

# Estimating Dynamic Macroeconomic Models: How Informative Are the Data?

Beltran, Daniel O., David Draper

Please cite paper as:  
Beltran, Daniel O., David Draper (2016). Estimating Dynamic Macroeconomic Models International Finance Discussion Papers 1175.

<http://dx.doi.org/10.17016/IFDP.2016.1175>



**International Finance Discussion Papers**  
Board of Governors of the Federal Reserve System

Number 1175  
August 2016

Board of Governors of the Federal Reserve System

International Finance Discussion Papers

Number 1175

August 2016

Estimating Dynamic Macroeconomic Models:  
How Informative Are the Data?

Daniel O. Beltran  
David Draper

NOTE: International Finance Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to International Finance Discussion Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors. Recent IFDPs are available on the Web at [www.federalreserve.gov/pubs/ifdp/](http://www.federalreserve.gov/pubs/ifdp/). This paper can be downloaded without charge from Social Science Research Network electronic library at <http://www.ssrn.com/>.

# Estimating dynamic macroeconomic models: How informative are the data?

Daniel O. Beltran<sup>1</sup>

*Federal Reserve Board of Governors, USA*

and David Draper

*University of California, Santa Cruz, USA*

August 2016

**Summary.** Central banks have long used dynamic stochastic general equilibrium (DSGE) models, which are typically estimated using Bayesian techniques, to inform key policy decisions. This paper offers an empirical strategy that quantifies the information content of the data relative to that of the prior distribution. Using an off-the-shelf DSGE model applied to quarterly Euro Area data from 1970:3 to 2009:4, we show how Monte Carlo simulations can reveal parameters for which the model’s structure obscures identification. By integrating out components of the likelihood function and conducting a Bayesian sensitivity analysis, we uncover parameters that are weakly informed by the data. The weak identification of some key structural parameters in our comparatively simple model should raise a red flag to researchers trying to draw valid inferences from, and to base policy upon, complex large-scale models featuring many parameters.

*Keywords:* Bayesian estimation, econometric modeling, Kalman filter, likelihood, local identification, Euro Area, MCMC, policy-relevant parameters, prior-versus-posterior comparison, sensitivity analysis.

*JEL codes:* C11, C18, F41.

## 1. Introduction

Large-scale time series models have played a major role for the last several decades in the setting of macro-economic policy in advanced and emerging economies. In particular, one class of such models — *dynamic stochastic general equilibrium* (DSGE) models (e.g., Kydland and Prescott (1982), Rotemberg and Woodford (1997)) — has gained increasing prominence in macroeconomics and econometrics over the past 25 years. DSGE models, which describe macro-level economic phenomena using micro-economic principles, are used in parallel with other models by central banks worldwide to inform policy decisions. These models typically examine inter-relationships over time among key economic indicators such as output, inflation, and interest rates.

When studying the properties of DSGE models, one needs to assign values to the parameters. Recognizing that maximum-likelihood estimation may well not be straightforward (because the likelihood function may contain many local maxima), and that estimation often produces counterintuitive results (because any DSGE model is a stylized — and thus misspecified — representation of real-world interactions between economic variables), researchers have long preferred to *calibrate* these models by placing point-mass distributions on parameter values, based on a-priori information or long-term features of the data. However, as DSGE models

---

<sup>1</sup>*Address for correspondence:* Federal Reserve Board, 20th and C St. NW, Mail Stop 38, Washington, DC. 20551. E-mail: daniel.o.beltran@frb.gov

have grown in complexity to incorporate more realistic features of the data, it has become less obvious how to calibrate many of the new parameters that have emerged. Furthermore, analyses of calibrated DSGE models are not always robust to alternative calibrations. Bayesian techniques are well suited to address this calibration problem, because they provide a formal way to estimate the parameters by combining prior information about them with the data, as viewed through the lens of the model being analyzed. This offers hope that calibration may no longer be needed, as long as the data do indeed have something to say about plausible parameter values.

As surveyed in Schorfheide (2011), researchers estimating DSGE models are generally aware that identification problems can prevent certain parameters from being consistently estimated (although, regrettably, most published Bayesian studies do not show plots of the prior and marginal posterior distributions, which may help to uncover parameters that are weakly informed by the data). Using limited-information methods, Canova and Sala (2009) demonstrate that many structural parameters in stylized New-Keynesian DSGE models are not identified. Iskrev (2010) and Komunjer and Ng (2011) develop necessary and sufficient rank conditions for assessing identifiability of DSGE model parameters. While these methods can be used to detect failure of identification coming from the structure of the DSGE model, they do not address the problem of making valid statistical inference in the presence of weak identification due to insufficient or inadequate data.

Guerron-Quintana et al. (2013) find that

- in weakly identified DSGE models, classical confidence sets and Bayesian credible sets will not coincide even asymptotically, and
- summary estimators of the posterior distribution such as the mean, median and mode may not be consistent.

Using the inverse of the likelihood ratio statistic and the inverse of the Bayes Factor (BF), they construct confidence intervals that are asymptotically valid from a frequentist point of view regardless of the strength of identification. However, creating the BF confidence set requires numerically performing pairwise tests over the entire parameter space, which they describe as “computationally challenging” because of the large number of parameters that are typically present in DSGE models. To reduce the computational burden, they approximate the BF intervals using Monte Carlo realizations of the prior and posterior distributions, the latter computed using Bayesian estimation methods. One may question the coverage accuracy of their BF confidence interval, given that it is based on a potentially inconsistent posterior distribution; the authors acknowledge that “the accuracy of the corresponding BF confidence sets can be sensitive to the choice of prior” (p. 30). Another drawback of their approach is that if one wishes to approximate the entire posterior distribution rather than just the upper and lower bounds of a single confidence set — thereby permitting, e.g., an examination of the tails of the distribution, which can be used to inform stress-test analysis or worse-case scenarios for economic decision-making — it is necessary to repeat the method with a potentially large number of confidence coefficients, which places an even greater computational burden on the approach.

When maximum-likelihood techniques are used to estimate DSGE models (e.g., Ireland (2003)), the parameter estimates come purely from the data (although, as we discuss below, if the likelihood function is essentially flat for a parameter, the precise “maximum” found by

numerical maximization may be largely arbitrary), and without (explicit) controversy over the role of priors. But if the data alone are not sufficient to identify all parameters of a model, the use of priors and Bayesian techniques is sensible, as long as the priors themselves are also sensible. The main message of this paper is that *if one is to achieve identification through the Bayesian approach, this should be done transparently*. That is, the Bayesian approach should not only be used to derive parameter estimates; it should also reveal which parameters are most sensitive to the prior specifications: if the model is to be used for policymaking, it is crucial to know how strongly the empirical results depend on the prior information. For researchers estimating DSGE models using Bayesian techniques, this paper offers an empirical strategy that strives to unveil the information content of the data relative to that of the prior distribution. Our approach is natural from a statistical perspective but, surprisingly, is not in routine use by investigators fitting complicated econometric models.

Before estimating the model, we check for lack of identification due to the model's structure, which could imply, for example, that a parameter is indistinguishable from another in terms of how it propagates the exogenous shocks. After repairing the model by removing unidentified parameters, we begin our empirical strategy by integrating out components of the likelihood function defined by the data. Our plots of the marginal likelihood densities derived from MCMC methods then reveal which parameters are weakly informed by the data.

*Notational convention.* In this paper the phrase *marginal likelihood* has a literal, rather than standard, meaning: we treat the normalized likelihood function as a density for the parameters, as if it were a posterior with uniform priors (which it is), and we integrate out elements of the parameter vector to obtain marginal distributions for the remaining parameters. (This is in contrast to the use of the phrase *marginal likelihood* to denote the numerator and denominator quantities in Bayes factors, a common (e.g., Bernardo and Smith (2000)) but entirely different calculation.)

After having identified the dimensions along which the likelihood function is relatively flat, we conduct a Bayesian sensitivity analysis using three sets of priors, which differ in information content from low to moderate. The Bayesian analysis complements the analysis of the likelihood function in helping to diagnose which parameters are weakly identified by the data, and also provides a reality check on whether the posterior estimates are being driven mainly by the prior or the data.

The plan for the remainder of the paper is as follows. In Section 2 we describe our data resources and sketch the model we fit, with more details on the model and the fitting process given in the Appendix. In Section 3 we estimate the model using artificial data, to check that the model's structure is not creating an identification problem. Section 4 provides results from likelihood analysis. In Section 5 we offer results from the Bayesian sensitivity analysis (with prior elicitation details given in Section A2 of the Appendix), and Section 6 concludes with a discussion.

## 2. Data resources and the Smets-Wouters model

Building on work by Christiano et al. (2005), Smets and Wouters (2003, hereafter SW) developed and estimated a large-scale dynamic stochastic general equilibrium model for the Euro Area. Using summary statistics such as marginal likelihoods, Bayes factors, and root-mean-squared errors, they showed that their DSGE model fits the data as well as standard non-theoretical VAR models and Bayesian VAR models estimated using the same data. This implies that their DSGE model does at least as good a job as the VAR models in predicting

the observable data series over the sample period. The model’s good fit largely derives from a number of frictions that generate persistence in the propagation of shocks, such as sticky prices and wages, external habit persistence, investment adjustment costs, and other features such as variable capacity utilization. SW describe the full non-linear model and derive a linearized version of it, which we summarize in Section A1 of the Appendix. Its basic structure involves 28 log-linearized time series that link output, inflation, real wages, investment, capital stock, hours worked, firms’ marginal costs, the real interest rate, and employment to ten structural shocks, six of which are allowed to be serially correlated.

The model has 32 parameters. The key structural parameters include the degree of relative risk aversion ( $\eta_C$ ), external habit-formation ( $h$ ), elasticity of work effort with respect to the real wage ( $\eta_L$ ), rigidity of goods prices ( $\xi_p$ ) and wages ( $\xi_w$ ), productivity persistence ( $\rho$ ) and three policy parameters relating to the lagged interest rate ( $\rho_i$ ), inflation ( $r_\pi$ ), and the output gap ( $r_y$ ).

The log-likelihood is computed using a Kalman filter (Hamilton (1994)); the Euro Area dataset is described in Fagan et al. (2001). The data comprise quarterly time series for seven key macroeconomic variables: real GDP, real consumption, real investment, real wages, employment, the GDP deflator, and the nominal interest rate. Following SW, we detrend real variables by their linear trend, and we detrend inflation and the nominal interest rate by the same linear trend in inflation. We extended the sample period used in SW to cover the most recent decade, so our data span the period 1970:3 to 2009:4, with the first 12 observations (3 years) being used to initialize the Kalman filter (later years could of course be included, but this data set is sufficient to make our methodological points).

The prior-posterior plots shown in SW suggest that data are not informative for 10 out of the 32 parameters, which makes the likelihood function nearly flat along these dimensions and yields marginal posteriors that closely resemble their prior distributions. Onatski and Williams (2010) also find evidence of weak parameter identification when they re-estimate the same model; when using Uniform priors they obtain “substantially different” parameter estimates.

In re-estimating the SW model, Onatski and Williams (2010) paid special attention to how strongly the empirical results depend on the prior assumptions. However, in using “less informative” (Uniform) priors they encountered two problems. First, they found “numerous local minima which confounded many optimization methods.” We also find that standard methods for maximizing the log-likelihood function are sharply inadequate with this model; we developed a new *maximization/adaptive-MCMC/maximization* algorithm to address this issue (see Section A3 of the Appendix). The second problem Onatski and Williams (2010) encountered is that for 9 of the 32 parameters, the maximum likelihood estimates are on the boundaries of the prior range. Onatski and Williams (2010) acknowledge this boundary problem but dismiss it by stating that “To the extent that we set our prior to reflect reasonable ranges of estimates, this is not troubling in itself. But it does suggest that the data may favor some parameter values which may be implausible from an economic viewpoint, and hence are not in the support of our prior” (p. 154). What is more worrying is that, because the data seem to favor parameter values that are far from the arbitrary prior boundaries they chose, their point estimates depend critically on their boundary assumptions — in other words, while Onatski and Williams (2010) sought to make their priors “less informative” by using Uniform distributions, the end result is that they inadvertently calibrated these parameters at the prior boundaries. By integrating over the likelihood function (instead of maximizing

it, as Onatski and Williams (2010) did), our estimation strategy will gauge the degree of parameter uncertainty, even for those parameters whose values seem implausible from an economic viewpoint.

### 3. Checking identifiability problems created by the model’s structure

One possibility for problems arising from the fitting of DSGE models is that the model’s structure could imply, for example, that one parameter is indistinguishable from another in terms of how they propagate the exogenous shocks to the state variables. Iskrev (2010) examines local identifiability by checking that the Jacobian matrix of the mapping, of the deep parameters to the parameters that determine the first two moments of the data, is not rank deficient. Because numerical derivatives tend to be inaccurate for highly non-linear functions, Iskrev develops a method for computing the Jacobian analytically. Iskrev applies this method to the Smets and Wouters (2007) model (an extended version of the model examined in this paper), and find that the Jacobian is rank deficient. Iskrev attributes the lack of identifiability to the similar roles played by the two curvature parameters for the goods and labor markets, and the Calvo wage and price parameters. These parameters play similar roles in the nonlinear version of the model, but become equivalent in the linearized version that is estimated. To fix the identifiability problem, Iskrev (2010) (as well as Smets and Wouters (2007)) calibrate these curvature parameters. Similarly, Komunjer and Ng (2011) obtain necessary and sufficient rank conditions for identifiability using the spectral density of the endogenous variables in the model. This approach requires extensive use of numerical derivatives and Kronecker-product matrices. Komunjer and Ng (2011) find that computation of the rank is sensitive to the tolerance level that is used to determine whether the eigenvalues are sufficiently small. They address this issue by performing the analysis using a variety of tolerance levels.

As recognized in Iskrev and Ratto (2010), computing the analytical derivatives of the Jacobian with respect to the deep parameters is computationally inefficient and requires a large amount of memory allocation, because sparse Kronecker-product matrices are used extensively. In this paper we take a simple, informal approach to diagnose identification problems that are inherent to the model’s structure, similar to that of Adolfson and Lindé (2011). The approach involves simulating the model to generate artificial data and then maximizing the resulting likelihood function (using our new maximization method detailed in Section A3 of the Appendix). In the case of our version of the SW model, this approach correctly reveals the parameters that are not identified due to the structure of the model.

For each simulation, we perform the following steps.

- (1) Randomly draw a vector of parameters  $\theta_{DGP}^*$  from the SW prior distribution.
- (2) Solve the DSGE using  $\theta_{DGP}^*$  to obtain the state-space representation.
- (3) Generate a sample of length  $(1,000 + T)$  of artificial data for the same set of observable variables we have actual data on, by simulating random draws for the IID shocks and feeding them into the state-space representation of the model equations found in step (2). We discarded the first 1,000 observations as burn-ins.
- (4) Given the artificial data of sample size  $T$ , search for the parameter vector  $\hat{\theta}$  that maximizes the log-likelihood.

(5) Compute absolute bias in percent as  $\left| \frac{\hat{\theta}}{\theta_{DGP}} - 1 \right| \times 100$ .

We performed 2,000 such simulations for a sample of  $T = 118$  (the number of quarters of data in the original SW dataset) and another 2,000 simulations with  $T = 1,000$  (a time series of obviously infeasible length). Table 1 gives the median of the absolute biases of the maximum likelihood estimates for both sets of simulations. When  $T = 118$ , 13 out of the 32 parameters have biases of 10 percent or more, with some of the largest biases found in  $\sigma_\pi$ ,  $\sigma_Q$ ,  $\sigma_L$ ,  $\sigma_I$ ,  $\sigma_b$ ,  $\psi$ ,  $r_y$ ,  $r_{\Delta y}$ , and  $\lambda_l$ . When we increased the sample size to  $T = 1,000$ , the bias was reduced for most parameters by 60 to 83 percent. However, as highlighted in bold font in the table, the increased sample size barely reduced the bias for  $\rho_\pi$  and  $\sigma_\pi$ . The small reduction in the bias of these two parameters suggests that the structure of the model is likely preventing them from being identified.  $\rho_\pi$  and  $\sigma_\pi$  govern the persistence and the standard deviation of the inflation objective shock that enters the interest rate equation, respectively; the inflation objective shock and the policy shock enter additively in the same equation. In principle, the two shocks should be identifiable, because the inflation objective shock is assumed to be autocorrelated in the data-generating process (DGP) while the policy-rule shock is not. However, when  $\rho$  is high the model can generate an autocorrelated interest rate even when the inflation objective shock is not autocorrelated, making it difficult to distinguish the two shocks (as evidenced by our simulation with  $T = 1,000$ ). SW address this identification problem by placing a tight prior on the autocorrelation parameters, whereas Onatski and Williams (2010) fix the problem by eliminating the policy rule shock. We take the latter approach, because we wish to compare the results of our estimation exercise with those of Onatski and Williams.

Although removing the policy-rule shock fixes the lack of identification inherent in the model’s structure, when taking the model to the actual data we may still find that some parameters are weakly identified. This could arise because we are using data on only seven observable variables to estimate a model that has 9 structural shocks (even after eliminating the policy-rule shock). SW attempt to identify the 10 shocks in their model by assuming that (a) they are uncorrelated with each other and (b) six of them are autocorrelated, treating the other four as white-noise processes. For the six autocorrelated shocks, SW impose fairly strong persistence by setting the prior mean on the autocorrelation coefficient to 0.85, with prior standard deviation of 0.10. However, without these priors, if the actual data favor low values for these autocorrelation parameters, some of the shocks will be difficult to identify. This is likely the case for  $\sigma_Q$ , which has a median absolute bias of 39 percent even with 1,000 observations. In the next Section we show how a close examination of the log-likelihood function can reveal which parameters are weakly informed by the data.

#### 4. Likelihood analysis

The “less informative” Uniform prior used by Onatski and Williams (2010) [OW], shown in column (1) of Table 2, restricts the range of values the parameters can take to what OW regard as the “plausible” region. To derive their point estimates, which are shown in column (2), OW maximize the log-likelihood function in the region defined by their prior bounds. As highlighted in bold, many of these estimates are at the prior boundary, suggesting that the data favor implausible values for these parameters. When using the same data and our own maximization algorithm (described in Section A3 of the Appendix), we obtain similar results (column 3) except for the Calvo employment parameter  $\xi_e$ , which gravitates in our results to the opposite prior boundary than that chosen by OW’s method. In column 4 we extend

Table 1: Median absolute biases (in percent) from 2,000 Monte Carlo simulations comparing the parameters  $\theta_{DGP}$  of the data-generating process with their maximum likelihood estimates  $\hat{\theta}$ , using  $T = 118$  and  $T = 1,000$  quarters of simulated data (see text for explanations of bold font).

| Parameter                      |                          | $T = 118$ | $T = 1000$ | Percent change |
|--------------------------------|--------------------------|-----------|------------|----------------|
| $\varphi$                      | Inverse adjustment cost  | 13        | 4          | -73            |
| $\lambda_C$                    | Risk aversion            | 13        | 3          | -72            |
| $h$                            | Habit persistence        | 3         | 1          | -71            |
| $\xi_w$                        | Calvo wages              | 3         | 1          | -71            |
| $\lambda_l$                    | Labor utility            | 23        | 6          | -75            |
| $\xi_p$                        | Calvo prices             | 3         | 1          | -72            |
| $\xi_e$                        | Calvo employment         | 4         | 1          | -73            |
| $\gamma_w$                     | Wage indexation          | 9         | 3          | -69            |
| $\gamma_p$                     | Price indexation         | 8         | 2          | -70            |
| $\psi$                         | Capital utilization cost | 32        | 5          | -83            |
| $\phi$                         | Fixed cost               | 3         | 1          | -77            |
| <i>Policy rule</i>             |                          |           |            |                |
| $r_\pi$                        | Inflation                | 9         | 2          | -74            |
| $r_{\Delta\pi}$                | Inflation gradient       | 8         | 2          | -71            |
| $\rho$                         | Lag interest rate        | 2         | 1          | -73            |
| $r_y$                          | Output gap               | 25        | 7          | -72            |
| $r_{\Delta y}$                 | Output gap gradient      | 24        | 6          | -75            |
| <i>Shocks, autocorrelation</i> |                          |           |            |                |
| $\rho_a$                       | Productivity             | 3         | 1          | -71            |
| $\rho_b$                       | Preference               | 3         | 1          | -77            |
| $\rho_G$                       | Government spending      | 4         | 1          | -75            |
| $\rho_L$                       | Labor supply             | 8         | 2          | -79            |
| $\rho_I$                       | Investment               | 6         | 1          | -75            |
| $\rho_\pi$                     | Inflation objective      | 13        | 10         | <b>-24</b>     |
| <i>Shocks, SD</i>              |                          |           |            |                |
| $\sigma_a$                     | Productivity             | 10        | 3          | -73            |
| $\sigma_b$                     | Preference               | 25        | 6          | -77            |
| $\sigma_G$                     | Government spending      | 5         | 2          | -61            |
| $\sigma_I$                     | Investment               | 53        | 15         | -71            |
| $\sigma_L$                     | Labor supply             | 49        | 12         | -75            |
| $\sigma_p$                     | Price markup             | 6         | 2          | -69            |
| $\sigma_w$                     | Wage markup              | 5         | 2          | -63            |
| $\sigma_R$                     | Interest rate            | 7         | 2          | -73            |
| $\sigma_Q$                     | Equity premium           | 97        | <b>39</b>  | -60            |
| $\sigma_\pi$                   | Inflation objective      | 98        | 87         | <b>-11</b>     |

Table 2: *Maximum-likelihood estimates of the parameters in the DSGE model (see text for explanations of column headings and bold font).*

|                                |                          | OW Prior     | OW Estimate  | Our Estimates        |                 |                 |
|--------------------------------|--------------------------|--------------|--------------|----------------------|-----------------|-----------------|
|                                |                          |              |              | OW Prior             |                 | Wider Prior     |
| Parameter                      |                          | Range<br>(1) | (2)          | Original Data<br>(3) | New Data<br>(4) | New Data<br>(5) |
| $\varphi$                      | Inverse adjustment cost  | 3.57–8.33    | 6.579        | 6.332                | 7.260           | 418.4           |
| $\lambda_C$                    | Risk aversion            | 1.00–4.00    | 2.178        | 2.953                | 3.299           | 998.7           |
| $h$                            | Habit persistence        | 0.40–0.90    | <b>0.400</b> | 0.400                | 0.400           | 0.000           |
| $\xi_w$                        | Calvo wages              | 0.65–0.85    | 0.704        | 0.708                | 0.760           | 0.870           |
| $\lambda_l$                    | Labor utility            | 1.00–3.00    | <b>3.000</b> | 3.000                | 3.000           | 842.3           |
| $\xi_p$                        | Calvo prices             | 0.40–0.93    | <b>0.930</b> | 0.930                | 0.930           | 0.979           |
| $\xi_e$                        | Calvo employment         | 0.40–0.80    | <b>0.400</b> | 0.800                | 0.800           | 0.874           |
| $\gamma_w$                     | Wage indexation          | 0.00–1.00    | <b>0.000</b> | 0.242                | 0.024           | 0.000           |
| $\gamma_p$                     | Price indexation         | 0.00–1.00    | 0.323        | 0.307                | 0.181           | 0.154           |
| $\psi$                         | Capital utilization cost | 2.80–10.00   | <b>2.800</b> | 2.800                | 10.00           | 0.297           |
| $\phi$                         | Fixed cost               | 1.00–1.80    | <b>1.800</b> | 1.800                | 1.800           | 2.000           |
| <i>Policy rule</i>             |                          |              |              |                      |                 |                 |
| $r_\pi$                        | Inflation                | 1.00–4.00    | <b>4.000</b> | 4.000                | 2.182           | 3.621           |
| $r_{\Delta\pi}$                | Inflation gradient       | 0.00–0.20    | 0.181        | 0.169                | 0.088           | 0.035           |
| $\rho$                         | Lag interest rate        | 0.60–0.99    | 0.962        | 0.957                | 0.945           | 0.989           |
| $r_y$                          | Output gap               | 0.00–1.00    | 0.062        | 0.000                | 0.158           | –0.195          |
| $r_{\Delta y}$                 | Output gap gradient      | 0.00–1.00    | 0.319        | 0.434                | 0.409           | 0.159           |
| <i>Shocks, autocorrelation</i> |                          |              |              |                      |                 |                 |
| $\rho_a$                       | Productivity             | 0.00–1.00    | 0.957        | 0.961                | 0.973           | 0.998           |
| $\rho_b$                       | Preference               | 0.00–1.00    | 0.876        | 0.913                | 0.897           | 0.953           |
| $\rho_G$                       | Government spending      | 0.00–1.00    | 0.972        | 0.901                | 0.934           | 0.956           |
| $\rho_L$                       | Labor supply             | 0.00–1.00    | 0.974        | 0.986                | 0.967           | 0.999           |
| $\rho_I$                       | Investment               | 0.00–1.00    | 0.943        | 0.967                | 0.914           | 0.799           |
| $\rho_\pi$                     | Inflation objective      | 0.00–1.00    | 0.582        | 0.746                | 0.791           | 0.224           |
| <i>Shocks, SD</i>              |                          |              |              |                      |                 |                 |
| $\sigma_a$                     | Productivity             | 0.00–6.00    | 0.343        | 0.542                | 0.543           | 0.633           |
| $\sigma_b$                     | Preference               | 0.00–4.00    | 0.240        | 0.220                | 0.278           | 12.9            |
| $\sigma_G$                     | Government spending      | 0.00–4.00    | 0.354        | 0.352                | 0.391           | 0.382           |
| $\sigma_I$                     | Investment               | 0.00–1.00    | 0.059        | 0.075                | 0.210           | 12.2            |
| $\sigma_L$                     | Labor supply             | 0.00–36.00   | 2.351        | 2.724                | 3.345           | 739.9           |
| $\sigma_p$                     | Price markup             | 0.00–2.00    | 0.172        | 0.197                | 0.253           | 0.240           |
| $\sigma_w$                     | Wage markup              | 0.00–3.00    | 0.246        | 0.267                | 0.207           | 0.196           |
| $\sigma_Q$                     | Equity premium           | 0.00–7.00    | <b>7.000</b> | 7.000                | 7.000           | 433.7           |
| $\sigma_\pi$                   | Inflation objective      | 0.00–1.00    | <b>1.000</b> | 0.716                | 1.000           | 2.645           |

the sample period by a decade and re-estimate the parameters using the same prior; many of the estimates remain at the prior boundary. It is interesting to check just how implausible the estimates become when the prior bounds are widened considerably (while still imposing some theoretical constraints). As shown in the last column, the estimates obtained when using the wider prior range are strikingly distant from the ones obtained using the OW prior. In particular, the estimates for  $\varphi$ ,  $\lambda_c$ ,  $\lambda_l$ ,  $\sigma_L$ , and  $\sigma_Q$  are clearly implausible.

Integrating over the likelihood function using MCMC methods reveals that those parameters whose maximum likelihood estimates are implausible are also weakly informed by the data. Figure 1 shows the marginal likelihood plots for the 31 parameters in our model. For each parameter, the shaded region represents the histogram of the MCMC draws from the likelihood function. The thick line just above the horizontal axis denotes the Onatski and Williams (2010) prior range, and the dashed line shows their point estimate. The wide range spanned by the histograms of  $\varphi$ ,  $\lambda_c$ ,  $\lambda_l$ ,  $\psi$ ,  $r_\pi$ ,  $\rho_\pi$ ,  $\sigma_B$ ,  $\sigma_I$ ,  $\sigma_L$ , and  $\sigma_Q$  suggests that these parameters are weakly informed by the data. These likelihood plots also reveal some contradictions with the prior. For example, the likelihood function strongly favors values for  $h$  between 0 and 0.4, which is in contrast with the Onatski and Williams (2010) prior range of 0.4 to 1. It is therefore not surprising that Onatski and Williams (2010) arrived at a point estimate of 0.4 for this parameter, effectively calibrating it at the prior boundary. The data also strongly favor values for  $\xi_p$ ,  $\xi_e$ , and  $\psi$  that are at odds with the prior used by Onatski and Williams (2010).

With insufficient data, some parameters that play similar roles in DSGE models can be difficult to identify. In the case of  $\lambda_c$ , the curvature parameter in the household's utility function, the histogram of the MCMC-based likelihood draws shown in Figure 1 ranges from 200 to 1000, well above the plausibility range of 1–4 typically used in the literature. In the linearized model  $\lambda_c$  governs how sensitive the household's optimal consumption choice is to the real interest rate and the preference shock. As shown in equation (1) of the Appendix, a high value of  $\lambda_c$  will dampen the response of consumption to the preference shock ( $\epsilon_t^b$ ). In the same equation, a high value for  $h$ , which governs the persistence of the habit formation in consumption, would also dampen the response of consumption to a preference shock. It is therefore not surprising that the data favor implausibly high values for  $\lambda_c$  that are balanced (in terms of the log-likelihood value) by values of  $h$  that are much lower than those typically used in the literature.

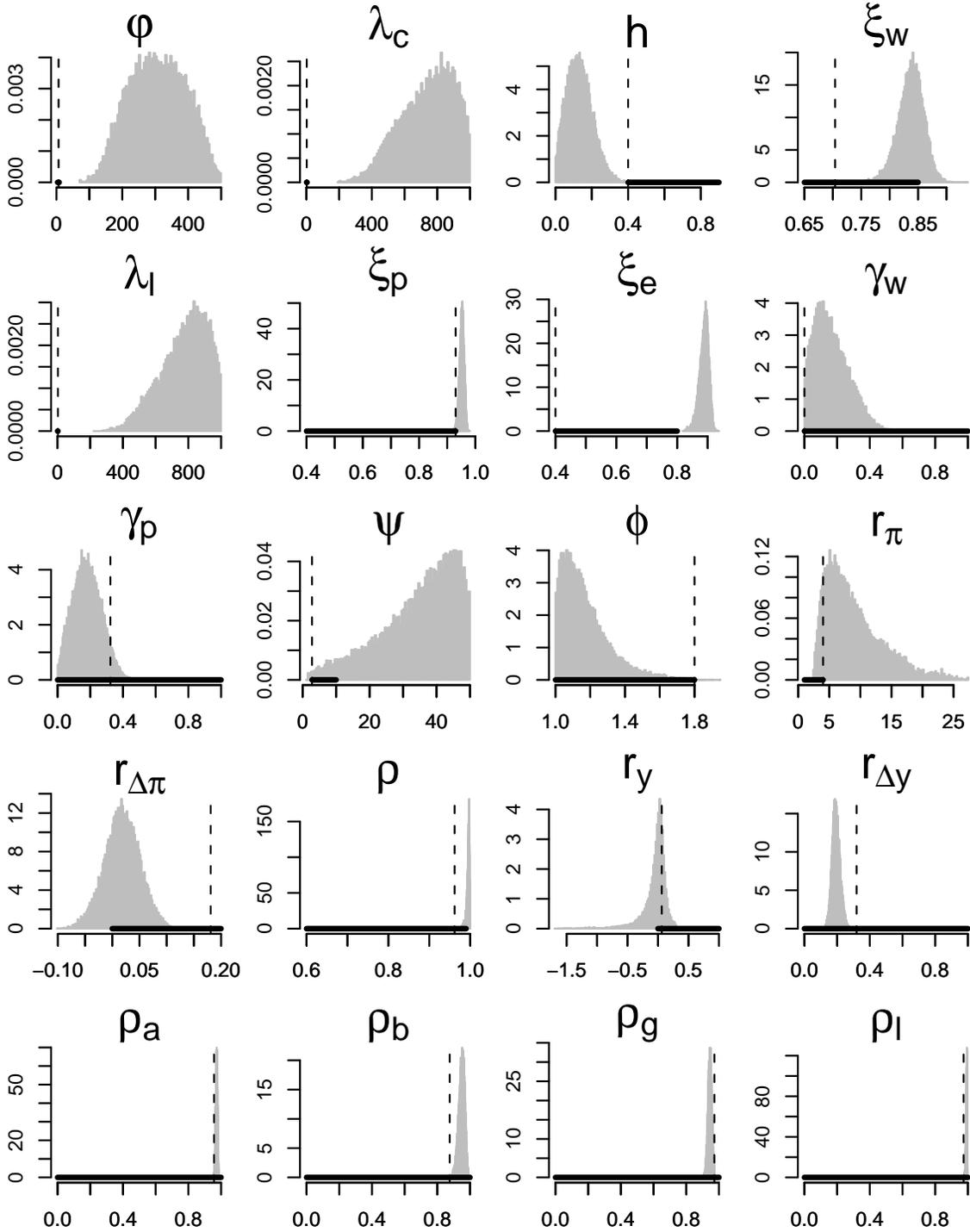
If we had strong prior knowledge that  $h$  should be closer to one (which we do), would such a prior help identify  $\lambda_c$ ?<sup>2</sup> One way to address this question is to examine the log-likelihood surface as a function of  $h$  and  $\lambda_c$ , while keeping the other parameters fixed at their maximum likelihood values. As shown in the top panel of Figure 2, when  $h$  is high (close to 1), the curvature of the log-likelihood surface increases with respect to  $\lambda_c$ , meaning that  $\lambda_c$  is indeed better identified when  $h$  is large. Because the SW and Onatski and Williams (2010) priors for  $h$  rule out low values for this parameter, this restriction likely helped them identify the curvature parameter in the household's utility function. The lesson to take from this example is that if one has good prior information for a parameter, one should use it, for two reasons (one obvious, the other less so): such a prior would not only help to identify the parameter in question, but may also help identify other parameters in the model as well.

If there is little prior information for a parameter, one must of course be careful in choosing the prior range when performing Bayesian estimation, especially if the parameter plays a key

---

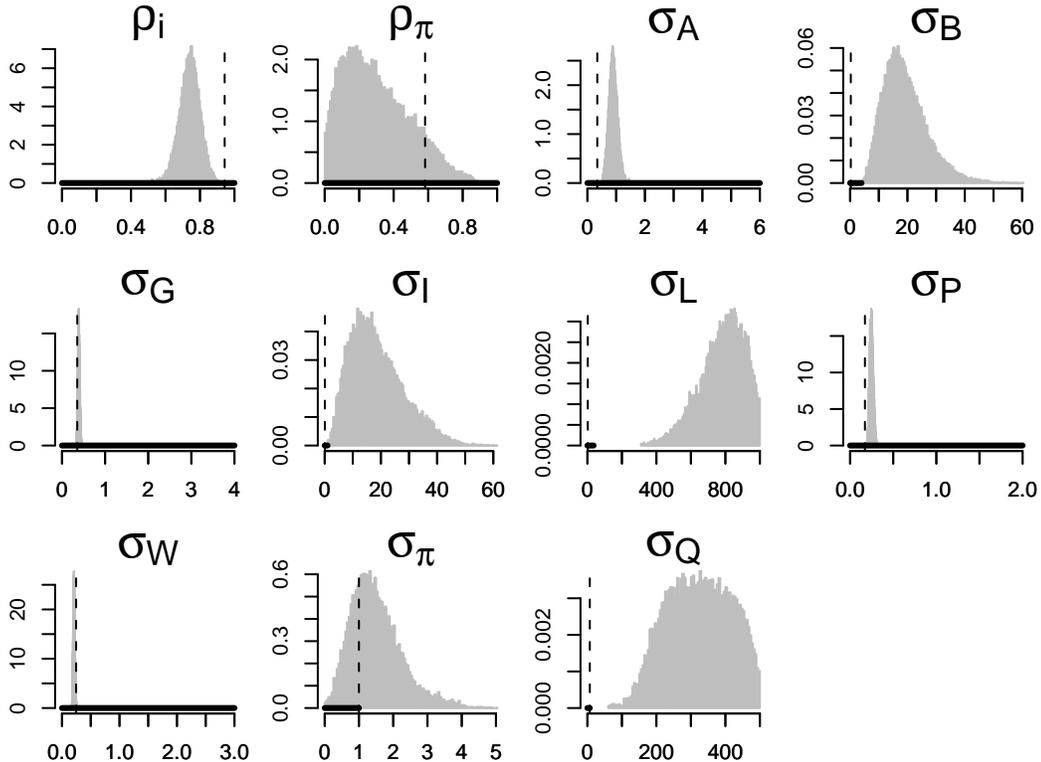
<sup>2</sup>See Section A2 of the Appendix for a discussion of studies that have estimated  $h$ .

Figure 1: *MCMC-based marginal likelihood plots for 20 of the 31 parameters.*



*Note:* The grey regions show histograms of the MCMC draws from the likelihood function. Thick horizontal lines denote the prior ranges from Onatski and Williams (2010), and the dashed vertical lines show the OW point estimates.

Figure 1: (continued). *MCMC-based marginal likelihood plots for the other 11 parameters.*



*Note:* The grey regions show histograms of the MCMC draws from the likelihood function. Thick horizontal lines denote the prior ranges from Onatski and Williams (2010), and the dashed vertical lines show the OW point estimates.

role in the model dynamics. One such parameter in the SW model is  $\psi$ , the elasticity of the capital utilization cost function. A low (respectively, high) value of  $\psi$  implies that the cost of utilizing capital increases slowly (rapidly) with its utilization rate. King and Rebelo (2000) show that variable capacity utilization makes the labor demand curve more elastic with respect to the real wage (or the marginal product of labor). Similarly, Francis and Ramey (2005) and Smets and Wouters (2007) show that capital adjustment costs can help explain the empirical finding of Gali (1999) that productivity shocks have a negative impact on hours worked.

The impact of productivity shocks on hours worked also depends on the elasticity  $\lambda_l^{-1}$  of work effort with respect to the real wage. The bottom panel of Figure 2 plots the likelihood surface as a function of  $\psi$  and  $\lambda_l$  while holding the other parameters fixed at their maximum likelihood values. The log-likelihood surface shows an inverse relationship between these two parameters; because they are poorly informed by the data (Figure 1), placing a strong prior on one of them will likely influence the estimate of the other.

To sum up, the histograms of the MCMC-based likelihood draws reveal that some parameters in the Onatski and Williams (2010) model cannot be identified by the data alone. Furthermore, the surface plots suggest that identification of some parameters could be achieved by placing a strong prior on other parameters. Both SW and Onatski and Williams (2010) achieve identification by incorporating priors into their analysis. However, as evidenced by

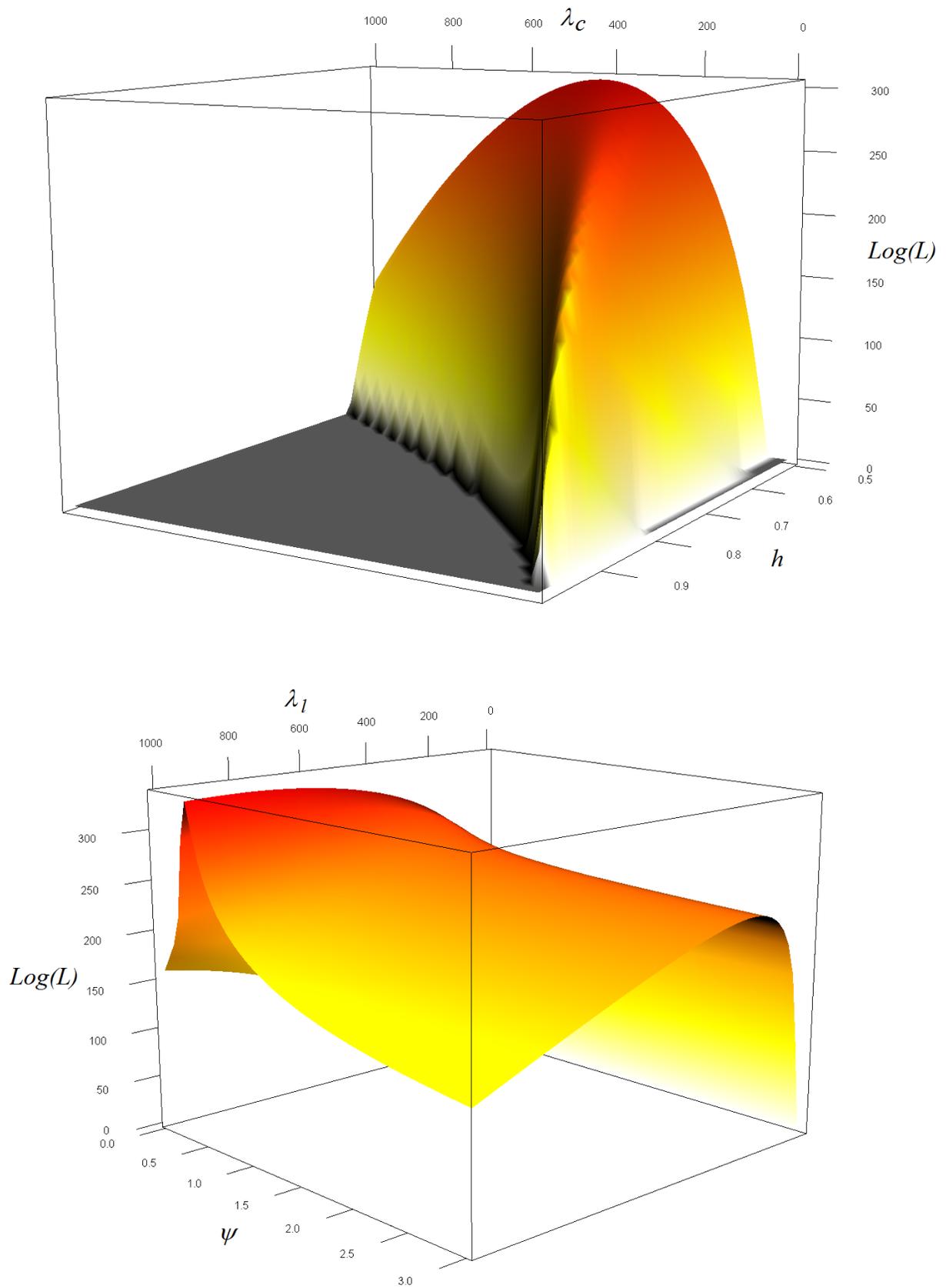


Figure 2: *Log likelihood as a function of  $\lambda_c$  and  $h$  (top) and  $\lambda_l$  and  $\psi$  (bottom).*

the disparity in their estimates, when identification is achieved through the use of priors, the results can be fragile. This is important because knowing how strongly the empirical results depend on the prior information is crucial if the model is to be used for policymaking.

## 5. Bayesian sensitivity analysis

When the data alone are not sufficient to identify all of the model’s parameters, using priors and Bayesian techniques is sensible, provided that the prior information is good. However, as noted in Section 1, if one is to achieve identification through the Bayesian approach, this should be done transparently. That is, the Bayesian estimation exercise should reveal which parameters are most sensitive to prior specification. To perform this sensitivity analysis, we estimated the SW model using three sets of priors: the SW (*informative*) prior, a looser (*somewhat informative*) version of their prior, and a *Uniform* prior. The bounds of the Uniform prior are wider than those used by Onatski and Williams (2010). In Section A2 of the Appendix we document the background literature we used in specifying the prior distributions for each parameter; the exact prior specifications are reported in Table 3.

For some parameters, there is practically no previous research on which to base a prior; often the priors for these parameters are chosen for convenience. In some cases, even though the data have little to say about a given parameter, its posterior estimate is used to inform the prior in subsequent studies. An example is the parameter that governs the elasticity of the capital utilization cost function ( $\psi$ ). Smets and Wouters (2007) normalize this parameter so that it lies in the unit interval and center their prior at 0.5 because they did not have any previous research on which to base it. As recognized by Smets and Wouters, their posterior estimate for  $\psi$  largely coincides with their prior, casting doubt on the insensitivity of this estimate to the prior. Even so, Onatski and Williams (2010) use the Smets and Wouters (2007) posterior estimate for  $\psi$  to inform the boundaries of their Uniform prior, only to arrive at a point estimate that is at the prior boundary. By using three sets of priors, we can determine how strongly our empirical results depend on the prior specifications.

Table 4 summarizes the marginal posterior distributions arising from the three priors, and Figure 3 plots the marginal posterior densities. The marginal posteriors for the weakly-identified parameters ( $\varphi$ ,  $\lambda_c$ ,  $\lambda_l$ ,  $\psi$ ,  $r_\pi$ ,  $\rho_\pi$ ,  $\sigma_B$ ,  $\sigma_I$ ,  $\sigma_L$ , and  $\sigma_Q$ ) vary tremendously when we change the prior, confirming the diagnosis from the previous analysis of the likelihood function. In particular, the posterior estimates of  $\psi$  and  $\phi$  crucially depend on the prior choice, which is unfortunate, because there is practically no information in the literature on which to base the prior. For parameters that are well-identified by the data, such as  $\rho_g$ ,  $\sigma_G$ , and  $\sigma_P$ , the marginal posterior is essentially invariant to the prior choice.

When trying to infer how informative the data are about the parameters of the model, it has become common practice in the econometric literature

- (a) to compare the moments of the posterior and prior distributions using just one set of priors, and
- (b) to conclude that the data provide substantial information if these moments differ substantially.

Our findings demonstrate that this approach could mislead one to believe that a parameter is well identified just because its posterior and prior distributions are different. For example, the SW prior for  $\varphi$  is concentrated on the interval (1.5, 6.5) (Table 3) with a mean of 4, and the

Table 3: *Prior distributions used in the Bayesian sensitivity analysis; see Section A2 of the Appendix for specification details.*

| Parameter( $\theta$ )              | Uniform |     | Somewhat Informative |             |             |      | Informative |             |             |      |
|------------------------------------|---------|-----|----------------------|-------------|-------------|------|-------------|-------------|-------------|------|
|                                    | Bound   |     | $p(\theta)$          | $E(\theta)$ | Percentiles |      | $p(\theta)$ | $E(\theta)$ | Percentiles |      |
|                                    | L       | U   |                      |             | 5%          | 95%  |             |             | 5%          | 95%  |
| $\varphi$ Inverse adjustment cost  | 1       | 100 | IG                   | 4           | 1.49        | 8.96 | N           | 4           | 1.53        | 6.47 |
| $\lambda_c$ Risk aversion          | 0       | 50  | IG                   | 1.5         | 0.36        | 4.07 | N           | 1           | 0.38        | 1.62 |
| $h$ Habit persistence              | 0       | 1   | B                    | 0.7         | 0.32        | 0.96 | B           | 0.7         | 0.52        | 0.85 |
| $\xi_w$ Calvo wages                | 0       | 1   | B                    | 0.75        | 0.47        | 0.95 | B           | 0.75        | 0.66        | 0.83 |
| $\lambda_l$ Labor utility          | 0       | 50  | IG                   | 1.5         | 0.36        | 4.07 | N           | 2           | 0.77        | 3.23 |
| $\xi_p$ Calvo prices               | 0       | 1   | B                    | 0.75        | 0.47        | 0.95 | B           | 0.75        | 0.66        | 0.83 |
| $\xi_e$ Calvo employment           | 0       | 1   | B                    | 0.5         | 0.04        | 0.96 | B           | 0.5         | 0.25        | 0.75 |
| $\gamma_w$ Wage indexation         | 0       | 1   | B                    | 0.5         | 0.04        | 0.96 | B           | 0.75        | 0.47        | 0.95 |
| $\gamma_p$ Price indexation        | 0       | 1   | B                    | 0.5         | 0.04        | 0.96 | B           | 0.75        | 0.47        | 0.95 |
| $\psi$ Capital utilization cost    | 1       | 100 | N                    | 10          | 1.78        | 18   | N           | 5           | 1.91        | 8.09 |
| $\phi$ Fixed cost                  | 1       | 2   | N                    | 1.5         | 1           | 2    | N           | 1.45        | 1.04        | 1.86 |
| <i>Policy rule</i>                 |         |     |                      |             |             |      |             |             |             |      |
| $r_\pi$ Inflation                  | 1       | 10  | N                    | 2           | 1           | 3    | N           | 1.7         | 1.54        | 1.86 |
| $r_{\Delta\pi}$ Inflation gradient | -1      | 1   | N                    | 0.3         | -0.03       | 0.63 | N           | 0.3         | 0.14        | 0.46 |
| $\rho$ Lag interest rate           | 0.5     | 1   | B                    | 0.8         | 0.38        | 1    | B           | 0.8         | 0.61        | 0.94 |
| $r_y$ Output gap                   | -1      | 1   | N                    | 0.13        | -0.04       | 0.29 | N           | 0.13        | 0.04        | 0.21 |
| $r_{\Delta y}$ Output gap gradient | -1      | 1   | N                    | 0.3         | -0.03       | 0.63 | N           | 0.06        | -0.02       | 0.14 |
| <i>Shocks, autocorrelation</i>     |         |     |                      |             |             |      |             |             |             |      |
| $\rho_a$ Productivity              | 0       | 1   | B                    | 0.85        | 0.41        | 1    | B           | 0.85        | 0.66        | 0.97 |
| $\rho_b$ Preference                | 0       | 1   | B                    | 0.85        | 0.41        | 1    | B           | 0.85        | 0.66        | 0.97 |
| $\rho_G$ Government spending       | 0       | 1   | B                    | 0.85        | 0.41        | 1    | B           | 0.85        | 0.66        | 0.97 |
| $\rho_L$ Labor supply              | 0       | 1   | B                    | 0.85        | 0.41        | 1    | B           | 0.85        | 0.66        | 0.97 |
| $\rho_I$ Investment                | 0       | 1   | B                    | 0.85        | 0.41        | 1    | B           | 0.85        | 0.66        | 0.97 |
| $\rho_\pi$ Inflation objective     | 0       | 1   | B                    | 0.85        | 0.41        | 1    | B           | 0.85        | 0.66        | 0.97 |
| <i>Shocks, SD</i>                  |         |     |                      |             |             |      |             |             |             |      |
| $\sigma_a$ Productivity            | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.4         | 0.19        | 0.77 |
| $\sigma_b$ Preference              | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.2         | 0.12        | 0.32 |
| $\sigma_G$ Government spending     | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.3         | 0.16        | 0.53 |
| $\sigma_I$ Investment              | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.1         | 0.07        | 0.14 |
| $\sigma_L$ Labor supply            | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 1           | 0.32        | 2.45 |
| $\sigma_p$ Price markup            | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.15        | 0.09        | 0.23 |
| $\sigma_w$ Wage markup             | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.25        | 0.14        | 0.43 |
| $\sigma_Q$ Equity premium          | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.4         | 0.19        | 0.77 |
| $\sigma_\pi$ Inflation objective   | 0       | 20  | E                    | 2           | 0.10        | 6    | IG          | 0.020       | 0.02        | 0.02 |

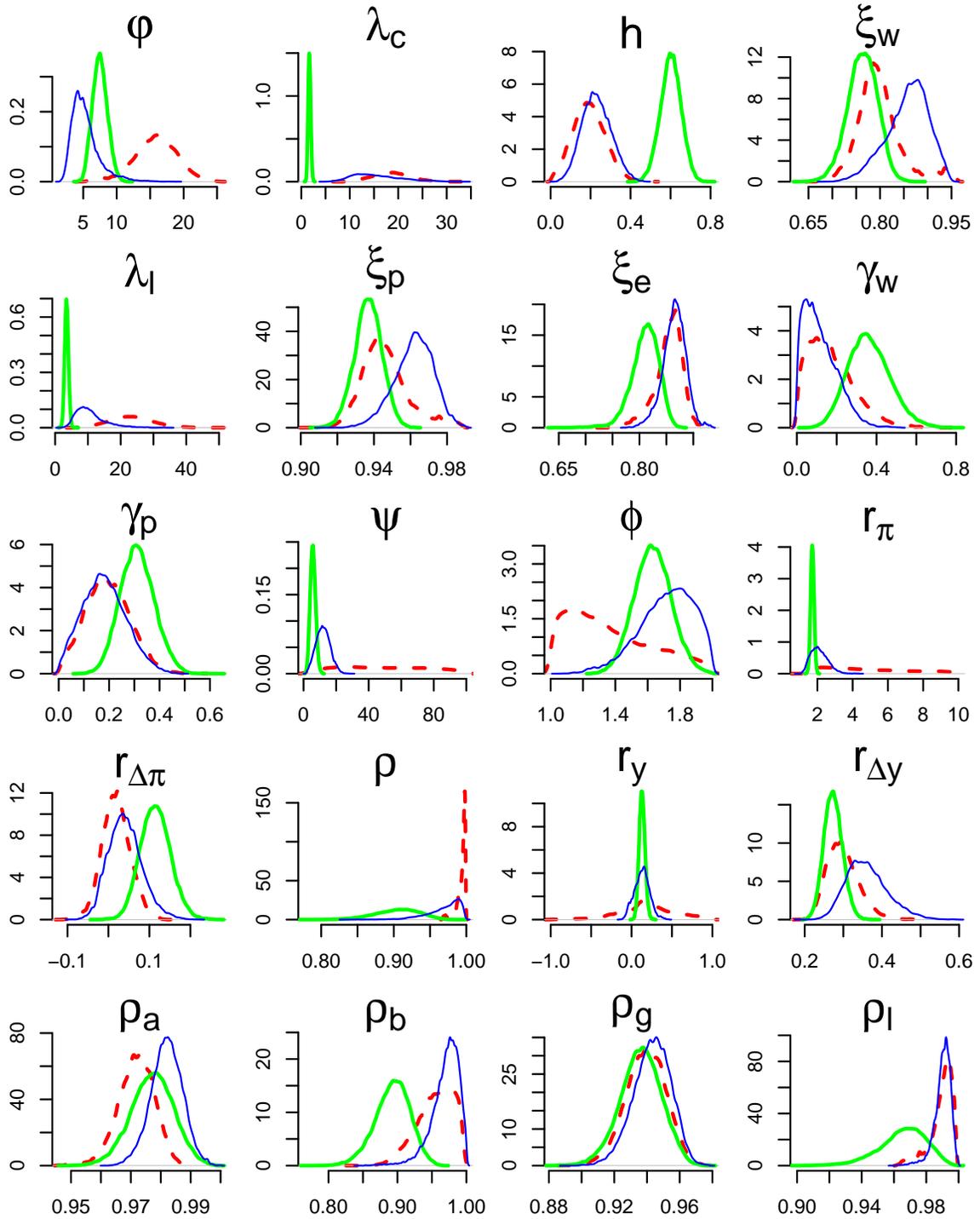
Notes: B = Beta, IG = Inverse-gamma, N = Normal, and E = Exponential; L = lower, U = upper; SD = standard deviation.

Table 4: *Summaries of marginal posterior distributions of the parameters under the three prior distributions described in Table 3.*

|                 | Uniform     |       |        | Somewhat Informative |       |        | Informative |       |        |
|-----------------|-------------|-------|--------|----------------------|-------|--------|-------------|-------|--------|
|                 | Percentiles |       |        | Percentiles          |       |        | Percentiles |       |        |
|                 | 2.5%        | 97.5% | Median | 2.5%                 | 97.5% | Median | 2.5%        | 97.5% | Median |
| $\varphi$       | 9           | 22    | 16     | 3                    | 11    | 5      | 5           | 10    | 7      |
| $\lambda_c$     | 12          | 26    | 19     | 8                    | 27    | 15     | 1           | 2     | 2      |
| $h$             | 0.04        | 0.35  | 0.19   | 0.1                  | 0.38  | 0.23   | 0.5         | 0.7   | 0.6    |
| $\xi_w$         | 0.72        | 0.94  | 0.79   | 0.76                 | 0.93  | 0.86   | 0.7         | 0.82  | 0.76   |
| $\lambda_l$     | 11          | 37    | 23     | 5                    | 23    | 10     | 2           | 5     | 3      |
| $\xi_p$         | 0.92        | 0.98  | 0.94   | 0.94                 | 0.98  | 0.96   | 0.92        | 0.95  | 0.94   |
| $\xi_e$         | 0.79        | 0.9   | 0.86   | 0.82                 | 0.9   | 0.87   | 0.76        | 0.85  | 0.81   |
| $\gamma_w$      | 0.01        | 0.44  | 0.15   | 0.01                 | 0.31  | 0.1    | 0.17        | 0.59  | 0.36   |
| $\gamma_p$      | 0.04        | 0.37  | 0.19   | 0.03                 | 0.36  | 0.18   | 0.18        | 0.45  | 0.31   |
| $\psi$          | 5.91        | 95.04 | 47.57  | 4.03                 | 20.65 | 11.88  | 2.74        | 9.13  | 5.71   |
| $\phi$          | 1.02        | 1.91  | 1.31   | 1.34                 | 1.97  | 1.74   | 1.41        | 1.86  | 1.63   |
| $r_\pi$         | 1.67        | 9.47  | 4.42   | 1.33                 | 3.17  | 2.05   | 1.52        | 1.9   | 1.71   |
| $r_{\Delta\pi}$ | -0.05       | 0.09  | 0.02   | -0.03                | 0.13  | 0.04   | 0.05        | 0.19  | 0.11   |
| $\rho$          | 0.97        | 1     | 1      | 0.91                 | 1     | 0.98   | 0.84        | 0.96  | 0.91   |
| $r_y$           | -0.55       | 0.88  | 0.19   | -0.05                | 0.32  | 0.13   | 0.06        | 0.2   | 0.13   |
| $r_{\Delta y}$  | 0.23        | 0.39  | 0.29   | 0.27                 | 0.48  | 0.36   | 0.23        | 0.32  | 0.27   |
| $\rho_a$        | 0.96        | 0.98  | 0.97   | 0.97                 | 0.99  | 0.98   | 0.96        | 0.99  | 0.98   |
| $\rho_b$        | 0.9         | 0.99  | 0.96   | 0.92                 | 1     | 0.97   | 0.84        | 0.94  | 0.89   |
| $\rho_G$        | 0.91        | 0.96  | 0.94   | 0.92                 | 0.96  | 0.94   | 0.91        | 0.96  | 0.94   |
| $\rho_L$        | 0.97        | 1     | 0.99   | 0.98                 | 1     | 0.99   | 0.93        | 0.99  | 0.97   |
| $\rho_I$        | 0.75        | 0.97  | 0.89   | 0.87                 | 0.99  | 0.96   | 0.88        | 0.98  | 0.94   |
| $\rho_\pi$      | 0.07        | 0.91  | 0.55   | 0.7                  | 1     | 0.94   | 0.71        | 0.96  | 0.88   |
| $\sigma_a$      | 0.48        | 0.94  | 0.68   | 0.43                 | 0.74  | 0.54   | 0.54        | 0.86  | 0.68   |
| $\sigma_b$      | 0.2         | 1.06  | 0.49   | 0.11                 | 0.57  | 0.23   | 0.19        | 0.47  | 0.3    |
| $\sigma_G$      | 0.35        | 0.44  | 0.39   | 0.35                 | 0.44  | 0.39   | 0.35        | 0.44  | 0.39   |
| $\sigma_I$      | 0.18        | 2.06  | 0.57   | 0.1                  | 0.5   | 0.18   | 0.06        | 0.31  | 0.13   |
| $\sigma_L$      | 15          | 20    | 19     | 8                    | 21    | 13     | 2           | 4     | 3      |
| $\sigma_p$      | 0.21        | 0.29  | 0.25   | 0.21                 | 0.29  | 0.24   | 0.2         | 0.27  | 0.23   |
| $\sigma_w$      | 0.18        | 0.24  | 0.21   | 0.18                 | 0.24  | 0.21   | 0.2         | 0.26  | 0.22   |
| $\sigma_\pi$    | 0.34        | 4.35  | 1.68   | 0.17                 | 3.72  | 1.29   | 0.26        | 1.3   | 0.56   |
| $\sigma_Q$      | 9           | 20    | 17     | 2                    | 11    | 5      | 6           | 12    | 9      |

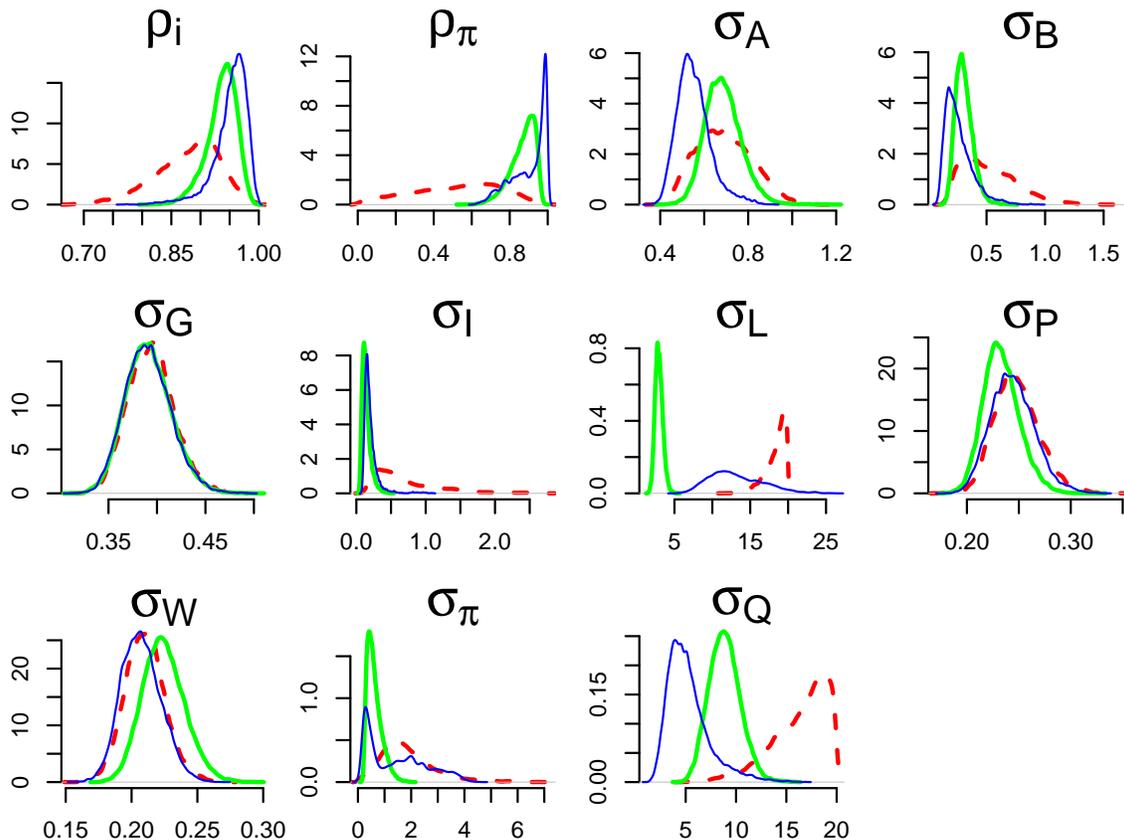
posterior with this prior has most of its mass in the interval (5, 10) (Table 4) with posterior median 7, which is not even in the prior 95% interval. However, it is incorrect to conclude that this difference is because the data swamped the prior. The huge interval spanned by the MCMC draws from the likelihood function for  $\varphi$  (Figure 1) and the high sensitivity of its posterior distribution to the prior choice (Figure 3) confirm that this parameter is actually weakly informed by the data. What has happened here is that the SW prior and the likelihood are sharply at odds with each other, with the prior concentrated on a region of extremely low likelihood, making the SW posterior an unreliable summary of the available information.

Figure 3: MCMC-based marginal posterior plots for 20 of the parameters.



*Note:* Dashed red lines give marginal posteriors under the Uniform priors, thin blue lines under the somewhat informative priors, and thick green lines under the informative priors of Smets and Wouters (2003).

Figure 3: (continued). MCMC-based marginal posterior plots for the remaining 11 parameters.



*Note:* Dashed red lines give marginal posteriors under the Uniform priors, thin blue lines under the somewhat informative priors, and thick green lines under the informative priors of Smets and Wouters (2003).

Ironically, this is an example in which the prior has swamped the data, by restricting the posterior to a range of values that the likelihood function regards as unrealistic.

From this (and other examples like it), we conclude that without a detailed analysis, e.g., of the type we advocate here — simulation with known parameter values to summarize small- and large-sample bias of parameter estimates, careful examination of the likelihood surface, and sensitivity analysis with a range of priors — it is difficult to draw correct conclusions about the amount of information the data provide for the parameters of DSGE models.

## 6. Conclusions and discussion

Recent contributions to the literature on parameter identification in DSGE models have provided several useful tools for flagging lack of identification due to the structure of the DSGE model (e.g. Iskrev (2010), and Komunjer and Ng (2011)). But existing research has paid little attention to the issue of weak identification in DSGE models that arises from insufficient or inadequate data. The main contribution of this paper is to provide an empirical strategy for estimating DSGE models that is aimed at obtaining reasonable parameter estimates through

the use of carefully chosen priors, while at the same time revealing the information content of the data relative to that of the prior distribution. Previous studies have focussed on obtaining reasonable estimates, which is fairly straightforward to achieve in a Bayesian setting if one uses reasonable priors to begin with. But the latter goal of transparency, which has largely been ignored in the DSGE literature, is crucial if the estimates are to be taken seriously for analysis of policy.

We illustrated our approach by estimating an off-the-shelf DSGE model, applied to the Euro Area over the period 1970–2009. Our empirical strategy has three parts: simulation to uncover lack of identification from the model’s structure, MCMC-based likelihood estimation to determine the dimensions along which the likelihood function is relatively flat, and Bayesian sensitivity analysis to gauge the information content of the data relative to that of the prior. Using this approach, we first found that several parameters were not identified because of the model’s structure. After eliminating these parameters from the model, we then took the model to the data and performed an MCMC-based likelihood analysis and a Bayesian sensitivity analysis. We found that roughly one third of the model’s parameters are weakly identified by the data. One such parameter is the elasticity of the capital utilization cost function ( $\psi$ ), which plays a key role in the dynamics of the model because it determines (among other things) the impact of productivity shocks on hours worked. Weak identification of this parameter is troubling because there is not much information in the literature on which to base a prior. Despite the lack of prior information regarding this parameter, previous studies have estimated it using informative priors, obtaining narrow posterior intervals that, in turn, have been used as the prior for subsequent studies (and these later studies have not acknowledged the weak level of data support for their priors).

When estimating DSGE models, it is sensible to incorporate priors by using Bayesian techniques when the data alone are not sufficient to identify all of the model’s parameters, provided that the prior information is credible. But unless this is done in a transparent manner, it is nearly impossible for subsequent readers to judge how well the posterior estimates are informed by the data. This is crucial if the estimates are to be used for policymaking, and also if the estimates will be used to inform priors in subsequent studies. We have shown how integrating over the likelihood function and performing a Bayesian sensitivity analysis can help diagnose weak identification arising from the data. This approach is natural from a statistical perspective but, surprisingly, is not in routine current use by investigators fitting complicated econometric models.

Our other noteworthy finding is that, in models of this type, the naive technique — examples of which may be readily found in econometrics — of {concluding that any parameter whose posterior and prior differ substantially must have been well informed by the data} can be sharply misleading; the parameter  $\theta_1$  in question may be nearly unidentified and yet exhibit strong prior-to-posterior movement, because the *prior* (not the likelihood) has exerted strong influence on the posterior (e.g., by inadvertently restricting the posterior to concentrate on a region of low likelihood).

### Acknowledgments

We would like to thank Federico Ravenna, Carl Walsh, Andrew Levin, Alejandro Justiniano, Doireann Fitzgerald, Luca Guerrieri, Abel Rodríguez and seminar participants at the Federal Reserve Board, the Bureau of Economic Analysis, and the Small Open Economies in a Globalized World conference for their valuable feedback. Grant Long and Zachary Kurtz provided excellent research assistance. The views expressed here are entirely those of the authors, and do not represent those of the Federal Reserve System or of any other person associated with the Federal Reserve System.

# Appendix

## A1. The Smets-Wouters Model

The Smets and Wouters (2003, hereafter SW) model consists of 28 time series equations that link output, inflation, real wages, investment, capital stock, hours worked, firms' marginal cost, the real interest rate, and employment to ten structural shocks, six of which are allowed to be serially correlated. The series are expressed as logarithmic deviations from the steady state. In order to induce intrinsic persistence in the propagation of shocks, the model features sticky wages and prices, external habit persistence in consumption, investment adjustment costs and variable capacity utilization with utilization costs. SW apply their model to data from the Euro Area; as noted in the main paper, we extended the sample period used in SW to cover the most recent decade, so our data span the period 1970:3 to 2009:4.

Household utility depends positively on consumption relative to an external habit stock, and negatively on labor supply. Utility maximization (subject to a budget constraint) implies the following optimal allocation of consumption over time  $C_t$ :

$$C_t = \frac{h}{1+h}C_{t-1} + \frac{1}{1+h}E_t C_{t+1} - \frac{1-h}{(1+h)\lambda_C}(i_t - E_t \pi_{t+1}) + \frac{1-h}{(1+h)\lambda_C}(\epsilon_t^b - E_t \epsilon_{t+1}^b), \quad (1)$$

in which  $h$  is the habit-persistence parameter;  $E_t$  is the rational expectation operator, which averages over uncertainty about future unexpected shocks (conditioning on the model structure, parameters and shock distribution);  $\lambda_C$  is the coefficient of relative risk aversion of households (i.e., the inverse of the intertemporal elasticity of substitution);  $i_t$  is the nominal interest rate;  $\pi_t$  is the rate of inflation; and  $\epsilon_t^b$  is an autoregressive preference shock (with first-order auto-correlation  $\rho_b$  and a mean-0 Gaussian error term with variance  $\sigma_b^2$ ), which affects the intertemporal substitution of households.

Because households supply differentiated labor in an imperfectly competitive market, they can set their own wages. However, only a fraction  $(1 - \xi_w)$  can adjust wages in period  $t$ ; when  $\xi_w = 0$ , wages are perfectly flexible. Households that cannot re-optimize their wages can partially index their wage to past inflation. The degree of wage indexation is determined by the parameter  $\gamma_w$ : when  $\gamma_w = 0$ , wages that cannot be re-optimized remain constant (i.e., in that case there is no wage indexation). Optimal wage setting implies that the real wage  $w_t$  evolves as follows:

$$w_t = \frac{\beta}{1+\beta}E_t w_{t+1} + \frac{1}{1+\beta}w_{t-1} + \frac{\beta}{1+\beta}E_t \pi_{t+1} - \frac{1+\beta\gamma_w}{1+\beta}\pi_t + \frac{\gamma_w}{1+\beta}\pi_{t-1} - \frac{\lambda_w(1-\beta\xi_w)(1-\xi_w)}{(1+\beta)[\lambda_w + (1+\lambda_w)\lambda_L]\xi_w} \left[ w_t - \lambda_L L_t - \frac{\lambda_C}{1-h}(C_t - h C_{t-1}) - \epsilon_t^L - \eta_t^w \right]. \quad (2)$$

Here  $\beta$  is the subjective discount factor;  $\lambda_L$  is the inverse of the elasticity of work effort with respect to the real wage;  $L_t$  is labor demand; the preference shocks to labor supply,  $\epsilon_t^L$ , are autocorrelated with first-order auto-correlation  $\rho_L$  and mean-0 Gaussian errors with variance  $\sigma_L^2$ ; the wage markup shocks  $\eta_t^w$  are IID  $N(0, \sigma_w^2)$ ; and  $\lambda_w$  is determined as follows: when wages are perfectly flexible, the real wage is the markup — equal to  $(1 + \lambda_w)$  — over the ratio of the marginal disutility of labor and the marginal utility of an additional unit of consumption.

The capital stock  $K_t$  grows with investment  $I_t$ , but shrinks every period by the depreciation rate  $\tau$ . The evolution of the capital stock is therefore

$$K_t = (1 - \tau)K_{t-1} + \tau I_{t-1}. \quad (3)$$

A typical household's optimal investment decision is given by

$$I_t = \frac{1}{1 + \beta} I_{t-1} + \frac{\beta}{1 + \beta} E_t I_{t+1} + \frac{\varphi}{1 + \beta} Q_t + \frac{\beta E_t \epsilon_{t+1}^I - \epsilon_t^I}{1 + \beta}, \quad (4)$$

where  $\varphi$  is the inverse of the investment adjustment costs and  $\epsilon_t^I$  is an autocorrelated investment shock (with first-order auto-correlation  $\rho_I$  and mean-0 Gaussian errors with variance  $\sigma_I^2$ ). Here  $Q_t$  is the current real value of the capital stock, with evolution given by

$$Q_t = -(i_t - E_t \pi_{t+1}) + \frac{1 - \tau}{1 - \tau + \bar{r}^k} E_t Q_{t+1} + \frac{\bar{r}^k}{1 - \tau \bar{r}^k} E_t r_{t+1}^k + \eta_t^Q, \quad (5)$$

where  $\bar{r}^k$  is the mean return on capital,  $r_t^k$  is the rental rate of capital and the  $\eta_t^Q$  are IID  $N(0, \sigma_Q^2)$  equity-premium shocks.

Labor demand  $L_t$  is assumed to follow the simple relation

$$L_t = -w_t + \left(1 + \frac{1}{\psi}\right) r_t^k + K_{t-1}, \quad (6)$$

where  $\psi$  measures the elasticity of the capital-utilization cost function. Because there are no consistently-measured data sets on aggregate hours worked in the Euro Area, SW used data on employment instead. Since employment typically responds more slowly to macroeconomic shocks, SW introduced an auxiliary equation linking employment to aggregate hours worked; in other words, employment  $e_t$  evolves according to

$$e_t = \beta e_{t+1} + \frac{(1 - \beta \xi_e)(1 - \xi_e)}{\xi_e} (L_t - e_t), \quad (7)$$

in which a fraction  $(1 - \xi_e)$  of the firms do not adjust employment in any given period.

Only a fraction  $(1 - \xi_p)$  of firms can optimally set prices each period, but those firms that do not re-optimize can still index their prices to past inflation. The degree of price indexation for firms that do not re-optimize is given by the parameter  $\gamma_p$ : when  $\gamma_p = 0$  there is no indexation, and prices of final goods produced by firms that do not re-optimize remain unchanged. Optimal price setting implies that the inflation rate evolves according to

$$\pi_t = \frac{\beta}{1 + \beta \gamma_p} E_t \pi_{t+1} + \frac{\gamma_p}{1 + \beta \gamma_p} \pi_{t-1} + \frac{(1 - \beta \xi_p)(1 - \xi_p)}{(1 + \beta \gamma_p) \xi_p} [\alpha r_t^k + (1 - \alpha)w_t - \epsilon_t^a + \eta_t^p]. \quad (8)$$

In this equation, the inflation rate depends on expected future inflation, past inflation, and the marginal cost of production;  $\alpha$  is the steady-state share of capital in total output,  $\epsilon_t^a$  is an autocorrelated technology shock with first-order auto-correlation  $\rho_a$  (and mean-0 Gaussian errors with variance  $\sigma_a^2$ ), and the  $\eta_t^p$  are IID  $N(0, \sigma_p^2)$  price-markup shocks.

The goods market equilibrium condition, which equates demand and supply of output  $Y_t$ , is

$$\begin{aligned} Y_t &= c_y C_t + g_y \epsilon_t^G + k_y (\tau I_t + \bar{r}^k \psi^{-1} r_t^k) \\ &= \phi [\epsilon_t^a + \alpha (K_{t-1} + \psi^{-1} r_t^k) + (1 - \alpha)L_t]. \end{aligned} \quad (9)$$

Here  $c_y$  is the steady-state ratio of consumption to output;  $g_y$  is the steady-state ratio of government spending to output;  $\epsilon_t^G$  is a government-spending shock, which follows an autoregressive process with first-order auto-correlation  $\rho_G$  (and mean-0 Gaussian errors with variance  $\sigma_G^2$ ); and  $k_y$  is the steady-state capital-output ratio. Equation (9) corrects a slight error that Onatski and Williams (2010) discovered in SW: the production of output is now specified on the right-hand side, where  $\phi$  is 1 plus the share of fixed cost in production.

Finally, to close the model SW adopt a monetary policy reaction function, whereby the monetary authority sets the nominal interest rate in response to inflation and the output gap. The output gap is defined as the difference between

- (a) actual output and
- (b) the output that would prevail under flexible prices and wages ( $Y_t^*$ ) and in the absence of the three “cost-push” shocks ( $\eta_t^w, \eta_t^Q, \eta_t^p$ ).

The policy rule is given by

$$i_t = \rho i_{t-1} + (1 - \rho) \left[ \bar{\pi}_t + r_\pi (\pi_{t-1} - \bar{\pi}_t) + r_y (Y_{t-1} - Y_{t-1}^*) \right] + r_{\Delta\pi} (\pi_t - \pi_{t-1}) + r_{\Delta y} \left[ Y_t - Y_t^* - (Y_{t-1} - Y_{t-1}^*) \right] + \eta_t^R, \quad (10)$$

in which  $\bar{\pi}_t$  is the inflation objective, which follows an autoregressive process with first-order auto-correlation  $\rho_\pi$  (and mean-0 Gaussian errors with variance  $\sigma_\pi^2$ );  $r_\pi$  is the policy response to deviations of lagged inflation from the inflation objective;  $r_y$  is the policy response to deviations in the lagged output gap;  $r_{\Delta\pi}$  is the policy response to current changes in inflation;  $r_{\Delta y}$  is the policy response to current changes in the output gap; and  $\eta_t^R$  is an IID  $N(0, \sigma_R^2)$  policy shock. Because the policy-rule shock is not identified, we follow Onatski and Williams (2010) and eliminate this shock when estimating the model.

In order to determine  $Y_t^*$  (the level of output that would prevail under flexible prices and wages), the model is supplemented with flexible-price versions of equations (1)–(9); see Appendix A1 for further details.

We fit the linear model implied by equations (1–10) above using the algorithm of Sims (2002), which relies on matrix eigenvalue decompositions. In order to derive the likelihood for the data, we write the model’s solution in state-space form,

$$\begin{cases} x_t = F x_{t-1} + Q z_t \\ y_t = H' x_t + v_t \end{cases}, \quad (11)$$

where  $z_t$  is the IID system noise and  $v_t$  is the IID measurement noise. The  $H$  matrix links the observed variables  $y_t$  to the state variables  $x_t$ ;  $F$  and  $Q$  are functions of the model’s parameters. The disturbances  $z_t$  and  $v_t$  are assumed to be Normally distributed with mean zero and covariance matrices  $Q Q'$  and  $R R'$ , respectively. Because our model includes no measurement errors, for us  $R R'$  is just a matrix of zeros. This is a dynamic linear model (West and Harrison (1999)), of which the Kalman filter is a special case. As shown in Hamilton (1994), the Kalman filter can be used to derive the sampling distribution of the data  $y_t$ , conditional on past observations  $\mathcal{Y}_{t-1} \equiv (y'_{t-1}, y'_{t-2}, \dots, y'_1)$ . The likelihood function is defined by the conditional sampling distribution

$$p(y_t | \mathcal{Y}_{t-1}) = (2\pi)^{-n/2} |H' P_{t-1} H + R R'|^{-1/2} \exp \left[ -\frac{1}{2} (y_t - H' \hat{x}_{t,t-1})' (H' P_{t-1} H + R R')^{-1} (y_t - H' \hat{x}_{t,t-1}) \right], \quad (12)$$

where  $\hat{x}_{t|t-1} \equiv \hat{E}(x_t|\mathcal{Y}_{t-1})$  is the linear least-squares forecast of the state vector based on the data observed through time  $(t-1)$  and  $P_{t|t-1}$  is the associated mean-squared-error (MSE) matrix, defined as  $P_{t|t-1} \equiv E[(x_t - \hat{x}_{t|t-1})(x_t - \hat{x}_{t|t-1})']$ .

## A2. Prior distribution elicitation

The DSGE literature has generally followed a sensible approach to specifying priors: the prior mean is typically chosen to match the median of the estimates obtained from previous studies, and the prior standard deviation is wide enough to include at least some of the more extreme values. SW follow this approach for the structural parameters in their model, and their priors were later adopted in numerous other studies (e.g., Levin et al. (2005), Negro et al. (2005), and Adolfson et al. (2007)).

However, in this field one should not necessarily place too much prior mass close to the means obtained from previous studies, because these estimates are sensitive to the choice of modeling assumptions, econometric techniques used, and the specific data sets employed. A DSGE model is a restricted vector-autoregressive (VAR) model in which the equation restrictions are based on economic theory. Because the parameter estimates are conditional on these restrictions, a prior “borrowed” from one model may not be consistent with the data when viewed through the lens of a different model.

If the empirical performance of the model is overly sensitive to different prior assumptions, the model is less useful for policy purposes. To gauge this sensitivity, we estimated the model using three sets of priors: the SW priors, which we refer to as the informative priors; a looser version of the SW priors, which we call the somewhat informative priors; and Uniform priors with fairly wide bounds (even wider than those adopted by Onatski and Williams (2010)).

First, we note that six of the parameters cannot be estimated, because they govern ratios of the different state variables in steady-state. Following SW and Onatski and Williams (2010), we calibrate these parameters at the same (standard) values chosen by SW. The quarterly depreciation rate of capital is set to  $\tau = 0.025$ , the subjective discount factor is pegged at  $\beta = 0.99$  (which implies an annual steady state real interest rate of 4 percent), the wage markup is set to  $\lambda_w = 0.5$ , the Cobb-Douglas production parameter is fixed at  $\alpha = 0.3$ , the steady-state share of consumption relative to output is pegged at  $c_y = 0.6$ , and the steady-state investment share is fixed at  $c_I = 0.22$  (which implies a steady-state ratio of capital to output of  $k_y = 2.2$ ).

For many parameters, the priors need to properly account for theoretical restrictions. For example, SW use inverse-Gamma distributions to specify the priors for the variances of the shocks, which guarantees that they are positive. For parameters that govern shares, and the autocorrelation parameters, the Beta distributions used by SW guarantee that these parameters lie in the unit interval.

- Few studies have shed light on the magnitude  $\varphi$  of the investment adjustment costs. The inverse of this parameter measures the elasticity of investment with respect to a 1 percent increase in the price of installed capital. Using impulse-response matching techniques, Altig et al. (2002) estimate the investment adjustment cost for the United States to be between 7.7 and 20, depending on which impulse responses they try to match. Christiano et al. (2005) obtain a lower estimate of 2.48 for the investment adjustment cost in the United States. These estimates translate into investment elasticities between 0.05 percent and 0.40 percent. For our informative prior, we follow SW and adopt a

normal distribution centered at 4 with standard deviation of 1.5 for  $\varphi$ . Our somewhat informative prior is an inverse-Gamma distribution with mean 4 and a much larger standard deviation of 3. Our Uniform prior ranges from 1 to 100 (implying investment elasticities from 0.01 to 1).

- Many economists prefer to use a constant-relative-risk-aversion (CRRA) value of unity as suggested by Arrow (1971), implying that a CRRA utility depends on the log of income, thus keeping the utility function bounded. For this reason, SW and others use a prior for the coefficient of relative risk aversion or the inverse intertemporal elasticity of substitution ( $\lambda_C$ ) that is centered at 1, with varying degrees of uncertainty. But, as surveyed in Kaplow (2005), more recent CRRA estimates from the financial economics literature often exceed 10. The informative (SW) prior for  $\lambda_c$  is a Normal distribution centered at 1 with standard deviation of 0.375. Our somewhat informative prior is an inverse-Gamma distribution with mean 1.5 and standard deviation of 3. The Uniform prior ranges from 0.001 to 50.
- The degree of habit persistence  $h$  is bounded between 0 and 1. Most empirical studies have found  $h$  to be greater than 0.6. Christiano et al. (2005) estimate an  $h$  of 0.63 for the United States; Fuhrer (2000) finds somewhat higher estimates of 0.8 and 0.9, and the highest estimates found in the literature are those of Bouakez et al. (2005), who estimate a value of 0.98. Our informative prior is a Beta distribution with a mean of 0.7 and standard deviation of 0.1. The somewhat informative prior has the same distribution and mean, but the standard deviation is 0.2. The Uniform prior ranges from 0.001 to 0.999.
- SW set the mean of the Calvo price and wage parameters ( $\xi_p$  and  $\xi_w$ ) to 0.75, so that the average length of the contract is about 1 year. This is in line with some of the estimates of Gali et al. (2001) for the Euro Area. Using monthly consumer price index (CPI) databases from 9 European countries, Dhyne et al. (2005) estimate a median price duration of 10.6 months in Europe. Similarly, a study by Angeloni et al. (2004) finds that European firms on average change prices once a year. Our informative prior is a Beta distribution with mean 0.75 and standard deviation of 0.05; our somewhat informative prior has the same distribution and mean, but a standard deviation of 0.15 instead. The Uniform prior ranges between 0.001 and 0.999.
- There is an entire literature devoted to estimating the inter-temporal elasticity of labor supply ( $\lambda_L^{-1}$ ), which plays an important role in explaining business cycles. Estimates from the micro literature are much lower than required by Real Business Cycle models to match certain “stylized facts” in the economy. In a meta-analysis of 32 micro-based empirical estimates of labor supply elasticities covering 7 European countries and the United States, Evers et al. (2006) find a mean of 0.24 (with a standard deviation of 0.42) for the elasticity of labor supply. Using a contract model, Ham and Reilly (2006) obtain much higher estimates ranging from 0.9 to 1.3. To achieve sensible results, values of 2 and higher are often used to calibrate DSGE models. Smets and Wouters (2007) center their prior at 2, but recognize that their posterior and prior are quite similar, suggesting that the data have little to say about this parameter. The SW informative prior for  $\lambda_L$  is a Normal distribution with mean 2 and standard deviation of 0.75; our somewhat

informative prior is an inverse-Gamma distribution with mean 1.5 and standard deviation 3. The Uniform prior ranges from 0.001 to 50.

- The Calvo-style employment parameter  $\xi_e$  is unique to this model and has no precedent in the literature. It is theoretically bound to the unit interval. Our informative prior is a Beta distribution with mean 0.5 and standard deviation of 0.15; the somewhat informative prior has the same distribution and mean, but a standard deviation of 0.3. The Uniform prior is bounded from 0.001 to 0.999.
- Estimates of the degree of price and wage indexation ( $\gamma_w$  and  $\gamma_p$ ) are also scarce. In a study of European inflation dynamics, Gali et al. (2001) find that backward-looking price setting “has been a relatively unimportant factor behind the dynamics of Euro Area inflation.” (p. 1256) This finding is consistent with values of  $\gamma_w$  and  $\gamma_p$  that are closer to zero. Ignoring this evidence, SW use a Beta distribution with mean 0.75 and standard deviation of 0.15 for their prior. Our somewhat informative prior is a Beta distribution with mean 0.5 and standard deviation 0.3. The Uniform prior has a lower bound of 0.001 and an upper bound of 0.999.
- There is not much reliable prior evidence for the parameter  $\psi$  governing the elasticity of the capital-utilization cost function. For the United States, Smets and Wouters (2007) obtain a 95% posterior interval for this elasticity of (1.4, 2.8), with a posterior mode of 1.9. But because their estimate largely coincides with their prior, we do not know if this parameter is well informed by the data or not. King and Rebelo (2000) use a much higher elasticity of 10 to obtain sensible results for their model simulations, but they did not estimate it from data. Christiano et al. (2005) encountered difficulties in estimating this elasticity (their estimation procedure resulted in implausibly high values), so they simply set it to 100. Our informative (Smets and Wouters (2007)) prior for  $\psi$  is a Normal distribution centered at 5 with standard deviation of 1.88. Our somewhat informative prior for  $\psi$  is Normally distributed with mean 10 and standard deviation 5. The Uniform prior ranges from 1 to 100.
- Smets and Wouters (2007) estimated the share of fixed costs in total production for the United States to be between 48 percent and 73 percent, which is notably higher than their prior centered at 25 percent. Because the share of fixed cost in production is equal to  $(\phi - 1)$ ,  $\phi$  is restricted to be between 1 and 2. Our informative prior for  $\phi$  is a Normal distribution centered at 1.45 with standard deviation of 0.25; our somewhat informative prior is a Normal distribution centered at 1.5 with standard deviation of 0.3. The Uniform prior is bounded between 1 and 2.
- Empirical studies have shown that the conduct of monetary policy in Europe (as described by the coefficients of an interest-rate feedback rule) is not much different than in the United States. For example, Gerlach and Schnabel (1999) estimate (in Europe) the coefficient on the output gap  $r_y$  in the policy rule equation to be 0.45 and the coefficient on inflation  $r_\pi$  to be 1.58, values that are statistically indistinguishable from those suggested by Taylor for the United States. However, Cochrane (2007) provided a theoretical argument for the lack of identification of the Taylor rule parameters in New Keynesian DSGE models; thus it is not surprising that posterior estimates for these parameters obtained in previous Bayesian studies have often coincided with their respective prior

distributions. Our somewhat informative priors for the five policy parameters  $\rho$ ,  $r_\pi$ ,  $r_{\Delta\pi}$ ,  $r_y$ , and  $r_{\Delta y}$  are based on coefficient estimates of Levin et al. (2006) and Smets and Wouters (2007) for the United States.

- For  $r_\pi$ , the informative prior (taken from SW) is  $N(1.7, 0.1^2)$ , our somewhat informative prior is  $N(2.0, 0.6^2)$ , and the Uniform prior ranges from 1 to 10, much wider than the prior range used by Onatski and Williams (2010).
  - For  $r_{\Delta\pi}$ , our informative prior is Gaussian with mean 0.3 and standard deviation 0.1; the somewhat informative prior is the same except that we doubled the standard deviation to 0.2; and the Uniform prior is bounded between  $-1$  and  $1$ .
  - Our informative prior for  $r_y$  is  $N(0.125, 0.05^2)$ , the somewhat informative prior is  $N(0.125, 0.1^2)$ , and the Uniform prior ranges from  $-1$  to  $1$ .
  - Our informative prior for  $r_{\Delta y}$  is Gaussian with mean 0.0625 and standard deviation 0.05, the somewhat informative prior is Gaussian with mean 0.3 and standard deviation 0.2, and the Uniform prior ranges from  $-1$  to  $1$ .
  - The informative prior for  $\rho$  is  $N(0.8, 0.1^2)$ , our somewhat informative prior is  $N(0.8, 0.2^2)$ , and the Uniform prior for  $\rho$  is bounded between 0 and 1.
- For the parameters that govern the persistence of the six autoregressive shocks in the model ( $\rho_a$ ,  $\rho_b$ ,  $\rho_G$ ,  $\rho_L$ ,  $\rho_I$ , and  $\rho_\pi$ ), there is not much information on which to base the priors. The one exception is the productivity shock ( $\rho_a$ ): there is ample evidence in the literature that productivity shocks are highly persistent in both Europe and the United States (see for example, Backus et al. (1992), Baxter and Crucini (1995), and Gruber (2002)). Following SW, our priors for these parameters have Beta distributions with mean 0.85 and standard deviations of 0.1. This prior spans the range of estimates obtained by Levin et al. (2006). Our somewhat informative prior has the same distribution and mean, but a standard deviation of 0.2. The Uniform prior for these parameters has a lower bound of 0.001 and an upper bound of 0.999.
  - Because there is little prior information on the standard deviations of the nine shocks in the model ( $\sigma_a$ ,  $\sigma_b$ ,  $\sigma_G$ ,  $\sigma_I$ ,  $\sigma_L$ ,  $\sigma_p$ ,  $\sigma_w$ ,  $\sigma_\pi$ , and  $\sigma_Q$ ), we allowed for a wide range of values in our somewhat-informative and Uniform prior specifications. The informative priors for these parameters (in the order listed above) are all inverse-Gamma distributions with means of 0.4, 0.2, 0.3, 0.1, 1, 0.15, 0.25, 0.02, and 0.4 (respectively), based on “previous estimation outcomes and trials with a very weak prior<sup>3</sup>”; following SW, we take a prior SD of 2 for all of these distributions. The somewhat informative priors for these 9 parameters are all Exponential distributions with mean 2, which implies a scale parameter of  $\beta = \frac{1}{2}$  and a standard deviation of 2. We chose to use the Exponential distribution for our somewhat informative prior to avoid placing a higher density around the means, because the available information about those means is not strong. The Uniform priors for these 9 shock parameters are all bounded between 0 and 20.

---

<sup>3</sup>There is some confusion as to how SW specify this prior in their paper: they say that their inverse-Gamma distribution has “a degree of freedom equal to 2.” The inverse-Gamma distribution is typically defined in terms of a shape parameter  $\alpha$  and a scale parameter  $\beta$ , with variance (in one of the two popular parameterizations)  $V(\theta) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$ . If their “degree of freedom” of 2 refers to the shape parameter, this would result in an improper prior with infinite variance, but the Dynare estimation code that SW shared with us suggests that they actually used a standard deviation of 2 for these priors.

### A3. Two maximization algorithms

We found that standard numerical gradient methods were unable to find the maximum when the log-likelihood function was nearly flat along several dimensions; other complications arise in fitting DSGE models because of the presence of many local modes, *cliffs* in the log-likelihood function at extreme parameter values, and regions in which the log-likelihood function is undefined because the model’s solution is indeterminate for certain parameter combinations. After experimenting with many different algorithms, we developed two new approaches of our own, which proved to be highly reliable in our experiments. Our first algorithm includes the following steps.

- (1) First, we choose an initial guess for the parameter vector from 1,000 function evaluations using random values drawn from uniform distributions with wide bounds for each parameter. Of these 1,000 random draws, we choose the parameter vector that generates the highest log-likelihood as the starting value for the algorithm.
- (2) The algorithm then loops through the following gradient-based and non-gradient-based optimization routines: simulated annealing (Belisle (1992)), the quasi-Newton “BFGS” method (Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970)), the Nelder and Mead (1965) simplex method, and the conjugate-gradient method of Fletcher and Reeves (1964). The optimized end value from one method (using a relative tolerance level of  $1.5 \cdot 10^{-8}$ ) is used as the starting value for the next method, and the entire loop is repeated until the improvement obtained by using a new method is less than 0.1 (on the log-likelihood scale). When this sequential maximization process is completed, we store the final parameter vector that resulted in the highest value of the function.
- (3) After storing this parameter vector, we start over and repeat steps (1) and (2) 20 times using 20 different initial guesses. The end result is a set of 20 parameter vectors with 20 corresponding “maximum” function values; we choose the best of these as the “apparent” maximum.

We used this *multiple-method maximization algorithm* in Section 3 of the paper, to obtain the 2,000 vectors of maximum-likelihood estimates that we compared with the actual data-generating parameter values.

In Section 4 of the paper we combined the multiple-method maximization algorithm with the adaptive MCMC algorithm described in the next Section of this Appendix, to perform an extremely thorough exploration of the log likelihood surface with our actual data set. The resulting *maximization/adaptive-MCMC/maximization* algorithm, the second of our two maximization methods, has the following steps:

- (4) We run the adaptive MCMC algorithm described in Section 4 below, using the “apparent” maximum from step (3) above as starting values; and
- (5) To conclude, we run step (2) again, using the means of the monitored draws from step (4) for each parameter as starting values.

Our multiple-method maximization algorithm (steps (1)–(3)), which is

- (a) not fast when single-threaded (it took about 5 hours of clock time, on a desktop computer with one 2.9GHz CPU, to obtain a single maximum of the log likelihood function with the data set studied in this paper) but
- (b) readily parallizable,

produced dramatic improvements in finding global, rather than local, maxima when combined with step (4): in an earlier version of the paper with a different data set, the maximum log likelihood at the end of step (3) was +421.26, and this was improved to +620.08 by step (4) and to +630.95 by step (5). The explanation for this remarkable difference was that we did not “cast our net wide enough” with the 20 initial guesses in step (3) — which succeeded in finding only a local mode of the log-likelihood surface — to find the probably-global maximum that was discovered with steps (4) and (5).

#### A4. Adaptive MCMC algorithm

We use an adaptive MCMC algorithm to obtain an efficient proposal distribution for simulating from both the likelihood function (Section 4 of the paper) and the posterior (Section 5 of the paper). Following Browne and Draper (2006), our adaptive MCMC method has three stages: adaptation, burn-in, and monitoring (we used a random-walk Metropolis sampler with a multivariate Gaussian proposal distribution).<sup>4</sup> The adaptation stage adjusts the covariance matrix of the proposal distribution every 2,500 iterations to be proportional to the posterior covariance matrix estimated from these iterations, with the scale factor adapted to achieve a target acceptance rate of 0.25 (Gelman et al. (1995)). The adaptation stage consists of 300,000 iterations, after which we fix the covariance matrix of the proposal distribution to that of the estimated covariance of the last 150,000. We then re-calibrate and fix the scale factor to achieve a target acceptance rate of 0.25. Following a burn-in period of 100,000 iterations (subsequent to adaptation), we then monitor the chain for 200,000 iterations; all inferences we make about the parameters come from this last set of 200,000 iterations from the monitoring phase.

## References

- Adolfson, M., S. Laseen, J. Linde, and M. Villani (2007). Bayesian estimation of an open economy DSGE model with incomplete pass-through. *Journal of International Economics* 72(2), 481–511.
- Adolfson, M. and J. Lindé (2011). Parameter identification in a estimated New Keynesian open economy model. Working Paper Series 251, Sveriges Riksbank (Central Bank of Sweden).
- Altig, D., L. J. Christiano, M. Eichenbaum, and J. Linde (2002, June). Technology shocks and aggregate fluctuations. Working papers, Northwestern University.
- Angeloni, I., L. Aucremanne, M. Ehrmann, J. Galí, A. Levin, and F. Smets (2004). Inflation persistence in the Euro Area: Preliminary summary of findings. Working paper, European Central Bank, National Bank of Belgium, CREI and Universitat Pompeu Fabra, Federal Reserve Board.
- Arrow, K. J. (1971). *Essays in the Theory of Risk Bearing.* , Chapter 3. Chicago: Markham Publishing Co.

---

<sup>4</sup>Because the efficiency of our MCMC algorithm is not our primary concern, we did not experiment with alternative MCMC algorithms. Readers interested in this topic may wish to refer to Chib and Ramamurthy (2010), who develop a Metropolis-Hastings algorithm that performs a random clustering of parameters at every iteration into an arbitrary number of blocks. Chib and Ramamurthy find that this algorithm is significantly more efficient than the conventional random-walk Metropolis-Hastings algorithm.

- Backus, D. K., P. J. Kehoe, and F. E. Kydland (1992). International real business cycles. *Journal of Political Economy* 100(4), 745–75.
- Baxter, M. and M. J. Crucini (1995, November). Business cycles and the asset structure of foreign trade. *International Economic Review* 36(4), 821–54.
- Belisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on  $r^d$ . *Journal of Applied Probability* 29(4), 885–895.
- Bernardo, J. M. and A. F. M. Smith (2000). *Bayesian Theory*. New York: Wiley.
- Bouakez, H., E. Cardia, and F. Ruge-Murcia (2005). The transmission of monetary policy in a multi-sector economy. Cahiers de recherche 2005-16, Universite de Montreal, Departement de sciences economiques.
- Browne, W. J. and D. Draper (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* 1(3), 473–550.
- Broyden, C. (1970). The convergence of a class of double-rank minimization algorithms. *IMA Journal of Applied Mathematics* 6(1), 76–90.
- Canova, F. and L. Sala (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* 56(4), 431–449.
- Chib, S. and S. Ramamurthy (2010, March). Tailored randomized block MCMC methods with application to DSGE models. *Journal of Econometrics* 155(1), 19–38.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113(1), 1–45.
- Cochrane, J. H. (2007). Determinacy and identification with Taylor rules. NBER Working Papers 13410, National Bureau of Economic Research, Inc.
- Dhyne, E., L. J. Alvarez, H. L. Bihan, G. Veronese, D. Dias, and J. Hof (2005, September). Price setting in the Euro area: some stylized facts from individual consumer price data. Working Paper Series 524, European Central Bank.
- Evers, M., R. A. de Mooij, and D. J. van Vuuren (2006, February). What explains the variation in estimates of labour supply elasticities? Tinbergen Institute Discussion Papers 06-017/3, Tinbergen Institute.
- Fagan, G., J. Henry, and R. Mestre (2001). An area-wide model (AWM) for the Euro Area. Working Paper Series 42, European Central Bank.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal* 13(3), 317–322.
- Fletcher, R. and C. Reeves (1964). Function minimization by conjugate gradients. *Computer Journal* 7(2), 149–154.
- Francis, N. and V. A. Ramey (2005). Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. *Journal of Monetary Economics* 52(8), 1379–1399.
- Fuhrer, J. C. (2000, June). Habit formation in consumption and its implications for monetary-policy models. *American Economic Review* 90(3), 367–390.
- Gali, J. (1999). Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations? *American Economic Review* 89(1), 249–271.
- Gali, J., M. Gertler, and J. D. Lopez-Salido (2001). European inflation dynamics. *European Economic Review* 45(7), 1237–1270.

- Gelman, A., G. Roberts, and W. Gilks (1995). Efficient Metropolis jumping rules. In J.M.Bernardo, J.O.Berger, A.P.Dawid, and A.F.M.Smith (Eds.), *Bayesian Statistics 5*. Oxford: Oxford University Press.
- Gerlach, S. and G. Schnabel (1999, October). The Taylor rule and interest rates in the EMU area. CEPR Discussion Papers 2271, C.E.P.R. Discussion Papers.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation* 24(109), 23–26.
- Gruber, J. W. (2002). Productivity shocks, habits, and the current account. International Finance Discussion Papers 733, Board of Governors of the Federal Reserve System (U.S.).
- Guerron-Quintana, P., A. Inoue, and L. Kilian (2013). Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics* 4(2), 197–229.
- Ham, J. C. and K. T. Reilly (2006, July). Using micro data to estimate the intertemporal substitution elasticity for labor supply in an implicit contract model. IEPR Working Papers 06.54, Institute of Economic Policy Research (IEPR).
- Hamilton, J. D. (1994). *Time Series Analysis*, Chapter 13. New Jersey: Princeton University Press.
- Ireland, P. N. (2003). Endogenous money or sticky prices? *Journal of Monetary Economics* 50(8), 1623–1648.
- Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics* 57(2), 189–202.
- Iskrev, N. and M. Ratto (2010). Analysing identification issues in DSGE models. Working paper, Bank of Portugal and European Commission.
- Kaplow, L. (2005). The value of a statistical life and the coefficient of relative risk aversion. *Journal of Risk and Uncertainty* 31(1), 23–34.
- King, R. G. and S. T. Rebelo (2000). Resuscitating real business cycles. NBER Working Papers 7534, National Bureau of Economic Research, Inc.
- Komunjer, I. and S. Ng (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica* 79(6), 1995–2032.
- Kydland, F. E. and E. C. Prescott (1982). Time to build and aggregate fluctuations. *Econometrica* 50(6), 1345–70.
- Levin, A. T., A. Onatski, J. Williams, and N. M. Williams (2006, April). *Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models*, pp. 229–312. MIT Press.
- Levin, A. T., A. Onatski, J. C. Williams, and N. Williams (2005, August). Monetary policy under uncertainty in micro-founded macroeconometric models. NBER Working Papers 11523, National Bureau of Economic Research, Inc.
- Negro, M. D., F. Schorfheide, F. Smets, and R. Wouters (2005). On the fit and forecasting performance of New Keynesian models. Working Paper Series 491, European Central Bank.
- Nelder, J. and R. Mead (1965). A simplex algorithm for function minimization. *Computer Journal* 7(4), 308–313.
- Onatski, A. and N. Williams (2010). Empirical and policy performance of a forward-looking monetary model. *Journal of Applied Econometrics* 25(1), 145–176.
- Rotemberg, J. and M. Woodford (1997). An optimization-based econometric framework for the evaluation of monetary policy. In *NBER Macroeconomics Annual 1997, Volume 12*, NBER Chapters, pp. 297–361. National Bureau of Economic Research, Inc.

- Schorfheide, F. (2011). Estimation and evaluation of DSGE models: progress and challenges. NBER Working Papers 16781, National Bureau of Economic Research, Inc.
- Shanno, D. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation* 24(111), 647–656.
- Sims, C. A. (2002, October). Solving linear rational expectations models. *Computational Economics* 20(1-2), 1–20.
- Smets, F. and R. Wouters (2003). An estimated dynamic stochastic general equilibrium model of the Euro area. *Journal of the European Economic Association* 1(5), 1123–1175.
- Smets, F. and R. Wouters (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* 97(3), 586–606.
- West, M. and J. Harrison (1999). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, Inc.