Identification and Persistence-Robust Exact Inference in DSGE Models *

Lynda Khalaf
Carleton University†

Zhenjiang Lin
University of Nottingham Ningbo China ‡

Abeer Reza
Bank of Canada §

December 23, 2016

Abstract

We develop a simulation-based finite-sample identification-robust confidence-set estimation method for Dynamic Stochastic General Equilibrium [DSGE] models by testing for insignificant discrepancies between relevant though possibly mis-specified auxiliary models from their DSGE-restricted population counterparts. Auxiliary forms include standard and forward and backward looking Vector Auto-Regressions [VAR], as well as vector error correction models.

*The views expressed in this article are of the authors' alone, and should not be attributed to the Bank of Canada. We thank Yuriy Gorodnichenko, James Hamilton, Alessandro Justiniano, Lutz Kilian, Sophocles Mavroeidis and Zhongjun Qu for helpful comments. We also thank seminar participants at the 14th Dynare conference, 2014 NBER-NSF time series conference, the 2015 Royal Economic Society meetings, Barcelona GSE Summer Forum, Société Canadienne de Science Économique, and the Bank of Canada. All errors and omissions are ours. A previous version of this paper was circulated under the title "On Attempts to Rescue Identification in DSGE Models: A Finite-Sample Exact Analysis".

†Corresponding author. Department of Economics, Carleton University. Mailing address: Department of Economics, Carleton University, Loeb Building 1125 Colonel By Drive, Ottawa, Ontario, K1S 5B6 Canada. Tel +1(613)-520-2600x8697; FAX: +1(613)-520-3906; email: lynda.khalaf@carleton.ca.

‡School of Economics, University of Nottingham Ningbo China. Mailing address: School of Economics, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, Zhejiang, 315100 China. Tel +86(574)-8818-0000x8461; FAX:+86(574)-8818-0125; email: zhenjiang.lin@nottingham.edu.cn.

§Bank of Canada. Mailing address: Bank of Canada, 234 Laurier Avenue West, Ottawa, Ontario, K1A 0G9 Canada. Tel +1(613)-782-8823; email: reza@bankofcanada.ca.

Our method controls coverage exactly by combining Indirect Inference with the Monte Carlo test method, regardless of sample size, model dimension, lag truncation order and identification of deep parameters. In contrast with most available identification-robust methods, exact or near unit roots are allowed, and a linearized state space solution is not required. The key ingredient is that the DSGE can be simulated once a finite dimensional parameter is specified, which covers typical (often calibration-validated) micro-founded models. Size and power properties and empirical relevance are illustrated via canonical real business cycle and New Keynesian models.

Keywords: Dynamic stochastic general equilibrium; Estimation uncertainty; Indirect inference; Exact confidence-set estimation; Monte Carlo test; New Keynesian model; Real Business Cycle model; finite sample inference.

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), Cochrane (2011), Schorfheide (2013)]. Under weak identification, many popular methods can deliver confidence bands or credible sets that are invalid even asymptotically 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 [Inoue and Rossi (2011),Tkachenko and Qu (2012, 2016), Moon and Schorfheide (2012), Qu (2014), Dufour, Khalaf and Kichian (2013), Guerron-Quintana, Inoue and Kilian (2013), Basturk, akmali, Ceyhan and Van Dijk (2014), Magnusson and Mavroeidis (2014), Andrews and Mikusheva (2015)].[1]

[1]Alternatives including limited information modeling and econometrics, survey-based expectations, or information stickiness are not covered in our analysis. See Mavroeidis (2005, 2010), Dufour, Khalaf and Kichian (2006), Kleibergen and Mavroeidis (2009), Coibion and Gorodnichenko (2010,

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, raises more fundamental concerns 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".

One way to read Cochrane (2011) is that the particular model in question is not identified for theoretical rather than for empirical reasons. "Not only might these problems exist, but theory predicts that most of them do exist". In models that are not unidentified "in principle", the econometric issues raised by Cochrane do not differ "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. Such central difficulties thus stem from persistence, which invalidates standard and even recently proposed identification-robust methods [Yang and Xu (2014)].

In this paper we propose a finite-sample simulation-based econometric method that delivers persistence and identification-robust size control, in a frequentist full-information rational expectations framework. The proposed method draws on two classes: (i) the above cited identification-robust methods and (ii) Indirect Inference [**InDInF**]. The latter method formally replaces complicated or intractable statistical functions by computer simulations.[2] More to the point from our perspective, Dridi, Guay and Renault (2007) suggest that calibrations, the common alternative to estimation and inference in macroeconomics, may formally be captured via InDInF. Gouriéroux, Phillips and Yu (2010) also show that the method may control persistence based biases specifically in dynamic panels.[3]

However, to the best of our knowledge, all available identification-robust meth-

2015), Mavroeidis, Plagborg-Mller and Stock (2014), Kapetanios, Khalaf and Marcellino (2016) and references therein for examples of such alternatives.

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

[3]Furthermore, 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).

ods are justified using asymptotic statistical theory, which is known to break down when persistence is high. 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] or sidestep near-boundary persistence. In contrast, macroeconomic series, especially time series that can be modeled with stable structures, are typically short[4] and near-unit roots cannot be ruled-out; see Mikusheva (2009) and Yang and Xu (2014) for recent practical references illustrating the latter problem. These issues combined open the door to spurious over-rejections, and bring to question the substantive results obtained by available asymptotically justified identification-robust methods.

With regards to InDInF broadly, confidence sets and hypotheses tests are also 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]

This paper introduces a simulation-based confidence-set estimation method based on InDInF for DSGE models that provides size-control regardless of sample length or persistence. We need neither the existence of a limiting distribution, nor for errors to be exclusively Gaussian, nor for (near or) unit roots to be ruled out. A linearized state space solution is also not necessary. All we require is the possibility of simulating data from the considered DSGE given its deep parameters. Through simulation studies based on a real business cycle and a New Keynesian model we show that our method has exact size and very good power for many key parameters. The prototypical New Keynesian model [Woodford (2003), Giannoni and Woodford (2004), Milani and Treadwell (2012)] is also considered to illustrate how our strategy works in practice.

InDInF relies on a binding function that links a model of interest to an aux-

[4]See *e.g.* the literature emphasizing pre and post-Volcker analyses of the Taylor rule, including Mavroeidis (2010), Benati (2008) and the references therein for a historical perspective on instabilities with inflation persistence.

[5]Exceptions include Dufour and Valéry (2009) for stochastic volatility models, and Dufour and Kurz-Kim (2010), Dufour, Khalaf and Beaulieu (2003) and Beaulieu, Dufour and Khalaf (2014) in Multivariate regressions and simple location-scale models with fat-tailed fundamentals.

iliary one for which a simple estimator is available. Minimum-distance estimation matches the observed to the calibrated criteria, which formally captures calibration and equilibrium, and attenuates persistence biases. In DSGEs, Vector Auto-Regressions [**VAR**] provide natural choices, since solved models can often be approximated via restricted VARs, for which the coefficients can be calibrated by simulation. Conformably, we first consider the popular unrestricted finite order VAR where, in contrast to Dufour et al. (2013), we account, exactly, for a mis-specified lag order. Secondly, we introduce a forward and backward looking VAR. To the best of our knowledge, this is the first paper that relies on leads and lags for identifying DSGE parameters. Perhaps the closest connection is the VAR with an Error Correction [**EC**] used *e.g.* by Del Negro, Schorfheide, Smets and Wouters (2007), since leads and lags and EC models are competing approaches with near non-stationary data. Our third auxiliary specification is thus a Vector EC model [**VECM**]. Finally, we assess whether leads in VECM improve mis-specification corrections.

The key ingredient to our methodology 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 devise a finite sample multi-stage simulation-based strategy to robustify InDInF on DSGEs. The outcome is a confidence set, derived by screening the observed-to-calibrated distance in question over relevant parameter values and retaining the ones with insignificant distances. Conceptually, this summarizes an infinite number of significance checks across the parameter space; if each check is valid at every check-point, the non-rejected set will reveal the full range of true parameters that fit the data. A non-informative possibly unbounded range confirms weak identification and all parameter values will be rejected if the model lacks fit. In contrast to usual methods, non-rejected values can thus be interpreted without taking an ex-ante stand on whether the model is the right one.

Rather than minimizing the observed-to-calibrated distance which makes sense from a point-estimation perspective, we thus view model-to-data matching as a search for the *insignificant* distances. The so-called test inversion general principle underlying confidence sets of this form conveys much more information than minimum distance and eschews standard errors. As a matter of fact, usual standard errors severely

under-state uncertainty when objective functions may have flat weakly-identified zones. The critical requirement is a test that is valid at every check-point. This paper introduces multi-level Monte Carlo Test [**MCT**] methods extending Dufour (2006), to validate the above distance measure at each tested point, in finite samples.

The remainder of the paper is organized as follows. Section 2 describes the statistical framework. The methodology is discussed in section 3. Section 4 documents the properties of our methodology via two fully micro-founded canonical DSGE models. Section 5 provides a practical example where a New Keynesian baseline model is taken to the data, and section 6 provides concluding comments.

2 Framework

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

$$X_t = \mathbf{F}\left(X_{t-1}, V_t, \vartheta\right), \tag{1}$$

$$Y_t = \mathbf{G}\left(X_{t-1}, V_t, \vartheta\right), \tag{2}$$

$$V_t = \mathbf{J}\left(V_{t-1}, \varepsilon_t, \vartheta\right), \tag{3}$$

where, X_t is a vector of possibly unobserved endogenous state variables; Y_t is a vector of observed endogenous variables of dimension $n^\star$; V_t is a vector of exogenous process driven by the vector of shocks ε_t, satisfying $E\varepsilon_t = 0$, $E\varepsilon_t\varepsilon_t' = \mathbf{I}$, and $E\varepsilon_t\varepsilon_{t-j}' = 0$ for $j \neq 0$; and ϑ is a vector of deep parameters. Functions $\mathbf{F}$ and $\mathbf{G}$ relate the exogenous driving processes and pre-determined state variables to current period endogenous variables, and comprise the solution of the DSGE model. $\mathbf{J}$ is a description of the law of motion of the exogenous process. The solution may or may not be linear in X_{t-1}, V_t and ϵ_t. However, most DSGE solutions are usually non-linear in ϑ.

We further assume that observables compatible with (1)-(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 (2015) are not required. All we need is the possibility of simulating data compatible with (1)-(3) imposing usual assumptions of the parameter space of ϑ.

InDInF, which exploits the fact that simulations can be easily drawn from the model whether likelihoods and/or scores are regular or not, provides interesting statistical objective functions in our context.[6]

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. Note that linearized versions of DSGE solutions in (1)-(3) can be expressed as a VARMA or infinite VAR model in the observables whose coefficients are nonlinear functions of ϑ. In that special case, which we consider in our examples below, 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]

$$\mathbf{\Gamma}(L, \vartheta, p)Y_t = z(\vartheta) + u_t, \tag{4}$$

where $\mathbf{\Gamma}(L, \vartheta, p) = \sum_{i=0}^{p} \Gamma_i(\vartheta)L^i$, $\Gamma_0(\vartheta) = \mathbf{I}$, L is a lag operator, p is the truncation order, $z(\vartheta)$ is a constant, and u_t is a vector of white noises with covariance matrix Ω.[8] From there on, and for a given parameter value $\vartheta = \vartheta_0$, regress the Ordinary Least Square [**OLS**] residual $\hat{u}_t(\vartheta_0) = \hat{\mathbf{\Gamma}}(L, \vartheta_0, p)Y_t - \hat{z}(\vartheta_0)$ on as many of the lags of Y_t 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

[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).

[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.

a VAR, which evacuates identification concerns.

In traditional estimation methodology a point estimate is found first and confidence intervals are then constructed, of the form {estimate ± standard error × critical point}. In contrast, "inverting" an identification- robust test, as in Dufour et al. (2013) and many other contributions to this literature, produces a confidence region assembling the ϑ_0 that are not rejected by the test in question at a certain level $\alpha^\star$. 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.

2.2 Auxiliary specifications

Our methodology - from a general perspective - is not restricted to (4) as the auxiliary model. We consider (4) as well as three other specifications, the first is a VAR with Leads and Lags [**VARLL**]:

$$\mathbf{\Gamma}^\dagger(L,\vartheta,p,q)Y_t = z^\dagger(\vartheta) + u_t, \tag{5}$$

where $\mathbf{\Gamma}^\dagger(L,\vartheta,p,q) = \sum_{i=-q}^{p} \Gamma^\dagger{}_i(\vartheta)L^i$, $\Gamma^\dagger_0(\vartheta) = \mathbf{I}$, q is the maximum number of leads, and $z^\dagger(\vartheta)$ is a constant. In this case, the future values of Y_t may have an impact on current Y_t, which may also better approximate the VARMA solution. We next consider a VECM as an alternative auxiliary specification:

$$\mathbf{\Phi}(L,\vartheta,p)\Delta Y_t = c(\vartheta) + \mathbf{\Phi_\Psi}(\vartheta)(\mathbf{\Psi}(\vartheta)LY_t) + u_t, \tag{6}$$

where $\mathbf{\Phi}(L,\vartheta,p) = \sum_{i=0}^{p-1} \Phi_i(\vartheta)L^i$, $\Phi_0(\vartheta) = \mathbf{I}$, $c(\vartheta)$ is a constant, $\mathbf{\Phi_\Psi}(\vartheta)(\mathbf{\Psi}(\vartheta)LY_t)$ is the error correction term, and $\mathbf{\Psi}(\vartheta)$ may or may not depend on ϑ. In line with the VAR case, we also consider a VECM with Leads and Lags [**VECMLL**]

$$\mathbf{\Phi}^\dagger(L,\vartheta,p,q)\Delta Y_t = c^\dagger(\vartheta) + \mathbf{\Phi}^\dagger_{\mathbf{\Psi}}(\vartheta)(\mathbf{\Psi}^\dagger(\vartheta)LY_t) + u_t, \tag{7}$$

where $\mathbf{\Phi}^\dagger(L,\vartheta,p,q) = \sum_{i=-q}^{p-1} \Phi^\dagger{}_i(\vartheta)L^i$, $\Phi^\dagger{}_0(\vartheta) = \mathbf{I}$, $c^\dagger(\vartheta)$ is a constant, $\mathbf{\Phi}^\dagger_{\mathbf{\Psi}}(\vartheta)(\mathbf{\Psi}^\dagger(\vartheta)LY_t)$ is the error correction term.

In contrast to Dufour et al. (2013) who suggest that Heteroskedastic and Autocor-

relation Consistent [**HAC**] procedures may be used to correct for the truncation lags, we use (4)-(7) so that the order p, q does not need to reflect a correct specification. Our methodology can accommodate exact unit roots, in addition to highly persistent stationary data. For ease of the presentation, we use (4) as an example to present our methodology below.

3 Inference Methodology

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

$$H_0(\vartheta_0) : \vartheta = \vartheta_0, \qquad \vartheta_0 \qquad \text{known.} \tag{8}$$

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 Monte Carlo simulations to N so that $\alpha^\star(N+1)$ is an integer, we obtain MCT 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. 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. We apply the Genetic Algorithm for this purpose.[9]

[9]See Khalaf and Lin (2015) for further insights and references on the performance of this opti-

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.

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.

3.1 Test statistics

The VAR(p) process in (4) is in the form of a Seemingly Unrelated Regressions [**SUR**] model, where each equation has the same explanatory variables. Hence $\mathbf{\Gamma}(L, \vartheta, p)$ and $z(\vartheta)$ can be estimated equation-by-equation disregarding underlying restrictions using OLS applied to (4) with the observed data. Let $\hat{\mathbf{\Gamma}}(L, \vartheta, p)$ and $\hat{z}(\vartheta)$ denote the OLS-based estimate over the $n^\star$ equations.

Next, obtain a population counterpart to $\hat{\mathbf{\Gamma}}(L, \vartheta, p)$ and $\hat{z}(\vartheta)$ which we will denote $\bar{\mathbf{\Gamma}}(L, \vartheta_0, p)$ and $\bar{z}(\vartheta_0)$. The following simulation-based algorithm is applied using the state-space DSGE model (1)-(3) imposing $\vartheta = \vartheta_0$.

Algorithm 1 calibrated counterparts of auxiliary parameters

1: For a given ϑ_0, generate $m = 1, ..., M$ simulated series $\{\tilde{Y}_{t,m}(\vartheta_0)\}_{t=1}^T$ using the state-space DSGE model (1)-(3) and random draws $\{\tilde{\varepsilon}_{t,m}\}_{t=1}^T$ compatible with (3);

2: Calculate the equation-by-equation OLS estimates using each simulated data set and collect them in $\tilde{\mathbf{\Gamma}}_m(L, \vartheta_0, p)$ and $\tilde{z}_m(\vartheta_0)$;

3: The average over the M paths

$$\begin{aligned} \bar{\mathbf{\Gamma}}(L, \vartheta_0, p) &= \frac{1}{M}\sum_{m=1}^{M} \tilde{\mathbf{\Gamma}}_m(L, \vartheta_0, p), \\ \bar{z}(\vartheta_0) &= \frac{1}{M}\sum_{m=1}^{M} \tilde{z}_m(\vartheta_0) \end{aligned}$$

provides our model consistent proxy for $\mathbf{\Gamma}(.)$ and $z(.)$.

mization method.

Discrepancies between $\{\hat{\mathbf{\Gamma}}(L,\vartheta,p),\hat{z}(\vartheta_0)\}$ and $\{\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0)\}$ will serve to assess $H_0(\vartheta_0)$. We use the LR distance measure to assess this discrepancy:

$$\Lambda(\vartheta_0) = |\hat{\Sigma}_W^0|/|\hat{\Sigma}_W| \tag{9}$$

where

$$W_t(Y,\vartheta) = \hat{\mathbf{\Gamma}}(L,\vartheta,p)Y_t - \hat{z}(\vartheta), \quad \hat{\Sigma}_W = \frac{1}{T-p-1}\sum_{t=p+1}^{T} W_t(Y,\vartheta)W_t'(Y,\vartheta), \tag{10}$$

$$W_t(Y,\vartheta_0) = \bar{\mathbf{\Gamma}}(L,\vartheta_0,p)Y_t - \bar{z}(\vartheta_0), \quad \hat{\Sigma}_W^0 = \frac{1}{T-p-1}\sum_{t=p+1}^{T} W_t(Y,\vartheta_0)W_t'(Y,\vartheta_0), \tag{11}$$

so $\hat{\Sigma}_W^0$ and $\hat{\Sigma}_W$ give the constrained (imposing $\vartheta = \vartheta_0$) and unconstrained sum of squared errors matrices. This statistic 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 MCT alternative. The F-approximation proceeds as follows:

$$\begin{aligned}
\mathcal{L}(\vartheta_0) &= \left(\frac{\mu\tau - 2\lambda}{(p+1)n^\star}\right)\frac{1-(\Lambda(\vartheta_0))^\tau}{(\Lambda(\vartheta_0))^\tau}, \\
\mu &= (T-p-1) - \frac{n^\star - p}{2}, \\
\lambda &= \frac{n^\star(p+1)-2}{4}, \\
\tau &= \begin{cases} \left[\frac{(p+1)^2 n^{\star 2}-4}{(p+1)^2+n^{\star 2}-5}\right]^{1/2}, & if \quad (p+1)^2 + n^{\star 2} - 5 > 0; \\ 1, & otherwise, \end{cases}
\end{aligned} \tag{12}$$

where $\mathcal{L}(\vartheta_0)$ has an approximate $F\left((p+1)n^\star, \mu\tau - 2\lambda\right)$ null distribution at level $\alpha^\star$.

3.2 Finite-sample p-values

The null distribution of $\Lambda(\vartheta_0)$, as well as the statistics introduced above, can be easily simulated, which justifies the application of MCTs [Dufour (2006), Beaulieu et al. (2014), Khalaf and Peraza (2014a,b) and Khalaf and Saunders (2014)]. When applied

to the above $\Lambda(\vartheta_0)$ statistic, for example, the MCT technique can be summarized as follows.

Algorithm 2 MCT technique

1: Obtain the simulation-based estimates underlying the considered statistic. Specifically, we implement the procedure described in Algorithm 1 of section 3.1 to obtain $\bar{\mathbf{\Gamma}}(L, \vartheta_0, p)$ and $\bar{z}(\vartheta_0)$ given ϑ_0. This population measure is generated only once, so the following steps are conditional on $\bar{\mathbf{\Gamma}}(L, \vartheta_0, p)$ and $\bar{z}(\vartheta_0)$.

2: Applying (10)-(11) and (9) to the data, find the observed value of the considered test statistic and denote it as $S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))$, where conditioning on the calibrated $\bar{\mathbf{\Gamma}}(.)$ and $\bar{z}(.)$ is emphasized.

3: Draw N *i.i.d.* samples of size T from the model (1)-(3) under $\vartheta = \vartheta_0$; these draws should be independent from those underlying $\bar{\mathbf{\Gamma}}(.)$ and $\bar{z}(.)$.

4: Using the same population measure as in $S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))$, and applying (10)-(11) and (9) to the simulated data, obtain N simulated values for the considered test statistic, denoted $S_1(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))$, ... , $S_N(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))$, where again, conditioning on the same value used for $\bar{\mathbf{\Gamma}}(.)$ and $\bar{z}(.)$ used in $S_0(.)$ is emphasized.

5: Compute a simulated p-value for the test statistic, using the rank of the observed statistic, relative to its simulated counterpart:

$$p_N(S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))\,) = \frac{NG_N(S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))\,) + 1}{N+1}\,, \tag{13}$$

$$G_N(S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))\,) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\left(S_i(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0)) \geq S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))\right)\,, \tag{14}$$

where $\mathbf{1}(C)$ is the indicator function associated with event C:

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

In other words, $NG_N(S_0(\bar{\mathbf{\Gamma}}(L, \vartheta_0, p), \bar{z}(\vartheta_0))\,)$ is the number of simulated values greater than or equal to S_0.

The null hypothesis is rejected at level $\alpha^\star$ by the test considered if the MCT

p-value so obtained is less than or equal to $\alpha^\star$. The MCT critical region is:

$$p_N(S_0\left(\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0)\right)) \leq \alpha^\star\,, 0<\alpha^\star<1,$$

which is exact for finite T and N, in the following sense.

Theorem 1. *In the context of model (1)-(3) under the null hypothesis (8), consider the test statistic (9) where $\bar{\mathbf{\Gamma}}(L,\vartheta_0,p)$ and $\bar{z}(\vartheta_0)$ are derived conforming with Algorithms 1 and $p_N(S_0\left(\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0)\right))$, its MCT p-value (13) derived using Algorithms 2. Then for finite T and N*

$$\mathsf{P}\left[p_N(S_0(\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0))\,) \leq \alpha^\star\right] = \alpha^\star \tag{15}$$

provided $\alpha^\star(N+1)$ is an integer.

Proof: Step (1) in Algorithm 2 ensures that observed statistic $S_0(\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0))$ and its simulated counterparts $S_1(\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0))$, ... , $S_N(\bar{\mathbf{\Gamma}}(L,\vartheta_0,p),\bar{z}(\vartheta_0))$ are exchangeable, since they rely on the same approximated population measure $\bar{\mathbf{\Gamma}}(L,\vartheta_0,p)$ and $\bar{z}(\vartheta_0)$. The considered statistics are also continuous. Thus (15) follows from Proposition 2.2 of Dufour (2006).■

Theorem 1 proves finite sample exactness regardless of: (i) the VAR truncation order and other possible mis-specification of the auxiliary model; (ii) the curse of dimensionality[10]; (iii) dynamic auxiliary models where lags violate the fixed-regressor assumption required for most exact finite-sample multivariate methods including the approximation used by Dufour et al. (2013); and, most importantly, (iv) persistence. We also note that a linearized state-space is not necessary as long as model consistent simulated data can be obtained given ϑ_0. This property circumvents important deficiencies with available methods.

The fact that exchangeability is sufficient to establish (15) from a finite-sample perspective is worth pointing out here since the statistic we propose relies on just one preliminary simulation. Exactness holds for all other statistics and all other auxiliary specifications, which we also estimate, using observed and calibrated data, by OLS; the error correction component in the VECM auxiliary model, that is $\mathbf{\Psi}(\vartheta)LY_t$ in

[10]When $n^\star$ is large relative to T, asymptotic approximations usually break down because of degrees-of-freedom crunches. Our methodology is valid for any $n^\star$ as long as the above defined determinants are regular.

(6), is usually imposed conforming with theory. Alternatively, if an expression for $\mathbf{\Psi}(\vartheta)LY_t$ is unavailable analytically, reduced rank regression estimation as in *e.g.* Johansen (1988, 1991) may be used instead of OLS, on observed and calibrated data. The models we consider below provide a tractable form for $\mathbf{\Psi}(\vartheta)LY_t$.

4 Simulation Evidence

In this section, we illustrate the properties of our MCT method via a simulation study conducted on two different prototypical DSGE models – a real business cycle model with flexible prices and a New Keynesian DSGE model with sticky prices. Although our method can accommodate non-linear solutions of DSGE models, the current exercise considers linearized solutions as they continue to be used most frequently in the literature and particularly in available work on identification-robust methods.

This exercise achieves three purposes. First, we establish that our method is (a) finite-sample exact and (b) persistence-robust. That is, our method provides correct size irrespective of the sample length or the degree of persistence for the time series processes involved. To show this, we generate data from the model under a null or "true" value of the underlying deep parameters, and analyze rejection frequencies using our method under the null as well as alternative parameter values. We then compare them to the rejection frequencies of an asymptotically based method, for which we pick the asymptotically F-distributed [denoted by **LRAsy**] statistic of Dufour et al. (2013). We show that our method provides correct size, while the asymptotically based method provides size distortion for (a) small samples, and (b) persistent shock processes.

Second, we demonstrate that our methodology can accommodate different auxiliary representations of the DSGE model. In particular, we show that DSGE models that have a VECM representation can be represented as such. VECM forms of DSGE models have received very little attention in the literature (with the important exception of Del Negro et al. (2007)). To the best of our knowledge, this is the first paper that demonstrates the relevance of VECM forms in the identification of DSGE model parameters. Moreover, we add leads and lags to auxiliary VAR or VECM forms and find that they contain important information on deep parameters related to structural shocks.

Finally, we shed light on the inherent non-linearities in the identification of DSGE

models. Specifically, the identifiability of deep parameters depend on the "true value" of these parameters. Identification of one parameter can vary depending on the "true value" of that or another deep parameter.

4.1 A Real Business Cycle Model with Permanent Shocks

We choose a prototypical RBC model from Del Negro and Schorfheide (2013) with flexible prices that is especially suited for our purposes. First, the model can accommodate both a temporary technology shock under a deterministic trend and a permanent one under a stochastic trend. This makes it particularly useful in demonstrating the performance of our method under both stationary and unit root processes. Second, the model provides explicit cointegration relationships between multiple variables from theory that permit a VECM representation alongside the usual VAR form. Finally, the lack of price stickiness means that the chosen model does not suffer from any additional identification problems documented for the New Keynesian case in the next section and in the literature (cf. Mavroeidis (2010) and Canova and Sala (2009)).

The optimality conditions that constitute the model can be summarized as follows:

$$\frac{1}{C_t} = \beta E\left[\frac{1}{C_{t+1}}\left(R_{t+1}+(1-\delta)\right)\right], \tag{16}$$

$$\frac{1}{C_t}W_t^\star = \frac{1}{B_t}\left(\frac{H_t}{B_t}\right)^{1/\nu}, \tag{17}$$

$$K_{t+1} = (1-\delta)\,K_t + I_t, \tag{18}$$

$$Y_t^\star = (A_t H_t)^\alpha\, K_t^{1-\alpha}, \tag{19}$$

$$W_t^\star = \alpha\frac{Y_t^\star}{H_t}, \tag{20}$$

$$R_t = (1-\alpha)\,\frac{Y_t^\star}{K_t}, \tag{21}$$

$$Y_t^\star = C_t + I_t. \tag{22}$$

Here, C_t, I_t and $Y_t^\star$ represent consumption, investment and output, respectively. R_t and $W_t^\star$ represent returns on capital and wages. K_t and H_t are capital and labour inputs into production, A_t represents aggregate technology, and B_t is an exogenous shock to labour supply. Equation (16) summarizes the representative agent's intertemporal consumption trade-off. Equation (17) determines labour supply. Equa-

tion (18) describes the capital accumulation. Equation (19) defines the production function. Equations (20) and (21) summarize the firms' optimal conditions for labour and capital demand. Finally, equation (22) specifies the resource constraint.

Log technology moves according to the following:

$$\begin{aligned} \ln A_t &= \ln A_0 + (\ln \gamma)\, t + \ln \tilde{A}_t, \\ \ln \tilde{A}_t &= \rho_a \ln \tilde{A}_{t-1} + \sigma_a \epsilon_{a,t}, \end{aligned}$$

where $\rho_a \in [0,1]$ and $\epsilon_{a,t} \sim i.i.d.N(0,1)$. This allows us two cases. If $0 \leq \rho_a \leq 1$, technology is trend stationary. On the other hand, if $\rho_a = 1$, $\ln A_t$ is a random-walk process with drift.

Exogenous labour supply shifts are assumed to follow a stationary process:

$$\ln B_t = (1-\rho_b) \ln B_* + \rho_b \ln B_{t-1} + \sigma_b \epsilon_{b,t}$$

where $0 \leq \rho_b \leq 1$, and $\epsilon_{b,t} \sim i.i.d.N(0,1)$. The model is detrended by dividing all trending variables by A_t, log-linearized, and solved using methods outlined in Sims (2002).

If $\rho_a = 1$, then the log-linearized technology variable is nonstationary. In this case, the model implies the following cointegration relationship:

$$\begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} \ln GDP_t \\ \ln I_t \end{bmatrix} = \ln \left[\frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta} \right] + \hat{I}_t - \hat{Y}_t^{\star}$$

where variables with hats denote log deviations around their detrended steady-state. Further details of the assumptions of the model and treatments regarding detrending and log-linearization can be found in Del Negro and Schorfheide (2013).

Collect all parameters of interest in ϑ, where

$$\vartheta = [\alpha, \gamma, \nu, \rho_a, \rho_b, \sigma_a, \sigma_b]'.$$

Figures 1 through 6 plot the rejection-frequency curves at the 95% confidence level for our MCT method for different sample sizes and different degrees of persistence in shock processes. For comparison, we also include rejection frequencies of the asymptotically based method in Dufour et al. (2013), which we denote by **LRAsy**.

Rejection-frequency curves are generated by varying the grid value of the parameters one at a time, while keeping all other at their null values. The null values are provided in the title of each figure with a subscript zero, and have been chosen from estimated posterior means reported in Del Negro and Schorfheide (2013). The rejection frequency of a parameter at their null value gives the size of the test, while the rejection frequency of the parameter away from the null value gives the power.

Figures 1 through 4 correspond to a sample size of 100. Many important quarterly time series for different countries begin only after the mid 1980's. Given the specific nature of the Great Recession of 2008, many studies often consider data prior to the great recession. Together, this restricts the available sample size to about 100. Moreover, many studies focus only on a subset of all available data. For example, there is an active literature focusing on comparing monetary policy across the "pre-Volcker" (1954:Q3 - 1879:Q4) and the "post-Volcker" or "Great Moderation" periods (1986:Q1 - 2007:Q4). Each of these periods correspond to a sample length close to 100. In contrast, Figures 5 and 6, consider a sample size of 300 – about one-and-half times the sample available for the Post-War U.S. economy.

A number of observations can be made. First, the MCT provides correct size regardless of the sample length. In contrast, the LRAsy method (solid lines with circles for the VECM case and dashed lines with circles for the VAR case) provides significant size distortions for empirically relevant small samples. However, Figure 5 shows that the LRAsy method approaches the correct size when the sample is large enough (albeit empirically irrelevant).

More importantly, the MCT method retains the correct size in the face of severe persistence. Figure 6 shows that even for the unit root case of $\rho_a = 1$, the MCT method provides correct size. In contrast, size in LRAsy is distorted when the persistence of the shock process (magnitude of ρ_a) is high.

Second, different auxiliary forms of the DSGE model retain different levels of information relating to the identifiability of a parameter. Figure 2 shows rejection frequencies using the MCT method for four different auxiliary formulations of the same DSGE model – a VECM form, a VAR form, and both of these enhanced with leads and lags.

Again, a number of observations can be made. First, the VECM formulation is more informative on the identification of certain parameters compared to the VAR formulation. That is, the rejection probability under the alternative (power of the

procedure) is higher for the VECM case than the VAR case. Not surprisingly, the advantage of the VECM manifests itself in parameters relevant to the common trend – the time trend parameter γ, the shock persistence parameters ρ_a and ρ_b, and shock standard error σ_b, compared to the VAR formulation. To the best of our knowledge, this is the first paper that highlights the importance of VECM formulations in the identification of DSGE model parameters.

Moreover, adding leads and lags to either the VAR or the VECM formulation (represented by dashed and solid lines with diamonds, respectively) provides important information on the identification of parameters, specifically those related to the standard deviation of shocks.

Finally, identification of a particular parameter depends crucially on the "true value" of that or another parameter. As long as they are produced under the same method, rejection-frequency curves can serve as a graphical device for measuring the relative depth of the identification problem inherent in parameter values of different models. If the rejection frequency of a parameter is low regardless of the distance from the "true value" from which data is generated, we can say that the parameter is hard to identify. In this light, compare Figures 1 and 2. An increase in the persistence of the shock processes ρ_a and ρ_b not only makes it harder to identify the persistence parameters in question, but also makes in harder to reject "wrong values" of parameters α and γ.

4.2 A New Keynesian Model with Sticky Prices

The New Keynesian model with sticky prices, inflation inertia and habit formation is widely considered to be more empirically relevant than the flexible price model studied above. Versions of the three equation New Keynesian DSGE model has been widely estimated by proponents of Bayesian and frequentist economists alike. Moreover, complications related to parameter identification in New Keynesian DSGE models have recently received significant attention in the literature (Canova and Sala (2009) Mavroeidis (2010) Cochrane (2011), Dufour et al. (2013), among others), making it particularly interesting and relevant for our case.

In this section, we conduct the same simulation study on a version of the New Keynesian model derived and estimated in Milani and Treadwell (2012). For book-length treatments of the model, see Woodford (2003) and Galí (2008).

The model consists of a New Keynesian Phillips Curve [**NKPC**], an Euler-equation based modified IS function, a Taylor rule followed by the monetary authority, and exogenous shock processes, summarized in log-linearized form below:

$$\begin{aligned}
\pi_t &= \frac{\beta}{1+\beta\gamma}E_t\pi_{t+1} + \frac{\gamma}{1+\beta\gamma}\pi_{t-1} \\
&\quad + \frac{(1-\alpha)(1-\alpha\beta)}{\alpha(1+\beta\gamma)}\left(\omega y_t + \frac{\sigma^{-1}}{1-\phi}\left(y_t - \phi y_{t-1}\right)\right) + \mu_t, \qquad (23)\\
y_t &= \frac{1}{1+\phi}E_t y_{t+1} + \frac{\phi}{1+\phi}y_{t-1} - \frac{\sigma(1-\phi)}{1+\phi}\left(R_t - E_t\pi_{t+1}\right) + g_t, \qquad (24)\\
R_t &= \rho R_{t-1} + (1-\rho)\left[\chi_\pi \pi_t + \chi_y y_t\right] + \zeta_t, \qquad (25)\\
\mu_t &= \rho_\pi \mu_{t-1} + \varepsilon_{\pi,t}, \qquad (26)\\
g_t &= \rho_y g_{t-1} + \varepsilon_{y,t}, \qquad (27)\\
\zeta_t &= \rho_R \zeta_{t-1} + \varepsilon_{R,t}. \qquad (28)
\end{aligned}$$

All variables are expressed as log deviations from their respective stationary steady-states. y denotes the output gap, π denotes inflation, and R denotes the nominal interest rate. The parameter γ represents the degree of price indexation to past inflation, while α represents the fraction of firms that cannot re-optimize 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 ζ are exogenous disturbances to the marginal cost of production, the Euler function, and the systematic policy rule. They evolve according to AR(1) processes with autoregressive coefficients ρ_i, for $i = \pi, y, R$, 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 identification-robust estimation of three-equation DSGE models focus on "semi-structural" versions of the above, 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".

Collect all parameters of interest in ϑ, where

$$\vartheta = [\phi, \sigma, \gamma, \alpha, \rho, \chi_\pi, \chi_y, \rho_\pi, \rho_y, \rho_R, \sigma_\pi, \sigma_y, \sigma_R]'. \tag{29}$$

Figures 7 and 8 show rejection frequencies for this model at the 95% confidence level similar to the previous subsection. The null values of the parameters, denoted with a zero subscript and noted in the titles of the sub figures, are chosen from posterior mean values estimated in Milani and Treadwell (2012). Since there is no cointegration relationship between the three variables considered in the model, only VAR forms of the auxiliary model is considered. The dashed and solid lines represent the MCT case with a VAR and VAR with leads and lags, respectively. The dashed lines with circles represent the LRAsy case with a VAR.

The main messages from the previous section carry through. Asymptotically based methods spuriously reject "true values" of the model parameter for finite samples. The size distortion worsens with the persistence of the underlying shock processes. This is shown starkly in Figure 8, where the shock persistence parameters ρ_y and ρ_R have been increased to near unit root. In contrast, the MCT method provides correct size regardless of sample length or persistence.

Augmenting the VAR auxiliary form with leads and lags provide better identification of shock standard deviations. It also provides for better identification of all parameters under persistent shocks (Figure 8). Finally, comparing Figures 7 with 8, we see that changing the "true value" of some parameters (in this case the shock persistence parameters) can worsen the identification of other parameters (for example the Taylor rule response parameters χ_π and χ_y).

In conclusion, we demonstrate in this section that (i) asymptotically motivated identification-robust methods can generate spurious rejections for the model, especially when sample length is small or when persistence of the underlying processes is high. In contrast, the MCT method retains correct size regardless of sample length or degree of persistence. (ii) VECM formulations and adding leads and lags in the auxiliary representation of DSGE models can be informative for the identification of

important parameters. And (iii) the inherent non-linearity of DSGE model parameters imply that the identification of one parameter can be deeply influenced by the "true value" of that or any other parameter in the model.

5 How does it work in practice?

Having demonstrated the properties of our method using a simulation study in the previous section, we now consider a practical application. We choose an application related to the literature on identification of monetary policy stance before and during the period of the "Great Moderation". Sample lengths available to address this issue is small and thus is susceptible to spurious over-rejections, as shown above.

The "Great Moderation" refers to the period between mid-1980's and the beginning of the global recession of 2008, when U.S. macroeconomic volatilities were much lower compared to other periods (Stock and Watson (2003)). Clarida et al. (2000) proposes that the decline in volatility is due to the U.S. monetary authority's adoption of an aggressive stance against inflation in the early 1980's.[11] 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 justified identification-robust settings.

We revisit this question in light of our findings from the previous section on the possibility of spurious rejections inherent in asymptotically justified methods. Specifically, we estimate and compare the upper and lower-bounds of the 95% confidence sets of the model parameters for the "pre-Great Moderation" subsample of 1954Q3 through 1985Q4, and "Great Moderation" subsample of 1986Q1 through 2007Q4.

Two 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 given below. The left hand side variables describe

[11] 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.

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.

Second, we employ a Genetic Algorithm to find the upper and lower bounds of a 95% confidence set by projecting the accepted sets one parameter at a time. 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 (ii) the simulated p-value implies a rejection at the 5% level. Khalaf and Lin (2015) show that the Genetic Algorithm provides a significantly more accurate result compared with *e.g.* grid search methods employed by Mavroeidis (2010) or Castelnuovo and Fanelli (2015).

Our results are stark. The model is rejected at the 95% confidence level when we use a VAR with leads and lags as an auxiliary representation. In contrast, the model is accepted when we use a simple VAR as auxiliary representation. However, the accepted parameter intervals are too large to provide any useful information. In other words, the model jumps from well identified but rejected to accepted but badly identified, depending on the auxiliary form used.

Two caveats should be kept in mind when interpreting these results. First, rejection or weak identification of the *specific* three-equation model provided in equations (23) through (28) does not imply a rejection of the building blocks of the New Keynesian theory. Second, different methods of bridging raw data to their DSGE model counterparts can have important implications for the estimated parameters. 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 three-equation DSGE model in the two-subsample periods, depending on whether the data is first differenced, de-trended, or matched with the model through bridge-equations that correct for non-stationarity. In particular, the VAR with leads and lags may be rejecting the measurement equations rather than the model itself.

Finally, as simulation results from the previous section and in Figure 8 suggests, weak identification of model parameters may follow from high persistence of the shock

processes. To gain further insight on this possibility, we look at a subset of the model parameters inspired by the choice of priors considered in Milani and Treadwell (2012) as a special example. In particular, we restrict the standard deviation of two of the shock processes, σ_π, σ_R to be equal to each other and normalize them to one.

The results of the projection-based upper and lower bounds for the VAR case of this special example are given in Tables 1 and 2 for the "pre-Great Moderation" and "Great Moderation" periods, respectively. First, we find that the model is accepted for both subsamples. To be precise, the *determinate* version of the model is accepted for both subsamples.

Second, we find that the model suffers from serious identification problems in both subsamples. For both subsamples, the entire search set is accepted for most parameters. The few parameters that are identified include the shock persistence parameters ρ_y and ρ_R, and the Calvo parameter α. Importantly, accepted sets of these parameters point to high persistence and a flatness in the slope of the NKPC. As shown in the simulation study in the previous section, persistent shock processes are expected to worsen the identification problem in a significant number of model parameters. This point is also made in an analytically tractable version of the model in Mavroeidis (2010). Therefore, the only parameters we find to be identified is expected to result in a lack of identification for the rest of the model parameters.

6 Conclusion

The paper develops a simulation-based finite-sample identification-robust confidence-set estimation method for DSGE models by combining Indirect Inference with the Monte Carlo test method. Confidence sets are obtained by inverting finite-sample tests that assess discrepancies between relevant, though possibly mis-specified, auxiliary forms from their DSGE-restricted population counterparts obtained by simulation. Auxiliary forms considered in this paper include standard and forward and backward looking VARs as well as VECMs. Our method controls coverage exactly, regardless of sample size, model dimension, lag truncation order, identification of deep parameters and persistence of underlying processes and data. A linearized state space solution is not required as long as the DSGE can be simulated once a finite dimensional parameter is specified. Size and power properties and empirical relevance are illustrated via canonical real business cycle and New Keynesian models.

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.

References

Andrews, I. and Mikusheva, A.: 2015, Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models, *Quantitative Economics* **6**(1), 123–152.

Basturk, N., akmali, C., Ceyhan, P. and Van Dijk, H.: 2014, Posterior-predictive evidence on us inflation using extended New Keynesian Phillips Curve models with non-filtered data, *Journal of Applied Econometrics* **29**, 1164–1181.

Beaulieu, M. C., Dufour, J. M. and Khalaf, L.: 2014, Exact confidence sets and goodness-of-fit methods for stable distributions, *Journal of Econometrics* **181**(1), 3–14.

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

Benati, L. and Surico, P.: 2009, VAR analysis and the great moderation, *American Economic Review* **99**(4), 1636–52.

Boivin, J. and Giannoni, M. P.: 2006, Has monetary policy become more effective?, *Review of Economics and Statistics* **88**(3), 445–462.

Calvet, L. E. and Czellar, V.: 2015, Through the looking glass: Indirect inference via simple equilibria, *Journal of Econometrics* **185**(2), 343 – 358.

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

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

Castelnuovo, E. and Fanelli, L.: 2015, Monetary policy indeterminacy and identification failures in the U.S.: Results from a robust test, *Journal of Applied Econometrics* **30**, 924–947.

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

Cochrane, J. H.: 2011, Determinacy and identification with Taylor rules, *Journal of Political Economy* **119**(3), 565 – 615.

Coibion, O. and Gorodnichenko, Y.: 2010, Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts, *NBER working papers*, National Bureau of Economic Research, Inc.

Coibion, O. and Gorodnichenko, Y.: 2015, Is the phillips curve alive and well after all? Inflation expectations and the missing disinflation, *American Economic Journal: Macroeconomics* **7**(1), 197–232.

Del Negro, M. and Schorfheide, F.: 2013, *Bayesian Macroeconometrics*, Oxford University Press.

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

Dominicy, Y. and Veredas, D.: 2013, The method of simulated quantiles, *Journal of Econometrics* **172**(2), 235–247.

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

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

Dufour, J. M. and Kurz-Kim, J. R.: 2010, Exact inference and optimal invariant estimation for the stability parameter of symmetric [alpha]-stable distributions, *Journal of Empirical Finance* **17**(2), 180–194.

Dufour, J. M. and 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, J. M. and Valéry, P.: 2009, Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models, *Journal of Econometrics* **150**(2), 193–206.

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

Fuleky, P. and Zivot, E.: 2014, Indirect inference based on the score, *Econometrics Journal* **17**(3), 383–393.

Galí, J.: 2008, Introduction to Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework, *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Introductory Chapters, Princeton University Press.

Gallant, A. R. and Tauchen, G.: 1996, Which moments to match?, *Econometric Theory* **12**(04), 657–681.

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

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

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

Gouriéroux, C., Phillips, P. C. and Yu, J.: 2010, Indirect inference for dynamic panel models, *Journal of Econometrics* **157**(1), 68–77.

Guay, A. and Scaillet, O.: 2003, Indirect inference, nuisance parameter, and threshold moving average models, *Journal of Business and Economic Statistics* **21**(1), 122–32.

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

Inoue, A. and Rossi, B.: 2011, Identifying the sources of instabilities in macroeconomic fluctuations, *Review of Economics and Statistics* **93**(4), 1186–1204.

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

Johansen, S.: 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* **12**, 231–254.

Johansen, S.: 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* **59**, 1551–1580.

Kapetanios, G., Khalaf, L. and Marcellino, M.: 2016, Factor based identification-robust inference in IV regressions, *Journal of Applied Econometrics* **31**, 821–842.

Khalaf, L. and Lin, Z.: 2015, Statistical test inversion and numerical projections with particle swarm optimization, *mimeo* .

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

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

Khalaf, L. and Saunders, C. J.: 2014, Indirect set inference for dynamic panel models, *mimeo* .

Kleibergen, F. and 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. and 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**(1), 120–128.

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

Lubik, T. A. and Schorfheide, F.: 2004, Testing for indeterminacy: An application to U.S. monetary policy, *American Economic Review* **94**(1), 190–217.

Magnusson, L. M. and Mavroeidis, S.: 2014, Identification Using Stability Restrictions, *Econometrica* **82**, 1799–1851.

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-Mller, M. and Stock, J. H.: 2014, Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve, *Journal of Economic Literature* **52**(1), 124–88.

Mikusheva, A.: 2009, Comment on "weak instrument robust tests in GMM and the new Keynesian Phillips curve", *Journal of Business and Economic statistics* **27**(3), 293–311.

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

Moon, H. R. and Schorfheide, F.: 2012, Bayesian and frequentist inference in partially identified models, *Econometrica* **80**, 755–782.

Qu, Z.: 2014, Inference in dynamic stochastic general equilibrium models with possible weak identification, *Quantitative Economics* **5**, 457–494.

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* D. Acemoglu, M. Arrelano and E. Deckel (eds), *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, Cambridge University Press, Cambridge (UK), pp. 184–230.

Sims, C. A.: 2002, Solving Linear Rational Expectations Models, *Computational Economics* **20**(1-2), 1–20.

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

Stock, J. H. and Watson, M. W.: 2003, Has the business cycle changed?, *Proceedings - Economic Policy Symposium - Jackson Hole* pp. 9–56.

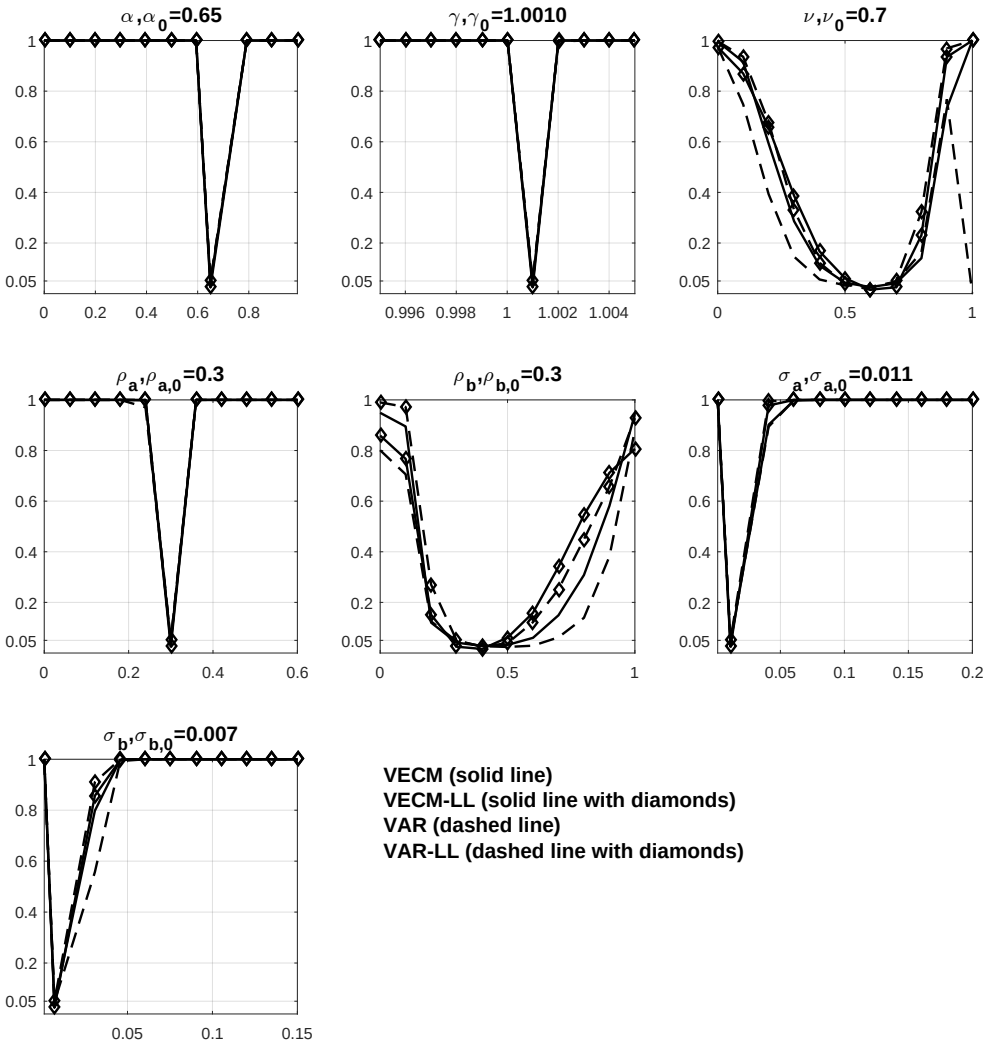
Tkachenko, D. and Qu, Z.: 2012, Identification and frequency domain quasimaximum likelihood estimation of linearized dynamic stochastic general equilibrium models, *Quantitative Economics* **3**, 95–132.

Tkachenko, D. and Qu, Z.: 2016, Global identification in dsge models allowing for indeterminacy, *Review of Economic Studies* p. to appear.

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

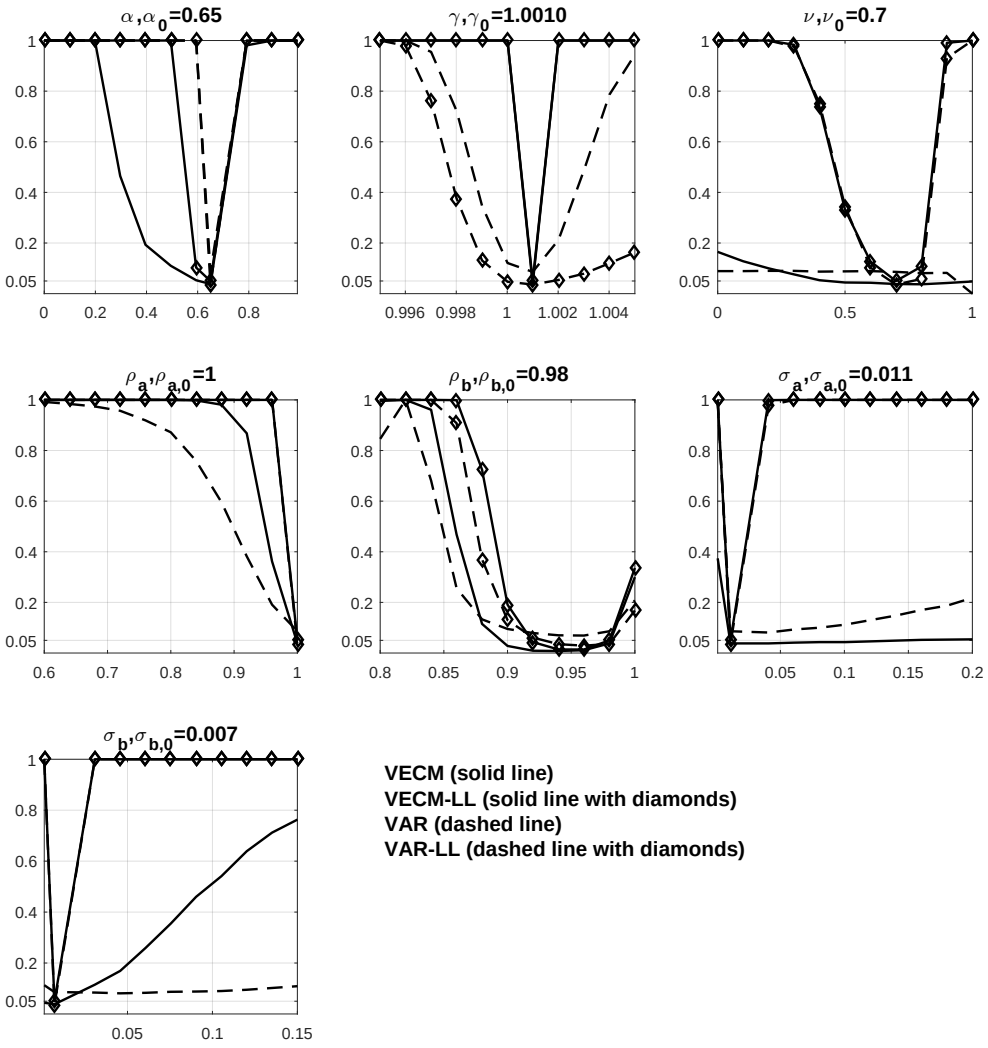
Yang, J. C. and Xu, K. L.: 2014, Estimation and Inference under Weak Identification and Persistence: An Application on Forecast-Based Monetary Policy Reaction Function, *Technical report*.

Figure 1: Rejection frequencies using the MCT method at the 5% significance level for small samples.



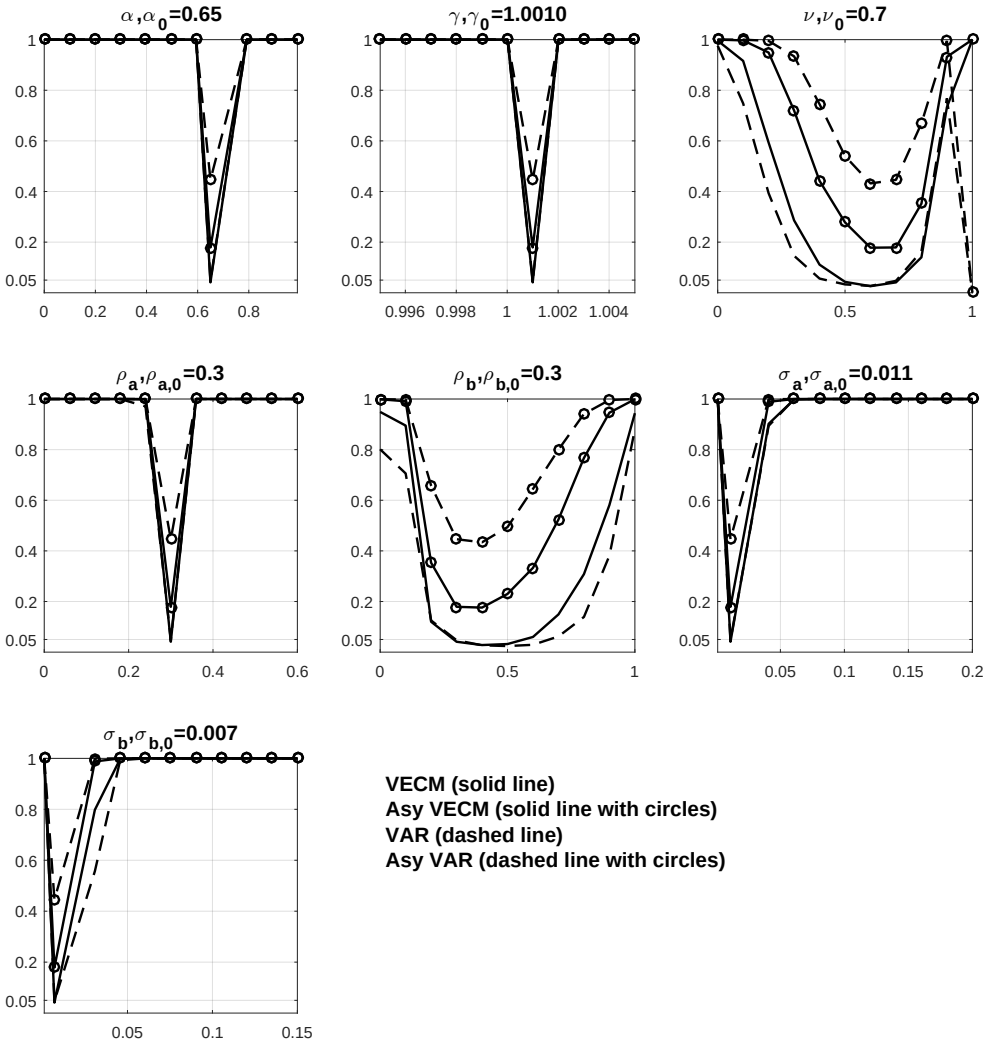
Note: Solid line represents VECM. Solid lines with diamonds represent VECMLL. Dashed lines represent VAR. Dashed lines with diamonds represent VARLL. $T = 100$. $\rho_a < 1$.

Figure 2: Rejection frequencies using the MCT method at the 5% significance level for small samples and unit root shocks.



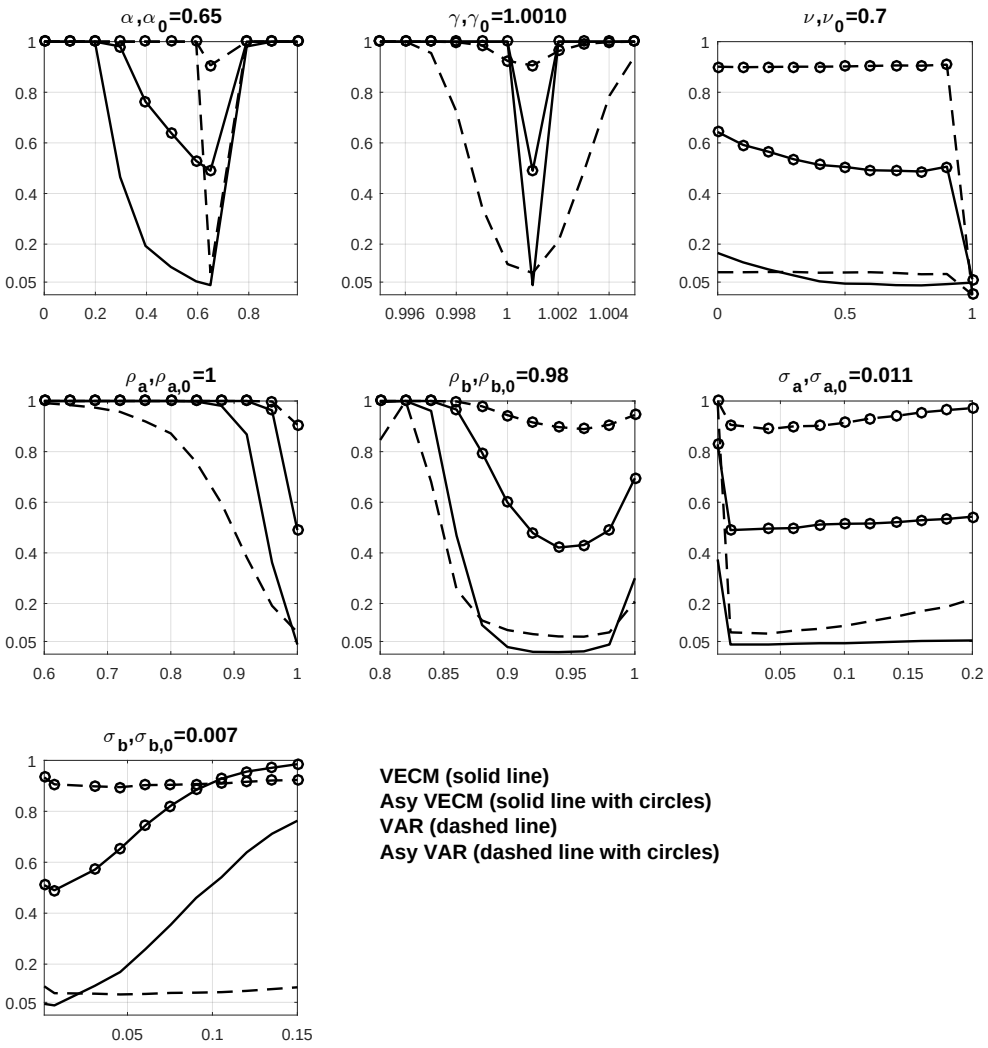
Note: Solid line represents VECM. Solid lines with diamonds represent VECMLL. Dashed lines represent VAR. Dashed lines with diamonds represent VARLL. $T = 100$. $\rho_a = 1$.

Figure 3: Comparison of rejection frequencies between the MCT and LRAsy methods at the 5% significance level for small samples.



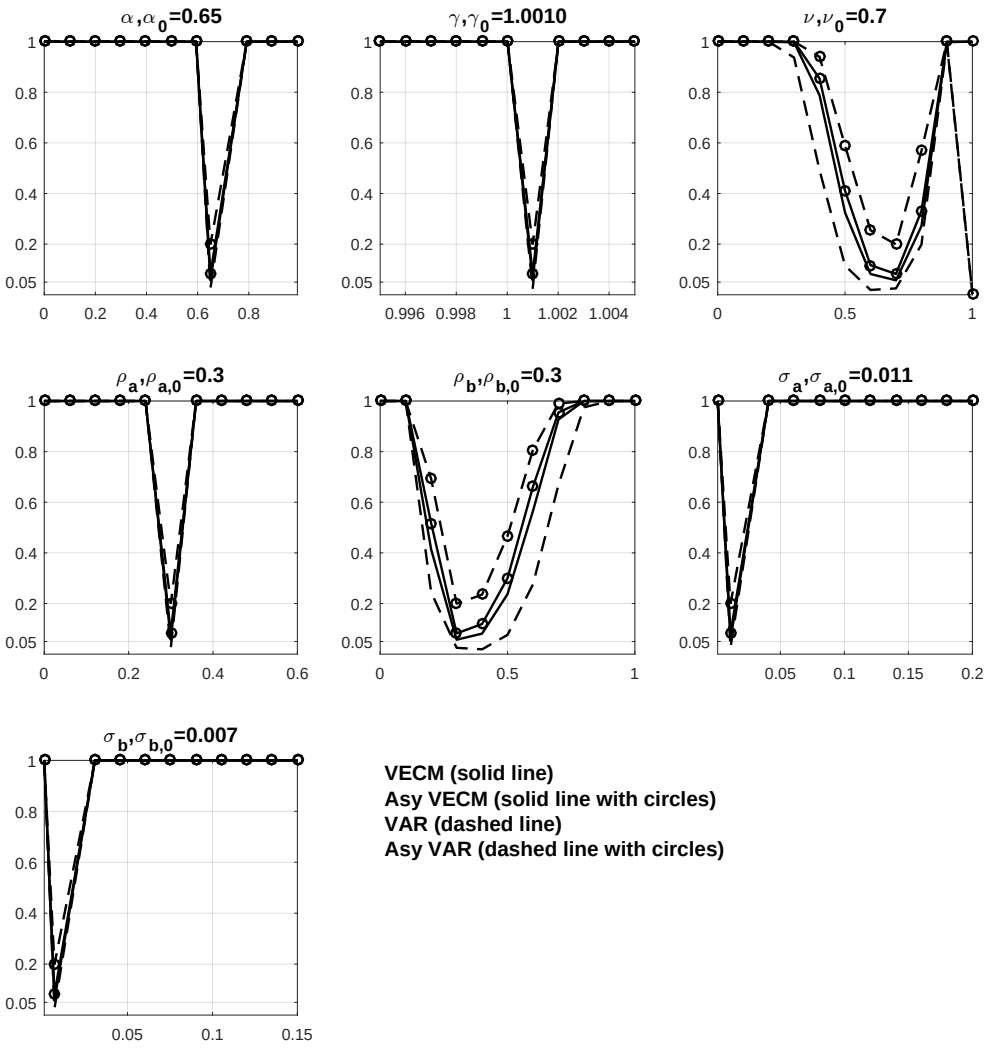
Note: Solid line represents MCT VECM. Solid lines with diamonds represents LRAsy VECM. Dashed lines represent MCT VAR. Dashed lines with diamonds represent LRAsy VAR. $T = 100$. $\rho_a < 1$.

Figure 4: Comparison of rejection frequencies between the MCT and LRAsy methods at the 5% significance level for small samples and unit root shocks.



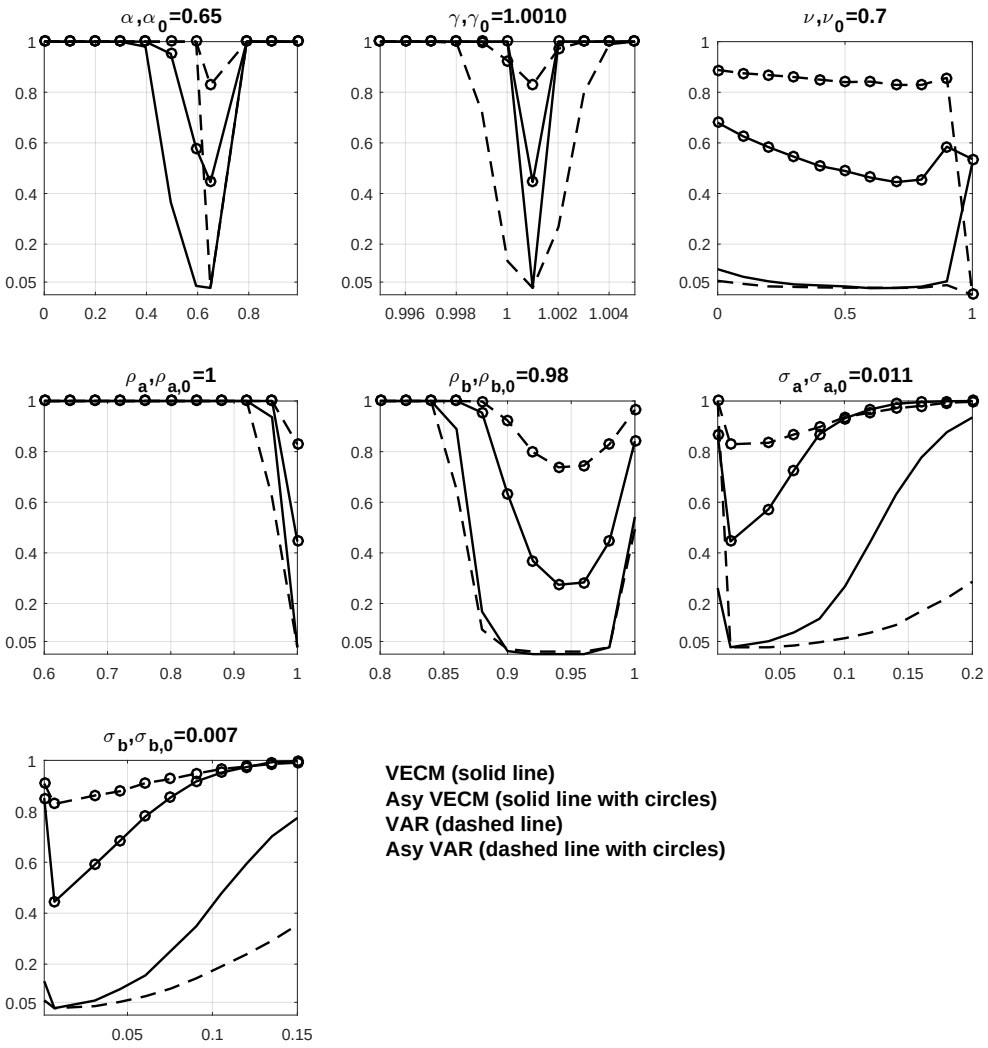
Note: Solid line represents MCT VECM. Solid lines with diamonds represents LRAsy VECM. Dashed lines represent MCT VAR. Dashed lines with diamonds represent LRAsy VAR. $T = 100$. $\rho_a = 1$.

Figure 5: Comparison of rejection frequencies between the MCT and LRAsy methods at the 5% significance level for large samples.



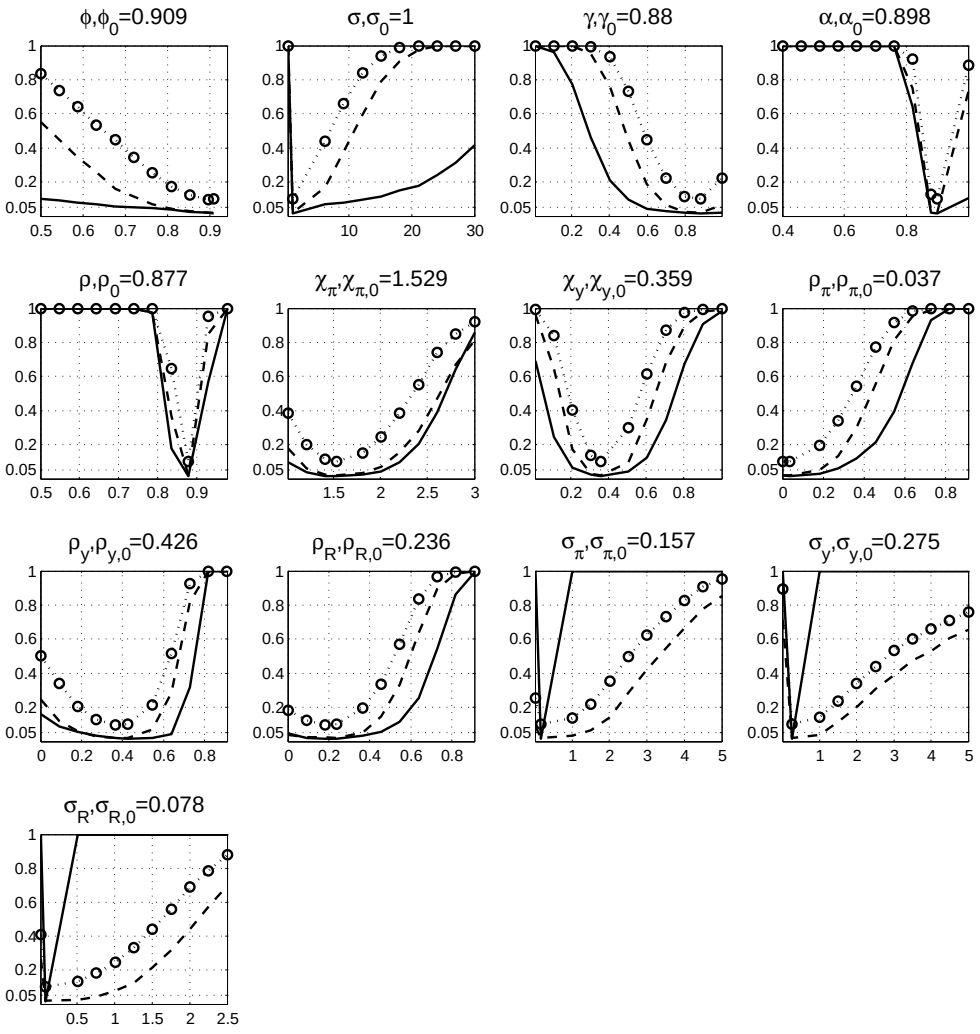
Note: Solid line represents MCT VECM. Solid lines with diamonds represents LRAsy VECM. Dashed lines represent MCT VAR. Dashed lines with diamonds represent LRAsy VAR. $T = 300$. $\rho_a < 1$.

Figure 6: Comparison of rejection frequencies between the MCT and LRAsy methods at the 5% significance level for large samples and unit root shocks.



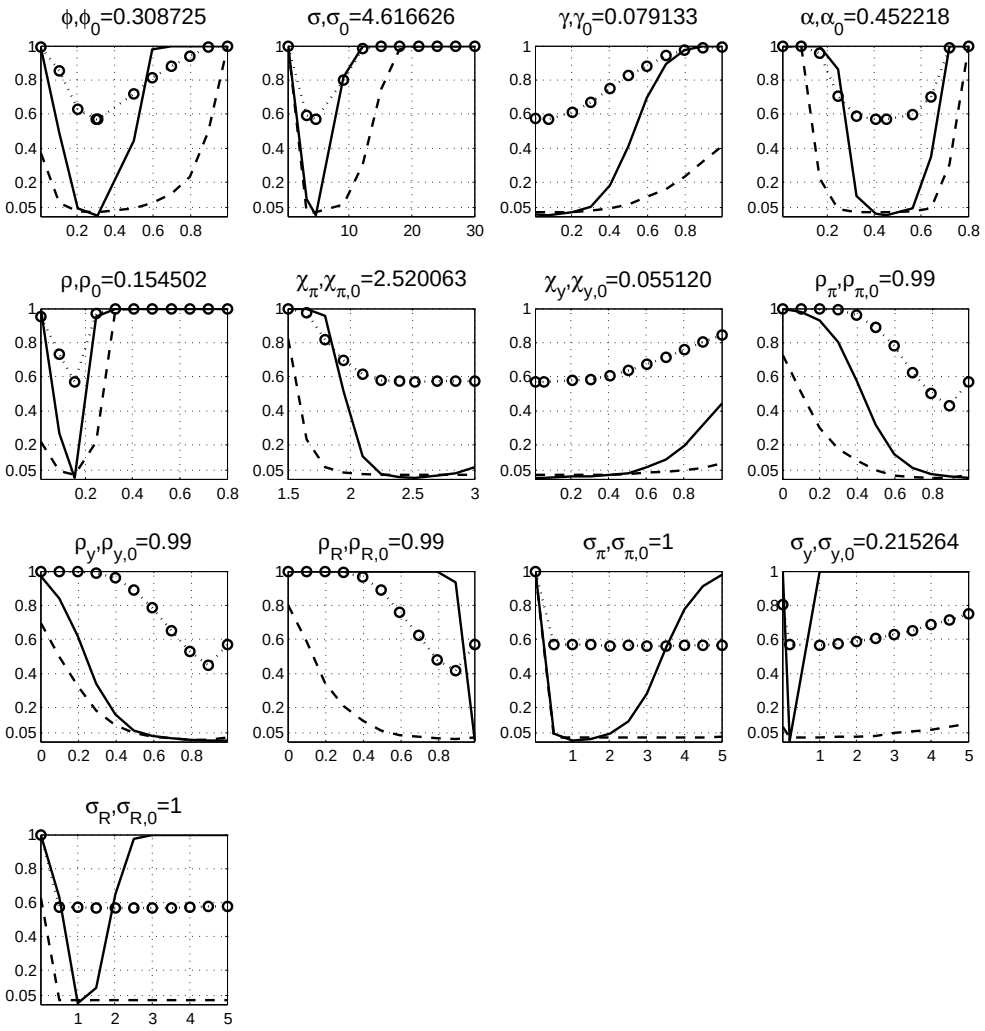
Note: Solid line represents MCT VECM. Solid lines with diamonds represents LRAsy VECM. Dashed lines represent MCT VAR. Dashed lines with diamonds represent LRAsy VAR. $T = 300$. $\rho_a = 1$.

Figure 7: Comparison of rejection frequencies in the New Keynesian model for MCT and LRAsy methods at the 5% significance level for small samples



Note: Dashed lines represent MCT VAR, and solid lines represent MCT VLL, circle-dot lines represent LRAsy VAR. T = 100.

Figure 8: Comparison of rejection frequencies in the New Keynesian model for MCT and LRAsy methods at the 5% significance level for small samples and persistent shocks.



Note: Dashed lines represent MCT VAR, and solid lines represent MCT VLL, circle-dot lines represent LRAsy VAR. T = 100.

Table 1: Projection-based 95% confidence sets, "pre-Great Moderation" data, $\sigma_\pi = \sigma_R = 1$, VAR vs.VLL

Coefficient	Search set	Least-rejected value	VAR 95% CI	VLL 95% CI
γ	[0.01, 0.99]	0.079133	[0.010000, 0.989319]	$\emptyset$
α	[0.01, 0.99]	0.452218	[**0.199834, 0.989899**]	$\emptyset$
σ	[0.20, 5.50]	4.616626	[0.205419, 5.458306]	$\emptyset$
ϕ	[0.01, 0.99]	0.308725	[0.010000, 0.985683]	$\emptyset$
ρ	[0.01, 0.99]	0.154502	[0.010000, 0.961050]	$\emptyset$
χ_π	[0.50, 3.00]	2.520063	[0.500000, 2.915174]	$\emptyset$
χ_y	[0.01, 1.00]	0.055120	[0.010000,1.000000]	$\emptyset$
ρ_π	[0.01, 0.99]	0.990000	[0.181789, 0.990000]	$\emptyset$
ρ_y	[0.01, 0.99]	0.990000	[**0.952433, 0.990000**]	$\emptyset$
ρ_R	[0.01, 0.99]	0.990000	[**0.981949, 0.990000**]	$\emptyset$
$\frac{\sigma_y}{\sigma_R}$	[0.01, 0.99]	0.215264	[0.638714, 0.989949]	$\emptyset$
Sup p-value		0.070000		0.010000

Table 2: Projection-based 95% confidence sets, "Great Moderation" data, $\sigma_\pi = \sigma_R = 1$, VAR vs.VLL

Coefficient	Search set	Least-rejected value	VAR 95% CI	VLL 95% CI
γ	[0.01, 0.99]	0.989752	[0.010000, 0.990000]	$\emptyset$
α	[0.01, 0.99]	0.462407	[**0.144809, 0.990000**]	$\emptyset$
σ	[0.20, 5.50]	1.977300	[0.200000, 5.499904]	$\emptyset$
ϕ	[0.01, 0.99]	0.096445	[0.010010, 0.988959]	$\emptyset$
ρ	[0.01, 0.99]	0.010224	[0.010000, 0.990000]	$\emptyset$
χ_π	[0.50, 3.00]	1.960051	[0.500000, 2.997760]	$\emptyset$
χ_y	[0.01, 1.00]	0.999935	[0.010000, 1.000000]	$\emptyset$
ρ_π	[0.01, 0.99]	0.989997	[0.010000, 0.990000]	$\emptyset$
ρ_y	[0.01, 0.99]	0.990000	[0.025103, 0.990000]	$\emptyset$
ρ_R	[0.01, 0.99]	0.989818	[**0.703749, 0.990000**]	$\emptyset$
$\frac{\sigma_y}{\sigma_R}$	[0.01, 0.99]	0.690177	[0.091410, 0.985605]	$\emptyset$
Sup p-value		0.090000		0.010000