Data-Driven Identification Constraints for DSGE Models*

MARKKU LANNE† and JANI LUOTO†

†Department of Political and Economic Studies, and HECER, University of Helsinki, Helsinki, Finland (e-mail: markku.lanne@helsinki.fi; jani.luoto@helsinki.fi)

Abstract

We propose imposing data-driven identification constraints to alleviate the multimodality problem arising in the estimation of poorly identified dynamic stochastic general equilibrium models under non-informative prior distributions. We also devise an iterative procedure based on the posterior density of the parameters for finding these constraints. An empirical application to the Smets and 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.

I. Introduction

Advances in Bayesian simulation methods have recently facilitated the estimation of relatively large-scale dynamic stochastic general equilibrium (DSGE) models. However, when using the commonly employed random walk Metropolis–Hastings (RWMH) algorithm, typically relatively tight prior distributions have to be assumed to tackle flat and multimodal posterior distributions arising from weak identification in these models (see e.g. Koop, Pesaran and Smith, 2013 and the references therein). This has the unfortunate consequence that the resulting posterior distributions may not have much to say about how well the structural model fits the data, but the priors are likely to be driving the results, which precludes us from learning about the parameters of the model from the data.

Under less informative priors, one potential solution to the problem of weak identification is offered by the so-called data-driven identifiability constraints put forth in the statistics literature (see Frühwirth-Schnatter, 2001) but, to the best of our knowledge, not applied to DSGE models. Such constraints can be found by inspection of the output of

JEL Classification numbers: C11, C32, C52, D58.

*We would like to thank Francesco Zanetti (the Editor), and an anonymous referee for useful comments. Financial support from the Academy of Finland (grants 268454 and 308628) is gratefully acknowledged. The first author also acknowledges financial support from CREATEES (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.
the posterior distribution. Subsequently, the restricted and unrestricted models may be compared to check the validity of the constraints. For instance, if two parameters seem to be weakly identified and always take an equal value with high probability, a model, where their equality is imposed, might be preferable. In practice, the constraints are set in an iterative procedure, where at each stage the posterior distribution of the parameters is inspected to find additional constraints, whose validity is then assessed by means of, say, Bayes factors and improvement in estimation accuracy. The iteration continues until no further acceptable constraints or signs of weak identification can be found.

It is important to distinguish our approach from specification searches (see e.g. Leamer, 1978), where the goal is to find the ‘true’ model, or to simplify or improve the current model. While our constraints may also lead to a simpler and more easily interpretable model, the ultimate objective is to alleviate problems arising from lack of identifiability. In other words, we take a DSGE model as given, but acknowledge the fact that these models tend to be poorly identified and, therefore, try to find constraints respecting the geometry of the posteriors to facilitate identification. In addition to improvements in estimation accuracy and probabilistic forecasts due to improved identification, data-driven identifiability constraints are indeed also likely to facilitate interpretation.

Our approach calls for an efficient estimation method that is capable of handling multimodality likely to be encountered at least in the unrestricted DSGE model. Such a procedure has recently been suggested by Herbst and Schorfheide (2014), who employed an adaptive sequential Monte Carlo (SMC) algorithm to estimate the Smets and Wouters (2007) model (SW model hereafter) based on relatively loose priors (see also Creal, 2007; Chib and Ramamurthy, 2010). Once the final model has been obtained, it is important to ensure that it has been accurately estimated (and in case of obvious inaccuracy, some of the constraints may be relaxed and alternative constraints entertained). To that end, Herbst and Schorfheide suggested running the SMC algorithm multiple times to obtain an approximation of the asymptotic variances of the parameter estimates. This is, unfortunately, computationally very costly in the case of a complex high-dimensional DSGE model, and, therefore, in practice only few runs (20 in Herbst and Schorfheide, 2014) are feasible, yielding a very imprecise measure of estimation accuracy.

To facilitate assessment of estimation accuracy, we propose to augment the SMC algorithm with a non-sequential importance sampling (IS) step, which has the advantage that numerical standard errors can be readily calculated without burdensome simulations. Moreover, convergence results for non-sequential IS are available (see Geweke (2005, Theorem 4.2.2), while the asymptotic properties of the adaptive SMC algorithm are not necessarily known. Hence, in addition to being computationally feasible in assessing the accuracy of the estimates, our procedure is theoretically well motivated. Of course, these two approaches are not substitutes. In particular, it may be a good idea to run the SMC algorithm a few times before the IS step to ensure that the SMC algorithm has visited the entire posterior. This is important because the IS step relies on the SMC approximation.

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 similar 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 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, implying different wage dynamics, also detected by Herbst and Schorfheide (2014), whose relative importance they found difficult to assess. Moreover, imposing the data-driven identification constraints improves estimation accuracy, and also forecast comparisons lend support to the restricted model.

In the next section, we discuss the SMC algorithm and explain, how any such algorithm can be augmented with an IS step. The estimation results of the SW model are reported in section ‘Estimation results’, 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 topics of section III. Forecasting with DSGE models and computing log predictive densities that can be used to rank competing models, are discussed in section IV, while in section ‘Forecast comparison’, we report the results of forecast comparison. Finally, section V concludes.

II. Sequential Monte Carlo estimation

The likelihood function of any DSGE model is a complicated nonlinear function of the vector of its structural parameters \( \theta \), 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 SMC algorithm for estimating DSGE models under non-informative priors (see also Creal, 2007 for estimating DSGE models by SMC methods). By virtue of their construction, these methods have been found to perform well for badly behaved posterior distributions (for further evidence on this, see e.g. Del Moral, Doucet and Jasra, 2006; Jasra, Stephens and Holmes, 2007; Durham and Geweke, 2014).

While SMC methods are well suited for estimating DSGE models, they involve the problem that the assessment of the accuracy of estimation is difficult 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, 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 densities on which we base our search for identification constraints, 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).

**Importance sampling step**

As discussed in the Introduction, in order to facilitate convenient computation of the numerical standard errors of the estimates, we propose to conclude the SMC run by non-sequential IS. In other words, given the estimates produced by the SMC algorithm, a final IS step is performed to obtain an approximation of the asymptotic variance of these estimates. The idea is to approximate the posterior distribution by a mixture of Student’s $t$-distributions. This approximation is then efficiently used as the IS density. The proposed procedure used in the final IS step 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 non-sequential 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 IS 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’s $t$-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 $\{\psi_n^i\}_{i=1}^N$. This mixture density is then used to calculate the 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 $\theta$ is approximated by a mixture of $J$ multivariate $t$ distributions:

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

where $t_k(\theta|\mu_j, V_j; v_j)$ is the density function of the $k$-variate $t$ distribution with mode $\mu_j$, (positive definite) scale matrix $V_j$, and degrees of freedom $v_j$, $\psi = (\mu_1', \ldots, \mu_J', \text{vech}(V_1)', \ldots, \text{vech}(V_J)', v_1, \ldots, v_J, \alpha_1, \ldots, \alpha_J)'$, and the mixing probabilities $\alpha_j$ sum to unity.

To obtain an efficient IS 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 equation (1) with respect to $\psi$. Because the elements of vector $\psi$ do not enter the posterior density $\pi_n(\theta)$, this is equivalent to maximizing
\[
\int [\log f(\theta|\psi)] \pi_{n}(\theta) d\theta = E[\log f(\theta|\psi)],
\] (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

\[
\frac{1}{N} \sum_{i=1}^{N} \log f(\theta^i|\psi),
\] (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 \( \psi \) in equation (1). Once the IS density has been obtained, it can be used to estimate any posterior quantity of interest, and the associated numerical standard errors (these are estimated by the standard formula). Of course, standard asymptotic results apply to these estimators (see e.g. Geweke, 2005, p. 114).

Hoogerheide et al. (2012) estimate the parameters \( \psi \) in equation (1) by maximizing a weighted variant of equation (3) (where the weights are standard importance weights) 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 equation (1). This also means that we do not have to calculate importance weights, when estimating the parameters \( \psi \) in equation (1). In our experience, a good approximation even for a pathological posterior can be obtained by setting the number of components \( J \) in the mixture distribution (1) sufficiently large. However, the estimates of the scale matrices \( V_j \) may become imprecise for some \( j \) for unreasonably large \( J \). In the empirical application in section ‘Estimation results’, we set \( J = 20 \).

It is worth noting that the EM algorithm may be sensitive to the starting values of \( \psi \). To solve the problem, we partition the particle approximation \( \{\theta^i\}_{i=1}^{N} \) into \( J \) clusters by an agglomerative hierarchical clustering algorithm (see e.g. Everitt et al., 2011), and then use the sample mean and covariance matrix of the particles in the \( j \)th cluster as initial values for \( \mu_j \) and \( V_j \) (\( j \in 1, \ldots, J \)). Prior to clustering, the particle approximation \( \{\theta^i\}_{i=1}^{N} \) is normalized and orthogonalized such that the elements of \( \theta \) have zero means and unit variances, and are uncorrelated. The distance between two normalized and orthogonalized particles is measured by the Euclidean distance between them. Moreover, the distance between two clusters is measured by Ward’s measure. A good initial value for the mixing probability \( \pi_j \) can be obtained by dividing the number of points in the \( j \)th cluster by \( N \). This procedure tends to be quick and very reliable.

Estimation results

We estimate the SW model on the same quarterly data set (from 1970:1–2004:4) as Smets and Wouters (2007) using an SMC algorithm augmented with a non-sequential 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, where information concerning its tuning parameters is also provided. To check whether the SMC algorithm has visited the entire posterior, we ran it for each DSGE model several times before the IS step. As already discussed in section ‘Importance sampling step’, the SMC method indeed turned out to deliver reliable approximations.

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 univariate 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 Schorfheide (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 deviations 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 deviations for the innovations, and hence a relatively large prior 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 Schorfheide (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. Here, as well as in all subsequent estimations, we have run the SMC algorithm several times before the IS step to make sure that it has visited the entire posterior. 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, which is reassuring from
### TABLE 2
Augmented sequential Monte Carlo (SMC) estimation results of the unrestricted Smets–Wouters model with the diffuse prior distributions

| Parameter | Mean  | [0.05, 0.95] | Post.SD | SD(mean) |
|-----------|-------|--------------|---------|---------|
| \( \sigma_a \) | 0.45  | [0.40, 0.49] | 0.03    | 0.0006  |
| \( \sigma_b \) | 0.25  | [0.21, 0.29] | 0.02    | 0.0004  |
| \( \sigma_g \) | 0.54  | [0.49, 0.60] | 0.03    | 0.0006  |
| \( \sigma_l \) | 0.47  | [0.39, 0.55] | 0.05    | 0.0008  |
| \( \sigma_r \) | 0.24  | [0.21, 0.26] | 0.01    | 0.0002  |
| \( \sigma_w \) | 0.26  | [0.22, 0.30] | 0.02    | 0.0004  |
| \( \rho_a \) | 0.97  | [0.95, 0.98] | 0.01    | 0.0002  |
| \( \rho_b \) | 0.16  | [0.04, 0.33] | 0.09    | 0.0015  |
| \( \rho_g \) | 0.98  | [0.97, 0.99] | 0.01    | 0.0001  |
| \( \rho_t \) | 0.70  | [0.59, 0.80] | 0.06    | 0.0011  |
| \( \rho_r \) | 0.06  | [0.02, 0.15] | 0.04    | 0.0007  |
| \( \rho_p \) | 0.91  | [0.84, 0.98] | 0.04    | 0.0007  |
| \( \rho_w \) | 0.64  | [0.18, 0.98] | 0.28    | 0.0062  |
| \( \mu_p \) | 0.82  | [0.61, 0.99] | 0.12    | 0.0020  |
| \( \mu_w \) | 0.58  | [0.08, 0.96] | 0.31    | 0.0071  |
| \( \varphi \) | 8.48  | [4.81, 13.08]| 2.47    | 0.0536  |
| \( \sigma_c \) | 1.70  | [1.35, 2.09] | 0.22    | 0.0044  |
| \( h(\lambda) \) | 0.70  | [0.62, 0.78] | 0.05    | 0.0011  |
| \( z_w \) | 0.94  | [0.83, 0.99] | 0.05    | 0.0009  |
| \( \sigma_l \) | 3.22  | [1.49, 5.52] | 1.24    | 0.0297  |
| \( \varepsilon_p \) | 0.72  | [0.62, 0.81] | 0.06    | 0.0010  |
| \( v_w \) | 0.75  | [0.43, 0.96] | 0.16    | 0.0024  |
| \( v_p \) | 0.13  | [0.02, 0.29] | 0.09    | 0.0015  |
| \( \psi \) | 0.75  | [0.50, 0.95] | 0.14    | 0.0021  |
| \( \phi \) | 1.79  | [1.58, 2.02] | 0.13    | 0.0024  |
| \( r_{\pi} \) | 2.75  | [2.12, 3.46] | 0.41    | 0.0065  |
| \( \rho \) | 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  |
| \( \pi \) | 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  |
| \( \gamma \) | 0.42  | [0.39, 0.44] | 0.02    | 0.0003  |
| \( \rho_{\Delta y} \) | 0.46  | [0.28, 0.64] | 0.11    | 0.0022  |
| \( \alpha \) | 0.18  | [0.15, 0.22] | 0.02    | 0.0004  |

**Notes:** The estimates are obtained by the SMC procedure augmented with the importance sampling step. The columns labelled 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. Following Herbst and Schorfheide (2014), we use 12,000 particles (i.e. \( N = 12,000 \)). The number of components \( J \) in equation (1) is set to 20, which should ensure that the resulting importance distribution provides an accurate approximation to the target posterior. The parameters of equation (1) are obtained using 500 iteration rounds in the EM algorithm. The IS results are based on 400,000 draws from equation (1).
the viewpoint of searching for the identification constraints. They are also much smaller than those reported by Herbst and Schorfheide (2014), which (given the similarity of their and our posterior distributions) most likely reflects the poor quality of their asymptotic variance estimates based on only 20 simulations. Hence, our results suggest that concluding the SMC run by non-sequential IS is worthwhile as far as assessing the accuracy of posterior estimates is concerned.

For comparison, we also estimated the SW model using our procedure augmented with the IS step assuming the informative priors of Smets and Wouters (2007). The posterior estimates are very close to those in Smets and Wouters (2007), with very small standard errors, indicating great estimation accuracy (see Table A1 in the online appendix). The marginal posteriors of Smets and Wouters (2007) also seemed to be strongly driven by their informative prior distributions such that the posteriors tended to very closely resemble the priors.

III. Identification

Dynamic stochastic general equilibrium models (including the SW model that we consider in this paper) tend to be poorly identified (see Koop et al., 2013 and the references therein). When taking them to data, the Monte Carlo (MC) output of the posterior distribution should, therefore, be systematically analysed 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 densities of the parameters for bad behaviour indicating lack of identification (such as multimodality, or flatness over a wide range of relevant parameter values even if there is only one mode). Once the badly behaved parameters have been singled out, the next step is to analyse 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 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 a potential multimodality problem,¹ 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. Finally, the numerical standard errors are computed by the non-sequential IS step. 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.

Inspection of the marginal posterior densities of the parameters reveals six badly behaved parameters that govern wage ($\xi_w, \iota_w, \rho_w, \mu_w$) and price ($\mu_p, \sigma_p$) stickiness (see Figure 2; the marginal posterior densities of the remaining parameters are depicted in

¹ Lack of identification may give rise to multimodal posterior densities of the parameters. This may, for instance, be the case if the data are unable to distinguish between the role of a DSGE model’s external and internal propagation mechanisms, as pointed out by Herbst and Schorfheide (2015, Ch. 4).
Identification constraints for DSGE models

Figures A1 and A2 in the online appendix). Also, Herbst and Schorfheide (2014) found multimodal features in the posterior distributions of these parameters.

Next, the bivariate density plots of all combinations of the marginal posteriors of these parameters are inspected to find out about the identification of these parameters. These plots are depicted in Figures A3–A5 in the online appendix, and in Figures 3 and 4. All of these plots indicate weak identification. As outlined in the Introduction, we continue by iteratively imposing identifiability constraints, one at a time, until no further acceptable constraints or signs of weak identification can be found. The order in which the constraints are imposed, may matter for the choice of the final model, and in some cases, it may be a good idea to experiment with a number of different orderings. However, it is advisable to
Figure 3. Joint posterior density of $\rho_w$ and $\mu_w$ under the diffuse priors in the unconstrained model.

Figure 4. Joint posterior density of $\mu_p$ and $\sigma_p$ under the diffuse priors in the unconstrained model.

© 2017 The Authors. *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd.
Figure 5. Joint posterior density of $\mu_p$ and $\sigma_p$ under the diffuse priors in the model with $\rho_w$ and $\mu_w$ constrained to zero, and $\sigma_p < 0.19$

confine oneself to constraints that have a clear interpretation, simplify the model, and can be feasibly imposed in practice.

Among the constraints suggested by the bivariate plots, the one involving the ARMA parameters $\rho_w$ and $\mu_w$ of the wage mark-up shock $\varepsilon_i^w = \rho_w \varepsilon_{i-1}^w + \eta_i^w - \mu_w \eta_{i-1}^w$, where $\eta_i^w \sim N(0, \sigma_w^2)$, in Figure 3 satisfies the criteria listed above. In particular, these parameters 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 $\rho_w$ and $\mu_w$ restricted to zero).

In addition, we consider two alternative constraints, namely, $\rho_w = 0$, and $\mu_w = 0$ separately. The Bayes factors based on the one-step-ahead log predictive densities in Table 4 slightly favour the model with only $\mu_w$ restricted to zero over the model with $\mu_w = \rho_w = 0$. However, because the estimate of $\rho_w$ lies very close to zero in the former model, we proceed with the model involving both restrictions, which affords a clearer interpretation (see the discussion at the end of this section).

It is worth noting that all three constraints on $\rho_w$ and $\mu_w$ produce well-behaved unimodal marginal posteriors for the parameters of the model with the exception of $\mu_p$ and $\sigma_p$ in the process of the price mark-up shock. Their bivariate posterior density plot for the unrestricted model in Figure 4 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 $\sigma_p$) suggests that the moving average part of the price mark-up shock is almost

---

2 As will be discussed in section IV, Bayes factors for comparison of alternative model specifications can be based on one-step-ahead log predictive densities that are closely connected to marginal likelihoods.
### TABLE 3

Augmented sequential Monte Carlo estimation results of the restricted Smets–Wouters model with the diffuse prior distributions

| Parameter | Mean   | [0.05, 0.95] | Post.SD | SD(mean) |
|-----------|--------|--------------|---------|----------|
| $\sigma_a$ | 0.44  | [0.40, 0.49] | 0.03    | 0.0002   |
| $\sigma_b$ | 0.25  | [0.20, 0.29] | 0.03    | 0.0002   |
| $\sigma_g$ | 0.55  | [0.49, 0.60] | 0.03    | 0.0002   |
| $\sigma_l$ | 0.47  | [0.39, 0.55] | 0.05    | 0.0004   |
| $\sigma_r$ | 0.24  | [0.21, 0.26] | 0.01    | 0.0001   |
| $\sigma_p$ | 0.13  | [0.10, 0.16] | 0.02    | 0.0002   |
| $\sigma_w$ | 0.30  | [0.27, 0.33] | 0.02    | 0.0001   |
| $\rho_a$   | 0.97  | [0.95, 0.98] | 0.01    | 0.0001   |
| $\rho_b$   | 0.16  | [0.04, 0.33] | 0.09    | 0.0011   |
| $\rho_g$   | 0.98  | [0.96, 0.99] | 0.01    | 0.0001   |
| $\rho_l$   | 0.71  | [0.60, 0.80] | 0.06    | 0.0006   |
| $\rho_r$   | 0.06  | [0.02, 0.14] | 0.04    | 0.0004   |
| $\rho_p$   | 0.94  | [0.87, 0.99] | 0.04    | 0.0003   |
| $\mu_p$    | 0.84  | [0.69, 0.94] | 0.08    | 0.0006   |
| $\sigma_c$ | 0.94  | [0.55, 1.35] | 0.24    | 0.0021   |
| $\beta(\lambda)$ | 0.84 | [0.79, 0.88] | 0.03    | 0.0003   |
| $\xi_w$    | 0.97  | [0.96, 0.99] | 0.01    | 0.0001   |
| $\xi_l$    | 3.41  | [1.66, 5.74] | 1.24    | 0.0114   |
| $\xi_p$    | 0.72  | [0.64, 0.79] | 0.05    | 0.0004   |
| $\lambda_w$| 0.86  | [0.72, 0.97] | 0.08    | 0.0010   |
| $\lambda_p$| 0.10  | [0.02, 0.24] | 0.07    | 0.0006   |
| $\psi$     | 0.79  | [0.56, 0.96] | 0.13    | 0.0014   |
| $\phi$     | 1.82  | [1.61, 2.05] | 0.13    | 0.0020   |
| $\beta$    | 2.72  | [2.09, 3.44] | 0.41    | 0.0031   |
| $\rho$     | 0.88  | [0.85, 0.91] | 0.02    | 0.0002   |
| $\beta_p$  | 0.14  | [0.08, 0.22] | 0.04    | 0.0004   |
| $\beta_{\lambda}$ | 0.27 | [0.21, 0.33] | 0.04    | 0.0003   |
| $\pi$      | 1.15  | [0.88, 1.42] | 0.16    | 0.0016   |
| $100(\beta^{-1} - 1)$ | 0.09 | [0.01, 0.24] | 0.07    | 0.0010   |
| $\gamma$   | 1.27  | [−3.62, 0.98] | 1.39    | 0.0120   |
| $\lambda_p$| 0.41  | [0.38, 0.44] | 0.02    | 0.0002   |
| $\lambda_{\lambda}$ | 0.45 | [0.27, 0.63] | 0.11    | 0.0016   |
| $\lambda_{\beta}$ | 0.18 | [0.15, 0.22] | 0.02    | 0.0002   |

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

non-invertible ($\mu_p$ 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 $\mu_p$ or $\sigma_p$. For instance, imposing $\sigma_p < 0.19$ results in a well-behaved joint posterior distribution for $\mu_p$ and $\sigma_p$ (see Figure 5).

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 3. 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 (see Figure A6 in the online appendix) suggests remaining bad behaviour, 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 two modes found by Herbst and Schorfheide (2014), the one implying the endogenous dynamics (related to $\xi_w$ and $\mu_w$) as the driver of the persistence of wages (their Mode 2) remains the only mode by construction since the parameter $\rho_w$ and $\mu_w$ related to the exogenous wage dynamics are constrained to zero. Moreover, in the model where only $\mu_w$ is restricted to zero, the posterior mean of $\rho_w$ is small (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 IV below, we provide further evidence in favour of the restricted SW model specification.

IV. 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 III.

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 $Y_n$, given its realized value $y_n$ 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 $S(y_n, p_n)$ 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 $Y_n$, $q_n$, 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 $y_n$. The logarithmic score $S(y_n, p_n) = \log p_n(y_n)$ 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

$$LS_h = \sum_{n=S+1}^{T} \log p \left( y_n | y_{1:(n-h)} \right), \quad (4)$$

3 It can also be seen from Tables 2 and 3 that the estimates of some key parameters in the restricted model are different from those in the weakly identified model. For instance, in the weakly identified model, the posterior probability that consumption and hours worked are complements is one (i.e. $\Pr(\sigma_c > 1 | y) = 1$; see Smets and Wouters, 2007, p. 589). This result also remains intact irrespective of the priors used (see Tables 2 and A1). However, when the identification constraints are imposed, the data lend strong support to log utility in consumption (i.e. $\sigma_c = 1$). In particular, under the constraints, the posterior mean of $\sigma_c$ in Table 3 is relatively close to one (0.94) and 95% of the posterior mass of $\sigma_c$ lies between 0.55 and 1.35.
where \( h \geq 1 \) is the forecasting horizon, \( S + 1 \) is the starting date of the forecast evaluation period, \( p(y_{n+h}|y_{1:n}) \) is the \( h \)-step-ahead predictive likelihood evaluated at the observed \( y_{n+h} \), and \( y_{1:n} = (y'_1, \ldots, y'_n) \).

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

\[
LS_1 = \sum_{n=S+1}^{T} \log p(y_n|y_{1:(n-1)})
\]

\[
= \log p(y_{(S+1):T}|y_{1:S}),
\]

where

\[
p(y_{(S+1):T}|y_{1:S}) = \frac{\int p(y_{1:T}|\theta)p(\theta)d\theta}{\int p(y_{1:T}|\theta)p(\theta)d\theta}
\]

\[
= \int p(y_{(S+1):T}|y_{1:S}, \theta)p(\theta|y_{1:S})d\theta.
\]

It is easy to see that quantity (5) has the same interpretation as the marginal likelihood, if \( y_{1:S} \) is interpreted as a training sample, that is, if \( p(\theta|y_{1:S}) \) is taken as the prior distribution of \( \theta \) (see e.g. Adolfson, Lindé and Villani, 2007; 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 \( p(y_{n+h}|y_{1:n}) \) for each model. Following Warne, Coenen and Christoffel (2017), we calculate these quantities using the IS estimator

\[
p(y_{n+h}|y_{1:n}) \approx \frac{1}{N} \sum_{i=1}^{N} p(y_{n+h}|y_{1:n}, \theta^i) \pi_n(\theta^i)/f(\theta|\psi),
\]

where, under Gaussianity, the conditional likelihood \( p(y_{n+h}|y_{1:n}, \theta) \) is given by

\[
p(y_{n+h}|y_{1:n}, \theta) = \frac{|\Sigma_{n+h}|^{-1/2}}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \epsilon_{n+h|n}^{'} \Sigma_{n+h}^{-1} \epsilon_{n+h|n} \right\},
\]

\( \epsilon_{n+h|n} = y_{n+h} - y_{n+h|n} \) 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 \( y_{n+h|n} \) 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 \( y_{1:n} \) (see e.g. Hamilton, 1994, p. 385).

The importance densities \( f(\theta|\psi) \) can be obtained by the procedure described in section ‘Importance sampling step’. However, if we are not interested in the numerical standard errors of the estimates of \( p(y_{n+h}|y_{1:n}) \) (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.
TABLE 4

Density forecasting results for 1970:1–2004:4

| Prior         | Constraints                   | $h = 1$  | $h = 4$  | $h = 8$  | $h = 12$ |
|---------------|-------------------------------|----------|----------|----------|----------|
| Informative   | −                             | −783.9   | −1,089.1 | −1,179.8 | −1,193.2 |
| Diffuse       | −                             | −752.9   | −1,050.4 | −1,157.9 | −1,175.2 |
| Diffuse       | $\rho_w = \mu_w = 0$         | −752.8   | −1,045.5 | −1,154.7 | −1,177.5 |
| Diffuse       | $\rho_w = 0$                 | −756.3   | −1,047.7 | −1,158.2 | −1,179.8 |
| Diffuse       | $\mu_w = 0$                  | −750.6   | −1,046.2 | −1,152.4 | −1,174.1 |
| Diffuse       | $\rho_w = \mu_w = 0, \sigma_p < 0.19$ | −749.6   | −1,042.3 | −1,150.8 | −1,173.2 |

Note: The figures are sums of $h$-step-ahead log predictive densities. The $h$-step-ahead log predictive densities are calculated at each time point as explained in the Appendix.

Forecast comparison

In order to gauge density forecasting performance, in particular the benefits of allowing for diffuse prior distributions with and without data-driven identification 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 1, 4, 8, and 12 quarters. As discussed in section IV, we rank the models using the $L_S h$ criterion (4). Recall that for $h = 1$, $L_S h$ 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 e.g. Kass and Raftery, 1995 for a detailed discussion on the Bayes factor).

The results are presented in Table 4. At all forecast horizons, the model estimated under the informative priors of Smets and Wouters (2007) performs the worst, while the model with the diffuse priors and the three identification constraints considered in section III 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 III.4

At the one-quarter forecast horizon, 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 favour of the model with diffuse priors. According to this measure of model fit, there is also strong evidence in favour of the data-driven identification 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 6.6.5

Our main motivation for conducting the forecast analysis is to ensure the feasibility of the data-driven identifiability constraints, and therefore, we only reported the pseudo-

4 Estimation results for inflation and GDP growth are reported in Tables A3 and A4 in the online appendix, respectively. The model involving all three identification constraints clearly turns out to outperform all other models in forecasting the GDP growth. For inflation, the model with only one constraint ($\mu_w = 0$) is the winner at all but the 12-quarter horizon, where, somewhat surprisingly, the unconstrained model estimated based on the informative prior produces the most accurate forecasts.

5 We also computed forecasts for a somewhat longer period, extending up to 2016:3. The results, reported in Table A2 in the online appendix, are in line with those in Table 4. It is only at the one-quarter horizon that the model involving the constraint $\rho_w = \mu_w = 0$ outperforms the model with the three identification constraints.
TABLE 5

Density forecasting results for 2005:1–2016:3

| Prior          | Constraints         | $h = 1$  | $h = 4$  | $h = 8$  | $h = 12$ |
|----------------|---------------------|----------|----------|----------|----------|
| Informative    | –                   | −272.7   | −439.9   | −516.0   | −553.3   |
| Diffuse        | –                   | −260.1   | −415.5   | −491.8   | −534.2   |
| Diffuse        | $\rho_w = \mu_w = 0$| −258.4   | −421.1   | −494.4   | −528.5   |
| Diffuse        | $\rho_w = 0$        | −267.8   | −427.9   | −496.8   | −534.1   |
| Diffuse        | $\mu_w = 0$         | −262.8   | −418.6   | −491.9   | −530.2   |
| Diffuse        | $\rho_w = \mu_w = 0$, $\sigma_p < 0.19$| −263.5   | −420.2   | −488.5   | −523.0   |

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

out-of-sample forecasts covering the estimation period 1970:1–2004:4 above. However, it might be interesting to extend the forecast comparison outside the estimation period to truly examine the forecast performance of the different models estimated under various priors, and to that end, in Table 5, we report results for the period 2005:1–2016:3. When interpreting them, it must be borne in mind that the forecast period is very different from the estimation period, involving, among other things, the recent financial crisis and a period with very low interest rates. Nevertheless, the identifiability constraints turn out quite useful in forecasting. In contrast to the pseudo-out-of-sample results in Table 4, the model involving all three constraints is the winner only at the 8- and 12-quarter horizons, while the constraints $\rho_w = \mu_w = 0$ lead to the most accurate forecasts at the one-quarter horizon. It is only at the one-year horizon that the model estimated with a diffuse prior and no constraints performs the best, but, most importantly, the model estimated with the informative prior performs the worst by a clear margin at all horizons.

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

V. Conclusion

In this paper, we have proposed the imposition of data-driven identifiability constraints to alleviate the multimodality problem encountered in estimating poorly identified DSGE models. Under diffuse prior distributions, such constraints facilitate interpretation, and improve estimation and forecast accuracy. They can be found by examining the posterior distribution of the parameters of the DSGE model being estimated in an iterative procedure, where constraints accepted by the data are repeatedly imposed until no further reasonable constraints can be found. The constraints are at each step validated by a number of measures, including estimation accuracy and Bayes factors.

Because finding the identifiability constraints is based on inspection of the posterior distribution of the parameters, it should be accurately estimated, and because estimation accuracy is a central measure in validating identifiability constraints, the estimation algorithm should facilitate convenient assessment of estimation accuracy. While the SMC algorithm of Herbst and Schorfheide (2014) is useful in estimating DSGE models under non-informative priors, it is computationally very burdensome in computing estimation
accuracy. Therefore, to facilitate the use of data-driven identifiability constraints, we propose augmenting it (or any SMC algorithm) with an IS step, which allows for efficient assessment of estimation accuracy, and whose asymptotic properties are well known.

In empirical analysis of the Smets and Wouters (2007) model on the same data set that they considered, we found their results strongly driven by their informative prior distributions. Assuming diffuse priors, we obtained results similar to those of Herbst and Schorfheide (2014) and were able to show that our estimates are very accurate. Inspection of the posterior distributions gave rise to identification constraints that were strongly supported by the data. In particular, we were able to restrict the wage markup shock to a white noise process. Under this constraint, we could conclude that the persistence of wages is driven by the endogenous dynamics.

The out-of-sample forecast results were 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 a wider set of competing models, including the so-called DSGE-V AR models found superior to DSGE models (see Del Negro and Schorfheide, 2013) would be of interest.

Appendix A: Sequential Monte Carlo Procedure

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

As discussed in section II, 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 $\{y_n\}_{n \in T}$ for $T=\{1,\ldots,T\}$, the likelihood function can be expressed as

$$p(y|\theta) = \prod_{n=1}^{T} \frac{|\Sigma_{n|n-1}|^{-1/2}}{(2\pi)^{p/2}} \exp \left\{ -\frac{1}{2} \epsilon_{n|n-1}' \Sigma_{n|n-1}^{-1} \epsilon_{n|n-1} \right\}, \quad (A1)$$

where $\epsilon_{n|n-1} = y_n - C x_{n|n-1}$ 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 $\theta$, and are calculated using the Kalman filter. For Bayesian analysis of DSGE models, equation (A1) is combined with a prior distribution denoted by $p(\theta)$.

Following Chopin (2004), we describe the SMC algorithm as an iterative method that produces an evolving particle system. At time $t \in L = \{1,\ldots,L\}$ ($L \leq T$), a particle system is defined as a collection of random variables $\{\theta'_j, w'_j\}_{j \in N}$ in a space of interest $\Theta \times \mathbb{R}^+$, and $N = \{1,\ldots,N\}$. The variables $\theta'_j$ and $w'_j$ are referred to as particles and their importance weights, respectively. The system $\{\theta'_j, w'_j\}_{j \in N}$ targets a given distribution $\pi_r$, in the sense that

$$\sum_{j=1}^{N} w'_j \varphi(\theta'_j) \rightarrow E_{\pi_r}(\varphi), \quad (A2)$$

© 2017 The Authors. Oxford Bulletin of Economics and Statistics published by Oxford University and John Wiley & Sons Ltd.
almost surely as $N \to \infty$, for any $\pi_t$-integrable function $\varphi$. The target distributions $\pi_t$ considered in this paper are defined on a common space $\Theta_0 = \Theta$. For example, for $L = T$ (i.e. $t \in T$), $\pi_t$ is the posterior density of $\theta$ given data up to time $t$: $\pi_t(\theta) = \pi(\theta | y_{1:t})$, where $y_{1:t} = (y_1', \ldots, y_t')$. However, following Durham and Geweke (2014), we introduce data in batches. In this case, $L \leq T$ and $\pi_t(\theta)$ is the posterior density of $\theta$ given data up to the most recent observation in the $r$th batch. In particular, let the integers $\tau_r, (t \in L)$ refer to the dates of these observations, such that $\tau_0 = 0 < \tau_1 < \cdots < \tau_L = T$. Then, from equation (A1), the kernel of the posterior $\pi_t(\theta)$ can be expressed as

$$
\pi(\theta | y_{1:t}) \propto p(\theta) \prod_{n=1}^{5t} \frac{|\Sigma^{-1}_{n-n-1}|^{-1/2}}{(2\pi)^{y/2}} \times \exp \left\{ -\frac{1}{2} \epsilon_{n-n-1}^{-1} \sum_{n-n-1}^{-\epsilon_{n-n-1}} \right\} (t \in \{1, \ldots, L\}).
$$

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

Next, we briefly describe the algorithm used to move the system $\{\theta_t, w_t\}_{t \in N}$ forward in time. For a more elaborate discussion on particle filters, we refer to Chopin (2004) and Del Moral et al. (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 $\{\theta_0, w_0\}_{t \in N}$ from the prior distribution $p(\theta)$. 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 $w_t'$ can be calculated using the importance weight function given by $w_t(\theta) = \pi_t(\theta) / \pi_{t-1}(\theta)$. From equation (A3), the kernel of the weight function can be expressed as

$$
\tilde{w}_t(\theta) = \prod_{n=\tau_{t-1}+1}^{\tau_t} \frac{|\Sigma^{-1}_{n-n-1}|^{-1/2}}{(2\pi)^{y/2}} \exp \left\{ -\frac{1}{2} \epsilon_{n-n-1}^{-1} \sum_{n-n-1}^{-\epsilon_{n-n-1}} \right\},
$$

where the quantities $x_{n-n-1}$ and $\Sigma_{n-n-1}$ ($n \in \{1, \ldots, \tau_t\}$) are obtained by the Kalman filter for each $\theta_{t-1}$ ($j \in N$). The weights are then normalized according to $w_t' = \tilde{w}_t / \sum_{i=1}^{N} \tilde{w}_t'$.  

2. Selection: The normalized weights from the correction step, $\{w_t'\}_{t \in N}$, are then used to produce a particle approximation $\{\hat{\theta}_{t-1}, 1\}_{t \in N}$ of $\pi_t(\theta)$. In particular, $\{\hat{\theta}_{t-1}\}_{t \in N}$ are simulated from the collection $\{\hat{\theta}_{t-1}, w_t'\}_{t \in N}$ using residual resampling. 

3. Mutation: The simulated particles $\{\hat{\theta}_{t-1}\}_{t \in N}$ are mutated according to $\hat{\theta}_{t-1} \sim p(\hat{\theta}_{t-1} | y_{1:t}, \theta_t)$ ($j \in N$), where the p.d.f. admits $\pi_t(\theta)$ as an invariant density. The vector $\theta_t$ contains the

\footnote{We use residual resampling, since it outperforms multinomial resampling in terms of precision of the particle approximation (see e.g. Chopin, 2004).}
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 \( \dot{\theta} \), are obtained online from the current population of particles.

This algorithm is discussed in Remark 1 of Del Moral et al. (2006), and in Durham and Geweke (2014). The convergence results for the particle system produced by the algorithm are established in Del Moral et al. (2006), where they assume that the sequences \( \{\tau_i\}_{i \in \mathbb{L}} \) and \( \{\dot{\theta}_t\}_{t \in \mathbb{L}} \) 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 \( \{\dot{\theta}_t, 1\}_{t \in \mathcal{N}} \) 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 \( \tau_i \) in equation (A3) must be set in such a way that the importance distributions \( \pi_{t-1}(\theta) \) approximate the target posteriors \( \pi_t(\theta) \) reasonably well for all \( t \in \{1, \ldots, L\} \) (see the discussion preceding (A4)). Otherwise, the particle approximation may be degenerate (see Del Moral et al., 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; 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 \( \gamma(\theta|y_{1:t-1}) \propto \pi(\theta|y_{1:t}) \) is obtained by introducing new data one observation at time into \( \gamma(\theta|y_{1:1:t-1}) \), until a stopping criterion, based on, say, the effective sample size (ESS) is met.

Degeneracy of the particles is usually monitored by the ESS: \( 1 \leq \left\{ \sum_{i=1}^{N}(w_i(\dot{\theta}_{t-1}))^2 \right\}^{-1} N \) (cf. equation (A4)). 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 \( N/2 \) (which is also used in the empirical application). The convergence results presented in Del Moral et al. (2012), suggest that equation (A2) holds almost surely also for the particle system generated by this adaptive algorithm.

As to 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 \( \dot{\theta}_t \) (\( j \in \mathcal{N} \), \( t \in \{1, \ldots, L\} \)) 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 \( \dot{\theta}_t \).

We follow the procedure of Chib and Ramamurthy (2010) to obtain the random blocks \( \theta_{1}^{N}, \theta_{2}^{N}, \ldots, \theta_{m}^{N} \), in the \( i \)th iteration. The algorithm is started by randomly permuting the \( \theta_{1}^{N}, \theta_{2}^{N}, \ldots, \theta_{m}^{N} \), where \( m \) is the number of the estimated parameters. The shuffled parameters are denoted by \( \theta_{\rho(1)}^{N}, \ldots, \theta_{\rho(m)}^{N} \), where \( \rho(1), \ldots, \rho(m) \) is a permutation of the integers \( 1, 2, \ldots, m \). Then, the blocks \( \theta_{l1}^{N}, \theta_{l2}^{N}, \ldots, \theta_{lb}^{N} \) are obtained recursively as follows. The first block \( \theta_{l1}^{N} \) is initialized at \( \theta_{\rho(1)}^{N} \). Each parameter \( \theta_{l1}^{(N)} \) in turn for \( l = 2, 3, \ldots, m \) is included in the first block with probability (tuning parameter) \( p_0 \) (in our empirical application, we set \( p_0 \) to 0.85), and used to start a new block with probability \((1-p_0)\). The procedure is repeated until each reshuffled error term is included in one of the blocks.

© 2017 The Authors. *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd.
The covariance matrices of the proposal distributions are constructed from the associated
elements of the sample covariance matrix of the current population of the particles, $V_{i,m}$. The matrix $V_{i,m}$ is further multiplied by an adaptive tuning parameter $0.1 \leq c_{i,m} \leq 1$, whose role is to keep the MH acceptance rate at 0.25 (see Durham and Geweke, 2014; Herbst and Schorfheide, 2014). In particular, $c_{i,m}$ is set at $c_{i,m-1} + 0.01$ if the acceptance rate is greater than 0.25, and at $c_{i,m-1} - 0.01$ otherwise. This procedure is repeated independently for each particle $\theta_i^t$ 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, p. 276). We calculate RNEs from predictive likelihoods, and stop mutating particles when the RNE exceeds a given threshold. The maximum number of MCMC iterations $M_{\text{max}}$ 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 $p(\mathbf{y}_{n+h} | \mathbf{y}_{1:n})$ for all $n \in S + 1, \ldots, T$, while the SMC algorithm described above produces a sequence of distributions $\pi_i(\theta) = p(\theta | \mathbf{y}_{1:t_i})$ for $t \in \mathcal{L}$, where the integers $\tau_1, \ldots, \tau_{T-1}$ ($\tau_T = T$) are computed online to optimize the performance of the sampler. To that end, given the sequence of the estimated posteriors $\pi_i(\theta)$ ($i \in 1, \ldots, L - 1$), we propose to simulate the posteriors $\pi_n(\theta)$ ($n \in \tau_i + 1, \ldots, \tau_i + 1 - 1$) by the sequential importance resampling algorithm of Rubin (1988). In particular, using $\pi_i(\theta)$ as the importance density, we obtain the following importance weight functions:

$$w_n(\theta) \propto \tilde{w}_n(\theta) = \prod_{l=\tau_i+1}^{n} \left| \frac{\Sigma_{l|l-1}}{2\pi^{d/2}} \exp \left\{ -\frac{1}{2} \epsilon_{l|l-1} \epsilon_{l|l-1} \right\} \right|^ {1/2}, \quad (n \in \tau_i + 1, \ldots, \tau_{i+1} - 1)$$

for each $t \in 1, \ldots, L - 1$ (see the discussion preceding (A4)). We then simulate the posterior distributions $\pi_n(\theta)$ from the particle approximation $\{\theta_i^t\}_{i \in N}$ of $\pi_i(\theta)$, using these weights $w_n(\theta')$. 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 et al. (2007). In this so-called sampling the future algorithm, for each of the $N$ draws from the posterior distribution of the parameters $\pi_n(\theta)$, $M$ (in the empirical application, $M = 8$) 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.

Final Manuscript Received: August 2017

References

Adolfson, M., Lindé, J. and Villani, M. (2007). ‘Forecasting performance of an open economy DSGE model’, *Econometric Reviews*, Vol. 26, pp. 289–328.

Cappé, O., Douc, R., Guillin, A., Marin, J. M. and Robert, C. P. (2008). ‘Adaptive importance sampling in general mixture classes’, *Statistics and Computing*, Vol. 18, pp. 447–459.

Chib, S. and Ramamurthy, S. (2010). ‘Tailored randomized block MCMC methods with application to DSGE models’, *Journal of Econometrics*, Vol. 155, pp. 19–38.
Supporting Information

Additional supporting information may be found in the online version of this article:

Table A1. Augmented SMC estimation results of the unrestricted SW model with the informative prior distributions of Smets and Wouters (2007).
Table A2. Density forecasting results for 1970:1–2016:3.
Table A3. Density forecasting results of inflation for 1970:1–2004:4.
Table A4. Density forecasting results of GDP growth for 1970:1–2004:4.
Figure A1. Marginal posterior densities of $\sigma_a$, $\sigma_b$, $\sigma_g$, $\sigma_i$, $\sigma_w$, $\rho_a$, $\rho_b$, $\rho_g$, $\rho_i$, $\rho_r$, $\varphi$, $\sigma_c$ and $h$ under the diffuse priors in the unconstrained model.
Figure A2. Marginal posterior densities of $\sigma_i$, $\xi_p$, $\iota_p$, $\psi$, $\Phi$, $r_\pi$, $\rho$, $r_y$, $r_{\Delta y}$, $\pi$, $100(1/\beta - 1)$, $l$, $\gamma$, $\rho_{ga}$, and $\alpha$ under the diffuse priors in the unconstrained model.
Figure A3. Joint posterior densities of $(\mu_w, \xi_w)$, $(\mu_w, \mu_p)$, $(\mu_w, \sigma_w)$ and $(\mu_w, \iota_w)$.
Figure A4. Joint posterior densities of $(\rho_w, \xi_w)$, $(\mu_p, \rho_w)$, $(\sigma_p, \rho_w)$ and $(\iota_w, \rho_w)$.
Figure A5. Joint posterior densities of $(\mu_p, \xi_w)$, $(\sigma_p, \xi_w)$, $(\iota_w, \xi_w)$, $(\iota_w, \sigma_p)$ and $(\mu_p, \iota_w)$.
Figure A6. Marginal posterior densities of $\xi_w$, $\iota_w$, $\mu_p$ and $\sigma_p$ under the diffuse priors in the restricted model.