Four Essays on Identification-Robust Numerical and Statistical Tools with Applications to Dynamic Stochastic General Equilibrium Models

by

Zhenjiang Lin

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Economics

Department of Economics

Carleton University

Ottawa, Ontario, Canada

Abstract

In this thesis, we propose identification-robust numerical and statistical confidence-set estimation tools for Dynamic Stochastic General Equilibrium [**DSGE**] models.

The first chapter introduces Particle Swarm Optimization [**PSO**] to econometrics with focus on test inversion and numerical projections. Empirically, the paper analyzes a three-equation New-Keynesian model. In contrast to PSO, the genetic algorithm, simulated annealing and even grid searches converge to local optima that suggest misleading economic decisions on: (i) the nature of the New-Keynesian Phillips Curve [**NKPC**], (ii) determinacy of monetary policies, and (iii) the persistence of the Taylor rule. Given this evidence and using PSO, the next three chapters introduce new and improved econometric methods for inference on DSGEs.

The second chapter documents the sensitivity of the identification of DSGE models to auxiliary assumptions on exogenous shock processes using the identification-robust inference procedure of Dufour, Khalaf and Kichian (2013). We find that adding lagged endogenous variables instead of assuming autoregressive shocks to capture external propagation improves the identification of important parameters, even when the slope of the NKPC is near zero. Also, the asymptotic Likelihood Ratio [**LR**] test is oversized in small sample, which provides motivation for the third chapter.

The third chapter proposes a finite-sample exact confidence-set estimation method using a LR distance measure for DSGEs. We demonstrate that our method has exact size in finite samples, and use it to estimate a canonical three-equation New-Keynesian

DSGE model. We find no conclusive evidence that the "pre-Great Moderation" era conforms to a passive monetary policy rule corresponding to an indeterminate model, as argued in the literature.

The final chapter extends the method proposed in the third chapter beyond the LR distance measure and approximate Vector Autoregression [**VAR**] context for DSGE models. We propose to use a VAR with Leads and Lags [**VARLL**] as the alternative reduced form and estimate the volatility parameters simultaneously together with other parameters.

Taken collectively, the results of this thesis suggest that far more attention needs to be paid to numerical precision as test inversion gains popularity in applied econometrics.

To my mother,

Jing Lin,

and my father,

Weixing Lin

Wish you health and happiness.

Acknowledgements

I gratefully thank my thesis supervisor, Lynda Khalaf, for her guidance, patience understanding and encouragement throughout the dissertation process. I especially enjoyed working with her in several projects which led to the major chapters of my thesis. I thank her for providing me with the foundation for becoming an econometrician. I look forward to productive collaborations with her in the future.

I am thankful to my committee members, Marcel-Cristian Voia, Hashmat Khan and Francesca Rondina, for insightful comments and questions.

I thank Patrick Coe, Till Gross, Christopher Gunn, Kim Huynh, Konstantinos Metaxoglou, Abeer Reza and many other faculties for offering me advises and helps during my job search.

I would also like to thank my colleagues and classmates, Anand Acharya, Maoqiong Lin, Beatriz Peraza Lopez and Charles Saunders for useful discussions on matters economic and otherwise. I am grateful to many faculties and staff of the Economics department at Carleton University for their encouragement and support.

Finally, I thank my parents and grandparents, for their faith in me and allowing me to be as ambitious as I wanted.

Table of Contents

List of Tables

List of Figures

Chapter 1

Statistical Test Inversion and Numerical Projections with Particle Swarm Optimization

1.1 Introduction

Important developments in economics and econometrics have often been driven by improvements and increased use of efficient numerical and computational tools. In econometrics, one important and recent research area relates to non-standard test inversion and projection-based inference methods, most of which can require technically complex computing. Projection-based inference methods have gained popularity following the introduction of non-standard tests in Instrumental Variable [**IV**] regression and with the Generalized Method of Moments [**GMM**], to account for weak-identification and/or weak instruments.[1] From the applied econometrics per-

[1]For a survey and a recent perspective, see Dufour (1997, 2003), Staiger and Stock (1997), Wang and Zivot (1998), Stock and Wright (2000), Kleibergen (2002), Stock, Wright and Yogo (2002), Moreira (2003), Dufour and Taamouti (2005, 2007), Kleibergen (2005), Andrews, Moreira and Stock (2006), Andrews and Stock (2007), Chaudhuri, Richardson, Robins and Zivot (2010), Stock (2010), Chaudhuri and Zivot (2011), Guggenberger, Kleibergen, Mavroeidis and Chen (2012), Guerron-

spective and despite the breadth of supportive theoretical work, very little guidance is available regarding implementing projections.

While nonlinear global optimization methods such as the Genetic Algorithm [**GA**] and Simulated Annealing [**SA**] are well known in econometrics, their properties have not been analyzed as yet for test inversion and projection-based purposes.[2] Furthermore, the Particle Swarm Optimization [**PSO**] method, that may be traced back to Kennedy and Eberhart (1995), seems to have escaped notice in this field despite its popularity in *e.g.* engineering and computer science. This chapter introduces PSO to statistical functions with focus on test inversion and numerical projections.

Inverting a test, which produces a joint confidence set for intervening parameters, means assembling the parameter values that are not rejected by this test at a given level. The least rejected parameter value can be treated as a point estimate. Projecting a joint confidence region, which produces confidence intervals for individual parameters, entails finding the smallest and largest values of each parameter component within this region. The individual confidence sets can be unbounded or empty, which would suggest identification or specification problems, respectively. In general, this involves a complicated and often high-dimensional optimization problem, which is formally described in section 1.2 of this chapter. Due to the non-convexity of the underlying joint confidence region, traditional *e.g.* gradient-based techniques are not appropriate for this problem. Stochastic search methods will work, yet their efficiency in this case remains to be documented.

To concretize concepts, consider the problem of approximating $\boldsymbol{\theta} = \underset{\theta \in \Theta}{argmin}\,[f(\theta)]$ where $f(\theta)$ is the considered objective function and Θ is a search space, and denote

Quintana, Inoue and Kilian (2013), Qu (2014).

[2]There are other nonlinear global optimization packages such as Nonlinear Optimization with the Mesh Adaptive Direct search [**NOMAD**, see Audet and Dennis, Jr. (2006)] and Nonlinear Interior point Trust Region Optimization [**KNITRO**, see http://www.ziena.com/knitro.htm]. The former is an algorithm designed for black-box optimization while the latter is a software package specialized for nonlinear optimization with continuous variables in continuous functions.

by $\tilde{\theta}^l$ the retained choice for $\boldsymbol{\theta}$ at iteration l. Typically, $\tilde{\theta}^l$ is determined accepting or rejecting a *single* possible value of θ, which we denote by θ^l, and the latter is updated recursively given some start-up value until successive accepted values stabilize.

In contrast, PSO works via *a group* [the so-called swarm] of possible values of θ per iteration, denoted $\theta^{i,l}$, $i = 1, ..., M$ [the so-called particles]. For each iteration l, let $pbest^i = \text{argmin}\,[f(\theta)]$ over values of θ within the $\{\theta^{i,0}, \theta^{i,1}, ..., \theta^{i,l}\}$ path and let $gbest = \text{argmin}\,[f(\theta)]$ over values of θ within the $pbest^i$, $i = 1, ..., M$ group. Each $\theta^{i,l}$ is updated following a recursion of the form $\theta^{i,l+1} = \theta^{i,l} + v^{i,l+1}$ where the additive step term $v^{i,l+1}$ is itself recursively determined as a function of $v^{i,l}$ as well as the distance between $\theta^{i,l}$ and each of $pbest^i$ and $gbest$. Before proceeding to the next iteration, the $\theta^{i,l+1}$ values are reshuffled according to some criterion which may lead to updating the $pbest^i$ and $gbest$ values. Convergence is confirmed when all the $pbest^i$, $i = 1, ..., M$, and $gbest$ cease to change by some pre-specified margin from an iteration to another.

With reference to this process, SA imposes $M = 1$ and a random updating of θ^l while acceptance criteria are not exclusively determined by $f(\theta^l)$, as uphill moves are not always rejected. The GA works with $M > 1$, however updating is restricted to *reshuffles* [permutations of $\theta^{i,l}$, $i = 1, ..., M$] and/or *mutations* [permutations of the components of $\theta^{i,l}$].

In this context, this chapter has two contributions. First, we document the performance of PSO relative to SA and GA using the following well-known tests: (i) the Anderson and Rubin test, for univariate *i.i.d.* IV regressions, which we denote the **IAR** test [Anderson and Rubin (1949), Dufour (1997), Staiger and Stock (1997), Dufour and Taamouti (2005)], and (ii) its generalized multivariate and nonlinear counterpart relaxing the *i.i.d.* assumption [Stock and Wright (2000)], which we denote the [**GAR**] test. Projections based on the IAR test admit an analytical solution which provides a well-known benchmark. In this case, results reveal that the numerical PSO

solution is reached with remarkable speed and perfectly matches its analytical counterpart. The GAR projections with PSO are compared against a grid search method, as well as SA and GA. Overall, we find that PSO is at least as accurate as existing tools and converges dramatically faster.

Second, the chapter contributes to the literature on the New-Keynesian model. We analyze a three-equation New-Keynesian model with nine parameters, using the GAR-type test (described below), U.S. data and instruments from Dufour et al. (2013).[3] Given the model's dimension, we use a step length of 0.3 to execute a grid-search reference check. Although seemingly coarse, the considered step does not misrepresent related applied methods. For example, Castelnuovo and Fanelli (2014) rely on five-million randomly drawn points within a ten-parameter search space. The grid we consider evaluates 65.7 million points for our nine-parameter case. Results can be summarized as follows. First, on comparing the optima we reach via PSO to their counterparts, we find that GA and SA as well as the considered grid fail to *escape* local optima. Second, the confidence limits to which PSO converges to suggest severe under-identification. Several fundamental factors that may compromise identification in this context are surveyed by *e.g.* Cochrane (2011) and Schorfheide (2013). Our findings with PSO are thus not unexpected. For the purpose of this chapter, more important than our specific under-identification result is the consequence of failing to detect its extent via popular toolkits such as GA and SA. Indeed, GA and SA converge to local optima that suggest misleading economic decisions on: (i) the nature of the New-Keynesian Phillips Curve [**NKPC**], (ii) determinacy of monetary policies,

[3]Recent work with focus on identification challenges among others in this context include Fuhrer and Rudebusch (2004), Mavroeidis (2004, 2005), Benati (2008), Dufour et al. (2006, 2010, 2013), Nason and Smith (2008), Canova and Sala (2009), Kleibergen and Mavroeidis (2009), Iskrev (2010), Magnusson and Mavroeidis (2010), Mavroeidis (2010), Cochrane (2011), Inoue and Rossi (2011),Komunjer and Ng (2011), Guerron-Quintana et al. (2013), Andrews and Mikusheva (2014), Basturk, Cakmakli, Ceyhan and van Dijk (2014), Castelnuovo and Fanelli (2014) and the surveys by Schorfheide (2013) and Mavroeidis et al. (2014).

and (iii) the persistence of the Taylor rule. The grid search converges to equally disappointing optima. Our results suggest that far more attention needs to be paid to numerical precision as test inversion gains popularity in applied econometrics.

The chapter is organized as follows. In section 1.2, we describe the test inversion and projection method. In section 1.3, we introduce PSO and its variants. In section 1.4, we report the empirical analysis and summarize the main results. In section 1.5 we make a conclusion.

1.2 Test inversion and projection

1.2.1 General framework

Consider a right-tailed test $G(\theta, \overline{\theta}, \Upsilon) > \overline{S}_\alpha$ where

$$\theta = [\theta_1, ..., \theta_D]'$$

is a D-dimensional vector of parameters of interest; $\bar{\theta}$ is a vector of fixed, calibrated parameters or parameters that can be partialled out; Υ is the observed data; $\overline{S}_\alpha$ is a cut-off point for a significance level α. Without loss of generality, we consider the case where $\overline{S}_\alpha$ does not vary with θ. The method can be easily adapted to the case where the cut-off point is a function of θ. The associated null hypothesis is:

$$H_0 : \theta = \theta_0, \quad \theta_0 \text{ known.}$$

Then test inversion requires finding a set

$$C_\theta(\alpha) = \{\theta_0 : G(\theta_0, \overline{\theta}, \Upsilon) \leq \overline{S}_\alpha\} \tag{1.2.1}$$

that collects values of θ_0 that are not rejected by the considered test at the α level, that is such that

$$G(\theta_0, \overline{\theta}, \Upsilon) \leq \overline{S}_\alpha. \tag{1.2.2}$$

This can be done numerically given an economically meaningful search space Θ for θ.

One can conduct specification checks using a numerical optimizer such as GA, SA or PSO before numerical test inversion. Define:

$$G_{min}(\overline{\theta}, \Upsilon) = \underset{\theta_0 \in \Theta}{min}\ G(\theta_0, \overline{\theta}, \Upsilon). \tag{1.2.3}$$

If $G_{min}(\overline{\theta}, \Upsilon) > \overline{S}_\alpha$, then $C_\theta(\alpha)$ is empty. On the contrary,

$$G_{max}(\overline{\theta}, \Upsilon) = \underset{\theta_0 \in \Theta}{max}\ G(\theta_0, \overline{\theta}, \Upsilon). \tag{1.2.4}$$

If $G_{max}(\overline{\theta}, \Upsilon) < \overline{S}_\alpha$, then $C_\theta(\alpha)$ is unbounded.

Projecting the region $C_\theta(\alpha)$ means finding the smallest and largest values of each component of θ which lies in the region $C_\theta(\alpha)$. Therefore the objective is to minimize and maximize each θ_d, $d = 1, ..., D$, subject to an inequality constraint (1.2.2).

We transform the constrained optimization into an unconstrained one by introducing a properly defined penalty function. Generally, penalty functions are classified into two categories: stationary and non-stationary. A stationary penalty function use fixed penalty values throughout the minimization, whereas a non-stationary penalty changes dynamically and is a function of the iteration number [see Joines and Houck (1994), Michalewicz (1999), Liu and Lin (2007)]. A stationary penalty rather than a non-stationary penalty is chosen since we find the former provides better results than the latter for the considered projection problem. Consider the minimization problem

for example. In this case, the transformed objective function is:

$$\min_{\theta} \theta_d + pP(\theta), \tag{1.2.5}$$

$$P(\theta) = \varrho(Q(\theta))Q(\theta)^{\gamma(Q(\theta))}, \tag{1.2.6}$$

$$\varrho(Q(\theta)) = a \times (1 - e^{-Q(\theta)}) + c, \tag{1.2.7}$$

$$Q(\theta) = max(0, (G(\theta, \overline{\theta}, \Upsilon) - \overline{S}_{\alpha})), \tag{1.2.8}$$

where p is a fixed penalty value; $Q(\theta)$ measures the violation of the constraints; $\varrho(Q(\theta))$ is a continuous increasing function of $Q(\theta)$; and $\gamma(Q(\theta))$ is the strength of the violation.[4] As usual, a, c and p are problem-dependent. An illustrative analysis on the robustness of these tuning parameters are presented in section 1.4. By this design, parameter values that are rejected by the considered test are automatically discarded through iterations. Notice the minimization problem is independent for each $d = 1, ..., D$, that is, there is no communication among these minimization problems. For each minimization problem, the initial solution vector θ is randomly drawn from the parameter search set Θ, which is determined based on the ex-ante economic theory. Similarly, the transformed objective function of a maximization problem is:

$$\min_{\theta} \ -\theta_d + pP(\theta). \tag{1.2.9}$$

If one is interested in finding the confidence intervals of a function of θ, *e.g.* the impulse response function $IR(\theta)$, simply replace the first component of the above objective function θ_d with $IR(\theta)$.

[4]For presentation simplicity, we drop the subscript zero from θ_0 in equations (1.2.5) to (1.2.8). $G(\theta, \overline{\theta}, \Upsilon)$ in equation (1.2.8) is the same as $G(\theta_0, \overline{\theta}, \Upsilon)$ in equation (1.2.2). $\gamma(Q(\theta)) = 1$ if $Q(\theta) \leq 1$; $\gamma(Q(\theta)) = 2$ if $Q(\theta) > 1$.

1.2.2 Specific application

The empirical application uses the univariate IAR and multivariate GAR statistics which we present formally here. It has been proven that both IAR and GAR tests have correct size whether the instruments are weak or strong.

The univariate IAR statistic

Consider a simple linear simultaneous equations system [see Dufour and Taamouti (2005)]:

$$y = Y\beta + U, \tag{1.2.10}$$

$$Y = X\Pi + V, \tag{1.2.11}$$

where y is a $T \times 1$ vector of left-hand side endogenous variable; T is the sample size; Y is a $T \times D$ matrix of right-hand side endogenous variables; X is a $T \times k$ matrix of exogenous regressors (instruments); β and Π are $D \times 1$ and $k \times D$ matrices of unknown parameters; U and V are vectors of disturbances.

Under the assumption that the components of U are *i.i.d.*, to test $H_0 : \beta = \beta_0$ (β_0 known), the IAR test statistic is given by:

$$IAR(\beta_0, X, Y) = \frac{(y - Y\beta_0)'[I_T - M(X)](y - Y\beta_0)/k}{(y - Y\beta_0)'M(X)(y - Y\beta_0)/(T - k)}, \tag{1.2.12}$$

$$M(X) = I_T - X(X'X)^{-1}X', \tag{1.2.13}$$

where I_T is a $T \times T$ identity matrix. An analytical solution exists only if the components of U are *i.i.d.* normal so that $IAR(\beta_0, X, Y)$ follows an $F(k, T - k)$ null distribution. Dufour and Taamouti (2005) derive a closed form solution of constructing individual Confidence Sets [**DTCS**]. Therefore, to examine the accuracy and relative execution speed, we can compare numerical solutions with the DTCS solu-

tion. To match the general notation in section 1.2.1, here $\theta_0 = \beta_0$, $\Upsilon = [X, Y]$, $G(\theta_0, \overline{\theta}, \Upsilon) = IAR(\beta_0, X, Y)$ and the fixed cut-off point $\overline{S}_\alpha = F_\alpha(k, T-k)$.

The multivariate GAR statistic

Dufour et al. (2013) propose a Limited Information Maximum Likelihood [**LIML**] test (a special case of the GMM-type statistic proposed by Stock and Wright (2000)), which is based on the moment conditions that the structural errors of the n structural equations are orthogonal to instruments.[5] Following their notations, consider the multivariate regression of structural errors $\epsilon_{jt}(Y, \theta)$, $j = 1, ..., n$, on instruments Z_t

$$\epsilon_t(Y, \theta) = \Pi Z_t + V_t, \quad \epsilon_t(Y, \theta) = [\epsilon_{1t}(Y, \theta), ..., \epsilon_{nt}(Y, \theta)]', \quad Z_t = [Z_{1t}, ..., Z_{nt}]', \tag{1.2.14}$$

where Y is the observed data, V_t is a vector of white noise and θ is a vector of parameters of interest. Denote A_j to be a $k_j \times k$ selection matrix such that $Z_{jt} = A_j Z_t$, then given the true θ, we have

$$A_j \Pi_j = 0, \tag{1.2.15}$$

where Π_j is the jth row of Π. Consider $H_0 : \theta = \theta_0$ (θ_0 known) and regression (1.2.14) can be stacked as

$$\epsilon(\theta_0) = zb + v, \quad z = (I_n \otimes Z), \quad Z = [Z_1, ..., Z_T]', \tag{1.2.16}$$

where $\epsilon(\theta_0)$ is a $nT \times 1$ vector of structural errors evaluated at θ_0, v is a $nT \times 1$ vector that stacks V_t and $b = vec(\Pi')$. If H_0 is true, then $Ab = 0$. Since the right-hand-side regressors of regression (1.2.16) are exogenous, we can apply standard statistics. If v

[5] This serves as a good stress test of weak identification.

is *i.i.d.*, the GAR statistic has an approximate $F(m, n(T-k))$ null distribution even with weak instruments [Dufour et al. (2013)]:

$$GAR(\theta_0, Y, Z) = \frac{n(T-k)}{m} \frac{(A\hat{b}(\theta_0))'[A(\hat{\Sigma}_V(\theta_0) \otimes (Z'Z)^{-1})A']^{-1}(A\hat{b}(\theta_0))}{(\epsilon(\theta_0) - z\hat{b}(\theta_0))'(\hat{\Sigma}_V^{-1}(\theta_0) \otimes I_n)(\epsilon(\theta_0) - z\hat{b}(\theta_0))}. \quad (1.2.17)$$

where $\hat{b}(\theta_0)$ and $\hat{\Sigma}_V(\theta_0)$ are the unrestricted OLS and covariance estimator from regression (1.2.16). In this case $\Upsilon = [Y, Z]$, $G(\theta_0, \overline{\theta}, \Upsilon) = GAR(\theta_0, Y, Z)$; and hence $\overline{S}_\alpha = F_\alpha(m, n(T-k))$.

1.3 PSO algorithm and its variants

PSO, developed by Kennedy and Eberhart (1995), is an unconstrained global optimization method inspired by bird flocking. In contrast to SA, where the iteration starts with one candidate solution, PSO starts with a group of candidates. To simplify presentation, define: $\theta^{i,l}$, the D-dimensional solution vector, where i indexes the ith candidate and l indicates the current iteration number; $v^{i,l}$, the D-dimensional step length; $pbest^i$, the historical optimum for i; $gbest$, the historical optimum for the group. Suppose that we want to minimize the objective function $f(\theta^{i,l})$, M solution vectors $\theta^{i,0}, i = 1, ..., M$, are randomly initialized in a predetermined D-dimensional search space Θ. Θ can integrate ex-ante economic theories. The algorithm involves evaluating the objective function and updating the solution vectors. The algorithm keeps "updating" and "evaluating" until a stopping criterion (*e.g.* maximum iterations) is met. Specifically, in each iteration l, the objective function $f(\theta^{i,l})$ is evaluated for $i = 1, ..., M$, and the associated function values are recorded. Then each candidate solution vector and the associated step length are updated: the new step length $\theta^{i,l+1}$ is calculated recursively based on the current solution vector $\theta^{i,l}$ and new step length $v^{i,l+1}$; while $v^{i,l+1}$ is obtained recursively based on the current step length $v^{i,l}$, the

distance between $pbest^i$ and $\theta^{i,l}$, and the distance between $gbest$ and $\theta^{i,l}$. Eventually all solution vectors are moving towards the global optimum. The flow chart that summarizes the procedures for the updating and evaluation of the M solution vectors in M_l iterations is presented in appendix A.2.1.

1.3.1 The standard updating procedure

As outlined in Shi and Eberhart (1998), the standard updating procedures of the step length and solution vector for $i = 1, ..., M$ are:

$$v^{i,l+1} = w\left(v^{i,l}\right) + w_1(pbest^i - \theta^{i,l}) + w_2(gbest - \theta^{i,l}), \quad v^{i,l} \in [-v_{max}, v_{max}], \quad (1.3.1)$$

$$\theta^{i,l+1} = \theta^{i,l} + v^{i,l+1}, \quad (1.3.2)$$

where, w_1 and w_2 are two independent uniformly distributed random weights in the range $[0, 2]$; w is the weight on the previous step length.[6] w can be deterministic or random. We follow Lin (2011), who performs a mathematical analysis on the relationship among w, w_1 and w_2 and proposes a functional form for w, with which the stability and speed of convergence of PSO are increased.[7] The vector v_{max} can be adjusted ex-ante to control the boundaries of step lengths:

$$v_{max} = \psi \times (\theta_{max} - \theta_{min}), \quad (1.3.3)$$

where ψ is a fraction in $(0, 1)$ (We choose $\psi = 0.3$ by a robust analysis, which is presented in section 1.4). v_{max} plays an important role in PSO. If v_{max} is too high, the solution vector may by pass the global optimum; if v_{max} is too low, it may get stuck in the local optimum.

[6]The range $[0, 2]$ is proposed by Kennedy and Eberhart (1995) and is generally accepted in the computer science literature.

[7]The technical discussion on w is presented in appendix A.2.2.

1.3.2 The hybrid updating procedure

One drawback of PSO is that the fast rate of information exchange among solution vectors results in the creation of similar solution vectors that increases the possibility of being trapped in local optima. To prevent this premature convergence, Lei (2012) introduces the GA operator "crossover" into PSO and shows the hybrid updating procedure has the best performance among several variants of PSO.[8] The basic idea is that between iteration l and $l+1$, add an intermediate procedure that reshuffles the randomly selected solution vectors to create new solution vectors, and then reevaluate the new ones and update the optimum if a better one is found. The technical procedures are presented in appendix A.2.3.

1.3.3 Boundary constraints satisfaction

The updated solution vectors may leave the predetermined boundaries. A straight forward method is to assign a high (infinite) objective function value to the solution vectors that cross the boundary so that these individuals are skipped [Bratton and Kennedy (2007)]. An alternative method is setting the solution vector back onto the boundary if it exceeds the boundary. However, if only solution vectors are modified, according to equation (1.3.2), the solution vectors may again leave the feasible region in the next iteration [Helwig, Branke and Mostaghim (2013)]. We find that for the considered projection problem, modifying the step length to zero in combination with setting the solution vector back onto the boundary has the overall best performance:

$$if \quad \theta^{i,l} > \theta_{max}, \quad \theta^{i,l+1} = \theta_{max} \quad and \quad v^{i,l+1} = 0, \tag{1.3.4}$$

$$if \quad \theta^{i,l} < \theta_{min}, \quad \theta^{i,l+1} = \theta_{min} \quad and \quad v^{i,l+1} = 0. \tag{1.3.5}$$

[8] Crossover is a a genetic operator that takes more than one parent solution and produces a child solution.

The robustness analysis is presented in section 1.4. The intuition behind this is linked to the weak identification: since the confidence band is potentially unbounded, checking whether the boundary of the search set is the optimal solution or not may increase the efficiency of the algorithm. The updated step length $v^{i,l}$ may also leave the feasible region $[-v_{max}, v_{max}]$, then randomly assign a new step length within the boundaries:

$$if \quad v^{i,l} > v_{max}, \quad v^{i,l+1} = r_0 \times v_{max}, \tag{1.3.6}$$

$$if \quad v^{i,l} < -v_{max}, \quad v^{i,l+1} = -r_0 \times v_{max}, \tag{1.3.7}$$

where r_0 is a uniformly distributed random number in $[0, 1]$.

1.4 Empirical analysis

For our empirical analysis, we focus on a prototypical New-Keynesian model [see Clarida, Galí and Gertler (1999), Lindé (2005) and Dufour et al. (2013)]

$$\pi_t = \omega_f E_t \pi_{t+1} + \omega_b \pi_{t-1} + \gamma y_t + \varepsilon_{\pi t}, \tag{1.4.1}$$

$$y_t = \beta_f E_t y_{t+1} + \sum_{j=1}^{4} (1 - \beta_f)\, \beta_{yj} y_{t-j} - \frac{1}{\beta_r} (R_t - E_t \pi_{t+1}) + \varepsilon_{yt}, \tag{1.4.2}$$

$$R_t = \sum_{j=1}^{3} \rho_j R_{t-j} + (1 - \rho_{sum}) [\gamma_\pi \pi_t + \gamma_y y_t] + \varepsilon_{Rt}, \tag{1.4.3}$$

where, for $t = 1, ..., T$, π_t is the aggregate inflation, y_t is a driving variable, R_t is the nominal interest rate, $\rho_{sum} = \sum_{j=1}^{3} \rho_j$, and $[\varepsilon_{\pi t}, \varepsilon_{yt}, \varepsilon_{Rt}]'$ is a zero-mean disturbance with finite variances. Fundamental structures of this form have been extensively studied in macroeconomics and form the building block of many other more complex models. Equation (1.4.1) is a NKPC on which identification-robust test inversion based evidence is available by now; see *e.g.* Mavroeidis (2004, 2005), Dufour et al.

(2006, 2010, 2013), Nason and Smith (2008), Canova and Sala (2009), Kleibergen and Mavroeidis (2009), Basturk, Cakmakli, Ceyhan and van Dijk (2014) and the survey by Mavroeidis et al. (2014). The restriction

$$\omega_b = (1 - \omega_f) \tag{1.4.4}$$

is more or less maintained empirically in this literature. Although no consensus view prevails, when the NKPC is estimated as a single equation [see *e.g.* Galí and Gertler (1999)], y_t is often represented by marginal costs. Instead, the output gap is used as a driver from a systems perspective, as (1.4.2) is a typical Euler equation for output and (1.4.3) is a Taylor rule. On the latter, identification-robust test inversion evidence, although less prevalent than with the NKPC, has been linked to determinacy of monetary policies; see *e.g.* Mavroeidis (2010), Cochrane (2011), Inoue and Rossi (2011) and the survey by Schorfheide (2013). Output equations have also been examined and identification problems documented, *e.g.* by Fuhrer and Rudebusch (2004) and Magnusson and Mavroeidis (2010).

More generally, identification-robust methods applied to such models may be categorized into two broad classes. The first one is a full-information method, which relies on restrictions strong enough to allow the existence of a rational-expectations solution and thus treats the equations as a Dynamic Stochastic General Equilibrium (closed) system [Iskrev (2010), Komunjer and Ng (2011), Dufour et al. (2013), Guerron-Quintana et al. (2013), Andrews and Mikusheva (2014)]. The second one which we discussed in section 1.2.2 is a limited-information approach that relies on weaker assumptions even though it remains system-based [Magnusson and Mavroeidis (2010), Dufour et al. (2013)]. In response to recent works that challenge unique rational expectation solutions [in particular, Cochrane (2011), Dufour et al. (2013) and Schorfheide (2013)], we focus on limited-information IV-based methods.

We use the same data and instrument set as Dufour et al. (2013) who consider a quarterly sample for the U.S. from 1962Q1 to 2005Q3; however, we focus on a subsample analysis post 1985, to reflect instability issues raised, in particular, by Benati (2008), on the NKPC, and Mavroeidis (2010), on the Taylor rule. The GDP deflator and the Federal funds rate are used for the price level and the short-run interest rate respectively, and the real-time output gap introduced by Dufour et al. (2010, 2013), is considered to ensure that lags of y_t remain valid instruments. In our single equation analysis of the NKPC, we also use the marginal cost as a driving variable. Finally, as in Lindé (2005) and Dufour et al. (2013), all our data is demeaned prior to estimation. The instrument set consists of lag one of π_t, lag one to four of y_t and lag one to three of R_t, as well as lag two and three of both wage and commodity price inflation. This set provides an illustrative analysis consistent with the literature.

Before conducting the formal optimization, one need to pin down the tuning parameters in the penalty function and PSO. The best way is learning by doing the experiments. Table 1.1 reports the empirical impact of different choices of tuning parameters on the optimization results. We vary one tuning parameter and keep others constant. The objective is to find the lower bound for γ_π given the search set defined in table 1.4. We find that the optimization result is not sensitive to the changes of the tuning parameters in the penalty function (1.2.5)-(1.2.8). Therefore we select $a = 150$, $c = 10$, $p = 625$.[9] For the tuning parameter ψ in (1.3.3) that controls the scale of the maximum step length v_{max}: if it is too small, the solution is trapped in the local extremum; if it is too large, the solution may by pass the global extremum. Therefore we select $\psi = 0.3$.

To find out the best boundary-handling technique, we consider two candidate methods for the solution vector if it exceeds the boundary: i) set the solution vector

[9]These values are proposed by Liu and Lin (2007).

back onto the boundary, denoted as *boundary*; ii) set the solution vector randomly back into the feasible region, denoted as *random*; in combination with three methods for modifying the step length: i) set the step length to zero, denoted as -z; ii) set the step length randomly back into the feasible region, denoted as -r; iii) keep the step length unmodified, denoted as -u. Table 1.2 documents the empirical performance of all above combinations and also the method proposed by Bratton and Kennedy (2007), denoted as *infinity*. The objective is to find the lower bound for β_f given the search set defined in table 1.4. We find that the *boundary-z* performs best.

After tuning and selecting the boundary-handling technique, we first apply the unrestricted IAR test to (1.4.1) with a marginal cost driver. This test subsumes that the time t expectation error is uncorrelated with $\varepsilon_{\pi,t}$. To see this, represent (1.4.1) as an IV regression where expectations of inflation are replaced with actual values plus errors:

$$\pi_{t+1} = E_t\pi_{t+1} + \upsilon_{\pi,t+1}, \tag{1.4.5}$$

which leads to

$$\pi_t = \omega_f\pi_{t+1} + \omega_b\pi_{t-1} + \gamma y_t + \varepsilon^*_{\pi,t}, \tag{1.4.6}$$

$$\varepsilon^*_{\pi,t} = \left(\varepsilon_{\pi,t} - \omega_f\upsilon_{\pi,t+1}\right). \tag{1.4.7}$$

Even if $\varepsilon_{\pi,t}$, $\varepsilon_{y,t}$, and $\varepsilon_{R,t}$ are contemporaneously uncorrelated and *i.i.d.* across time, the compounded disturbances $\varepsilon^*_{\pi,t}$ can be autocorrelated at lag one, if $\varepsilon_{\pi,t}$ is contemporaneously correlated with $\upsilon_{\pi,t}$. The Anderson-Rubin test disregarding autocorrelation in $\varepsilon^*_{\pi,t}$ may thus be restrictive, yet we report it to "benchmark" our code since analytical formula for projections are available [Dufour and Taamouti (2005)] for this case. In addition, the LIML point estimator minimizes the inverted statistic and also

admits a well-known analytical solution.[10]

Table 1.3 reports these estimates and confidence sets at 5% level for γ and ω_f, projecting the inflation-lag out since with *i.i.d.* disturbances predetermined variables can be assumed exogenous.[11] In addition to analytical results, we derive point and set estimates completely ignoring their analytical counterparts, using SA, GA and PSO. All algorithms are coded in MATLAB and are tested on a Intel Core i7-4790 CPU with 3.6 GHz and 12 GB RAM machine, running 64-bit version of Windows. Convergence time is reported seconds. PSO and GA match the numerical solution up to four decimal points, while SA is accurate up to one decimal point. In execution speed, PSO is sixty-nine times faster than GA and 872 times faster than SA.

From the pure statistical perspective, the confidence sets are bounded which suggests the instruments are not uninformative. However, from the macroeconomic perspective, the confidence sets are quite wide which is not unexpected. We next perform a multivariate limited-information analysis. With the real-time output gap as the driving variable, we invert the multivariate GAR statistic in section (1.2.2). Results should be interpreted recalling the following comments.

Comment 1. Analytical benchmarks are unavailable. We apply SA, GA and PSO as well as a naive grid search method with a coarse step set to 0.3. This grid search is not intended as the solution of reference. Quite the opposite, we aim to document that unless grid searches are very fine which is often prohibitive as illustrated here, misleading solutions (detailed below) can be reached. Castelnuovo and Fanelli (2014) conduct numerical test inversion using five-million randomly drawn points within a ten-parameter search space. In our case, the coarse grid search evaluates 65.7 million

[10] Analytical solutions are available only for this case, since the underlying inequality takes the quadric form. Heteroscedasticity and Autocorrelation Consistent [**HAC**] weighting matrices which intervene in the test's generalized counterpart [Stock and Wright (2000)] complicate the inequality in a non-standard way, so the quadric form no longer applies.

[11] These results are conditional on the statistics.

points for a nine-parameter case.

Comment 2. Wider confidence intervals should *not* be interpreted as poor performance. Recall that we are assessing the optimizer rather than the statistical test here. By the definition of projections, a lower (upper) left (right) limit confirmed as a non-rejected value is an improved minimum (maximum). Given the available literature on the considered models, identification problems are expected and partly justify our focus on this application. Indeed, testing solvers when local optima are expected raises an interesting challenge.

Results are reported in Table 1.4. Despite its coarseness, the grid search is completed in 26047 seconds, while the slowest competing algorithm (SA here) converges in 8446 seconds. Given the dimension of the problem, a small refinement of the considered step would increase cost exponentially. Stochastic search algorithms are much more efficient, with PSO outperforming competing methods in both precision and speed.

Table 1.4 further confirms that different optimizers produce strikingly different estimation results. PSO returns the entire search set, which suggests that no parameter can be identified via the considered statistic and search set . This result is consistent with: (i) empirical conclusions of Mavroeidis (2010) on post-Volcker data, and (ii) theoretical "impossibility" arguments by Cochrane (2011) regarding determinacy. PSO thus delivers results corroborating these works, while competing methods suggest spurious identification. Before commenting on the latter, it is useful to briefly summarize the rationale of both Mavroeidis (2010) and Cochrane (2011) in support of weak identification; see also Schorfheide (2013).

From a methodological perspective, an *i.i.d.* assumption on error terms of equations such as the ones we study implies that lags of commonly used predictors are valid but possibly weak instruments; less restrictive assumptions complicate finite-

sample distributions of available statistics and may invalidate instruments. From a deep modeling perspective, identification problems suggest that policy did not necessarily satisfy the Taylor principle post Volcker and may have remained inactive instead. Since identification problems are thus expected ex-ante, precision problems with numerical projections are not unlikely. As emphasized above, our empirical analysis was motivated by such considerations. The precision differences we observe are nevertheless quite sizable and have important substantive implications. Indeed, in contrast with PSO, the GA, SA and the naive grid search algorithms identify some parameters, results we know are spurious given Comment 2 above. Specifically, the algorithms disagree on at least five key parameters.

1. On ω_f, SA suggests the parameter is greater than 0.5, which means that the inflation is more forward than backward looking. The grid search shows that zero can be refuted, while GA refutes extreme persistence. PSO confirms that such evidence over the post-Volcker period is inconclusive. Published conclusions on the nature of the NKPC [see e.g. Galí and Gertler (1999), Benati (2008) and Dufour et al. (2013)] need to be qualified in view of the numerical precision problems we document.

2. On γ, the grid search confirms that monetary policy has no effect on inflation. SA shows the effect is rather small. PSO and GA confirm that the effect of the forcing variable is in fact indeterminate. Schorfheide (2013) documents conflicting conclusions about the impact of considered forcing variables in this literature. We show that numerical issues may in fact exacerbate such conflicts.

3. On β_f, GA confirms that output is more backward than forward looking. PSO, SA and the grid search do not support this conclusion. Once again, spurious identification is suggested and may be attributed to imperfect optimization.

4. On the Taylor rule parameter ρ_{sum}, the grid search suggests high persistence; SA confirms some persistence while the GA refutes persistence; PSO disagrees with all these rejections. Given the importance of persistence in this literature, the conflict we find is quite noteworthy.

5. On the Taylor rule parameter γ_π, GA shows the parameter is greater than one, which matches the conventional assumptions on determinacy. The other algorithms provide no supportive evidence. Recall that a unique rational expectations solution was not imposed here, and as is well-known, the evidence conveyed via γ_π in favor or against determinacy is unlikely to be the whole story. Our results nevertheless assess the potential contribution of numerical problems to reported empirical conflicts.

On balance, our findings suggest that numerical problems cannot be ruled out when interpreting available empirical work and enduring conflicts in this literature. Fragility of empirical estimates may be linked, among others, to considered solvers.

1.5 Conclusion

This chapter introduces PSO to test inversion and numerical projections. A numerical comparison among PSO, GA and SA are conducted and the superiority of PSO for test inversion and related projection is documented. We show that PSO maintains stability and accuracy when the objective function is ill-behaved. Through the analysis on the New-Keynesian model, we show that proper decision on the significance of the key variables rely on an extremely fine grid search which can be numerically very costly. We recommend PSO for checking and building the projection-based confidence sets.

Bibliography

Anderson TW, Rubin H. 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* **20**: 46–63.

Andrews DWK, Moreira MJ, Stock JH. 2006. Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* **74**: 715–752.

Andrews DWK, Stock JH. 2007. Testing with many weak instruments. *Journal of Econometrics* **138**: 24–46.

Andrews I, Mikusheva A. 2014. Maximum likelihood inference in weakly identified DSGE models. *Quantitative Economics* : to appear.

Audet C, Dennis, Jr J. 2006. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization* **17**: 188–217.

Basturk N, Cakmakli C, Ceyhan P, van Dijk HK. 2014. Posterior-predictive evidence on US inflation using extended New Keynesian Phillips Curve models with non-filtered data. *Journal of Applied Econometrics* : to appear.

Benati L. 2008. Investigating inflation persistence across monetary regimes. *Quarterly Journal of Economics* **123**: 1005–1060.

Bratton D, Kennedy J. 2007. Defining a standard for particle swarm optimization. In *Proceedings of IEEE Swarm Intelligence Symposium*. 120–127.

Canova F, Sala L. 2009. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* **54**: 431–449.

Castelnuovo E, Fanelli L. 2014. Monetary policy indeterminacy and identification failures in the U.S.: Results from a robust test. *Journal of Applied Econometrics* : to appear.

Chaudhuri S, Richardson T, Robins J, Zivot E. 2010. A new split-sample score test in linear instrumental variables regression. *Econometric Theory* **26**: 1820–1837.

Chaudhuri S, Zivot E. 2011. A new method of projection-based inference in GMM with weakly identified nuisance parameters. *Jounal of Econometrics* **164**: 239–251.

Clarida R, Galí J, Gertler M. 1999. The science of monetary policy: A New Keynesian perspective. *Journal of Economic Literature* **37**: 1661–1707.

Cochrane JH. 2011. Determinacy and identification with Taylor rules. *Journal of Political Economy* **119**: 565–615.

Dufour JM. 1997. Some impossibility theorems in econometrics, with application to structural and dynamic models. *Econometrica* **65**: 1365–1389.

Dufour JM. 2003. identification, weak instruments and statistical inference in econometrics. *Canadian Journal of Economics* **36**: 767–808.

Dufour JM, Khalaf L, Kichian M. 2006. Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis. *Journal of Economic Dynamics and Control* **30**: 1707–1727.

Dufour JM, Khalaf L, Kichian M. 2010. On the precision of Calvo parameter estimates in structural NKPC models. *Journal of Economic Dynamics and Control* **34**: 1582–1595.

Dufour JM, Khalaf L, Kichian M. 2013. Identification-robust analysis of DSGE and structural macroeconomic models. *Journal of Monetary Economics* **60**: 340–350.

Dufour JM, Taamouti M. 2005. Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica* **73**: 1351–1365.

Dufour JM, Taamouti M. 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics* **139**: 133–153.

Fuhrer J, Rudebusch G. 2004. Estimating the Euler equation for output. *Journal of Monetary Economics* **51**: 1133–1153.

Galí J, Gertler M. 1999. Inflation dynamics: A structural econometric analysis. *Journal of Monetary Economics* **44**: 195–222.

Guerron-Quintana P, Inoue A, Kilian L. 2013. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics* **4**: 197–229.

Guggenberger P, Kleibergen F, Mavroeidis S, Chen L. 2012. On the asymptotic sizes of subset Anderson-Rubin and Lagrange multiplier tests in linear instrumental variables regression. *Econometrica* **80**: 2649–2666.

Helwig S, Branke J, Mostaghim S. 2013. Experimental analysis of bound handling techniques in particle swarm optimization. In *Proceedings of IEEE Transactions on Evolutionary Computation*. 259–271.

Inoue A, Rossi B. 2011. Testing for weak identification in possibly nonlinear models. *Journal of Econometrics* **161**: 246–261.

Iskrev N. 2010. Local identification in DSGE models. *Journal of Monetary Economics* **57**: 189–202.

Joines J, Houck C. 1994. On the use of non-stationary penalty functions to solve non-linear constrained optimization problems with GAs. In *Proceedings of the First IEEE International Conference on Evolutionary Computation*, volume 2. 579–584.

Kennedy J, Eberhart RC. 1995. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, volume 4. 1942–1948.

Kleibergen F. 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* **70**: 1781–1803.

Kleibergen F. 2005. Testing parameters in GMM without assuming that they are identified. *Econometrica* **73**: 1103–1123.

Kleibergen F, Mavroeidis S. 2009. Weak instrument robust tests in GMM and the New Keynesian Phillips Curve. *Journal of Business and Economic Statistics* **27**: 293–311.

Komunjer I, Ng S. 2011. Dynamic identification of DSGE models. *Econometrica* **79**: 1995–2032.

Lei X. 2012. *Swarm Intelligent Optimization Algorithms and Their Applications*. Science Press, Beijing. In Chinese.

Lin WX. 2011. Modified particle swarm optimization algorithm with fast convergence. *Journal of System Simulation* **24**: 83–94. In Chinese.

Lindé J. 2005. Estimating New-Keynesian Phillips Curves: A full information maximum likelihood approach. *Journal of Monetary Economics* **52**: 1135–1149.

Liu JL, Lin JH. 2007. Evolutionary computation of unconstrained and constrained problems using a novel momentum-type particle swarm optimization. *Engineering Optimization* **39**: 287–305.

Magnusson LM, Mavroeidis S. 2010. Identification-robust minimum distance estimation of the New Keynesian Phillips Curve. *Journal of Money, Credit and Banking* **42**: 465–481.

Mavroeidis S. 2004. Weak identification of forward-looking models in monetary economics. *Oxford Bulletin of Economics and Statistics* **66**: 609–635.

Mavroeidis S. 2005. Identification issues in forward-looking models estimated by GMM with an application to the Phillips curve. *Journal of Money, Credit and Banking* **37**: 421–449.

Mavroeidis S. 2010. Monetary policy rules and macroeconomic stability: Some new evidence. *American Economic Review* **100**: 491–503.

Mavroeidis S, Plagborg-Mollor M, Stock JH. 2014. Empirical evidence on inflation expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature* **52**: 124–188.

Michalewicz Z. 1999. *Genetic Algorithms + Data Structures = Evolution Program*. Springer, Berlin.

Moreira MJ. 2003. A conditional likelihood ratio test for structural models. *Econometrica* **71**: 1027–1048.

Nason J, Smith G. 2008. Identifying the New Keynesian Phillips Curve. *Journal of Applied Econometrics* **23**: 525–551.

Qu Z. 2014. Inference in DSGE models with possible weak identification. *Quantitative Economics* **5**: 457–494.

Schorfheide F. 2013. Estimation and evaluation of DSGE models: Progress and challenges. In Acemoglu D, Arrelano M, Deckel E (eds.) *Advances in Economics*

and Econometrics: Theory and Applications, Tenth World Congress. Cambridge University Press, Cambridge (UK), 184–230.

Shi Y, Eberhart RC. 1998. A modified particle swarm optimization. In *Proceedings of the IEEE International Conference on Evolutionary Computation*. 69–73.

Staiger D, Stock JH. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**: 557–586.

Stock JH. 2010. The other transformation in econometric practice: Robust tools for inference. *Journal of Economic Perspectives* **24**: 83–94.

Stock JH, Wright JH. 2000. GMM with weak identification. *Econometrica* **68**: 1097–1126.

Stock JH, Wright JH, Yogo M. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* **20**: 518–529.

Wang J, Zivot E. 1998. Inference on structural parameters in instrumental variable regression with weak instruments. *Econometrica* **66**: 1389–1404.

Table 1.1: Choice of tuning parameters

Tuning parameters in the penalty function (1.2.5)-(1.2.8)						
a	1	100	150	600	10000	100000
$f(\theta)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
c	0	5	10	100	10000	100000
$f(\theta)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p	1	100	625	2500	10000	1000000
$f(\theta)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tuning parameter that controls the scale of v_{max} in (1.3.3)						
ψ	0.01	0.1	0.3	0.5	0.8	0.9
$f(\theta)$	0.4462	0.0158	0.0000	0.0000	0.0000	0.0753

Note: The objective is to find the lower bound for γ_π given the search set defined in table 1.4.

Table 1.2: Different boundary-handling techniques

	$f(\theta)$
boundary-z:	0.000000
boundary-r	0.200607
boundary-u	0.212893
random-z	0.000013
random-r	0.000013
random-u	0.000028
infinity	0.805391

Note: The objective is to find the lower bound for β_f given the search set defined in table 1.4.

Table 1.3: Projection-based inference for equation (1.4.1) via inverting IAR

	DTCS	SA	GA	PSO
γ	[-0.0792,0.0710]	[-0.0776,0.0703]	[-0.0792,0.0710]	[-0.0792,0.0710]
ω_f	[0.1663,1.7735]	[0.1688,1.7349]	[0.1663,1.7735]	[0.1663,1.7735]
min-IAR	0.4635	0.4642	0.4635	0.4635
sec		1048	84	1.2

Note: Marginal cost measure is used as the driving variable, the structural restriction on ω_f is relaxed. The population size is 100 for GA and 80 for PSO.

Table 1.4: Projection-based multi-equation inference, post-Volcker sample

Coefficient	Search set	Grid Search	SA	GA	PSO
ω_f	[0,0.99]	[0.3,0.9]	[0.5116,0.9900]	[0.0000,0.8055]	[0.0000,0.9900]
γ	[0,0.99]	[0.0,0.0]	[0.0000,0.0131]	[0.0000,0.9900]	[0.0000,0.9900]
β_f	[0,1]	[0.0,0.9]	[0.0000,1.0000]	[0.0000,0.2544]	[0.0000,1.0000]
β_r	[1,30]	[1.3,29.8]	[2.0415,30.0000]	[1.5575,30.0000]	[1.0000,30.0000]
γ_π	[0,2.5]	[0.0,2.4]	[0.0000,2.5000]	[1.4846,2.5000]	[0.0000,2.5000]
γ_y	[0,1.5]	[0.0,0.3]	[0.0000,1.5000]	[0.0000,1.5000]	[0.0000,1.5000]
ρ_1	[0,2]	[0.9,1.8]	[0.0000,1.8515]	[0.0000,2.0000]	[0.0000,2.0000]
ρ_2	[-1,1]	[-1.0,0.2]	[-1.0000,0.0044]	[-1.0000,1.0000]	[-1.0000,1.0000]
ρ_{sum}	[0,1]	[0.9,0.9]	[0.3992,1.0000]	[0.0000,0.4251]	[0.0000,1.0000]
sec		26047	8446	1013	968

Note: The model (1.4.1)-(1.4.3) is estimated with the real-time output gap measure using the limited information method proposed by Dufour et al. (2013). The step for the naive grid search is 0.3. The population size is 100 for GA and 80 for PSO.

Chapter 2

Implication of Auxiliary Assumptions on the Identification of DSGE Models

2.1 Introduction

The estimation of Dynamic Stochastic General Equilibrium [**DSGE**] models suffer from non-identifiability [Canova and Sala (2009), Schorfheide (2013)], whereby the optimum point in the econometrician's objective function does not map uniquely to the true value for the deep parameter of interest.[1] As a result, estimated DSGE models are often uninformative for important parameters such as the slope coefficient of the Phillips curve, or the response coefficients in the monetary policy rule. At the same time, models with different policy implications may fit the data equally well [Del Negro and Schorfheide (2008)]. Problems in identification stem from the fact

[1] Under weak identification, a large subset of alternate values for a given deep parameter, alongside the true value, is equally consistent with the optimum point. Consequently, classical estimation tools such as maximum likelihood or generalized method of moments, may yield biased and inconsistent results, while the Bayesian posterior mode can no longer be considered a consistent estimator of the true parameter vector [Guerron-Quintana et al. (2013)].

that solutions of linearized DSGE models have a state-space representation that is a complicated and nonlinear function of underlying deep parameters. Identification of key model parameters might depend on true values of other parameters [Mavroeidis (2010)], or on arbitrary assumptions on exogenous shock processes that are auxiliary to the core model of interest [Schorfheide (2013)].

This chapter documents the sensitivity of the identification of a canonical three-equation DSGE model to auxiliary assumptions on shock structures using the identification robust estimation methodology of Dufour et al. (2013). Specifically, we compare two versions of the prototypical three-equation New-Keynesian DSGE model from Lindé (2005) that are identical in its core mechanism and parametrization, but differ in auxiliary assumptions on external propagation – one assumes that shocks to the output gap and interest rates follow an autoregressive framework, while the other allows for adhoc lags to the part of the model, and assumes errors to be white noises.

The Dufour et al. (2013) method derives confidence regions for parameters that are asymptotically valid even in the presence of weak identification by finding the set of parameter values that are not rejected by a certain test procedure. Weak identification then manifests itself as non-informative, possibly unbounded, confidence regions. Instead of calculating confidence regions, we simulate population data from two models, and calculate the frequency of rejection across the parameter space in the neighborhood of its null value. A low rejection frequency for a particular parameter away from the null (*e.g.* low power for the test) implies that different values for that parameter fit the data equally well. In other words, the parameter is weakly identified. In this sense, the procedure serves as a graphical indication of identification that complements recently proposed identification pre-tests [Iskrev (2010) and Komunjer and Ng (2011)] that provide a binary rank test for identification. Our method is, to some extent, informative about the degree of identification.

We confirm the well-known phenomenon that when the true parameter value of the slope coefficient of the New-Keynesian Phillips Curve [**NKPC**] is near zero, then a number of important parameters, such as the Taylor rule response coefficients, are weakly identified [Mavroeidis (2010)]. Additionally, we document that using lags instead of Autoregressive [**AR**] shocks to model persistence improves the identification of these parameters, even when the slope of the NKPC is near zero.

Another contribution of this chapter is in conducting simulation studies of the Dufour et al. (2013) procedure, which was not done in the original paper. We show that with a larger sample size, the power of the procedure improves, while the relative identification weaknesses of different parameters are preserved.

The remainder of the chapter is organized as follows. Section 2.2 describes the two models for comparison. Section 2.3 briefly describes the Dufour et al. (2013) statistic and the simulation setup. Section 2.4 describes the basic results and provides some intuition for the change in identification. Section 2.5 provides concluding comments and offers further research ideas.

2.2 Model

Modeling effort in the DSGE literature correctly focuses on the specification of the endogenous propagation mechanism, while putting less emphasis on the specification for exogenous propagation. While most DSGE models specify exogenous shocks as AR(1) processes, recent empirical work using Bayesian techniques has also considered Autoregressive Moving-Average [**ARMA**] shocks [Smets and Wouters (2007)] or higher-order autoregressive processes [Del Negro and Schorfheide (2009)]. In the growing literature concerned with estimating DSGE models, or parts thereof, in a classical setting, however, it is common to include lagged endogenous variables in an adhoc manner to remove autocorrelation in the error terms [Lindé (2005), Mavroeidis

(2010), Dufour et al. (2013)]. The implied underlying assumption is that as long as the core endogenous propagation mechanism remains true to the spirit of micro-founded models, specifications of the exogenous propagation are interchangeable, or less important.[2] From a closed form DSGE perspective, however, these assumptions are far from innocuous. Rather, lagged values of endogenous variables enter as state variables in DSGE models, with important implications for parameter identification.

We document the identification sensitivity of these assumptions by comparing two versions of the prototypical three-equation New-Keynesian DSGE model with the same core mechanism that appear in Lindé (2005), and is similar to Dufour et al. (2013), Mavroeidis (2010) and Clarida et al. (1999). The first model [**MAR**] treats exogenous shocks as AR(1) processes, and is given as:

$$\pi_t = \omega_f E_t \pi_{t+1} + (1-\omega_f)\,\pi_{t-1} + \gamma y_t + \varepsilon_{\pi,t}, \tag{2.2.1}$$

$$y_t = \beta_f E_t y_{t+1} - \frac{1}{\beta_r}\left(R_t - E_t \pi_{t+1}\right) + (1-\beta_f)\, y_{t-1} + u_{y,t}, \tag{2.2.2}$$

$$R_t = (1-\rho)\left[\gamma_\pi \pi_t + \gamma_y y_t\right] + \rho R_{t-1} + u_{R,t}, \tag{2.2.3}$$

where y_t is the output gap, R_t is the nominal interest rate, and π_t is inflation. The exogenous shocks, $u_{y,t}$, and $u_{R,t}$ follow autoregressive processes given by:

$$u_{y,t} = \rho_y u_{y,t-1} + \sigma_y \varepsilon_{y,t}, \tag{2.2.4}$$

$$u_{R,t} = \rho_R u_{R,t-1} + \sigma_R \varepsilon_{R,t}. \tag{2.2.5}$$

Here, $[\varepsilon_{\pi,t}, \varepsilon_{y,t}, \varepsilon_{R,t}]'$ are zero-mean Gaussian disturbances with the variance-covariance

[2]For example, Lindé (2005) estimates the NKPC with adhoc lags in a laboratory setup using simulated data from a model with AR(1) shocks.

matrix given by:

$$\Omega = \begin{bmatrix} \sigma_\pi^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_R^2 \end{bmatrix}.$$

Equation (2.2.1) is a hybrid version of the NKPC [Clarida et al. (1999)] where the parameters ω_f, and γ_y are non-linear functions of the degree of price stickiness, θ, the degree of price indexation or 'backwardness' in price setting, φ, and the discount factor, Λ, as follows:

$$\begin{aligned} \gamma_y &= \frac{(1-\theta)(1-\varphi)(1-\Lambda\theta)}{\theta + \varphi[1-\theta(1-\Lambda)]}, \\ \omega_f &= \frac{\Lambda\theta}{\theta + \varphi[1-\theta(1-\Lambda)]}. \end{aligned}$$

Equation (2.2.2) is a version of the IS curve, while equation (2.2.3) is the monetary policy reaction function, where the parameter ρ measures policy inertia.

The second model [**MLag**] adds lagged endogenous variables in equations (2.2.1) through (2.2.3) to capture the exogenous propagation mechanism inherent in (2.2.4) and (2.2.5) as follows:

$$\pi_t = \omega_f E_t \pi_{t+1} + (1-\omega_f)\pi_{t-1} + \gamma y_t + \varepsilon_{\pi,t}, \tag{2.2.6}$$

$$y_t = \beta_f E_t y_{t+1} - \frac{1}{\beta_r}(R_t - E_t\pi_{t+1}) + (1-\beta_f)\sum_{j=1}^{4}\beta_j y_{t-j} + \varepsilon_{y,t}, \tag{2.2.7}$$

$$R_t = \left(1 - \sum_{j=1}^{3}\rho_j\right)[\gamma_\pi \pi_t + \gamma_y y_t] + \sum_{j=1}^{3}\rho_j R_{t-j} + \varepsilon_{R,t}, \tag{2.2.8}$$

where the shocks $[\varepsilon_{\pi,t}, \varepsilon_{y,t}, \varepsilon_{R,t}]'$ are the same zero-mean disturbances with variance-covariance matrix Ω.

2.3 Estimation and simulation setup

We compare the identification properties of the two models above by generating rejection frequencies on simulated data in a laboratory setting using the identification robust methodology of Dufour et al. (2013). The methodology involves collecting parameter values that are not statistically significantly different from a null value under some test. Weak identification, in this case, manifests itself as wide, possibly unbounded, confidence regions. We generate simulated data under a common null value of the deep parameter vector from the two models MAR and MLag above, which serves as our population data. We then generate rejection frequencies around the neighborhood of the null values of the parameters. A low rejection frequency for a particular parameter away from the common null implies that different values of that parameter fit the data equally well. Under the same test and the same common true null value, a lower rejection frequency for a particular parameter in one model vis-à-vis the other is then indicative of weak identification.

In the remaining section, we briefly describe the test procedure, and our simulation setup. For details about the identification robust test inversion methodology used here, refer to Dufour et al. (2013).

Let the state-space representation of the DSGE model be given as a function of the deep parameter vector ϑ as follows:

$$X_t = A(\vartheta)X_{t-1} + B(\vartheta)V_t, \tag{2.3.1}$$

$$Y_t^\star = C(\vartheta)X_{t-1} + D(\vartheta)V_t, \tag{2.3.2}$$

$$V_t = H(\vartheta)V_{t-1} + \varepsilon_t, \tag{2.3.3}$$

where $Y_t^\star$ is a vector of observed variables, X_t is a vector of possibly unobserved state variables, V_t is an autoregressive process, and ϑ is a vector that collects all the deep

parameters of the model. Uhlig (1995), among others, provides a method of solving linearized DSGE models into the solution matrices $A(\cdot), B(\cdot), C(\cdot), D(\cdot)$ and $H(\cdot)$, and obtaining simulations, $\tilde{Y}_t$, for the observables from the model.

A solved DSGE given in (2.3.1) through (2.3.3) can be expressed as a Vector Autoregressive Moving-Average [**VARMA**] or infinite Vector Autoregression [**VAR**] model in the observables, whose coefficients are nonlinear functions of the deep parameters ϑ. This can in turn be approximated via the finite-order VAR:

$$\tilde{Y}_t \;=\; \sum_{j=1}^{p} \tilde{\Gamma}_j(\vartheta)\tilde{Y}_{t-j} + \epsilon_t$$

which serves as an auxiliary model representation of the DSGE in question. The Dufour et al. (2013) methodology involves collecting all parameter values ϑ, such that the distance between the auxiliary representation $\tilde{\Gamma}_j(\vartheta)$, and its data counterpart $\hat{\Gamma}_j$, estimated via OLS from the observed data, Y_t, is statistically insignificant under some test. The procedure therefore has an indirect inference interpretation [Smith (1993)].

For a given null value of the deep parameter, ϑ_0, the test statistic in question takes the Likelihood Ratio [**LR**] form, has an approximate F distribution under the null, and is calculated as follows [Dufour et al. (2013)]:

$$\begin{aligned}
\Lambda(\vartheta_0) &= \frac{|\hat{\Sigma}^0_W|}{|\hat{\Sigma}_W|}, \\
\mathcal{L}(\vartheta_0) &= \left(\frac{\mu\tau - 2\lambda}{Kn^\star}\right)\frac{1-(\Lambda(\vartheta_0))^\tau}{(\Lambda(\vartheta_0))^\tau}, \\
\mu &= (T-K) - \frac{n^\star - K + 1}{2}, \\
\lambda &= \frac{n^\star K - 2}{4}, \\
\tau &= \begin{cases} \left[\frac{K^2 n^{\star 2}-4}{K^2+n^{\star 2}-5}\right]^{1/2}, & if \quad K^2 + n^{\star 2} - 5 > 0; \\ 1, & otherwise, \end{cases}
\end{aligned}$$

where $n^\star$ is the dimension of the observed data Y_t, $K = n^\star p$, $\hat{\Sigma}_W^0$ and $\hat{\Sigma}_W$ summarizes the information from the auxiliary models $\tilde{\Gamma}_j(\vartheta_0)$, and $\hat{\Gamma}_j$, and are calculated as follows:

$$\begin{array}{rclcrcl} \hat{\Sigma}_W & = & \frac{1}{T-p}\sum_{t=p+1}^{T} W_t(Y)W_t'(Y) & ; & W_t(Y) & = & Y_t - \sum_{j=1}^{p}\hat{\Gamma}_j Y_{t-j}, \\ \hat{\Sigma}_W^0 & = & \frac{1}{T-p}\sum_{t=p+1}^{T} W_t(Y,\vartheta_0)W_t'(Y,\vartheta_0) & ; & W_t(Y,\vartheta_0) & = & Y_t - \sum_{j=1}^{p}\tilde{\Gamma}_j(\vartheta_0) Y_{t-j}. \end{array}$$

The rejection frequencies for the test is derived in the following steps:

1. For a given vector of parameter value ϑ_0, simulate one set of observable data, $\tilde{Y}_t^1$, using the DSGE model given in (2.3.1) through (2.3.3).

2. Treating $\tilde{Y}_t^1$ as population data, calculate the test statistic $\mathcal{L}^1$.

3. Taking each point in the parameter set $\left\{\vartheta^L < \vartheta_0 < \vartheta^U\right\}$ in the neighborhood of ϑ_0 as the true value, calculate a sequence of M statistics $\mathcal{L}(\vartheta)^m, m = 1, ..., M$ and compare them to the appropriate cutoff point in the F-distribution at the α-level.

4. The rejection frequency is calculated as the number of null rejections at the α-level as a proportion of M.

The population null values used in the simulation is summarized in table 2.1, and taken from Lindé (2005).

2.4 Results and intuition

Figure 2.1 shows the rejection frequencies of the LR test for the common parameters of both models at the 5% level of significance. The solid line represents the MLag model, while the dashed line represents the MAR model. The first panel in the figure

Table 2.1: Population parameter values for simulation

Parameter	MAR	MLag
ω_f	0.781	0.781
γ	0.002	0.002
β_f	0.373	0.373
β_r	28.57	28.57
γ_π	1.296	1.296
γ_y	0.417	0.417
$\rho = \sum_{j=1}^{3} \rho_j$	0.892	0.892
ρ_y	0.5	-
ρ_R	0.8	-
ρ_1	-	1.042
ρ_2	-	-0.357
ρ_3	-	0.207
$\beta_j, j = 1, 2, 3, 4$	-	0.25
σ_π	1.012	1.012
σ_y	0.333	0.333
σ_R	0.431	0.431

shows rejection frequencies for a sample size T of 175, while the second panel increases the sample size to 500.

The rejection frequencies for each parameter is calculated keeping the other parameters at their respective population values. Population values for the parameter are given in the titles of each sub-figure. For the point in the graph where the parameter value equal to the population value, both the alternative and the null value of the parameter is the same, and the rejection frequency gives the size of the test. At other points of the parameter space, the rejection frequency corresponds to the power of the test.

A number of observations can be made. First, note that the MLag model is better identified than the MAR model for all the common parameters. Second, the power of the test, as well as the size improves with T. This supports the asymptotic validity

of the Dufour et al. (2013) method.

Most importantly, we see that the intertemporal elasticity of substitution parameter β_r, and the monetary policy reaction function parameters γ_π and γ_y, are very weakly identified for the MAR case, but not for the MLag case. In face, the methodology cannot distinguish between any of the alternative parameter values for β_r and γ_y shown in the figure when the external propagation of the model is modeled as AR(1) shocks. This is consistent with the well-known result that when the slope of the NKPC, γ, is close to zero, then these parameters become weakly identified [Mavroeidis (2010)]. Figure 2.2 reproduces the results with a higher population value for γ. As we can see, a steeper slope for the NKPC implies significant improvement in identification for a number of parameters. Note, however, that the MLag model is able to convey important information about the true parameter space, even in the presence of a nearly-flat NKPC.

2.5 Conclusion

This chapter shows that different auxiliary assumptions have important impact on the identification of DSGE models. We find that adding lagged endogenous variables instead of assuming autoregressive shocks to capture external propagation improves the identification of important parameters, even when the slope of the NKPC is near zero. Our finding confirms the argument of Schorfheide (2013), page 23: "*the assumption that monetary policy errors are i.i.d. sequences provide identification in the sense that lagged inflation and output can serve as instrumental variables in the estimation of the policy rule coefficients. This source of identification vanishes if errors are allowed to be serially correlated.*" We also find that the asymptotic LR test is oversized in small sample, which provides motivation for finite-sample improvement in the third chapter.

Bibliography

Canova F, Sala L. 2009. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* **54**: 431–449.

Clarida R, Galí J, Gertler M. 1999. The science of monetary policy: A New Keynesian perspective. *Journal of Economic Literature* **37**: 1661–1707.

Del Negro M, Schorfheide F. 2008. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics* **55**: 1191–1208.

Del Negro M, Schorfheide F. 2009. Monetary policy analysis with potentially misspecified models. *American Economic Review* **99**: 1415–50.

Dufour JM, Khalaf L, Kichian M. 2013. Identification-robust analysis of DSGE and structural macroeconomic models. *Journal of Monetary Economics* **60**: 340–350.

Guerron-Quintana P, Inoue A, Kilian L. 2013. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics* **4**: 197–229.

Iskrev N. 2010. Local identification in DSGE models. *Journal of Monetary Economics* **57**: 189–202.

Komunjer I, Ng S. 2011. Dynamic identification of DSGE models. *Econometrica* **79**: 1995–2032.

Lindé J. 2005. Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics* **52**: 1135–1149.

Mavroeidis S. 2010. Monetary policy rules and macroeconomic stability: Some new evidence. *American Economic Review* **100**: 491–503.

Schorfheide F. 2013. Estimation and evaluation of DSGE models: Progress and challenges. In Acemoglu D, Arrelano M, Deckel E (eds.) *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*. Cambridge University Press, Cambridge (UK), 184–230.

Smets F, Wouters R. 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* **97**: 586–606.

Smith A. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**: 63–84.

Uhlig H. 1995. A toolkit for analyzing nonlinear dynamic stochastic models easily. *mimeo* .

Figure 2.1: Rejection frequencies when γ is small: the MAR (dashed line) and MLag (solid line) models

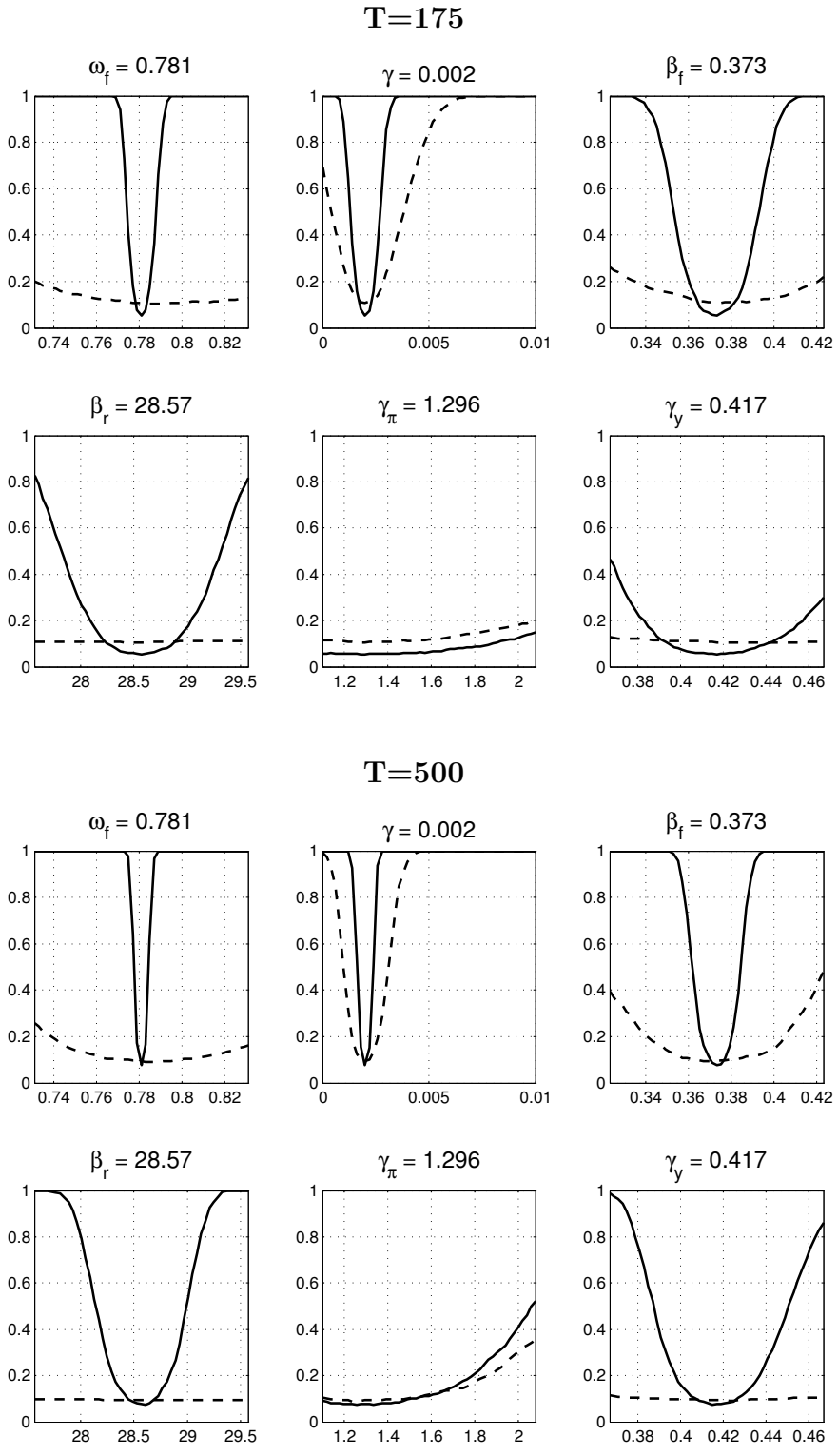
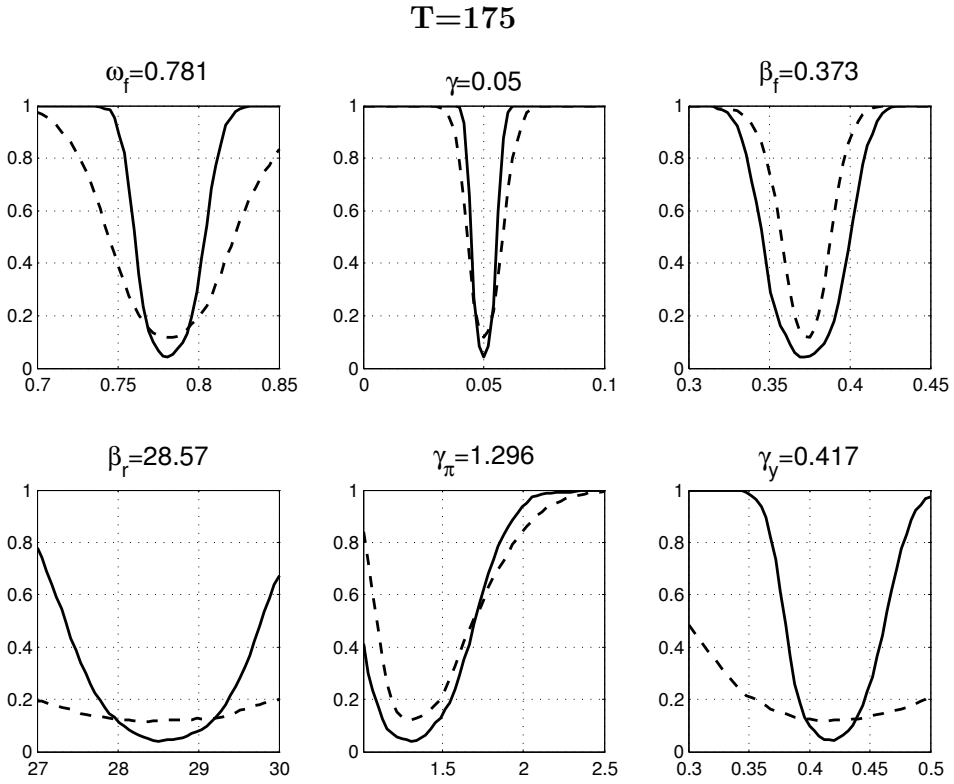


Figure 2.2: Rejection frequencies when γ is large: the MAR (dashed line) and MLag (solid line) models



Chapter 3

On Attempts to Rescue Identification in DSGE Models: A Finite-Sample Exact Analysis

3.1 Introduction

Estimated Dynamic Stochastic General Equilibrium [**DSGE**] models are now ubiquitous among academics and policy-makers interested in empirical macroeconomic research and quantitative policy analysis. Regardless of the estimation strategy, however, weak identification is pervasive in such models, making inference about the deep parameters a major concern [Canova and Sala (2009), Schorfheide (2013), Cochrane (2011)]. Under weak identification, many popular methods can deliver confidence bands or credible sets that are invalid even asymptotically [Guerron-Quintana, Inoue and Kilian (2013)], and thereby run the risk of misleading practitioners interested in determining policy implications. Such concerns have prompted efforts in the literature to understand the causes of identification failure, at least locally [Iskrev (2010),

Komunjer and Ng (2011)] and to develop identification-robust econometric methods that are valid whether identification restrictions hold or not [Andrews and Mikusheva (2013), Dufour, Khalaf and Kichian (2013), Guerron-Quintana et al. (2013) and Qu (2013)].

The literature has pointed out several pervasive "culprits" in the available models and data that undermine identification, including nonlinearities, and the properties of disturbances as well as of expectations [Schorfheide (2013)]. Cochrane (2011), on the other hand, suggests more fundamental "culprits" arising from determinacy, and suggests that the literature's "many attempts to rescue identification" may be theoretically bound to fail. In particular, Cochrane's analysis of determinacy within New-Keynesian models suggests that theory produces estimating equations that is "a snake pit for econometricians", so empirical work "must throw out important elements of the theory in order to identify parameters". The problems illustrated by Cochrane do not differ econometrically, or said differently, "symptom-wise", from the above "culprits". For example, usual exogenous or predetermined variables are either invalid, that is, cannot be orthogonal to persistent disturbances, or weakly informative, when *i.i.d.* disturbances are imposed. The central difficulty analyzed by Cochrane (2011) is that such "culprits" are not empirical concerns, but are hardwired theoretically. "Not only might these problems exist, but theory predicts that most of them do exist".

Arguments as in Cochrane (2011) may thus suggest that identification can only be salvaged by imposing theoretically unfounded assumptions. Are the current vintage of DSGE models therefore worth taking to the data? In this chapter we provide a finite-sample simulation-based analysis of such questions.

The econometric methods we focus on draw on two classes: (i) the above mentioned identification-robust methods and (ii) Indirect Inference [**InDInF**]. Methods of

the former class have recently been used to re-investigate the validity of deep parameters and their implications on important economic questions embedded in the structure of modern DSGE models, such as the degree of forward-looking behavior in New-Keynesian Phillips Curves [**NKPC**, Dufour, Khalaf and Kichian (2006), Mavroeidis (2005), Kleibergen and Mavroeidis (2009), Kapetanios, Khalaf and Marcellino (2015)], and the implications of monetary policy stance and determinacy implied by Taylor-type monetary policy rules [Mavroeidis (2010), Inoue and Rossi (2011), Castelnuovo and Fanelli (2015)]. Such methods typically deliver post-estimation information on weak-identification in the form of possibly unbounded confidence intervals, and in some cases like ours, information regarding lack of fit in the form of empty confidence sets.

InDInF, on the other hand, formally replaces complicated or intractable statistical functions by computer simulations.[1] More to the point from our perspective, Dridi, Guay and Renault (2007) suggest that calibrations, the common alternative to estimation and inference in macro-economics, may formally be captured via InDInF.[2]

However, to the best of our knowledge, all available identification-robust methods are justified using asymptotic statistical theory. Monte-Carlo experiments conducted to examine the performance of these methods often consider sample sizes that are beyond empirical relevance in macroeconomics [see *e.g.* Kleibergen and Mavroeidis (2009) and Magnusson and Mavroeidis (2014) who consider 1000 and 2000 observations]. In contrast, macroeconomic series, especially time series that can be modeled with stable structures, are typically short.[3] Furthermore, asymptotic methods may

[1]See Smith (1993), Gouriéroux, Monfort and Renault (1993), Gouriéroux and Monfort (1997) and Gallant and Tauchen (1996) for leading references.

[2]Furthermore, Gouriéroux, Phillips and Yu (2010) show that the method may control persistence based biases specifically in dynamic panels. Guay and Scaillet (2003) demonstrate advantages with unidentified nuisance parameters, and more recently, Calvet and Czellar (2015) revisit equilibrium models. See also Li (2010), Dominicy and Veredas (2013) and Fuleky and Zivot (2014).

[3]See *e.g.* the literature emphasizing pre and post-Volcker analyses of the Taylor rule, including Benati (2008) and the references therein for a historical perspective on instabilities with inflation

suffer from size problems due to the usual multivariate times-series issues, since solved DSGEs are (restricted) Vector Autoregressive Moving-Average [**VARMA**] processes. Whether approximated by finite VARs or not, VARMA processes are hard to estimate notwithstanding DSGE restrictions. Dynamics and the curse of dimensionality cause standard asymptotic approximations to perform poorly even in linear VARs.[4] These issues combined open the door to spurious over-rejections, and brings to question the substantive results obtained by available asymptotically justified identification-robust methods.

With regards to InDInF as well, confidence sets and hypotheses tests are typically justified using standard asymptotic arguments. For example, regularity conditions are derived to ensure consistency and asymptotic normality of the estimate, leading to Student-t type confidence intervals. Such approaches [*e.g.* Dufour and Taamouti (2007)] usually require identification, which suggests that despite promises, traditional InDInF may also suffer from the weak identification curse. To the best of our knowledge, provably finite-sample InDInF based methods are scarce.[5]

In view of the Cochrane critique, we thus ask whether the above cited successes with estimation and inference in some DSGEs are driven by small sample over-rejections. This chapter makes two contributions that address the issues above. First, we introduce a simulation-based confidence-set estimation method based on InDInF for DSGE models that provides size-control regardless of sample length. We need neither the existence of a limiting distribution, nor for errors to be exclusively Gaussian. All we require is the possibility of simulating the distribution of the relevant test statistic under the null hypothesis. Through a simulation study based on a three-

persistence.

[4]The dimensionality issue refers to the degree-of-freedom loss caused by the increase of the number of parameters/equations.

[5]Exceptions include Dufour and Valéry (2009) for stochastic volatility models, and Dufour and Kurz-Kim (2010) and Dufour, Khalaf and Beaulieu (2003), Beaulieu, Dufour and Khalaf (2013) in Multivariate regressions and simple location-scale models with fat-tailed fundamentals.

equation New-Keynesian model, we document problems with asymptotic methods and show that in contrast, out method has exact size and very good power for many key parameters.

On the empirical side, we revisit a canonical three-equation New-Keynesian sticky-price DSGE model [Woodford (2003), Giannoni and Woodford (2004), Milani and Treadwell (2012)] and estimate the deep parameters of the closed-form model imposing a unique, determinate rational-expectations solution using our size-corrected finite-sample exact method for the post-war U.S. economy, and focus on the Taylor rule parameters that have implications for determinacy. In contrast to Mavroeidis (2010), Dufour et al. (2013), and Castelnuovo and Fanelli (2015), we find that a unique, determinate version of the three-equation model is accepted by data from both the "Great Moderation" and "pre-Great Moderation" regimes, once we control for possible spurious over-rejections in small samples. Consistent with the analysis by Cochrane (2011), however, we find that identification is a major problem for most parameters of the model, and especially so for the coefficients in the Taylor rule. The only parameter that shows some identification can be mapped to a very flat slope for the NKPC, which has been argued to be at the heart of weak identification for the remaining model parameters from a mathematical, rather than statistical point of view [Mavroeidis (2010)]. Consequently, we find no evidence that the "pre-Great Moderation" era conforms to a passive monetary policy rule corresponding to an indeterminacy region in the three-equation model, as argued by Clarida, Galí and Gertler (2000), and supported by Mavroeidis (2010) using asymptotically valid identification-robust methods on single equations, and Castelnuovo and Fanelli (2015) using similar methods on closed-form DSGE models. Curiously, however, we find that the non-rejected parameter subset allows for the inflation-coefficient in the Taylor rule to be below one, while still meeting the determinacy conditions as formulated by Blanchard

and Kahn (1980).

As mentioned earlier, the only requirement we have is that the DSGE model can be simulated once a finite dimensional parameter, denoted ϑ in what follows, is specified. Note that calibration exercises that are popular in the DSGE literature depend on drawing such simulated samples from the model to derive population measures, including impulse responses, associated with specific choices for ϑ. Building on this property, we propose to invert statistics that test a specific value for ϑ, say ϑ_0, against an unrestricted relevant reduced form. An unrestricted VAR is considered as one of the most popular reduced forms. Inverting a test of $\vartheta = \vartheta_0$ at a given level $\alpha^\star$ consists in collecting, numerically or analytically, the ϑ_0 values that are not rejected using a considered test at a considered level. For example, given a right-tailed test statistic $S(\vartheta_0)$ with $\alpha^\star$-level cut-off point S_c, test inversion involves solving, over ϑ_0, the inequality $S(\vartheta_0) < S_c$. The solution of this inequality is a parameter space subset, denoted $\text{CS}(\vartheta_0; \alpha^\star)$, that satisfies the level constraint, that is: the probability that $\text{CS}(\vartheta_0; \alpha^\star)$ covers the true parameter value is greater than or equal to $1 - \alpha^\star$. Identification robustness in finite samples requires a convenient choice of S_c so that the latter level property holds for any given sample size regardless of identification. Exact critical values (S_c in our context) are hard to derive analytically for the models at hand. Furthermore, there is no reason to expect that a useful parameter-invariant critical point exists that achieves good power. In contrast, non-rejection can be assessed via a simulation based p-value, derived through the Monte-Carlo [**MC**] test method, which will provide an exact assessment of the critical point as ϑ_0 varies.

The statistic we consider evaluates the distance between the reduced-form parameters that summarize information in observed data, and their DSGE-restricted population counterpart, which we propose to obtain by simulation. Our method, which we denote as [**LRMC**] in what follows, focuses on a Monte-Carlo evalua-

tion of the LR-type criteria in multivariate regressions. We compare our method with the asymptotically justified method of Dufour et al. (2013), which we denote as [**LRAsy**] in what follows, that exploits the asymptotic F-based approximation of the LR-statistic in regressions with fixed covariates. This measure, derived from multivariate least-squares, implies that the determinant of the residual variance/covariance matrix efficiently summarizes the restricted and unrestricted statistical information on available data. The ratio of these determinants thus captures the distance under consideration. Rather than minimizing this distance which makes sense from a point-estimation perspective, we view model-to-data matching as a search for the *insignificant* distances. Said differently, we collect the set of parameter values for which the defined DSGE does not significantly differ from the considered reduced form (*e.g.* the unrestricted VAR). Our method is not restricted to the considered reduced form nor to the considered distance measure.

The remainder of the chapter is organized as follows. Section 3.2 describes the statistical framework and methodology. Section 3.3 establishes the finite-sample exactness our method, and demonstrates possible spurious rejections of an asymptotically justified method through a simulation study on a fully micro-founded, canonical three-equation New-Keynesian DSGE model. Section 3.4 takes the same model to post-war U.S. data and demonstrates identification inherent to the model and provides the substantive argument that there is no conclusive evidence that "pre-Great Moderation" U.S. data is compatible with a passive monetary policy regime that implies an indeterminate representation of the model. We argue that at the heart of the lack of identification is a flat NKPC. Section 3.5 provides concluding comments and offers further research ideas.

3.2 Framework and methodology

Consider the general state-space representation of a DSGE model:

$$X_t = A(\vartheta) X_{t-1} + B(\vartheta) V_t, \tag{3.2.1}$$

$$Y_t^\star = C(\vartheta) X_{t-1} + D(\vartheta) V_t, \tag{3.2.2}$$

$$V_t = H(\vartheta) V_{t-1} + \varepsilon_t, \tag{3.2.3}$$

where, for $t = 1, ..., T$, X_t is a vector of possibly unobserved state variables; $Y_t^\star$ is a vector of observed variables; V_t is an autoregressive process; ϑ is a vector of deep parameters in the log-linearized DSGE models; ε_t is a vector of the structural economic shocks satisfying $E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_t' = \Sigma$, and $E\varepsilon_t\varepsilon_{t-j}' = 0$ for $j \neq 0$. Σ is a positive definite matrix and denote by J the Cholesky decomposition of Σ such that $\Sigma = JJ'$. Collect the parameters of J in the vector θ. Therefore $\psi = [\vartheta', \theta']'$. Uhlig (1995) (among others) provides a method of obtaining the matrices $A(.), B(.), C(.), D(.)$ and $H(.)$ from log-linearized DSGE models with deep parameter ϑ. We further assume that observables compatible with (3.2.1)-(3.2.3) can be obtained by simulation for a given parameter value $\vartheta = \vartheta_0$. Our purpose is to derive a joint and simultaneous confidence set for ϑ, maintaining a full-information perspective.

A likelihood function may - or may not - be tractable in this context. We also do not require that the associated score is tractable. In particular, assumptions on the likelihood or the score as in Guerron-Quintana et al. (2013) and Andrews and Mikusheva (2013) are not required. All we need is the possibility of simulating data compatible with (3.2.1)-(3.2.3) imposing usual assumptions of the parameter space of ϑ. InDInF which exploits the fact simulations can be easily drawn from the model whether likelihoods and/or scores are regular or not, provides an interesting statistical

objective function in our context.[6]

3.2.1 Background

InDInF requires an underlying auxiliary model that "matches" the structural model and may be easily fit to available data. Matching does not presume that the auxiliary model is correctly specified. Instead, a binding function that is not necessarily tractable should exist that links the parameters of the auxiliary to the postulated model. A closed form for the binding function is not used explicitly and may hold in limits as well. For example, the well-know Yule-Walker equations provide a natural binding function for moving average models estimated via AR auxiliary regression. This same rationale suggests a VAR as an auxiliary model for DSGEs. Typically, a solved DSGE as in (3.2.1)-(3.2.3) can be expressed as a VARMA or infinite VAR model in the observables whose coefficients are nonlinear functions of ϑ, in which case a finite-order VAR provides a natural auxiliary regression.

For this same problem, Dufour et al. (2013) propose an inference method that does not require a closed likelihood yet respects full-information principles. The method builds on the underlying infinite VAR solution approximated via the finite-order VAR model.[7]

$$Y_t = \sum_{j=1}^{p} \Gamma_j(\vartheta) Y_{t-j} + u_t. \tag{3.2.4}$$

From there on, and for a given parameter value $\vartheta = \vartheta_0$, regress $Y_t - \sum_{j=1}^{p} \Gamma_j(\vartheta_0) Y_{t-j}$ on as many of the Y_{t-j} as implied by the structure. Then, $\vartheta = \vartheta_0$ implies that these regressors should be jointly insignificant. Testing this restriction involves a regular zero-restriction in a VAR, which evacuates identification concerns.

[6]Refer to Gouriéroux et al. (1993) and Smith (1993) for a general discussion of IndInF, and for Dridi et al. (2007) with regards to DSGEs.

[7]For further conditions on such approximations, see Fernández-Villaverde, Rubio-Ramrez, Sargent and Watson (2007), Ravenna (2007).

Note that the VAR may - or may not - be demeaned. Demeaning all variables corresponds to allowing for unrestricted constants in the model studied, that is, it allows to express explained variables in deviation with respect to (potentially non-zero) unknown equilibrium values. Constraints on equation constants are evacuated by demeaning and thus are not accounted for.

In traditional estimation methodology a point estimate is found first *e.g.* via maximum likelihood or InDInF and confidence intervals are then constructed, of the form {estimate $\pm$ standard error $\times$ critical point}. In contrast, Dufour et al. (2013) build a confidence region assembling the ϑ_0 that are not rejected by the above described test at a certain level $\alpha^\star$. Formally, this exercise is called "inverting" the proposed tests. In contrast to intervals, confidence sets so obtained can be unbounded. If objective functions are almost flat, most values in the parameter space would not be rejected, reflecting weak identification. The sets can also be empty, which implies that the structural model is rejected at the considered level. It is worth noting that Guerron-Quintana et al. (2013) and Andrews and Mikusheva (2013) also proceed via test inversion arguments (with more stringent assumptions on likelihoods and scores). This chapter addresses - among others, three important deficiencies with available methods: (i) the truncation order; (ii) the curse of dimensionality, and (iii) the fact that instruments are lags, which violates the fixed-regressor assumption required for most exact finite-sample multivariate methods.

Our methodology - from a general perspective - is not restricted to (3.2.4) as the auxiliary model. However, to set focus, and to compare results to existing identification-robust methods, we first maintain (3.2.4) as our "reduced form". In contrast to Dufour et al. (2013) who suggest that Heteroskedastic and Autocorrelation Consistent [**HAC**] procedures may be used to correct for the truncation lags, we use (3.2.4) so that the order p does not need to reflect a correct specification.

3.2.2 The methodology: a brief overview

We proceed by inverting a test statistic for the hypothesis that fixes $\vartheta = \vartheta_0$ to a known value

$$H_0\left(\vartheta_0\right) : \vartheta = \vartheta_0, \qquad \vartheta_0 \quad \text{known.} \tag{3.2.5}$$

A complete description of our methodology thus requires: (i) defining the test statistic we propose to invert, (ii) obtaining identification-robust p-values for this statistic, and (iii) characterizing the numerical inversion solution. Steps (i) - (iii) are discussed in what follows.

Formally, given a right-tail test $S(\vartheta_0)$ and setting the number of MC simulations to N so that $\alpha^\star(N+1)$ is an integer, we obtain MC p-values, denoted $p_N(\vartheta_0)$, such that for finite T and finite N,

$$\mathsf{P}\big[p_N(\vartheta_0) \leq \alpha^\star\big] = \alpha^\star.$$

$p_N(\vartheta_0)$ is formally defined in section 3.2.4. Inverting a test, which produces a joint confidence set for intervening parameters, means assembling the parameter values ϑ_0 that are not rejected by this test at a given level. For example, given a right-tail test statistic $S(\vartheta_0)$ and associated exact p-values $p_N(\vartheta_0)$ at $\alpha^\star$ level, we aim to collect the ϑ_0 for which $p_N(\vartheta_0) > \alpha^\star$. The set of parameters that satisfies this inequality is a joint confidence set with level no less than $1 - \alpha^\star$. The least rejected parameter values (associated with the highest p-value) can be treated as point estimates. Note, however, that by the very nature of weak identification, the least rejected parameter values can be a set, rather than a unique point.

Projecting a joint confidence region, which produces confidence intervals for individual parameters, entails finding the smallest and largest values of each parameter component within this region. The procedures and tools are provided in the first

chapter. Concretely, we iterate over ϑ_0 to obtain the minimum and maximum of the function $a'\vartheta_0$, where a is a selection matrix, such that $p_N(\vartheta_0) > \alpha^\star$; *e.g.* $a = [1, 0, ...0]'$ leads to a confidence interval on the first component of ϑ. We impose a stable and unique solution to the DSGE, as in Blanchard and Kahn (1980).

The shape of the joint confidence region can be highly irregular and non-convex, therefore the test inversion and related projection have to be conducted numerically. The individual confidence sets can be unbounded or empty, which would suggest identification or specification problems, respectively.

3.2.3 Test statistic

Let $\Gamma(\vartheta) = [\Gamma_1(\vartheta), ..., \Gamma_p(\vartheta)]$ and $Y_{t,p} = vec([Y_{t-1}, ..., Y_{t-p}])$. Rewrite model (3.2.4) as:

$$Y_t = \Gamma(\vartheta)Y_{t,p} + u_t, \tag{3.2.6}$$

where $\Gamma(\vartheta)$ is a $n^\star \times K$ matrix, $Y_{t,p}$ is a $K \times 1$ vector and u_t is a vector of white noises with covariance matrix Ω.[8] This VAR(p) process is the form of Seemingly Unrelated Regressions [**SUR**] model where each equation has the same explanatory variables. Hence $\Gamma(\vartheta)$ can be estimated equation-by-equation disregarding underlying restrictions using Ordinary Least Square [**OLS**] applied to (3.2.6) with the observed data. Let $\hat{\Gamma}$ denote the $n^\star \times K$ matrix of OLS coefficients for the $n^\star$ equations.

Next, obtain a population counterpart to $\hat{\Gamma}$, which we will denote $\Gamma(\vartheta_0)$. The following simulation-based algorithm is applied using the state-space DSGE model (3.2.1) to (3.2.3) imposing $\vartheta = \vartheta_0$.

1. Draw a sequence of random errors $\{\tilde{\varepsilon}_{t,m}\}_{t=1}^T$ from equation (3.2.3) imposing $\theta = \theta_0$. Suppose M such sequences are produced, so m indicates the number of

[8] Given the true process is VARMA, if p is mis-specified, it is possible that u_t is no longer a vector of white noises and Ω may depend on ϑ.

simulations.

2. For a given ϑ_0, generate simulated series $\{\tilde{Y}_{t,m}(\vartheta_0)\}_{t=1}^{T}$ using the state-space DSGE model (3.2.1) to (3.2.3).

3. Calculate the equation-by-equation OLS coefficients and collect them in $\tilde{\Gamma}_m(\vartheta_0)$ for each simulation and obtain the average for M paths:

$$\bar{\Gamma}(\vartheta_0) = \frac{1}{M}\sum_{m=1}^{M}\tilde{\Gamma}_m(\vartheta_0). \tag{3.2.7}$$

Based on model (3.2.6), $\hat{\Gamma}$ and $\bar{\Gamma}(\vartheta_0)$ with observed data, construct:

$$W_t(Y,\vartheta) = Y_t - \hat{\Gamma}Y_{t,p}, \quad \hat{\Sigma}_W = \frac{1}{T-p}\sum_{t=p+1}^{T} W_t(Y,\vartheta)W_t'(Y,\vartheta), \tag{3.2.8}$$

$$W_t(Y,\vartheta_0) = Y_t - \bar{\Gamma}(\vartheta_0)Y_{t,p}, \quad \hat{\Sigma}_W^0 = \frac{1}{T-p}\sum_{t=p+1}^{T} W_t(Y,\vartheta_0)W_t'(Y,\vartheta_0), \tag{3.2.9}$$

where $\hat{\Sigma}_W^0$ and $\hat{\Sigma}_W$ give the constrained (imposing $\vartheta = \vartheta_0$) and unconstrained sum of squared errors matrices. Discrepancies between $\hat{\Sigma}_W^0$ and $\hat{\Sigma}_W$ will serve to assess $H_0\,(\vartheta_0)$. We use

$$\Lambda(\vartheta_0) = |\hat{\Sigma}_W^0|/|\hat{\Sigma}_W| \tag{3.2.10}$$

to assess this discrepancy.

Note that simulations from the DSGE model (3.2.1) through (3.2.3) described above depend on imposing $\theta = \theta_0$. The following theorem (the proof is given in the appendix using the companion VAR(1) process in Canova (2007)) shows that the statistic we invert is invariant to the error variance, θ, provided that the shocks are *i.i.d.*, as assumed in the vast majority of the DSGE literature (with important exceptions, such as Smets and Wouters (2007)).

Theorem 1 *If $\Sigma = J'J$ is diagonal and invertable, then the null distribution of the statistic (3.2.10) does not depend on the volatility parameter vector θ (θ collects the diagonal elements of J).*

In this case, we can fix θ to any value θ_0. In the following sections, we impose Σ to be the identity matrix.

The statistic presented in (3.2.10) admits an F-based approximation under the null in regressions with fixed covariates that was used by Dufour et al. (2013). Since Dufour et al. (2013) did not provide supportive simulation evidence, it would be useful to assess this approximation to motivate our LRMC alternative. The F-approximation (LRAsy) proceeds as follows:

$$\begin{aligned}
\mathcal{L}(\vartheta_0) &= \left(\frac{\mu\tau - 2\lambda}{Kn^\star}\right)\frac{1-(\Lambda(\vartheta_0))^\tau}{(\Lambda(\vartheta_0))^\tau}, \qquad (3.2.11)\\
\mu &= (T-K) - \frac{n^\star - K + 1}{2},\\
\lambda &= \frac{n^\star K - 2}{4},\\
\tau &= \begin{cases} \left[\frac{K^2 n^{\star 2}-4}{K^2+n^{\star 2}-5}\right]^{1/2}, & if \quad K^2 + n^{\star 2} - 5 > 0;\\ 1, & otherwise, \end{cases}
\end{aligned}$$

where $\mathcal{L}(\vartheta_0)$ has an approximate $F\left(Kn^\star, \mu\tau - 2\lambda\right)$ null distribution at level $\alpha^\star$.

3.2.4 Finite-sample p-values

The null distribution of $\Lambda(\vartheta_0)$ can be easily simulated, which justifies the application of MC tests [Dufour (2006), Beaulieu et al. (2013), Khalaf and Peraza (2014a,b) and Khalaf and Saunders (2014)]. The general MC test methodology proceeds as follows.

Let S_0 denote the test statistic calculated from the observed data set; generate N replications $S_1, \ldots, S_N$ of the test statistic S so that $S_0, S_1, \ldots, S_N$ are exchange-

able. Given the latter series, compute $p_N(S_0)$ where

$$p_N(S_0) = \frac{NG_N(S_0) + 1}{N+1}, \quad G_N(S_0) = G_N[S_0; S(N)] = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(S_i \geq S_0), \tag{3.2.12}$$

$S(N) = [S_1, \ldots, S_N]'$ and $\mathbf{1}(C)$ is the indicator function associated with event C:

$$\mathbf{1}(C) = 1, \text{ if event } C \text{ holds;}$$
$$= 0, \text{ otherwise.}$$

In other words, $NG_N(S_0)$ is the number of simulated values greater than or equal to S_0. The MC critical region is: $p_N(S_0) \leq \alpha^\star$, $0 < \alpha^\star < 1$.

If the distribution of S is continuous and $\alpha^\star(N+1)$ is an integer, then

$$\mathsf{P}\left[p_N(S_0) \leq \alpha^\star\right] = \alpha^\star.$$

When applied to the above $\Lambda(\vartheta_0)$ statistic, the MC test technique can be summarized as follows.

1. Obtain the simulation-based estimates underlying the considered statistic. Specifically, we implement procedures described in section 3.2.3 to obtain $\bar{\Gamma}(\vartheta_0)$ given ϑ_0. This population measure is generated only once, so the following steps are conditional on $\bar{\Gamma}(\vartheta_0)$;

2. Applying (3.2.8)-(3.2.9) and (3.2.10) to the data, find the observed value of the considered test statistic;

3. Draw N *i.i.d.* samples of size T from the model (3.2.1)-(3.2.3) under $\vartheta = \vartheta_0$;

4. Using the same population measure derived in step 1, and applying (3.2.8)-(3.2.9) and (3.2.10) to the simulated data, obtain N simulated values for the

considered test statistic;

5. Compute a simulated p-value for the test statistic, using the rank of the observed statistic, relative to its simulated counterpart; see (3.2.12). The null hypothesis is rejected at level $\alpha^\star$ by the test considered if the MC p-value so obtained is less than or equal to $\alpha^\star$.

Because the above MC test procedure involves two levels of simulations (a first one to approximate the population measure, and a second one to get the test statistic), it is a two-stage MC test [see Beaulieu et al. (2013)]. It is important to emphasize a key step in the above algorithm: the observed and simulated statistics must rely on the same approximated population measure; in this way, the observed and simulated statistics are (by construction) exchangeable under the null hypothesis, which yields size control [see Dufour et al. (2003), Dufour (2006) and Beaulieu et al. (2013)]. The underlying simulations are non-independent but remain exchangeable, which is sufficient to ensure exactness as shown by Dufour (2006). In addition to minimizing noise which may affect power, using the same approximated population measure implies important execution cost savings. The fact that exchangeability is sufficient from a finite-sample perspective is worth pointing out here since the statistic we propose relies on just one preliminary simulation.

3.3 Simulation study

In this section, we illustrate the finite-sample properties of the LRMC test through a simulation study on a fully micro-founded, prototypical three-equation New-Keynesian DSGE model with inflation inertia and habit formation [Giannoni and Woodford (2004), Milani and Treadwell (2012)]. The primary objective of this exercise is to establish finite-sample exactness of our method. We do so by generating data from the

model under a null, or "true-value" of parameters, and analyze the rejection frequency of our method under alternative parameter values, and compare it to the same for an asymptotically based method, for which we pick the asymptotically F-distributed LRAsy statistic.

Another objective of this exercise is to demonstrate that the identification of the deep parameters varies depending on the "true value". To this effect, we compare the rejection frequencies of our method, as well as the Dufour et al. (2013) method, for the same model with two different null values. The version of the prototypical model presented in Milani and Treadwell (2012), and estimated using Bayesian techniques, proves particularly suitable for this purpose since it comes equipped with two empirically justifiable nulls for us to compare – one corresponding to the prior mean, and the other to the posterior median.

The model consists of a NKPC, an Euler-equation based modified IS function, a Taylor rule followed by the monetary authority, and a description of the exogenous shock processes, as given below:

$$\begin{aligned}
\pi_t &= \frac{\beta}{1+\beta\gamma}E_t\pi_{t+1} + \frac{\gamma}{1+\beta\gamma}\pi_{t-1} \\
&+ \frac{(1-\alpha)(1-\alpha\beta)}{\alpha(1+\beta\gamma)}\left(\omega y_t + \frac{\sigma^{-1}}{1-\phi}\left(y_t - \phi y_{t-1}\right)\right) + \mu_t, && (3.3.1)\\
y_t &= \frac{1}{1+\phi}E_t y_{t+1} + \frac{\phi}{1+\phi}y_{t-1} - \frac{\sigma(1-\phi)}{1+\phi}\left(R_t - E_t\pi_{t+1}\right) + g_t, && (3.3.2)\\
R_t &= \rho R_{t-1} + (1-\rho)\left[\chi_\pi \pi_t + \chi_y y_t\right] + \nu_t, && (3.3.3)\\
\mu_t &= \rho_\pi \mu_{t-1} + \varepsilon_{\pi,t}, && (3.3.4)\\
g_t &= \rho_y g_{t-1} + \varepsilon_{y,t}, && (3.3.5)\\
\nu_t &= \rho_R \nu_{t-1} + \varepsilon_{R,t}, && (3.3.6)
\end{aligned}$$

where the variable y denotes the output gap, π denotes inflation, and R denotes the

nominal interest rate. The parameter γ denotes the degree of price indexation to past inflation, while α represents the fraction of firms that cannot reoptimize their prices in a given period due to Calvo-type price rigidities. σ represents the consumer's intertemporal elasticity of substitution, ϕ denotes the degree of habit persistence in the consumer's utility function. Parameters χ_π and χ_y denote the monetary authorities' response to inflation and the output gap, respectively, while ρ represents the degree of monetary policy inertia. μ, g, and ν are exogenous disturbances to the marginal cost of production, the Euler function, and the systematic policy rule, and evolve according to a AR(1) process with autoregressive coefficients ρ_i, for $i = \mu, g, \nu$, and are affected by *i.i.d.* fundamental Gaussian shocks $\varepsilon_i \sim N(0, \sigma_i^2)$. In what follows, the two remaining parameters of the model, namely the discount factor, β, and the elasticity of marginal cost to income, ω is kept fixed at 0.99 and 2 respectively. A full derivation of the model from micro-foundations can be found in Milani and Treadwell (2012).

Note that many available studies of frequentist, or identification-robust estimation of three-equation DSGE models focus on "semi-structural" versions of the above model where complex non-linear expressions of deep-parameters are simplified to reduced form linear parameters [Lindé (2005), Benati and Surico (2009), Dufour et al. (2013), Mavroeidis et al. (2014), Castelnuovo and Fanelli (2015)], that can be easier to identify. However, in the spirit of Cochrane (2011), we stick to the micro-founded model to illustrate the identification problems inherent in versions where we do *not* "throw out important elements of the theory in order to identify parameters".[9]

Collect all parameters of interest in $\psi = [\vartheta', \theta']'$, where

$$\vartheta = [\phi, \sigma, \gamma, \alpha, \rho, \chi_\pi, \chi_y, \rho_\pi, \rho_y, \rho_R]', \theta = [\sigma_\pi, \sigma_y, \sigma_R]'. \tag{3.3.7}$$

[9]We have, however, done the simulation exercise on semi-structural models as well, and found that the main message of this chapter is preserved there as well.

As theorem 1 shows, if Σ is diagonal and invertible, the null distribution of the statistic that we propose to invert is invariant to Σ. In this case, one can set Σ to be the identity matrix for simplicity.

Figures 3.1 through 3.4 plot the rejection-frequency curves for the parameters for different sample sizes, and different null values at the 5% level.[10] In each subplot, the rejection frequency of a parameter at their null value (provided in the title with a subscript 0) gives the size of our test, while the rejection frequency of the parameter away from the null value gives the power. As long as they are produced under the same method, rejection-frequency curves can serve as a graphical device for ascertaining the depth of the identification problem inherent in the model and parameter value. If the rejection frequency of a parameter is low regardless of the distance from the "true value" from which the data is generated, we can say that the parameter is hard to identify.

With this in mind, we compare rejection frequencies for a sample size of 103, which corresponds roughly to the length of the "pre-Great Moderation" (1954:Q3 - 1984:Q4) and "Great Moderation" (1985:Q1 - 2006:Q4) subsamples considered in the literature, and in the following section of this chapter, and for a sample size of 175, which corresponds to the length of the full sample used by Dufour et al. (2013) (1962:Q1 - 2005:Q3). A number of observations can be made.

First, the LRMC has the correct size, regardless of sample length. For each of the figures presented, the rejection frequency of the LRMC at the null value, ϑ_0, is exactly 5%. In comparison, the LRAsy has a higher frequency of rejection, especially for small samples. This shows that spurious rejection is a real problem for asymptotically based methods. Comparing figures 3.1 with 3.2, and figures 3.3 with 3.4, however, we can see that for the larger sample size, the F-distributed LRAsy delivers admirable precision,

[10]The rejection-frequency curves are generated by varying the grid value of the parameters one at a time, while keeping all others at their null value.

while requiring considerably lower computational intensity.

Second, some parameters have serious identification problems. The Taylor rule parameters, regardless of null value or sample size, are hard to identify, followed by the AR parameters in the exogenous disturbances.

Third, the identification of the deep parameters vary depending on the "true value". Comparing figures 3.1 and 3.2 with figures 3.3 and 3.4, we see that rejection frequencies can vary for the same parameter, depending on the null. See *e.g.* the deterioration in the identification of the parameters ϕ and γ denoting the degree of habit formation and price indexation. Not only do the identification of the parameters change as their own "true value" changes, but changing the "true value" of one parameter can have serious effects on the identification of other parameters. As an example, note that higher values of the Calvo rigidity parameter, α, and habit persistence parameter, ϕ, corresponding to the second set of null values shown in figures 3.3 and 3.4, make the identification of the intertemporal elasticity of substitution parameter, σ, much harder, even though the "true value" of this parameter itself has not changed.

This graphically demonstrates the identification problems raised by Mavroeidis (2010) and Cochrane (2011) in a simulated environment. In particular, Mavroeidis (2010) analytically solves a purely forward looking version of the three-equation New-Keynesian DSGE model, and demonstrates that parameter identification is compromised when the slope of the NKPC is flat, or when the persistence of the exogenous disturbances are high. The second set of null values considered in figures 3.3 and 3.4 do exactly that: a higher value for the Calvo parameter, α, or the price indexation parameter, γ, flattens the slope of the NKPC, correspondingly making other parameters harder to identify. As a preamble to the next section, we note that the second set of null values are taken from the posterior medians derived by Milani and Treadwell

(2012) through Bayesian methods. The data, according to the results in Milani and Treadwell (2012), then, favors parameter values that foster lack of identification.

In conclusion, we demonstrate in this section that (i) asymptotically motivated identification-robust methods can generate spurious rejections for the model, particularly in small samples. (ii) If the true parameter value behind the data generation process implies a flat NKPC, identification of the remaining parameters become more challenging, particularly in small samples.

3.4 Empirical analysis

The results of the previous section suggests that available asymptotic methods might allow spurious rejections of the model parameters, or the entire model, especially for small samples. With that in mind, we now turn to take the model to post-war U.S. data and make our empirical contribution. We focus on the question of determinacy and identification before and during the period of the "Great Moderation" that has been extensively studied in the literature.

Stock and Watson (2003) documented that macroeconomic volatilities in the G7 countries, including the U.S. reduced markedly after 1984, calling this period the "Great Moderation". One possible explanation for this phenomenon popularized in the literature by Clarida et al. (2000) is that the U.S. monetary authority started reacting aggressively to changes in inflation in determining its short-term policy interest rate, following the leadership of Paul Volcker as Chairman of the Federal Reserve during the early 1980's period. In a New-Keynesian framework, such a switch from a passive to active monetary policy can bring an economy from an indeterminacy region, where movements in macroeconomic variables can occur due to sun-spot fluctuations, to a determinate region, where active monetary policy succeeds in anchoring private agents' inflation expectations, which in turn, induces stability in the economy. This

narrative has subsequently been supported by Lubik and Schorfheide (2004) using Bayesian methods, Boivin and Giannoni (2006), and Benati and Surico (2009) in a VAR setting, and Mavroeidis (2010), Inoue and Rossi (2011), and Castelnuovo and Fanelli (2015) in asymptotically valid identification-robust settings.[11]

We revisit this question in light of Cochrane (2011)'s critique that determinacy in New-Keynesian models can lead to lack of identification, and our findings from the previous section on the possibility of spurious rejections inherent in asymptotically justified methods. Specifically, we estimate the upper and lower-bounds of the 95% confidence sets of the "pre-Great Moderation" subsample of 1954Q3 through 1985Q4, and "Great Moderation" subsample of 1986Q1 through 2007Q4.

Three application details need to be discussed before we proceed further. First, we match the DSGE model, that deal in stationary variables, with first-differenced data using the measurement equations as follows, where the left hand side variables describe the data, and the right-hand-side variables denote their model-equivalent expressions.

$$\begin{aligned} \text{Output growth (annualized \%)} &= 400\left(y_t - y_{t-1} + \varepsilon_{y,t}\right), \\ \text{Inflation (annualized \%)} &= 400\left(\pi_t + \ln \pi_*\right), \\ \text{Interest rates (annualized \%)} &= 400\left(R_t + \ln R_*\right). \end{aligned}$$

The three series of data are downloaded via Haver Analytics. Note, however, that different methods of bridging raw data to their DSGE model counterparts can have important implications for the estimated parameters. In particular, Canova (2014) shows that one can reach drastically different conclusions while comparing Taylor rule coefficients that measure the aggressiveness of monetary policy stance in a similar

[11] In particular, Castelnuovo and Fanelli (2015) suggest that an empty set resulting from an asymptotically justified identification-robust method for the "pre-Great Moderation" subsample is a proof of indeterminacy in that period.

three-equation DSGE model in the two-subsample periods, depending on whether the data is first differenced, or de-trended data, or whether one allows for non-stationarities in the level data through bridge-equations. We leave a proper analysis of the sensitivity of data matching to outcomes in the identification-robust method for future research.

Second, we employ numerical tools introduced in the first chapter to find the upper and lower bounds through a projection method. In particular, we optimize an objective function that imposes a high penalty on the distance statistic we invert if (i) determinacy conditions are not met, or if (ii) the simulated p-value implies a rejection at the 5% level. The first chapter shows that the proposed algorithm provides a significantly more accurate result compared with *e.g.* grid search methods employed by Mavroeidis (2010) or Castelnuovo and Fanelli (2015).[12] We first search for the unconditional joint upper and lower bounds via projection. One possible problem with the projection method is that there is no guarantee that the upper and lower bounds contain a compact set of parameter values. In particular, there can be large areas within this set that are rejected under the null.

The results of the unconditional upper and lower bounds are given in Tables 3.1 and 3.2 for the "pre-Great Moderation" and "Great Moderation" period, respectively.[13] For comparison, we also provide the bounds generated by the Dufour et al. (2013) asymptotic LRAsy.

First, note that the determinate rational expectations version of the three-equation New-Keynesian DSGE Model is rejected by LRAsy for the "pre-Great Moderation" subsample, while it is accepted for the "Great Moderation" subsample. This matches

[12]Castelnuovo and Fanelli (2015) conduct their analysis on 5 million 9-dimensional points on a tight grid selected randomly, implying a non-zero chance of multiple draws of the same point. In contrast, in the first chapter we conduct an analysis with 65.7 million controlled non-random, and hence unique 9-dimensional points to show the limitations of grid method.

[13]These results are conditional on the statistics and search sets.

with the results found in Castelnuovo and Fanelli (2015) for a reduced-form semi-structural version of the model, who take this as evidence that the first subsample is incompatible with a determinacy region. However, correcting for small sample size, we find that the model is accepted for both subsamples. To be precise, the *determinate* version of the model is accepted for both subsamples. As such, we find no evidence that the first sub-sample suffers from determinacy problems – a result that is in stark contrast with Mavroeidis (2010) and Castelnuovo and Fanelli (2015).

Second, we find that the model suffers from serious identification problems in both subsamples. For both subsamples, the entire search set where we conduct our analysis is returned as accepted for every parameter. Nothing is identified in the data. In contrast, LRAsy provides minor identification at extreme points for a number of parameters. In an effort to find more information about the lack of identification in the finite-sample exact case, we propose the following: first we look for the point that is least rejected by the sample (*e.g.* has the highest p-value), allowing for the possibility that the highest p-value can correspond to a set of points, rather than a unique point. Note that the very essence of weak identification is that such "point-estimates" can be spuriously determined, and often misleading. None-the-less, we proceed to look for the upper and lower bound for each parameter, one at a time, while holding the others at their least-rejected values. These conditional upper and lower bound results are provided in Tables 3.3 and 3.4.

This brings us to our third finding: the only parameter that shows some identification is the Calvo parameter, α, which, in turn, makes for a flat NKPC. We know from the illustration by Mavroeidis (2010) and our results in the previous section, a flat NKPC makes identification difficult for the remaining parameters. This is consistent with the results from LRAsy in Table 3.2, as well as the Bayesian estimates in Milani and Treadwell (2012).

Finally, we find the curious result that the Taylor rule coefficient on inflation that is crucial in measuring policy stance, allows values that are below one. The lower bounds for this variable is around 0.67 for the "pre-Great Moderation" period, and around 0.86 for the "Great Moderation" period. Moreover, note that we have explicitly controlled for determinacy. We do not pass a candidate parameter vector to the next step if determinacy condition [Blanchard and Kahn (1980)] is not met. It turns out that the high parameter values of α allow for the determinacy conditions to be met for Taylor rule coefficient well below one. This serves as a reminder to practitioners that interpreting determinacy, or the Taylor principle itself, in terms of the singular focus on the value of the inflation-response parameter alone can be misleading.

3.5 Conclusion

Estimation of determinate, rational-expectations consistent, closed-form New Keynesian DSGE models run the risk of misinforming practitioners about appropriate parameter values due to pervasive problems related to lack of identification. Available identification-robust methods can provide results that are asymptotically valid under weak-identification, but may spuriously over-reject parameter-values, and entire models, when taken to data with the small samples concretely available in macroeconomics. Considerable effort in the literature on the determinacy or lack-thereof in New-Keynesian DSGE models *e.g.* has focused on the "pre-Great Moderation" subsample in the U.S., that may yield as little as 100 quarterly observations. To correct for possible spurious over-rejection in samples of such small length, this chapter proposes a simulation-based finite-sample exact confidence-set estimation method based on InDInF. Using our method, we show that contrary to the findings in the literature, both the "pre-Great Moderation" and "Great Moderation" periods are

equally consistent with a determinate region of a fully micro-founded three-equation New-Keynesian DSGE model. However, we find that identification of any parameter is severely challenged.

Statistical test outcomes in general cannot disentangle micro-founded identification difficulties from similar symptoms arising from shock dynamics or trend specifications. Nevertheless, the InDInF MC method we propose in this chapter may be generalized to facilitate such diagnostics. We used VAR-based metrics as measures of empirical fit, since such metrics provide a well-understood prototypical tool-kit. Whichever other metrics one chooses to match, we show how its "observed" value can be contrasted with its "model coherent" counterpart endogenously, that is, adjusting the latter as well as the resulting "critical distance" to the data generating parameters. Exactness is achieved this way via just two nested simulations, which opens up many promising research avenues.

Bibliography

Andrews I, Mikusheva A. 2013. Maximum likelihood inference in weakly identified DSGE models. *Quantitative Economics* : to appear.

Beaulieu MC, Dufour JM, Khalaf L. 2013. Exact confidence sets and goodness-of-fit methods. *Journal of Econometrics* : to appear.

Benati L. 2008. Investigating inflation persistence across monetary regimes. *Quarterly Journal of Economics* **123**: 1005–1060.

Benati L, Surico P. 2009. VAR analysis and the great moderation. *American Economic Review* **99**: 1636–52.

Blanchard OJ, Kahn CM. 1980. The solution of linear difference models under rational expectations. *Econometrica* **48**: 1305–11.

Boivin J, Giannoni MP. 2006. Has monetary policy become more effective? *Review of Economics and Statistics* **88**: 445–462.

Calvet LE, Czellar V. 2015. Through the looking glass: Indirect inference via simple equilibria. *Journal of Econometrics* **185**: 343 – 358.

Canova F. 2007. *Methods for Applied Macroeconomic Research*. Princeton University Press, 2nd edition.

Canova F. 2014. Bridging DSGE models and the raw data. *Journal of Monetary Economics* **67**: 1–15.

Canova F, Sala L. 2009. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* **54**: 431–449.

Castelnuovo E, Fanelli L. 2015. Monetary policy indeterminacy and identification failures in the U.S.: Results from a robust test. *Journal of Applied Econometrics* : to appear.

Clarida R, Galí J, Gertler M. 2000. Monetary policy rules and macroeconomic stability: Evidence and some theory. *Quarterly Journal of Economics* **115**: 147–180.

Cochrane JH. 2011. Determinacy and identification with Taylor rules. *Journal of Political Economy* **119**: 565 – 615.

Dominicy Y, Veredas D. 2013. The method of simulated quantiles. *Journal of Econometrics* **172**: 235–247.

Dridi R, Guay A, Renault E. 2007. Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics* **136**: 397–430.

Dufour JM. 2006. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. *Journal of Econometrics* **133**: 443–478.

Dufour JM, Khalaf L, Beaulieu MC. 2003. Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models. *Oxford Bulletin of Economics and Statistics* **65**: 891–906.

Dufour JM, Khalaf L, Kichian M. 2006. Inflation dynamics and the New Keynesian Phillips curve: An identification robust econometric analysis. *Journal of Economic Dynamics and Control* **30**: 1707–1727.

Dufour JM, Khalaf L, Kichian M. 2013. Identification-robust analysis of DSGE and structural macroeconomic models. *Journal of Monetary Economics* **60**: 340–350.

Dufour JM, Kurz-Kim JR. 2010. Exact inference and optimal invariant estimation for the stability parameter of symmetric [alpha]-stable distributions. *Journal of Empirical Finance* **17**: 180–194.

Dufour JM, Taamouti M. 2007. Further results on projection-based inference in IV regressions with weak, collinear or missing instruments. *Journal of Econometrics* **139**: 133–153.

Dufour JM, Valéry P. 2009. Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models. *Journal of Econometrics* **150**: 193–206.

Fernández-Villaverde J, Rubio-Ramrez JF, Sargent TJ, Watson MW. 2007. ABCs (and Ds) of understanding VARs. *American Economic Review* **97**: 1021–1026.

Fuleky P, Zivot E. 2014. Indirect inference based on the score. *Econometrics Journal* **17**: 383–393.

Gallant AR, Tauchen G. 1996. Which moments to match? *Econometric Theory* **12**: 657–681.

Giannoni M, Woodford M. 2004. Optimal Inflation-Targeting Rules. In *The Inflation-Targeting Debate*, NBER Chapters. National Bureau of Economic Research, Inc, 93–172.

Gouriéroux C, Monfort A. 1997. *Simulation-based Econometric Methods*. Oxford University Press.

Gouriéroux C, Monfort A, Renault E. 1993. Indirect inference. *Journal of Applied Econometrics* **8**: 85–118.

Gouriéroux C, Phillips PC, Yu J. 2010. Indirect inference for dynamic panel models. *Journal of Econometrics* **157**: 68–77.

Guay A, Scaillet O. 2003. Indirect inference, nuisance parameter, and threshold moving average models. *Journal of Business and Economic Statistics* **21**: 122–32.

Guerron-Quintana P, Inoue A, Kilian L. 2013. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics* **4**: 197–229.

Inoue A, Rossi B. 2011. Identifying the sources of instabilities in macroeconomic fluctuations. *Review of Economics and Statistics* **93**: 1186–1204.

Iskrev N. 2010. Local identification in DSGE models. *Journal of Monetary Economics* **57**: 189–202.

Kapetanios G, Khalaf L, Marcellino M. 2015. Factor based identification-robust inference in IV regressions. *Journal of Applied Econometrics* : to appear.

Khalaf L, Peraza B. 2014a. Finite sample simulation-based inference for moving average processes. *mimeo* .

Khalaf L, Peraza B. 2014b. Persistence robust simulation-based inference for autoregressive moving average processes. *mimeo* .

Khalaf L, Saunders CJ. 2014. Indirect set inference for dynamic panel models. *mimeo* .

Kleibergen F, Mavroeidis S. 2009. Weak instrument robust tests in GMM and the New Keynesian Phillips curve. *Journal of Business and Economic Statistics* **27**: 293–311.

Komunjer I, Ng S. 2011. Dynamic identification of DSGE models. *Econometrica* **79**: 1995–2032.

Li T. 2010. Indirect inference in structural econometric models. *Journal of Econometrics* **157**: 120 – 128.

Lindé J. 2005. Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics* **52**: 1135–1149.

Lubik TA, Schorfheide F. 2004. Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review* **94**: 190–217.

Magnusson LM, Mavroeidis S. 2014. Identification using stability restrictions. *Econometrica* : to appear.

Mavroeidis S. 2005. Identification issues in forward-looking models estimated by GMM with an application to the Phillips curve. *Journal of Money, Credit and Banking* **37**: 421–449.

Mavroeidis S. 2010. Monetary policy rules and macroeconomic stability: Some new evidence. *American Economic Review* **100**: 491–503.

Mavroeidis S, Plagborg-Mollor M, Stock JH. 2014. Empirical evidence on inflation expectations in the New Keynesian Phillips curve. *Journal of Economic Literature* : to appear.

Milani F, Treadwell J. 2012. The effects of monetary policy news and surprises. *Journal of Money, Credit and Banking* **44**: 1667–1692.

Qu Z. 2013. Inference in DSGE models with possible weak identification. *Quantitative Economics* : to appear.

Ravenna F. 2007. Vector autoregressions and reduced form representations of DSGE models. *Journal of Monetary Economics* **54**: 2048–2064.

Schorfheide F. 2013. Estimation and evaluation of DSGE models: Progress and challenges. In Acemoglu D, Arrelano M, Deckel E (eds.) *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*. Cambridge University Press, Cambridge (UK), 184–230.

Smets F, Wouters R. 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review* **97**: 586–606.

Smith A. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**: 63–84.

Stock JH, Watson MW. 2003. Has the business cycle changed? *Proceedings - Economic Policy Symposium - Jackson Hole* : 9–56.

Uhlig H. 1995. A toolkit for analyzing nonlinear dynamic stochastic models easily. *mimeo* .

Woodford M. 2003. *Interest and prices*. Princeton University Press.

Figure 3.1: Rejection frequency: the LRAsy (dashed line) and LRMC (solid line) tests at level 5%, taking the prior mean from Milani and Treadwell (2012) as null, $T = 103$, $p = 4$

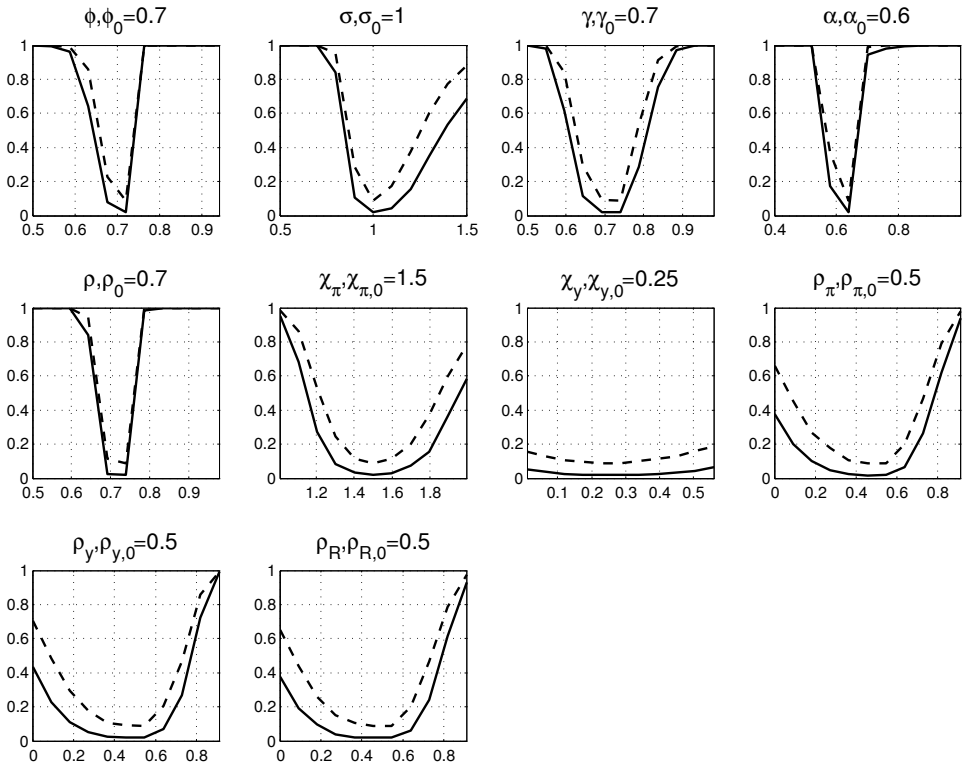


Figure 3.2: Rejection frequency: the LRAsy (dashed line) and LRMC (solid line) tests at level 5%, taking the prior mean from Milani and Treadwell (2012) as null, $T = 175$, $p = 4$

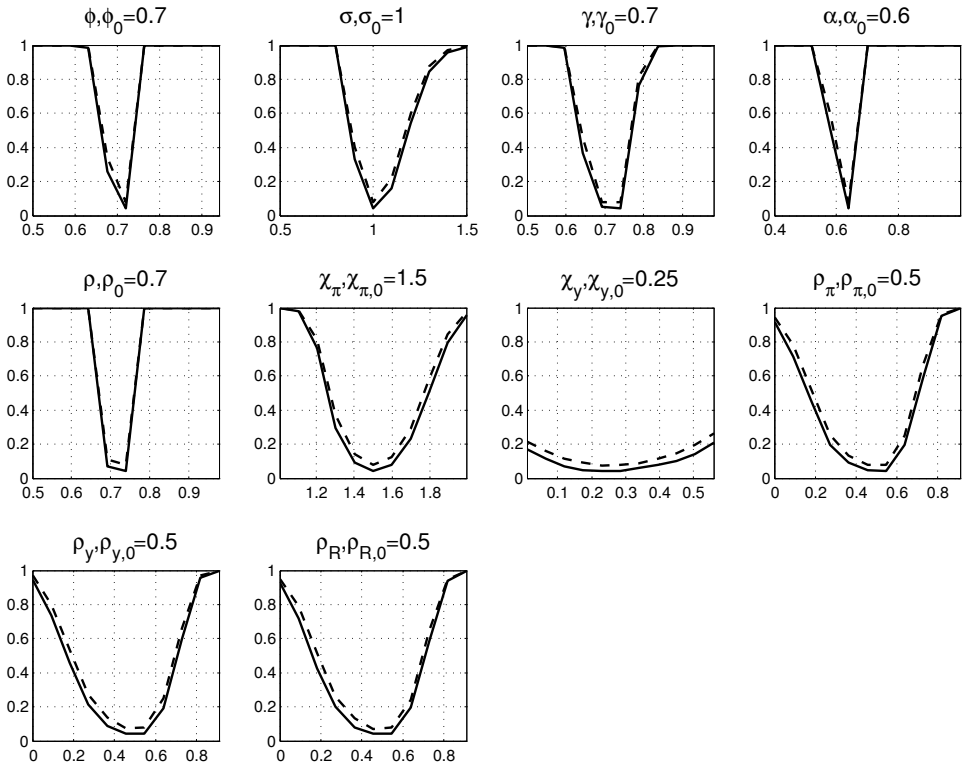


Figure 3.3: Rejection frequency: the LRAsy (dashed line) and LRMC (solid line) tests at level 5%, taking the posterior median from Milani and Treadwell (2012) as null, $T = 103$, $p = 4$

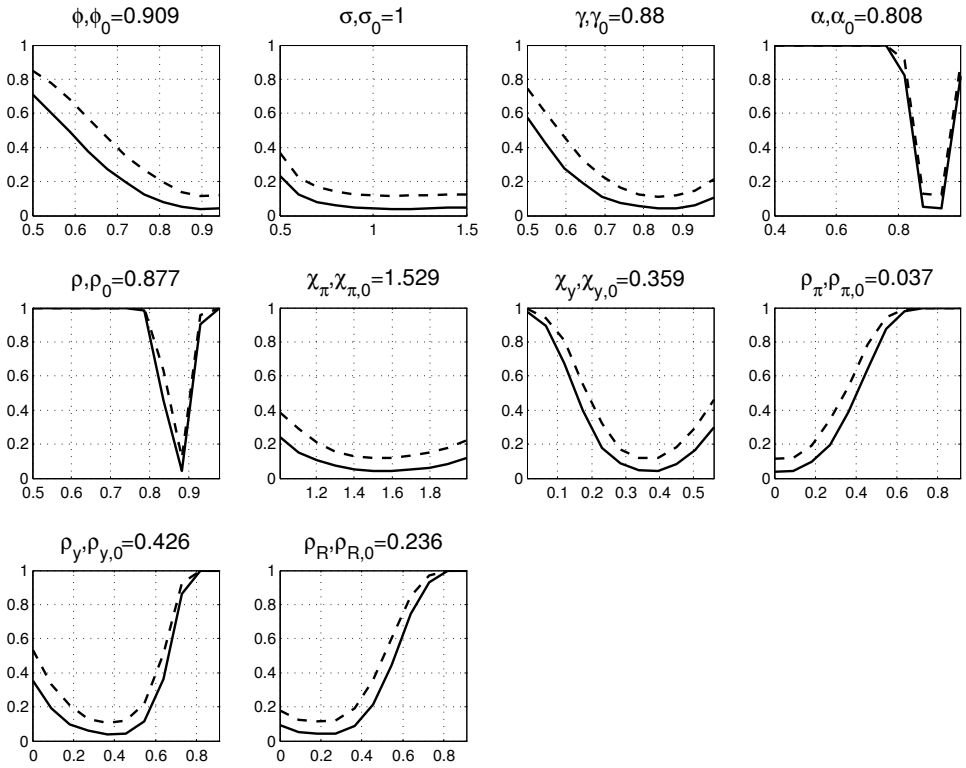


Figure 3.4: Rejection frequency: the LRAsy (dashed line) and LRMC (solid line) tests at level 5%, taking the posterior median from Milani and Treadwell (2012) as null, $T = 175$, $p = 4$

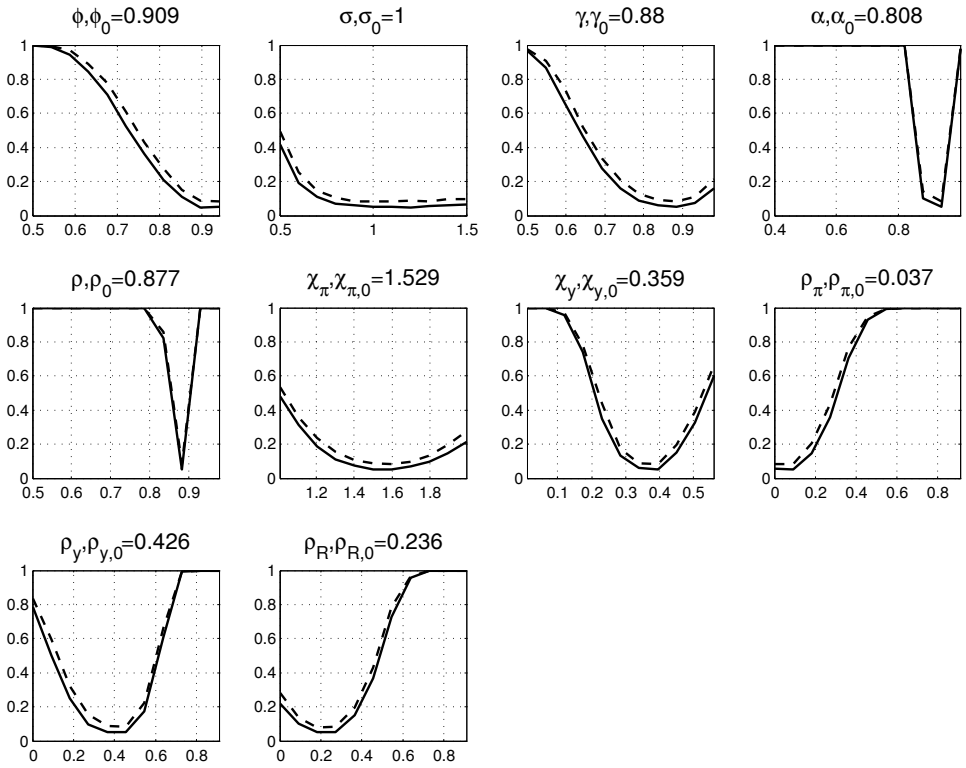


Table 3.1: Projection-based 95% confidence sets, "pre-Great Moderation" data

Coefficient	Search set	LRMC	LRAsy
γ	[0, 1]	[0.000000, 1.000000]	$\emptyset$
α	[0, 1]	[0.000000, 1.000000]	$\emptyset$
σ	[0.5, 1.5]	[0.500000, 1.500000]	$\emptyset$
ϕ	[0, 1]	[0.000000, 0.999999]	$\emptyset$
ρ	[0, 1]	[0.000000, 0.989504]	$\emptyset$
χ_π	[0, 2]	[0.000000, 2.000000]	$\emptyset$
χ_y	[0, 1]	[0.000000, 1.000000]	$\emptyset$
ρ_π	[0, 1]	[0.000000, 1.000000]	$\emptyset$
ρ_y	[0, 1]	[0.000000, 1.000000]	$\emptyset$
ρ_R	[0, 1]	[0.000000, 1.000000]	$\emptyset$

Note: LRMC refers to the finite-sample exact Monte Carlo method, and LRAsy refers to the Dufour et al. (2013) asymptotically F-distributed method.

Table 3.2: Projection-based 95% confidence sets, "Great Moderation" data

Coefficient	Search set	LRMC	LRAsy
γ	[0, 1]	[0.000000, 1.000000]	[0.000000, 1.000000]
α	[0, 1]	[0.000000, 1.000000]	[0.722394, 1.000000]
σ	[0.5, 1.5]	[0.500000, 1.500000]	[0.500000, 1.500000]
ϕ	[0, 1]	[0.000000, 1.000000]	[0.049264, 1.000000]
ρ	[0, 1]	[0.000000, 0.994594]	[0.313304, 0.984394]
χ_π	[0, 2]	[0.000000, 2.000000]	[0.000000, 2.000000]
χ_y	[0, 1]	[0.070217, 1.000000]	[0.087996, 1.000000]
ρ_π	[0, 1]	[0.000000, 1.000000]	[0.000000, 0.908576]
ρ_y	[0, 1]	[0.000000, 1.000000]	[0.043373, 1.000000]
ρ_R	[0, 1]	[0.000000, 1.000000]	[0.229440, 1.000000]

Note: LRMC refers to the finite-sample exact Monte Carlo method, and LRAsy refers to the Dufour et al. (2013) asymptotically F-distributed method.

Table 3.3: Projection-based 95% confidence sets for conditional on least rejected point, "pre-Great Moderation" data

Coefficient	Search set	Least-rejected value	95% confidence interval
γ	[0.01, 0.99]	0.984004	[0.010000, 0.990000]
α	[0.01, 0.99]	0.989501	[0.607158, 0.990000]
σ	[0.20, 5.50]	3.445347	[0.200000, 5.500000]
ϕ	[0.01, 0.99]	0.492969	[0.010000, 0.990000]
ρ	[0.01, 0.99]	0.856695	[0.010000, 0.974849]
χ_π	[0.50, 3.00]	1.026900	[0.677229, 2.482081]
χ_y	[0.01, 1.00]	1.000000	[0.001000, 1.000000]
ρ_π	[0.01, 0.99]	0.990000	[0.883021, 0.990000]
ρ_y	[0.01, 0.99]	0.272322	[0.010000, 0.990000]
ρ_R	[0.01, 0.99]	0.010275	[0.010000, 0.990000]

Table 3.4: Projection-based 95% confidence sets for conditional on least rejected point, "Great Moderation" data

Coefficient	Search set	Least-rejected value	95% confidence interval
γ	[0.01, 0.99]	0.981321	[0.120312, 0.982424]
α	[0.01, 0.99]	0.990000	[0.989589, 0.990000]
σ	[0.20, 5.50]	0.740003	[0.200000, 5.500000]
ϕ	[0.01, 0.99]	0.599157	[0.010000, 0.990000]
ρ	[0.01, 0.99]	0.692894	[0.010000, 0.990000]
χ_π	[0.50, 3.00]	0.881122	[0.873660, 3.000000]
χ_y	[0.01, 1.00]	0.455642	[0.428731, 1.000000]
ρ_π	[0.01, 0.99]	0.989014	[0.918695, 0.990000]
ρ_y	[0.01, 0.99]	0.386604	[0.010000, 0.990000]
ρ_R	[0.01, 0.99]	0.010000	[0.010000, 0.990000]

Chapter 4

Exact Inference in DSGE Models: Leads and Lags, Different Statistics

4.1 Introduction

This chapter extends the finite-sample exact simulation-based confidence set estimation method proposed in the third chapter beyond the multivariate Likelihood Ratio [**LR**] distance measure and approximate Vector Autoregression [**VAR**] context for Dynamic Stochastic General Equilibrium [**DSGE**] models. Assessing estimation uncertainty caused by weak identification is a major concern in these models, regardless of the estimation strategy. As a result, available empirical works on key and popular models have not yet produced a serious consensus view; see *e.g.* the large scale recent survey by Mavroeidis, Plagborg-Mollor and Stock (2014) on the New-Keynesian Phillips Curve [**NKPC**].

Progresses in econometrics have been made regards to weak identification. Rather than characterizing identification restrictions, which are documented to be highly nonlinear and non-standard [see Canova and Sala (2009), Iskrev (2010) and Komunjer and Ng (2011)], identification-robust econometric methods that are valid whether

identification restrictions hold or not have gained popularity [see Andrews and Mikusheva (2013), Dufour, Khalaf and Kichian (2013), Guerron-Quintana, Inoue and Kilian (2013) and Qu (2013)]. In contrast to typically intractable and/or rarely reliable identification pre-tests, robust methods often deliver post-estimation information on weak-identification, in the form of unbounded confidence intervals. Some robust methods, which we endorse in the present chapter, also deliver information regarding lack-of-fit, in the form of empty confidence sets.

Most proposed identification-robust methods rely on asymptotic properties. However, macroeconomic series are rather limited, which suggest that finite-sample improvements are still required in this literature. In the third chapter, we propose finite-sample *confidence set* estimation methods for any DSGE that can be simulated once a finite dimensional parameter vector is specified. Four econometric principles are interacted to obtain exactness: (i) duality principles linking exact tests to exact confidence intervals via test inversion; (ii) the Monte Carlo [**MC**] test principle [see Dufour (2006), Dufour and Khalaf (2002a,b, 2003), Dufour and Kiviet (1996, 1998), Dufour, Khalaf, Bernard and Genest (2004), Dufour, Khalaf and Beaulieu (2003), Beaulieu, Dufour and Khalaf (2013), Khalaf and Peraza (2014a,b) and Khalaf and Saunders (2014)], (iii) the Indirect Inference [**InDInF**] principle, and (iv) the multivariate LR distance measure.

The MC method in general is not restricted to the VAR-based reduced form nor to the LR distance measure that we introduced in the third chapter. This chapter thus consider various non-LR distance measures such as the Wald-type statistic with identity weighting matrix [**Wald**],[1] the Lawley-Hotelling [**LH**] trace criterion, the Bartlett-Nanda-Pillai [**BNP**] trace criterion and the Maximum Root [**MR**] criterion.[2]

[1] Although inefficient in some contexts, this is common with InDInF; see *e.g.* Gouriéroux et al. (1993).

[2] For references, see Rao (1973), Anderson (1984) and Dufour and Khalaf (2002b).

The performance of above distance measures are studied in section 4.3.

Since solved DSGEs are (restricted) Vector Autoregressive Moving-Average processes [**VARMA**], in the third chapter, we use finite VARs as the approximated reduced form. Furthermore, we prove that if the variance-covariance matrix of the structural errors, denoted Σ, is diagonal, then the null distribution of the LR statistic is invariant to Σ, so the volatility parameters are evacuated from the estimation procedure. In other papers, the estimation of Σ is avoided via calibration or partialling-out procedures [see Lindé (2005), Dufour et al. (2013) and Castelnuovo and Fanelli (2015)]. However, the volatility parameters play an important role in *e.g.* impulse response analysis. We propose to use VAR with Leads and Lags [**VARLL**] as the alternative reduced form and estimate the volatility parameters simultaneously together with other parameters. To the best of our knowledge, this is the first paper that relies on leads and lags for identifying DSGEs. Perhaps the closest connection is the VAR with an error correction used by Del Negro, Schorfheide, Smets and Wouters (2007), since leads and lags and Error Correction Models [**ECM**] are competing approaches with near non-stationary data. In this chapter we use leads and lags and maintain a finite-sample frequentist approach.

A simulation study extending the design we introduced in the third chapter shows that with VARLL as the reduced form, the MC test has good power for inference on the volatility parameters. The MC method allows us to choose any reduced form and simulatable statistics. We view our results as further motivation to explore non-standard distance measures.

The chapter is organized as follows. In section 4.2, we describe the statistical framework and methodology. In section 4.3, we report the results of the simulation study. In section 4.4 we male a conclusion.

4.2 Framework and methodology

Follow the notation of the third chapter, consider the general state-space representation of a DSGE model:

$$X_t = A(\psi) X_{t-1} + B(\psi) V_t, \tag{4.2.1}$$

$$Y_t^\star = C(\psi) X_{t-1} + D(\psi) V_t, \tag{4.2.2}$$

$$V_t = H(\psi) V_{t-1} + \varepsilon_t, \tag{4.2.3}$$

where, for $t = 1, ..., T$, X_t is a vector of possibly unobserved state variables; $Y_t^\star$ is a vector of observed variables; V_t is an autoregressive process; ψ is a vector of deep parameters in the log-linearized DSGE models; ε_t is a vector of the structural economic shocks satisfying $E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_t' = \Sigma$, and $E\varepsilon_t\varepsilon_{t-j}' = 0$ for $j \neq 0$. Σ is a positive definite matrix and denote by J the Cholesky decomposition of Σ such that $\Sigma = JJ'$. Collect the parameters of J in the vector θ. Therefore $\vartheta = [\psi', \theta']'$. Uhlig (1995) (among others) provides a method of obtaining the matrices $A(.), B(.), C(.), D(.)$ and $H(.)$ from log-linearized DSGE models with deep parameter ψ. We further assume that observables compatible with (4.2.1)-(4.2.3) can be obtained by simulation for a given parameter vector $\vartheta = \vartheta_0$. Our purpose is to construct a joint and simultaneous confidence set for ϑ, maintaining a full-information perspective.

Dufour et al. (2013) propose an inference method that does not require a closed likelihood yet respects full-information principles. The method builds on the underlying infinite VAR solution approximated via the finite-order VAR model.[3]

$$Y_t = \Gamma(\vartheta) Y_{t,p} + u_t, \tag{4.2.4}$$

[3]For further conditions on such approximations, see Fernández-Villaverde, Rubio-Ramrez, Sargent and Watson (2007), Ravenna (2007).

where $\Gamma(\vartheta)$ is a $n^\star \times K$ matrix, $Y_{t,p} = vec([Y_{t-1}, ..., Y_{t-p}])$ is a $K \times 1$ vector and u_t is a vector of white noises with covariance matrix Ω.

4.2.1 Test statistic

We propose to add leads to the VAR(p) model (4.2.4) as follows

$$Y_t = \Gamma(\vartheta) Y_{t,p,l} + u_t, \tag{4.2.5}$$

where $Y_{t,p,l} = vec([Y_{t+l}, ..., Y_{t+1}, Y_{t-1}, ..., Y_{t-p}])$. In this case, the future values of Y_t may have an impact on current Y_t. VARLL serves as a better approximation of the VARMA auxiliary model than VAR and it can better capture the information on MA part.

This VAR(p, l) process is the form of Seemingly Unrelated Regressions [**SUR**] model and can be estimated equation-by-equation disregarding underlying restrictions using Ordinary Least Square [**OLS**] with the observed data. Zero or sign restriction can be used at this stage as well. Let $\hat{\Gamma}$ denote the $n^\star \times K$ matrix of OLS coefficients for the $n^\star$ equations. Its population counterpart is obtained using the state-space DSGE model (4.2.1) to (4.2.3) imposing $\vartheta = \vartheta_0$

$$\bar{\Gamma}(\vartheta_0) = \frac{1}{M} \sum_{m=1}^{M} \tilde{\Gamma}_m(\vartheta_0), \tag{4.2.6}$$

where $\tilde{\Gamma}_m(\vartheta_0)$ is the simulated data under DSGE restrictions and M is the number of

such sequences. Based on model (4.2.4), $\hat{\Gamma}$ and $\bar{\Gamma}(\vartheta_0)$ with observed data, construct:

$$W_t(Y,\vartheta) = Y_t - \hat{\Gamma}Y_{t,p,l}, \quad \hat{\Sigma}_W = \frac{1}{T-p-l}\sum_{t=p+l+1}^{T} W_t(Y,\vartheta)W_t'(Y,\vartheta), \tag{4.2.7}$$

$$W_t(Y,\vartheta_0) = Y_t - \bar{\Gamma}(\vartheta_0)Y_{t,p,l}, \quad \hat{\Sigma}^0_W = \frac{1}{T-p-l}\sum_{t=p+l+1}^{T} W_t(Y,\vartheta_0)W_t'(Y,\vartheta_0), \tag{4.2.8}$$

where $\hat{\Sigma}^0_W$ and $\hat{\Sigma}_W$ give the constrained (imposing $\vartheta = \vartheta_0$) and unconstrained sum of squared errors matrices.

This chapter consider various non-LR distance measures. The LH trace criterion, the BNP trace criterion and the MR criterion are functions of the roots $r_1, r_2, ..., r_q$ of the equation

$$|\hat{\Sigma}_W - r\hat{\Sigma}^0_W| = 0. \tag{4.2.9}$$

In particular,

$$\text{LH}(\vartheta_0) = \sum_{i=1}^{q}(1-r_i)/r_i, \tag{4.2.10}$$

$$\text{BNP}(\vartheta_0) = \sum_{i=1}^{q}(1-r_i), \tag{4.2.11}$$

$$\text{MR}(\vartheta_0) = \max_{1\le i\le q}(1-r_i)/r_i. \tag{4.2.12}$$

We also consider the Wald-type statistic:

$$\text{Wald}(\vartheta_0) = (vec(\hat{\Gamma}) - vec(\bar{\Gamma}(\vartheta_0)))'(vec(\hat{\Gamma}) - vec(\bar{\Gamma}(\vartheta_0))). \tag{4.2.13}$$

Here an identity matrix rather than the usual positive definite weighting matrix is used following the idea of Gouriéroux et al. (1993) who argue the loss in efficiency with respect to an optimal estimator is so small that it can be disregarded.

4.2.2 Estimation of Σ

In the third chapter, we show that if (i) VAR(p) process (4.2.4) is the reduced form; (ii) Σ is diagonal and invertible, then the null distribution of the Wilks measure is invariant to Σ. In that case, Σ is set to be the identity matrix and evacuated from the estimation procedure. However, if one would like to estimate Σ simultaneously, then there is an incentive to modify the reduced form. We propose to use VARLL as the preferred reduced form representation for the underlying data generating process [see Dufour and Torrès (2000)] and in section 4.3 we show that Σ is well identified.

4.3 Simulation study

In this section, we illustrate the finite-sample size and power of the proposed method through a simulation study extending the design we introduced in the third chapter. The primary objective of this exercise is to show (i) the benefit we obtain on the identification of Σ when VARLL is applied as the reduced form; (ii) the performance of different statistics. We do so by generating data from the model under a null, and analyze the rejection frequency of our method under the null and alternative parameter values.

The model consists of a NKPC, an Euler-equation based modified IS function, a Taylor rule followed by the monetary authority, and a description of the exogenous shock processes [Giannoni and Woodford (2004), Milani and Treadwell (2012)], as

given below:

$$\begin{aligned}
\pi_t &= \frac{\beta}{1+\beta\gamma}E_t\pi_{t+1} + \frac{\gamma}{1+\beta\gamma}\pi_{t-1} \\
&+ \frac{(1-\alpha)(1-\alpha\beta)}{\alpha(1+\beta\gamma)}\left(\omega y_t + \frac{\sigma^{-1}}{1-\phi}\left(y_t - \phi y_{t-1}\right)\right) + \mu_t, \qquad (4.3.1)\\
y_t &= \frac{1}{1+\phi}E_t y_{t+1} + \frac{\phi}{1+\phi}y_{t-1} - \frac{\sigma(1-\phi)}{1+\phi}\left(R_t - E_t\pi_{t+1}\right) + g_t, \qquad (4.3.2)\\
R_t &= \rho R_{t-1} + (1-\rho)\left[\chi_\pi \pi_t + \chi_y y_t\right] + \nu_t, \qquad (4.3.3)\\
\mu_t &= \rho_\pi \mu_{t-1} + \varepsilon_{\pi,t}, \qquad (4.3.4)\\
g_t &= \rho_y g_{t-1} + \varepsilon_{y,t}, \qquad (4.3.5)\\
\nu_t &= \rho_R \nu_{t-1} + \varepsilon_{R,t}, \qquad (4.3.6)
\end{aligned}$$

where the variable y denotes the output gap, π denotes inflation, and R denotes the nominal interest rate. μ, g, and ν are exogenous disturbances and evolve according to a AR(1) process with autoregressive coefficients ρ_i, for $i = \mu, g, \nu$, and are affected by *i.i.d.* fundamental Gaussian shocks $\varepsilon_i \sim N(0, \sigma_i^2)$. In what follows, the two remaining parameters of the model, namely the discount factor, β, and the elasticity of marginal cost to income, ω is kept fixed at 0.99 and 2 respectively. A full derivation of the model from micro-foundations can be found in Milani and Treadwell (2012). Collect all parameters of interest in $\vartheta = [\phi, \sigma, \gamma, \alpha, \rho, \chi_\pi, \chi_y, \rho_\pi, \rho_y, \rho_R, \sigma_\pi, \sigma_y, \sigma_R]'$.

Figures 4.1 through 4.4 plot the rejection frequency curves for each parameters for different sample sizes at the 5% level.[4] In each sub-plot, the rejection frequency of a parameter at their null value (provided in the title with a subscript 0) gives the size of our method, while the rejection frequency of the parameter away from the null value gives the power.

Figure 4.1 and 4.2 show that using the VARLL as the reduced form, the proposed

[4]The rejection frequency curves are generated by varying the grid value of the parameters one at a time, while keeping all others at their null value.

method has good power for inference on the volatility parameters without considerable loss of power for inference on parameters other than the ones that map to the forward relative to backward looking coefficients in the model. Therefore, for practitioners interested in inference on θ, adding leads and lags can be a useful option. The power loss on non-volatility parameters is not unexpected as the dimensionality is increased by three. More important is the fact that only forward looking terms seem to be affected. In view of the important literature on the joint significance of forward and backward looking terms [summarized *e.g.* in Benati (2008), Gali, Gertler and David Lopez-Salido (2005) and the conclusion of Cochrane (2011)], this result is noteworthy and calls for deeper analysis.

To show that the proposed method is not limited to a particular statistic, we conduct a simulation study based on reduced form VARLL with $p = 4, l = 2$ and compare the performance of different statistics. Figure 4.3 and 4.4 show that none of these statistics is dominating and the improvement in power is limited.

4.4 Conclusion

We extend the finite-sample exact simulation-based confidence set estimation method proposed the third chapter beyond a VAR reduced form and LR measures. We propose to use VAR with leads and lags as the reduced form to simultaneously estimate the volatility parameters.[5] We also show that various statistics can by applied to our method and we document their performance in the simulation study. The strength of the MC method is that it offers us the infinite choices of reduced forms and simulatable statistics.[6] With regards to the DSGE under consideration, we find that simultaneous inference on the volatility parameters may cost precision with the for-

[5] In future work we may also simultaneously estimate the correlations of the shocks.

[6] These include the score tests in Qu (2013), Andrews and Mikusheva (2013) and Guerron-Quintana et al. (2013).

ward/backward looking terms. We view this result as exploratory, yet worthy to report, in view of the conclusion of Cochrane (2011): "*If inflation is, in fact, stabilized in modern economies by interest rate targets interacted with backward-looking IS and Phillips curves, economists really have no idea why this is so.*"

Bibliography

Anderson TW. 1984. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 2nd edition.

Andrews I, Mikusheva A. 2013. Maximum likelihood inference in weakly identified DSGE models. *Quantitative Economics* : to appear.

Beaulieu MC, Dufour JM, Khalaf L. 2013. Exact confidence sets and goodness-of-fit methods. *Journal of Econometrics* : to appear.

Benati L. 2008. Investigating inflation persistence across monetary regimes. *Quarterly Journal of Economics* **123**: 1005–1060.

Canova F, Sala L. 2009. Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* **54**: 431–449.

Castelnuovo E, Fanelli L. 2015. Monetary policy indeterminacy and identification failures in the U.S.: Results from a robust test. *Journal of Applied Econometrics* : to appear.

Cochrane JH. 2011. Determinacy and identification with Taylor rules. *Journal of Political Economy* **119**: 565 – 615.

Del Negro M, Schorfheide F, Smets F, Wouters R. 2007. On the fit of New Keynesian models. *Journal of Business and Economic Statistics* **25**: 123–143.

Dufour JM. 2006. Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. *Journal of Econometrics* **133**: 443–478.

Dufour JM, Khalaf L. 2002a. Exact tests for contemporaneous correlation of disturbances in seemingly unrelated regressions. *Journal of Econometrics* **106**: 143–170.

Dufour JM, Khalaf L. 2002b. Simulation based finite and large sample tests in multivariate regressions. *Journal of Econometrics* **111**: 303–322.

Dufour JM, Khalaf L. 2003. Finite sample tests in seemingly unrelated regressions. In Giles D (ed.) *Computer-Aided Econometrics*. Marcel Dekker: New York, 11–35.

Dufour JM, Khalaf L, Beaulieu MC. 2003. Exact skewness-kurtosis tests for multivariate normality and goodness-of-fit in multivariate regressions with application to asset pricing models. *Oxford Bulletin of Economics and Statistics* **65**: 891–906.

Dufour JM, Khalaf L, Bernard JT, Genest I. 2004. Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. *Journal of Econometrics* **122**: 317–347.

Dufour JM, Khalaf L, Kichian M. 2013. Identification-robust analysis of DSGE and structural macroeconomic models. *Journal of Monetary Economics* **60**: 340–350.

Dufour JM, Kiviet JF. 1996. Exact tests for structural change in first-order dynamic models. *Journal of Econometrics* **70**: 39–68.

Dufour JM, Kiviet JF. 1998. Exact inference methods for first-order autoregressive distributed lag models. *Econometrica* **66**: 79–104.

Dufour JM, Torrès O. 2000. Markovian processes, two-sided autoregressions and

finite-sample inference for stationary and nonstationary autoregressive processes. *Journal of Econometrics* **99**: 255–289.

Fernández-Villaverde J, Rubio-Ramrez JF, Sargent TJ, Watson MW. 2007. ABCs (and Ds) of understanding VARs. *American Economic Review* **97**: 1021–1026.

Gali J, Gertler M, David Lopez-Salido J. 2005. Robustness of the estimates of the hybrid New Keynesian Phillips curve. *Journal of Monetary Economics* **52**: 1107–1118.

Giannoni M, Woodford M. 2004. Optimal Inflation-Targeting Rules. In *The Inflation-Targeting Debate*, NBER Chapters. National Bureau of Economic Research, Inc, 93–172.

Gouriéroux C, Monfort A, Renault E. 1993. Indirect inference. *Journal of Applied Econometrics* **8**: 85–118.

Guerron-Quintana P, Inoue A, Kilian L. 2013. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics* **4**: 197–229.

Iskrev N. 2010. Local identification in DSGE models. *Journal of Monetary Economics* **57**: 189–202.

Khalaf L, Peraza B. 2014a. Finite sample simulation-based inference for moving average processes. *mimeo* .

Khalaf L, Peraza B. 2014b. Persistence robust simulation-based inference for autoregressive moving average processes. *mimeo* .

Khalaf L, Saunders CJ. 2014. Indirect set inference for dynamic panel models. *mimeo* .

Komunjer I, Ng S. 2011. Dynamic identification of DSGE models. *Econometrica* **79**: 1995–2032.

Lindé J. 2005. Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics* **52**: 1135–1149.

Mavroeidis S, Plagborg-Mollor M, Stock JH. 2014. Empirical evidence on inflation expectations in the New Keynesian Phillips curve. *Journal of Economic Literature* : to appear.

Milani F, Treadwell J. 2012. The effects of monetary policy news and surprises. *Journal of Money, Credit and Banking* **44**: 1667–1692.

Qu Z. 2013. Inference in DSGE models with possible weak identification. *Quantitative Economics* : to appear.

Rao CR. 1973. *Linear Statistical Inference and its Applications*. Wiley, New York, 2nd edition.

Ravenna F. 2007. Vector autoregressions and reduced form representations of DSGE models. *Journal of Monetary Economics* **54**: 2048–2064.

Uhlig H. 1995. A toolkit for analyzing nonlinear dynamic stochastic models easily. *mimeo* .

Figure 4.1: Rejection frequency: the VAR($p = 4$) (dashed line) and VARLL($p = 2, l = 2$) (solid line) with LR criterion at level 5%, $T = 103$,

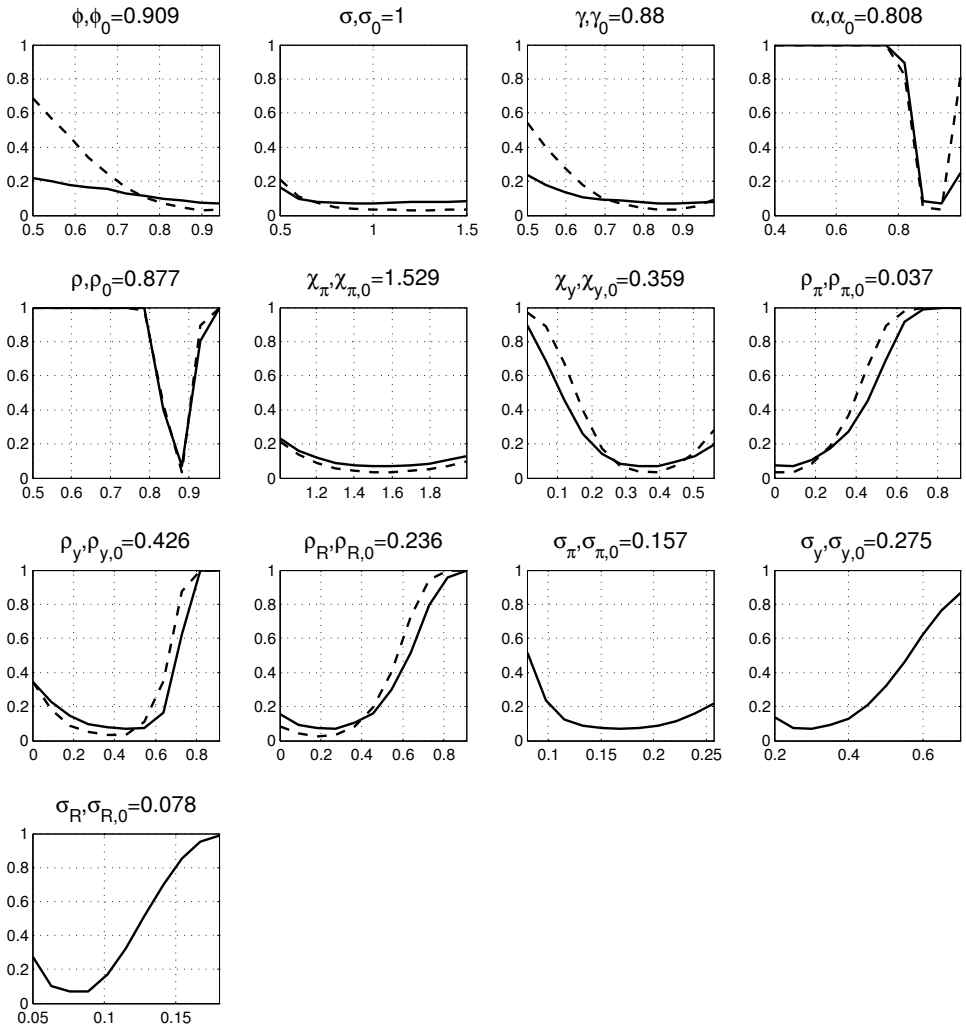


Figure 4.2: Rejection frequency: the VAR($p = 4$) (dashed line) and VARLL($p = 2, l = 2$) (solid line) with LR criterion at level 5%, null 2, $T = 175$, $p = 2$ and $l = 2$

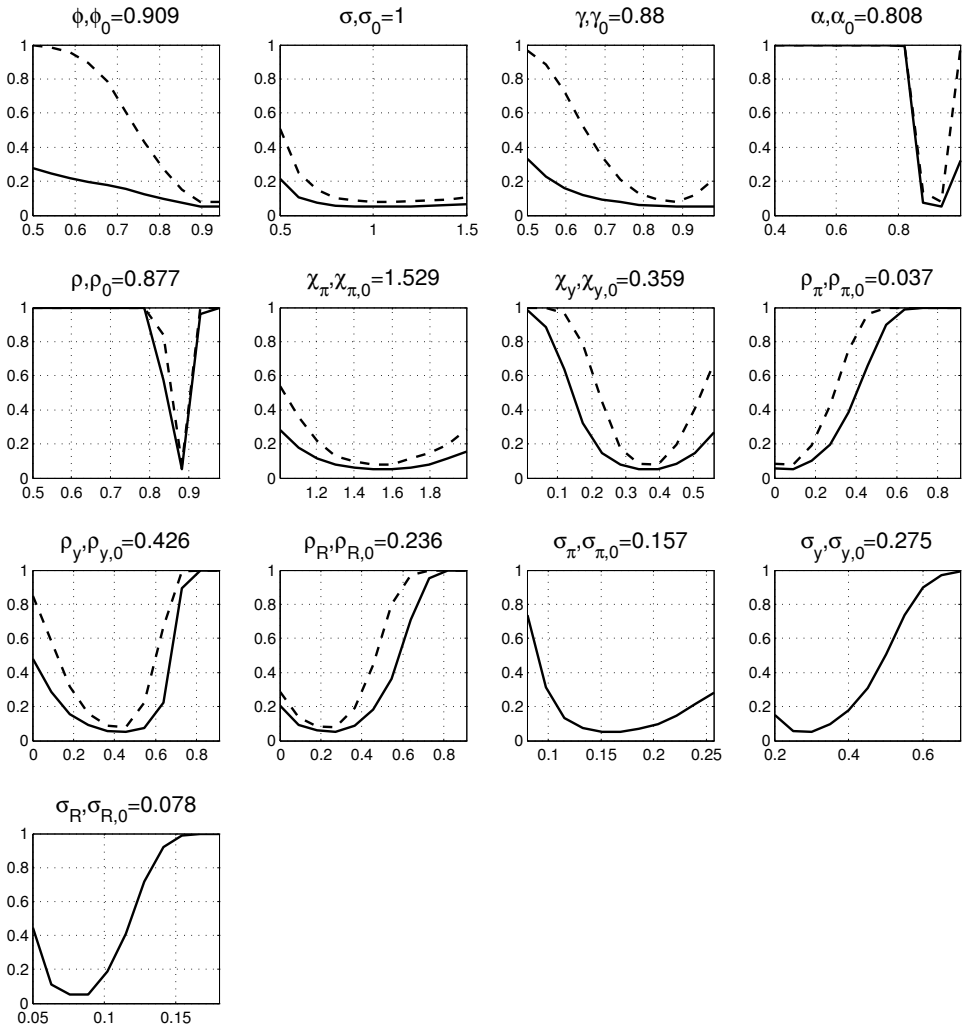


Figure 4.3: Rejection frequency: MC method (apply the LR criterion (solid line), the LH trace criterion (dashed line), the BNP trace criterion (dotted line), the MR criterion (dash-dot line) and the Wald criterion (circle)) at level 5%, $T = 103$, $p = 4$, $l = 2$

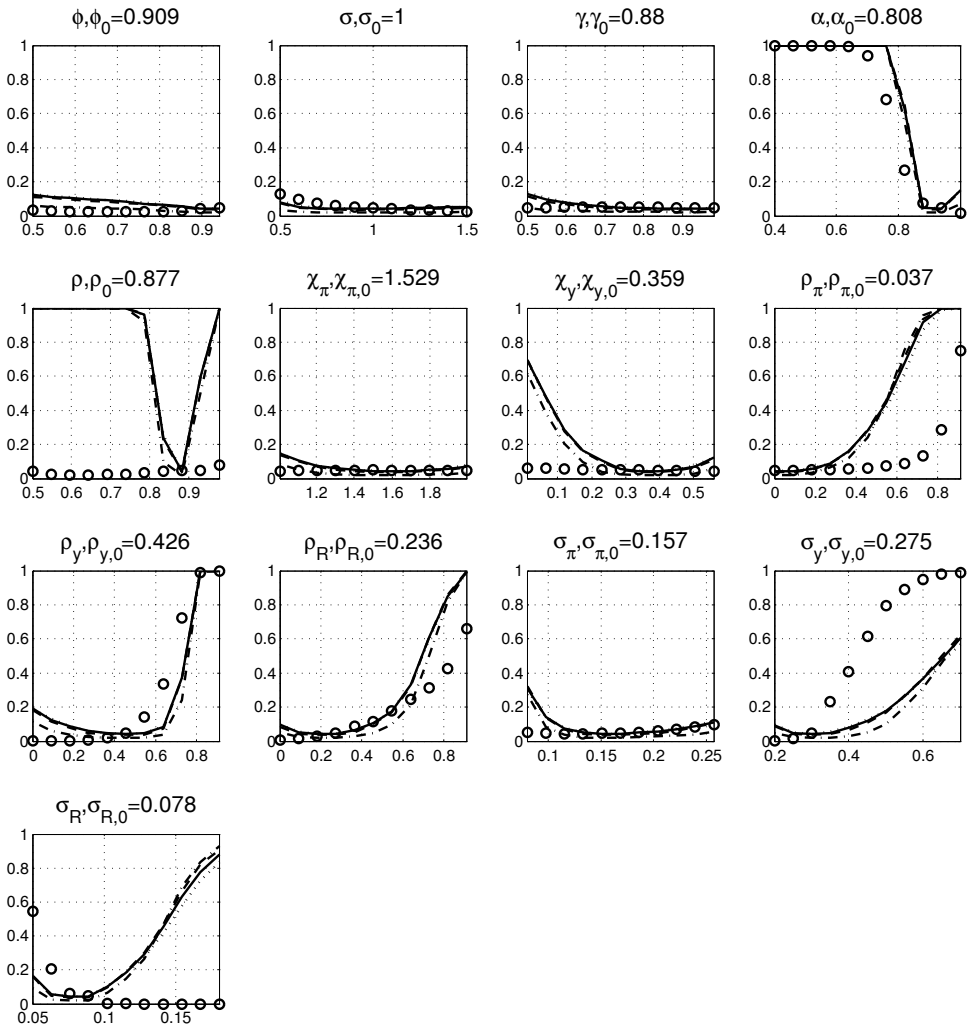
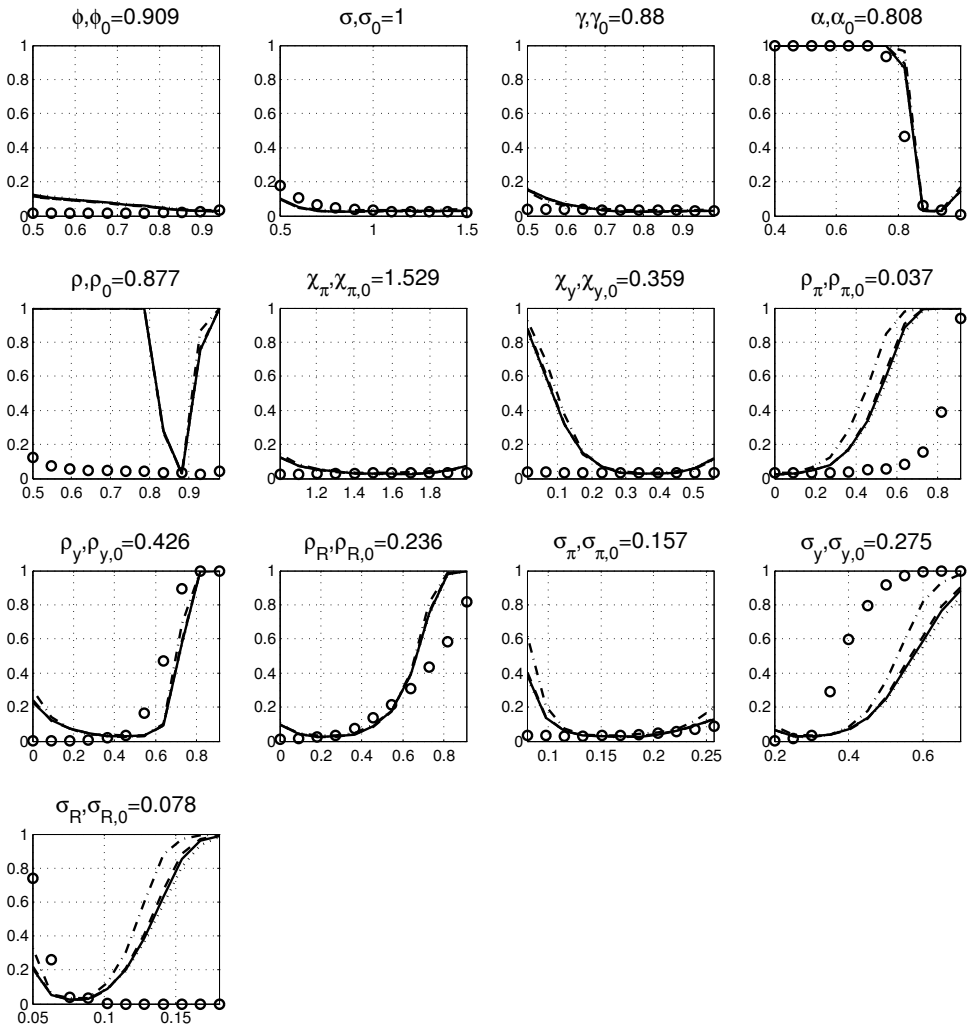


Figure 4.4: Rejection frequency: MC method (apply the LR criterion (solid line), the LH trace criterion (dashed line), the BNP trace criterion (dotted line), the MR criterion (dash-dot line) and the Wald criterion (circle)) at level 5%, $T = 175$, $p = 4$, $l = 2$



Appendix A

Appendix for Chapter 1

A.1 Competing algorithms, similarities and differences

A.1.1 Brief introduction of SA and GA

The simulated annealing is inspired by the annealing process in metallurgy. At the initial state, a candidate solution s is drown randomly within the search space. Then a new candidate s' is drawn from the neighborhood of s. Consider the minimization problem, if s' has a lower objective function value, jump to s'; otherwise, jump to s' with a probability. The probability depends on the current state temperature. The temperature is declining slowly through iterations. A jump always occurs if the temperature is high, which helps the global search at the beginning. However, as the temperature is cooling down, the jump becomes more and more unlikely. The process is repeated until some stopping criteria are met.

The genetic algorithm is inspired by "natural selection". The evolution starts from a randomly drawn population of candidate solutions (chromosomes). The objective function value of each chromosome is evaluated and a group (parents) of chromosomes

that have lower objective function values are selected. The parents are recombined (crossover) and mutated with some probabilities to create next generation (children), which will be used in the next iteration. The process is repeated until some stopping criteria are met.

A.1.2 Similarities of SA, GA and PSO

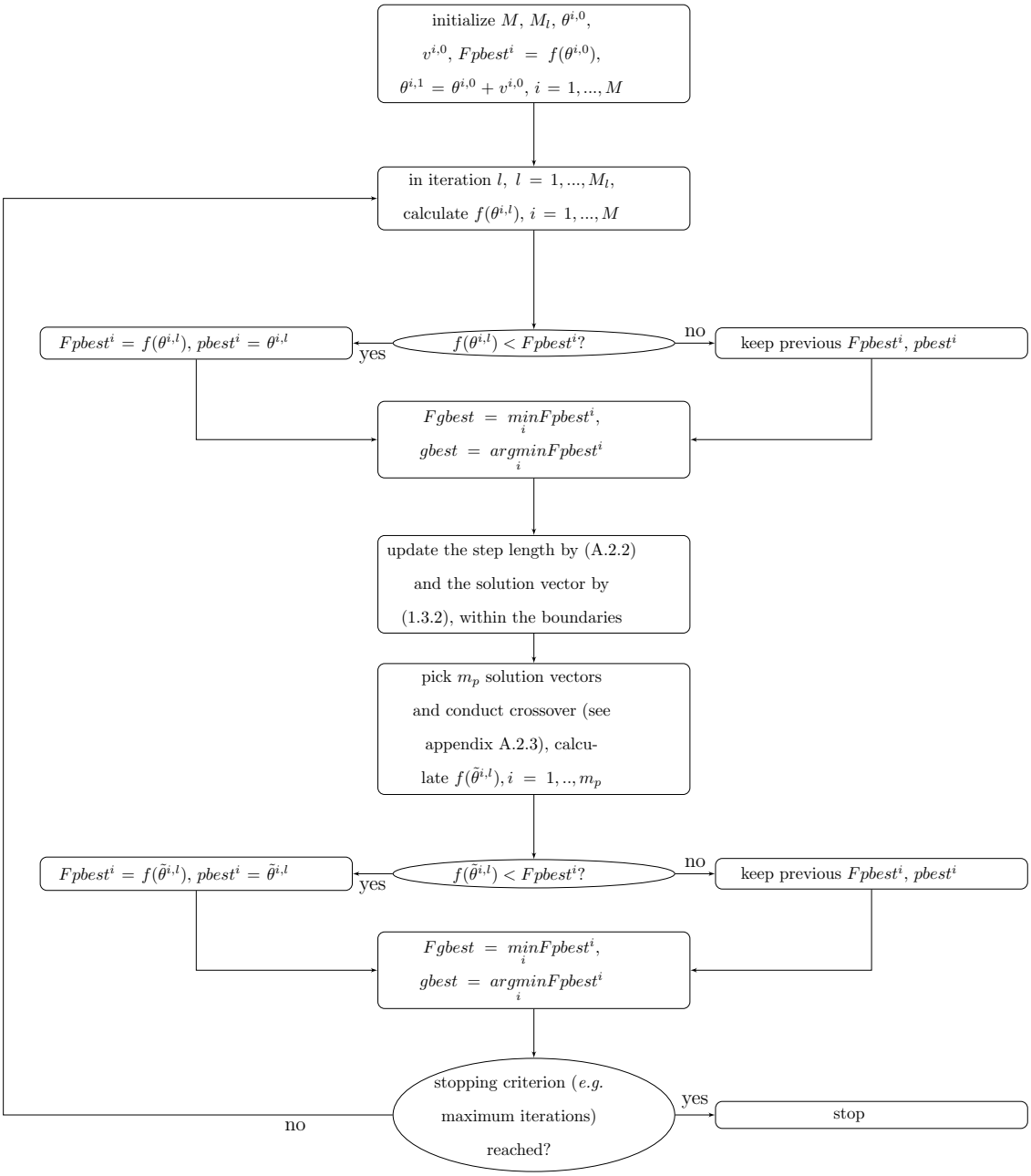
They are all iterative stochastic search methods which need less assumptions than conventional methods. Each candidate solution is initialized randomly, then objective function evaluation, candidate updating and best objective function value updating are involved. All three algorithms require the tuning of some problem-dependent parameters and there is no general way for finding best choices for a given problem.

A.1.3 Key differences among SA, GA and PSO

The key concept in SA is accepting worse solutions with decreasing probabilities as the temperature declines; while in GA and PSO, however, worse solutions are never accepted. GA and PSO start with a group of candidate solutions while the SA is not. The key difference between GA and PSO is that the candidate solutions in latter exchange information endogenously. In PSO, each candidate solution stores and utilizes its own information as well as the shared information and adjusts its behavior adaptively. The behavior of each solution vector will affect the group and vice versa. The mutual best responses will lead to a Nash-type equilibrium, which may explain the convergence of PSO. However, in GA, the information exchange among solution vectors is exogenous. They are similar to the products in a factory: products with low quality (bad candidate solutions) are destroyed; high-quality products are selected and redesigned (genetic operators such as cross over and mutation) in order to achieve higher qualities.

A.2 Procedures

A.2.1 Flow chart: PSO



A.2.2 Deterministic versus random weight in step length updating

Equation (A.2.1) defines the weight on the previous step length as a linearly decreasing function of current iteration l:

$$w = w_{max} - \frac{l(w_{max} - w_{min})}{iter_{max}}, \tag{A.2.1}$$

where, $w_{max} = 0.9$ is the initial weight; $w_{min} = 0.4$ is the final weight; $iter_{max}$ is the maximum number of iteration. The initial w is large. Hence the step length is relatively large to ensure a global exploration. The final w is small so that a local search is conducted at the end of search. In this design w is deterministic.

Lin (2011) argues that w should be linked with w_1 and w_2 and he proposes an improved formula for w that increases the stability and speed of convergence of PSO. The step length is modified as:

$$v^{i,l+1} = (1 + \varphi - \sqrt{2\varphi + \varphi^2})v^{i,l} + w_1(pbest^i - \theta^{i,l}) + w_2(gbest - \theta^{i,l}), \tag{A.2.2}$$

$$\varphi = w_1 + w_2, \quad \varphi \in (0, 4]. \tag{A.2.3}$$

The author replaces the deterministic weight w with a function of φ and performs a mathematical analysis on the relationship between w and φ in complex space and gives a stable region for φ. In this paper we adopt the random weight in step length updating proposed by Lin (2011).

A.2.3 The hybrid updating procedures

Lei (2012) proposes that after each iteration, randomly pick m_p solution vectors from the M solution vectors, conduct crossover based on their current solution vectors,

$pbest^i$ and top m_p best solution vectors (denoted as $bs^{i,l}, i = 1, .., m_p$) ranked by evaluating the associated objective function values within this iteration, and generate m_p new solution vectors. Evaluate the objective function value of each new solution vector and compare to the one of top m_p best solution vectors. Replace the according solution vector if the new one has better function value. The pseudo-code is as follows:

- $d_1 = floor(D/3)$, $d_2 = floor(2D/3)$
- Initialize integer $m_p \in [1, M]$
- **For** each $i = 1$: m_p
- Conduct crossover $\tilde{\theta}^{i,l} = [pbest^i(1 : d_1), \theta^{i,l}(d_1 + 1 : d_2), bs^{i,l}(d_2 + 1 : D)]'$
- **If** $f(\tilde{\theta}^{i,l}) < Fpbest^i$, then $Fpbest^i = f(\tilde{\theta}^{i,l})$ and $pbest^i = \tilde{\theta}^{i,l}$
- **If** $f(\tilde{\theta}^{i,l}) < Fgbest$, then $Fgbest = f(\tilde{\theta}^{i,l})$ and $gbest = \tilde{\theta}^{i,l}$
- **End for**

Appendix B

Appendix for Chapter 3

B.1 Proof of theorem 1

Proof. Transform the VAR(p) process (3.2.4) to the companion VAR(1) process [see Canova (2007)]. Let $\mathbb{Y}_t = [Y_t, Y_{t-1}, ..., Y_{t-p+1}]'$; $\mathbb{U}_t = [u_t, 0, ..., 0]'$ and

$$\Gamma = \begin{bmatrix} \Gamma_1(\vartheta) & \Gamma_2(\vartheta) & ... & \Gamma_p(\vartheta) \\ I_{n^\star} & 0 & ... & 0 \\ 0 & ... & I_{n^\star} & 0 \end{bmatrix}.$$

Then the companion form representation is

$$\mathbb{Y}_t = \Gamma \mathbb{Y}_{t-1} + \mathbb{U}_t, \tag{B.1.1}$$

where $\mathbb{Y}_t$, $\mathbb{U}_t$ are $n^\star p \times 1$ vectors and Γ is a $n^\star p \times n^\star p$ matrix. Let $\mathbb{Y} = [\mathbb{Y}_1, \mathbb{Y}_2, ..., \mathbb{Y}_T]'$; $\mathbb{U} = [\mathbb{U}_1, \mathbb{U}_2, ..., \mathbb{U}_T]'$. Then (B.1.1) is

$$\mathbb{Y} \;=\; \mathbb{Y}_{-1}\Gamma + \mathbb{U}, \tag{B.1.2}$$

where $\mathbb{Y}$ is a $T \times n^{\star}p$ matrix of observables, $\mathbb{Y}_{-1}$ is the first lag of $\mathbb{Y}$. Let $\Xi = I_p \otimes \Sigma = I_p \otimes (J'J)$. Post-multiply the data $\mathbb{Y}$ by $G = I_p \otimes (J^{-1})$

$$\mathbb{Y}G = \mathbb{Y}_{-1}GG^{-1}\Gamma G + \mathbb{U}G. \tag{B.1.3}$$

Redefine the above equation

$$\mathbb{Y}^{\star} = \mathbb{Y}^{\star}_{-1}\Gamma^{\star} + \mathbb{U}^{\star}. \tag{B.1.4}$$

Since G is diagonal and invertable, then

$$\Gamma^{\star} = G^{-1}\Gamma G = \Gamma. \tag{B.1.5}$$

Similarly, the population parameter $\bar{\Gamma}(\vartheta_0)$ does not depend on θ_0. The considered statistic defined in equation (3.2.10) is a ratio of determinants, so $|G|$ can be factored out and will be canceled out:

$$\hat{\mathbb{U}}^{\star\prime}\hat{\mathbb{U}}^{\star} = \mathbb{Y}^{\star\prime}M^{\star}\mathbb{Y}^{\star} = G'\mathbb{Y}'M\mathbb{Y}G \tag{B.1.6}$$

$$\hat{\mathbb{U}}^{\star\prime}_0\hat{\mathbb{U}}^{\star}_0 = G'(\mathbb{Y}' - \bar{\Gamma}'(\vartheta_0)\mathbb{Y}'_{-1})(\mathbb{Y} - \mathbb{Y}_{-1}\bar{\Gamma}(\vartheta_0))G \tag{B.1.7}$$

$$\Lambda = |\hat{\mathbb{U}}^{\star\prime}_0\hat{\mathbb{U}}^{\star}_0|/|\hat{\mathbb{U}}^{\star\prime}\hat{\mathbb{U}}^{\star}| \tag{B.1.8}$$

$$= \frac{|G'||\mathbb{Y}'M\mathbb{Y}||G|}{|G'||(\mathbb{Y}' - \bar{\Gamma}'(\vartheta_0)\mathbb{Y}'_{-1})(\mathbb{Y} - \mathbb{Y}_{-1}\bar{\Gamma}(\vartheta_0))||G|} \tag{B.1.9}$$

$$= \frac{|\mathbb{Y}'M\mathbb{Y}|}{|(\mathbb{Y}' - \bar{\Gamma}'(\vartheta_0)\mathbb{Y}'_{-1})(\mathbb{Y} - \mathbb{Y}_{-1}\bar{\Gamma}(\vartheta_0))|} \tag{B.1.10}$$

where $M^{\star} = I_T - \mathbb{Y}^{\star}_{-1}(\mathbb{Y}^{\star\prime}_{-1}\mathbb{Y}^{\star}_{-1})^{-1}\mathbb{Y}^{\star\prime}_{-1}$ and $M = I_T - \mathbb{Y}_{-1}(\mathbb{Y}'_{-1}\mathbb{Y}_{-1})^{-1}\mathbb{Y}'_{-1}$. ■