DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Estimation of DSGE Models under Diffuse Priors and Data-Driven Identification Constraints

Markku Lanne and Jani Luoto

CREATES Research Paper 2015-37

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Estimation of DSGE Models under Diffuse Priors and Data-Driven Identification Constraints*

Markku Lanne
University of Helsinki and
CREATES

Jani Luoto
University of Helsinki

Abstract

We propose a sequential Monte Carlo (SMC) method augmented with an importance sampling step for estimation of DSGE models. In addition to being theoretically well motivated, the new method facilitates the assessment of estimation accuracy. Furthermore, in order to alleviate the problem of multimodal posterior distributions due to poor identification of DSGE models when uninformative prior distributions are assumed, we recommend imposing data-driven identification constraints and devise a procedure for finding them. An empirical application to the Smets-Wouters (2007) model demonstrates the properties of the estimation method, and shows how the problem of multimodal posterior distributions caused by parameter redundancy is eliminated by identification constraints. Out-of-sample forecast comparisons as well as Bayes factors lend support to the constrained model.

Keywords: Particle filter, importance sampling, Bayesian identification
JEL Classification: D58, C11, C32, C52.

*Financial support from the Academy of Finland is gratefully acknowledged. The first author also acknowledges financial support from CREATES (DNRF78) funded by the Danish National Research Foundation, while the second author is grateful for financial support from the Research Funds of the University of Helsinki. Part of this research was done while the second author was visiting the Bank of Finland, whose hospitality is gratefully acknowledged.

1 Introduction

Advances in Bayesian simulation methods have recently facilitated the estimation of relatively large-scale dynamic stochastic general equilibrium (DSGE) models. Because the commonly employed random walk Metopolis-Hastings (RWMH) algorithm has a number of shortcomings in this context, alternative algorithms have recently started to gain ground (see, for example, Creal (2007), Chib and Ramamurthy (2010), and Herbst and Schorfheide (2014)).

Besides the well-known problem of potentially very slow convergence to the posterior distribution, the RWMH algorithm is ill-suited for handling multimodality typically arising in poorly identified DSGE models, when diffuse priors are assumed (see, e.g., Koop et al. (2013)). Because of the latter property, relatively tight prior distributions have had to be assumed in the previous literature. An unfortunate consequence of very informative priors is that the resulting posterior distributions may not have much to say about how well the structural model fits the data. In other words, in that case, the priors are likely to be driving the results, precluding us from learning about the parameters of the model from the data.

Herbst and Schorfheide (2014) have recently addressed these issues and employed an adaptive sequential Monte Carlo (SMC) algorithm to estimate the Smets and Wouters (2007) model (SW model hereafter) based on uninformative priors. They also used simulations to assess the accuracy of the SMC algorithm compared to the RWMH algorithm, and not surprisingly, found the former much more accurate. However, their approximation of the asymptotic variances of the parameter estimates, used to measure the approximation error of the estimates, is based on running the SMC algorithm multiple times, which is computationally very costly in the case of a complex high-dimensional DSGE model. Therefore, they computed the numerical standard errors from only 20 runs, and they are thus likely to be very inaccurate.

In this paper, we propose to augment the SMC algorithm with a nonsequential Importance Sampling (IS) step, which has the advantage that numerical standard errors can be readily calculated without burdensome simulations. Moreover, convergence results for nonsequential IS are available, while the asymptotic properties of the adaptive SMC al-

gorithm are not necessarily known. Hence, in addition to being computationally feasible in assessing the accuracy of the estimates, our procedure is theoretically well motivated.

In order to alleviate the multimodality problem discussed above, we propose to impose so-called data-driven identifiabilty constraints and devise a procedure of finding them. This idea of such constraints was put forth by Frühwirth-Schnatter (2001), and to the best of our knowledge, it has not been applied to DSGE models. They may lead to a superior model, and their validity can be checked by comparing the restricted and unrestricted models, say, by Bayes factors. For instance, when estimating the SW model, Herbst and Schorfheide (2014) found two modes, implying different wage dynamics, and by data-driven identification constraints we are able to discard one of them.

We estimate the SW model on the same data set as Smets and Wouters (2007) and with diffuse priors that are slightly different from those assumed by Herbst and Schorfheide (2014). Our augmented SMC method yields very accurate estimates similiar to those of Herbst and Schorfheide (2014) based on both the RWMH and SMC methods and diffuse priors, but very different from those of Smets and Wouters (2007) based on tight prior distributions. We also estimate the model with the tight priors assumed by Smets and Wouters (2007), and find accurate posterior estimates very close to theirs. Comparison of the estimates based on the informative and uninformative priors reveals that the results of Smets and Wouters (2007) are indeed strongly driven by their informative prior distributions.

Closer inspection of the posterior densities of the parameters of the SW model suggests redundancy in the parametrization of the SW model. In particular, restricting the ARMA wage markup shock to a white noise process (and the standard deviation of the price markup shock, σ_p, to a relatively small value $\sigma_p < 0.19$) leads to an improved model, and we are able to conclude that the persistence of wages is driven by the endogenous instead of exogenous dynamics. In other words, by imposing these constraints accepted by the data, we can rule out one of the two modes discussed above, whose relative importance Herbst and Schorfheide (2014) found difficult to assess. Moreover, imposing the data-driven identification constraints improves estimation accuracy, and also out-of-sample forecast comparisons lend strong support to the restricted model.

In the next section, we discuss the SMC algorithm and explain, how any such algorthim

can be augmented with an IS step. The estimation results of the SW model are reported in Subsection 2.2, where they are also compared to the corresponding results in Smets and Wouters (2007), and Herbst and Schorfheide (2014). The data-driven identification constraints and the procedure for finding them, as well as the related empirical results are the topic of Section 3. Forecasting with DSGE models and computing log predictive densities that can be used to rank competing models, are discussed in Section 4, while in Subsection 4.1, we report the results of forecast comparison. Finally, Section 5 concludes.

2 Sequential Monte Carlo Estimation

The likelihood function of any DSGE model is a complicated nonlinear function of the vector of its structural parameters θ, and, hence, potentially badly behaved. This complicates the estimation of the posterior distribution of the parameters and, if not properly handled, makes commonly used approaches such as the RWMH algorithm ill-suited for this purpose. One solution is to assume tight prior distributions, but as discussed in the Introduction, this may be problematic for a number of reasons. Therefore, it might be advisable to employ less informative priors, but this, of course calls for an appropriate estimation method.

Recently Herbst and Schorfheide (2014) have proposed using a Sequential Monte Carlo (SMC) algorithm (particle filter) for estimating DSGE models under noninformative priors (see also Creal (2007) for estimating DSGE models by SMC methods). The particle filter is a recursive algorithm that approximates an evolving distribution of interest by a system of simulated particles and their weights. This system is carried forward over time using multiple resampling and importance sampling (IS) steps. By virtue of their construction, these methods have been found to perform well for badly behaved posterior distributions (for further evidence on this, see, for example, Del Moral, Doucet and Jasra (2006), Jasra, Stephens and Holmes (2007), and Durham and Geweke (2014)).

While SMC methods are well suited for estimating DSGE models, they also involve at least two problems to which we propose a solution in this paper. First, to obtain a sequence of good importance densities the algorithm must in practice be tuned online using the current population of particles, and the properties of the estimates calculated

from the output of such adaptive samplers are not necessarily known (see Durham and Geweke (2014) for a detailed discussion). In particular, there are only very few results stating their convergence to the associated posterior quantities.

Another, and from the practical point of view potentially more seriour problem is related to to the assessment of the accuracy of estimates, which is dificult for SMC methods because of the complexity of the asymptotic variances. One solution, applied by Herbst and Schorfheide (2014), among others, is to estimate the asymptotic variance from multiple runs of the SMC algorithm. While the time cost of this approach may be relatively small for simple univariate models, whose likelihoods can be readily calculated in parallel using graphics processing units (GPUs), it can be unreasonably high for complex high dimensional models such as the SW model. For this reason, numerical standard errors, needed to assess the quality of the posterior estimates, have been based on a relatively small number of runs and are thus likely to be very inaccurate (for instance, Herbst and Schorfheide (2014) only run the algorithm 20 times).

2.1 Importance Sampling Step

As discussed in the Introduction, in order to solve the two problems of the SMC approach applied to DSGE models, we propose to conclude the SMC run by nonsequential IS, for which convergence results are available and numerical standard errors can be readily calculated. The idea is to approximate the posterior distibution by a mixture of Student-t distributions. This approximation is then efficiently used as the importance sampling density. The proposed procedure closely resembles that of Hoogerheide, Opschoor, and van Dijk (2012), and it delivers the efficient mixture components by minimizing the Kullback-Leibler divergence from the target posterior, which is approximated by the simulated particles. As IS algorithms are highly parallelizable, the time cost of the proposed procedure is marginal compared to a single SMC run. It is important to notice that we propose to use nonsequential IS as a complement to SMC methods mainly because assessing the numerical accuracy of SMC estimates is challenging. Indeed, it is our experience that adaptive SMC algorithms are able to deliver reliable approximations of even very pathological posterior distributions also in cases where no formal asymptotic convergence result

is available.

Any SMC algorithm can be augmented by an importance sampling step. For a description of the SMC algorithm as applied to a DSGE model, including the SW model, we refer the interested reader to the Appendix, or, e.g., Herbst and Schorfheide (2014). Here we only explain how to construct a mixture of Student-t distributions approximation to the target posterior distribution $\pi_n(\theta)$ (for data up to any period $n \in \{1, \ldots, T\}$), from its particle approximation $\{\theta_n^i\}_{i=1}^N$. This mixture density is then used to calculate the Importance Sampling (IS) estimates of the posterior quantities of interest. The proposed procedure closely resembles that of Hoogerheide et al. (2012), and we refer to their paper for a more detailed discussion (see also Cappé et al. (2008)).

The posterior distribution of the parameters of interest θ is approximated by a mixture of J multivariate t distributions:

$$f(\theta \,|\psi) = \sum_{j=1}^{J} \alpha_j t_k(\theta \,|\mu_j, V_j; \nu_j), \tag{1}$$

where $t_k(\theta \,|\mu_j, V_j; \nu_j)$ is the density function of the k-variate t distribution with mode μ_j, (positive definite) scale matrix V_j, and degrees of freedom ν_j, $\psi = (\mu_1', \ldots, \mu_J',\ \mathrm{vech}(V_1)', \ldots, \mathrm{vech}(V_J)', \nu_1, \ldots, \nu_J, \alpha_1, \ldots, \alpha_J)'$, and the mixing probabilities α_j sum to unity.

To obtain an efficient importance sampling density that enables us to accurately approximate the posterior, we minimize the Kullback–Leibler divergence $\int \pi_n(\theta) \log \frac{\pi_n(\theta)}{f(\theta|\psi)} d\theta$ between the target posterior distribution $\pi_n(\theta)$ and $f(\theta \,|\psi)$ in (1) with respect to ψ. Because the elements of vector ψ do not enter the posterior density $\pi_n(\theta)$, this is equivalent to maximizing

$$\int [\log f(\theta \,|\psi)]\, \pi_n(\theta)\, d\theta = E[\log f(\theta \,|\psi)], \tag{2}$$

where E denotes the expectation with respect to the posterior distribution $\pi_n(\theta)$. A simulation-consistent estimate of expression (2) is given by

$$\frac{1}{N} \sum_{i=1}^{N} \log f\left(\theta^i \,|\psi\right), \tag{3}$$

where the particle approximation $\{\theta^i\}_{i=1}^N$ is taken as a sample from the posterior distribution $\pi_n(\theta)$. Following Hoogerheide et al. (2012), we use the Expectation Maximization

(EM) algorithm to maximize (3) with respect to the parameters of the mixture distribution ψ in (1). Once the importance sampling density has been obtained, it can be used to estimate any posterior quantity of interest, and the associated numerical standard errors.

Hoogerheide et al. (2012) maximize a weighted variant of (3) in their bottom-up procedure, which iteratively adds components to the mixture (1), starting with one multivariate t distribution. Conversely, we start with a reasonably large number of distributions and remove the (nearly) singular ones (i.e., those with (nearly) singular covariance matrices and very small probability weights). This can be done because the particle approximation $\{\theta^i\}_{i=1}^N$ provides a very accurate description of the posterior distribution. In other words, in our case it is sufficient to represent the information in $\{\theta^i\}_{i=1}^N$ in terms of the mixture distribution in (1).

It is worth noting that the EM algorithm may be sensitive to the starting values of ψ. To solve the problem, we partition the particle approximation $\{\theta^i\}_{i=1}^N$ into J clusters by an agglomerative hierarchical clustering algorihtm (see, for example, Everitt et al. (2011)), and then use the sample mean and covariance matrix of the particles in the jth cluster as initial values for μ_j and V_j ($j \in 1, \ldots, J$).[1] A good initial value for the mixing probability α_j can be obtained by dividing the number of points in the jth cluster by N. This procedure tends to be quick and very reliable.

2.2 Estimation Results

We estimate the SW model on the same quarterly data set as Smets and Wouters (2007) using an SMC algorithm augmented with a nonsequential IS step, and compare the results to the RWMH estimates of Smets and Wouters (2007), and the RWMH and SMC estimates of Herbst and Schorfheide (2014). The SMC algorithm is described in the Appendix.

While Smets and Wouters (2007) assumed tight prior distributions, in estimation by SMC methods less informative priors are entertained. In particular, for the parameters that are defined on the unit interval, we consider logit transformations and assume univari-

[1]Prior to clustering, the particle approximation $\{\theta^i\}_{i=1}^N$ is normalized and orthogonalized such that the elements of θ have zero means and unit variances, and are uncorrelated. The distance between two normalized and orthogonalized particles is measures by the Euclidean distance between them. Moreover, the distance between two clusters is measured by Ward's measure.

ate normal prior distributions for the transformed parameters. We operate on transformed parameters in order to enhance the performance of the sampler. In contrast, Herbst and Schthorfheide (2014) operate on the original parameters and assume uniform prior distributions for the parameters on the unit interval. In line with Smets and Wouters (2007), we assume inverse-gamma prior distributions for the standard errors of the innovations of the structural shocks. However, we set the prior hyperparameters such that the prior means and variances equal 0.5 and 1, respectively, instead of the values 0.1 and 4 that they considered. The differences between the priors are clearly visible in the density plots depicted in Figure 1. More than 90% of the probability mass of the prior distribution of Smets and Wouters (2007) lies below 0.2, which suggests rather small standard errors for the innovations. It is also worth noting that this prior distribution has a relatively large variance, 4, because of its very long but thin right tail. Our prior has a smaller variance, 1, but it is considerably less leptokurtic than theirs. As far as the other parameters of the model are concerned, we follow Herbst and Schthorfheide (2014) in scaling the associated prior variances of Smets and Wouters (2007) by a factor of three. Our diffuse priors are described in detail in Table 1.

In Table 2, we report the estimation results based on our procedure with diffuse priors. The posterior means, their standard deviations and the 5th and 95th percentiles of the posterior densities of all parameters lie very close to those that Herbst and Schorfheide (2014) obtained by both the RWMH and SMC methods assuming their diffuse priors, but different from those of Smets and Wouters (2007) based on tight prior distributions. The numerical standard errors of the posterior means of the parameters reported in the rightmost column of Table 2 are remarkably small, indicating very accurate estimation. They are also much smaller than those reported by Herbst and Schorfheide (2014), which may reflect the poor quality of their asymptotic variance estimates based on only 20 simulations rather than the superiority of our estimation procedure augmented with the IS step. In any case, our results suggest that concluding the SMC run by nonsequential IS is worthwhile as far as assessing the accuracy of posterior estimates is concerned. It is, however, important to note that all the SMC and IS estimates are very close to each other.

For comparison, in Table 3 we report the results obtained by our procedure augmented

with the importance sampling step assuming the informative priors of Smets and Wouters (2007). The posterior estimates are very close to those in Smets and Wouters (2007), and their numerical standard errors are also very small, indicating great estimation accuracy. By comparing the results of Table 3 to those in Table 2, we can see that the marginal posteriors of Smets and Wouters (2007) are strongly driven by their informative prior distributions.

3 Identification

Especially in models that are known to be poorly identified, such as the SW model, the Monte Carlo (MC) output of the posterior distribution should be systematically analyzed for the information concerning the identification of the parameters of the underlying structural model. To this end, we recommend first visually inspecting the marginal posterior distributions of the parameters for bad behavior (such as multimodality or fat or long tails). Once the badly behaved parameters have been singled out, the next step is to analyze the bivariate posterior density plots of all pairs of these parameters. In our experience, an efficient estimation algorithm produces dense areas in the density plots that may reveal to what extent the associated structural parameters are identified. As will be seen, this information can then be useful in deriving so-called data-driven identifiability constraints on the parameters of the DSGE model (see Frühwirth-Schnatter (2001)). Because these constraints respect the geometry of the data-driven posterior distribution, they help us to learn about the parameters from the data, and are likely to improve the quality of probabilistic forecasts. Moreover, they may be useful in alleviating the multimodality problem often encountered in estimating DSGE models, and they are also likely to improve estimation accuracy. After estimating the restricted model, the procedure may be repeated until the data suggests no further identification constraints.[2]

Inspection of the marginal posterior densities of the parameters reveals six badly be-

[2]For example, in the case of a multimodal posterior density, we may be able to rule out some of the modes (subspaces) by comparing different restricted and unrestricted models using, say, Bayes Factors. This procedure is demonstrated below.

haved parameters that govern wage (ξ_w, ι_w, ρ_w, μ_w) and price (μ_p, σ_p) stickiness.[3] Also Herbst and Schorfheide (2014) found multimodal features in the posterior distributions of these parameters. The bivariate density plots of all combinations of the marginal posteriors of these parameters are next inspected to find out about the identification of these parameters. As an example, Figure 2 depicts the density plot of the ARMA parameters ρ_w and μ_w of the wage mark-up shock $\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w$, where $\eta_t^w \sim N(0, \sigma_w^2)$. It indicates weak identification of these parameters. In particular, they take almost equal values with high probability, giving rise to potential for redundant parameterization. Given the properties of the ARMA model, with $\rho_w \approx \mu_w$, the wage mark-up shock might be better described as a white noise process (with ρ_w and μ_w restricted to zero). In addition, we consider two alternative constraints, namely, $\rho_w = 0$, and $\mu_w = 0$ separately.[4]

It is worth noting that all three contraints on ρ_w and μ_w produce well-behaved unimodal marginal posteriors for the parameters of the model with the exception of μ_p and σ_p in the process of the price mark-up shock. Their bivariate posterior density plot for the unrestricted model in Figure 3 reveals two separate modes. The dominating mode ($\mu_p \approx 0.85$, $\sigma_p \approx 0.14$) lies close to the one in Smets and Wouters (2007), while the other mode (with a greater value of σ_p) suggests that the moving average part of the price mark-up shock is almost noninvertible (μ_p has about 23% of its probability mass very close to unity). In order to restrict attention only to the neighbourhood of the dominating mode, we may place different constraints on μ_p or σ_p. For instance, imposing $\sigma_p < 0.19$ results in a well-behaved joint posterior distribution for μ_p and σ_p (see Figure 4).

The posterior means and standard deviations of the parameters of the restricted model (with $\rho_w = \mu_w = 0$ and $\sigma_p < 0.19$) are presented in Table 4. Imposing the constraint clearly improves estimation accuracy; all the numerical standard errors are smaller than those of the unrestricted SW model in Table 2, in some cases even substantially so. None of the plots of marginal posterior distributions (not shown) suggests remaining bad behavior, and hence there seems to be no need for further data-driven identification constraints. In particular, no signs of the multimodality problem discussed above can be seen. Of the

[3]The results are not shown to save space, but they are available upon request.

[4]Detailed estimation results of these restricted models are not presented to save space, but they are available upon request.

two modes found by Herbst and Schorfheide (2014), the one implying the endogenous dynamics (related to ξ_w and ι_w) as the driver of the persistence of wages (their Mode 2) remains the only mode by construction since the parameter ρ_w and μ_w related to the exogenous wage dynamics are constrained to zero. However, in the model with $\mu_w = 0$, the posterior mean of ρ_w is only 0.13 (with standard error 0.06), which also lends support to the endogenous dynamics as the driver of the persistence of wages. In Section 4 below, we provide further evidence in favor of the restricted SW model specification.

4 Forecasting

There is by now a large literature on forecasting with DSGE models (for a recent survey, see Del Negro and Schorfheide (2013)). While forecasting is not the main focus of this paper, we present a number of density forecasting results in order to provide further information on model fit and the usefulness of the data-driven identification constraints imposed on the SW model in Section 3.

Following Adolfson, Lindé, and Villani (2007), and Geweke and Amisano (2011, 2012), among many others, we tackle the problem of assessing the quality of the probabilistic forecasts for a random vector $\mathbf{Y}_n$, given its realized value $\mathbf{y}_n$ using scoring rules. Scoring rules are carefully reviewed in Gneiting and Raftery (2007), and we refer to their paper for a more detailed discussion on the topic. In the following, we briefly describe the scoring rule used in this paper.

Let p_n denote a forecaster's predictive distribution. A scoring rule $S(\mathbf{y}_n, p_n)$ can be considered a reward that the forecaster seeks to maximize. It is said to be strictly proper if the expected score under the distribution of $\mathbf{Y}_n$, q_n, is maximized by the choice $p_n = q_n$. It is further termed local if it depends on the density of the forecaster's predictive distribution p_n only through its realized value $\mathbf{y}_n$. The logarithmic score $S(\mathbf{y}_n, p_n) = \log p_n(\mathbf{y}_n)$ is known to be the only scoring rule with these desirable properties, and, therefore, we use it to assess the quality of the probabilistic forecasts (see also Parry, Dawid, and Lauritzen (2012) for a discussion on the so-called order-m proper local scoring rules). In particular,

we rank the competing models by the sum of the h-step-ahead log predictive densities

$$LS_h = \sum_{n=S+1}^{T} \log p\left(\mathbf{y}_n \left| \mathbf{y}_{1:(n-h)} \right.\right), \tag{4}$$

where $h \geq 1$ is the forecasting horizon, $S+1$ is the starting date of the forecast evaluation period, $p\left(\mathbf{y}_{n+h} \left| \mathbf{y}_{1:n} \right.\right)$ is the h-step-ahead predictive likelihood evaluated at the observed $\mathbf{y}_{n+h}$, and $\mathbf{y}_{1:n} = (\mathbf{y}_1', \ldots, \mathbf{y}_n')$.

The close connection of LS_h to the marginal likelihood, when $h = 1$ facilitates the interpretation of the forecasting results (see, for example, Kass and Raftery (1995)). To see the connection, write (4) for $h = 1$ as

$$\begin{aligned} LS_1 &= \sum_{n=S+1}^{T} \log p\left(\mathbf{y}_n \left| \mathbf{y}_{1:(n-1)} \right.\right) \\ &= \log p\left(\mathbf{y}_{(S+1):T} \left| \mathbf{y}_{1:S} \right.\right), \end{aligned}$$

where

$$\begin{aligned} p\left(\mathbf{y}_{(S+1):T} \left| \mathbf{y}_{1:S} \right.\right) &= \frac{\int p\left(\mathbf{y}_{1:T} \left| \theta \right.\right) p\left(\theta\right) d\theta}{\int p\left(\mathbf{y}_{1:S} \left| \theta \right.\right) p\left(\theta\right) d\theta} \\ &= \int p\left(\mathbf{y}_{(S+1):T} \left| \mathbf{y}_{1:S}, \theta \right.\right) p\left(\theta \left| \mathbf{y}_{1:S} \right.\right) d\theta. \end{aligned} \tag{5}$$

It is easy to see that quantity (5) has the same interpretation as the marginal likelihood, if $\mathbf{y}_{1:S}$ is interpreted as a training sample, that is, if $p\left(\theta \left| \mathbf{y}_{1:S} \right.\right)$ is taken as the prior distribution of θ (see, for example, Adolfson et al. (2007), and Geweke and Amisano (2010) for a detailed discussion).

To rank the competing forecasting models by LS_h, we need to evaluate the h-step-ahead predictive likelihoods $p\left(\mathbf{y}_{n+h} \left| \mathbf{y}_{1:n} \right.\right)$ for each model. Following Warne, Coenen, and Christoffel (2013), we calculate these quantities using the importance sampling estimator

$$p\left(\mathbf{y}_{n+h} \left| \mathbf{y}_{1:n} \right.\right) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{p\left(\mathbf{y}_{n+h} \left| \mathbf{y}_{1:n}, \theta^i \right.\right) \pi_n\left(\theta^i\right)}{f\left(\theta \left| \psi \right.\right)}, \tag{6}$$

where, under Gaussianity, the conditional likelihood $p\left(\mathbf{y}_{n+h} \left| \mathbf{y}_{1:n}, \theta \right.\right)$ is given by

$$p\left(\mathbf{y}_{n+h} \left| \mathbf{y}_{1:n}, \theta \right.\right) = \frac{\left|\Sigma_{n+h|n}\right|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{ -\frac{1}{2} \epsilon_{n+h|n}' \Sigma_{n+h|n}^{-1} \epsilon_{n+h|n} \right\},$$

$\epsilon_{n+h|n} = \mathbf{y}_{n+h} - \mathbf{y}_{n+h|n}$ is the h-step-ahead forecast error, and $\Sigma_{n+h|n}$ is the h-step-ahead mean squared error of the forecast. The h-period-ahead forecasts $\mathbf{y}_{n+h|n}$ with $\Sigma_{n+h|n}$ are calculated in a standard fashion from the filter estimates of the state variable and the associated state variable covariance matrix, both based on the data $\mathbf{y}_{1:n}$ (see, for example, Hamilton (1994, 385)).

The importance densities $f(\theta|\psi)$ can be obtained by the procedure described in Section 2.1. However, if we are not interested in the numerical standard errors of the estimates of $p(\mathbf{y}_{n+h}|\mathbf{y}_{1:n})$ (not typically reported), we may also evaluate (6) using the SMC approximation of the posterior distribution of the parameters by setting $f(\theta^i) = \pi_n(\theta^i)$. At least for the SW model, the SMC and IS approximations of the posterior density of the parameters lie very close to each other. Thus, it may be reasonable to save computing time by estimating (6) using $\pi_n(\theta)$ as an importance distribution.

4.1 Forecast Comparison

In order to gauge density forecasting performance, in particular the benefits of allowing for diffuse prior distributions with and without data-driven indentification constraints, we compute pseudo out-of-sample forecasts from a number of models for the period 1970:1 to 2004:4. The forecasts are computed recursively, using an expanding data window starting at 1966:1. We consider the forecast horizons of one, four, eight, and twelve quarters. As discussed in Subsection 4, we rank the models using the LS_h criterion (4). Recall that for $h = 1$, LS_h is very closely connected to the log of the marginal likelihoods (see the discussion preceding (5)), and, hence, they can be interpreted using Bayes factors (see, for example, Kass and Raftery (1995) for a detailed discussion on the Bayes factor).

The results are presented in Table 5. At all forecast horizons, the model estimated under the informative priors of Smets and Wouters (2007) performs the worst, while the diffuse priors with the three identification constraints considered in Section 3 is the clear winner. It is worth pointing out that among the identification constraints, the restriction $\rho_w = 0$ gets the least support, while the restriction $\mu_w = 0$ leads to the second-best outcome at three out of the four forecast horizons considered. These findings thus lend further support to the identification constraints imposed in Section 3.

Twice the logarithmic Bayes factor of the model with the diffuse priors against the model estimated under the informative priors of Smets and Wouters (2007) is 62, providing very strong evidence in favor of the model with diffuse priors. According to this measure of model fit, there is also strong evidence in favor of the data-driven indentification constraints, with twice the logarithmic Bayes factor of the model with $\rho_w = \mu_w = 0$ and $\sigma_p < 0.19$ against the unrestricted model with the diffuse priors around 6.6.

All in all, the forecast results provide supporting evidence in favor of the SW model estimated using the diffuse priors. Moreover, the data-driven identification constraints are seen to considerably improve forecast accuracy over and above the model with the diffuse priors alone.

5 Conclusion

In this paper, we have revisited the problem of estimating Bayesian DSGE models under diffuse prior distributions. In most of the previous literature, tight prior distributions have been assumed, and it is only the introduction of sequential Monte Carlo methods that has made it possible to entertain diffuse priors. The main problem with very informative priors is their tendency to drive the empirical results, which inhibits learning about the parameters of the model from the data. Diffuse priors, on the other hand, tend to yield multimodal posterior distributions, which can be handled by means of sequential Monte Carlo methods.

We propose augmenting any sequential Monte Carlo algorithm with an importance sampling step, which facilitates efficient assessment of estimation accuracy, and whose asymptotic properties are well known. We also recommend examining the posterior distributions of the parameters of a DSGE model for data-driven identification constraints that can be imposed to avoid redundancies and improve fit. The validity of such constraints can readily be checked by Bayes factors and out-of-sample forecast comparisons.

In empirical analysis of the Smets and Wouters (2007) model on the same data set that they considered, we find that their results are strongly driven by their informative prior distributions. Assuming diffuse priors, we obtain results similar to those of Herbst and Schorfheide (2014) and are able to show that our estimates are very accurate. Inspection

of the posterior distributions gives rise to identification constraints that are strongly supported by the data. In particular, we are able to restrict the wage markup shock to a white noise process. Under this constraint, we can conclude that the persistence of wages is driven by the endogenous dynamics.

The out-of-sample forecast results are encouraging. While we have in this paper concentrated on estimation and used forecast comparisons mostly as a device in model selection, in future work it might be interesting to examine more systematically the usefulness of data-driven identification constraints in forecasting. In particular, comparisons to wider set of competing models, including the so-called DSGE-VAR models found superior to DSGE models (see Del Negro and Schorfheide (2013)) would be of interest.

Appendix

In this appendix, we describe the SMC procedure employed in the text. We start by introducing the likelihood function and then explain the particle filter.

As discussed in Section 2, we focus on dynamic stochastic general equilibrium (DSGE) models, whose dynamics are adequately captured by the first-order approximation around the non-stochastic steady state. The solution of such a DSGE model can be written as a system of state and observation equations, whose likelihood function can be evaluated using the Kalman filter. Defining a $T \times p$ vector y, obtained by combining the $p \times 1$ observation vectors $\{\mathbf{y}_n\}_{n\in\mathcal{T}}$ for $\mathcal{T}=\{1,\ldots,T\}$, the likelihood function can be expressed as

$$p\left(\mathbf{y}\,|\theta\right) = \prod_{n=1}^{T} \frac{\left|\Sigma_{n|n-1}\right|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\epsilon'_{n|n-1}\,\Sigma^{-1}_{n|n-1}\,\epsilon_{n|n-1}\right\}, \tag{A.1}$$

where $\epsilon_{n|n-1} = \mathbf{y}_n - \mathbf{C}\mathbf{x}_{n|n-1}$ is the Gaussian forecast error based on the best forecast $x_{n|n-1}$ of the state variable x_n, C is a selection matrix of ones and zeros, and $\Sigma_{n|n-1}$ is the associated mean squared error. The quantities $x_{n|n-1}$ and $\Sigma_{n|n-1}$ are complicated nonlinear functions of the model parameters θ, and are calculated using the Kalman filter. For Bayesian analysis of DSGE models, (A.1) is combined with a prior distribution denoted by $p\left(\theta\right)$.

Following Chopin (2004), we describe the particle filter as an iterative method that produces an evolving particle system. At time $t \in \mathcal{L} = \{1, \ldots, L\}$ $(L \leq T)$, a particle system is defined as a collection of random variables $\left\{\theta_t^j, w_t^j\right\}_{j \in \mathcal{N}}$ in a space of interest $\Theta_t \times R^+$, and $\mathcal{N} = \{1, \ldots, N\}$. The variables θ_t^j and w_t^j are referred to as particles and their importance weights, respectively. The system $\left\{\theta_t^j, w_t^j\right\}_{j \in \mathcal{N}}$ targets a given distribution π_t, in the sense that

$$\sum_{j=1}^{N} w_t^j \varphi\left(\theta_t^j\right) \to \mathbb{E}_{\pi_t}\left(\varphi\right), \tag{A.2}$$

almost surely as $N \to \infty$, for any π_t-integrable function φ. The target distributions π_t considered in this paper are defined on a common space $\Theta_t = \Theta$. For example, for $L = T$ (i.e., $t \in \mathcal{T}$), π_t is the posterior density of θ given data up to time t: $\pi_t\left(\theta\right) = \pi\left(\theta \left| \mathbf{y}_{1:t} \right.\right)$, where $\mathbf{y}_{1:t} = \left(\mathbf{y}_1', \ldots, \mathbf{y}_t'\right)$. However, following Durham and Geweke (2014), we introduce data in batches. In this case, $L \leq T$ and $\pi_t\left(\theta\right)$ is the posterior density of θ given data up to the most recent observation in the tth batch. In particular, let the integers τ_t $(t \in \mathcal{L})$ refer to the dates of these observations, such that $\tau_0 = 0 < \tau_1 < \cdots < \tau_L = T$. Then, from (A.1), the kernel of the posterior $\pi_t\left(\theta\right)$ can be expressed as

$$\begin{aligned} \pi\left(\theta \left| \mathbf{y}_{1:\tau_t} \right.\right) \quad \propto \quad & p\left(\theta\right) \prod_{n=1}^{\tau_t} \frac{\left|\Sigma_{n|n-1}\right|^{-1/2}}{\left(2\pi\right)^{p/2}} \\ & \times \exp\left\{-\frac{1}{2} \epsilon_{n|n-1}' \Sigma_{n|n-1}^{-1} \epsilon_{n|n-1}\right\} \quad \left(t \in \{1, \ldots, L\}\right). \end{aligned} \tag{A.3}$$

It is worth noting that Herbst and Schorfheide (2014) use a sequence of artificial intermediate distributions $\left[\pi_T\left(\theta\right)\right]^{\phi_k}$ for $\phi_0 = 0 < \phi_1 < \cdots < \phi_K = 1$, to obtain the posterior $\pi_T\left(\theta\right)$ directly. Their approach may be inefficient in our case, since the whole sequence of distributions $\pi_t\left(\theta\right)$ is required for the calculation of LS_h in (4).

Next, we briefly describe the algorithm used to move the system $\left\{\theta_t^j, w_t^j\right\}_{j \in \mathcal{N}}$forward in time. For a more elaborate discussion on particle filters, we refer to Chopin (2004) and Del Moral, Doucet, and Jasra (2006). The evolution of the particle system comprises three iterative operations, namely, correction, selection, and mutation, in Chopin's (2004) terminology, and the particle filter consists in repeating these three steps until $t = L$ (i.e., $\tau_L = T$). The algorithm is first initialized by drawing N particles $\left\{\theta_0^j\right\}_{j \in \mathcal{N}}$ from the prior distribution $p\left(\theta\right)$. Then, for all $t \in \{1, \ldots, L\}$, the following steps are repeated until $t = L$ (i.e., $\tau_L = T$).

1. Correction: The correction step involves calculating weights that reflect the density of the particles in the current iteration. These weights w_t^j can be calculated using the importance weight function given by $w_t(\theta) = \pi_t(\theta)/\pi_{t-1}(\theta)$. From (A.3), the kernel of the weight function can be expressed as

$$\widetilde{w}_t(\theta) = \prod_{n=\tau_{t-1}+1}^{\tau_t} \frac{\left|\Sigma_{n|n-1}\right|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\epsilon'_{n|n-1}\Sigma_{n|n-1}^{-1}\epsilon_{n|n-1}\right\}, \tag{A.4}$$

   where the quantities $\mathbf{x}_{n|n-1}$ and $\Sigma_{n|n-1}$ $(n \in \{1,\ldots,\tau_t\})$ are obtained by the Kalman filter for each θ_{t-1}^j $(j \in \mathcal{N})$. The weights are then normalized according to $w_t^j = \widetilde{w}_t^j / \sum_{i=1}^N \widetilde{w}_t^i$.

2. Selection: The normalized weights from the correction step, $\left\{w_t^j\right\}_{j\in\mathcal{N}}$, are then used to produce a particle approximation $\left\{\widehat{\theta}_{t-1}^j, 1\right\}_{j\in\mathcal{N}}$ of $\pi_t(\theta)$. In particular, $\left\{\widehat{\theta}_{t-1}^j\right\}_{j\in\mathcal{N}}$ are simulated from the collection $\left\{\theta_{t-1}^j, w_t^j\right\}_{j\in\mathcal{N}}$ using residual resampling.[5]

3. Mutation: The simulated particles $\left\{\widehat{\theta}_{t-1}^j\right\}_{j\in\mathcal{N}}$ are mutated according to $\theta_t^j \sim p(\widehat{\theta}_{t-1}^j|\, \mathbf{y}_{1:\tau_t}, \vartheta_t)$ $(j \in \mathcal{N})$, where the p.d.f. admits $\pi_t(\theta)$ as an invariant density. The vector ϑ_t contains the parameters of the p.d.f. $p(\cdot)$, involving the distinct elements of the covariance matrix of the Metropolis-Hastings (MH) proposal distribution. The elements of ϑ_t are obtained online from the current population of particles.

This algorithm is discussed in Remark 1 of Del Moral, Doucet and Jasra (2006), and in Durham and Geweke (2014). The convergence results for the particle system produced by the algorithm are established in Del Moral, Doucet and Jasra (2006), where they assume that the sequences $\{\tau_t\}_{t\in\mathcal{L}}$ and $\{\vartheta_t\}_{t\in\mathcal{L}}$ are known. It is important to notice that the particles are mutated after the selection step. One advantage of mutating after resampling is that the accuracy of the particle approximation $\left\{\theta_t^j, 1\right\}_{j\in\mathcal{N}}$ of $\pi_t(\theta)$, used in forecasting, can be readily improved by increasing the number of Markov Chain Monte Carlo (MCMC) iterations, M, in the mutation phase, where M is a positive integer.

In applications, the integers τ_t in (A.3) must be set in such a way that the importance distributions $\pi_{t-1}(\theta)$ approximate the target posteriors $\pi_t(\theta)$ reasonable well for all $t \in \{1,\ldots,L\}$ (see the discussion preceding (A.4)). Otherwise, the particle approximation

[5]We use residual resampling, since it outperforms multinomial resampling in terms of precision of the particle approximation (see, for example, Chopin (2004)).

may be degenerate (see Del Moral, Doucet and Jasra (2006)). Obviously, when $\tau_t = t$ (i.e., $\tau_1 = 1, \ldots, \tau_L = T$), i.e., resampling is performed T times, the successive posterior distributions are as close to each other as possible, resulting in potentially good importance distributions. However, this approach is not usually ideal in terms of precision of the particle approximation, because resampling increases the variance of the estimates and reduces the number of distinct particles (see Chopin (2004) and Del Moral, Doucet and Jasra (2012)). Therefore, resampling should be used only when necessary for preventing degeneracy of the particles. Hence, we use the following (adaptive) recursive procedure of Durham and Geweke (2014) to produce $\tau_1, \ldots, \tau_L$. At each cycle $t \in \{1, \ldots, L\}$, conditional on the previous cycles, the posterior density kernel $\pi\left(\theta \left| \mathbf{y}_{1:\tau_t} \right.\right) \propto \gamma\left(\theta \left| \mathbf{y}_{1:\tau_t} \right.\right)$ is obtained by introducing new data one observation at time into $\gamma\left(\theta \left| \mathbf{y}_{1:\tau_{t-1}} \right.\right)$, until a stopping criterion, based on, say, the effective sample size (ESS) is met. Degeneracy of the particles is usually monitored by the ESS: $1 \leq \left\{\sum_{i=1}^{N} \left(w_t\left(\theta_{t-1}^{i}\right)\right)^2\right\}^{-1} \leq N$ (cf. (A.4)). As the ESS takes small values for a degenerate particle approximation, we introduce new data one observation at a time until the ESS drops below a particular threshold, such as $N/2$. The convergence results presented in Del Moral, Doucet and Jasra (2012), suggest that (A.2) holds almost surely also for the particle system generated by this adaptive algorithm.

As for the mutation step, the sampling is performed by the randomized block Metropolis-Hastings (MH) method of Chib and Ramamurthy (2010), where at each MCMC iteration $m \in \{1, \ldots, M\}$, the parameters θ_t^j ($j \in \mathcal{N}$, $t \in 1, \ldots, L$) are first randomly clustered into an arbitrary number of blocks, and then simulated one block at a time using an MH step, carried out using a Gaussian random walk proposal distribution for each block of θ_t^j.[6] The covariance matrices of the proposal distributions are constructed from the associated

[6]We follow the procedure of Chib and Ramamurthy (2010) to obtain the random blocks $\theta_1^{j,N}, \theta_2^{j,N}, \ldots, \theta_{b_j}^{j,N}$, in the ith iteration. The algorithm is started by randomly permuting $\theta_1^{j,N}, \ldots, \theta_m^{j,N}$, where m is the number of the estimated parameters. The shuffled parameters are denoted by $\theta_{\rho(1)}^{j,N}, \ldots, \theta_{\rho(m)}^{j,N}$, where $\rho(1), \ldots, \rho(m)$ is a permutation of the integers $1, 2, \ldots, m$. Then, the blocks $\theta_1^{j,N}, \theta_2^{j,N}, \ldots, \theta_{b_j}^{j,N}$ are obtained recursively as follows. The first block $\theta_1^{j,N}$ is initialized at $\theta_{\rho(1)}^{j,N}$. Each parameter $\theta_{\rho(l)}^{j,N}$ in turn for $l = 2, 3, \ldots, m$, is included in the first block with probability (tuning parameter) p_θ, and used to start a new block with probability $(1 - p_\theta)$. The procedure is repeated until each reshuffled error term is included in one of the blocks.

elements of the sample covariance matrix of the current population of the particles, $\mathbf{V}_{t,m}$. The matrix $\mathbf{V}_{t,m}$ is further multiplied by an adaptive tuning parameter $0.1 \leq c_{t,m} \leq 1$, whose role is to keep the MH acceptance rate at 0.25 (see Durham and Geweke (2014), and Herbst and Schorfheide (2014)). In particular, $c_{t,m}$ is set at $c_{t,m-1} + 0.01$ if the acceptance rate is greater than 0.25, and at $c_{t,m-1} - 0.01$ otherwise. This procedure is repeated independently for each particle θ_t^j until the particles are clearly distinct. Following Durham and Geweke (2014), we use the relative numerical efficiency (RNE) as a measure of particle divergence (see Geweke (2005,276)). We calculate RNEs from predictive likelihoods, and stop mutating particles when the RNE exeeds a given threshold. The maximum number of MCMC iterations $M_{\max}$ is set at 50.

In order to assess the quality of the probabilistic forecasts h periods ahead of the structural models of interest, we need $p(\mathbf{y}_{n+h} | \mathbf{y}_{1:n})$ in (6) for all $n \in S+1, \ldots, T$, while the SMC algorithm described above produces a sequence of distributions $\pi_t(\theta) = \pi(\theta | \mathbf{y}_{1:\tau_t})$ for $t \in \mathcal{L}$, where the integers $\tau_1, \ldots, \tau_{L-1}$ ($\tau_L = T$) are computed online to optimize the performance of the sampler. To that end, given the sequence of the estimated posteriors $\pi_t(\theta)$ ($t \in 1, \ldots, L-1$), we propose to simulate the posteriors $\pi_n(\theta)$ ($n \in \tau_t+1, \ldots, \tau_{t+1}-1$) by the sequential importance resampling algorithm of Rubin (1988). In particular, using $\pi_t(\theta)$ as the importance density, we obtain the following importance weight functions:

$$\begin{aligned} w_n(\theta) &\propto \widetilde{w}_n(\theta) \\ &= \prod_{l=\tau_t+1}^{n} \frac{\left|\Sigma_{l|l-1}\right|^{-1/2}}{(2\pi)^{p/2}} \exp\left\{-\frac{1}{2}\epsilon'_{l|l-1}\Sigma^{-1}_{l|l-1}\epsilon_{l|l-1}\right\}, \quad (n \in \tau_t+1, \ldots, \tau_{t+1}-1) \end{aligned}$$

for each $t \in 1, \ldots, L-1$ (see the discussion preceding (A.4)). We then simulate the posterior distributions $\pi_n(\theta)$ from the particle approximation $\left\{\theta_t^j\right\}_{j \in \mathcal{N}}$ of $\pi_t(\theta)$, using these weights $w_n(\theta^j)$. The resulting posterior estimates of $\pi_n(\theta)$ for $n \in S+1, \ldots, T$, can also be used to sample the joint predictive distributions $p(\mathbf{Y}_{n+1}, \ldots, \mathbf{Y}_{n+h})$, as described in Adolfson, Lindé, and Villani (2007). In this so-called sampling the future algorithm, for each of the N draws from the posterior distribution of the parameters $\pi_n(\theta)$, M future paths of $\mathbf{Y}_{n+1}, \ldots, \mathbf{Y}_{n+h}$ are simulated. This sample of $N \times M$ draws from the posterior predictive distribution, can then be used to calculate the posterior quantities of interest, such as point forecasts.

References

Adolfson, M., Linde, J., and Villani, M. (2007), "Forecasting Performance of an Open Economy DSGE Model," *Econometric Reviews*, 26, 289–328.

Amisano, G., and Geweke, J. (2013), "Prediction Using Several Macroeconomic Models," Working Paper Series 1537, European Central Bank.

Cappé, O., Douc, R., Guillin, A., Marin, J.M., and Robert, C.P. (2008), "Adaptive Importance Sampling in General Mixture Classes," *Statistics and Computing*, 18, 447–459.

Chib, S., and Ramamurthy, S. (2010), "Tailored Randomized Block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 155, 19–38.

Chopin, N. (2004), "Central Limit Theorem for Sequential Monte Carlo Methods and Its Application to Bayesian Inference," *Annals of Statistics*, 32, 2385–2411.

Creal, D. (2007), "Sequential Monte Carlo Samplers for Bayesian DSGE Models," Working paper, Vrije Universitiet, Amsterdam.

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society, Series B*, 68, 411–436.

Del Moral, P, Doucet, A. and Jasra, A. (2012), "On Adaptive Resampling Strategies for Sequential Monte Carlo Methods," *Bernoulli*, 18, 252–278.

Del Negro, M., and Schorfheide, F. (2013), "DSGE Model-Based Forecasting," in *Handbook of Economic Forecasting* (Vol. 2A), ed. G. Elliott, G., and A. Timmermann, Amsterdam: North-Holland, pp. 57–140.

Durham, G., and Geweke, J. (2014), "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," *Advances in Econometrics*, 34, 1–44.

Everitt, B., Landau, S., Leese, M., and Stahl, D. (2011), *Cluster Analysis*, Chichester: Wiley.

Fruhwirth-Schnatter, S. (2001), "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models," *Journal of the American Statistical Association*, 96, 194–209.

Geweke, J. (1989), "Bayesian Inference in Econometric-Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, New Jersey: Wiley.

Geweke, J. and Amisano, G. (2010), "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, 26, 216–230.

Geweke, J. and Amisano, G. (2011), "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

Geweke, J. and Amisano, G. (2012), "Prediction with Misspecified Models," *American Economic Review*, 102, 482–486.

Gneiting, T. and Raftery, A.E. (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association*, 102, 359–378.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton, New Jersey: Princeton University Press.

Herbst, E. and Schorfheide, F. (2014), "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 29, 1073–1098.

Hoogerheide, L., Opschoor, A., and van Dijk, H.K. (2012), "A Class of Adaptive Importance Sampling Weighted EM Algorithms for Efficient and Robust Posterior and Predictive Simulation," *Journal of Econometrics*, 171, 101–120.

Jasra, A., Stephens, D.A., and Holmes, C.C. (2007), "On Population-Based Simulation for Static Inference," *Statistics and Computing*, 17, 263–279.

Kass, R.E., and Raftery, A.E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Kloek, T. and van Dijk, H. (1978), "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46, 1–19

Koop, G, Pesaran, M.H., and Smith, R.P. (2013), "On Identification of Bayesian DSGE Models," *Journal of Business & Economic Statistics*, 31, 300–314.

Parry, M., Dawid, A.P., and Lauritzen, S. (2012), "Proper Local Scoring Rules," *Annals of Statistics*, 40, 561–592.

Rubin, D. (1988), "Using the SIR Algorithm to Simulate Posterior Distributions," in *Bayesian Statistics*, ed. J.M. Bernardo, M. H. DeGroot, D.V. Lindley, and A.F.M. Smith, Oxford: Oxford University Press, pp. 395–402.

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

Warne, A., Coenen, G., and Christoffel, K. (2013), "Predictive Likelihood Comparisons with DSGE and DSGE-VAR Models," Working Paper Series 1536, European Central Bank.

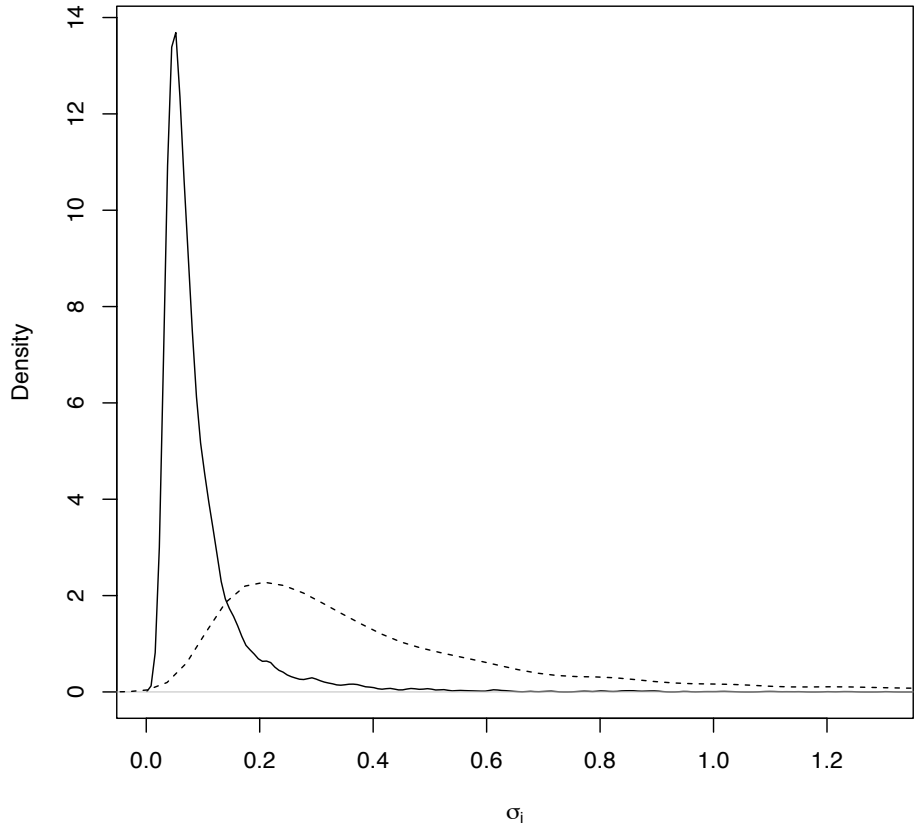

Figure 1: Priors for σ_i ($i \in \{a, b, g, I, r, p, w\}$); solid line: the SW prior, dashed line: our diffuse prior.

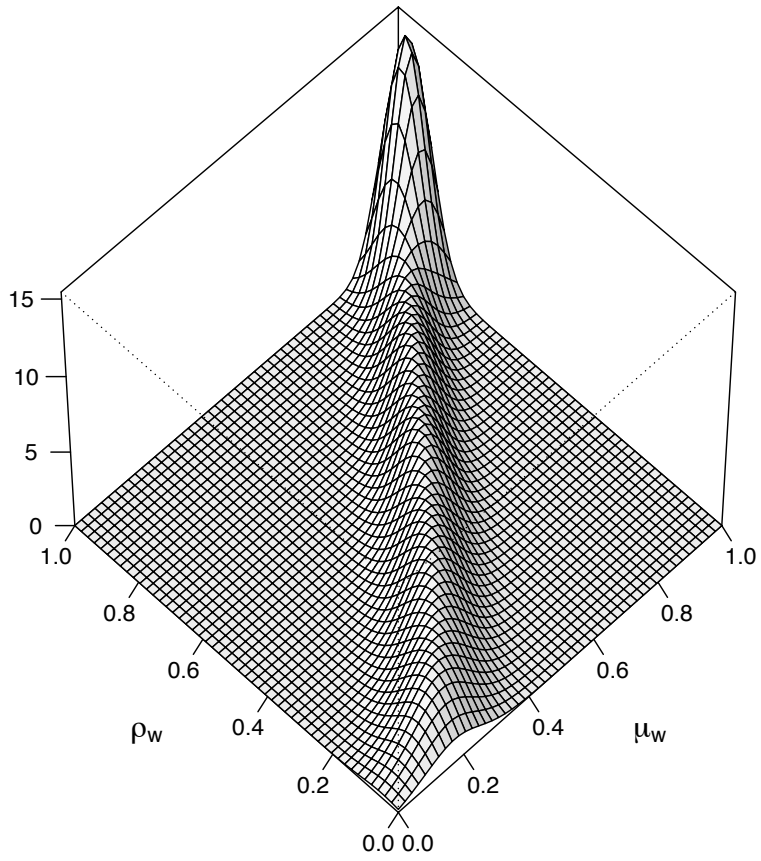

Figure 2: Joint posterior density of ρ_w and ζ_w under the diffuse priors in the unconstrained model.

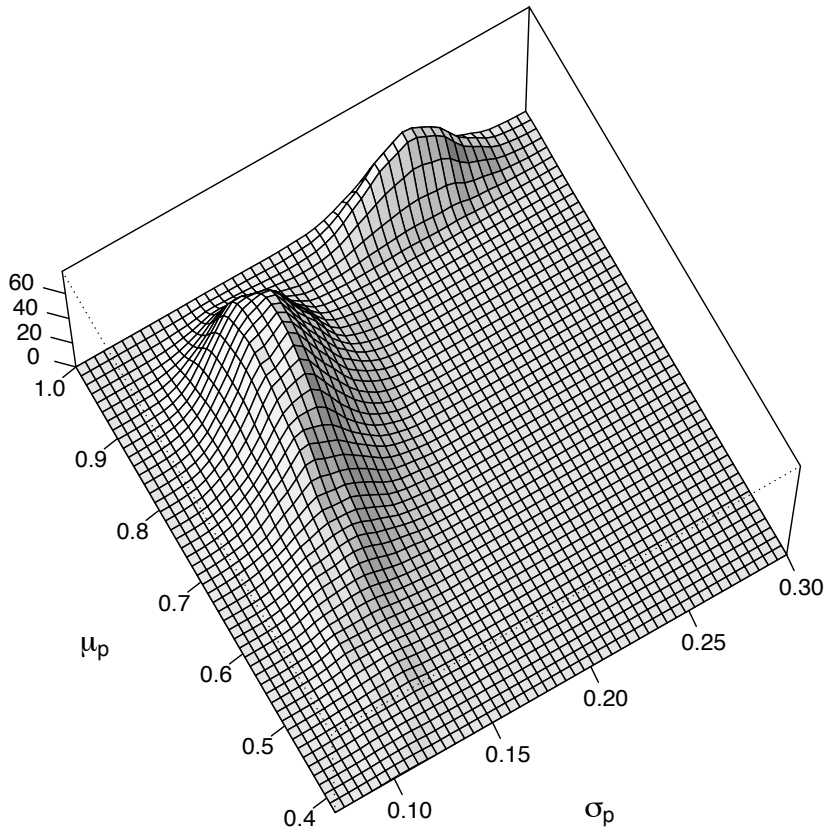

Figure 3: Joint posterior density of μ_p and σ_p under the diffuse priors in the unconstrained model.

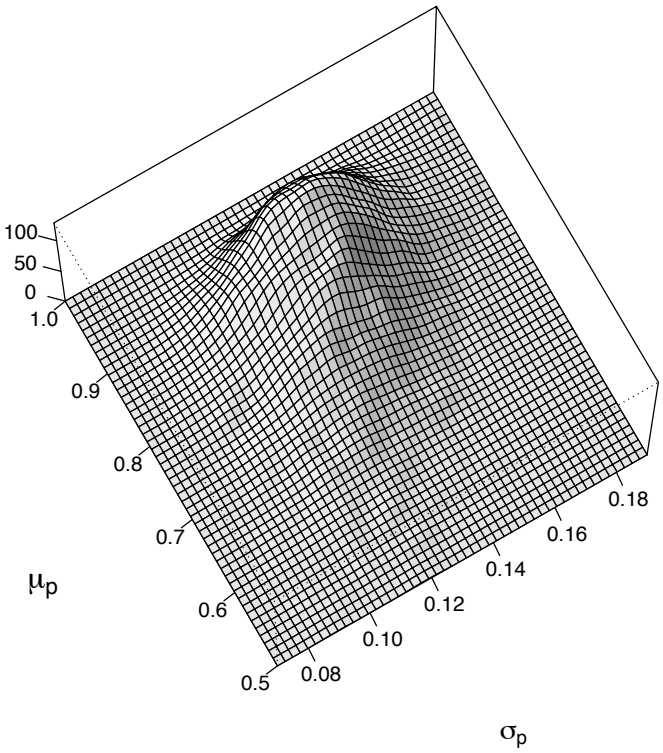

Figure 4: Joint posterior density of μ_p and σ_p under the diffuse priors in the model with ρ_w and μ_w constrained to zero.

Table 1: Diffuse prior distributions.

Parameter	Transformation	Distribution	Mean	SD
σ_a	-	Invgamma	0.50	1.00
σ_b	-	Invgamma	0.5	1.00
σ_g	-	Invgamma	0.50	1.00
σ_l	-	Invgamma	0.50	1.00
σ_r	-	Invgamma	0.50	1.00
σ_ρ	-	Invgamma	0.50	1.00
σ_w	-	Invgamma	0.50	1.00
ρ_a	Logit	Normal	0.00	1.70
ρ_b	Logit	Normal	0.00	1.70
ρ_g	Logit	Normal	0.00	1.70
ρ_l	Logit	Normal	0.00	1.70
ρ_r	Logit	Normal	0.00	1.70
ρ_ρ	Logit	Normal	0.00	1.70
ρ_w	Logit	Normal	0.00	1.70
μ_ρ	Logit	Normal	0.00	1.70
μ_w	Logit	Normal	0.00	1.70
φ	-	Normal	4.00	4.50
σ_c	-	Normal	1.50	1.11
$h(\lambda)$	Logit	Normal	0.00	1.70
ξ_w	Logit	Normal	0.00	1.70
σ_l	-	Normal	2.00	2.25
ξ_p	Logit	Normal	0.00	1.70
ι_w	Logit	Normal	0.00	1.70
ι_p	Logit	Normal	0.00	1.70
ψ	Logit	Normal	0.00	1.70
ϕ	-	Normal	1.25	0.36
r_π	-	Normal	1.50	0.75
ρ	Logit	Normal	0.00	1.70
r_y	-	Normal	0.12	0.15
$r_{\Delta y}$	-	Normal	0.12	0.15
π	-	Gamma	0.62	0.30
$100(\beta^{-1}-1)$	-	Gamma	0.25	0.30
l	-	Normal	0.00	6.00
γ	-	Normal	0.40	0.30
ρ_{ga}	Logit	Normal	0.00	1.70
α	-	Normal	0.30	0.15

Table 2: Augmented SMC estimation results of the unrestricted SW model with the diffuse prior distributions.

Parameter	Mean	[0.05, 0.95]	Post.SD	SD(mean)
σ_a	0.45	[0.40 , 0.49]	0.03	0.0006
σ_b	0.25	[0.21 , 0.29]	0.02	0.0004
σ_g	0.54	[0.49 , 0.60]	0.03	0.0006
σ_l	0.47	[0.39 , 0.55]	0.05	0.0008
σ_r	0.24	[0.21 , 0.26]	0.01	0.0002
σ_ρ	0.15	[0.11 , 0.25]	0.04	0.0009
σ_w	0.26	[0.22 , 0.30]	0.02	0.0004
ρ_a	0.97	[0.95 , 0.98]	0.01	0.0002
ρ_b	0.16	[0.04 , 0.33]	0.09	0.0015
ρ_g	0.98	[0.97 , 0.99]	0.01	0.0001
ρ_l	0.70	[0.59 , 0.80]	0.06	0.0011
ρ_r	0.06	[0.02 , 0.15]	0.04	0.0007
ρ_ρ	0.91	[0.84 , 0.98]	0.04	0.0007
ρ_w	0.64	[0.18 , 0.98]	0.28	0.0062
μ_ρ	0.82	[0.61 , 0.99]	0.12	0.0020
μ_w	0.58	[0.08 , 0.96]	0.31	0.0071
φ	8.48	[4.81 , 13.08]	2.47	0.0536
σ_c	1.70	[1.35 , 2.09]	0.22	0.0044
$h(\lambda)$	0.70	[0.62 , 0.78]	0.05	0.0011
ξ_w	0.94	[0.83 , 0.99]	0.05	0.0009
σ_l	3.22	[1.49 , 5.52]	1.24	0.0297
ξ_p	0.72	[0.62 , 0.81]	0.06	0.0010
ι_w	0.75	[0.43 , 0.96]	0.16	0.0024
ι_p	0.13	[0.02 , 0.29]	0.09	0.0015
ψ	0.75	[0.50 , 0.95]	0.14	0.0021
ϕ	1.79	[1.58 , 2.02]	0.13	0.0024
r_π	2.75	[2.12 , 3.46]	0.41	0.0065
ρ	0.88	[0.84 , 0.92]	0.02	0.0004
r_y	0.15	[0.08 , 0.22]	0.04	0.0007
$r_{\Delta y}$	0.27	[0.21 , 0.33]	0.04	0.0007
π	1.11	[0.85 , 1.39]	0.17	0.0034
$100(\beta^{-1}-1)$	0.08	[0.01 , 0.22]	0.07	0.0010
l	-0.94	[-3.26 , 1.32]	1.39	0.0249
γ	0.42	[0.39 , 0.44]	0.02	0.0003
ρ_{ga}	0.46	[0.28 , 0.64]	0.11	0.0022
α	0.18	[0.15 , 0.22]	0.02	0.0004

The estimates are obtained by the SMC procedure augmented with the importance sampling step. The columns labeled Mean, [0.05, 0-95], Post.SD, and SD(mean) contain the mean, the 5th and 95th percentiles and standard deviation of the posterior distribution, and the numerical standard error of the mean of the respective parameter, respectively.

Table 3: Augmented SMC estimation results of the unrestricted SW model with informative prior.

Parameter	Mean	[0.05, 0.95]	Post.SD	SD(mean)
σ_a	0.46	[0.42 , 0.51]	0.03	0.0002
σ_b	0.24	[0.20 , 0.28]	0.02	0.0001
σ_g	0.53	[0.48 , 0.58]	0.03	0.0002
σ_l	0.45	[0.38 , 0.54]	0.05	0.0003
σ_r	0.25	[0.22 , 0.27]	0.02	0.0001
σ_ρ	0.14	[0.11 , 0.17]	0.02	0.0001
σ_w	0.24	[0.21 , 0.28]	0.02	0.0002
ρ_a	0.96	[0.94 , 0.98]	0.01	0.0001
ρ_b	0.22	[0.09 , 0.37]	0.09	0.0005
ρ_g	0.98	[0.96 , 0.99]	0.01	0.0001
ρ_l	0.71	[0.61 , 0.81]	0.06	0.0006
ρ_r	0.15	[0.05 , 0.26]	0.06	0.0004
ρ_ρ	0.89	[0.80 , 0.96]	0.05	0.0004
ρ_w	0.97	[0.94 , 0.99]	0.02	0.0001
μ_ρ	0.71	[0.52 , 0.84]	0.10	0.0008
μ_w	0.84	[0.72 , 0.92]	0.06	0.0004
φ	5.71	[4.07 , 7.51]	1.04	0.0082
σ_c	1.40	[1.19 , 1.64]	0.14	0.0009
$h(\lambda)$	0.71	[0.64 , 0.78]	0.04	0.0002
ξ_w	0.70	[0.59 , 0.81]	0.07	0.0005
σ_l	1.85	[0.98 , 2.84]	0.57	0.0028
ξ_p	0.65	[0.56 , 0.74]	0.05	0.0003
ι_w	0.58	[0.36 , 0.78]	0.13	0.0006
ι_p	0.25	[0.11 , 0.40]	0.09	0.0005
ψ	0.53	[0.34 , 0.71]	0.11	0.0006
ϕ	1.61	[1.48 , 1.74]	0.08	0.0007
r_π	2.05	[1.77 , 2.34]	0.18	0.0009
ρ	0.81	[0.77 , 0.85]	0.02	0.0002
r_y	0.09	[0.05 , 0.13]	0.02	0.0001
$r_{\Delta y}$	0.22	[0.18 , 0.27]	0.03	0.0001
π	0.81	[0.64 , 0.99]	0.11	0.0005
$100(\beta^{-1}-1)$	0.17	[0.08 , 0.27]	0.06	0.0004
l	0.46	[-1.37 , 2.29]	1.11	0.0058
γ	0.43	[0.41 , 0.46]	0.01	0.0001
ρ_{ga}	0.52	[0.37 , 0.67]	0.09	0.0006
α	0.19	[0.16 , 0.22]	0.02	0.0001

See notes to Table 2.

Table 4: Augmented SMC estimation results of the restricted SW model with the diffuse prior distributions.

Parameter	Mean	[0.05, 0.95]	Post.SD	SD(mean)
σ_a	0.44	[0.40 , 0.49]	0.03	0.0002
σ_b	0.25	[0.20 , 0.29]	0.03	0.0002
σ_g	0.55	[0.49 , 0.60]	0.03	0.0002
σ_l	0.47	[0.39 , 0.55]	0.05	0.0004
σ_r	0.24	[0.21 , 0.26]	0.01	0.0001
σ_ρ	0.13	[0.10 , 0.16]	0.02	0.0002
σ_w	0.30	[0.27 , 0.33]	0.02	0.0001
ρ_a	0.97	[0.95 , 0.98]	0.01	0.0001
ρ_b	0.16	[0.04 , 0.33]	0.09	0.0011
ρ_g	0.98	[0.96 , 0.99]	0.01	0.0001
ρ_l	0.71	[0.60 , 0.80]	0.06	0.0006
ρ_r	0.06	[0.02 , 0.14]	0.04	0.0004
ρ_ρ	0.94	[0.87 , 0.99]	0.04	0.0003
μ_ρ	0.84	[0.69 , 0.94]	0.08	0.0006
φ	7.78	[4.37 , 12.22]	2.37	0.0209
σ_c	0.94	[0.55 , 1.35]	0.24	0.0021
$h(\lambda)$	0.84	[0.79 , 0.88]	0.03	0.0003
ξ_w	0.97	[0.96 , 0.99]	0.01	0.0001
σ_l	3.41	[1.66 , 5.74]	1.24	0.0114
ξ_p	0.72	[0.64 , 0.79]	0.05	0.0004
ι_w	0.86	[0.72 , 0.97]	0.08	0.0010
ι_p	0.10	[0.02 , 0.24]	0.07	0.0006
ψ	0.79	[0.56 , 0.96]	0.13	0.0014
ϕ	1.82	[1.61 , 2.05]	0.13	0.0020
r_π	2.72	[2.09 , 3.44]	0.41	0.0031
ρ	0.88	[0.85 , 0.91]	0.02	0.0002
r_y	0.14	[0.08 , 0.22]	0.04	0.0004
$r_{\Delta y}$	0.27	[0.21 , 0.33]	0.04	0.0003
π	1.15	[0.88 , 1.42]	0.16	0.0016
$100(\beta^{-1}-1)$	0.09	[0.01 , 0.24]	0.07	0.0010
l	-1.27	[-3.62 , 0.98]	1.39	0.0120
γ	0.41	[0.38 , 0.44]	0.02	0.0002
ρ_{ga}	0.45	[0.27 , 0.63]	0.11	0.0016
α	0.18	[0.15 , 0.22]	0.02	0.0002

See notes to Table 2. Estimation is based on the following constraints: $\rho_w = \mu_w = 0$ and $\sigma_p < 0.19$.

Table 5: Density forecasting results.

Prior	Constraints	$h = 1$	$h = 4$	$h = 8$	$h = 12$
Informative	-	-783.9	-1089.1	-1179.8	-1193.2
Diffuse	-	-752.9	-1050.4	-1157.9	-1175.2
Diffuse	$\rho_w = \mu_w = 0$	-752.8	-1045.5	-1154.7	-1177.5
Diffuse	$\rho_w = 0$	-756.3	-1047.7	-1158.2	-1179.8
Diffuse	$\mu_w = 0$	-750.6	-1046.2	-1152.4	-1174.1
Diffuse	$\rho_w = \mu_w = 0$, $\sigma_p < 0.19$	-749.6	-1042.3	-1150.8	-1173.2

The figures are sums of h-step-ahead log predictive densities.

Research Papers 2013

2015-20: Henri Nyberg and Harri Pönkä: International Sign Predictability of Stock Returns: The Role of the United States

2015-21: Ulrich Hounyo and Bezirgen Veliyev: Validity of Edgeworth expansions for realized volatility estimators

2015-22: Girum D. Abate and Niels Haldrup: Space-time modeling of electricity spot prices

2015-23: Eric Hillebrand, Søren Johansen and Torben Schmith: Data revisions and the statistical relation of global mean sea-level and temperature

2015-24: Tommaso Proietti and Alessandra Luati: Generalised partial autocorrelations and the mutual information between past and future

2015-25: Bent Jesper Christensen and Rasmus T. Varneskov: Medium Band Least Squares Estimation of Fractional Cointegration in the Presence of Low-Frequency Contamination

2015-26: Ulrich Hounyo and Rasmus T. Varneskov: A Local Stable Bootstrap for Power Variations of Pure-Jump Semimartingales and Activity Index Estimation

2015-27: Rasmus Søndergaard Pedersen and Anders Rahbek: Nonstationary ARCH and GARCH with t-distributed Innovations

2015-28: Tommaso Proietti and Eric Hillebrand: Seasonal Changes in Central England Temperatures

2015-29: Laurent Callot and Johannes Tang Kristensen: Regularized Estimation of Structural Instability in Factor Models: The US Macroeconomy and the Great Moderation

2015-30: Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Testing for Level Shifts in Fractionally Integrated Processes: a State Space Approach

2015-31: Matias D. Cattaneo, Michael Jansson and Whitney K. Newey: Treatment Effects with Many Covariates and Heteroskedasticity

2015-32: Jean-Guy Simonato and Lars Stentoft: Which pricing approach for options under GARCH with non-normal innovations?

2015-33: Nina Munkholt Jakobsen and Michael Sørensen: Efficient Estimation for Diffusions Sampled at High Frequency Over a Fixed Time Interval

2015-34: Wei Wei and Denis Pelletier: A Jump-Diffusion Model with Stochastic Volatility and Durations

2015-35: Yunus Emre Ergemen and Carlos Velasco: Estimation of Fractionally Integrated Panels with Fixed Effects and Cross-Section Dependence

2015-36: Markku Lanne and Henri Nyberg: Nonlinear dynamic interrelationships between real activity and stock returns

2015-37: Markku Lanne and Jani Luoto: Estimation of DSGE Models under Diffuse Priors and Data-Driven Identification Constraints