

Mickelsson, Glenn

Working Paper

Estimation of DSGE models: Maximum likelihood vs. Bayesian methods

Working Paper, No. 2015:6

Provided in Cooperation with:

Department of Economics, Uppsala University

Suggested Citation: Mickelsson, Glenn (2015) : Estimation of DSGE models: Maximum likelihood vs. Bayesian methods, Working Paper, No. 2015:6, Uppsala University, Department of Economics, Uppsala,
<https://nbn-resolving.de/urn:nbn:se:uu:diva-270200>

This Version is available at:

<https://hdl.handle.net/10419/168391>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



UPPSALA
UNIVERSITET

Department of Economics

Working Paper 2015:6

Estimation of DSGE models: Maximum Likelihood vs. Bayesian methods

Glenn Mickelsson

Department of Economics
Uppsala University
P.O. Box 513
SE-751 20 Uppsala
Sweden
Fax: +46 18 471 14 78

Working paper 2015:6
December 2015
ISSN 1653-6975

ESTIMATION OF DSGE MODELS: MAXIMUM LIKELIHOOD VS. BAYESIAN METHODS

GLENN MICKELSSON

Estimation of DSGE models: Maximum Likelihood vs. Bayesian methods*

Glenn Mickelsson[†]

Department of Economics, Uppsala University

December 22, 2015

Abstract

DSGE models are typically estimated using Bayesian methods, but a researcher may want to estimate a DSGE model with full information maximum likelihood (FIML) so as to avoid the use of prior distributions. A very robust algorithm is needed to find the global maximum within the relevant parameter space. I suggest such an algorithm and show that it is possible to estimate the model of Smets and Wouters (2007) using FIML. Inference is carried out using stochastic bootstrapping techniques. Several FIML estimates turn out to be significantly different from the Bayesian estimates and the reasons behind those differences are analyzed.

Keywords: Bayesian methods, Maximum likelihood, Business Cycles, Estimate DSGE models

JEL classification: C11, E32, E37, E32

*Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

[†]Department of Economics, Uppsala University, Box 513, SE-751 20 Uppsala, Sweden,
glenn.mickelsson@nek.uu.se

1 Introduction

Since the seminal papers by Smets and Wouters (2003) and Christiano et al. (2005) it has been popular to use Bayesian methods when estimating dynamic stochastic general equilibrium (DSGE) models. Typically, Bayesian methods are implemented by postulating prior distributions of the parameters and estimating the model with maximum likelihood¹.

Still, there may be situations when estimation of a DSGE model using full information maximum likelihood (FIML) is desirable. Prior information may be lacking and the author may want to identify flaws in the model by finding the parameters that get unreasonable estimates when the data speaks freely.

When trying to estimate the model with FIML one encounters the problem that the likelihood function may be flat and have many local maximums. Local search algorithms are fast but they will almost certainly get stuck in one of the local maximums and thereby make the researcher draw the wrong conclusions. Current global search algorithms are usually time consuming and require a lot of input information beside the minimum and maximum values of the parameters. Examples of these additional inputs are population size and cross over ratio for the genetic algorithm, temperature level and cooling down pace for the simulated annealing algorithm and so forth. Usually, there are no clear guidelines for how to chose these inputs even though they can be crucial for the outcome. This makes current global search algorithms inconvenient when estimating big systems such as a medium scale DSGE model.

I propose a global solution algorithm that requires no other input than maximum and minimum values of the parameters. The algorithm is compared with other commonly used global algorithms. The comparison is made for both tricky mathematical functions and estimation on artificial data. On average, my algorithm outperforms the other algorithms in both comparisons.

After the algorithm comparison I use the data from Smets and Wouters (2007) (SW) and estimate their model using FIML. For presentation of the original model I refer to their paper. I find that there are quite many of the parameter estimates in SW that are outside

¹The likelihood function is usually constructed by the Kalman (1960) filter.

the 99 % confidence interval that I find when I use FIML. These include the investment adjustment cost ϕ , constant relative risk aversion σ_c , habit formation in consumption h , wage stickiness ξ_w , capital utilization ψ , etc. This demonstrates that prior distributions have substantial effects on the posterior estimates even though there is enough information in the data to generate reasonable parameter estimates.

In an earlier study by Ireland (2004) a DSGE model was estimated with FIML as well but his model was much smaller and only six structural parameters was estimated compared to this study which estimates 36 structural parameters. Andreasen (2009) evaluated the success rate when estimating larger DSGE models and also compared different algorithms. His method assumes that the shocks are known and for that reason his method cannot be used on real data. However, the most successful algorithm in that study is part of the algorithm comparison in this paper.

In section 2, my proposed algorithm is explained and in Section 3 it is compared to other global algorithms. In Section 4 the fixed parameters and the parameter space for the DSGE estimation are presented. The results are presented in Section 5 together with diagnostic tests and simulated distributions of the confidence intervals. Lastly in Section 6, there is a discussion about the differences between the FIML estimates and the Bayesian estimates in SW. I focus on the parameter estimates in SW which are outside of the 99 % confidence interval of the FIML estimation and try to understand the reasons behind the differences.

2 A fast and robust global algorithm

In many applications, we have little prior information about parameter values but we can delimit some range of parameter values that make economic sense. Parameter values outside this range are not part of an economically meaningful interpretation of the data. The idea of the proposed algorithm is that it should be fast and robust, not rely on derivatives and not require any other input than maximum and minimum for the parameter values.

Specifically, the algorithm should not depend on starting values that are chosen by the researcher. The suggested algorithm makes its own choice of starting values that are spread out in the pre-specified, finite parameter space. In this way it differs from simulated an-

nealing, the genetic algorithm and the evolutionary algorithm since they all require the user to give inputs such as temperature, the population size, number of generations, a starting vector and so forth.

It should be emphasized that estimation by FIML with bounds on the parameters is not the same thing as Bayesian estimation with uniform priors. Any prior will push the estimate towards the mean of the prior distribution which in the uniform case is the mean of the interval. With FIML estimation, the exact values of the boundaries matter only if the parameter ends up on the boundary of the interval. As we will see, this happens in relatively few cases in the economic application that is presented below.

This section describes how the algorithm works and the procedure is illustrated with a two parameter example. A general and mathematical description is available in the Appendix. The two parameters are denoted $p_1, p_2 \in [-10, 10]$, thus $min = -10$ and $max = 10$ for both p_1 and p_2 . The number of parameters is denoted P and is equal to two in this example.

First, $ceil(1.1 \times P) = 3$ vectors are created². The three vectors are equally spread out over $(min + (max - min) * 0.15, max - (max - min) * 0.15)$. The resulting vectors are shown in matrix A:

$$A = \begin{bmatrix} -7 & 0 & 7 \\ -7 & 0 & 7 \end{bmatrix}. \quad (1)$$

Then two more vectors are added that are closer to the minimum and maximum values. $(min + (max - min) * 0.05, max - (max - min) * 0.05)$. Those are shown in matrix B:

$$B = \begin{bmatrix} -9 & 9 \\ -9 & 9 \end{bmatrix}. \quad (2)$$

All of these vectors are summarized in matrix C for convenience:

$$C = \begin{bmatrix} -9 & -7 & 0 & 7 & 9 \\ -9 & -7 & 0 & 7 & 9 \end{bmatrix}. \quad (3)$$

The dots in Figure 1 show the vectors in matrix C.

²*ceil* means round up to the nearest integer

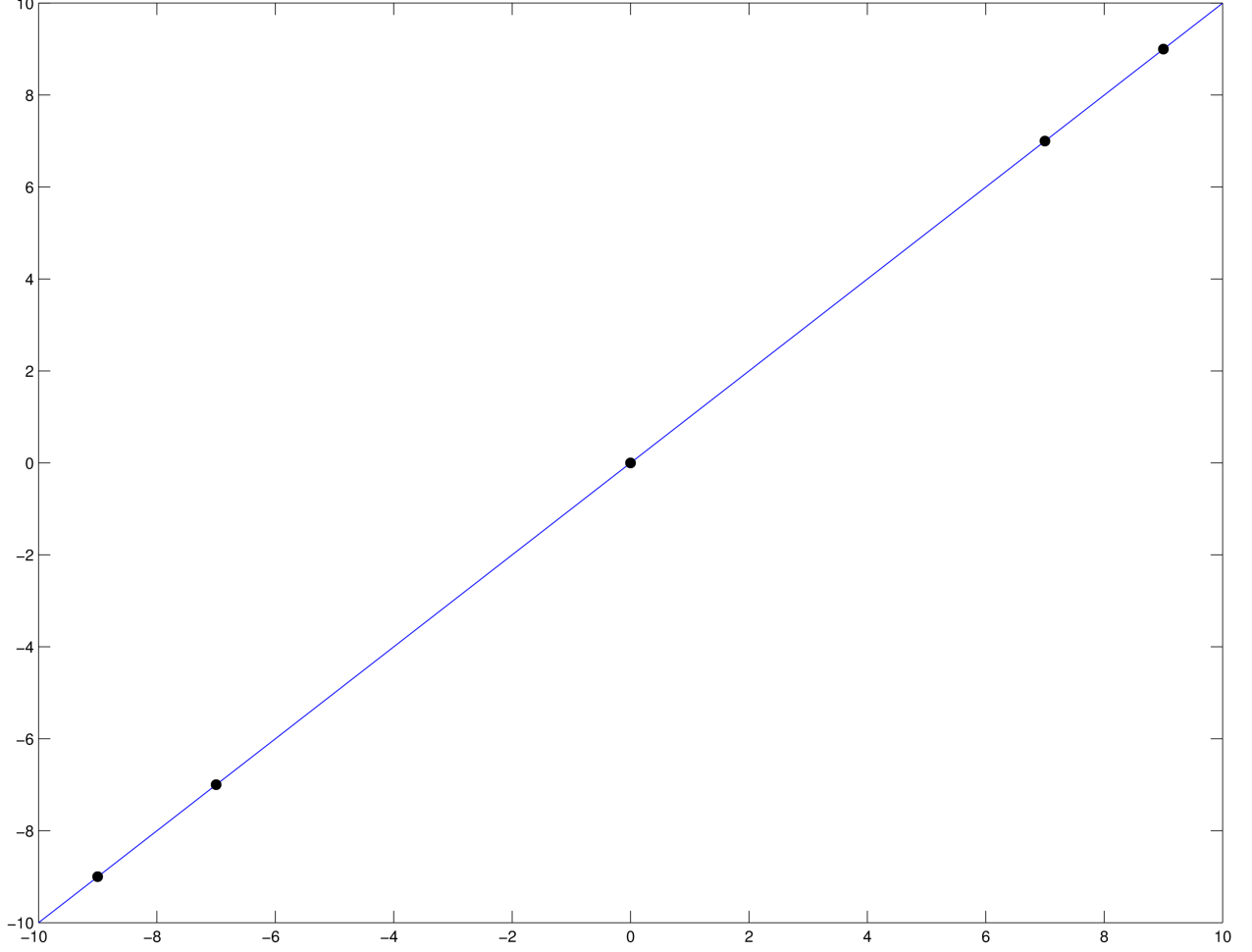


Figure 1: $\text{ceil}(1.1 \times P) + 2 = 5$ vectors that cover the parameter space (Matrix C)

2.1 Step 1 - Repeated three times

Around each vector in Figure 1 another vector is stochastically chosen within the circle radius $(-(\max - \min) * 0.1, (\max - \min) * 0.1)$ from the vectors closest to origo and $(-(\max - \min) * 0.05, (\max - \min) * 0.05)$ for the remaining vectors. This is demonstrated with vectors 1 to 5 in Figure 2 where the circles represents the maximum and minimum just mentioned. The reason for the stochastic part is that it is impossible to find a pattern which works for all kinds of problems. A stochastic part is therefore needed in order make up for the huge variation in target functions and their shapes.

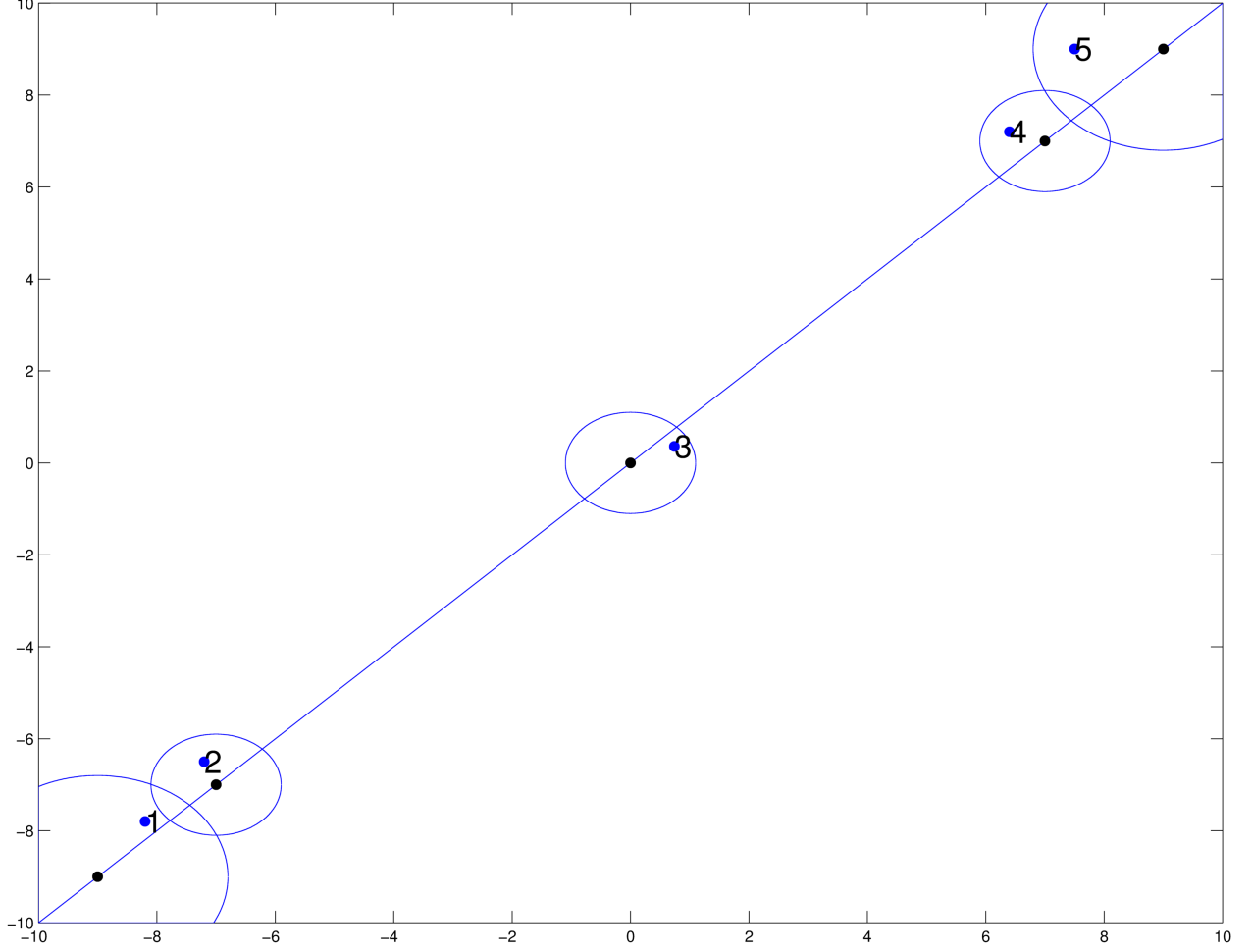


Figure 2: Vectors 1 to 5 are stochastically chosen within their respective stochastic space and those are the ones that will be part of the estimation

Next, around vectors 1 to 5, $2 \times P = 4$ vectors are created where each vector element is replaced by its minimum and maximum parameter space value one at the time (see Figure 3). To demonstrate this we start by doing it for vector 1, $(-8.2, 7.8)$. In matrix D we see vector 1 together with the four additional vectors.

$$D = \begin{bmatrix} -8.2 & -10 & 10 & -8.2 & -8.2 \\ -7.8 & -7.8 & -7.8 & -10 & 10 \end{bmatrix} \quad (4)$$

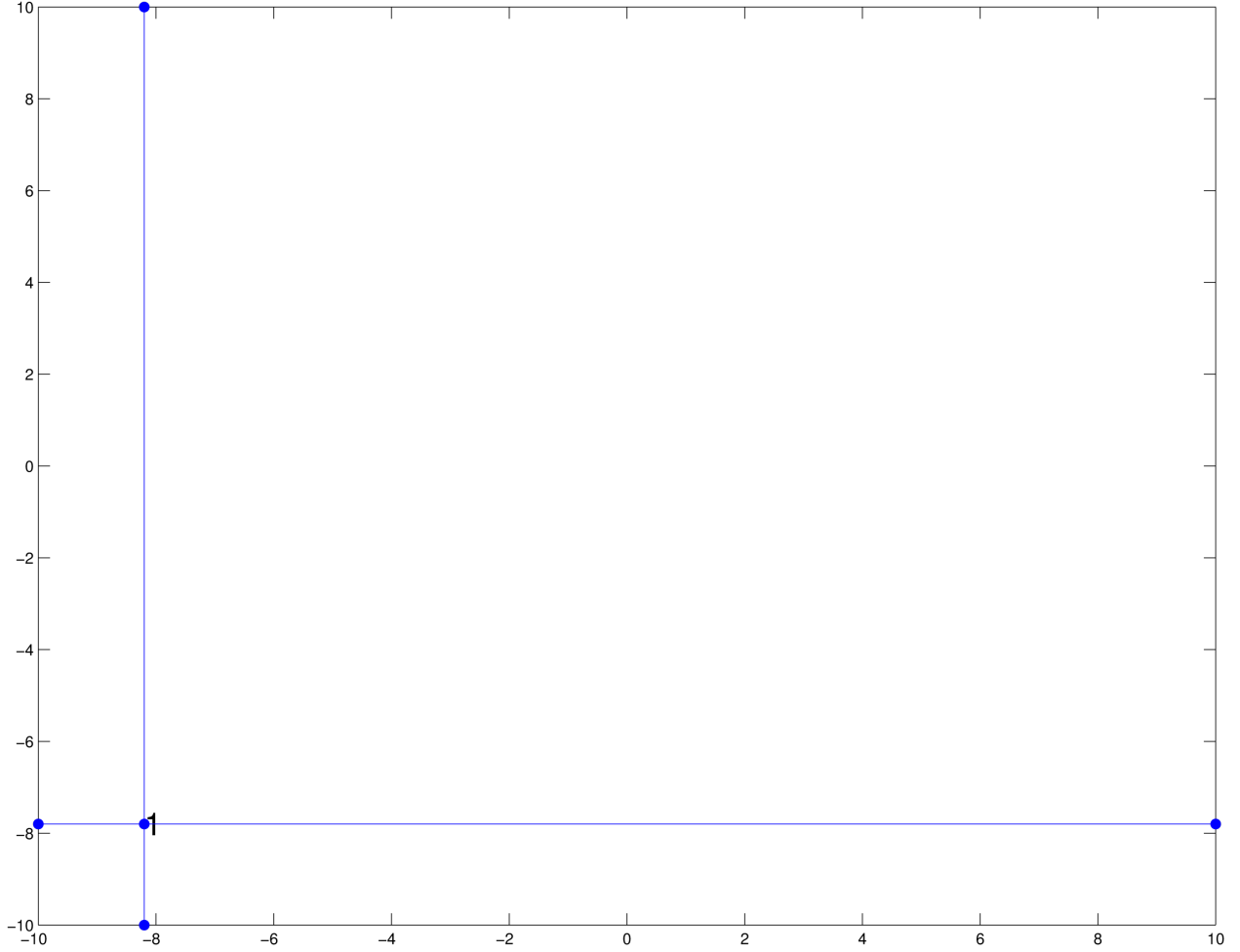


Figure 3: $2 \times P$ vectors created around vector 1 where vector 1's elements are replaced with the maximum and minimum value of the parameter space $[-10, 10]$ one at the time.

Vectors 1 to 5 will be run through the simplex Nelder and Mead (1965) algorithm. Roughly speaking, the idea of the Nelder and Mead (1965) method is to move the vectors according to a schedule that is supposed to approach the optimum (minimum). After every $5 \times P$ th turn four new vectors will be created in the same manner as in Figure 3 only this time it will be around the best vector (vector with the lowest value) thus far. This entire procedure continues until the absolute logarithmic change of the best value is smaller than one; see equation (5). When the algorithm has stopped it saves that vector and continues with vector 2 in Figure 2.

$$|\ln |X_{j+1}^{best}| - \ln |X_j^{best}|| < 1 \quad (5)$$

When this is done there are five vectors which are supposed to be one step closer to the optimum. All of these five vectors are run through the Nelder and Mead (1965) method until the absolute logarithmic change of the best value is smaller than 10^{-3} . The resulting vector is saved and step 1 is repeated two more times. This will eventually give us vectors, V_1 , V_2 and V_3 .

2.2 Step 2

When Step 1 is finished there are three vectors in total, V_1 , V_2 and V_3 . These three vectors are run through the Nelder and Mead (1965) method until the absolute logarithmic change of the best value is smaller than 10^{-9} . This vector is then run through the same procedure as in Figure 3 only this time the process is restarted every $360 \times P$ th turn. The vector with the lowest value when the improvement is lower than 10^{-9} is the final one!

3 Algorithm comparisons

This section compares the algorithm of this paper with some other algorithms that are popular in the global optimization literature.

3.1 Algorithms

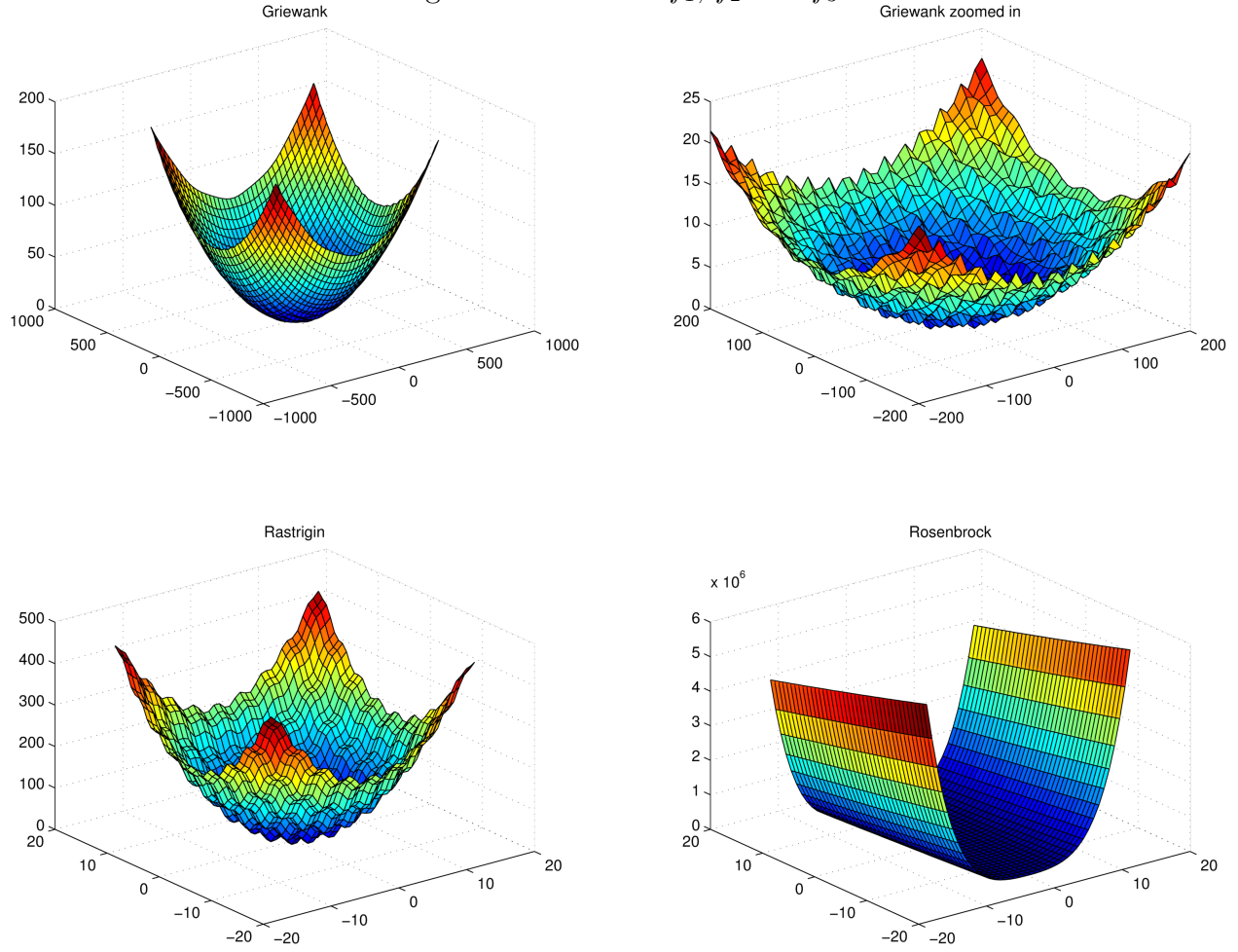
Two alternative algorithms are popular global optimization algorithms in Matlab. Those are the genetic algorithm (GA) by Holland (1975) which mimics the evolution of organisms akin to Charles Darwin's "natural selection" principle. The other one is the simulated annealing algorithm (SA) by Kirkpatrick (1984) which was used in an econometric study by Goffe et al. (1994). SA mimics the process of hardening metal where the degree is high in the beginning and gets lower and lower until the process is stopped. Additionally there is another branch of global algorithms which build on the idea of a swarm, such as an ant colony or a bee

swarm. This branch is called particle swarm optimization (PSO). Among those there is one interesting algorithm, namely the artificial bee colony (ABC) algorithm by Karaboga and Basturk (2007) which will also be used for comparison. The evolutionary algorithm (CMA-ES) in Andreassen (2009) is also part of the the comparison.

3.2 Minimizing tricky functions

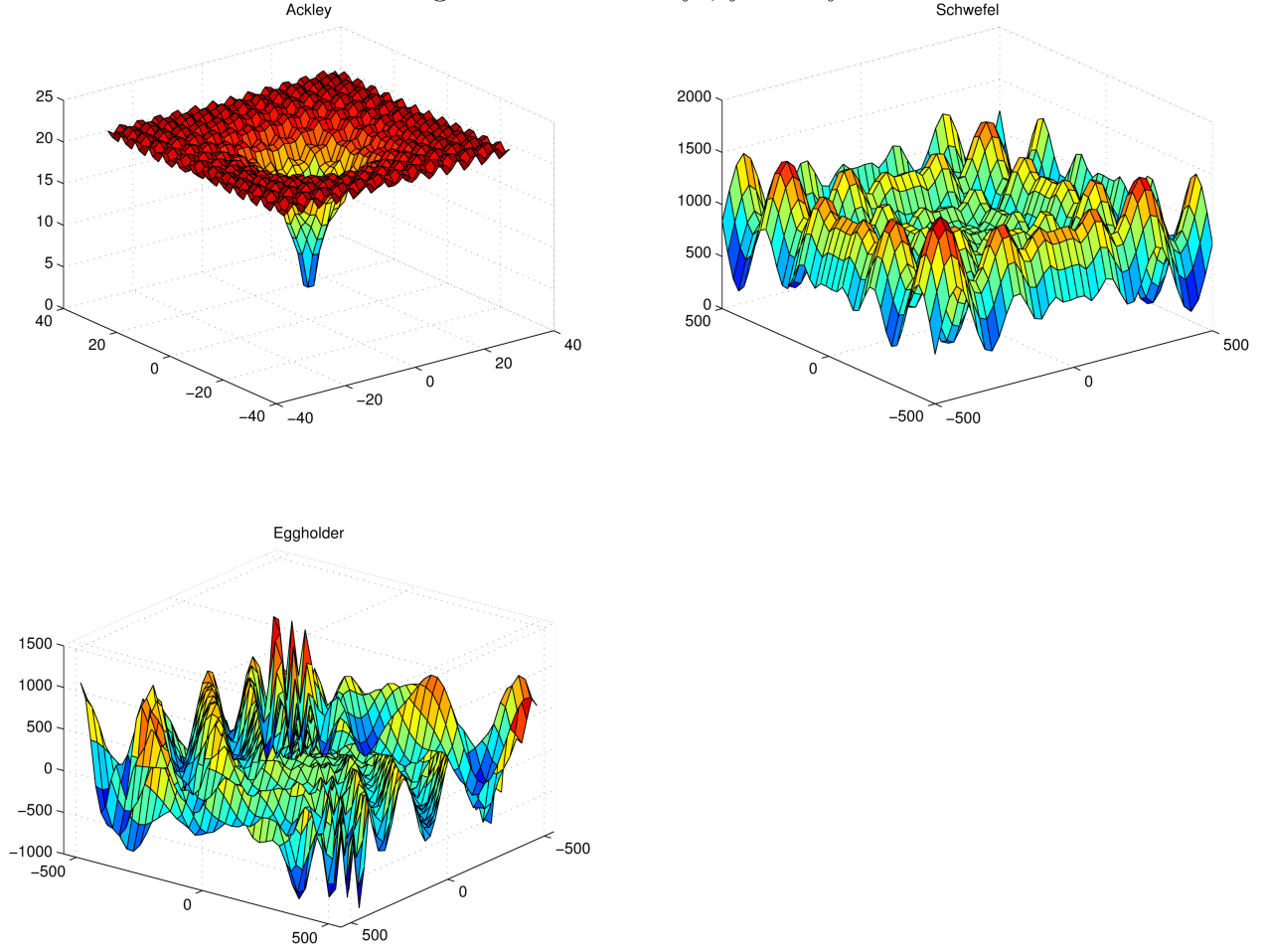
To test the algorithms we need some complicated functions where the true minimum is known. Below, some functional forms, their optimum values and their bounds are presented. Additionally, a two dimensional plot is shown in order to get a visual idea of the difficulty of the optimization problem. The first five benchmark functions of the comparison are the same as those used by Karaboga and Basturk (2007). The number of parameters in this comparison are the same as in SWs model, $P = 36$, where P is the total number of parameters to be estimated.

Figure 4: Functions f_1 , f_2 and f_3



Note: f_1 is the Griewank function, f_2 is the Rastrigin function and f_3 is the Rosenbrock function

Figure 5: Functions f_4 , f_5 and f_6



Note: f_4 is the Ackley function, f_5 is the Schwefel function and f_6 is the Eggholder function

The first function is the Griewank function:

$$f_1(\vec{p}) = \left(\frac{1}{4000} \sum_{i=1}^P p_i^2 \right) - \prod_{i=1}^P \cos \left(\frac{p_i}{\sqrt{i}} \right) + 1 \quad (6)$$

$$f_1(\vec{0}) = 0 \quad p \in [-600, 600]$$

At distance the Griewank function looks like a Sphere function but when zooming in it becomes evident that it consists of many small local optimums.

The next function is the Rastrigin function where there instead of small local minimums are bigger local minimums which are visible already at first sight:

$$\begin{aligned}
f_2(\vec{p}) &= \sum_{i=1}^P (p_i^2 - 10 \cos(2\pi p_i) + 10) \\
f_2(\vec{0}) &= 0 \quad p \in [-15, 15]
\end{aligned} \tag{7}$$

Next we have the Rosenbrock function which looks simpler than the former ones but it has a big flat area around its global minimum which makes it harder for the algorithms to converge:

$$\begin{aligned}
f_3(\vec{x}) &= \sum_{i=1}^P 100(p_i^2 - p_{i+1})^2 + (1 - p_i)^2 \\
f_3(\vec{1}) &= 0 \quad p \in [-15, 15]
\end{aligned} \tag{8}$$

Ackley is the 4th function:

$$\begin{aligned}
f_4(\vec{p}) &= 20 + e - 20e^{\left(-0.2\sqrt{\frac{1}{P} \sum_{i=1}^P p_i^2}\right) - e^{\frac{1}{P} \sum_{i=1}^P \cos(2\pi p_i)}} \\
f_4(\vec{0}) &= 0 \quad p \in [-32.768, 32.768]
\end{aligned} \tag{9}$$

Schwefel is the 5th function and now we have a much harder function with no clear bowl-shaped pattern where it is even hard to see the minimum in a two dimensional plot:

$$\begin{aligned}
f_5(\vec{p}) &= P \times 420.9687 + \sum_{i=1}^P -p_i \sin |p_i| \\
f_5(420.9687, 420.9687 \dots P) &= 0 \quad p \in [-500, 500]
\end{aligned} \tag{10}$$

The 6th function, the eggholder function, is also hard to optimize since it has many local minimums and no clear bowl-shaped pattern. The basic eggholder function is two-dimensional but since all the tests will be based on a problem with the same dimensions as when we estimate a medium sized DSGE model by maximum likelihood estimation we must have 36 dimensions. To accomplish this the two dimensional eggholder is multiplied by itself until the number of dimensions is equal to 36. Its functional form is displayed in equation (11)

$$\begin{aligned}
g(p_i, p_{i+1}) &= -(p_{i+1} + 47) \sin \left(\sqrt{|p_{i+1} + \frac{p_i}{2} + 47|} \right) - p_i \sin \left(\sqrt{|p_i - (p_{i+1} + 47)|} \right) \\
f_6(\vec{p}) &= \prod_{i=1}^{P-1} (1 + 959.64066 + g(p_i, p_{i+1})) \\
f_6(512, 404.2319 \dots P) &= 1 \quad p \in [-512, 512]
\end{aligned} \tag{11}$$

Last but not least. In order to construct a very challenging problem for the algorithms, a multiplicative combination of the Eggholder and the Griewank function is created according to equation (12):

$$f_7(\vec{p}) = f_1(\vec{p})f_6(\vec{p}) \quad (12)$$

$$f_7(512, 404.2319, 0, 0 \dots P) = 1 \quad p \in [-512, 512, 600, -600]$$

3.3 Results

The comparison was carried out with 36 parameters for each function. Each function in turn was estimated ten times and the average value is presented in Table 1. The reason for repeating the estimation ten times per algorithm is because they all have a stochastic component. The computing time in Table 2 is the total computing time for the ten repetitions.

Table 1: Function values

Function	SA	GA	Bee swarm	CMA-ES	New algorithm	Solution
$f_1(\vec{x})$	9.053	$3.795e - 05$	0.0007411	$8.951e - 09$	0.003638	0
$f_2(\vec{x})$	840.9	0.199	$2.064e - 12$	46.46	11.14	0
$f_3(\vec{x})$	10.74	0.03602	0.2788	$9.644e - 09$	33.54	0
$f_4(\vec{x})$	7.445	0.0001339	$4.574e - 14$	$1.893e - 06$	0.002241	0
$f_5(\vec{x})$	53.41	1888	13.1	6931	$-9.767e - 09$	0
$f_6(\vec{x})$	$8.413e + 24$	$7.93e + 41$	$3.576e + 34$	$1.606e + 48$	13.98	1
$f_7(\vec{x})$	$3.985e + 32$	$5.011e + 24$	$2.626e + 15$	$7.86e + 34$	8.839	1

Note: The lower the better. SA=Simulated Annealing, GA=Genetic Algorithm, ABC=Artificial Bee Colony, CMA-ES=Evolutionary algorithm.

Table 2: Total computing time in minutes

Function	SA	GA	Bee swarm	CMA-ES	New algorithm
$f_1(\vec{x})$	4.11	1.37	0.68	0.28	2.49
$f_2(\vec{x})$	3.31	2.14	1.20	0.21	2.26
$f_3(\vec{x})$	82.15	26.64	1.81	1.44	2.50
$f_4(\vec{x})$	2.29	2.78	0.70	0.15	1.63
$f_5(\vec{x})$	3.19	1.99	1.12	0.40	2.50
$f_6(\vec{x})$	118.56	84.29	5.66	6.40	19.74
$f_7(\vec{x})$	124.36	39.96	7.92	5.00	28.26

Note: The lower value the better. SA=Simulated Annealing, GA=Genetic Algorithm, ABC=Artificial Bee Colony, CMA-ES=Evolutionary algorithm in Andreassen (2009).

All algorithms do a decent job for the simpler functions (functions 1 to 5) except for SA which gets 840.9 for the *Rastrigin* function. For the more complicated *Eggholder* and *Eggholder – Griewink* functions only the new algorithm manages to get a reasonable values (below 14) while the others get values above 2×10^{15} .

3.4 Estimation of the DSGE model

A more relevant comparison is to see how the algorithms perform when maximizing the likelihood function for the DSGE model in Smets and Wouters (2007). In order to make the comparison generate simulated samples as if the model was correctly specified. This implies that we know the true vector of parameters which makes it interesting to look at the squared deviation measure in equation (13)

$$100 \sqrt{\frac{1}{N} \sum_{i=0}^N \left(\frac{x_{i,est} - x_{i,true}}{x_{i,max} - x_{i,min}} \right)^2} \quad (13)$$

For each of the samples I let the algorithms search for the highest likelihood value. The likelihood value of each algorithm and sample is shown in Table 3, the squared deviation of the estimated parameter vector and the true parameter is shown in Table 4

Table 3: Artificial samples: maximum likelihood value

Sample	SA	GA	Bee swarm	CMA-ES	New algorithm
S1	-1558	-804.6	-819.2	-1e+40	-804
S2	-1790	-931	-819.3	-874.4	-860.1
S3	-2350	-765.9	-766.8	-907.9	-756.4
S4	-1457	-886.6	-827.7	-814.1	-783.8
S5	-1857	-792.9	-771.8	-1e+40	-764.2
S6	-1953	-844.8	-830.7	-1.001e+05	-803.8
S7	-1667	-835	-840.5	-959.6	-923.4
S8	-2447	-794.4	-808.4	-778.6	-777
S9	-1958	-807.1	-815.2	-816.8	-798.5
S10	-2237	-806.5	-786.4	-2.678e+04	-777
S11	-2638	-889.1	-796.7	-888	-769.2
S12	-1713	-842.8	-847.1	-806.9	-795.4
S13	-1693	-837.2	-792.5	-856.1	-835.2
S14	-1907	-939.5	-827.7	-1e+40	-793.5
S15	-2164	-806.1	-797.5	-876.2	-828.2
S16	-1810	-776.2	-788.4	-919.4	-773.4
S17	-2585	-785.2	-796.6	-824.7	-814.1
S18	-2074	-771.6	-800.8	-792.4	-796.2
S19	-1391	-803.5	-796.7	-1e+40	-786.6
S20	-1267	-791.6	-796	-1.315e+04	-789.1
Wins	0	3	3	0	14

Note: The higher the better. SA=Simulated Annealing, GA=Genetic Algorithm, ABC=Artificial Bee Colony, CMA-ES=Evolutionary algorithm in Andreasen (2009).

The new algorithm outperforms the others since it produces the highest likelihood value in 70% of the samples and gets closest to the true parameter vector in 50% of the samples. The algorithm that comes second (The Genetic algorithm) has an average computing time of about 2 hours per sample compared to the new algorithm which has an average computing

time of 40 minutes.

Table 4: Artificial samples: Sum of squared deviations

Sample	SA	GA	Bee swarm	CMA-ES	New algoritm
S1	34.4	9.808	22.47	54.53	7.674
S2	33.5	23.52	8.853	30.14	22.72
S3	43.43	11.77	14.31	29.47	7.652
S4	28.94	17.61	16.11	23.42	5.286
S5	43.09	16.33	17.68	49.7	18.76
S6	38.88	17.54	15.43	47.56	6.162
S7	35.65	17.17	19.73	24.08	29.89
S8	43.73	13.81	26.26	3.745	4.308
S9	36.73	10.39	20.78	26.11	5.41
S10	43.13	22.84	9.642	40.79	19.06
S11	48.83	24.3	14.82	34.57	5.539
S12	39.05	21.04	18.04	19.06	7.968
S13	39.41	17.1	8.036	30.84	22.69
S14	35.74	28.84	25.65	42.69	6.191
S15	42.88	16.7	6.486	34.8	15.84
S16	40.37	7.542	20.03	35.92	6.894
S17	42.78	8.149	19.4	19	13.88
S18	37.38	5.503	17	26.27	25.79
S19	32.59	12.02	14.93	53.44	9.517
S20	20.32	11.11	22	44.23	20.37
Wins	0	5	4	1	10

Note: The lower the better. SA=Simulated Annealing, GA=Genetic Algorithm, ABC=Artificial Bee Colony, CMA-ES=Evolutionary algorithm.

4 Estimation of the Smets and Wouters (2007) model on real data

This section presents the parameters that are calibrated, followed by the intervals for the estimated parameters. Lastly, there is a short description of the data.

4.1 Fixed parameters

Following SW there are five parameters that are kept fixed. Capital depreciation is set to $\delta = 0.025$ and the government spending to GDP ratio is fixed at $g_y = 0.18$. The steady state mark up for wage setters λ_w is not identified and set to 1.5. Lastly the Kimball (1995) aggregators in the goods and labor market, ϵ_p and ϵ_w , are both set to 10. The fixed parameters are summarized in Table 5

Table 5: Parameters of the model that re not estimated		
Description	Parameter	Value
Depreciation	δ	0.025
Government spending-GDP ratio	g_y	0.18
Labor markup	λ_w	1.5
Kimball aggregator goods market	ϵ_p	10
Kimball aggregator labor market	ϵ_w	10

Note: Parameters of the model that re not estimated

4.2 Intervals for estimated parameters

As discussed above, the algorithm, needs a prespecified finite parameter space so we need to determine bounds for the parameters based on economic reasoning. These bounds are presented in Table 6.

Habit persistence h , the calvo probabilities ξ_w and ξ_p , wage and price indexation ι_w and ι_p , capital utilization ψ , interest rate smoothing in the Taylor rule ρ and the capital share in production α are all theoretically bounded between 0 and 1. For the persistances in the

autoregressive shock processes $\rho_a, \rho_b, \rho_g, \rho_I, \rho_r, \rho_p, \rho_w$ and ρ_{ga} they have to be stationary and therefore belong to the interval $[-0.999, 0.999]$. Steady state hours worked \bar{l} is defined as the logarithmic deviation from the logarithmic average of hours worked. It is therefore normalized to zero and is allowed to deviate 20 percent above and below its average.

Regarding the remaining parameters such as the standard deviation of the shock processes, σ_a to σ_w , investment adjustment cost ϕ , the relative risk aversion σ_c etc. are all positive but there are no obvious upper bound. The strategy chosen here is to specify sufficiently broad intervals and then check how quickly the likelihood function worsens as the parameter in question approaches the borders of the interval. As will be evident from Figures 6 to 8 below, the parameter intervals in Table 6 and 7 are more than enough.

Table 6: Parameter intervals

Description	Parameter	Interval
Investment adjustment cost	ϕ	$[0.01, 100.00]$
Constant relative risk aversion	σ_c	$[0.25, 20.00]$
Habit persistence	h	$[0.00, 1.00]$
Calvo parameter labor market	ξ_w	$[0.00, 1.00]$
Inverse Frisch elasticity	σ_l	$[0.01, 20.00]$
Calvo parameter goods market	ξ_p	$[0.00, 1.00]$
Indexation labor market	ι_w	$[0.00, 1.00]$
Indexation goods market	ι_p	$[0.00, 1.00]$
Capital utilization	ψ	$[0.00, 1.00]$
Fixed production cost	Φ	$[0.00, 30.00]$
T.R reaction to inflation	r_π	$[1.01, 100.00]$
T.R interest rate smoothing	ρ	$[0.00, 1.00]$
T.R reaction to output	r_y	$[0.00, 30.00]$
T.R reaction to Δ_y	$r_{\Delta y}$	$[0.00, 2.00]$
Steady state inflation rate	$\bar{\pi}$	$[-30.00, 30.00]$

Table 7: Parameter intervals continue

Description	Parameter	Interval
Steady state inflation rate	$\bar{\pi}$	$[-30.00, 30.00]$
Sub. discount rate times 100	$100(\beta^{-1} - 1)$	$[-8.00, 4.00]$
Steady state hours worked	\bar{l}	$[-30.00, 30.00]$
Steady state growth rate	$\bar{\gamma}$	$[0.00, 2.00]$
Capital share in production	α	$[0.01, 0.99]$
Technology shock	σ_a	$[0.00, 5.00]$
Risk premium shock	σ_b	$[0.00, 5.00]$
Gov. expenditure shock	σ_g	$[0.00, 5.00]$
Investment shock	σ_l	$[0.00, 5.00]$
Monetary policy shock	σ_r	$[0.00, 5.00]$
Price shock	σ_p	$[0.00, 5.00]$
Wage shock	σ_w	$[0.00, 5.00]$
AR term technology shock	ρ_a	$[-1.00, 1.00]$
AR term risk premium shock	ρ_b	$[-1.00, 1.00]$
AR term gov. shock	ρ_g	$[-1.00, 1.00]$
AR term investment shock	ρ_l	$[-1.00, 1.00]$
AR term MP shock	ρ_r	$[-1.00, 1.00]$
AR term price shock	ρ_p	$[-1.00, 1.00]$
AR term wage shock	ρ_w	$[-1.00, 1.00]$
MA term price shock	μ_p	$[-2.00, 2.00]$
MA term wage shock	μ_w	$[-2.00, 2.00]$
Correlation between a and g	ρ_{ga}	$[-1.00, 1.00]$
Target function value	Ω	$[0.00, Inf]$

4.3 Data

The data are the same as in Smets and Wouters (2007), that is, US quarterly data between 1947 to 2004.

The observables are, the logarithmic change in total real GDP (Δy_t), the logarithmic change in total real consumption (Δc_t), the logarithmic change in total real investment (Δi_t), the logarithmic change in the wage rate (Δw_t), deviations of hours worked from an average work week (Δl_t), the logarithmic change in the GDP deflator ($\Delta \pi_t$) and the federal funds rate (r_t). As in Smets and Wouters (2007) the first 71 observations are excluded which results in a sample of 156 observations.

5 Results

This section starts with a diagnostic test in order to verify that the estimates are indeed at the optimum. To check this I increase and decrease the value of one parameter at the time while the remaining parameters are fixed at the FIML estimate. After the "Diagnostics" section I present the simulated distributions that I use to calculate the confidence intervals. Lastly, the FIML estimates and their confidence intervals are presented together with the corresponding Bayesian estimates in SW.

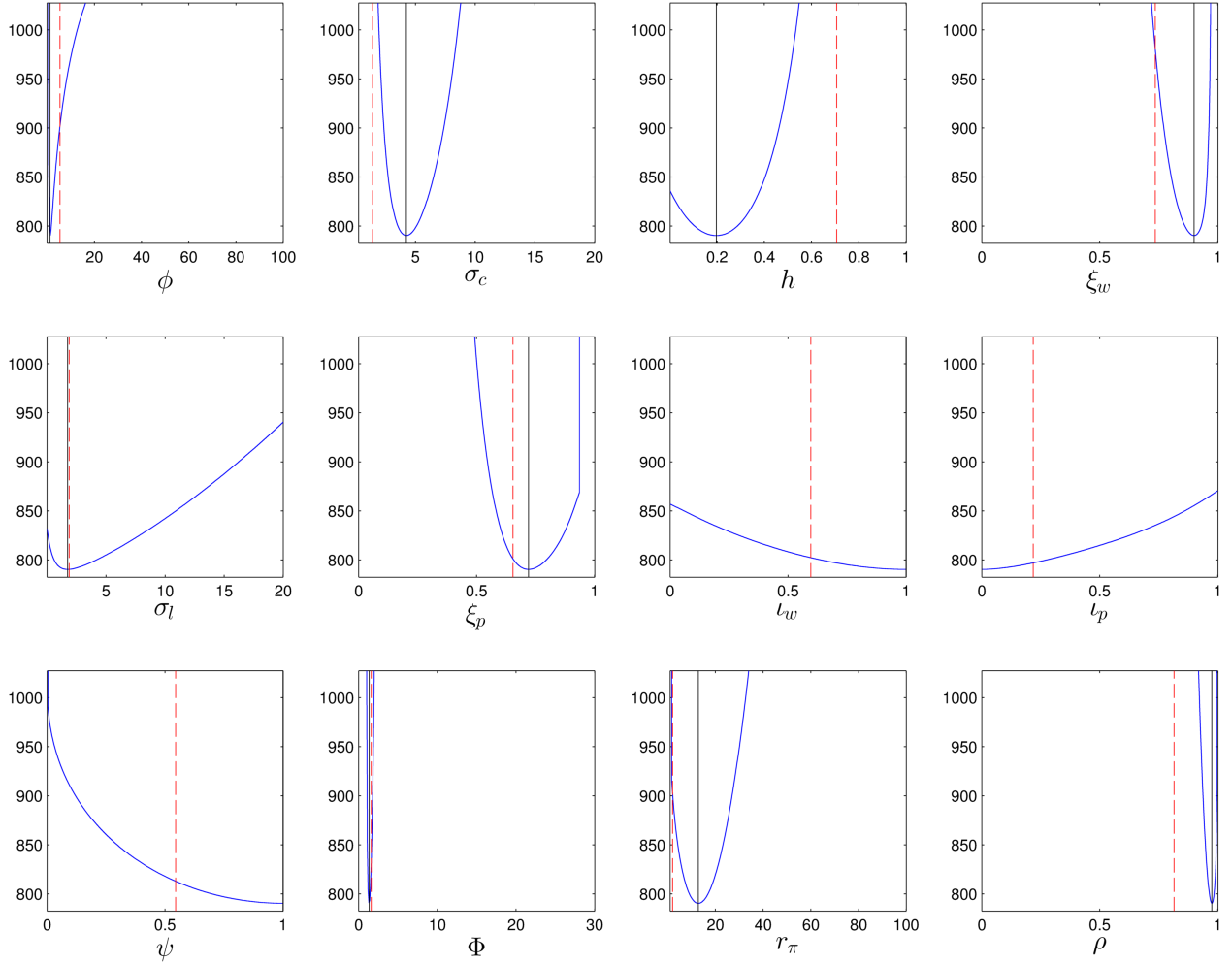
5.1 Diagnostics

The FIML estimates in Table 8 are evaluated by decreasing and increasing one parameter at the time while the remaining parameters are kept constant at the FIML estimate. I multiply the likelihood function by minus one and plot the value. If a minimum has in fact been reached the resulting plot should have its lowest point at the estimate (solid vertical line) and the objective function should increase as the parameter moves away from that value in either direction. I have chosen the same vertical scale for all graphs so that one can see whether the likelihood function is more flat in some dimensions. This can be confirmed by looking at Figures 6 to 8 where the vertical solid line is the estimate. The dashed vertical lines show SW's posterior mode estimates.

These diagnostic plots are also interesting since they indicate whether or not the prespecified intervals of the parameters in Table 6 are too small or not. According to the diagnostics, virtually all intervals have more than enough margin so that the minimum is at an interior point in the interval. The exceptions are parameters that are bounded based on economic

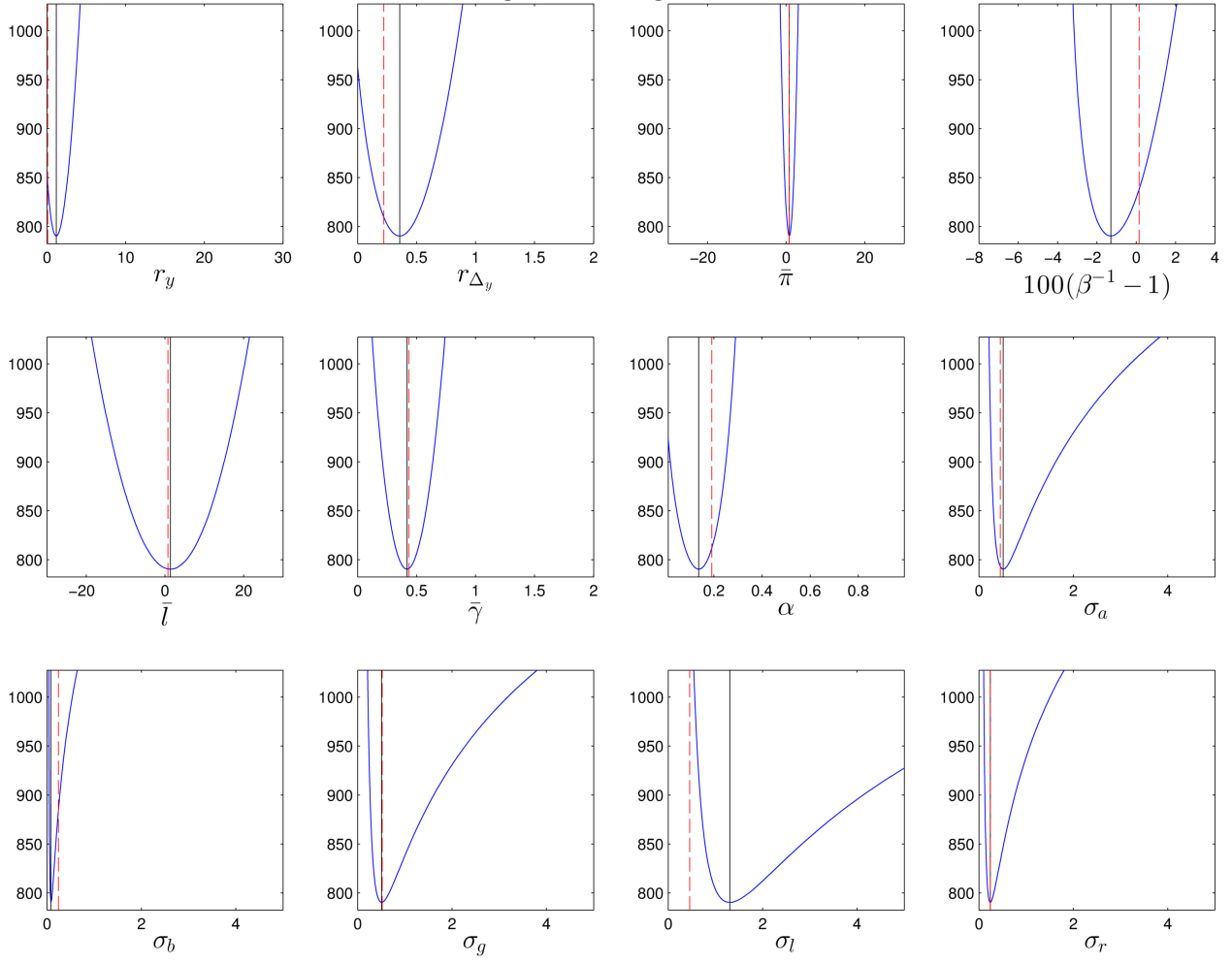
theory or for stationarity reasons. Worth mentioning here are the indices, ι_w , ι_p and capital utilization, ψ , as well as the persistence for some of the AR(1) processes, ρ_a , ρ_g , ρ_I and ρ_w . Those seem to be at or near a corner solutions and this issue is discussed in Section 6 where we analyze the differences between Bayesian and FIML estimates. Except for those, all parameters seem to be at an optimum where the target function clearly worsens as the parameter value moves away from its FIML estimate.

Figure 6: Diagnostics



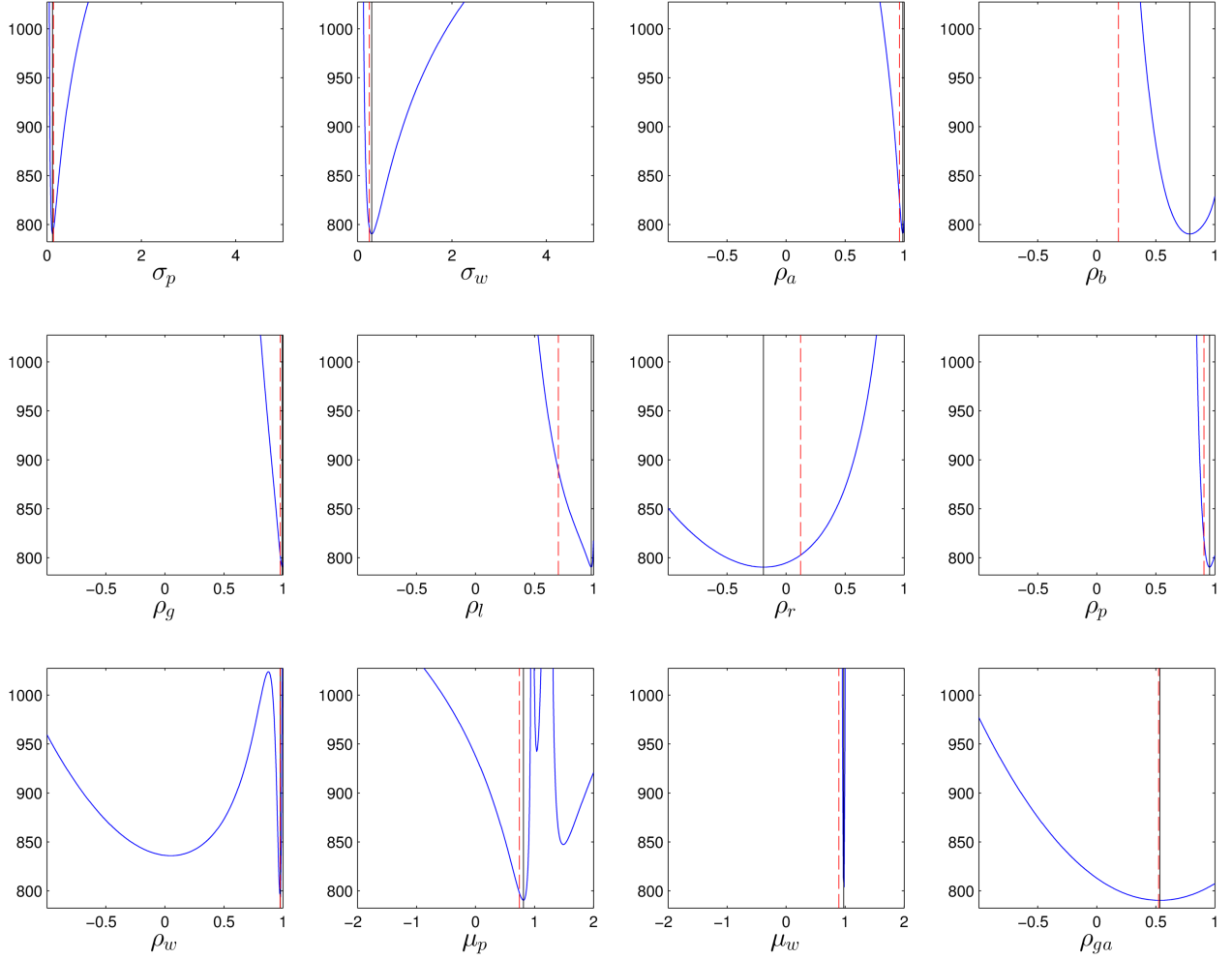
Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007)

Figure 7: Diagnostics



Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007)

Figure 8: Diagnostics



Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007)

5.2 Simulated distributions

I simulate the distributions in order to calculate the confidence interval of the FIML estimates. In SW it is assumed that the shocks have the properties in

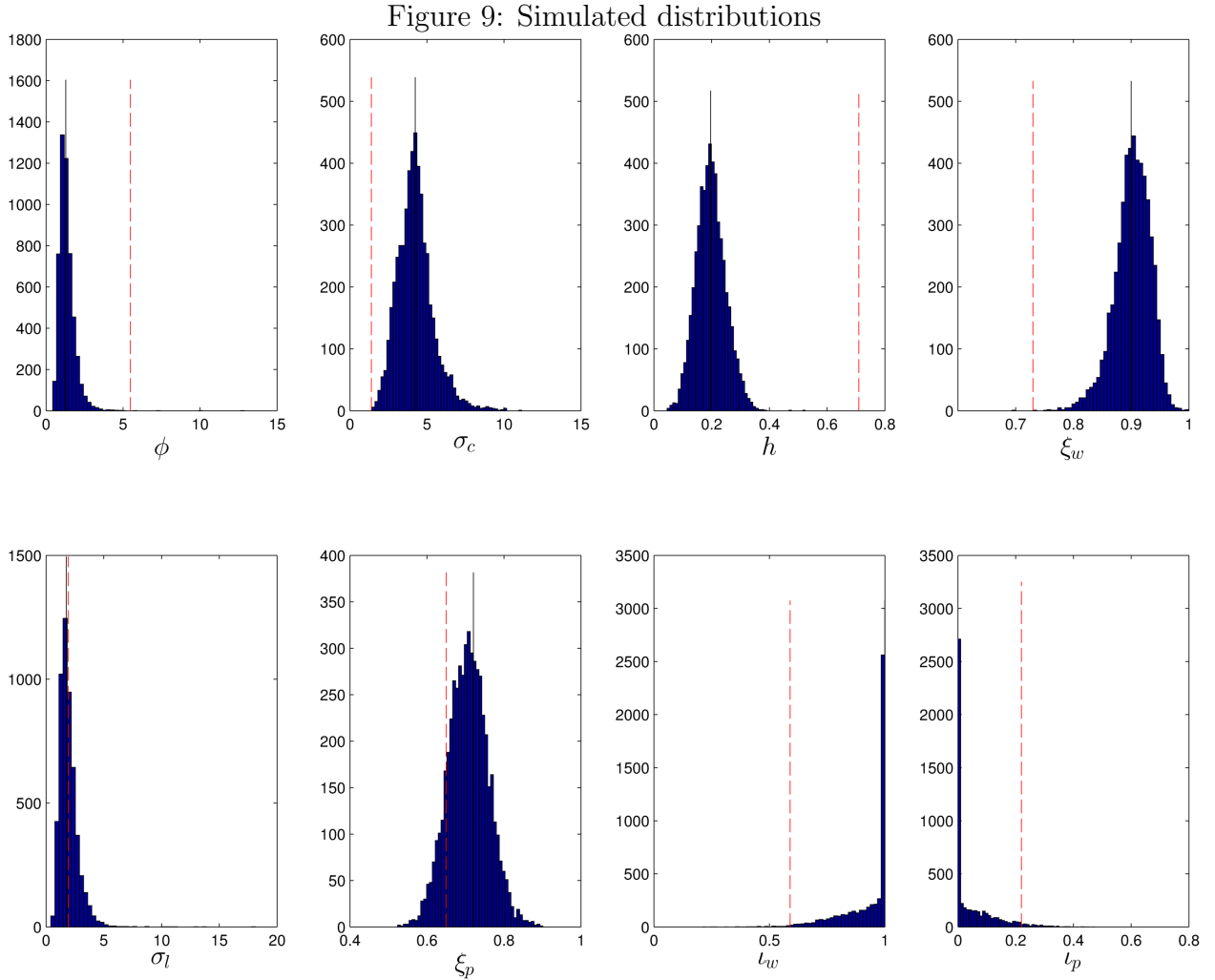
$$\vec{\varepsilon}_t \sim \mathcal{N}(0, \Sigma) \quad (14)$$

where

$$\vec{\varepsilon}_t = (\varepsilon_t^a, \varepsilon_t^b, \varepsilon_t^g, \varepsilon_t^I, \varepsilon_t^r, \varepsilon_t^p, \varepsilon_t^w) \quad (15)$$

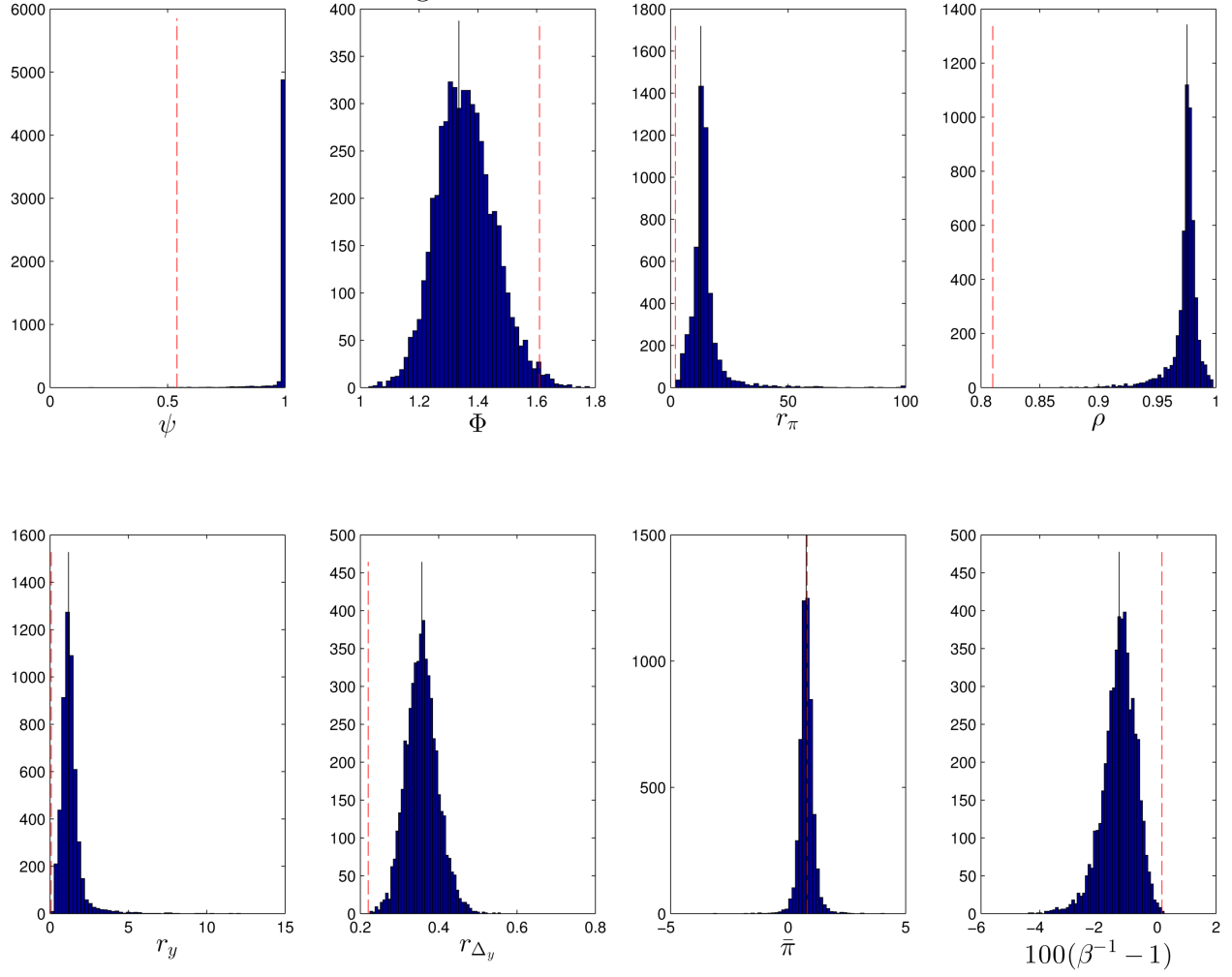
and where the diagonal elements of Σ consist of the variances in Table 9 and the off diagonal elements are zero.

Thus, all stochastic shocks follow a normal distribution with zero mean and they are identically and independently distributed. Based on SW's assumption regarding Σ and the estimates in Table 8 and Table 9 column "FIML" new samples are simulated with the same sample size (230) as the original data. Then the optimal vector is estimated for each sample. This procedure is repeated until the distributions in Figure 9 to Figure 13 have converged. For convergence criteria see the Appendix.



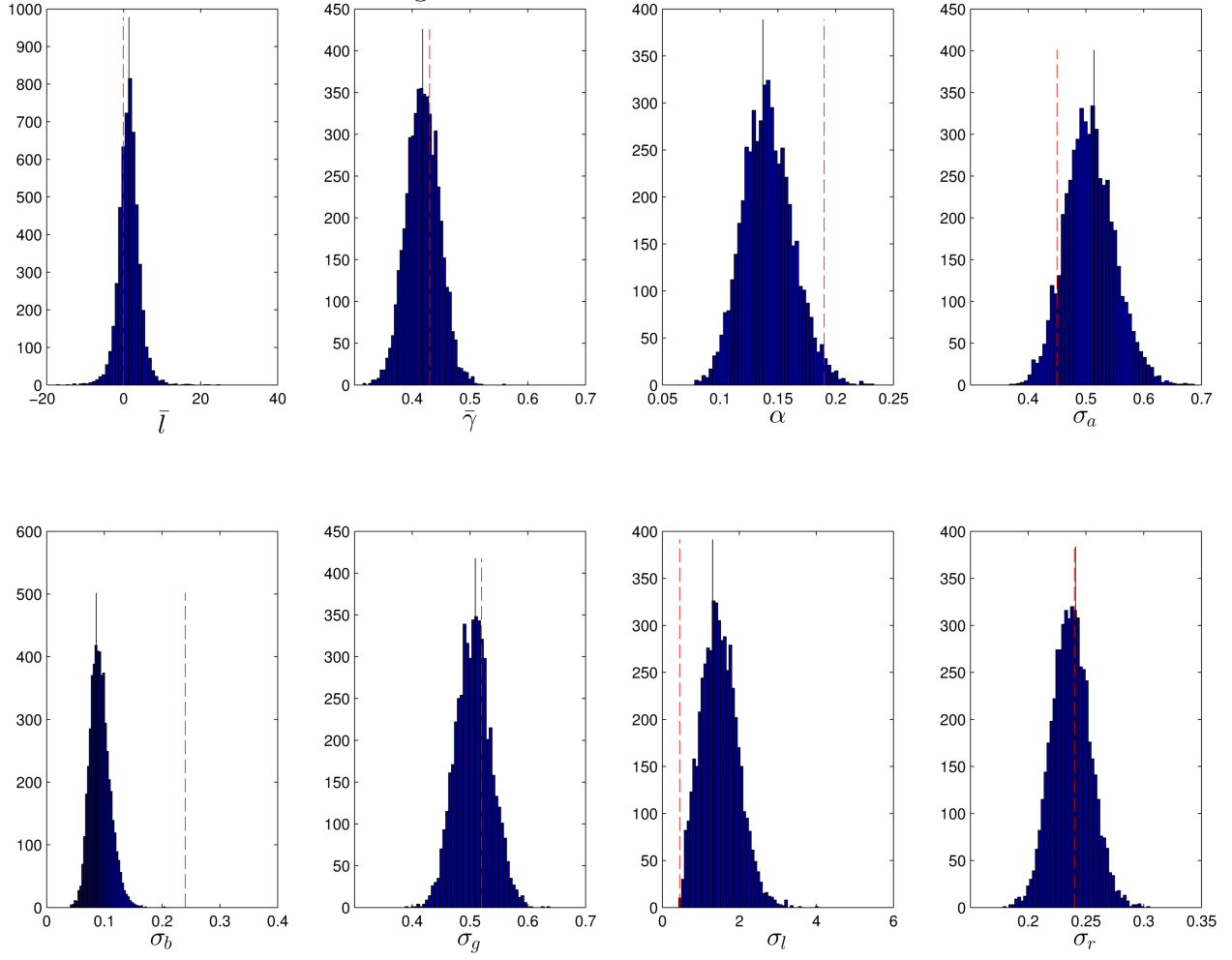
Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007). The number of simulated samples=4400

Figure 10: Simulated distributions



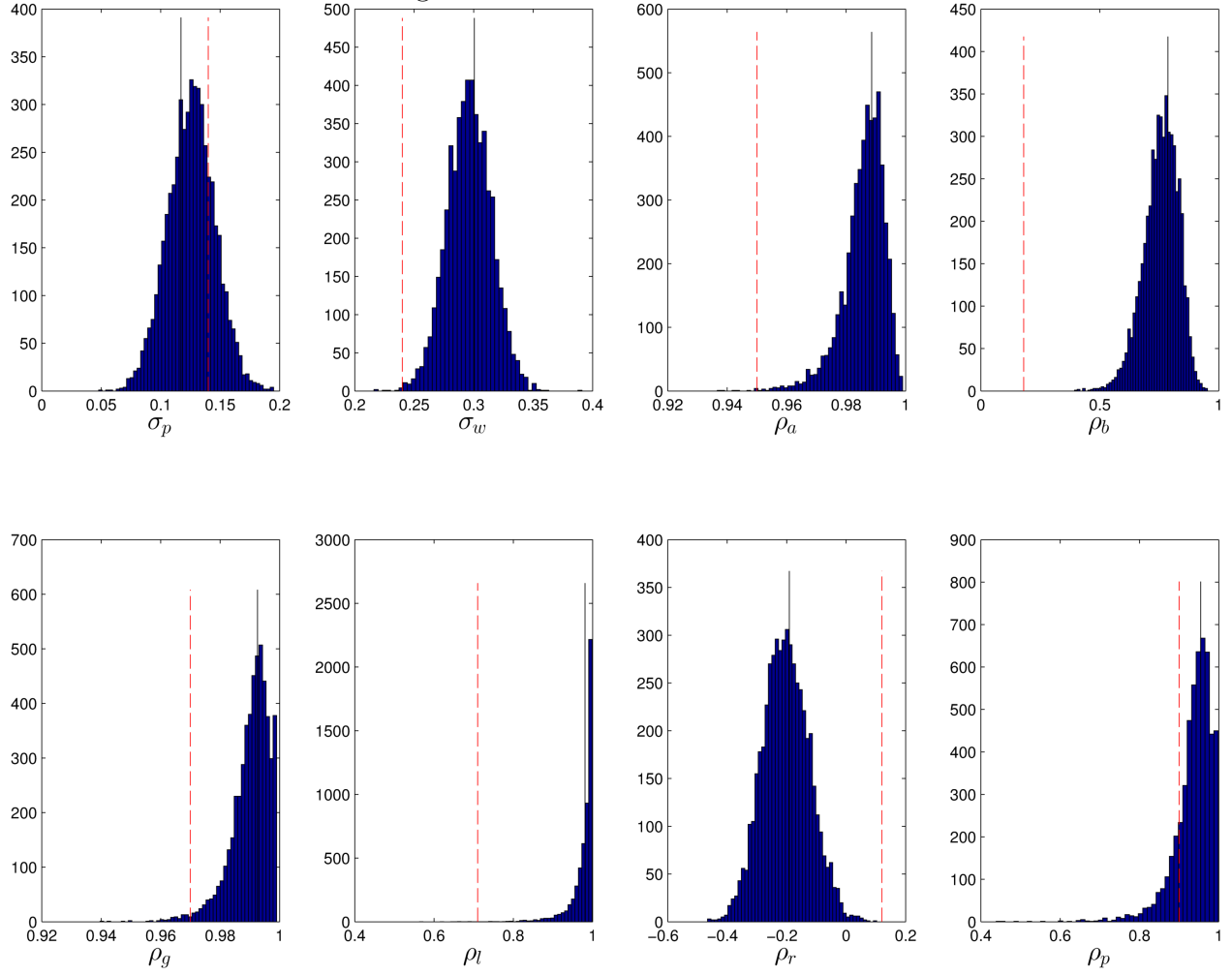
Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007). The number of simulated samples=4400

Figure 11: Simulated distributions



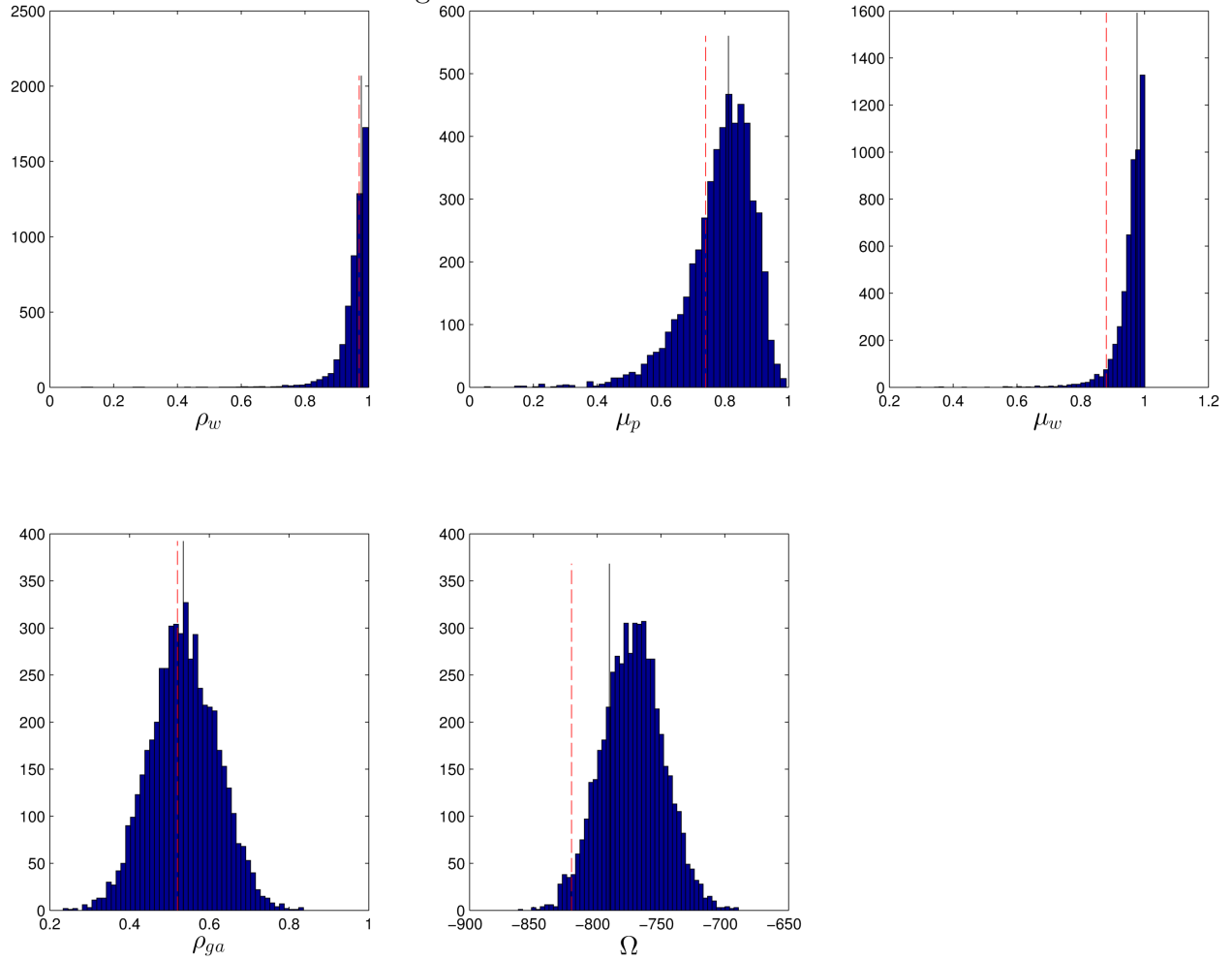
Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007). The number of simulated samples=4400

Figure 12: Simulated distributions



Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007). The number of simulated samples=4400

Figure 13: Simulated distributions



Note: The solid line is the FIML estimate and the dashed line is the posterior mode estimate from Smets and Wouters (2007). The number of simulated samples=4400

5.3 FIML versus Bayesian estimation - Estimates

Tables 8 and 9 show the mode estimate, its confidence interval and SWs posterior estimates. *, ** and *** indicate if SWs posterior mode estimates are outside of the 90%, 95% and 99% confidence intervals around the FIML estimates.

Table 8: Estimation results

Description	Parameter	FIML	95% FIML	SW	95% SW
Investment adjustment cost	ϕ	1.28	[0.56, 2.05]	5.48***	[3.97, 7.42]
Constant relative risk aversion	σ_c	4.24	[2.11, 6.39]	1.39***	[1.16, 1.59]
Habit persistence	h	0.20	[0.10, 0.29]	0.71***	[0.64, 0.78]
Calvo parameter labor market	ξ_w	0.90	[0.84, 0.97]	0.73***	[0.60, 0.81]
Inverse Frisch elasticity	σ_l	1.75	[0.60, 3.43]	1.92	[0.91, 2.78]
Calvo parameter goods market	ξ_p	0.72	[0.61, 0.81]	0.65	[0.56, 0.74]
Indexation labor market	ι_w	1.00	[0.71, 0.99]	0.59**	[0.38, 0.78]
Indexation goods market	ι_p	0.00	[0.00, 0.20]	0.22**	[0.10, 0.38]
Capital utilization	ψ	1.00	[0.97, 0.99]	0.54***	[0.36, 0.72]
Fixed production cost	Φ	1.34	[1.10, 1.52]	1.61***	[1.48, 1.73]
T.R reaction to inflation	r_π	12.88	[3.21, 22.77]	2.03***	[1.74, 2.33]
T.R interest rate smoothing	ρ	0.98	[0.95, 1.00]	0.81***	[0.77, 0.85]
T.R reaction to output	r_y	1.18	[0.14, 2.08]	0.08***	[0.05, 0.12]
T.R reaction to Δ_y	r_{Δ_y}	0.36	[0.28, 0.44]	0.22***	[0.18, 0.27]
Steady state inflation rate	$\bar{\pi}$	0.77	[0.24, 1.40]	0.81	[0.61, 0.96]
Sub. discount rate times 100	$100(\beta^{-1} - 1)$	-1.29	[-2.22, 0.00]	0.16**	[0.07, 0.26]
Steady state hours worked	\bar{l}	1.38	[-3.44, 6.77]	-0.10	[-1.30, 2.32]
Steady state growth rate	$\bar{\gamma}$	0.42	[0.36, 0.47]	0.43	[0.40, 0.45]
Capital share in production	α	0.14	[0.10, 0.18]	0.19**	[0.16, 0.21]

Note: FIML=Full information maximum likelihood mode estimates. SW=The Bayesian posterior estimates in Smets and Wouters (2007). SW's estimates outside of the FIML confidence intervals are denoted *=90%, **=95% and ***=99% respectively.

Table 9: Estimation results

Description	Parameter	FIML	95% FIML	SW	95% SW
Technology shock	σ_a	0.51	[0.43, 0.60]	0.45	[0.41, 0.50]
Risk premium shock	σ_b	0.09	[0.05, 0.12]	0.24***	[0.19, 0.27]
Gov. expenditure shock	σ_g	0.51	[0.46, 0.57]	0.52	[0.48, 0.58]
Investment shock	σ_l	1.30	[0.46, 2.27]	0.45***	[0.37, 0.53]
Monetary policy shock	σ_r	0.24	[0.21, 0.27]	0.24	[0.22, 0.27]
Price shock	σ_p	0.12	[0.09, 0.16]	0.14	[0.11, 0.16]
Wage shock	σ_w	0.30	[0.26, 0.33]	0.24***	[0.20, 0.28]
AR term technology shock	ρ_a	0.99	[0.97, 1.00]	0.95***	[0.94, 0.97]
AR term risk premium shock	ρ_b	0.79	[0.63, 0.92]	0.18***	[0.07, 0.36]
AR term gov. shock	ρ_g	0.99	[0.98, 1.00]	0.97**	[0.96, 0.99]
AR term investment shock	ρ_l	0.98	[0.92, 0.99]	0.71***	[0.61, 0.80]
AR term MP shock	ρ_r	-0.19	[-0.35, -0.05]	0.12***	[0.04, 0.24]
AR term price shock	ρ_p	0.95	[0.86, 0.99]	0.90	[0.80, 0.96]
AR term wage shock	ρ_w	0.98	[0.88, 0.99]	0.97	[0.94, 0.99]
MA term price shock	μ_p	0.81	[0.60, 0.98]	0.74	[0.54, 0.85]
MA term wage shock	μ_w	0.98	[0.88, 0.99]	0.88*	[0.75, 0.93]
Correlation between a and g	ρ_{ga}	0.53	[0.38, 0.70]	0.52	[0.37, 0.66]
Target function value	Ω	-790.37	[-811.76, -715.17]	-820.12**	$[-Inf, Inf]$

*Note:*FIML=Full information maximum likelihood mode estimates. SW=The Bayesian posterior estimates in Smets and Wouters (2007). SW's estimates outside of the FIML confidence intervals are denoted *=90%, **=95% and ***=99% respectively.

6 Analysis of differences between Bayesian and FIML estimates

In this section we look more closely at the cases where SW's parameter estimates are outside of the 99 % confidence intervals from the FIML estimation. For each parameter we first examine what role it plays in the model and then we try to understand why the FIML and SW's estimates are different. FIML and SW's estimates are denoted p_{FIML} and p_{SW} respectively for parameter p .

The purpose of this exercise is not to criticize the study by SW, but to use it as an example of a model that has been estimated by Bayesian methods and to show that the choice of priors had important effects on the posterior estimates. The use of priors is not innocuous. Presumably, the same applies to many other studies that have used Bayesian estimation.

6.1 Investment adjustment cost ϕ

Households own the capital and log linearizing the household's optimality condition for investment we get

$$i_t = \frac{1}{(1 + \beta\gamma^{(1-\sigma_c)})\gamma^2\phi} q_t + \frac{1}{1 + \beta\gamma^{(1-\sigma_c)}} i_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1 + \beta\gamma^{(1-\sigma_c)}} E_t i_{t+1} + \varepsilon_t^i \quad (16)$$

where q_t is the real value of the existing capital stock, defined as

$$q_t = \frac{1 - \delta}{R_*^k + 1 - \delta} E_t q_{t+1} + \frac{R_*^k}{R_*^k + 1 - \delta} E_t r_{t+1}^k - (r_t - E_t \pi_{t+1}) - \varepsilon_t^b \quad (17)$$

q_t is determined by a weighted average of the expected real value of capital, $E_t q_{t+1}$, and the expected future rental rate of capital, $E_t r_{t+1}^k$. The weights are $\frac{1-\delta}{R_*^k+1-\delta}$ and $\frac{R_*^k}{R_*^k+1-\delta}$ respectively, where R_*^k is the steady state rental rate of capital and ε_t^b is a shock to the risk premium of investment.

If the expected future real value of capital increases the household will save more in capital at a decreasing rate determined by ϕ . The reason for this is that a higher ϕ means that a bigger share of investment is wasted due to the adjustment cost. The FIML estimates

suggests a lower investment cost as $\phi_{FIML} = 1.28$ compared to $\phi_{SW} = 5.48$. The coefficient in front of q_t in equation (16) shows how many percentages investment goes up if the value of real capital goes up by one percent, *ceteris paribus*. The answer to that question is 0.39 percent for the FIML estimate and 0.09 percent for SW. That SW get a much higher value of ϕ is not surprising since they have the prior $\phi \sim N(4, 1.5)$ which has a center far above the FIML estimate. SW base their prior of choice on estimates from Christiano et al. (2005) whose investment model is simpler with fewer parameters and where investment depends on the real rental price of capital instead of the real value of capital. The analogous equation in Christiano et al. (2005) for investment is

$$i_t = i_{t-1} + \frac{1}{\phi} \sum_{j=0}^{\infty} \beta^j E_{t-1} r_{t+j}^k \quad (18)$$

They estimate ϕ by matching empirical impulse responses of their model and get the estimate $\phi = 2.48$ which corresponds to a value of 0.4 percent for the q_t coefficient in equation (16). From this we can conclude that SW used a somewhat higher prior value of ϕ than justified by the estimate in Christiano et al. (2005).

However, neither the real rental price of capital, r_t^k , nor the real value of capital, q_t , can be directly observed in the data. Simply speaking, the reason for adding adjustment costs in investment is to make the model match empirical fluctuations in investment. The standard deviation of the change in investment, $\sigma_{\Delta i}$, divided by the standard deviation of the change in output, $\sigma_{\Delta y}$, is an interesting measure of investment fluctuations since it is decreasing in ϕ . In the data, $\frac{\sigma_{\Delta i}}{\sigma_{\Delta y}} = 2.60$. When simulating the DSGE model using the FIML and SW estimates we get 2.62 and 2.13 respectively. This strengthens the argument that SW used a relatively a high prior value for ϕ .

6.2 The subjective discount factor β , relative risk aversion σ_c and habit persistence in consumption h

In order to reduce the income effect on labor supply SW use a utility function where consumption and labor are complements when relative risk aversion, σ_c , is greater than unity. This makes the income effect in labor supply independent of σ_c ; for a discussion of the role

of the utility function see Basu and Kimball (2002).

Households maximize utility

$$E_t \sum_{s=0}^{\infty} \beta^s \frac{(C_{j,t+s} - h\bar{C}_{j,t+s-1})^{1-\sigma_c}}{1-\sigma_c} \exp\left(\frac{\sigma_c-1}{1+\sigma_l} L_{j,t+s}^{1+\sigma_l}\right) \quad (19)$$

with respect to consumption $C_{j,t}$, hours worked $L_{j,t}$, investment $I_{j,t}$, bond holdings $B_{j,t}$ and capital utilization $Z_{j,t}$ and subject to the budget constraint

$$C_{j,t+s} + I_{j,t+s} + \frac{B_{j,t+s}}{\varepsilon_t^b R_{t+s} P_{t+s}} - T_{t+s} \leq \frac{B_{j,t+s-1}}{P_{t+s}} + \frac{W_{j,t+s}^h L_{j,t+s}}{P_{t+s}} + \frac{R_{t+s}^k Z_{j,t+s} K_{j,t+s-1}}{P_{t+s}} - a(Z_{j,t+s}) K_{j,t+s-1} + \frac{Div_{t+s}}{P_{t+s}} \quad (20)$$

and the capital accumulation equation

$$K_{j,t} = (1-\delta) + \varepsilon_t^i \left[1 - S \frac{I_{j,t}}{I_{j,t-1}}\right] I_{j,t}. \quad (21)$$

The subjective discount factor β decides how the household values the future relative to today. The estimated parameter that indirectly decides β is defined as

$$x = 100(\beta^{-1} - 1). \quad (22)$$

Solving for β we get

$$\beta = \frac{100}{x + 100}. \quad (23)$$

The estimates of x imply $\beta_{SW} = 0.998$ and $\beta_{FIML} = 1.013$ respectively. To investigate the plausibility of the estimates let us investigate the household's Euler equation in steady state

$$\beta = \frac{\gamma^{\sigma_c} \Pi 100}{100 + r} \quad (24)$$

where γ is the steady state growth rate of the trend variables³ in the model, Π is the steady state inflation rate and r is the steady state nominal interest rate. Based on values in the

³The trend variables in SWs data set are investment, consumption, real GDP and the real wage rate.

data, the average quarterly growth rate of the trend variables gives $\gamma = 1.0045$, the quarterly average growth rate in inflation is 1% and the quarterly average interest rate is 1.68%. Using these averages together with the estimates of σ_c gives $\beta_{SW} = 0.9995$ and $\beta_{FIML} = 1.012$ respectively. This implies that both estimates make sense. Note however, that by restricting β , σ_c is indirectly restricted since both the growth rate, γ , and the inflation rate, Π , are decided by the data.

SW restrict β to be between 0.98 and 1 although equation (24) clearly demonstrates that having $\beta > 1$ is possible in a model with growth and price level growth⁴. The assumption made by SW restricts σ_c to be between -3 and 1.48. Negative values of σ_c are not theoretically possible but restricting σ to be at most 1.48 on the other hand might be a bit low which will be discussed later.

The log-linearized Euler equation is

$$c_t = \frac{1}{1 + \frac{h}{\gamma}} \left[\frac{h}{\gamma} c_{t-1} + E_t c_{t+1} + \frac{\sigma_c - 1}{\sigma_c} \frac{W_*^h L_*}{C_*} (l_t - E_t l_{t+1}) - \frac{1 - \frac{h}{\gamma}}{\sigma_c} (r_t - E_t \pi_{t+1} + \varepsilon_t^b) \right] \quad (25)$$

Let us first discuss how c_t relates to lagged consumption, c_{t-1} . The effect that c_{t-1} has on c_t is decided by the habit parameter h . More specifically the coefficient relating c_t to c_{t-1} is

$$\epsilon_c \equiv \frac{dc_t}{dc_{t-1}} = \frac{h/\gamma}{1 + h/\gamma} \quad (26)$$

SW have a much higher habit formation, $h_{SW} = 0.71$ compared to $h_{FIML} = 0.2$, so that $\epsilon_{c,FIML} = 0.1660$ and $\epsilon_{c,SW} = 0.4141$ respectively.

A key parameter that decides the elasticity of hours worked, l_t , with respect to current consumption c_t is the risk aversion parameter σ_c . The higher the σ_c , given that $\sigma_c > 1$, the more are consumption and hours worked complements which is evident in equation (27)

$$\epsilon_{c,l} \equiv \frac{dc_t}{dl_t} = \frac{\sigma_c - 1}{\sigma_c} \frac{W_*^h L_*}{C_*} \quad (27)$$

Since $\sigma_{c,FIML} = 4.24$ the FIML estimate suggests a stronger complementary effect and when plugging in numbers we get $\epsilon_{c,l,FIML} = 0.6417$ and $\epsilon_{SW,c,l} = 0.2334$. Additionally, σ_c affects

⁴This means that the sum of the utility function is finite as long as equation (24) holds

the households sensitivity to interest rate changes together with the habit parameter h . Both σ_c and h weaken this sensitivity which is evident in equation (28)

$$\epsilon_{c,r} \equiv \frac{dc_t}{dr_t} = -\frac{\gamma/h - 1}{\gamma/h + 1} \frac{1}{\sigma_c} \quad (28)$$

$\epsilon_{c,r}$ is decreasing in both σ_c and h which means that a high h and a low σ_c can generate a similar interest rate sensitivity as a low h and a high σ_c . This is the reason why the FIML and SW estimates are similar; $\epsilon_{c,r,FIML} = -0.1575$ and $\epsilon_{c,r,SW} = -0.1236$ respectively. Thus, if the risk adjusted real interest rate goes up by 1 percent, current consumption goes down by about 0.14 percent for both estimates.

Why do the SW estimates suggest higher habit formation and lower risk aversion compared to FIML?

Referring to the discussion about β in the beginning of this section. When β is restricted to be at most 1, σ_c is implicitly restricted to be at most 1.48. Using the value for $\epsilon_{c,r}$, which is around -0.14, implicitly restricts h to be at least 0.66; see equation (29).

$$h = \gamma \frac{1 + \sigma_c \epsilon_{c,r}}{1 - \sigma_c \epsilon_{c,r}} \quad (29)$$

Furthermore, SW's priors for risk aversion and habit formation are $\sigma_c \sim N(1.4, 0.37)$ and $h \sim \beta(0.7, 0.1)$ respectively which also steers the results towards a strong habit formation and low risk aversion. SW justify their prior choice of h with the fact that the data suggests a high autocorrelation in consumption and to habit formation studies such as Fuhrer (2000) and McCallum and Nelson (1999).

6.3 Wage stickiness (ξ_w) and labor and goods market indexation (ι_w and ι_p)

Unions set wages according to a sticky price mechanism a la Calvo (1983). The wage rate W_{t+s} depends on $\tilde{W}_t(l)$, the wage rate of members belonging to union l that are allowed to change their wage in period t , and an index which is a weighted sum of past price inflation $\Pi_{i=1}^s \gamma \pi_{t+i-1}^{\iota_w}$ and steady state price inflation π_* according to (30)

$$W_{t+s}(l) = \tilde{W}_t(l) \Pi_{i=1}^s \gamma \pi_{t+i-1}^{\iota_w} \pi_*^{1-\iota_w} \quad (30)$$

The FIML estimate $\iota_w = 1$ means that past price inflation gets all the weight and steady state inflation none. The union, which is allowed to change its wage decides an optimum wage, $\tilde{W}_t(l)$, that will last until they are allowed to change it again. The probability of a wage change each period is $1 - \xi_w$ so the duration of the wage contract is $\frac{1}{1-\xi_w}$ periods.

Firms set their prices according to a Calvo (1983) mechanism where the price evolves according to.

$$P_{t+s} = \tilde{P}_t(i) \Pi_{j=1}^s \gamma \pi_{t+j-1}^{\iota_p} \pi_*^{1-\iota_p} \quad (31)$$

The FIML estimate $\iota_p = 0$ means that past inflation plays no role and the economy ends up with a regular forward looking Phillips curve; see for instance Gali (2009).

The most important parameter for wage-price dynamics is wage stickiness ξ_w . The estimates are $\xi_{w,SW} = 0.74$ and $\xi_{w,FIML} = 0.9$ which correspond to wage contract durations of 3.8 and 10 quarters respectively. The estimate by SW is significantly lower than the FIML estimate and this is most likely due to the prior used by SW: $\xi_w \sim \beta(0.5, 0.1)$. SW assume that price stickiness is similar to wage stickiness and base their prior on the micro evidence on price stickiness from Bils and Klenow (2002). There are not that many studies of wage stickiness but one is Barattieri et al. (2010) where they look at wage surveys and find that ξ_w is between 0.82 and 0.95 which corresponds to a duration between 5.6 and 20 quarters. This is in line with the FIML estimate and suggests that SWs estimate, $\xi_w = 0.74$, is too low because of an inappropriately chosen prior.

Higher wage stickiness implies a much slower pass-through of the marginal rate of substitution into wages which in turn decreases the variance of the wage. This is the most dominant effect on the wage dynamics. The indexation parameter ι_w has only a small effect in comparison.

6.4 Capital utilization ψ

Capital utilization ψ makes it possible to change production in the short run without hiring/firing nor changing the capital stock. The role of capital utilization can be explained by the log linearized production function

$$y_t = \phi_p (\alpha k_t^s + (1 - \alpha)l_t + \varepsilon_t^a) \quad (32)$$

where k_t^s is the capital stock that is actually used in production and can be temporary increased/decreased when the household orders a higher/lower utilization rate of current installed capital k_{t-1} . Utilized capital relative to stock in excess of installed capital, $k_t^s - k_{t-1}$, is determined by

$$k_t^s - k_{t-1} = \left(\frac{1}{\psi} - 1 \right) r_t^k \quad (33)$$

The utilization rate is positively correlated with the real rental rate of capital r_t^k since a higher r_t^k makes it more profitable to increase utilization. ψ determines the cost of utilizing capital and is normalized to be between 0 and 1. If $\psi = 1$ capital is infinitely expensive to utilize and the utilization channel is completely shut down so that $k_t^s = k_{t-1}$. This is what the FIML estimation suggests since $\psi_{FIML} = 1$. On the contrary, SW have capital utilization as an important feature with $\psi_{SW} = 0.54$. SW's estimate is significantly lower than the FIML estimate, so let us figure out why this is the case.

A low ψ dampens the effect that output has on the real rental rate of capital which in turn dampens the effect on marginal cost. This is the main reason why Christiano et al. (2005) introduced capital utilization of capital in the first place. How can we explain the difference between the FIML and SW results?

Since wage stickiness is higher for the FIML estimate, $\xi_{w,FIML} = 0.9$, compared to $\xi_{w,SW} = 0.73$, wage pass-through is low which makes the reaction in marginal cost to production low as well despite a big effect of production on the rental rate of capital. This effect can be seen in equation (34)

$$mc_t = \alpha r_t^k + (1 - \alpha)w_t \quad (34)$$

The reason why changes in wages have a dominant influence on marginal cost is because the income share of capital is much smaller than the income share of labor, $\alpha_{FIML} = 0.14$ and

$\alpha_{SW} = 0.19$. Since SW use the sticky wage prior $\xi_w \sim \beta(0.5, 0.1)$ the real wage response is relatively strong and the model is therefore forced to use utilization of capital in order to dampen the total effect of production on marginal cost and prices. Additionally, SW use the prior $\psi \sim \beta(0.5, 0.15)$ for capital utilization which also steers the estimate towards more variation in capital utilization and lower wage stickiness.

6.5 Fixed production cost Φ

The fixed production cost enters in the production function

$$y_t = Z_t k_{t-1}^\alpha L_t^{(1-\alpha)} - \Phi \quad (35)$$

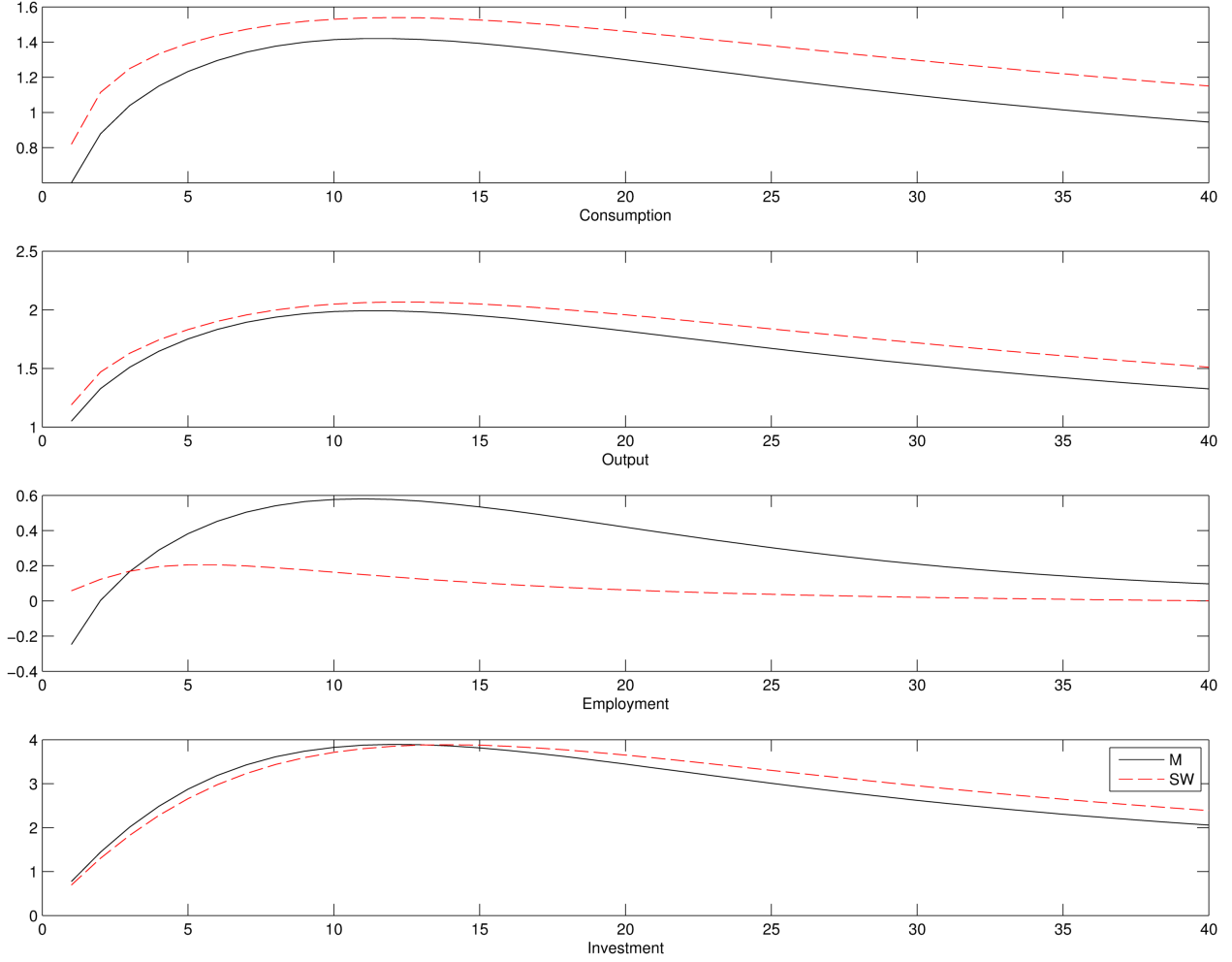
Rewriting the production function in terms of logarithmic deviations from steady state gives

$$\hat{y}_t = \left(1 + \frac{\Phi}{y^*}\right) \left(\alpha \hat{k}_t^s + (1 - \alpha) \hat{l}_t + \varepsilon_t^a\right) \quad (36)$$

where $\hat{y}_t = y_t - y_*$ and $*$ means the variable in steady state.

The reason why \hat{y}_t is increasing in Φ is because a higher fixed production cost decreases y_t which in turn increases the dynamic response measured in percent. This is demonstrated in Figure 14 below when subjecting the economy to the risk shock ε_t^b .

Figure 14: Impulse response of the risk shock



Note: Impulse response of a the risk shock ε_t^b where the solid line is the FIML estimate and the dashed line is the mode estimate from Smets and Wouters (2007)

6.6 Taylor rule parameters r_π , ρ , r_y and $r_{\Delta y}$

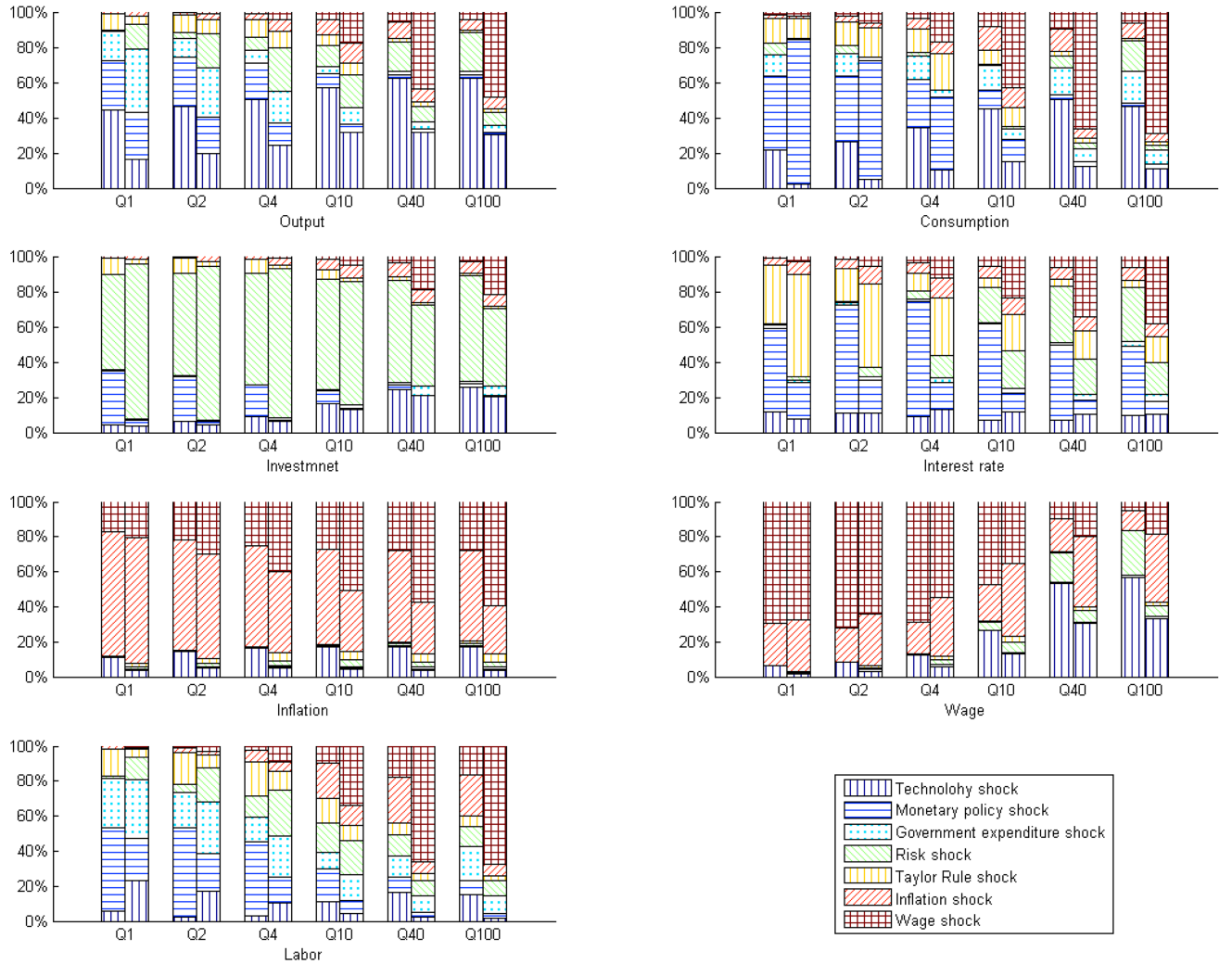
The FIML estimates of all the Taylor rule parameters are much higher than SW's estimates. Still, the central bank's reaction to shocks in the model is weaker compared to the estimates of SW due to a much higher interest rate smoothing, $\rho_{FIML} = 0.98$ and $\rho_{SW} = 0.81$.

6.7 The stochastic processes

The FIML estimates have larger shock variances and few of the AR(1) processes get really high persistences where $\rho_{a,FIML}$, $\rho_{g,FIML}$, $\rho_{I,FIML}$ and $\rho_{w,FIML}$ get values close to one.

This indicates that detrending the model assuming that the nonstationary variables fluctuate around a linear trend in logs might not be correct. Other detrending methods should therefore be considered such as detrending relative to technological progress $\hat{Y}_t = \frac{Y_t}{A_t}$. This is done by for instance Christiano et al. (2011). This way the trend is allowed to change dynamically over the business cycle.

Figure 15: Variance decomposition



Note: The bar to the left is the maximum likelihood estimate and the bar to the right is the posterior mode estimate in Smets and Wouters (2007)

The relatively high persistence of the technology shock, according to the FIML estimate, is probably the reason why this shock explains a bigger share of this variance of output and consumption.

7 Conclusion

It is popular to estimate DSGE models using Bayesian methods. This allows incorporation of prior information but in many cases there is little prior information available which makes full information maximum likelihood (FIML) an attractive alternative. FIML allows the researcher to easier find flaws in the model and generates estimates in cases where prior distributions are missing. However, estimating a DSGE model using FIML is much more challenging computationally.

In this paper I suggest an algorithm that makes it possible to estimate a medium size DSGE model using (FIML) without the need for priors. The method is applied to the model by Smets and Wouters (2007) and stochastic bootstrapping techniques are used to make inference. The algorithm outperforms other commonly used global algorithms.

I find that several of the FIML estimates are significantly different from those of Smets and Wouters (2007). I discuss these differences and find that some of the priors can be questionable. One example is the assumption that the discount factor has to be smaller than one. In a model with consumption growth and risk aversion it is possible to get a stationary solution even if the discount rate is larger than one which is what I get when I estimate the model by FIML.

In this analysis I have used the model by Smets and Wouters (2007) as an example of a medium size DSGE model, but the general insight should apply also to other studies. Priors are not innocuous and can have a big impact on the results even in cases where there is a lot of information in the data. When we have a weak basis for our priors, it may be interesting to let the data speak freely.

References

- Andreasen, Martin Mller (2009), “How to Maximize the Likelihood Function for a DSGE Model.” *Computational Economics*, 35, 127–154, URL <http://link.springer.com/article/10.1007/s10614-009-9182-6>.
- Barattieri, Alessandro, Susanto Basu, and Peter Gottschalk (2010), “Some Evidence on the Importance of Sticky Wages.” Working Paper 16130, National Bureau of Economic Research, URL <http://www.nber.org/papers/w16130>.
- Basu, Susanto and Miles Kimball (2002), “Long-run labor supply and the elasticity of intertemporal substitution for consumption.” *manuscript, University of Michigan*, URL <https://files.nyu.edu/cps272/public/ReadingGroup/Shenghao030207.pdf>.
- Bils, Mark and Peter J. Klenow (2002), “Some Evidence on the Importance of Sticky Prices.” Working Paper 9069, National Bureau of Economic Research, URL <http://www.nber.org/papers/w9069>.
- Calvo, Guillermo A. (1983), “Staggered prices in a utility-maximizing framework.” *Journal of Monetary Economics*, 12, 383–398, URL <http://www.sciencedirect.com/science/article/pii/0304393283900600>.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans (2005), “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy.” *Journal of Political Economy*, 113, 1–45, URL <http://www.jstor.org/stable/10.1086/426038>.
- Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin (2011), “Introducing financial frictions and unemployment into a small open economy model.” *Journal of Economic Dynamics and Control*, 35, 1999–2041, URL <http://www.sciencedirect.com/science/article/pii/S0165188911001710>.
- Fuhrer, J. C. (2000), “Habit formation in consumption and its implications for monetary-policy models.” *American Economic Review*, 367–390, URL <http://www.jstor.org/stable/10.2307/117334>.

- Gali, Jordi (2009), *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press.
- Goffe, William L., Gary D. Ferrier, and John Rogers (1994), “Global optimization of statistical functions with simulated annealing.” *Journal of Econometrics*, 60, 65–99, URL <http://www.sciencedirect.com/science/article/pii/0304407694900388>.
- Holland, John H. (1975), *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*, volume viii. U Michigan Press, Oxford, England.
- Ireland, Peter N. (2004), “A method for taking models to the data.” *Journal of Economic Dynamics and Control*, 28, 1205–1226, URL <http://www.sciencedirect.com/science/article/pii/S0165188903000800>.
- Kalman, R. E. (1960), “A New Approach to Linear Filtering and Prediction Problems.” *Journal of Basic Engineering*, 82, 35–45, URL <http://dx.doi.org/10.1115/1.3662552>.
- Karaboga, Dervis and Bahriye Basturk (2007), “A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm.” *Journal of Global Optimization*, 39, 459–471, URL <http://link.springer.com/article/10.1007/s10898-007-9149-x>.
- Kimball, Miles S. (1995), “The Quantitative Analytics of the Basic Neomonetarist Model.” Working Paper 5046, National Bureau of Economic Research, URL <http://www.nber.org/papers/w5046>.
- Kirkpatrick, Scott (1984), “Optimization by simulated annealing: Quantitative studies.” *Journal of Statistical Physics*, 34, 975–986, URL <http://link.springer.com/article/10.1007/BF01009452>.
- McCallum, Bennett T. and Edward Nelson (1999), “Nominal income targeting in an open-economy optimizing model.” *Journal of Monetary Economics*, 43, 553–578, URL <http://www.sciencedirect.com/science/article/pii/S0304393299000057>.

Nelder, J. A. and R. Mead (1965), “A Simplex Method for Function Minimization.” *The Computer Journal*, 7, 308–313, URL <http://comjnl.oxfordjournals.org/content/7/4/308>.

Smets, Frank and Raf Wouters (2003), “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association*, 1, 1123–1175, URL <http://www.jstor.org/stable/40004854>.

Smets, Frank and Rafael Wouters (2007), “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *The American Economic Review*, 97, 586–606, URL <http://www.jstor.org/stable/30035013>.

A Algorithm description

A.1 Global part

Each parameter p has its min and max value, p_{min} and p_{max} . First let us introduce the new min and max definitions

$$p_{min_c} = p_{min} + (p_{max} - p_{min})c \quad (37)$$

and

$$p_{max_c} = p_{max} - (p_{max} - p_{min})c \quad (38)$$

where c is a small number and the idea is to have bounds that are a little bit apart from the bounds than p_{min} and p_{max} .

The next step is to create N vectors where each vector has the dimension $P \times 1$ where P is the number of parameters. Each element of every vector is generated with the following function

$$x_p^i = p_{min_c} + (p_{max_c} - p_{min_c}) \frac{i}{N} \quad (39)$$

All the resulting N vectors are stored in a matrix

$$\begin{bmatrix} x_1^1 & x_1^2 & \cdots & x_1^N \\ x_2^1 & x_2^2 & \cdots & x_2^N \\ \vdots & \vdots & \ddots & \vdots \\ x_P^1 & x_P^2 & \cdots & x_P^N \end{bmatrix} \quad (40)$$

Each element is multiplied with the stochastic process

$$\varepsilon_p^i \sim \text{uniform}(1 - \delta, 1 + \delta) \quad (41)$$

where δ is a small number and the idea is to get a small stochastic variance around all parameters.

$$X_\varepsilon = \begin{bmatrix} \varepsilon_1^1 x_1^1 & \varepsilon_1^2 x_1^2 & \cdots & \varepsilon_1^N x_1^N \\ \varepsilon_2^1 x_2^1 & \varepsilon_2^2 x_2^2 & \cdots & \varepsilon_2^N x_2^N \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_P^1 x_P^1 & \varepsilon_P^2 x_P^2 & \cdots & \varepsilon_P^N x_P^N \end{bmatrix} \quad (42)$$

Additionally two more vectors are added with the first one very close to origo and the second one very close to the maximum value of all parameters. This in order to catch the minimum and maximum parts of the space as well.

$$\vec{X}^{min} = \begin{pmatrix} \varepsilon_1^{min} \\ \varepsilon_2^{min} \\ \vdots \\ \varepsilon_P^{min} \end{pmatrix}, \varepsilon_p^{min} \sim \text{uniform}(p_{min}, p_{min} + (p_{max} - p_{min})\delta) \quad (43)$$

and

$$\vec{X}^{max} = \begin{pmatrix} \varepsilon_1^{max} \\ \varepsilon_2^{max} \\ \vdots \\ \varepsilon_P^{max} \end{pmatrix}, \varepsilon_p^{max} \sim \text{uniform}(p_{max} - (p_{max} - p_{min})\delta, p_{max}) \quad (44)$$

$$X = \begin{bmatrix} \vec{X}^{min} & \vec{X}^{max} & X_\varepsilon \end{bmatrix}_{Px2+N} \quad (45)$$

Each vector in matrix 45 are evaluated the following way. First matrix A is created around vector X_i

$$A = \begin{bmatrix} \vec{X} & \vec{A}_{min} & \vec{A}_{max} \end{bmatrix} = \begin{bmatrix} x_1 & x_1^{min} & x_1 & \cdots & x_1 & x_1^{max} & x_1 & \cdots & x_1 \\ x_2 & x_2 & x_2^{min} & \cdots & x_2 & x_2 & x_2^{max} & \cdots & x_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_P & x_P & x_P & \cdots & x_P^{min} & x_P & x_P & \cdots & x_P^{max} \end{bmatrix}_{(P,1+2P)}$$

Matrix A is then run through the Nelder and Mead (1965) simplex method until

$$\Delta \log X^* < 1$$

where X^* is the vector with the best function value. Additionally, the process is restarted every $5 \times P$ th time around the vector with the best function value thus far in order not to get stuck in a local minimum.

This step is repeated for each one of the vectors which result in $2 + N$ vectors. These vectors are then sorted in order of the best value. The $P + 1$ vectors that have the best value are saved in a new matrix and are run through the Nelder and Mead (1965) simplex method. The resulting vector is then saved.

This entire process is repeated three times.

A.2 Local part

Three vectors are saved as a result of the global part, these three vectors are in turn run through the Nelder and Mead (1965) simplex method. This creates an output vector which in turn is used to create matrix A as in the local part and is run through the Nelder and Mead (1965) simplex method. That resulting vector is the final vector!

A.3 The Nelder and Mead (1965) algorithm - applied to this example

The algorithm consists of the following steps.

Step 1. Order all vectors in Matrix D (equation (??)) in the order of their target values. Calculate the center of the $P-1$ best vectors. From the center walk approximately in the

opposite direction from the worst vector. If this reflection vector is better than the best vector investigate this direction further by going further in the same direction (extend). If the extended vector is not better than the reflection vector replace the worst vector with the reflection vector. If the extended vector is better than the reflection vector replace the worst vector with that one instead. Then repeat step 1.

If the reflection vector is not better than the second worst vector but better than the worst vector make an outside contraction. That is, use the same idea as reflection but make a smaller step. Replace the worst vector if it is better than the reflection point and repeat step 1.

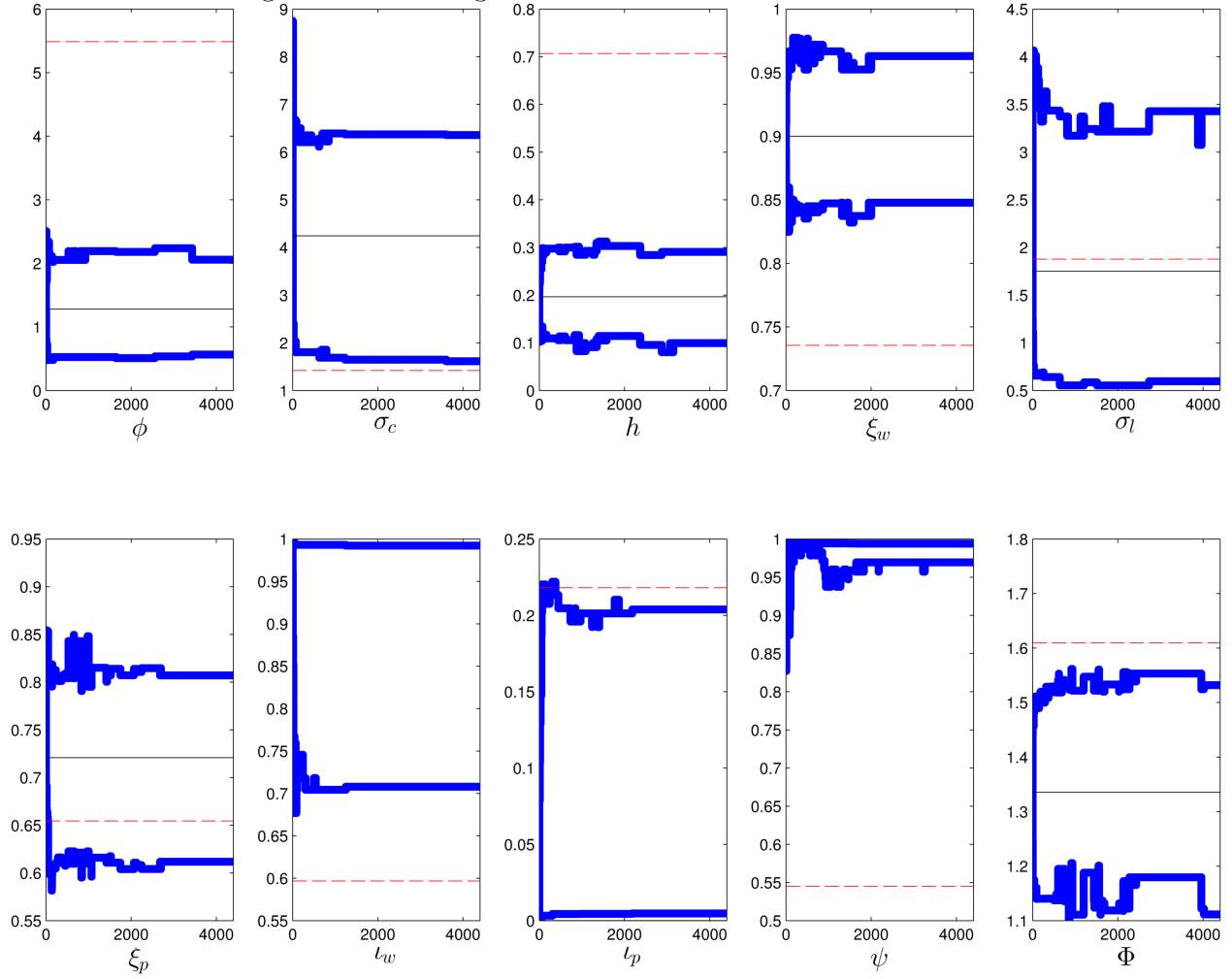
If the outside contraction vector is not better than the reflection vector create a weighted average with the center and the worst vector. If this vector is better than the worst vector replace it and repeat step 1.

If it is not better than the worst vector perform a shrink where all vectors take one step closer to the best vector. Repeat step 1.

B Convergence diagnostics of the histograms

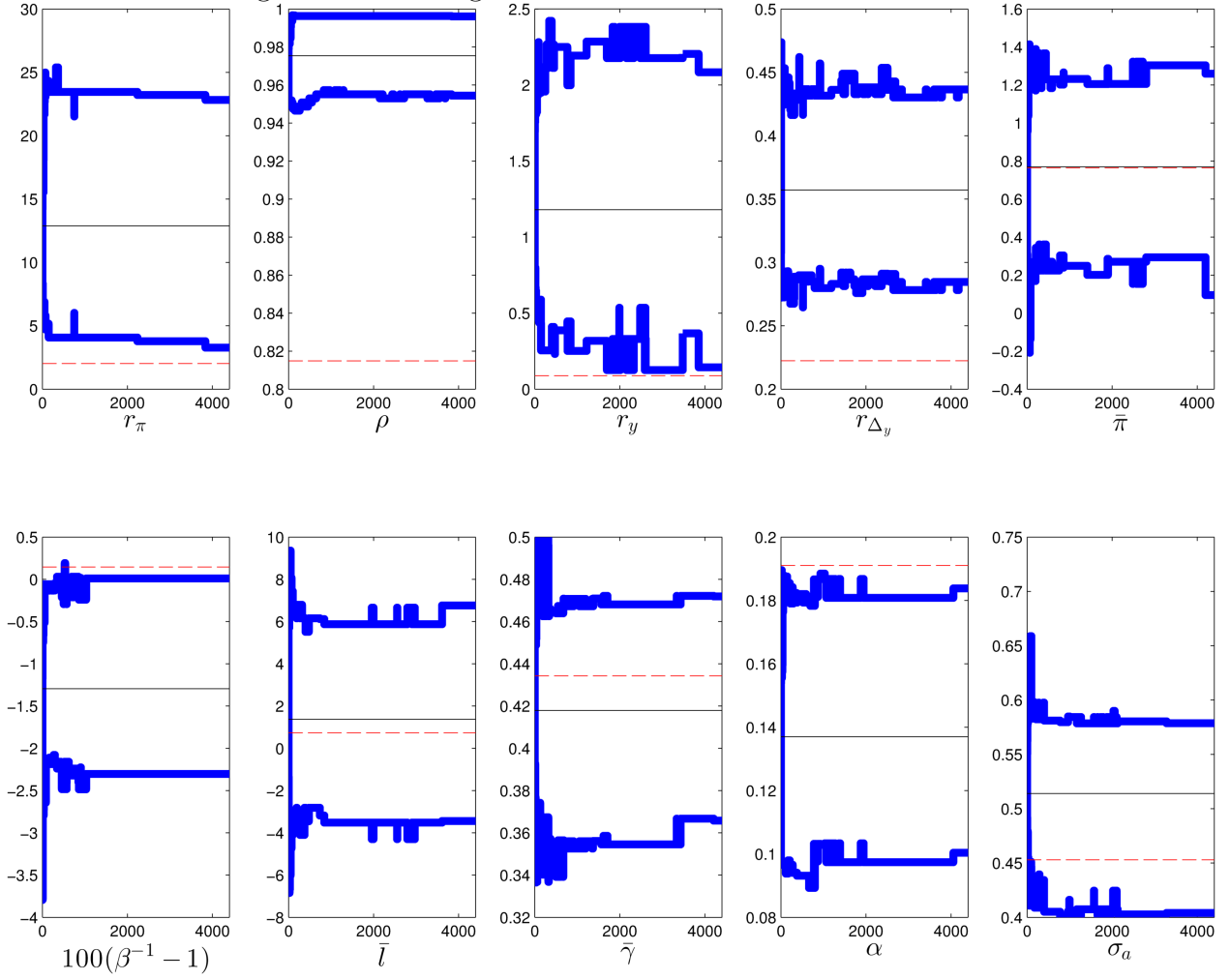
The thick horizontal lines in Figure 16 to 19 below show $P_{min}^{95\%}(i)$ and $P_{max}^{95\%}(i)$ for each variable. $P_{min}^{95\%}(i)$ and $P_{max}^{95\%}(i)$ is the lower and upper bound of the 95 % confidence interval respectively for parameter P based on i simulations. i is the value of the x-axis and $i \in 1 \dots 4400$. The more simulations (higher i) the more certain should the interval become.

Figure 16: Convergence of the distributions in Section 5.2



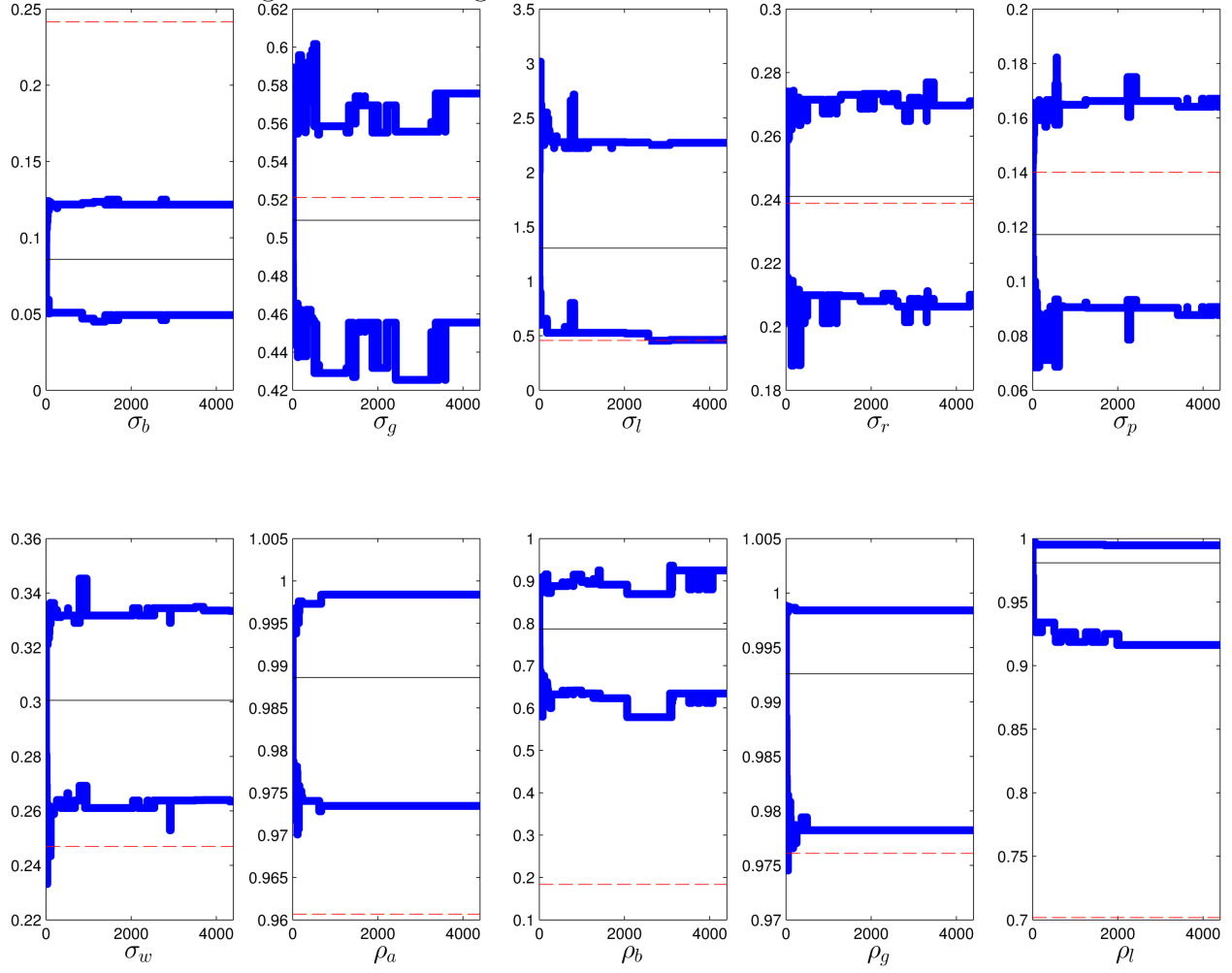
Note: The solid horizontal line is the FIML estimate, the dashed line the SW estimate and the thick horizontal lines show $P_{min}^{95\%}(i)$ and $P_{max}^{95\%}(i)$ of parameter i respectively.

Figure 17: Convergence of the distributions in Section 5.2



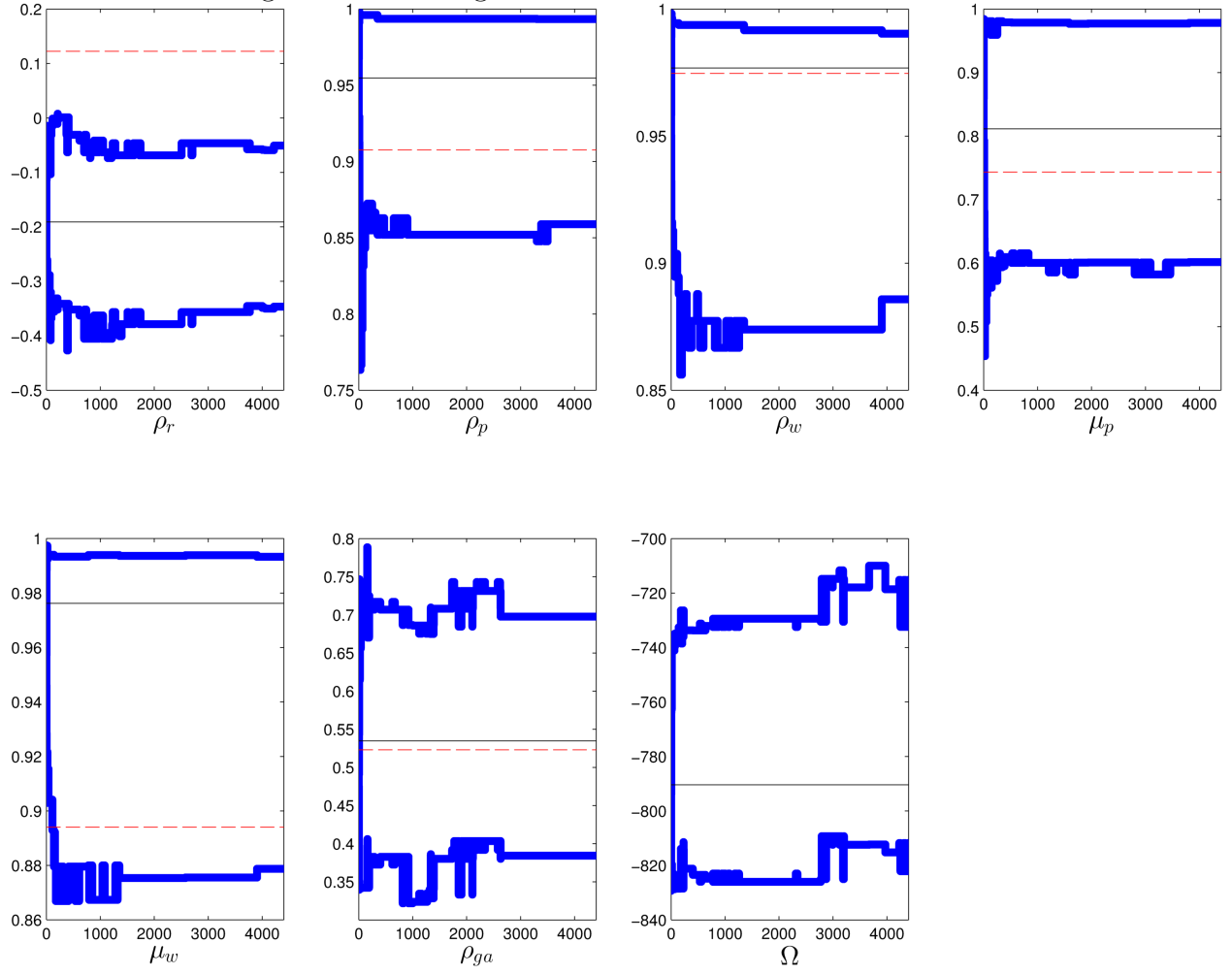
Note: The solid horizontal line is the FIML estimate, the dashed line the SW estimate and the thick horizontal lines show $P_{min}^{95\%}(i)$ and $P_{max}^{95\%}(i)$ of parameter i respectively.

Figure 18: Convergence of the distributions in Section 5.2



Note: The solid horizontal line is the FIML estimate, the dashed line the SW estimate and the thick horizontal lines show $P_{min}^{95\%}(i)$ and $P_{max}^{95\%}(i)$ of parameter i respectively.

Figure 19: Convergence of the distributions in Section 5.2



Note: The solid horizontal line is the FIML estimate, the dashed line the SW estimate and the thick horizontal lines show $P_{min}^{95\%}(i)$ and $P_{max}^{95\%}(i)$ of parameter i respectively.

WORKING PAPERS*

Editor: Nils Gottfries

- 2013:7 Che-Yuan Liang, Optimal Inequality behind the Veil of Ignorance. 26 pp.
- 2013:8 Pia Fromlet, Monetary Policy Under Discretion Or Commitment? -An Empirical Study. 57 pp.
- 2013:9 Olof Åslund and Mattias Engdahl, Open borders, transport links and local labor markets. 41 pp.
- 2013:10 Mohammad Sepahvand, Roujman Shahbazian and Ranjula Bali Swain, Time Investment by Parents in Cognitive and Non-cognitive Childcare Activities. 31 pp.
- 2013:11 Miia Bask and Mikael Bask, Social Influence and the Matthew Mechanism: The Case of an Artificial Cultural Market. 13 pp
- 2013:12 Alex Solis, Credit access and college enrollment. 54 pp
- 2013:13 Alex Solis, Does Higher Education Cause Political Participation?: Evidence From a Regression Discontinuity Design. 48 pp.
- 2013:14 Jonas Kolsrud, Precaution and Risk Aversion: Decomposing the Effect of Unemployment Benefits on Saving. 37 pp.
- 2013:15 Helena Svaleryd, Self-employment and the local business cycle. 28 pp.
- 2013:16 Tobias Lindhe and Jan Södersten, Distortive Effects of Dividend Taxation. 22 pp.
- 2013:17 Jonas Poulsen, After Apartheid: The Effects of ANC Power. 42 pp.
- 2013:18 Magnus Gustavsson, Permanent versus Transitory Wage Differentials and the Inequality-Hours Hypothesis. 11 pp.
- 2013:19 Lovisa Persson, Consumption smoothing in a balanced budget regim. 35 pp.
- 2013:20 Linuz Aggeborn, Voter Turnout and the Size of Government. 50 pp.
- 2013:21 Niklas Bengtsson, Stefan Peterson and Fredrik Sävje, Revisiting the Educational Effects of Fetal Iodine Deficiency. 48 pp
- 2013:22 Michihito Ando, How Much Should We Trust Regression-Kink-Design Estimates? 69 pp.
- 2013:23 Bertil Holmlund, What do labor market institutions do? 25 pp.
- 2014:1 Oscar Erixson and Henry Ohlsson, Estate division: Equal sharing as choice, social norm, and legal requirement. 45 pp.

* A list of papers in this series from earlier years will be sent on request by the department.

- 2014:2 Eva Mörk, Anna Sjögren and Helena Svaleryd, Parental unemployment and child health. 35 pp.
- 2014:3 Pedro Carneiro, Emanuela Galasso and Rita Ginja, Tackling Social Exclusion: Evidence from Chile. 87 pp.
- 2014:4 Mikael Elinder och Lovisa Persson, Property taxation, bounded rationality and housing prices. 38 pp.
- 2014:5 Daniel Waldenström, Swedish stock and bond returns, 1856–2012. 51 pp.
- 2014:6 Mikael Carlsson, Selection Effects in Producer-Price Setting. 34 pp.
- 2014:7 Ranjula Bali Swain and Fan Yang Wallentin, The impact of microfinance on factors empowering women: Differences in regional and delivery mechanisms in India's SHG programme. 28 pp.
- 2014:8 Ashim Kumar Kar and Ranjula Bali Swain, Competition, performance and portfolio quality in microfinance markets. 33 pp.
- 2014:9 Shyam Ranganathan, Ranjula Bali Swain and David J.T. Sumpter, A DYNAMICAL SYSTEMS APPROACH TO MODELING HUMAN DEVELOPMENT. 42 pp.
- 2014:10 Shyam Ranganathan and Ranjula Bali Swain, Analysing Mechanisms for Meeting Global Emissions Target - A Dynamical Systems Approach. 40 pp.
- 2014:11 Chuan-Zhong Li and Ranjula Bali Swain, Growth, Water Resilience, and Sustainability: A DSGE Model Applied to South Africa. 21pp.
- 2015:1 Evelina Bonnier, Jonas Poulsen, Thorsten Rogall and Miri Stryjan, Preparing for Genocide: Community Work in Rwanda. 40 pp.
- 2015:2 Linna Martén, Political Bias in Court? Lay Judges and Asylum Appeals. 38 pp.
- 2015:3 Daniel Waldenström, The National Wealth of Sweden, 1810–2014. 31 pp.
- 2015:4 Daniel Waldenström, Wealth-income ratios in a small, late-industrializing, welfare-state economy: Sweden, 1810–2014. 41 pp.
- 2015:5 Mattias Nordin, Local Television, Citizen Knowledge and Political Accountability: Evidence from the U.S. Senate. 45 pp.
- 2015:6 Glenn Mickelsson, Estimation of DSGE models: Maximum Likelihood vs. Bayesian methods. 51 pp.

See also working papers published by the Office of Labour Market Policy Evaluation
<http://www.ifau.se/> ISSN 1653-6975