc.	Posterior mean
$\sigma^{-}1$	4.4469	17.3490	9.5278
Θ_p	6.8017	14.3987	10.5628
γ	0.1664	0.5168	0.2802
ω	0.6540	0.8880	0.7669
ρ_r	0.2892	0.4901	0.3932
γ_y	0.0063	0.0678	0.0322
γ_π	1.0045	1.1920	1.0844
ρ_a	0.7717	0.9229	0.8550
ρ_g	0.8319	0.9491	0.8921
σ_a	0.0108	0.0323	0.0211
σ_m	0.0040	0.0055	0.0047
σ_p	0.6504	2.9236	1.6528
σ_g	0.0831	0.3150	0.1729

Table 8: Posterior Estimates with DS: ($\sigma^{?-1}$, $\Theta_p^?$ at 2)

Parameters	0.01 perc.	0.99 perc.	Posterior Mean
σ^{-1}	10.3426	23.6727	15.7359
Θ_p	3.1205	4.9284	3.9258
γ	0.4376	0.8156	0.6270
ω	0.7361	0.9786	0.8935
ρ_r	0.3380	0.5534	0.4539
γ_y	0.0594	0.1315	0.0930
γ_π	1.0176	1.2788	1.1297
ρ_a	0.6856	0.8270	0.7601
ρ_g	0.8435	0.8928	0.8724
σ_a	0.0080	0.0149	0.0109
σ_m	0.0040	0.0057	0.0048
σ_g	0.2015	0.4277	0.2993
σ_p	0.1438	0.3621	0.2282

Table 9: Posterior Estimates with DS: ($\sigma^{?-1}$, $\Theta_p^?$ at 5)

Parameters	0.01 perc.	0.99 perc.	Posterior Mean
σ^{-1}	4.7907	11.6364	7.3624
Θ_p	8.1756	10.1361	9.1578
γ	0.2786	0.5678	0.4165
ω	0.6984	0.9504	0.8188
ρ_r	0.0862	0.3692	0.2263
γ_y	0.0264	0.1202	0.0740
γ_π	1.0019	1.1973	1.0908
ρ_a	0.6520	0.9066	0.8142
ρ_g	0.8087	0.9217	0.8646
σ_a	0.0085	0.0173	0.0119
σ_m	0.0060	0.0082	0.0073
σ_g	0.0962	0.2134	0.1401
σ_p	1.0877	1.4887	1.3153

Table 10: Benchmark priors: simulated data

Parameters	0.01 perc.	0.99 perc.	Posterior mean
σ^-1	2.6439	12.9031	6.1677
Θ_p	5.1776	9.1008	6.7599
γ	0.3489	0.6236	0.4664
ω	0.6568	0.8689	0.7554
ρ_r	0.3141	0.4448	0.3836
γ_y	0.0092	0.0814	0.0444
γ_π	0.9997	1.1653	1.0627
ρ_a	0.8281	0.9420	0.8888
ρ_g	0.7483	0.8920	0.8176
σ_a	0.0088	0.0199	0.0138
σ_m	0.0044	0.0055	0.0049
σ_p	0.3809	1.2081	0.6743
σ_g	0.0590	0.2625	0.1284

E. Convergence diagnostics

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,

Table 11: IRF-prior on block II: K = 0.05 (simulated data)

Parameters	0.01 perc.	0.99 perc.	Posterior mean
$\sigma^{-}1$	2.6833	13.0484	6.3809
Θ_p	5.6056	9.9593	7.5941
γ	0.3333	0.5923	0.4474
ω	0.6575	0.8691	0.7562
ρ_r	0.3169	0.4470	0.3844
γ_y	0.0029	0.0798	0.0405
γ_π	0.9989	1.1708	1.0583
ρ_a	0.8407	0.9547	0.9030
ρ_g	0.7601	0.9014	0.8206
σ_a	0.0092	0.0258	0.0146
σ_m	0.0044	0.0055	0.0049
σ_p	0.4497	1.4290	0.8557
σ_g	0.0591	0.2635	0.1320

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 8: Cumsum plots : benchmark priors

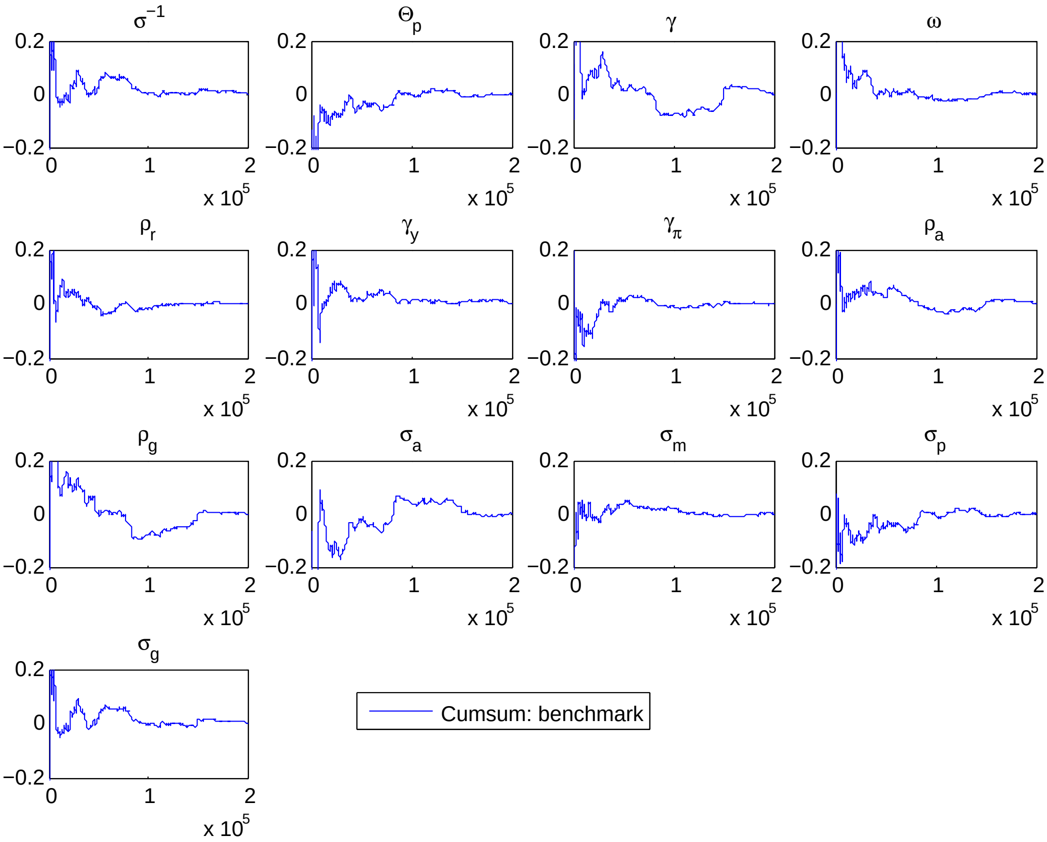


Figure 9: Cumsum plots : DS prior, DS1 experiment

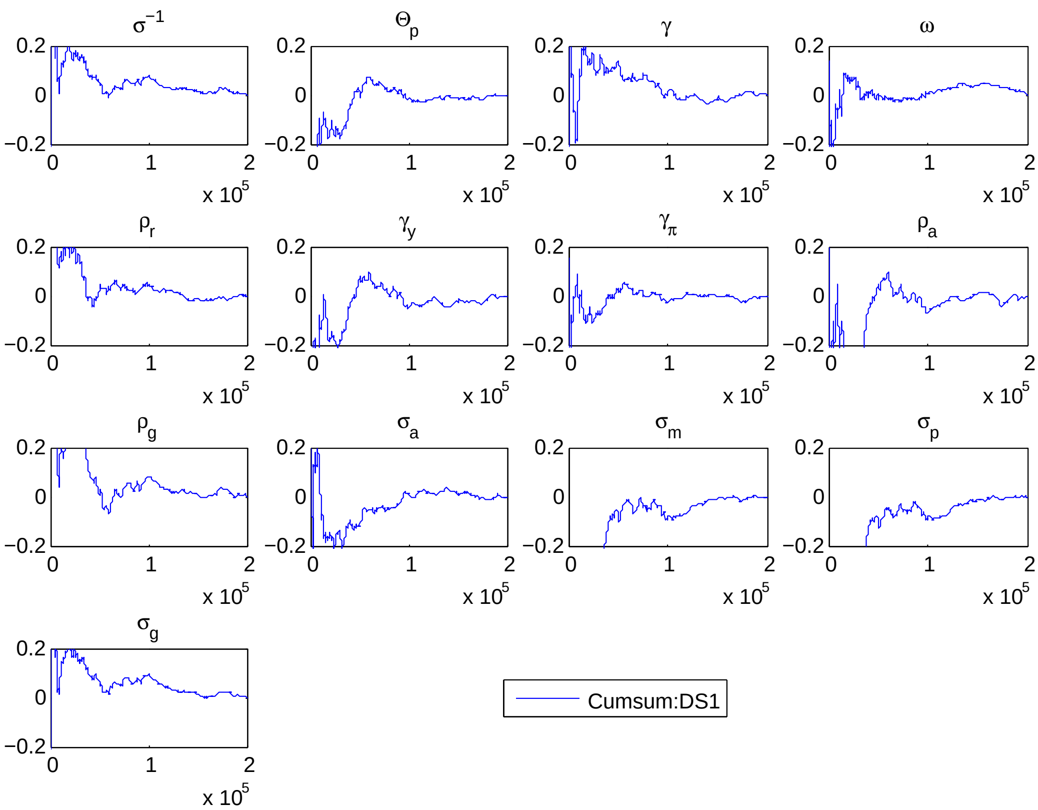


Figure 10: Cumsum plots : DS prior, DS2 experiment

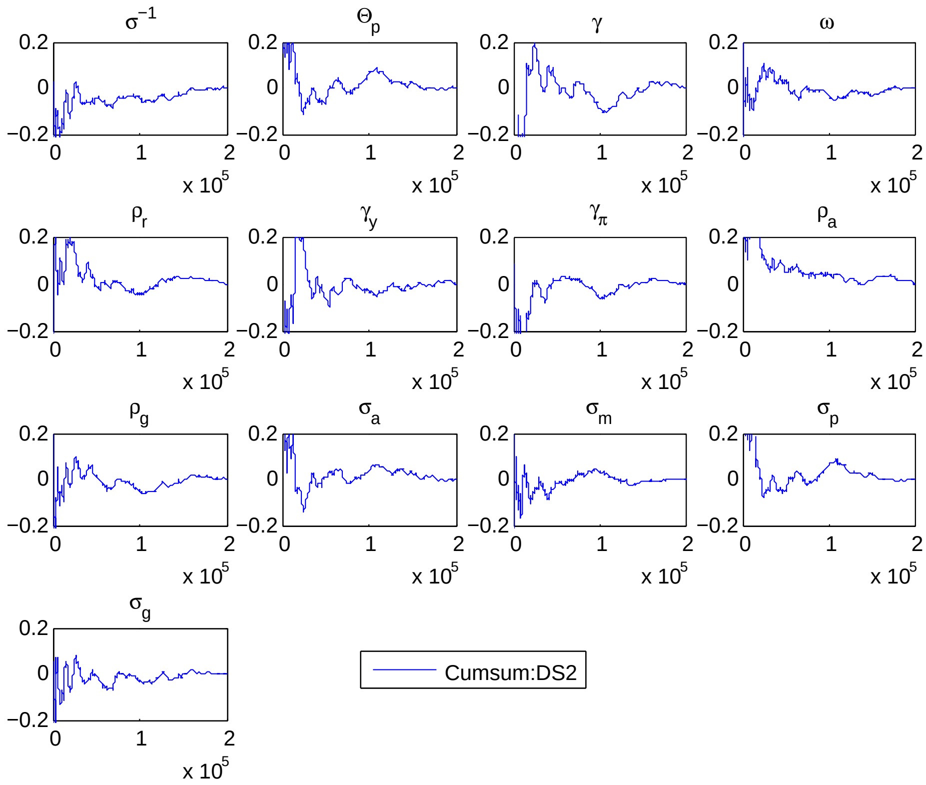


Figure 11: Cumsum plots: IRF-priors, block 1

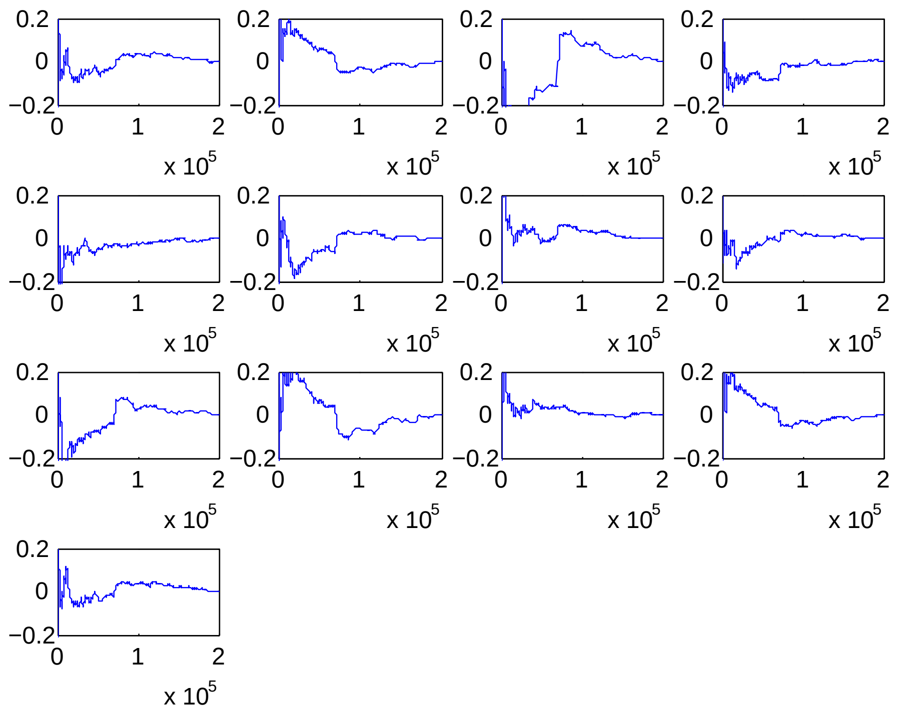
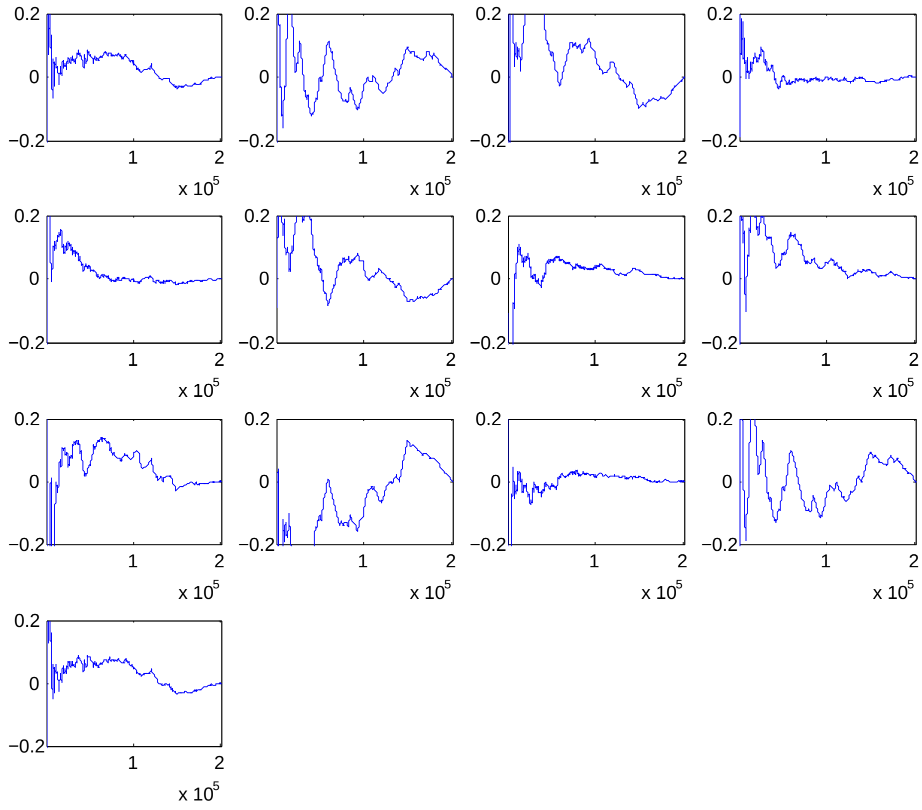


Figure 12: Cumsum plots: IRF-priors, block 2



Bibliography

AN, S., AND F. SCHORFHEIDE (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, pp. 187–192.

CANOVA, F., AND L. SALA (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56(4), 431–449.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

DEL NEGRO, M., AND F. SCHORFHEIDE (2008): "Forming Priors for DSGE Models and How it Matters for Nominal Rigidities," *Journal of Monetary Economics*, 55(7), 1191–1208.

DURBIN, J., AND S. J. KOOPMAN (2001): *Time series analysis by state space methods*. Oxford University Press.

KOOP, G., M. H. PESARAN, AND S. M. POTTER (1996): "Impulse Response Analysis in Nonlinear Multivariate Models," *Journal of Econometrics*, 114(74), 119–147.

MOELLER, J., A. N. PETTITT, K. K. BERTHELSEN, AND R. W. REEVES (2006): "An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants," *Biometrika*, 93, 451–458.

RABANAL, P., AND J. F. RUBIO-RAMIREZ (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, 52, 1151–1166.

SALTELLI, A., S. TARANTOLA, F. CAMPOLONGO, AND M. RATTO (2004): *Sensitivity Analysis in Practice*. Wiley-VCH.

SCHORFHEIDE, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15(6), 645–70.

SIMS, C. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2).

SIMS, C., AND T. ZHA (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113–1155.

SMETS, F., AND R. WOUTERS (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

TAYLOR, J. (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.

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