Bootstrapping DSGE models

Giovanni Angelini, Giuseppe Cavaliere, Luca Fanelli

Department of Statistical Sciences & School of Economics, Management and Statistics, University of Bologna

February 2016

PRELIMINARY VERSION

Abstract

This paper explores the potential of bootstrap methods in the empirical evaluation of dynamic stochastic general equilibrium (DSGE) models and, more generally, in linear rational expectations models that feature unobservable (latent) components. We consider two dimensions. First, under the regularity conditions that permit standard likelihood-based inference (henceforth strong identification), the bootstrap Quasi-Maximum Likelihood (QML) estimator of the structural parameters replicates the (Gaussian) asymptotic distribution of the QML estimator. The bootstrap is consistent and we can keep the probability of false rejections of the cross-equation restrictions DSGE models imply on their state space representation under strict control and attenuate the extent of the over-rejection phenomenon that characterizes likelihood-ratio tests. Second, we build a novel, computationally straightforward, bootstrap-based approach to detect misspecification, including weak identification. We show that under strong identification and bootstrap consistency, a test statistic based on a set of bootstrap repetitions of the QML estimator (and associated quasi-likelihood ratio test) approximates the Gaussian (Chi-square) distribution, while when the regularity conditions for inference does not hold as e.g. it happens when (part of) the structural parameters are weakly identified, this result is no longer valid with the proposed test statistic diverging to infinity. Therefore, we can rely on simple normality tests associated with the bootstrap repetitions of the structural parameter estimators in order to evaluate how close or distant is the estimated model from the case of strong identification. Our Monte Carlo experimentations suggest that the bootstrap plays an important role along both dimensions and represents a promising evaluation tool of the cross-equation restrictions and, under certain conditions, of the strength of identification. An empirical illustration based on U.S. quarterly observations on the Great Moderation period is also provided.

Keywords: ABCD form, Bootstrap, Cross-equation restrictions, DSGE, LR test, State space model, Strong identification; Weak identification.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are liner(ized) rational expectations models currently used by central banks and academicians to evaluate macroeconomic policies and predict the stance of the business cycle. These models are stylized representations of the economy and are misspecified in several dimensions (An and Schorfheide, 2007). They are typically treated by econometricians as restricted but parametrically incomplete representations of the actual data (Diebold *et al.* 1998). DSGE models imply highly nonlinear restrictions on the state space representation they generate. These restrictions, hereafter denoted 'cross-equation restrictions' (CER), are the 'hallmark' of rational expectations models (Hansen and Sargent, 1980, 1981) and represent the 'natural' metric through which these models should be evaluated empirically (Hansen, 2014). In the frequentist setting, DSGE models are usually rejected when evaluated through the CER. This is one of the explanations of why practitioners typically prefer the Bayesian approach.

Following the original intuition of Mankiw and Shapiro (1987) who focused on orthogonality restrictions, Bekaert and Hodrick (2001) have shown that in rational expectations models commonly employed tests of the CER based on asymptotic distributions may lead to severe size distortions and power losses in finite samples. In particular, the empirical size of these tests tends to exceed markedly the prefixed nominal type-I error, inducing practitioners to falsely conclude that their models are too simple to capture the complex probabilistic nature of the data. If standard tests behave poorly in small samples, inference based on standard asymptotic distribution theory is no longer reliable and practitioners might falsely reject models that might actually have good empirical contents.

As is known, under certain conditions, bootstrap methods can provide useful inferential refinements (Horowitz, 2001).[1] Cho and Moreno (2006) and Bårdsen and Fanelli (2015) have exploited bootstrap methods for testing the CER implied by monetary DSGE models, documenting substantial finite-sample size improvements under the null. However, these authors consider only cases where all endogenous variables of the system are observed, or can be easily proxied by observables. Moreover, finite-order vector autoregressive (VAR) representations are employed, while it is now well known that VAR representations with a finite number of lags

[1] In macroeconometric analysis, bootstrap methods are typically used either to build confidence bands for impulse response functions computed in structural vector autoregressions (Kilian, 1998), or to obtain confidence intervals for the structural parameters of dynamic macro models (Cho and Moreno, 2006). They can also be conveniently use to determine the cointegration rank and test hypotheses of interest in cointegrated systems, see Cavaliere *et al.* (2012), Cavaliere *et al.* (2015) and Boswijk *et al.* (2015). Only seldom they have been used to improve the small sample performance of tests of the CER in linear(ized) rational expectations models.

are the exception rather than the rule in DSGE models (Ravenna, 2007; Franchi and Vidotto, 2013; Franchi and Paruolo 2014). It is also worth noticing that Cho and Moreno (2006) and Bårdsen and Fanelli (2015) implicitly assume that the regularity conditions which permit standard inference - henceforth generically denoted with the term 'strong identification' - are valid. Conversely, the recent literature suggests that 'sample' or 'weak' identification problems, defined as situations where the likelihood function of the system does not fully respect the usual regularity conditions necessary for standard inference, are an important issue in DSGE models. See Canova and Sala (2009), Dufour *et al.* (2009, 2013), Kleibergen and Mavroeidis (2009), Mavroeidis (2005, 2010), Guerron-Quintana *et al.* (2013), Andrews and Mikusheva (2014), Qu (2014) and Castelnuovo and Fanelli (2015), among others.

In the current literature, little is known about the asymptotic properties of bootstrap methods in DSGE models and how to implement these methods to test the CER. Moreover, little is known on how the bootstrap performs in practice and nothing is known about the performance of the bootstrap in weakly identified DSGE models. In this paper we fill the gap and explore the potential of bootstrap methods in the empirical evaluation of DSGE models, along at least two dimensions.

First, we develop a (time-domain) bootstrap-based approach for quasi-likelihood ratio (QLR) tests of the CER implied by DSGE models and, more generally, by linear(ized) rational expectations models involving unobservable (latent) components. To do so, we extend Stoffer and Wall's (1991) nonparametric bootstrap approach for state space models to the case of DSGE models. In the case of strong identification, we generalize and strengthen Stoffer and Wall's (1991) result by proving formally that the (restricted) bootstrap is consistent when the CER are tested by QLR tests. In this scenario, the asymptotic distribution of the quasi-maximum likelihood (QML) estimator of the structural parameters can be estimated accurately by the bootstrap. Bootstrap confidence intervals for the structural parameters or for impulse responses can be built and interpreted in the conventional way. Importantly, not only the (either standard or bootstrap) QLR test is asymptotically pivotal and chi-square distributed, but the bootstrap also reduces the discrepancy between actual and nominal probabilities of type-I error. It turns out that our bootstrap approach to DSGE models has the potential to attenuate the over-rejection phenomenon that characterizes tests of the CER when first-order asymptotic approximations are used.

Second, we show the novel result that the bootstrap can also be implemented in order to assess how far the estimated DSGE model is from the case of strong identification. Our starting point is that, when the regularity conditions for standard likelihood-based inference are not satisfied as it e.g. happens because of weak identification in some part of the parameter space, the asymptotic distribution of the QML estimator of the structural parameters is no longer Gaussian, with the asymptotic distribution of the associated QLR test for the CER being neither asymptotically pivotal nor chi-square distributed. Andrews and Cheng (2012) develop a general non-standard asymptotic theory for nonlinear models and extremum estimators which covers cases where lack of identification and/or weak identification occurs in part of the parameter space. Unfortunately, they assume the validity of a parameterization of the model which is not always easy to check for all DSGE models of interest. Our crucial point is that in these cases the bootstrap is still useful and can potentially be used to detect some types of misspecification, including weak identification.[2] A simple, 'descriptive' indicator of weak identification is given by empirical 'distance' between the analytical standard errors associated with the QML estimates of the structural parameters and the bootstrap standard errors. Indeed, in strongly identified models, the bootstrap standard errors estimate consistently, conditionally on the original data, the asymptotic standard errors, while this result is no longer true when standard inference is no longer valid. More important, we propose a test statistic, based on an arbitrary number, say B, of bootstrap realizations of the QML estimator (or the associated QLR test) to test the hypothesis that the model is strongly identified against a generic alternative of failure of the regularity conditions for standard inference. The test statistic is essentially a standard normality (or a test for chi square distribution when inference is based on the QLR test) test statistic, which is therefore straightforward to compute in practice. We derive sufficient conditions on the number of bootstrap repetitions B for the test statistic to have a well-defined asymptotic distribution under the null of strong identification. Crucially, we also show that this result is no longer valid and the test statistic diverges with probability one when strong identification fails. Hence, our misspecification test is consistent. In summary, we can use the bootstrap replicates of the QML estimator of the structural parameters and simple normality tests to evaluate how close or distant is the estimated DSGE model from the case of strong identification.

On the practical side, our approach requires (i) the estimation of the state space representation associated with the DSGE model on the original sample, with and without the CER, (ii) the computation of the QLR test for the CER, and (iii) the application of a (non)parametric bootstrap algorithm which imposes the null of the CER. Our bootstrap algorithm provides: (a) the bootstrap p-value associated with the QLR test, (b) an estimate of the distribution of the bootstrap QML estimator of the structural parameters and (c) an estimate of the distribu-

[2] Since in weakly identified DSGE model we do not have asymptotically pivotal estimators and test statistics, the common wisdom is that the bootstrap 'does not work'. While the fact that the bootstrap does not work under non-standard conditions is not always necessarily true (see e.g. Cavaliere et al. 2015 for an example), our intuition is that the distribution of the bootstrap QML estimator of the structural parameters is informative and useful also under weak identification.

tion of the bootstrap QML estimator of the parameters of the state space representation of the DSGE model in which the CER are not imposes. The bootstrap distributions in (b) and (c) is then analyzed by standard normality tests. We show how the information in (a)-(b)-(c) can be processed to analyze the DSGE model. In particular, when the practitioner has the suspect that weak identification in part of the structural parameters is the only possible source of misspecification, the diagnostic analysis in (b) and (c) is crucial to decide whether standard inference can be applied, or if it is convenient to resort to identification-robust methods along the lines suggested by Dufour *et al.* (2009, 2013), Kleibergen and Mavroeidis (2009), Guerron-Quintana *et al.* (2013), Qu (2014) and Andrews and Mikusheva (2015).

We investigate the usefulness of our approach by a set of Monte Carlo experiments. As weak identification is the type of misspecification of main interest in the literature on DSGE models, we follow Andrews and Cheng (2012) and take the workhorse autoregressive moving average process of order one (ARMA(1,1)) as data generating process. Our simulation studies confirm that our novel bootstrap-based mispecification test detects weak identification reasonably well. An empirical illustration based on the estimation of An and Schorfheide's (2007) monetary DSGE model on U.S. quarterly data is also provided. Our analysis confirms that our bootstrap approach works fairly well in detecting strong and weak identification. Moreover, we can conclude that the empirical evidence against this class of models seems to be less dramatic than commonly thought, once model evaluation is based on our nonparametric bootstrap approach.

This paper is organized as follows. Section 2 introduces the reference structural DSGE model and the underlying assumptions, and discusses its state space representations. Section 3 focuses on the QLR test for the CER. Section 3.1 introduces the testing problem and Section 3.2 characterizes the concepts of strong and weak identification we refer in this paper. Section 4 summarizes our bootstrap approach to DSGE models. Section 4.1 presents the bootstrap algorithm, Section 4.2 proves its asymptotic validity and Section 4.3 derives our novel bootstrap-based test to detect strong/weak identification. Section 4.4 frames our approach within the existing literature. Section 5 explores the finite sample performace of our approach by some Monte Carlo simulations that assume the ARMA(1,1) model as data generating process. Section 6 illustrates how our approach works on actual data by taking the An and Schorfheide's (2007) DSGE monetary model to U.S. quartely data. Section 7 contains some concluding remarks and directions for future research. Appendix A contains technical details and proofs. An associated Technical Supplement complements the results of the paper in several dimensions.[3]

[3] The Technical Supplement is available at http://www.rimini.unibo.it/fanelli/TS_bootstrapping_DSGE.pdf

2 Structural model, state space representations and assumptions

Let $Z_t{:=}(Z_{1,t}, Z_{2,t}, \cdots, Z_{n_z,t})'$ be a $n_z \times 1$ vector of endogenous, possibly unobserved variables at time t, which can be interpreted as deviation from corresponding steady state values. We assume that, after log-linearization and, for $t = 1, ..., T$, the structural form of the DSGE model reads as

$$\Gamma_0 Z_t = \Gamma_f E_t Z_{t+1} + \Gamma_b Z_{t-1} + \Pi \eta_t \tag{1}$$

$$\eta_t = R\eta_{t-1} + \omega_t \quad , \quad \omega_t \sim \text{WN}(0, \Sigma_\omega). \tag{2}$$

In eq.s (1)-(2), $\Gamma_i{:=}\Gamma_i(\theta)$, $i = 0, f, b$ are $n_z \times n_z$ matrices whose elements depend on the vector of structural parameters θ, $\Pi{:=}\Pi(\theta)$ is an $n_z \times n_\omega$ matrix of full-column rank ($n_\omega \leq n_z$), whose elements may depend on θ but which essentially selects the shocks that enter the equations, η_t is a $n_\omega \times 1$ vector of autoregressive disturbances, $R{:=}R(\theta)$ is an $n_\omega \times n_\omega$ stable diagonal matrix, ω_t is the $n_\omega \times 1$ vector of white noise structural shocks with covariance matrix $\Sigma_\omega{:=}\Sigma_\omega(\theta)$. Here, ω_t is adapted to the sigma-field $\mathcal{F}_t$, where $\mathcal{F}_t$ is the agents' information set at time t and $E_t Z_{t+1}{:=}E(Z_{t+1} \mid \mathcal{F}_t)$. The term ω_t will be referred to as the vector of fundamental structural shocks, and its covariance matrix Σ_ω can be either diagonal or non-diagonal. The initial condition Z_0 are treated as given. Finally, the true value of θ is denoted by θ_0 and is assumed to be an interior point of the compact parameter space $\mathcal{P}$.

The multivariate linear(ized) rational expectations model described by system (1)-(2) nests a large class of DSGE models currently used in policy and business cycle analysis. Our starting point is system (1)-(2) rather than its state space representation (which is derived below), essentially because our approach can cover a general class of linear rational expectations models used in macroeconomics and finance. The next example considers the prototype monetary DSGE model discussed in the literature, see Komunger and Ng (2011), Qu and Tkachenko (2012) and Qu (2014), among others.

Example 1 [An and Schorfheide's (2007) model] With a slight change of notation, An and Schorfheide's (2007) DSGE model is given by the equations:

$$x_t = E_t x_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(r_t - E_t \pi_{t+1} - E_t z_{t+1}), \tag{3}$$

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(x_t - g_t), \tag{4}$$

$$c_t = x_t - g_t, \tag{5}$$

$$r_t = \rho_r r_{t-1} + (1 - \rho_r)\psi_1 \pi_t + (1 - \rho_r)\psi_2(x_t - g_t) + \varepsilon_{r,t} \quad , \quad \varepsilon_{r,t} \sim \text{WN}(0, \sigma_r^2) \tag{6}$$

$$g_t = \rho_g g_{t-1} + \varepsilon_{g,t} \quad , \quad \varepsilon_{g,t} \sim \mathrm{WN}(0, \sigma_g^2) \tag{7}$$

$$z_t = \rho_z z_{t-1} + \varepsilon_{z,t} \quad , \quad \varepsilon_{z,t} \sim \mathrm{WN}(0, \sigma_z^2). \tag{8}$$

eq. (3) is a forward-looking output-gap equation and x_t is the output gap; eq. (4) is a purely forward-looking New-Keynesian Phillips Curve (NKPC) with slope $\kappa{:=}\frac{\tau(1-\nu)}{\nu\phi^*\bar{\pi}^2}$ and π_t is the inflation rate; eq. (5) is a consumption equation and c_t is consumption; eq. (6) is the monetary policy rule and r_t is the policy rate; finally, eq.s (7)-(8) maintain that the aggregate supply (g_t) and demand (z_t) disturbances are autoregressive processes. The vector of structural parameters is given by $\theta{:=}(\tau, \beta, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, \sigma_z^2, \sigma_g^2, \sigma_r^2)'$, $\dim(\theta) = 11$. We refer to An and Schorfheide (2007) for a derivation and discussion of the system in eq.s (3)-(8). It is seen that the consumption equation (5) does not bear any independent information on θ other that contained in the other equations, hence it can be dropped from the structural equations without any loss of information. In terms of the notation in eq.s (1)-(2) we have: $Z_t{:=}(x_t, \pi_t, r_t)'$, $(n_z = 3)$, $\eta_t{:=}(z_t, g_t, \varepsilon_{r,t})'$, $\omega_t{:=}(\varepsilon_{z,t}, \varepsilon_{g,t}, \varepsilon_{r,t})'$ $(n_\omega = 3)$ and

$$\Gamma_0 := \begin{pmatrix} 1 & 0 & \tau^{-1} \\ -\kappa & 1 & 0 \\ -(1-\rho_r)\psi_2 & -(1-\rho_r)\psi_1 & 1 \end{pmatrix}, \Gamma_f := \begin{pmatrix} 1 & \tau^{-1} & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\Gamma_b := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \rho_r \end{pmatrix} \quad , \quad \Pi := \begin{pmatrix} \tau^{-1}\rho_e & (1-\rho_g) & 0 \\ 0 & -\kappa & 0 \\ 0 & -(1-\rho_r)\psi_2 & 1 \end{pmatrix} \tag{9}$$

$$R := \begin{pmatrix} \rho_e & 0 & 0 \\ 0 & \rho_g & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad , \quad \Sigma_\omega := \begin{pmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_g^2 & 0 \\ 0 & 0 & \sigma_r^2 \end{pmatrix}.$$

The derivation of the rational expectations solution of system (1)-(2) and the associated state space representation in 'minimal form' is discussed in detail in the Technical Supplement. We assume that a unique and stable solution exists.

ASSUMPTION 1 [Determinacy] Given system (1)-(2), θ_0 is an interior point of the determinacy region $\mathcal{P}^D$ of the compact parameter space $\mathcal{P}$, $\mathcal{P}^D \subseteq \mathcal{P}$.

Assumption 1 implies that in correspondence of θ_0, the stochastic process that solves system (1)-(2) is covariance stationary and depends neither on extra parameters other than θ, nor on additional stochastic shocks other than the fundamental shocks ω_t, see Lubik and Schorfheide

(2003, 2004), Fanelli (2012) and Castelnuovo and Fanelli (2015). We prefer to rule out indeterminate equilibria because the occurrence of extra parameters and shocks other than fundamental shocks complicates identification issues considerably. Let $y_t := (y_{1,t}, y_{2,t}, \cdots, y_{n_y,t})'$ be the $n_y \times 1$ vector of (de-meaned) observed variables. A consequence of Assumption 1 is that for $\theta = \theta_0$, the stochastic process that generates the observables, $\{y_t\}$, is covariance stationary. We strengthen this condition by postulating that $\{y_t\}$ is also ergodic.

In the Technical Supplement we show that under Assumption 1, the so-called ABCD form associated with the determinate solution of the DSGE model in eq.s (1)-(2) can be expressed in the form

$$\underset{n_m \times 1}{Z_{m,t}} = \underset{n_m \times n_m}{A(\tau_\theta)} \underset{n_m \times 1}{Z_{m,t-1}} + \underset{n_m \times n_\omega}{B(\tau_\theta)} \underset{n_\omega \times 1}{\omega_t} \tag{10}$$

$$\underset{n_y \times 1}{y_t} = \underset{n_y \times n_m}{C(\tau_\theta)} \underset{n_m \times 1}{Z_{m,t-1}} + \underset{n_y \times n_\omega}{D(\tau_\theta)} \underset{n_\omega \times 1}{\omega_t} \tag{11}$$

where $Z_{m,t}$ is the n_m-dimension sub-vector of Z_t that contains the candidate 'minimal' states of the system, $A(\tau_\theta)$, $B(\tau_\theta)$, $C(\tau_\theta)$ and $D(\tau_\theta)$ are matrices of parameters that depend on θ (hence indirectly on Γ_0,Γ_f, Γ_b, Π and R) through the relationship $\tau_\theta = g(\theta)$, where $g(\cdot)$ is a nonlinear vector function in which τ_θ can be thought of as the vector of 'reduced form' coefficients that depend nonlinearly on θ under the CER. Obviously, $\tau_{\theta_0} = g(\theta_0)$. Assumption 1 ensures that the matrix $A(\tau_\theta) = A(g(\theta))$ in eq. (10) is stable[4], but this does not necessarily imply that $A(\tau_\theta)$ is also invertible. In most applications, however, it is found that $A(\tau_\theta)$ is invertible. Throughout the paper we also impose, without loss of generality, that $A(\tau_\theta)$ is non-singular.

The ABCD form in eq.s (10)-(11) is the candidate minimal state space representation associated with the DSGE model. It is minimal if and only if the system is controllable and observable, i.e. if the following rank conditions are valid

$$\text{rank}(C^0(\tau_{\theta_0})) = n_m = \text{rank}(O^0(\tau_{\theta_0})) \tag{12}$$

where

$$C^0(\tau_\theta) := (B(\tau_\theta), A(\tau_\theta)B(\tau_\theta), ..., A(\tau_\theta)^{n_m-1}B(\tau_\theta)) \quad , O^0(\tau_\theta) := \begin{pmatrix} C(\tau_\theta) \\ C(\tau_\theta)A(\tau_\theta) \\ \vdots \\ C(\tau_\theta)A(\tau_\theta)^{n_m-1} \end{pmatrix}$$

are respectively the controllability (C^0) and observability (O^0) matrices.

ASSUMPTION 2 [Minimality] The state-space system (10)-(11) is minimal, i.e. such that the condition in eq. (12) is valid.

[4] Throughout the paper we use the term 'stable' to denote a matrix that has all eigenvalues inside the unit circle in the complex plane.

In general, under Assumptions 1–2, the state-space system (10)-(11) admits a stationary VARMA-type representation for y_t, see e.g. Hannan and Deistler (1988). The minimality condition mimics the left-coprime condition we typically impose on (or assume in) VARMA processes (see e.g. Lütkepohl, 2005, p. 452). For cases in which the $D(\tau_\theta)$ matrix in eq.s (10)-(11) is square ($n_y = n_\omega$), it is possible to derive conditions that ensure the fundamentalness of the ABCD form (Fernández-Villaverde *et al.* 2007), i.e. a representation where the state $Z_{m,t}$ is fully recoverable from $y_1, ..., y_t$, and a finite-order VAR representation for y_t exists (Ravenna, 2007; Franchi and Vidotto, 2013; Franchi and Paruolo, 2015). We do not impose any of these conditions, in the sense that for suitable re-formulations of the state vector and the process through which the shocks propagate in the system, the specified DSGE model and the bootstrap approach we present in this paper can potentially accommodate foresight phenomena about technology and government spending, anticipated shocks and non-fundamental representations along the lines of e.g. Leeper *et al.* (2013).

Our last assumption regards the (local) identifiability of the structural parameters θ from system (10)-(11). Following Komunjer and Ng (2011), θ is identified if and only if

$$\text{rank}(\mathcal{M}(\tau_{\theta_0})) = \dim(\theta_0) + n_m^2 + n_\omega^2 \tag{13}$$

where

$$\mathcal{M}(\theta) := \begin{pmatrix} \frac{\partial vecA(\tau_\theta)}{\partial\theta'} & A(\tau_\theta)' \otimes I_{n_m} - I_{n_m} \otimes A(\tau_\theta) & 0_{n_m^2 \times n_\omega^2} \\ \frac{\partial vecB(\tau_\theta)}{\partial\theta'} & B(\tau_\theta)' \otimes I_{n_m} & I_{n_\omega} \otimes B(\tau_\theta) \\ \frac{\partial vecC(\tau_\theta)}{\partial\theta'} & -I_{n_m} \otimes C(\tau_\theta) & 0_{n_y n_m \times n_\omega^2} \\ \frac{\partial vecD(\tau_\theta)}{\partial\theta'} & 0_{n_y n_m \times n_m^2} & I_{n_\omega} \otimes D(\tau_\theta) \\ \frac{\partial vech\Sigma_\omega(\tau_\theta)}{\partial\theta'} & 0_{\frac{1}{2}n_y(n_y+1) \times n_m^2} & -2d_{n_\omega}^{+}(\Sigma_\omega(\tau_\theta) \otimes I_{n_\omega}) \end{pmatrix}.$$

Here $d_{n_u}^{+} := (d_{n_\omega}' d_{n_\omega})^{-1} d_{n_\omega}'$ is the Moore-Penrose inverse of the duplication matrix d_{n_ω}, i.e. the $n_\omega^2 \times \frac{1}{2}n_\omega(n_\omega + 1)$ matrix such that $d_{n_\omega} vech(\Sigma_\omega) = vec(\Sigma_\omega)$, and it is maintained that the matrices $A(\tau_\theta)$, $B(\tau_\theta)$, $C(\tau_\theta)$ and $D(\tau_\theta)$ have continuos first and second derivatives with respect to θ.

ASSUMPTION 3 [Identification] The state-space system (10)-(11) is 'locally identified', i.e. such that the functions $vec(A(g(\theta)))$, $vec(B(g(\theta)))$, $vec(C(g(\theta))$ and $vec(D(g(\theta)))$ have continuos first and second derivatives with respect to θ, the rank condition in eq. (13) is valid in a neighborhood $\mathcal{N}_{\theta_0}$ of θ_0.

Komunjer and Ng (2011) show that for the case $n_y \leq n_\omega$, the information matrix of the system is nonsingular if and only if the condition in eq. (13) is satisfied. Using the terminology in Canova and Sala (2009), Assumption 3 is a statement about the 'population' identification of the DSGE model. However, identification problems in a system of variables featuring highly nonlinear restrictions might also involve the relationship between the structural parameters and the sample objective function, which might display 'small' curvature in certain regions of the parameter space. We clarify these issues in the next sections.

Finally, we consider the counterpart of the state space representation in eq.s (10)-(11) without imposing the CER. Indeed, in the next sections state space representation of the DSGE model which does not incorporate the CER is needed in order to compute QLR tests for the CER.[5] Guerron-Quintana *et al.* (2013) have shown that it is in principle possible to couple the ABCD form with a counterpart where τ_θ is replaced with a vector of parameters, say τ_u, which collapses to τ_θ when the CER are imposed, see also Angelini and Fanelli (2015). We posit that, associated with the ABCD form in eq.s (10)-(11), it exists a state space representation for $Z_{m,t}$ and y_t, given by

$$\underset{n_m\times 1}{Z_{m,t}} = \underset{n_m\times n_m}{A(\tau_u)}\ \underset{n_m\times 1}{Z_{m,t-1}} + \underset{n_m\times n_\omega}{B(\tau_u)}\ \underset{n_\omega\times 1}{\omega_t^u} \tag{14}$$

$$\underset{n_y\times 1}{y_t} = \underset{n_y\times n_m}{C(\tau_u)}\ \underset{n_m\times 1}{Z_{m,t-1}} + \underset{n_y\times n_\omega}{D(\tau_u)}\ \underset{n_\omega\times 1}{\omega_t^u}\ ,\ \ t=1,...,T, \tag{15}$$

where ω_t^u is an $n_\omega \times 1$ vector of white noise structural shocks with diagonal covariance matrix Σ_{ω^u}, and τ_u is a vector of coefficients such that $\dim(\tau_u) = \dim(\tau_\theta)$, $\dim(\tau_u) > \dim(\theta)$. System (14)-(15) characterizes the first- and second-order moments of the data, net of the parametric restrictions stemming from the theory.

ASSUMPTION 4 [State space representation wihout the CER] System (14)-(15) is such that:

(4.i) the matrix $A(\tau_u)$ is stable and invertible, and the functions $vec(A(\tau_u))$, $vec(B(\tau_u))$, $vec(C(\tau_u)$ and $vec(D(\tau_u))$ have continuous first and second derivatives with respect to τ_u;

(4.ii) is in minimal form, in the sense that the condition in eq. (12) is valid once τ_{θ_0} is replaced by $\tau_{u,0}$, where $\tau_{u,0}$ is the true value of τ_u and is an interior point of the compact parameter space $\mathcal{P}_\tau$;

(4.iii) is identified in the sense that the condition in eq. (13) is valid once τ_{θ_0} is replaced by $\tau_{u,0}$.

[5] Finding an 'unrestricted' state space representation of the DSGE model is not a trivial task because of the difficulties associated with finding an identified minimal state space form that plays the same role reduced form models have in the context of simultaneous systems of equations, see e.g. Schorfheide (2010), Komunjer and Ng. (2011) Guerron-Quintana *et al.* (2013), Qu (2014) and Andrews and Mikusheva (2015).

Together with the (assumed fixed) initial conditions for the state, under Assumption 4.i-4.iii, system (14)-(15) provides a complete characterization of first- and second-order moment properties of the data. Also in this case, the stochastic process that generates the observables, $\{y_t\}$, is covariance stationary and, we add, ergodic.

3 Test of the cross-equation restrictions

The ABCD form in eq.s (10)-(11) and the system (14)-(15) can be regarded as the state space interfaces of the DSGE model. It is seen that the ABCD form in eq.s (10)-(11) is nested within system (14)-(15), in the sense that system (14)-(15) collapses to the ABCD form in eq.s (10)-(11) once the elements of the vector τ_u are replaced with the elements of the vector $\tau_\theta = g(\theta)$, where $g(\cdot)$ is a nonlinear differentiable function such that the Jacobian $(\nabla_\theta \tau_\theta)' \equiv \frac{\partial g(\theta)}{\partial \theta'}$ has full-column rank $\dim(\theta)$ in a neighborhood of θ_0. Our Technical Supplement discussed in detail the relationship between τ_u and τ_θ. We have all the ingredients to compute a QLR test of the CER implied by the DSGE model.

In Section 3.1 we briefly address the testing problem and in Section 3.2 we characterize the concepts of strong and weak identification and their consequences on inference.

3.1 Testing problem

We address the testing problem

$$H_0 : \tau_u = \tau_\theta = g(\theta) \quad \text{against} \quad H_1 : \tau_u \neq \tau_\theta \tag{16}$$

by a QLR test. The null H_0 incorporates the CER and the state space representation of the DSGE model under H_0 is given by system (10)-(11). Instead, the state space representation of the DSGE model under H_1 corresponds to system (14)-(15). To compute a QLR test of the CER it is necessary to maximize the likelihood associated with both systems. Let $\hat{\tau}_{u,T}$ and $\hat{\theta}_T$ ($\hat{\tau}_{\theta,T} = g(\hat{\theta}_T)$) be the QML estimates of the state space model obtained under H_1 and H_0, respectively. The estimation problems are based on Gaussian innovation errors and the Kalman-filter are considered in the Technical Supplement. For the purposes of the present analysis, it is important to recall that under H_0, the innovation form representation (Anderson and Moore, 1979) in (steady-state form) associated with the ABCD form in eq.s (10)-(11) is summarized by the expressions

$$\hat{Z}_{m,t+1|t} = A(\tau_\theta)\hat{Z}_{m,t|t-1} + K(\tau_\theta)\, \epsilon_t^0 \tag{17}$$

$$y_t = H^{+}(\tau_\theta)\hat{Z}_{m,t|t-1} + \epsilon_t^0 \tag{18}$$

where $K{=}K(\tau_\theta)$ is the (steady-state) Kalman gain, $H^{+}(\tau_\theta){:=}C(\tau_\theta)A(\tau_\theta)^{-1}$ and $\hat{Z}_{m,t|t-1}{:=}E(Z_{m,t} \mid \mathcal{F}^{y}_{t-1})$, where $\mathcal{F}^{y}_{t}{:=}\sigma(y_t, ..., y_1) \subset \mathcal{F}_t$ is the information set based on the observable variables up to time t. ϵ^0_t are the innovation errors with covariance matrix $\Sigma_{\epsilon^0,t}$, $t = 1, ..., T$. We use the superscript '0' for ϵ^0_t and $\Sigma_{\epsilon^0,t}$ to remark that the representation in eq.s (17)-(18) is obtained under the null H_0 which imposes the CER. Under the alternative H_1, the innovation form representation is similar to system (17)-(18) but τ_θ needs to be replaced with τ_u, hence $\epsilon^u_t{:=}y_t - H^{+}(\tau_u)\hat{Z}_{m,t|t-1}$, and the implied covariance matrix is $\Sigma_{\epsilon^u,t}, t = 1, ..., T$.

Let $\ell_{o,T}(\hat{\tau}_{u,T})$ and $\ell_{o,T}(\hat{\tau}_{\hat{\theta},T})$ be the log-likelihoods obtained under H_1 and H_0, respectively. The QLR test is given by

$$QLR_T(\hat{\theta}_T){:=} - 2[\ell_{o,T}(\hat{\tau}_{\theta,T}) - \ell_{o,T}(\hat{\tau}_{u,T})]. \tag{19}$$

The asymptotic properties of the tests statistics $QLR_T(\hat{\theta}_T)$ are intimately related to the asymptotic properties of $\hat{\theta}_T$ $(\hat{\tau}_{\theta,T})$ and $\hat{\tau}_{u,T}$ and these crucially depend on whether the regularity conditions for inference are valid in the estimated DSGE model.

3.2 Strong and weak identification: characterization

While the DSGE model is unidentified if the rank condition in Assumption 3 fails, the validity of Assumption 3 does not necessarily rule out cases in which the log-likelihood of the system $\ell_{o,T}(g(\theta))$ does not satisfy the standard regularity conditions that permit 'standard inference', where by this term we denote the situation in which $\hat{\theta}_T$ is consistent, asymptotically Gaussian and quadratic forms derived from $\hat{\theta}_T$ are asymptotically χ^2-distributed. More precisely, the DSGE model can satisfy the local identification condition in Assumption 3 at the point $\theta = \theta_0$, but $\ell_{\circ,T}(\theta)$ might exhibit local maxima (and minima) and nearly flat surfaces in some directions of the parameter space. This concept is often termed, borrowing terminology from the literature on instrumental variables and generalized methods of moments, 'weak identification' (Staiger and Stock, 1997; and Stock and Wright 2000).

When the standard regularity conditions for inference hold (see Technical Supplement), we can think about the scenario in which Assumptions 1-3 are also valid for a $\theta_\dagger$ in $\mathcal{P}^D$ that replaces θ_0 and such that

$$\hat{\theta}_T \longrightarrow_p \theta_\dagger \tag{20}$$

and

$$T^{1/2}Z_T \overset{d}{\longrightarrow}_p N\left(0_{dim(\theta)\times 1}, I_{dim(\theta)}\right), \tag{21}$$

where $T^{1/2}Z_T{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\theta}_T,T}(\hat{\theta}_T - \theta_\dagger)$. The quantity $\check{\mathcal{I}}_{\hat{\theta}_T,T}$ in Z_T is a consistent, (a.s.) positive definite estimator of the inverse of the covariance matrix

$$V_{\theta_\dagger} := \left(\mathcal{I}^{2D}_{\theta_\dagger,\infty} \left(\mathcal{I}^{OP}_{\theta_\dagger,\infty} \right)^{-1} \mathcal{I}^{2D}_{\theta_\dagger,\infty} \right)^{-1} \tag{22}$$

where

$$\mathcal{I}^{OP}_{\theta_\dagger,\infty} := \lim_{T\to\infty} \frac{1}{T} \mathcal{I}^{OP}_{\theta_\dagger,T} \quad , \quad \mathcal{I}^{OP}_{\theta_\dagger,T} := E\left(\nabla_\theta \ell_{\circ,T}(\theta_\dagger) \times \nabla_\theta \ell_{\circ,T}(\theta_\dagger)' \right) \tag{23}$$

$$\mathcal{I}^{2D}_{\theta_\dagger,\infty} := \lim_{T\to\infty} \frac{1}{T} \mathcal{I}^{2D}_{\theta_\dagger,T} \quad , \quad \mathcal{I}^{2D}_{\theta_\dagger,T} := -\, E\left(\nabla^2_{\theta\theta} \ell_{\circ,T}(\theta_\dagger) \right). \tag{24}$$

Here and throughout the paper, $\nabla_\theta \ell_{\circ,T}(\theta_\dagger) \equiv \left. \frac{\partial \ell_{\circ,T}(\theta)}{\partial \theta} \right|_{\theta=\theta_\dagger}$ and $\nabla^2_{\theta\theta} \ell_{\circ,T}(\theta_\dagger) \equiv \left. \frac{\partial^2 \ell_{\circ,T}(\theta)}{\partial \theta \partial \theta'} \right|_{\theta=\theta_\dagger}$.

The (locally) identified parameter vector $\theta_\dagger$ plays in eq.s (20)-(22) the role of 'pseudo-true' value of the structural parameters which minimizes the Kullback-Leibler information criterion (White, 1982). Under standard regularity conditions, the QML estimator $\hat{\theta}_T$ converges to $\theta_\dagger$ when the CER are imposed in estimation. When the CER are true (i.e. the determinate solution of the DSGE model belongs to the data generating process), $\theta_\dagger = \theta_0$ in eq.s (20)-(22) and $V_{\theta_\dagger} = V_{\theta_0}$ reads as the asymptotic covariance matrix of $T^{1/2}(\hat{\theta}_T - \theta_0)$, i.e. $V_{\theta_0} := \lim_{T\to\infty} Var(T^{1/2}\hat{\theta}_T)$. Furthermore, if the CER are true and the innovation errors that characterize the Kalman filter algorithm are actually Gaussian, both matrices $\mathcal{I}^{OP}_{\theta_0,T}$ and $\mathcal{I}^{2D}_{\theta_0,T}$ derived from eq.s (23)-(24) for $\theta_\dagger = \theta_0$ characterize the system's information matrix and V_{θ_0} collapses to the expression $V_{\theta_0} := \left(\mathcal{I}^{2D}_{\theta_0,\infty} \right)^{-1}$. The quantity:

$$\check{\mathcal{I}}^{OP}_{\hat{\theta}_T,T} := \sum_{t=1}^{T} \nabla_\theta \ell_t(\hat{\theta}_T) \times \nabla_\theta \ell_t(\hat{\theta}_T)' \ , \tag{25}$$

known as 'incremental observed information' and the quantity:

$$\check{\mathcal{I}}^{2D}_{\hat{\theta}_T,T} := -\sum_{t=1}^{T} \nabla^2_{\theta\theta} \ell_t(\hat{\theta}_T), \tag{26}$$

known as 'observed information' (evaluated at $\hat{\theta}_T$), are both consistent (interchangeable) estimates of the system's information matrix. It turns out that one can use

$$\check{\mathcal{I}}_{\hat{\theta}_T,T} = \check{\mathcal{I}}^{2D}_{\hat{\theta}_T,T} \left(\check{\mathcal{I}}^{OP}_{\hat{\theta}_T,T} \right)^{-1} \check{\mathcal{I}}^{2D}_{\hat{\theta}_T,T} \tag{27}$$

in eq. (21), or $\check{\mathcal{I}}_{\hat{\theta}_T,T} = \check{\mathcal{I}}^{2D}_{\hat{\theta}_T,T}$ (or $\check{\mathcal{I}}_{\hat{\theta}_T,T} = \check{\mathcal{I}}^{OP}_{\hat{\theta}_T,T}$) when it is known that the innovations errors that feed the Kalman filter are Gaussian.

Similar considerations apply to the estimation of τ_u under Assumption 4 and H_1, see Watson (1989). It follows that under Assumptions 1-4 (and the regularity conditions (TS-A.i)-(TS-A.viii) reviewed in the Technical Supplement), standard arguments imply that the test statistic $QLR_T(\hat{\theta}_T)$ in eq. (19) is asymptotically χ^2-distributed with k =dim(τ_u)-dim(θ) degree of freedom (henceforth χ^2_k) when H_0 is true and diverges when H_1 is true. Thus, the null H_0 is rejected when $QLR_T(\hat{\theta}_T) \geq c^{\eta}_{\chi^2_k}$, where $c^{\eta}_{\chi^2_k}$ is the η-level cut-off point associated with the χ^2_k-distribution and $0<\eta<1$ is the pre-fixed nominal level of significance (or type-I error) of the test. Notice, however, also in strongly identified DSGE models, the χ^2_k distribution might represent a poor approximation in finite samples.

Conversely, in weakly identified DSGE models, the consistency result in eq. (20) and the asymptotic normality result in eq. (21) are no longer guaranteed, see, among others, Andrews and Cheng (2012) and Andrews and Mikusheva (2015). Andrews and Mikusheva (2015) show through examples that under weak identification, the appropriately normalized quadratic variation of the score converges to fixed positive definite matrix, while the Hessian converges in distribution to a random matrix. Thus, under weak identification, one can expect large disparities between different estimators of information also if the model respects White's (1982) information matrix equality. Also when the innovations errors are Gaussian, it is not necessarily true that the matrices in eq.s (25) and (26) evaluated at the point $\theta = \theta_0$ are consistent (and interchangeable) estimates of the information matrix. Andrews and Mikusheva (2015) observe that in these cases 'it is unwise to estimate the information matrix using an estimator of θ' such as e.g. the quantities $\check{\mathcal{I}}^{2D}_{\hat{\theta}_T,T}$ and $\check{\mathcal{I}}^{OP}_{\hat{\theta}_T,T}$ in eq.s (25)-(26).

Summing up, the class of weakly identified DSGE models we have in mind in this paper, must not be considered misspecified in the 'conventional' sense discussed in e.g. Inoue *et al.* (2015), An and Shorfheide (2007), Section 3.1 and Angelini and Fanelli (2015). Rather, weak identification can be associated with situations in which the convergences in eq.s (20)-(21) are no longer valid and suitably normalized versions of the matrix $\check{\mathcal{I}}^{2D}_{\theta_0,T}$ converge weakly to a random matrix. Conventionally, throughout the paper we use the terms 'strong identification' and 'strongly identified DSGE model' to denote the situations in which the regularity conditions that permit standard inference on θ are valid. Conversely, failure of these conditions, as it happens e.g. in 'weakly identified' DSGE models, does not permit the 'conventional' asymptotic expansions of the likelihood, with the consequence that irrespective of whether the CER are true (i.e. H_0 is valid) or not (i.e. H_1 is valid), the test statistics $T^{1/2}Z_T$ deviates asymptotically from the Gaussian distribution.

The different asymptotic behaviour of the statistic $T^{1/2}Z_T$ under strong identification and when strong identification fails is at the basis of our bootstrap approach to the evaluation of DSGE models. In the next section we prove that in strongly identified DSGE models, the bootstrap analog of $T^{1/2}Z_T$, $T^{1/2}Z^*_T{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\theta}_T,T}(\hat{\theta}^*_T - \hat{\theta}_T)$, is also asymptotically Gaussian and the bootstrap is consistent. Instead, when strong identification fails $T^{1/2}Z_T$ is far from normal also asymptotically and also its bootstrap analog $T^{1/2}Z^*_T$ is not asymptotically Gaussian.

4 Bootstrap approach

We take the point of view of a practitioner who is concerned with (i) the estimation of the structural parameters of the DSGE model in eq.s (1)-(2) and (ii) the empirical evaluation of the estimated model by the testing the hypothesis H_0 against H_1 in eq.(16). We assume that given the estimation sample $y_1, ..., y_T$, the model estimation of θ (τ_θ) (under H_0) and τ_u (under H_1) has been completed and the QML estimates $\hat{\theta}_T$ ($\hat{\tau}_{\theta,T} = g(\hat{\theta}_T)$) and $\hat{\tau}_{u,T}$ are available. The practitioner does not know *a priori* whether the DSGE model meets the regularity conditions (Technical Supplement) that permit standard inference.

Our inferential strategy is based on the following idea. Computed the QLR test in eq. (19) on the sample y_1, ..., y_T, we compute its bootstrap analog, denoted by $QLR_T^*(\hat{\theta}_T^*)$. This can be done by a standard Monte Carlo algorithm as described in Section 4.1. Henceforth 'asterisks' denote bootstrap analogs of estimators and statistics; for instance, $\hat{\theta}_T^*$ denotes be estimator of θ obtained on the bootstrap sample $y_1^*, ..., y_T^*$. In order to compute the bootstrap p-value associated to the bootstrap version of the QLR test, it is required to generate an arbitrarily number of bootstrap samples, B say, under the null of the DSGE model, i.e. with the CER in H_0 imposed (meaning that θ is fixed at the QML estimate $\hat{\theta}_T$). Then, on each generated sample the state space representation of the model is estimated both under the CER (H_0) and unrestrictedly (H_1). In Section 4.2 we can prove that in strongly identified DSGE models the bootstrap is consistent. Thus, $\hat{\theta}_T^*$ should, conditional on the original data, be normally distributed for large values of T and B. Accordingly, each element of $\hat{\theta}_T^*$ is expected, conditional on the original data, to conform to the Gaussian distribution. Similarly, the distribution of $QLR_T^*(\hat{\theta}_T^*)$ should, conditionally on the original data, be 'close to' the χ_k^2 distribution for large values of T and B. In DSGE models for which the regularity condition for standard inference fail, instead, $\hat{\theta}_T^*$ does not conform to the multivariate normal.[6] In Section 4.3 we discuss how these considerations can be used to design a bootstrap-based misspecification test for the estimated model. In Section 4.4 we compare our approach with the existing literature.

4.1 Bootstrap algorithm

The QLR test statistic, $QLR_T(\hat{\theta}_T)$, is computed as in eq. (19). Our nonparametric bootstrap procedure is adapted from Stoffer and Wall's (1991) algorithm, see also Cavanaugh and Shumway

[6] The non-normality of $\hat{\theta}_T^*$ might be due to the non-normality of some of its elements. The converse, instead, is not necessarily true: even if the hypothesis of normality is tenable for each of the elements of $\hat{\theta}_T^*$ individually, this does not necessarily imply multivariate normality of $\hat{\theta}_T^*$, since there are non-normal multivariate distributions that have normal marginals.

(1997), and is given by the following steps:

1. Given the innovation residuals $\hat{\epsilon}_t^0 := y_t - H^+(\hat{\tau}_{\theta,T})\hat{Z}_{m,t|t-1}$ and the estimated covariance matrix $\hat{\Sigma}_{\epsilon^0,t} = \Sigma_{\epsilon^0,t}(\hat{\tau}_{\theta,T})$ produced by the estimation of the DSGE model under the CER (H_0), construct the standardized innovations as

$$\hat{e}_t^0 := \hat{\Sigma}_{\epsilon^0,t}^{-1/2}\ \hat{\epsilon}_t^{0,c}\ ,\quad t = 1, ..., T,$$

where $\hat{\Sigma}_{\epsilon^0,t}^{-1/2}$ is the inverse of the square-root matrix of $\hat{\Sigma}_{\epsilon^0,t}$ and $\hat{\epsilon}_t^{0,c}$, $t = 1, ..., T$, are the centered residuals $\hat{\epsilon}_t^{0,c} := \hat{\epsilon}_t^0 - T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^0$;

2. Sample, with replacement, T times from $\hat{e}_1^0, \hat{e}_2^0, ..., \hat{e}_T^0$ to obtain the bootstrap sample of standardized innovations $\hat{e}_1^{0,*}, \hat{e}_2^{0,*}, ..., \hat{e}_T^{0,*}$;

3 Mimicking the innovation form representation of the DSGE model in eq.s (17)-(18), the bootstrap sample $y_1^*, y_2^*, ..., y_T^*$ is generated recursively by solving, for $t = 1, ..., T$, the system

$$\begin{pmatrix} \hat{Z}_{m,t+1|t}^* \\ y_t^* \end{pmatrix} = \begin{pmatrix} A(\hat{\tau}_{\theta,T}) & 0_{n_m \times n_y} \\ H^+(\hat{\tau}_{\theta,T}) & 0_{n_y \times n_y} \end{pmatrix} \begin{pmatrix} \hat{Z}_{m,t|t-1}^* \\ y_{t-1}^* \end{pmatrix} + \begin{pmatrix} K(\hat{\tau}_{\theta,T})\hat{\Sigma}_{\epsilon^0,t}^{1/2} \\ \hat{\Sigma}_{\epsilon^0,t}^{1/2} \end{pmatrix} \hat{e}_t^{0,*} \tag{28}$$

with initial condition $\hat{Z}_{m,t+1|t}^* = \hat{Z}_{m,1|0}$;

4 From the generated pseudo-sample $y_1^*, y_2^*, ..., y_T^*$, estimate the DSGE model under H_0 obtaining the bootstrap estimator $\hat{\theta}_T^*$ ($\hat{\tau}_{\theta,T}^* = g(\hat{\theta}_T^*)$) and the associated log-likelihood $\ell_{o,T}^*(\hat{\tau}_{\theta,T}^*)$, and estimate the DSGE model under H_1 obtaining the bootstrap estimator $\hat{\tau}_{u,T}^*$ and the associated log-likelihood $\ell_{o,T}^*(\hat{\tau}_{u,T}^*)$; the bootstrap QLR test for the CER is defined by:

$$QLR_T^*(\hat{\theta}_T^*) := -2[\ell_{o,T}^*(\hat{\tau}_{\theta,T}^*) - \ell_{o,T}^*(\hat{\tau}_{u,T}^*)]; \tag{29}$$

5 Steps 2-4 are repeated B times in order to obtain B bootstrap realizations of $\hat{\theta}_T$ and $\hat{\tau}_{u,T}$, say $\{\hat{\theta}_T^{*,1}, \hat{\theta}_T^{*,2}, ..., \hat{\theta}_T^{*,B}\}$ and $\{\hat{\tau}_{u,T}^{*,1}, \hat{\tau}_{u,T}^{*,2}, ..., \hat{\tau}_{u,T}^{*,B}\}$, and the B bootstrap realizations of the associated bootstrap QLR test, $\{QLR_T^{*,1}, QLR_T^{*,2}, ..., QLR_T^{*,B}\}$, where $QLR_T^{*,b} := QLR_T^{*,b}(\hat{\theta}_T^{*,b})$, $b = 1, ..., B$;

6 The bootstrap p-value of the test of the CER is computed as

$$\widehat{p}_{T,B}^* := \frac{B \cdot \hat{G}_{T,B}^*(QLR_T(\hat{\theta}_T)) + 1}{B+1}, \tag{30}$$

where $\hat{G}_{T,B}^*(x) := B^{-1}\sum_{b=1}^{B}\mathbb{I}\{QLR_T^{*,b} > x\}$, $\mathbb{I}\{\cdot\}$ being the indicator function;

7 The bootstrap QLR test for the CER at the $100\eta\%$ (nominal) significance level rejects H_0 if $\widehat{p}_{T,B}^* \leq \eta$.

A few remarks are in order.

Remark 1 The bootstrap algorithm features a 'restricted' resampling scheme. Steps 1-3 refer to the generation of pseudo-bootstrap samples under the null H_0, i.e. under the CER implied by the DSGE model. The procedure can be easily adapted to the case of an 'unrestricted' resampling scheme, i.e. the bootstrap pseudo-samples might also be generated without imposing the null. In that case, in step 1 one considers and resamples the residuals and associated covariance matrix obtained from the estimation of the state space model without imposing the CER, $\hat{\epsilon}_t^u := y_t - H^{+}(\hat{\tau}_{u,T})\hat{Z}_{m,t|t-1}$ and $\hat{\Sigma}_{\epsilon^u,t} = \Sigma_{\epsilon^u,t}(\hat{\tau}_{u,T})$, and then replaces $\hat{\tau}_{\theta,T}$ with $\hat{\tau}_{u,T}$ in step 3. Moreover, the algorithm is 'nonparametric' in the sense that in step 1 the bootstrap innovations are obtained by re-sampling the estimated residuals. However, if the normality hypothesis is correct, the parametric version of the bootstrap algorithm requires ignoring steps 1 and 2 and starting from the step 3, with the $\hat{e}_t^{0,*}$ now taken as independent random draws from the $N(0_{n_y \times 1}, I_{n_y})$ distribution.

Remark 2 It is important that in the execution of the step 4 the bootstrap pseudo-samples $y_1^*, y_2^*, ..., y_T^*$ for which it is found that the estimated matrices $A(g(\hat{\theta}_T^*))$ (estimation under H_0, i.e. the CER) and $A(\hat{\tau}_{u,T}^*)$ (estimation under H_1, i.e. without the CER) are unstable are discarded and replaced with pseudo-samples that respect the stability condition, until B valid replications are obtained.

Remark 3 When the $D(\tau_\theta)$ matrix in eq.s (10)-(11) is square ($n_y = n_\omega$) and invertible, it exists a finite-order VAR representation for y_t (Section 2). In this case, the bootstrap algorithm can be adapted such that it is the VAR representation of the DSGE model which is used to estimate the parameters and bootstrap the residuals, see e.g. Bårdsen and Fanelli (2005) and Cho and Moreno (2006).

4.2 Asymptotic validity

Let $T^{1/2}Z_T^* := T^{1/2}\check{\mathcal{I}}_{\hat{\theta}_T,T}^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$ be the bootstrap analog of $T^{1/2}Z_T := T^{1/2}\check{\mathcal{I}}_{\hat{\theta}_T,T}^{1/2}(\hat{\theta}_T - \theta_0)$, where $\check{\mathcal{I}}_{\hat{\theta}_T,T}$ is a consistent estimator of the inverse of the asymptotic covariance matrix V_{θ_0}, see eq. (33). The QML estimator $\hat{\theta}_T$ plays in $T^{1/2}Z_T^*$ the role of a 'pseudo-true parameter'. From a computational viewpoint, measures of $\hat{\theta}_T^*$ can be obtained through the algorithm presented in previous Section, which generates the B repetitions $\hat{\theta}_T^{*,1}, ..., \hat{\theta}_T^{*,B}$.

Using regularity conditions by Ljung and Caines (1979), Stoffer and Wall (1991) prove that their bootstrap algorithm delivers (a) an asymptotic distribution for $T^{1/2}Z_T^*$ which is the same as that of $T^{1/2}Z_T$, i.e. multivariate Gaussian and (b) consistent bootstrap standard errors.

Reinterpreting these results in light of the notation discussed in Appendix A and the regularity conditions summarized in the Technical Supplement, Stoffer and Wall's (1991) result (a) can be restated, assuming that the CER are true ($\theta_{\dagger}$=θ_0), in the form

$$\hat{\theta}_T^* - \hat{\theta}_T \xrightarrow{p^*}_p 0_{\dim(\theta)\times 1} \tag{31}$$

$$T^{1/2} Z_T^* \xrightarrow{d^*}_p N\left(0_{dim(\theta)\times 1}, I_{\dim(\theta)}\right) \tag{32}$$

where '$\xrightarrow{p^*}_p$' denotes $\hat{\theta}_T^*$ converges 'in P^*, in probability', while '$\xrightarrow{d^*}_p$' denotes converge in 'conditional distribution, in probability' (see Appendix A). Stoffer and Wall's (1991) result (b) can be stated by:

$$Var^*(T^{1/2}\hat{\theta}_T^*) - Var(T^{1/2}\hat{\theta}_T) \xrightarrow{p^*}_p 0_{\dim(\theta)\times\dim(\theta)} \tag{33}$$

where $Var^*(\cdot)$ denotes conditional (on the original sample) variance (Appendix A). Note that a 'computationally convenient' measure of $Var^*(\hat{\theta}_T^*)$ can be obtained, for sufficiently large B, from the bootstrap replicates $\hat{\theta}_T^{*,1}$, $\hat{\theta}_T^{*,2}$, ..., $\hat{\theta}_T^{*,B}$ as

$$\widehat{Var}^*(\hat{\theta}_T^*) := \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\theta}_T^{*,b} - \overline{\hat{\theta}}_T^*\right)\left(\hat{\theta}_T^{*,b} - \overline{\hat{\theta}}_T^*\right)' \quad , \quad \overline{\hat{\theta}}_T^* := \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_T^{*,b}.$$

The square root of the elements on the main diagonal of $\widehat{Var}^*(\hat{\theta}_T^*)$ delivers the 'bootstrap standard errors' practitioners routinely report along with QML estimates of the structural parameters.

We have all the ingredient to prove that the bootstrap is consistent. The next proposition establishes bootstrap consistency by showing that the convergences in eq.s (31)-(33) are valid in strongly identified DSGE models and, moreover, that the bootstrap is consistent. For x:=$(x_1, ..., x_{\dim(\theta)})' \in \mathbb{R}^{\dim(\theta)}$ and K_T:=$T^{1/2}Z_T$=$(K_{1,T}, ..., K_{\dim(\theta),T})'$, let $G_T(x)$:=$P(K_{1,T} \leq x_1, ..., K_{\dim(\theta),T} \leq x_{\dim(\theta)})$ be the cumulative distribution function (CDF) of K_T, and given K_T^*:=$T^{1/2}Z_T^*$=$(K_{1,T}^*, ..., K_{\dim(\theta),T}^*)'$, let
$G_T^*(x)$:=$P^*(K_{1,T}^* \leq x_1, ..., K_{\dim(\theta),T}^* \leq x_{\dim(\theta)})$ be the CDF of K_T^* induced by the bootstrap sampling process conditional on the sample.

Proposition 1 [Consistency] Consider the DSGE model in eq.s (1)-(2) with ABCD representation in eq.s (10)-(11), Assumptions 1-3, the set of regularity conditions (A.i*)-(A.viii*) and Lemma 1 and Lemma 2 of Appendix A. Assume further that the CER are true, i.e. H_0 is valid and $\theta_{\dagger}$=θ_0. Then the following results holds as $T \longrightarrow \infty$:
(i) $\hat{\theta}_T^*$ satisfies the convergences in eq.s (31)-(33);
(ii)

$$\sup_{x\in\mathbb{R}^{\dim(\theta)}} |G_T^*(x) - G_\infty(x)| \longrightarrow_p 0 \tag{34}$$

where $G_\infty(x)$=$\Phi(x)$ and $\Phi(x)$ is the CDF of the standardized multivariate Gaussian.

Proof: see Appendix A.

Some remarks are in order.

Remark 4 Proposition 1 generalizes Stoffer and Wall's (1991) main result on state space models. Other than formalizing the fact that the asymptotic distribution of $T^{1/2}Z_T^{*}{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\theta}_T,T}(\hat{\theta}_T^{*}-\hat{\theta}_T)$ is multivariate Gaussian as the asymptotic distribution of $T^{1/2}Z_T{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\theta}_T,T}(\hat{\theta}_T-\theta_0)$ and that the bootstrap standard errors consistently estimate the analytic standard errors of the QML estimator, it further proves the consistency of the bootstrap. Thus, Proposition 1 ensures that one can build and interpret the bootstrap confidence intervals for the structural parameters and for the impulse responses implied by strongly identified DSGE models in the 'conventional' way.

Remark 5 By the same arguments, it is possible to prove that under the analog of the regularity conditions of Proposition 1, also the statistic $T^{1/2}U_T^{*}{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\tau}_{u,T},T}(\hat{\tau}_{u,T}^{*}-\hat{\tau}_{u,T})$, which is the bootstrap analog of $T^{1/2}U_T{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\tau}_{u,T},T}(\hat{\tau}_{u,T}-\tau_{u,0})$, is asymptotically multivariate Gaussian in strongly identified DSGE models. Therefore, by using standard arguments it can be shown that in strongly identified DSGE models, the asymptotic distribution of the bootstrap QLR test is χ_r^2 under the null H_0.

In the next section we show how the consistency result of Proposition 1 can be exploited to design a novel, computational straightforward, bootstrap-based mispecification test for an estimated DSGE model.

4.3 A novel mispecification test

Other than producing the bootstrap p-value for the QLR test of the CER (as well as standard errors for the estimated parameters), the bootstrap algorithm presented in section 4.1 can also play a relevant role in assessing how 'close or distant' is the estimated DSGE model from the case where all regularity conditions for inference hold.

Consider the CDF $G_T^{*}(x)$ defined in the previous section. To simplify exposition we temporarily assume that $K_T^{*}{:=}T^{1/2}Z_T^{*}$ is a scalar.[7] We can estimate $G_T^{*}(x)$ by using

$$G_{T,B}^{*}(x){:=}\frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\{K_T^{*,b}\leq x\}$$

[7] Obviously, also $K_T^{*}{:=}QLR_T^{*}(\hat{\theta}_T^{*})$ is a possible choice.

where $K_T^{*,b}$, $b = 1, ..., B$, are independent draws from $G_T^*(x)$ (the conditional distribution of K_T^*, given the original sample). For any fixed $x \in \mathbb{R}$, by the strong law of large numbers, as $B \to \infty$,

$$G_{T,B}^*(x) \longrightarrow G_T^*(x), \text{ a.s.}$$

and, moreover, by the CLT

$$B^{1/2}(G_{T,B}^*(x) - G_T^*(x)) \xrightarrow{d} N(0, V_{G,T}(x)) \tag{35}$$

where $V_{G,T}(x) := G_T^*(x)(1 - G_T^*(x))$.

Consider now the statistic $SN_B(x) := B^{1/2}(G_{T,B}^*(x) - \Phi(x))$. This statistic represents the (normalized) distance between the estimated (over B repetitions) bootstrap distribution $G_{T,B}^*(x)$ and its theoretical limiting distribution arising when standard regularity conditions hold. We can decompose $SN_B(x)$ as follows:

$$\begin{aligned} SN_B(x) &= B^{1/2}(G_{T,B}^*(x) - G_T^*(x)) + B^{1/2}(G_T^*(x) - \Phi(x)) \\ &= B^{1/2}(G_{T,B}^*(x) - G_T^*(x)) + \left(\frac{B}{T}\right)^{1/2} T^{1/2}(G_T^*(x) - \Phi(x)). \end{aligned} \tag{36}$$

The first term on the left-hand side of eq. (36) satisfies, as the number of bootstrap repetitions B grows, the CLT in eq. (35). Regarding the leading factor of the second term, $T^{1/2}(G_T^*(x) - \Phi(x))$, under our regularity conditions the consistency result in Proposition 1 can be strengthened by using the asymptotic pivotatility of the bootstrap statistic (see also the discussion in Kilian, 1998, and Bose, 1988), the term $T^{1/2}(G_T^*(x) - \Phi(x))$ in the right-hand side of eq. (36) is of $O_p(1)$ in strongly identified DSGE models. Therefore, the term $B^{1/2}(G_T^*(x) - \Phi(x))$ is of $O_p((\frac{B}{T})^{1/2})$ and converges to zero provided $B = o(T)$. This implies that $SN_B(x)$ is asymptotically Gaussian if both $B, T \to \infty$ under the condition

$$BT^{-1} \to 0.$$

Conversely, in the absence of the consistency result of Proposition 1, the second term on the right hand side of eq. (36) does not vanishes asymptotically and $SN_B(x)$ diverges for large enough B. This for instance happens in weakly identified DSGE models. These arguments can easily be extended to the case where $K_T^* := T^{1/2} Z_T^*$ is a $\dim(\theta) \times 1$ vector.

The asymptotic behaviour of $SN_B(x)$ under strong identification as well as under cases where the regularity conditions for standard inference fail, suggests a computational straightforward bootstrap-based misspecification test for DSGE models. The test can be constructed as follows. For a proper choice of B and T, and for any $i = 1, ..., \dim(\theta)$, we can apply standard normality tests to the sequence of bootstrap repetitions $\hat{\theta}_{i,T}^{*,1}$, $\hat{\theta}_{i,T}^{*,2}$, ..., $\hat{\theta}_{i,T}^{*,B}$ obtained using the bootstrap algorithm of section 4.1. The same can be done, for any $j = 1, ..., \dim(\tau_u)$, on the bootstrap repetitions $\hat{\tau}^{*,1}_{u,j,T}$, $\hat{\tau}^{*,2}_{u,j,T}, ..., \hat{\tau}^{*,B}_{u,j,T}$ obtained from the estimation of the DSGE model without imposing the CER. Then, one would expect not to reject normality on $\hat{\theta}^{*,1}_{i,T}$, $\hat{\theta}^{*,2}_{i,T}$, ..., $\hat{\theta}^{*,B}_{i,T}$ and on $\hat{\tau}^{*,1}_{u,j,T}$, $\hat{\tau}^{*,2}_{u,j,T}$, ..., $\hat{\tau}^{*,B}_{u,j,T}$ in DSGE models where all regularity conditions for standard inference hold. Conversely, one would expect to reject normality on $\hat{\theta}^{*,1}_{i,T}$, $\hat{\theta}^{*,2}_{i,T}$, ..., $\hat{\theta}^{*,B}_{i,T}$ and/or on $\hat{\tau}^{*,1}_{u,j,T}$, $\hat{\tau}^{*,2}_{u,j,T}$, ..., $\hat{\tau}^{*,B}_{u,j,T}$ when the estimated DSGE model does not meet the regularity conditions for standard inference, as it happens e.g. under weak identification (or unidentification). We explore the empirical performance of this bootstrap-based misspecification test for DSGE models in Sections 5 and 6 below.

4.4 Connections with the literature

To our knowledge, there are only a few of papers in the literature that use bootstrap methods in DSGE models. As already observed in the Introduction, Chow and Moreno (2006) and Bårdsen and Fanelli (2015) exploit the potential of the bootstrap in the empirical analysis of dynamic structural models but limit their attention to the case in which all state variables are observed and do not address the consistency of the bootstrap. Bekaert and Hodrick (2001) and Fanelli and Palomba (2011) show empirically that bootstrap methods lead to substantial finite-sample size improvements in tests of the CER in the class of present value models used in financial and macroeconomic analysis, but do not prove neither the consistency of the bootstrap, nor whether it provides valid asymptotic refinements. Moreover, the class of DSGE models we consider in this paper in not covered by their analysis.

Fève *et al.* (2009) suggest a minimum-distance Structural VAR (SVAR) approach for DSGE models in which in a first-step the model is estimated and tested by the overidentification test delivered by the minimum-distance program, and in a second-step simulated versions of the overidentification test are obtained by bootstrapping the SVAR residuals. Hence, in Fève's *et al.* (2009) bootstrap resampling involves the residuals of the auxiliary SVAR and not the residuals of the DSGE model. Similarly, Le *et al.* (2011) combine the bootstrap with indirect inference methods for DSGE models. Compared to this literature, the distinctive feature of our approach is that we apply the bootstrap to the state space representation of the DSGE model, resampling the innovation residuals. Moreover, we also show that the bootstrap can be used constructively also in the situations in which the standard 'regularity conditions' for inference are not valid. In this respect, Zhan (2014) is the only example in the literature where the bootstrap is used as a diagnostic tool to detect 'weak instruments' in instrumental variables regressions.

Focusing on the more general setup of state space models, the only study which addresses the problem of bootstrap maximum likelihood estimation in state space models is the already mentioned article by Stoffer and Wall (1991), see also Berkowitz and Kilian (2000) and Stoffer and Wall (2004). Cavanaugh and Shumway (1997) exploite Stoffer and Wall's (1991) algorithm to develop a bootstrap-based corrected variant of Akaike Information Criterion (AIC) to small-sample state-space model selection. We borrow from Stoffer and Wall (1991) the design of our bootstrap algorithm, but we generalize and strengthen their results by proving that the bootstrap is consistent and reduces the tendency of QLR tests to overreject the CER in the strong identification case. Moreover, we further show that the bootstrap can be used constructively to detect misspecification in state space models, including the case of weak identification.[8]

5 Monte Carlo study

In this section, we investigate the empirical performance of our bootstrap approach to DSGE models through a set of Monte Carlo experiments. We are particularly interested in the empirical performance of our novel bootstrap approach to strong/weak identification detection, see Section 4.3. We focus on the ARMA(1,1) model as data generating process, which is particularly suited to characterize strong and weak identification.

Let y_t be a scalar that obeys the ARMA(1,1) model:

$$y_t = (\pi + \beta)y_{t-1} + \omega_t - \pi\omega_{t-1} \quad , \quad \omega_t \sim iidN(0,1), \quad t = 1, ..., T \tag{37}$$

where y_0 and ω_0 are given, and the vector of parameters before any restriction is imposed is τ_u:=$(\tau_1, \tau_2)'$:=$(\pi, \pi+\beta)'$, where β can be interpreted as the difference between the autoregressive ($\tau_2 = \pi + \beta$) and moving average ($\tau_1 = \pi$) coefficients. It is seen that when $\beta = 0$ ($\tau_2 = \tau_1$), the model collapses to

$$y_t = \omega_t$$

and the moving average parameter π is not identified. In this case, our Assumptions 2-3 are violated and the conditions for the consistency of the QML estimator are not satisfied. When β is close to zero (but different from zero so that Assumptions 2-3 still hold), the likelihood function of the ARMA(1,1) model is relatively flat in the direction of π. This model satisfies the parameterization of Andrews and Cheng (2012). Andrews and Mikusheva (2015) show that modeling the parameter β by the embedding $\beta_T = C/T^{1/2}$, where C is a constant, in the weak

[8] Interestingly, the 'nearly a parameter redundant ARMA(2,2) process' estimated by Stoffer and Wall (1991) in their Monte Carlo section (Section 4.1) can be interpreted as a weakly identified model. Stoffer and Wall (1991, p.0126) observe: 'By employing the bootstrap, we obtain vital information concerning the problem with model specification due to near parameter redundancy when sample size are small'. We formalize such an idea by introducing a novel test for strong/weak identification detection based on the bootstrap.

identification case, suitably normalized versions of the measures of information $\mathcal{I}^{OP}_{\theta_0,T}$ and $\mathcal{I}^{2D}_{\theta_0,T}$ converge in the limit to different quantities, in particular:

$$K_T(\mathcal{I}^{OP}_{\theta_0,T} - \mathcal{I}^{2D}_{\theta_0,T})K_T' \to_p M_{\dim(\theta)\times\dim(\theta)}, \quad M_{\dim(\theta)\times\dim(\theta)} \text{ random}$$

where K_T is a normalization (typically diagonal) matrix. The equation above implies that $\mathcal{I}^{OP}_{\theta_0,T}$ and $\mathcal{I}^{2D}_{\theta_0,T}$ are no longer interchangeable measures of information even if White's (1982) information matrix equality $E\left(\check{\mathcal{I}}^{OP}_{\theta_0,T}(\theta_0) - \check{\mathcal{I}}^{2D}_{\theta_0,T}(\theta_0)\right) = 0_{\dim(\theta)\times\dim(\theta)}$ is valid.

We assume that the model in eq. (37) is stationary and invertible and consider the testing problem

$$H_0'': \pi = 0.4 \quad \text{vs} \quad H_1'': \pi \neq 0.4. \tag{38}$$

Under H_0'', the (minimal) state-space representation associated with eq. (37) is given by

$$Z_{m,t} = \underbrace{\begin{pmatrix} 0.4+\theta & 1 \\ 0 & 0 \end{pmatrix}}_{A(\tau_\theta)} Z_{m,t-1} + \underbrace{\begin{pmatrix} 1 \\ -0.4 \end{pmatrix}}_{B(\tau_\theta)} \omega_t \tag{39}$$

$$y_t = \underbrace{(1,0)}_{H^+(\tau_\theta)} Z_{m,t} \tag{40}$$

hence $\theta = \beta$ is the vector (scalar) of structural parameters that appears under the null. The associated representation without imposing the restriction (i.e. under H_1'') is given by

$$Z_{m,t} = \underbrace{\begin{pmatrix} \pi+\beta & 1 \\ 0 & 0 \end{pmatrix}}_{A(\tau_u)} Z_{m,t-1} + \underbrace{\begin{pmatrix} 1 \\ -\pi \end{pmatrix}}_{B(\tau_u)} \omega_t \tag{41}$$

$$y_t = \underbrace{(1,0)}_{H^+(\tau_u)} Z_{m,t} \tag{42}$$

where τ_1=π and τ_2=$(\pi+\beta)$ satisfy the conditions -1<τ_1<1, -1<τ_2<1 that ensure stationarity and invertibility of the ARMA(1,1) counterpart. In this specific example, the relationship $\tau_\theta = g(\theta)$ implied by the null hypothesis H_0'' is linear and can be summarized by the expression:

$$\tau_u = \begin{pmatrix} \tau_1 \\ \tau_2 \end{pmatrix} = \begin{pmatrix} \pi \\ \pi+\beta \end{pmatrix} = \text{under } H_0'' = \begin{pmatrix} 0.4 \\ 0.4+\theta \end{pmatrix} = \tau_\theta.$$

We generate samples of length $T = 100$ and $T = 250$ from the ARMA (1,1) model under the null H_0'' in eq. (38), $M = 5000$ times, setting y_0=$\omega_0 = 0$ and assuming a Gaussian distribution for ω_t. For each replication, a sample of $T + 200$ observations is actually generated and the first 200 observations are then discarded. We consider different scenarios that depend on the values assumed by the parameter $\theta(=\beta)$. As in Andrews and Cheng (2012), we select $\theta(=\beta)$ from the set {-0.76, -0.05}, hence results are partly comparable with their work. The data generating process which corresponds to the case of strong identification is obtained with $\beta(=\theta)$=-0.76, while $\beta(=\theta)$=-0.05 characterizes the weak identification scenario. In the Technical Supplement we also report results for the case of unidentification. ($\theta(=\beta)$=0).

For each considered data generating process and on each generated sample, we test the hypothesis H_0'' against H_1'' in eq. (38), applying the algorithm discussed in Section 4.1 and using $B = T - 1$ bootstrap replications to compute the bootstrap p-value associated with the QLR test. We run a restricted bootstrap procedure (i.e. the bootstrap-pseudo samples are generated under H_0'') except where explicitly indicated. Both under H_0'' and H_1'', the log-likelihood is maximized by the BFGS method, imposing that the optimization parameter spaces for the MA and AR coefficients are constrained to [-0.85, 0.85] and [-0.90, 0.90], respectively.[9] Since in this experiment the specified likelihood is correctly Gaussian, throughout this section we use the terms 'ML estimator' and 'LR test' in place of 'QML estimator' and 'QLR test'.

Instead, to evaluate the empirical size of the normality tests for strong/weak identification, we apply our bootstrap algorithm for different choices of B in the set $\{19, 29, 39, 49, T-1\}$. The so-obtained bootstrap repetitions $\hat{\theta}_T^{*,1}, ..., \hat{\theta}_T^{*,B}$ (under H_0'') and $\hat{\pi}_T^{*,1}, ..., \hat{\pi}_T^{*,B}$ and $\hat{\beta}_T^{*,1}, ..., \hat{\beta}_T^{*,B}$ (under H_1'') are used to compute Henze and Zirkler's (1990) normality test at the 5% nominal level of significance.[10] The normality tests are calculated using both the restricted and unrestricted bootstrap procedure, i.e. generating the bootstrap samples imposing both the null H_0'' and not imposing the null, respectively.

Strong identification

We start our investigation from the case $\beta_0(=\theta_0)$=-0.76 which characterizes a strongly identified ARMA(1,1) process. Estimation and testing results are summarized in Table 1. We notice that the ML estimates of the parameters π and β under H_1'' are substantially similar to their bootstrap counterparts and tend to converge to their true population values as T increases. For both T=100 and T=250, the Hessian-based standard errors associated with $\hat{\pi}_T$ and $\hat{\beta}_T$ are similar to the bootstrap standard errors. The similarity between sample estimates and bootstrap estimates is also observed for $\hat{\beta}_T(=\hat{\theta}_T)$ when the null H_0'' is imposed. For both T=100 and T=250, the rejection frequency of the LR test for H_0'' against H_1'' is close to the nominal 5% level.

[9] All results in this and the next sections have been obtained in Matlab. Codes are available on request to the authors.

[10] Obviously, there are many available normality tests that might be applied. The main feature of Henze and Zirkler's (1990) test is that according to our and other simulation studies (see e.g. Szèkely and Rizzo, 2005), it proves to be less sensitive against heavy tail alternatives relative to other normality tests such as e.g. Doornik and Hansen's (1998) 'ominibus' test (which is explicitly built on the estimated kurtosis of the distribution under investigation). Even if we impose 'mild' bounds on the optimization parameter space, a normality test which is less sensitive to heavily tails seems appropriate in the context of DSGE models.

The bootstrap counterpart of the LR test displays similar rejection frequencies. In particular, the bootstrap distribution of the LR test seems to match perfectly the χ^2_1 approximation. (More on this in the Technical Supplement).

The empirical rejection frequencies of the normality tests associated with the bootstrap repetitions of the structural parameter estimators are reported in Tables 2a-2d, respectively. Tables 2a-2b summarize the results obtained with the restricted bootstrap while Tables 2c-2d, which are confined in the Technical Supplement to save space, summarize the results obtained with the unrestricted bootstrap. The analysis in Section 4.3 suggests that $B = o(T)$ for the test to be asymptotically valid, and this fact seems to be fully reflected in the simulation results which show that the lower B relative to T, the closer is the empirical size to the 5% nominal level of significance.

Overall, the results in Table 1 and Tables 2a-2d show that in strong identified models, the bootstrap works in the expected direction. In particular, our simple test for strong identification displays a good finite sample size coverage for proper choices of B relative to T.

Weak identification

We move to the case ($\beta_0(=\theta_0)$=-0.05), which characterizes a weakly identified 'with near cancelling roots' ARMA(1,1) process. Estimation and testing results are summarized in Table 3. In this case, the ML estimator of π is not consistent (Andrews and Cheng 2012) and this is fully reflected in the estimates of Table 3. However, under H_0'' the parameter π is fixed, while $\beta(=\theta)$ plays the role of nuisance parameter in the testing problem of H_0'' against H_1''. Andrews and Mikusheva (2015) show that $\hat{\beta}_T(=\hat{\theta}_T)$ is consistent and asymptotically normal under the null, and this is also fully reflected in the results of Table 3. For both T=100 and T=250, we also observe a substantial mismatch between the ML estimates of π and β and the average of the bootstrap replicates under H_1''. The mismatch between analytic (Hessian-based) and bootstrap standard errors seems to increase with T. Instead, the ML estimates of $\beta(=\theta)$ under H_0'' are substantially similar to the average of the bootstrap ML estimates and tend to converge to the true population values as T increases, consistently with Andrews and Mikusheva's (2015) finding. Interestingly, for both T=100 and T=250, the rejection frequency of the LR test for H_0'' against H_1'' and of its bootstrap analog is close to the nominal 5% level, suggesting that in this specific case, the χ^2_1 distribution represents still a good approximation, as also confirmed by Andrews and Cheng (2012).

The empirical rejection frequencies of the normality tests associated with the bootstrap repetitions of the structural parameter estimators are reported in Tables 4a-4d for different values of B, respectively. Tables 4a-4b summarize the results obtained with the restricted bootstrap and Tables 4c-4d, which are reported in the Technical Supplement to save space, summarize the results obtained with the unrestricted bootstrap. Notably, in this scenario, our test seems to fully detect the normality of $\hat{\beta}_T^*(=\hat{\theta}_T^*)$ under H_0'', in line with Andrews and Mikusheva's (2015) result. Conversely, and as expected, the rejection frequencies of the normality test of $\hat{\pi}_T^*$ and $\hat{\theta}_T^*$ under H_1'' behave similarly to empirical powers as B increases (relative to T). For values of B for which we had a satisfactorily empirical size control under strong identification, the rejection frequency of the test lies in the range 30%-54%. If we combine the information of these tests with the discrepancy observed between analytic standard errors and bootstrap standard errors, our approach seems to detect weak identification reasonably well.

Overall, the results in Table 3 and Tables 4a-4d show that in weakly identified DSGE models the bootstrap reflects deviations from normality reasonably well, therfore it is still informative and useful for practitioners.

6 Empirical illustration

In this section we apply our bootstrap approach to estimate and evaluate An and Schorfheide's (2007) DSGE model introduced in the Example 1 of Section 2 on U.S. data. We employ quarterly data relative to the 'Great Moderation' sample 1984Q2-2008Q3. The starting date of our estimation and evaluation sample, 1984Q2, is justified by McConnell and Pérez-Quirós (2000), who find a break in the variance of the U.S. output growth in 1984Q1. The ending date is instead motivated by the fact that, with data after 2008Q3, it would be hard to identify a 'conventional' monetary policy shock with our structural model during the well known zero lower bound (ZLB) episodes. The three observable variables in $y_t{:=}(\tilde{x}_t, \pi_t, R_t)'$ are measured as follows. The output gap, $\tilde{x}_t$, is computed as percent log-deviation of the real GDP with respect to the potential output estimated by the Congressional Budget Office. The inflation rate, π_t, is the quarterly growth rate of the GDP deflator. For the short-term nominal interest rate, R_t, we consider the effective Federal funds rate expressed in quarterly terms (averages of monthly values). The source of the data is the Federal Reserve Bank of St. Louis' web site.

We consider two estimable versions of the model. In the former, denoted M1, the vector of unknown parameters is ${}^1\theta{:=}(\kappa, \rho_r, \sigma_g^2)'$ ($\dim({}^1\theta){=}3$) and all other structural parameters of system (3)-(8) are calibrated at the values reported in the upper panel of Table 1 in Komunjer and Ng (2011) and are assumed to be known by the econometrician. In the latter, denoted M2, the vector of unknown parameters is ${}^2\theta{:=}(\kappa, \psi_1, \rho_r, \sigma_e^2, \sigma_g^2, \sigma_r^2)'$ ($\dim({}^2\theta){=}6$), and all other parameters that characterize system (3)-(8) are always calibrated at the values reported in the upper panel of Table 1 in Komunjer and Ng (2011), and are assumed to be known by the econometrician. Our empirical results are conditional upon the calibrated values of the parameters which are not estimated directly from the data.

The main difference between model M1 and model M2 is that M2 estimates the Central Bank's long run response to inflation ψ_1 from the data and the variance of all three fundamental shocks of the system, while M1 treats ψ_1 (other than the Central Bank's long run response to output gap ψ_2) and the variances σ_e^2 and σ_r^2 as fixed. Cochrane (2011) has recently argued that in New Keynesian monetary models, the parameters of Taylor-type rules like that in eq. (6) are not identifiable, while Komujer and Ng (2011) have shown that in the DSGE model in eq.s (3)-(8) it is necessary to fix one among ρ_r, ψ_1 and ψ_2 to achieve (population) identification. Moreover, the recent empirical literature suggests that it is hard to identify the parameters of Taylor-type policy rules on the Great Moderation sample, see e.g. Mavroeidis (2010) and Castelnuovo and Fanelli (2015). Thus it is reasonably to expect identification issues in model M2 relative to M1.

Before discussing estimation issues and the test for the CER, we run preliminary multivariate normality tests on the innovation residuals resulting from the (Gaussian) Kalman filter estimation of the representation of the DSGE model without the CER. The multivariate normality test for the innovation residuals obtained from the state space model under H_1 has p-value 0.081. Moreover, the largest eigenvalue of the matrix $A(\hat{\tau}_u)$ is 0.947, which suggests that the system features persistent variables. Considering the state space model under the CER (H_0), in model M1 the multivariate normality test for the innovation residuals has p-value 0.70 and for model M2 p-value 0.27. Based on these preliminary evidences, we can interpret our QML estimators as ML estimators and the QLR test for the CER as LR test. Throughout this section we use the terms 'ML estimator' and 'LR test' in place of 'QML estimator' and 'QLR test'.

Estimation is carried out by combining the Kalman filter with the 'CMA-ES' likelihood maximization algorithm (Andreasen, 2010). For both models, the bootstrap standard errors associated with the QML estimates of the structural parameters and the bootstrap p-values associated with the LR tests are computed using B=999 replications, following the algorithm summarized in Section 4.1. The normality tests on the bootstrap repetitions of the structural parameter estimators are based on B=29 repetitions. These choices are motivated by the fact that our mispecification test requires B=$o(T)$. As in the Monte Carlo experiment, we use Henze and Zirkler's (1990) normality test. Results are summarized in Table 5.

The upper panel of Table 5 refers to the version M1 of An and Schorfheide's (2007) DSGE model. We notice that in this case, the average of the bootstrap parameter estimator ${}^1\hat{\theta}_T^*$ (fourth column) across bootstrap repetitions is close to the ML estimate computed on the original sample ${}^1\hat{\theta}_T$ (second column), the main difference characterizing the policy rule inertia parameter ρ_r. Similarly, the bootstrap standard errors (fifth column) tend to reproduce the Hessian-based standard errors (third column) and, again, the main distance between the two measures of variability seems to characterize the inertia parameter ρ_r. Using asymptotic critical values (taken from the χ^2_{10} distribution, where r=dim(τ_u)−dim($^1\theta$)=13-3=10), the LR test rejects the CER at the 5% nominal significance level. Instead, the (restricted) bootstrap p-value associated with the LR test is equal to 0.053 and is significant considering the usual measure of variability, $\left(\frac{\widehat{p}^*_{T,B}(1-\widehat{p}^*_{T,B})}{B}\right)^{1/2}$, where $\widehat{p}^*_{T,B}$ is given in eq. (30). Moreover, it is seen that the normality of the bootstrap repetitions of the structural parameter estimators is not rejected for any parameter at the 5% nominal level of significance. We can safely conclude that version M1 of the estimated An and Schorfheide's (2007) DSGE model is strongly identified and, notably, is not rejected at the 5% nominal level of significance when the bootstrap version of the LR test for the CER is considered.

The lower panel of Table 5 refers to the version M2 of An and Schorfheide's (2007) DSGE model. We observe that in this case, the magnitudes of the ML estimates of κ (slope of the Phillips curve) and of σ^2_g (variance of the shock in the Phillips curve) are substantially different from the magnitudes obtained with model M1. Moreover, the ML estimate of ψ_1 lies on the (upper) boundary of the optimization parameter space (see the Notes of Table 5 for detail). Compared to model M1, the discrepancy between the average across bootstrap repetitions of $^2\hat{\theta}^*_T$ (fourth column) and the ML estimate computed on the original sample $^2\hat{\theta}_T$ (second column) is larger and can be essentially ascribed to the estimate of ρ_r (but the discrepancy is more pronounced compared to what observed in model M1) and σ^2_g. The bootstrap standard errors (fifth column) diverge substantially from the Hessian-based standard errors (third column). Using asymptotic critical values (taken from the χ^2_7 distribution, where r=dim(τ_u)−dim($^2\theta$)=13-6=7), the LR test rejects the CER at the 5% nominal significance level. Also the (restricted) bootstrap p-value rejects the CER. The normality of the bootstrap repetitions of the structural parameter estimators is rejected for all parameters at the 5% nominal level of significance, except for $\hat{\sigma}^{2*}_g$, whose normality test displays a p-value of 0.14. Thus for model M2 we have clear evidence that, compared to the model M1, the estimation of the additional parameter of the policy rule seems to affect the regularity conditions for standard inference, inducing weak identification issues. It turns out that we can not interpret the asymptotic p-value and the bootstrap p-value computed for the LR test for the CER in the conventional way. In this case, it is more safe to evaluate the CER implied by model M2 implicitly by computing identification-robust confidence sets for $^2\theta$ along the lines discussed in e.g. Guerron-Quintana *et al.* (2013).

7 Conclusions

In this paper we have explored the potential of bootstrap methods in the empirical evaluation of DSGE models. We have shown that with out bootstrap approach, the inference on the CER implied by DSGE models is considerably more accurate and the risk of false rejections is under strict control if the DSGE model is strongly identified. We have further proposed a novel bootstrap misspecification approach which can be used to analyze the strength of identification of the estimated model when all other possible causes of misspecification are under control, which seems to be promising, based on the evidences produced in the paper, and straightforward to apply.

An important dimension that it is worth exploring is the role of bootstrap resampling in the construction of identification-robust confidence intervals for the structural parameters. The approach suggested in the current literature to deal with weakly identified DSGE models requires test inversion techniques, see e.g. Guerron-Quintana *et al.* (2013), Dufour *et al.* (2009, 2013) and Andrews and Mikusheva (2015). In DSGE models, test inversions typically require the use of numerical procedures (grid testing). In these cases, bootstrap methods can potentially be applied to produce more accurate identification-robust confidence sets for the structural parameters, albeit the combination of numerical inversion of tests with Monte Carlo simulations can be computationally costly. We have left the exploration of this important topic to future research.

A Appendix: consistency of the bootstrap in strongly identified DSGE models

In this Appendix, we provide heuristic proofs of Proposition 1 and Proposition 2. Our heuristic proof includes two small preliminary sections in which we fix the notation necessary to understand the bootstrap setting, and define the objective functions.

Notation

We use $P(\cdot)$, $E(\cdot)$ and $Var(\cdot)$, respectively, to denote probability, expectations and variance under the true probability distribution of the data. We use $P^*(\cdot)$, $E^*(\cdot)$ and $Var^*(\cdot)$, respectively, to denote probability, expectations and variance induced by the bootstrap resampling process, conditional on the original sample. Thus, $P^*(\cdot)$, $E^*(\cdot)$ and $Var^*(\cdot)$ denote stochastic quantities with respect to the probability distribution of the data. To fix ideas, define, for any $\delta > 0$, $p_T{:=}P\left(\left\|\hat{\theta}_T - \theta_0\right\| > \delta\right)$ and $p_T^*{:=}P^*\left(\left\|\hat{\theta}_T^* - \hat{\theta}_T\right\| > \delta\right)$, where $\hat{\theta}_T^*$ is the bootstrap analog of the QML estimator $\hat{\theta}_T$, and $\|\cdot\|$ is the Euclidean norm. While with the conventional notation '

$\hat{\theta}_T \longrightarrow_p \theta_0$' we mean that the (deterministic) sequence $\{p_T\}$ converges to zero ($p_T \longrightarrow 0$), with the notation ' $\hat{\theta}_T^* \xrightarrow{p^*}_p \hat{\theta}_T$ ', which reads '$\hat{\theta}_T^*$ convergences in P^* in probability to $\hat{\theta}_T$', we mean that the (stochastic) sequence $\{p_T^*\}$ converges in probability to zero ($p_T^* \longrightarrow_p 0$).

Likewise, for any $\delta > 0$, define $s_T := P\left(\sup_{\theta\in\mathcal{P}^D}\left|\hat{Q}_T(\theta) - Q_0(\theta)\right| > \delta\right)$, where $\hat{Q}_T(\theta)$ and $Q_0(\theta)$ are the criterion functions defined below, and let $s_T^* := P^*\left(\sup_{\theta\in\mathcal{P}^D}\left|\hat{Q}_T^*(\theta) - Q_0(\theta)\right| > \delta\right)$, where $\hat{Q}_T^*(\theta)$ is the bootstrap analog of $\hat{Q}_T(\theta)$ given the original sample, see below. While the concept '$\hat{Q}_T(\theta)$ converges uniformly in probability to $Q_0(\theta)$' implies that the (deterministic) sequence $\{s_T\}$ converges to zero ($s_T \longrightarrow 0$), the concept '$\hat{Q}_T^*(\theta)$ converges uniformly in P^* in probability to $Q_0(\theta)$' means that the stochastic sequence $\{s_T^*\}$ converges to zero in probability ($s_T^* \longrightarrow_p 0$). Alternative useful notations are $\sup_{\theta\in\mathcal{P}^D}\left|\hat{Q}_T(\theta) - Q_0(\theta)\right| \rightarrow_p 0$ (uniform convergence in probability) and $\sup_{\theta\in\mathcal{P}^D}\left|\hat{Q}_T^*(\theta) - Q_0(\theta)\right| \xrightarrow{p^*}_p 0$ (uniform convergence in P^* in probability), respectively.

Finally, given the scalar stochastic sequence $\{W_T\}$ and the random variable W, consider the CDFs $G_{W,T}(x) := P(W_T \leq x)$ and $G_W(x) := P(W \leq x)$. The conventional notation '$W_T \xrightarrow{d}_p W$' means that $G_{W,T}(x) \longrightarrow G_W(x)$ for each x at which $G_W(x)$ is continuos. Let W_T^* be the bootstrap analog of W_T. Given the sequence $\{W_T^*\}$ and the CDF $G_{W^*,T}^*(x) := P^*(W_T^* \leq x)$, we say that W_T^* 'converges in conditional distribution D^* in probability to W', denoted '$W_T^* \xrightarrow{d^*}_p W$', if $G_{W^*,T}^*(x) \xrightarrow{p} G_W(x)$ for each x at which $G_W(x)$ is continuos. This concept can be extended to the multivariate framework in the conventional way.

Objective functions: definitions

Consider $\ell_{\circ,T}(\theta) = \ell_T(\tau_\theta(\theta))$ the essential part of the log-likelihood function associated with the DSGE model, see Section 3.1. We define the normalized log-likelihood:

$$\hat{Q}_T(\theta) := \frac{1}{T}\ell_{\circ,T}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\ell_t(\theta) \tag{43}$$

where $\ell_t(\theta) := l(y_t \mid \mathcal{F}_{t-1}^y \ ; \ \theta) := -\left\{\log\det(\Sigma_{\epsilon^0,t}) + e_t^{0\prime}e_t^0\right\}$, $e_t^0 := \Sigma_{\epsilon^0,t}^{-1/2}\epsilon_t^0$ (see Technical Supplement). Hence, the QML estimator $\hat{\theta}_T$ solves $\hat{\theta}_T = \arg\max_{\theta\in\mathcal{P}^D} \hat{Q}_T(\theta)$. By construction, $\hat{Q}_T(\theta)$ is twice continuosly differentiable in a neighborhood $\mathcal{N}_{\theta_0}$ of θ_0. Associated with $\hat{Q}_T(\theta)$ we have

$$Q_T(\theta) := E(\hat{Q}_T(\theta)) = E\left(\frac{1}{T}\sum_{t=1}^{T}\ell_t(\theta)\right) = \frac{1}{T}\sum_{t=1}^{T}E\left(\ell_t(\theta)\right) = E\left(\ell_t(\theta)\right) \equiv Q_0(\theta) \tag{44}$$

where the last equality holds because of Assumption 1 (recall that Assumption 1 implies that $\{y_t\}$ is weakly stationary and we added the hypothesis of ergodicity).

In the bootstrap setting, the analog of $\hat{Q}_T(\theta)$ is given by

$$\hat{Q}_T^*(\theta) := \frac{1}{T}\ell_{\circ,T}^*(\theta) = \frac{1}{T}\sum_{t=1}^{T}\ell_t^*(\theta) \tag{45}$$

where $\ell_t^*(\theta){:=}l^*(y_t^* \mid \mathcal{F}_{t-1}^{y^*};\theta) = -\left\{\log\det(\Sigma_{\epsilon^0,t}) + e_t^{0*\prime}e_t^{0*}\right\}$, and $\hat{e}_t^{0,*}$, $t = 1, ..., T$, is a bootstrap sample of standardized innovations, which can be obtained e.g. through the algorithm discussed in Section 4.1. Obviously, $\hat{\theta}_T^*{:=}\arg\max_{\theta\in\mathcal{P}^D}\hat{Q}_T^*(\theta)$. Associated with $\hat{Q}_T^*(\theta)$ we have

$$Q_T^*(\theta){:=}E^*(\hat{Q}_T^*(\theta)) = E^*\left(\frac{1}{T}\sum_{t=1}^{T}\ell_t^*(\theta)\right) = \frac{1}{T}\sum_{t=1}^{T}E^*\left(\ell_t^*(\theta)\right) = E^*\left(\ell_t^*(\theta)\right) = Q_0^*(\theta) \tag{46}$$

where, again, the last equality holds because the covariance stationarity and ergodicity of $\{y_t\}$ is inherited by $\{y_t^*\}$ under the bootstrap algorithm discussed in Section 4.1.

Consistency of the bootstrap QML estimator

To prove the consistency in eq. (31), we proceed by showing that we have bootstrap analogs of the conditions (TS-A.i)-(TS-A.iv) reviewed in the Technical Supplement that are valid for the criterion function $\hat{Q}_T^*(\theta)$ evaluated at $\hat{\theta}_T$.

We need the Lemma that follows.

Lemma 1 (Stoffer and Wall, 1991)

$$Q_T^*(\theta) = \hat{Q}_T(\theta) \;\text{ for all } \theta\in\mathcal{P}^D$$

Proof: Stoffer and Wall (1991).

By Lemma 1, $\hat{\theta}_T{=}\arg\max_{\theta\in\mathcal{P}^D} Q_0^*(\theta)$, hence $\hat{\theta}_T$ is the unique maximizer of $Q_0^*(\theta)$ in a neighborhood $\mathcal{N}_{\hat{\theta}_T}$ of $\hat{\theta}_T$. This condition is the bootstrap analog of (TS-A.ii) when (TS-A.ii) is restricted in a neighborhood $\mathcal{N}_{\theta_0}$ of θ_0. Always by Lemma 1, $Q_0^*(\theta)$ is continuos in θ since so is $\hat{Q}_T(\theta)$. This is the bootstrap analog of (TS-A.iii). If in the presence of stationary and ergodic processes $\{\ell_t(\theta)\}$ satisfies a UWLLN, also the bootstrap counterpart $\{\ell_t^*(\theta)\}$ will satisfy a UWLLN in P^* in probability, i.e.

$$P^*\left(\sup_{\theta\in\mathcal{P}^D}\left|\hat{Q}_T^*(\theta) - Q_0^*(\theta)\right|\right) \xrightarrow{p} 0 \text{ as } \; T\to\infty$$

hence by Lemma 1:

$$P^*\left(\sup_{\theta\in\mathcal{P}^D}\left|\hat{Q}_T^*(\theta) - \hat{Q}_T(\theta)\right|\right) \xrightarrow{p} 0 \text{ as } \; T\to\infty. \tag{47}$$

eq. (47) is the bootstrap analog of (TS-A.iv).

Summing up, we have the following bootstrap analogs of the regularity conditions (TS-A.i)-(TS-A.iv):

(A.i*) $\mathcal{P}^D$ is compact;

(A.ii*) $\hat{\theta}_T$ is the unique maximizer of $Q_0^*(\theta)$ in a neighborhood $\mathcal{N}_{\hat{\theta}_T}$ of $\hat{\theta}_T$;
(A.iii*) $Q_0^*(\theta)$ is continuos in θ;
(A.iv*) $\hat{Q}_T^*(\theta)$ converge uniformily in P^* in probability to $\hat{Q}_T(\theta)$, i.e. it holds eq. (47).
It turns out that

$$\arg \max_{\theta \in \mathcal{N}_{\hat{\theta}_T} \cap \mathcal{P}^D} \hat{Q}_T^*(\theta) =: \hat{\theta}_T^* \xrightarrow{p^*}_p \hat{\theta}_T := \arg \max_{\theta \in \mathcal{N}_{\hat{\theta}_T} \cap \mathcal{P}^D} \hat{Q}_T(\theta). \tag{48}$$

This proves the convergence in eq. (31) of part (i) of Proposition 1.

Asymptotic normality of the bootstrap QML estimator

To prove the asymptotic normality of $T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$, we consider a mean value expansion of $\nabla_\theta \hat{Q}_T^*(\hat{\theta}_T^*) = 0_{dim(\theta)\times 1}$ around the QML estimator $\hat{\theta}_T$, obtaining

$$0_{\dim(\theta)\times 1} = \nabla_\theta \hat{Q}_T^*(\hat{\theta}_T) + \nabla_{\theta\theta}^2 \hat{Q}_T(\bar{\theta}_T^*)(\hat{\theta}_T^* - \hat{\theta}_T)$$

where $\bar{\theta}_T^*$ is on the segment connecting the points $\hat{\theta}_T^*$ and $\hat{\theta}_T$ of $\mathcal{P}^D$. Rearranging terms,

$$T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T) = -\left\{\nabla_{\theta\theta}^2 \hat{Q}_T^*(\bar{\theta}_T^*)\right\}^{-1} T^{1/2} \nabla_\theta \hat{Q}_T^*(\hat{\theta}_T)$$

$$= -\left\{T^{-1} \sum_{t=1}^T \nabla_{\theta\theta}^2 \ell_t^*(\bar{\theta}_T^*)\right\}^{-1} T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T). \tag{49}$$

To prove that the term on the right in eq. (49) converges to the Gaussian, we need a bootstrap analog of the CLT in (TS-A.vi) and we have to prove that the implied long run covariance matrix is asymptotically indistinguishable from $\mathcal{B}(\hat{\theta}_T) := \lim_{T\to\infty} Var\left(T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t(\theta_0)\right)$. Instead, to prove that the term on the left in eq. (49) converges in P* in probability to $\mathcal{A}(\hat{\theta}_T)^{-1}$ we need a bootstrap analog of the uniform convergence result of the Hessian matrix in (TS-A.vii) and a bootstrap analog of the nonsingularity condition (TS-A.viii).

As concerns the bootstrap analog of the CLT in (TS-A.vi), if $T^{1/2}\nabla_\theta \hat{Q}_T^*(\hat{\theta}_T) = T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T)$ satisfies a CLT, then we have:

$$T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T) \xrightarrow{d^*}_p N(0_{\dim(\theta)\times 1}, \mathcal{B}^*) \quad , \quad \mathcal{B}^* := \mathcal{B}^*(\hat{\theta}_T) := \lim_{T\to\infty} Var^* \left(T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T)\right). \tag{50}$$

This result can be further specialized by using the Lemma that follows, which is a reformulation of Lemma 2 in Stoffer and Wall (1991).

Lemma 2 (Stoffer and Wall, 1991)

$$\mathcal{B}^*(\hat{\theta}_T) = Var^* \left(T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T)\right) \xrightarrow{p^*}_p = Var\left(T^{-1/2} \sum_{t=1}^T \nabla_\theta \ell_t(\hat{\theta}_T)\right) = \mathcal{B}(\hat{\theta}_T)$$

Proof: The proof follows from Stoffer and Wall (1991) changing the 'almost surely' convergence with '$\xrightarrow{p^*}_p$'.

As concerns the bootstrap analog of the uniform convergence result of the Hessian matrix in (TS-A.vii), if similarly to $\left\{\nabla^2_{\theta\theta}\ell_t(\theta)\right\}$ also the process $\left\{\nabla^2_{\theta\theta}\ell^*_t(\hat{\theta}_T)\right\}$ satisfies a UWLLN, then

$$\sup_{\theta\in\mathcal{N}_{\theta_0}}\left\|\nabla^2_{\theta\theta}\hat{Q}^*_T(\hat{\theta}_T)-\mathcal{A}^*(\hat{\theta}_T)\right\|\xrightarrow{p^*}_p 0$$

where

$$\mathcal{A}^*(\hat{\theta}_T){:=}\lim_{T\to\infty}E^*\left(-\nabla^2_{\theta\theta}\hat{Q}^*_T(\hat{\theta}_T)\right)=\lim_{T\to\infty}E^*\left(-\frac{1}{T}\sum_{t=1}^{T}\nabla^2_{\theta\theta}\ell^*_t(\hat{\theta}_T)\right)$$
$$=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E^*\left(-\nabla^2_{\theta\theta}\ell^*_t(\hat{\theta}_T)\right)=E^*\left(-\nabla^2_{\theta\theta}\ell^*_t(\hat{\theta}_T)\right)=-\nabla^2_{\theta\theta}\hat{Q}_T(\hat{\theta}_T)$$

Summing up, we have the following bootstrap analogs of the regularity conditions (TS-A.v)-(TS-A.viii):

(A.v*) $\hat{\theta}_T$ is in int$(\mathcal{P}^D)$;

(A.vi*) $T^{1/2}\nabla_\theta\hat{Q}^*_T(\hat{\theta}_T)=T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\ell^*_t(\hat{\theta}_T)$ satisfies the CLT:

$$T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\ell^*_t(\hat{\theta}_T)\xrightarrow{d^*}_p N(0_{\dim(\theta)\times1},\mathcal{B}(\hat{\theta}_T))\quad,\quad\mathcal{B}(\hat{\theta}_T){:=}\lim_{T\to\infty}Var\left(T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\ell^*_t(\hat{\theta}_T)\right);\tag{51}$$

(A.vii*) $\sup_{\theta\in\mathcal{N}_{\hat{\theta}_T}}\left\|\nabla^2_{\theta\theta}\hat{Q}^*_T(\hat{\theta}_T)-\mathcal{A}(\hat{\theta}_T)\right\|\xrightarrow{p*}_p 0$.

(A.viii*) $\mathcal{A}(\hat{\theta}_T)$ is nonsingular.

Coming back to the mean value expansion in eq. (49), we have that the first term on the left converges in P^* in probability to $\mathcal{A}(\hat{\theta}_T)$ because of (A.vii*)-(A.viii*) and the result in eq. (48). The right-hand side term satisfies the convergence in eq. (51) by (A.vi*). It follows that

$$T^{1/2}(\hat{\theta}^*_T-\hat{\theta}_T)\xrightarrow{d^*}_p N(0_{\dim(\theta)\times1},\mathcal{A}(\hat{\theta}_T)^{-1}\mathcal{B}(\hat{\theta}_T)\mathcal{A}(\hat{\theta}_T)^{-1}).\tag{52}$$

This proves the convergence in eq.s (32)-(33) of part (i) of Proposition 1.

The asymptotic normality in eq. (52) implies that the statistic $K^*_T{:=}T^{1/2}Z^*_T{:=}T^{1/2}\check{\mathcal{I}}^{1/2}_{\hat{\theta}_T,T}(\hat{\theta}^*_T-\hat{\theta}_T)$, with has CDF $G^*_T(x){:=}P^*(K^*_{1,T}\leq x_1,...,K^*_{\dim(\theta),T}\leq x_{\dim(\theta)})$ for finite T, converges asymptotically to $G^*_\infty(x)=\Phi(x)$. From Polya's theorem we obtain the consistency in eq. (34) of part (ii) of Proposition 1.

References

An, S., Schorfheide, F. (2007), Bayesian analysis of DSGE models, *Econometric Reviews* 26, 113-172.

Anderson, B.D.O. and Moore, J. (1979), *Optimal filtering*, Englewood Cliffs, NJ: Prentice-Hall.

Andreasen, M. M. (2010), How to maximize the likelihood function for a DSGE model, *Computational Economics* 35, 127-154.

Andrews, D.W.K., Cheng, X. (2012), Estimation and inference with weak, semi-strong, and strong identification, *Econometrica* 80, 2153-2211.

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

Angelini, G., Fanelli, L. (2015), Misspecification and expectations correction in New-Keynesian DSGE models, *Oxford Bulletin of Economics and Statistics*, forthcoming.

Bårdsen, G., Fanelli, L. (2015), Frequentist evaluation of small DSGE models, *Journal of Business and Economic Statistics* 33, 307-322.

Bekaert G., Hodrick, R. (2001), Expectations hypotheses tests, *Journal of Finance* 56, 1357-1394.

Berkowitz, J., Kilian, L. (2000), Recent developments in bootstrapping time series, *Econometric Reviews* 19, 1-48.

Bose, A. (1988), Edgeworth correction by bootstrap autoregressions, *Annals of Statistics* 16, 1709-1722.

Boswijk, P., Cavaliere, G., Rahbek, A. and Taylor, R. A.M. (2015), Inference on Co-integration Parameters in Heteroskedastic Vector Autoregressions, *Journal of Econometrics*, forthcoming.

Canova F., Sala, L. (2009), Back to square one: identification issues in DSGE models, *Journal of Monetary Economics* 56, 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* 30,857-1010.

Cavaliere, G., Rahbek, A., and Taylor R. A. M. (2012), Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models, *Econometrica* 80,1721-1740.

Cavaliere, G., Rahbek, A. and Nielsen, H. B. (2015), Bootstrap testing of hypotheses on co-integration relations in VAR models, *Econometrica* 83, 813–831.

Cavaliere, G., Nielsen, H.B. and Rahbek, A. (2015), On the consistency of bootstrap testing for a parameter on the boundary of the parameter space, mimeo.

Cavanaugh, J.E. and Shumway, R.H. (1997), A bootstrap variant of AIC for state-space model selection, *Statistica Sinica* 7, 473-496.

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

Cho, S., Moreno, A. (2006), A small-Sample study of the New-Keynesian macro model, *Journal of Money Credit and Banking* 38, 1462-1482.

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

Diebold, F. X., Ohanian, L. E., Berkowitz, J. (1998), Dynamic equilibrium economies: a framework for comparing models and data, *Review of Economic Studies* 65, 433-452.

Doornik, J.A., Hansen, H. (2008), An omnibus test for univariate and multivariate normality, *Oxford Bulletin of Economics and Statistics* 70 (Supplement), 927-939.

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.

Fanelli, L. (2012), Determinacy, indeterminacy and dynamic misspecification in linear rational expectations models, *Journal of Econometrics* 170, 153-163.

Fanelli, L., Palomba, G. (2011), Simulation-based tests of forward-looking models under VAR learning Dynamics, *Journal of Applied Econometrics* 26, 762-782.

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

Fève, P., Matheron, J. and Sahuc, J-G. (2009), Minimum distance estimation and testing of DSGE model from structural VARs, *Oxford Bulletin of Economics and Statistics* 71, 883-894.

Franchi, M., Vidotto, A. (2013), A check for finite order VAR representations of DSGE models, *Economic Letters* 120, 100-103.

Franchi, M., Paruolo, P. (2014), Minimality of state space solutions of DSGE models and existence conditions for their VAR represenation, *Computational Economics*, forthcoming.

Guerron-Quintana P, Inoue A, Kilian L. (2013), Frequentist inference in weakly identified DSGE models, *Quantitiative Economics* 4, 197-229.

Hannan, E.J., Deistler, M. (1988), *The Statistical Theory of Linear Systems*, SIAM ed. Wiley, New York.

Hansen, L.P. (2014), Nobel Lecture: Uncertainty outside and inside economic models, *Journal of Political Economy*, 122, 945-987.

Hansen, L. P., Sargent, T. J. (1980), Formulating and estimating dynamic linear rational expectations models, *Journal of Economic Dynamics and Control* 2, 7-46.

Hansen, L. P., Sargent, T. J. (1981), Linear rational expectations models for dynamically interrelated variables, in: R.E.Jr. Lucas and T. J. Sargent, (Eds.), *Rational Expectations and Econometric Practise*. University of Minnesota Press, Minneapolis, pp. 127-156.

Henze, N., Zirkler, B. (1990), A class of invariant and consistent tests for multivariate normality, *Communications in Statistics. Theory and Methods* 19, 3595-36-17.

Horowitz, J.L. (2001), The bootstrap, *Handbook of Econometrics* Vol. 5, (eds.) Heckman, J.J and E. Leamer, Elsevier Science, pp. 3159-3226.

Inoue, A., Kuo, C-H. and Rossi, B. (2015), Indentifying the sources of model misspecification, Working Paper.

Iskrev, N (2008), Evaluating the information matrix in linearized DSGE models, *Economic Letters* 99, 607-610.

Iskrev, N (2010), Local identification in DSGE models, *Journal of Monetary Economics* 57, 189-202.

Kilian, L. (1998), Small-sample confidence intervals for impulse response functions, *Review of Economic and Statistics* 80, 218-230.

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 dynamic stochastic general equilibrium models, *Econometrica* 79, 1995-2032.

Le, V.P.M, Meenagh, D., Minford, P. and Wickens, M. (2011), How much nominal rigidity is there in the US economy? Testing a new Keynesian DSGE model using indirect inference, *Journal of Economic Dynamics and Control* 35, 2078-2140.

Leeper, E. M., Walker, T. B. and Yang, S.-C. S. (2013), Fiscal Foresight and Information Flows. *Econometrica* 81, 1115–1145.

Ljung, L., Caines, P.E. (1979), Asymptotic normality of predictor error estimators for approximate system models, *Stochastics* 3, 29-46.

Lütkepohl, H. (2005), *New introduction to multiple time series*, Springer, Berlin.

Mankiw, N.G., Shapiro, M. D. (1986), Do we reject too often? *Economic Letters* 20, 139-145.

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, 421-448.

Mavroeidis, S. (2010), Monetary policy rules and macroeconomic stability: some new evidence, *American Economic Review* 100, 491-503.

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

Qu, Z., Tkachenko, D. (2012), Identification and frequency domanin quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models, *Quantitative Economics* 3, 95-132.

Ravenna, F. (2007), Vector autoregressions and reduced form representations of DSGE models, *Journal of Monetary Economics* 54, 2048-2064.

Schorfheide, F. (2010), Estimation and evaluation of DSGE models: Progress and challenges, NBER Working Paper n. 141.

Stoffer, D.S. and Wall, K. (1991), Bootstrapping state-space models: Gaussian maximum likelihood estimation and the Kalman filter, *Journal of the American Statistical Association* 86, 1024-1033.

Stoffer, D.S. and Wall, K. (2004), Resampling in state space models, in *State Space and Unobserved Component Models: Theory and Applications*, (eds.) A.C. Harvey, S.J. Koopman and N. Shepard, Chap. 9, Cambridge University Press, Cambridge.

Staiger, D., Stock, J.H. (1997), Instrumental variables regressions with weak instruments, *Econometrica* 65, 557-586.

Stock, J.H., Wright, J. H. (2000), GMM with weak identification, *Econometrica* 68, 1055-1096.

Watson, M. W. (1989), Recursive solution methods for dynamic linear rational expctations models, *Journal of Econometrics* 41, 65-89.

White, H. (1982), Maximum likelihood estimation in misspecified models, *Econometrica* 50, 1-25.

Zhan, Z. (2014), Detecting weak identification by bootstrap, mimeo.

TABLE 1. MC experiment. Strong identification. ML estimates and rejection freq. of LR test.

DGP: ARMA(1,1), $\pi_0 = 0.4$, $\beta_0 = -0.76$, $H_0'': \pi = 0.4$ vs $H_1'': \pi \neq 0.4$						
	T=100					
	MONTE CARLO			BOOTSTRAP (B=T-1)		
	Under H_0''	Under H_1''		Under H_0''	Under H_1''	
	$\hat{\beta}_T(=\hat{\theta}_T)$	$\hat{\pi}_T$	$\hat{\beta}_T$	$\hat{\beta}_T^*(=\hat{\theta}_T^*)$	$\hat{\pi}_T^*$	$\hat{\beta}_T^*$
Average	-0.7542	0.4089	-0.7597	-0.7495	0.4089	-0.7541
s.e.	0.0949	0.1487	0.1040	0.0924	0.1508	0.1012
Hessian-based s.e.	0.0948	0.1418	0.1022	0.0961	0.1586	0.1077
	Rej(LR_T)=0.052			Rej(LR_T^*)=0.040		
	T=250					
Average	-0.7588	0.4028	-0.7602	-0.7574	0.4039	-0.7591
MC s.e.	0.0585	0.0893	0.0632	0.0588	0.0903	0.0632
Hessian-based s.e.	0.0594	0.0976	0.0674	0.0598	0.0984	0.0667
	Rej(LR_T)=0.049			Rej(LR_T^*)=0.0470		

NOTES: Results are based on M=1000 Monte Carlo replications. 'Average' is the average of the ML estimator across Monte Carlo replications. 'MC s.e.' indicates the standard errors computed across Monte Carlo replications. 'Hessian-based s.e.' indicates the average across Monte Carlo replications of the standard errors computed from the Hessian matrix. 'Rej($\cdot$)' denotes rejection frequency (across Monte Carlo simulations). LR_T is the LR test for H_0'' vs H_1''. LR_T^* is the bootstrap analog of LR_T and is obtained with B=$T-1$ replication.

TABLE 2a. MC experiment. Strong identification. Rejection freq. of normality tests, restricted bootstrap.

DGP: ARMA(1,1), π_0= 0.4, β_0= -0.76 , H_0'' : $\pi = 0.4$vs H_1'': $\pi \neq 0.4$				
		T=100		
	B=19	B=29	B=49	B=T-1
$\hat{\theta}_T^*(=\hat{\beta}_T^*)$ under H_0''	0.051	0.068	0.073	0.106
$\hat{\pi}_T^*$ under H_1'':	0.083	0.093	0.103	0.112
$\hat{\beta}_T^*$ under H_1'':	0.050	0.052	0.051	0.046

NOTES: Results are based on M=1000 Monte Carlo replications. Rejection frequencies refer to Henze and Zirkler's (1990) normality test.Normality tests tests are computed using the 5% nominal level of significance.

TABLE 2b. MC experiment. Strong identification. Rejection freq. of normality tests, restricted bootstrap.

DGP: ARMA(1,1), π_0= 0.4, β_0= -0.76 , H_0'' : $\pi = 0.4$vs H_1'': $\pi \neq 0.4$				
		T=250		
	B=19	B=29	B=49	B=T-1
$\hat{\theta}_T^*(=\hat{\beta}_T^*)$ under H_0''	0.049	0.054	0.077	0.091
$\hat{\pi}_T^*$ under H_1'':	0.062	0.065	0.064	0.087
$\hat{\beta}_T^*$ under H_1'':	0.044	0.051	0.075	0.066

NOTES: Results are based on M=1000 Monte Carlo replications. Rejection frequencies refer to Henze and Zirkler's (1990) normality test.Normality tests tests are computed using the 5% nominal level of significance.

TABLE 3. MC experiment. Weak identification. ML estimates and rejection freq. of LR test.

DGP: ARMA(1,1), $\pi_0 = 0.4$, $\beta_0 = -0.05$, $H_0'' : \pi = 0.4$ vs $H_1'' : \pi \neq 0.4$						
			T=100			
	MONTE CARLO			BOOTSTRAP (B=T-1)		
	Under H_0''	Under H_1''		Under H_0''	Under H_1''	
	$\hat{\beta}_T (=\hat{\theta}_T)$	$\hat{\pi}_T$	$\hat{\beta}_T$	$\hat{\beta}^*_T (=\hat{\theta}^*_T)$	$\hat{\pi}^*_T$	$\hat{\beta}^*_T$
Average	-0.0584	0.1856	-0.0626	-0.0666	0.2142	-0.0154
s.e.	0.0949	0.4865	0.1149	0.0927	0.4813	0.1085
Hessian-based s.e.	0.0950	0.4362	0.0939	0.0964	2.3251	0.2864
	Rej(LR_T)=0.0453			Rej(LR^*_T)=0.041		
			T=250			
Average	-0.0537	0.2021	-0.0586	-0.0576	-0.0598	-0.0074
s.e.	0.0593	0.4700	0.0672	0.0591	0.4699	0.6750
Hessian-based s.e.	0.0596	0.4362	0.0624	0.0600	2.1502	0.1588
	Rej(LR_T)=0.0532			Rej(LR^*_T)=0.0480		

NOTES: Results are based on M=1000 Monte Carlo replications. 'Average' is the average of the ML estimator across Monte Carlo replications. 'MC s.e.' indicates the standard errors computed across Monte Carlo replications. 'Hessian-based s.e.' indicates the average across Monte Carlo replications of the standard errors computed from the Hessian matrix. 'Rej(·)' denotes rejection frequency (across Monte Carlo simulations). LR_T is the LR test for H_0'' vs H_1''. LR^*_T is the bootstrap analog of LR_T and is obtained with B=$T-1$ replication.

TABLE 4a. MC experiment. Weak identification. Rejection freq. of normality tests, restricted bootstrap.

DGP: ARMA(1,1), $\pi_0 = 0.4$, $\beta_0 = -0.05$, $H_0'' : \pi = 0.4$vs H_1'': $\pi \neq 0.4$				
		T=100		
	B=19	B=29	B=49	B=T-1
$\hat{\theta}_T^* (=\hat{\beta}_T^*)$ under H_0''	0.051	0.055	0.059	0.088
$\hat{\pi}_T^*$ under H_1'':	0.313	0.544	0.817	0.980
$\hat{\beta}_T^*$ under H_1'':	0.087	0.130	0.197	0.371

NOTES: Results are based on M=1000 Monte Carlo replications. Rejection frequencies refer to Henze and Zirkler's (1990) normality test.Normality tests tests are computed using the 5% nominal level of significance.

TABLE 4b. MC experiment. Weak identification. Rejection freq. of normality tests, restricted bootstrap.

DGP: ARMA(1,1), $\pi_0 = 0.4$, $\beta_0 = -0.05$, $H_0'' : \pi = 0.4$vs H_1'': $\pi \neq 0.4$				
		T=250		
	B=19	B=29	B=49	B=T-1
$\hat{\theta}_T^* (=\hat{\beta}_T^*)$ under H_0''	0.060	0.068	0.077	0.081
$\hat{\pi}_T^*$ under H_1'':	0.332	0.513	0.812	0.999
$\hat{\beta}_T^*$ under H_1'':	0.096	0.139	0.200	0.715

NOTES: Results are based on M=1000 Monte Carlo replications. Rejection frequencies refer to Henze and Zirkler's (1990) normality test.Normality tests tests are computed using the 5% nominal level of significance.

TABLE 5. An and Shorfheide's (2007) DSGE model on U.S. quarterly data

		SAMPLE 1984Q2-2008Q3 (T=98)		BOOTSTRAP		NORMALITY	
model	M1:	${}^1\hat{\theta}_T$	s.e. (Hessian)	${}^1\hat{\theta}^*_T$	s.e		
	$\hat{\kappa}_T$	0.123	0.018	0.124	0.021	$\hat{\kappa}^*_T$	p-value=0.297
	$\hat{\rho}_{r,T}$	0.835	0.017	0.758	0.030	$\hat{\rho}^*_{r,T}$	p-value=0.416
	$\hat{\sigma}^2_{g,T}$	0.5862	0.100	0.568	0.100	$\hat{\sigma}^{2*}_{g,T}$	p-value=0.422
		LR_T=176.11 asym. p-value=0.00		boot. p-value=0.053			
model	M2:	${}^2\hat{\theta}_T$	s.e. (Hessian)	${}^2\hat{\theta}^*_T$	s.e.		
	$\hat{\kappa}_T$	0.698	0.216	0.692	0.037	$\hat{\kappa}^*_T$	p-value=0.00
	$\hat{\psi}_{1,T}$	2.000	0.282	1.940	0.147	$\hat{\psi}^*_{1,T}$	p-value=0.00
	$\hat{\rho}_{r,T}$	0.776	0.017	0.609	0.077	$\hat{\rho}^*_{r,T}$	p-value=0.02
	$\hat{\sigma}^2_{r,T}$	0.029	0.004	0.065	0.029	$\hat{\sigma}^{*2}_{r,T}$	p-value=0.01
	$\hat{\sigma}^2_{g,T}$	0.311	0.048	0.315	0.054	$\hat{\sigma}^{*2}_{g,T}$	p-value=0.141
	$\hat{\sigma}^2_{z,T}$	0.016	0.003	0.022	0.009	$\hat{\sigma}^{*2}_{z,T}$	p-value=0.009
		LR_T=83.25 asym. p-value=0.00		boot. p-value=0.00			

NOTES: The log-likelihood maximization under both H_0 (the CER) and H_1 (without the CER) is obtained by the combining the Kalman filter with the 'CMA-ES' algorithm with the following bounds on the optimization paramater space: [0.04, 0.7] for κ, [0.5, 0.95] for $\rho_r$and [0.5, 2] for ψ_1. The column 'BOOTSTRAP' reports the average across B=999 bootstrap repetitions of the ML estimator of the structural parameters and the associated bootstrap standard errors. Asymptotic p-values associated with the LR tests for the CER are taken from the χ^2_r distribution with r=10 for model M1 and r=7 for model M2. The bootstrap p-values associated with the LR tests are computed using B=999 repetitions generated under the null of the CER. The normality test is the Henze and Zirkler's (1990) test and is computed using B=29 bootstrap replications of the ML estimators.