Bootstrap inference and diagnostics in state space models: With applications to dynamic macro models

Published:
DOI: http://doi.org/10.1002/jae.2843

Angelini, G., Cavaliere, G., & Fanelli, L. (2022). Bootstrap inference and diagnostics in state space models: With applications to dynamic macro models. Journal of Applied Econometrics, 37(1), 3-22.

https://doi.org/10.1002/jae.2843

BOOTSTRAP INFERENCE AND DIAGNOSTICS IN STATE SPACE MODELS: WITH APPLICATIONS TO DYNAMIC MACRO MODELS

By Giovanni Angelini[a], Giuseppe Cavaliere[a,b*]
and Luca Fanelli[a]

First draft: March 2020. First revision: November 2020; This version: April 2021.

Abstract

This paper investigates the potentials of the bootstrap as a tool for inference on the parameters of macroeconometric models which admit a state space representation. We consider a bootstrap estimator of the parameters of state space models and show that the bootstrap realizations of this estimator, usually employed to approximate asymptotic confidence intervals, p-values and critical values of tests, can be also constructively used to build a test for forms of misspecifications which invalidate asymptotic normality. The test evaluates how 'close or distant' the estimated state space model is from the case where asymptotic inference based on the Gaussian distribution applies. We derive sufficient conditions on the number of bootstrap repetitions, B, relative to the number of sample observations, T, for the test statistic to have a well-defined asymptotic distribution under the null. Throughout the paper we focus on the state space form of small-scale monetary dynamic stochastic general equilibrium (DSGE) models and investigate the usefulness of our approach through Monte Carlo experiments and empirical illustrations based on U.S. quarterly data. Results show that (i) bootstrapping the state space form provides highly reliable inference, and (ii) the suggested test detects weakly identified parameters reasonably well in finite samples.

Keywords: Bootstrap, State space models, Dynamic macroeconomic models, DSGE, Quasi-Maximum Likelihood, Normality test, Weak identification.

JEL Classification: C32, C51, E30, E50.

[a]Department of Economics, University of Bologna, Italy. [b] Department of Economics, University of Exter Business School, UK. *Corresponding author. Correspondence to: Giuseppe Cavaliere, email: giuseppe.cavaliere@unibo.it.

1 Introduction

State space models provide flexible representations of time series models, see Hannan and Deistler (1988), Caines (1988), Harvey (1989), Durbin and Koopman (2001) and Commandeur and Koopman (2007). Combined with Kalman filtering techniques, these models permit the estimation of the parameters of interest of, among others, first-order solutions of dynamic stochastic general equilibrium (DSGE) models, dynamic factor models, stochastic volatility models and affine term structure models. Understanding how the bootstrap performs in state space models is therefore crucial to envisage to what extent inference can be improved in a variety of models. Nevertheless, the literature on the bootstrap in these models is still scant, see e.g. Stoffer and Wall (1991) and Berkowitz and Kilian (2000).

This paper investigates the potential of the bootstrap in state space macroeconometric models along an important dimension, i.e. as a diagnostic tool to check whether conditions for asymptotic inference based on the Gaussian distribution hold. To do so, we adapt Stoffer and Wall's (1991) nonparametric bootstrap algorithm and consider the bootstrap Quasi-Maximum Likelihood (QML) estimator of the time-invariant parameters of state space models, and show that the distribution of the bootstrap QML estimator can be informative and useful also in situations in which the likelihood function is 'not well behaved'. The test is based on a number, say B, of realizations of the bootstrap QML estimator of the parameters and is essentially a standard normality test, hence straightforward to compute in practice. Contrary to standard bootstrap asymptotics where the number of bootstrap repetitions can be taken arbitrarily large, we derive sufficient conditions on B relative to the number of sample observations, T, for the test statistic to have a well-defined asymptotic distribution under the null that standard regularity conditions hold.

The suggested 'omnibus' test of model misspecification controls size in situations in which the QML estimator of the parameters is asymptotically Gaussian, and is expected to have power against situations where the QML estimator deviates asymptotically from the Gaussian distribution. Main (albeit not exhaustive) sources of violations of the asymptotic normality of the QML estimator include unidentified and/or weakly identified parameters, parameters which lie near the boundaries of the parameter space, infinite higher order moments in the distribution of innovation errors and nonstationary variables. Importantly, by design the test does not capture other types of misspecifications of the state space model which typically affect the consistency of the QML estimator but not its asymptotic normality, such as e.g. the omission of relevant variables/shocks and propagation mechanisms and the imposition of wrong parametric restrictions.

The suggested test is particularly useful for at least two reasons. First, the practitioner is not required to take a stand on the causes of failure of asymptotic normality. For instance, imagine that the suspect is that the parameters of the state space model are weakly identified. In this case, the practioner is not required to know a-priori which are the weakly and the strongly identified parameters. Second, despite there exists a large literature on identification-robust methods in structural dynamic macro models (e.g. Kleibergen and Mavroeidis, 2009), there is also a substantial lack of easily implementable tools to measure the strength of identification. We cover this gap and provide a general and computationally straightforward bootstrap-based method to assess the quality of the inference in an estimated state space model.

DSGE models are prominent examples of dynamic macro models whose equilibria can be represented in state space form. They are stylized descriptions of the economy and are widely used to evaluate macroeconomic policies or to predict the stance of the business cycle. It is well recognized that the sampling distribution of estimators of DSGE structural parameters tends to be non-normal and/or pile up on the boundary of the theoretically admissible parameter space as reflection of weakly identified parameters[1] and/or solution multiplicity; see An and Schorfheide (2007), Canova and Sala (2009) and Morris (2017). Identification-robust methods of inference for DSGE models have been developed in e.g. Guerron-Quintana *et al.* (2013), Dufour *et al.* (2013), Qu (2014), Andrews and Mikusheva (2015) and Guerron-Quintana *et al.* (2017). For these reasons, throughout the paper we investigate the performance of the bootstrap in state space models through the lens of small-scale monetary DSGE models, keeping in mind that the suggested test has power not only against weakly identified parameters but also against other possible sources of asymptotic non-normality.

We investigate the empirical properties of our bootstrap approach by a set of Monte Carlo experiments and an empirical illustration based on U.S. quarterly data. In both cases we consider the state space form associated with the small-scale monetary DSGE model analyzed in Guerron-Quintana *et al.* (2013). Simulation and empirical results point out that a proper combined use of univariate and multivariate normality tests provides highly reliable inference in correctly specified models. In these cases bootstrap standard errors track closely their non-bootstrap counterparts and the empirical coverage probability of bootstrap confidence intervals is close to nominal size. Simulation and

[1] This phenomenon mirrors the weak identification problem studied in the instrumental variable literature (Staiger and Stock, 1997) and in the generalized method of moments literature (Stock and Wright, 2000).

empirical evidence also suggests that the proposed test detects situations characterized by weakly identified parameters reasonably well in finite samples, no matter whether the structural shocks are non-Gaussian. This is an important result which suggests that bootstrapping the state space form of dynamic macro models is advantageous for practitioners as they can evaluate the reliability of Gaussian asymptotic inference through simple normality tests at small computational costs.

Two aspects of our approach are worth mentioning. First, the suggested test can be easily implemented in state space models for which estimation through Kalman filtering does not represent a major issue, which is typically the case in small-scale models. Second, our diagnostic test can be interpreted as a pretest in the sense that while 'standard' methods based on the Gaussian distribution can be applied when the null is not rejected, non-standard methods of inference are required otherwise.

1.1 Related literature

Our analysis exploits results in Stoffer and Wall (1991) on bootstrap consistency in state space models. However, to prove that the bootstrap QML estimator replicates the asymptotic distribution of the QML estimator, Stoffer and Wall (1991) rely on the regularity conditions reported in Ljung and Caines (1979) which, unfortunately, can hardly be framed and checked in the class of models applied in empirical macroeconomics and finance. We revisit and reinterpret bootstrap consistency by relying on assumptions which are specific to DSGE models and therefore are more easily understandable and interpretable in the context of dynamic macro models.

Ours is not the first application of the bootstrap to DSGE models. Guerron-Quintana *et al.* (2017) develop a new theory for impulse response matching estimation of DSGE models based on the bootstrap, see also Fève *et al.* (2009). Cho and Moreno (2006) and Bårdsen and Fanelli (2015*a*) apply bootstrap methods in small-scale New-Keynesian DSGE models that have a finite-order vector autoregressive (VAR) representation. Le *et al.* (2011) combine the use of bootstrap methods with indirect inference techniques for DSGE models; see also Khalaf *et al.* (2019). All these contributions, however, are based on VAR approximations of the DSGE equilibrium. The novel feature of our approach, when focusing on DSGE models, is that the bootstrap involves directly the innovation form representation of the model. This exempts practitioners from putting the DSGE equilibrium in VAR form and from choosing which moments or features of the data to match with the theoretical model as, e.g., in Hall *et al.* (2012) and Guerron-Quintana *et al.* (2017).

A test for the null hypothesis of 'strong identification' against weak identification in nonlinear dynamic macro models has been also developed by Inoue and Rossi (2011). Their test, however, does not apply to models featuring unobservable (latent) components and does not involve the bootstrap.

There are earlier (but few) contributions in the literature where the use of the bootstrap as diagnostic tool has been advocated. In the statistical literature, Beran (1997) suggests diagnostic plots for detecting bootstrap failure in regression models, considering however a setup which can not be easily reconciled with the features of dynamic macro models. In the econometric literature, Zhan (2018) has shown in the context of instrumental variable regressions that a substantial difference between the distribution of the standardized Two-Stage-Least-Squares estimator and the Gaussian distribution indicates the existence of weak instruments.[2] Zhan's (2018) approach is peculiar to instrumental variable regressions and requires a preliminarily conventional definition of weak and strong instruments along the lines of Staiger and Stock (1997). Our approach is more general: it covers the broad class of econometric models which can be represented in state space form involving time-invariant parameters, and reads as an omnibus test for checking whether conditions for Gaussian asymptotic inference are supported. Importantly, we do not need any preliminary (and arbitrary) definition of model misspecification, in the sense that it is the bootstrap distribution to inform the practitioner on the extent of deviations from the case in which standard regularity conditions are at work.[3]

Finally, in the literature on Structural VARs identified with external instruments (proxy-SVARs or SVARs-IV) recently popularized by Mertens and

[2] The use of plots of the bootstrap distribution of estimated parameters of interest as 'graphical diagnostic tool' may be also found in Bårdsen and Fanelli (2015*b*), Figures 1-2. These authors contrast the bootstrap distribution of the estimators of the structural parameters of a DSGE models with the Gaussian distribution, and ascribe the discrepancy they observe for some parameters to identification issues. Also Stoffer and Wall (1991) plot the bootstrap distribution of the estimated parameters of a 'nearly redundant' ARMA(2,2) model represented in state space form, see their Figures 1 and 2, and observe that the bootstrap provides (p.1028) 'vital information concerning the problems with model specification due to near parameter redundancy when sample sizes are small'.

[3] There exist only few studies in the bootstrap literature where the applicability of the bootstrap is discussed in situations where not all regularity conditions for inference are assumed to hold. While the results in Moreira *et al.* (2004) suggest that the bootstrap might be valid in some weak identification cases, more recently Dovonon and Gonçalves (2017) address the bootstrap estimation of the standard test of overidentification restrictions in the generalized method of moments framework when the model is globally identified but the rank condition is not valid, a situation referred to as lack of first-order local identification. Instead Cavaliere *et al.* (2017) analyze bootstrap consistency in testing problems where a parameter is on the boundary of the parameter space.

Ravn (2013) and Stock and Watson (2018), Angelini *et al.* (2021) formalize a boostrap-based test of instrument relevance.

1.2 Structure of the paper

This paper is organized as follows. Section 2 introduces the state space representations and summarizes the assumptions under which the QML estimator of the parameters is asymptotically Gaussian. Section 3 discusses bootstrap inference and Section 4 presents our bootstrap-based test of model misspecification. Section 5 explores the finite sample performance of our test by a set of Monte Carlo simulations based on Guerron-Quintana *et al.* (2013)'s small-scale monetary DSGE model. Section 6 applies Guerron-Quintana *et al.* (2013)'s DSGE model to U.S. quarterly data and investigates the reliability of Gaussian asymptotic inference. Section 7 contains some concluding remarks. Notation, technical proofs and additional Monte Carlo and empirical results are provided in an online Supplementary Material, SM henceforth.

2 Representations, assumptions and asymptotic inference

In this section we focus on models admitting a state space representation that can be cast in the form:

$$\underset{n_z\times 1}{Z_t} = \underset{n_z\times n_z}{A(\theta)}\ \underset{n_z\times 1}{Z_{t-1}} + \underset{n_z\times n_\omega}{B(\theta)}\ \underset{n_\omega\times 1}{\omega_t} \tag{1}$$

$$\underset{n_y\times 1}{y_t} = \underset{n_y\times n_z}{C(\theta)}\ \underset{n_z\times 1}{Z_{t-1}} + \underset{n_y\times n_\omega}{D(\theta)}\ \underset{n_\omega\times 1}{\omega_t}\ . \tag{2}$$

known as the 'ABCD representation' (Fernández-Villaverde *et al.* 2007). Here Z_t is a $n_z \times 1$ vector of endogenous state variables, $y_t := (y_{1,t}, y_{2,t}, \cdots, y_{n_y,t})'$ is a $n_y \times 1$ vector of (demeaned) observed variables and ω_t is a $n_\omega \times 1$ vector of shocks with covariance matrix $\Sigma_\omega := \Sigma_\omega(\theta)$, where $\Sigma_\omega(\theta)$ can be diagonal or 'full'. The matrices $A(\theta)$, $B(\theta)$, $C(\theta)$, $D(\theta)$ and $\Sigma_\omega(\theta)$ depend nonlinearly on the $n_\theta \times 1$ vector of (time-invariant) parameters $\theta \in \Theta$ (Θ being a compact subset of $\mathbb{R}^{n_\theta}$). The true value of θ is denoted by θ_0. (Bootstrap) inference on θ_0 is the object of interest of this paper.

The state space representation in (1)-(2) is general enough to cover DSGE models, VARMA models and dynamic factor models. Given a sample of T observations $\{y_1, ..., y_T\}$, an equivalent representation of (1)-(2), useful for estimation purposes, is the *innovation form* which, for $t = 1, ..., T-1$, can be written as

$$\hat{Z}_{t+1|t+1} = A(\theta)\hat{Z}_{t|t} + K_t(\theta)\ \epsilon_{t+1} \tag{3}$$

$$y_{t+1} = C(\theta)\hat{Z}_{t|t} + \epsilon_{t+1} \tag{4}$$

where $K_t(\theta)$ is the Kalman gain, $\epsilon_t = y_t - C(\theta)\hat{Z}_{t-1|t-1}$ are the innovation errors with covariance matrix $\Sigma_{\epsilon,t+1}(\theta)$, and $\hat{Z}_{t|t}$ defined as $\hat{Z}_{t|t} := E(Z_t \mid \mathcal{F}^y_{t,1})$ for $\mathcal{F}^y_{t,1} := \sigma(y_t, ..., y_1) \subseteq \mathcal{F}^y_{t,-\infty}$. The initial condition $\hat{Z}_{1|1}$ is fixed in the statistical analysis. The mapping that links the parameterization ($A(\theta)$, $B(\theta), C(\theta)$, $D(\theta)$, $\Sigma_\omega(\theta)$) in (1)-(2) and the parameterization ($A(\theta)$, $K_t(\theta)$, $C(\theta)$, $\Sigma_{\epsilon,t+1}(\theta)$) in (3)-(4) is explicitly derived in Hansen and Sargent (2005). The covariance matrix $\Sigma_{\epsilon,t+1}(\theta)$ in (3)-(4) obeys

$$\Sigma_{\epsilon,t+1}(\theta) = C(\theta)P_{t|t}(\theta)\,C(\theta)' + D(\theta)\Sigma_\omega(\theta)D(\theta)'\ ,\ t = 1, ..., T-1 \tag{5}$$

with $P_{t|t}(\theta) := E((Z_t - \hat{Z}_{t|t})(Z_t - \hat{Z}_{t|t})' \mid \mathcal{F}^y_t)$ and $P_{0|0}$ being given. In general, the matrices $\Sigma_{\epsilon,t}(\theta)$, $K_t(\theta)$ and $P_{t|t}(\theta)$ in the innovation form are updated recursively through the standard Gaussian Kalman recursions, and due to regularity conditions stated below, as t grows, these matrices converge (exponentially fast) to time-invariant counterparts $\Sigma_\epsilon(\theta)$ and $K(\theta)$ and $P(\theta)$. The corresponding *time-invariant* (steady state) *innovation form* (Anderson and Moore 1979; Hansen and Sargent, 2005) is

$$\hat{Z}_{t+1|t+1} = A(\theta)\hat{Z}_{t|t} + K(\theta)\epsilon_{t+1} \tag{6}$$

$$y_{t+1} = C(\theta)\hat{Z}_{t|t} + \epsilon_{t+1}. \tag{7}$$

In (6)-(7), $K(\theta)$ is the steady state Kalman gain and the innovation errors $\epsilon_t = y_t - C(\theta)\hat{Z}_{t-1|t-1}$ can be interpreted by considering the quantity $\hat{Z}_{t|t} := E(Z_t \mid \mathcal{F}^y_{t,-\infty})$ as the optimal predictor of Z_t, based on the filtration $\mathcal{F}^y_{t,-\infty} := \sigma(y_t, ..., y_1, ...)$. The (steady state) innovation variance is therefore $\Sigma_\epsilon(\theta) := E(\epsilon_t\epsilon_t')$. Hereafter we call the representation (6)-(7) 'AKC form'.

We now consider, in Assumptions A1–A5 below, a set of regularity conditions on the state space model which permit 'standard' asymptotic and bootstrap inference on the parameters θ.

Assumption A1. For every $\theta \in \Theta$:
(i) For all t, s, $E(\omega_t) = 0$, $E(\omega_t\omega_s') = \Sigma_\omega(\theta)\,\mathbb{I}(t = s)$.
(ii) For every $z \in \mathbb{C}$, $\det(I_{n_z} - A(\theta)\,z) = 0$ implies $|z| > 1$.

Assumption A1*(i)* requires the shocks ω_t to be white noise with unconditional covariance matrix $\Sigma_\omega(\theta)$. Assumption A1*(ii)* implies that the matrix $A(\theta)$ in (1)-(2) is stable (i.e. with eigenvalues inside the unit disk) and, combined with Assumption A1*(i)*, that the stochastic process that generates $\{y_t\}$ is covariance stationary and ergodic. This condition subsumes that all the necessary variable transformations have been performed such that the variables of the state space model are stationary. The stability condition in Assumption A1(ii) guarantees that the AKC form (6)-(7) can be written as the innovation form (3)-(4) with the Kalman gain $K_t(\theta)$ and the innovation covariance matrix $\Sigma_{\epsilon,t+1}(\theta)$ replaced with their steady state counterparts $K(\theta)$ and $\Sigma_\epsilon(\theta)$, respectively. As in Komunjer and Ng (2011, p.2007), we also consider the following assumptions that guarantee the (population) local identifiability of θ.

Assumption A2. With $\Lambda(\theta) := (vec(A(\theta))', vec(K(\theta))', vec(C(\theta))', vech(\Sigma_\epsilon(\theta))')'$, it holds that, for every $\theta \in \Theta$:
(i) $\Lambda(\theta)$ is continuously differentiable on Θ;
(ii) $D(\theta)\Sigma_\epsilon(\theta)D(\theta)'$ is nonsingular;
(iii) The matrices

$$\begin{gathered}\left(K(\theta), A(\theta)K(\theta), ..., A^{n_z-1}(\theta)K(\theta)\right)\\ \left(C(\theta)', A(\theta)'C(\theta)', ..., A^{n_z-1}(\theta)'C(\theta)'\right)\end{gathered}$$

have full row rank.

Assumption A3. Define the matrix

$$\Delta(\theta) := \begin{pmatrix} \frac{\partial vec(A(\theta))}{\partial\theta'} & A(\theta)'\otimes I_{n_z} - I_{n_z}\otimes A(\theta) \\ \frac{\partial vec(C(\theta))}{\partial\theta'} & K(\theta)'\otimes I_{n_z} \\ \frac{\partial vec(K(\theta))}{\partial\theta'} & -I_{n_z}\otimes C(\theta) \\ \frac{\partial vec(\Sigma_\epsilon(\theta))}{\partial\theta'} & 0_{(n_y(n_y+1)/2)\times n_z^2} \end{pmatrix}.$$

It holds that $\Delta(\theta_0)$ has full column rank $n_\theta + n_z^2$ and is regular in the neighborhood $\mathcal{N}_\delta(\theta_0) := \{\theta \in \Theta : ||\theta - \theta_0|| \leq \delta\}$ for some $\delta > 0$.

Assumption A2*(i)* is a standard differentiability condition. Assumption A2*(ii)*, along with Assumption A1*(ii)*, ensures that the covariance matrix associated with the innovation errors in system (6)-(7) exists. Assumption A2*(iii)* ensures that the system in (6)-(7) is 'minimal', in the sense that Z_t does not contain more states than strictly necessary to fully characterize the dynamics of the system. Minimality mimics the left-coprime condition typically imposed on (or assumed in) VARMA processes (see e.g. Lütkepohl, 2005, p. 452). Importantly, Assumption A2 implies that y_t admits a Wold representation in terms of ϵ_t, specifically

$$y_t = H_\epsilon(L, \theta_0)\epsilon_t \tag{8}$$

where

$$H_{\epsilon}(z,\theta) := I_{n_y} + C(\theta)(I_{n_z} - A(\theta) z)^{-1} K(\theta) \tag{9}$$

is square and invertible for $|z| > 1$, a condition known as *left-invertibility* (Komunjer and Ng, 2011). In this case, the innovations ϵ_t in (8) are fundamental (meaning that ϵ_t is spanned by $\mathcal{F}^y_{t,-\infty}$) and have nonsingular covariance matrix $\Sigma_\epsilon(\theta)$ for every θ.

Assumption A3 is a necessary and sufficient rank condition for identification which ensures that θ_0 is locally identified from the complete set of autocovariances $\Gamma_k := Cov(y_t, y_{t-k})$, $k = 0, \pm 1, ...$, of $\{y_t\}$; see Definition 1 in Komunjer and Ng (2011) or Definition 1 in Qu and Tkachenko (2012) for an equivalent formulation in terms of the spectral density of $\{y_t\}$. Komunjer and Ng (2011) show that the system information matrix is nonsingular iff the rank condition in Assumption A3 holds. Another implication of Assumptions A2-A3 is that, for all $\theta \neq \theta_0$ in the neighborhood $\mathcal{N}_\delta(\theta_0) := \{\theta \in \Theta : ||\theta - \theta_0|| \leq \delta\}$ for some $\delta > 0$, it holds that $H_\epsilon(z,\theta) \neq H_\epsilon(z,\theta_0)$ on a subset of $\{z \in \mathbb{C} : |z| = 1\}$ of positive Lebesgue measure. This condition is crucial to establishing the asymptotic properties of standard and bootstrap estimators considered in the paper, denoted with $\hat{\theta}_T$ and $\hat{\theta}^*_T$, respectively.

It is worth stressing that the necessary and sufficient conditions in Assumption A3 refer to the (population) local identifiability of θ, not to its global identifiability.[4]

In order to derive the asymptotic properties of $\hat{\theta}_T$, we also introduce some conditions on the innovation errors and on the smoothness of the function $\Lambda(\theta) := (vec(A(\theta))', vec(K(\theta))', vec(C(\theta))', vech(\Sigma_\epsilon(\theta))')'$. Assumption A4 below involves the higher-order moments of the innovation errors $\epsilon_t := \epsilon_t(\theta_0)$ evaluated at the true parameter value θ_0 in the AKC form representation where, by construction, $E(\epsilon_t \mid \mathcal{F}^y_{t-1,-\infty}) = 0_{n_y \times 1}$ (a.s.), hence $\{\epsilon_t, \mathcal{F}^y_{t,1}\}$ is a martingale difference sequence.

Assumption A4. The innovation errors ϵ_t associated with the AKC form in (6)-(7) satisfy:
(i) $E(\epsilon_t \epsilon_t' \mid \mathcal{F}^y_{t-1,1}) = \Sigma_\epsilon(\theta_0)$ (a.s.);
(ii) $E||\epsilon_t||^\nu < \infty$ for some $\nu \geq 4$.

Assumption A5 focuses on $\Lambda(\theta)$.

Assumption A5. The function $\Lambda(\theta)$ is thrice differentiable in the neighborhood $\mathcal{N}_\delta(\theta_0) := \{\theta \in \Theta : ||\theta - \theta_0|| \leq \delta\}$ for some $\delta > 0$.

[4] We refer to Qu and Tkachenko (2017) and Kocięcki and Kolasa (2018) for global identification in a class of models which can be represented as in (1)-(2).

Assumption A4*(i)* rules out conditional heteroskedasticity,[5] while Assumption A4*(ii)* ensures that the innovation disturbances ϵ_t have finite fourth-order moments. Assumption A5 is a technical condition which extends the differentiability of the function $\Lambda(\theta)$ up to the third-order in the neighborhood $\mathcal{N}_\delta(\theta_0)$.

Throughout the paper we consider the case $n_\omega \geq n_y$, which implies that there are at least as many shocks as observable variables. This is known as 'nonsingular case', see Komunjer and Ng (2011); models for which $n_\omega < n_y$ can be covered by adding artificially $n_v := n_y - n_\omega$ measurement errors v_t and rewriting system (1)-(2) by replacing ω_t with the n_u-dimensional vector $u_t := (\omega_t', v_t')$ $(n_u{:=}n_y + n_v)$ so that nonsingularity is automatically restored.

We now briefly discuss the estimation of the structural parameters θ. This is a necessary step in order to prove the bootstrap consistency in state space models for which standard asymptotic inference applies. Let $L_T(\theta)$ be the log-likelihood function computed from system (3)-(4), as given by

$$L_T(\theta) := \sum_{t=1}^{T} \ell_t(\theta) \tag{10}$$

where, under the auxiliary assumption of Gaussian innovation errors,

$$\ell_t(\theta) := l(y_t \mid \mathcal{F}_{t-1,1}^y; \theta) = -\{\log\det(\Sigma_{\epsilon,t}(\theta)) + \epsilon_t(\theta)' \Sigma_{\epsilon,t}(\theta)^{-1} \epsilon_t(\theta)\} \tag{11}$$

with $\epsilon_t(\theta)$ denoting the ϵ_t term already defined in (3)-(4) (we now stress explicitly its dependence on θ). The QML estimator of θ solves the problem

$$\hat{\theta}_T := \arg\max_{\theta \in \mathcal{T}} L_T(\theta) \tag{12}$$

[5] The conditional homoskedasticity hypothesis in Assumption A4(i) implies that also the shocks ω_t in the original formulation (1)-(2) of the state space model are conditionally homoskedastic. To see this, observe that from the innovation form (3)-(4) and from standard Gaussian Kalman recursions it is possible to derive the expression

$$E(\epsilon_t\epsilon_t' \mid \mathcal{F}_{t-1,1}^y) = C(\theta)P_{t-1|t-1}(\theta)\, C(\theta)' + D(\theta)E(\omega_t\omega_t' \mid \mathcal{F}_{t-1,1}^y)D(\theta)'$$

which, as t grows and $P_{t-1|t-1}(\theta)$ collapses to $P(\theta)$ in the *time-invariant* (steady state) *innovation form*, qualifies to

$$E(\epsilon_t\epsilon_t' \mid \mathcal{F}_{t-1,1}^y) = C(\theta)P(\theta)\, C(\theta)' + D(\theta)E(\omega_t\omega_t' \mid \mathcal{F}_{t-1,1}^y)D(\theta)'.$$

The last expression shows that Assumption A4(i) could be alternatively derived by assuming conditional homoskedasticity for the shocks ω_t in (1)-(2). It is worth remarking that specification tests can be naturally implemented after estimation of the innovation form (this would essentially require to test for conditional homoskedasticity and for higher order moments). For these reasons our preference is to place assumptions on ϵ_t, rather than on ω_t.

where $\mathcal{T} \subseteq \Theta$ is the user-chosen optimizing (compact) set and $L_T(\theta)$ is maximized recursively through Gaussian Kalman filtering.

The proposition that follows establishes the convergence of $\hat{\theta}_T$ and $W_T := T^{1/2}(\hat{\theta}_T - \theta_0)$. We implicitly maintain the assumption that the user-chosen maximization set $\mathcal{T}$ belongs to the neighborhood $\mathcal{N}_\delta(\theta_0)$ of θ_0. See Lemma 1 in Qu and Tkachenko (2012) for an equivalent set of conditions. Proofs are in the accompanying supplement, SM.

Proposition 1 *Consider the ABCD form in (1)-(2), the QML estimator of θ defined in (12) and Assumptions A1-A4. Then, as $T \to \infty$:*
(i) $\hat{\theta}_T \xrightarrow{p} \theta_0$;
(ii) Provided $\theta_0 \in int(\mathcal{T})$, the interior of $\mathcal{T}$, and Assumption A5 holds, then

$$W_T := T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0_{n_\theta \times 1}, \Omega_0\right),$$

where $\Omega_0 := \left(\mathcal{A}_0\mathcal{B}_0^{-1}\mathcal{A}_0'\right)^{-1}$ *with* $\mathcal{B}_0 := \lim_{T\to\infty} T^{-1}\mathcal{B}_{0,T}$, $\mathcal{B}_{0,T} := E(\nabla_\theta L_T(\theta_0) \times \nabla_\theta L_T(\theta_0)')$, *and* $\mathcal{A}_0 := \lim_{T\to\infty} T^{-1}\mathcal{A}_{0,T}$, $\mathcal{A}_{0,T} := E(-\nabla^2_{\theta\theta} L_T(\theta_0))$.[6]

Some remarks are in order.

Remark 2.1 Proposition 1 ensures that when the state space model is correctly specified (up to the probability model), $\hat{\theta}_T$ is consistent for θ_0 and $W_T := T^{1/2}(\hat{\theta}_T - \theta_0)$ is asymptotically Gaussian. When the innovation errors are actually Gaussian, the information matrix equivalence $\mathcal{A}_0 + \mathcal{B}_0 = 0$ holds and the asymptotic covariance matrix of W_T collapses to $\Omega_0 := \mathcal{A}_0^{-1}$. Consistent analytic standard errors for the estimated parameters are taken from the main diagonal of the matrix $\hat{\Omega}_T := \hat{\mathcal{A}}_T \hat{\mathcal{B}}_T^{-1} \hat{\mathcal{A}}_T$, where $\hat{\mathcal{B}}_T := T^{-1}\sum_{t=1}^T \nabla_\theta \ell_t(\hat{\theta}_T) \times \nabla_\theta \ell_t(\hat{\theta}_T)'$ and $\hat{\mathcal{A}}_T := -T^{-1}\sum_{t=1}^T \nabla^2_{\theta\theta}\ell_t(\hat{\theta}_T)$.

Remark 2.2 Proposition 1 is based on a maintained assumption of 'correct specification' of the state space model. Actually, the convergence facts in Proposition 1 can be extended to the case in which the true parameter value θ_0 is replaced with a vector $\theta_\dagger$ interpreted, along the lines of White (1982), as the (non-random) 'pseudo-true' parameter value. Indeed, there are forms of misspecification of the state space model which affect the consistency of $\hat{\theta}_T$ but not the asymptotic normality of $T^{1/2}(\hat{\theta}_T - \theta_\dagger)$ (e.g. the omission of important propagation mechanisms or relevant variables, the imposition of

[6] Under Assumptions A1-A5, the incremental observed information matrix $\hat{\mathcal{B}}_T := T^{-1}\sum_{t=1}^T \nabla_\theta \ell_t(\hat{\theta}_T) \times \nabla_\theta \ell_t(\hat{\theta}_T)'$ (evaluated at $\hat{\theta}_T$), and the observed information matrix $\hat{\mathcal{A}}_T := -T^{-1}\sum_{t=1}^T \nabla^2_{\theta\theta}\ell_t(\hat{\theta}_T)$ (evaluated at $\hat{\theta}_T$), estimate $\mathcal{B}_0$ and $\mathcal{A}_0$, respectively. The asymptotic variance matrix of $\hat{\theta}_T$ can be estimated by $\hat{\Omega}_T := \hat{\mathcal{A}}_T \hat{\mathcal{B}}_T^{-1} \hat{\mathcal{A}}_T$.

invalid parametric restrictions, etc.). The diagnostic test we develop in the next sections is designed to have power against forms of misspecification of the state space model which depend on violations of the conditions in Assumptions A1-A5, therefore it has no power against other types of model misspecification.

An appealing feature of Proposition 1 is that the asymptotic normality of the QML estimator of θ is derived by circumventing some of the involved regularity conditions considered by e.g. Ljung and Caines (1979), Caines (1988) or Harvey (1989, pp.128-130), which can be hard to check in the class of dynamic models used in macroeconomics and finance.

3 Bootstrap inference

In this section we discuss bootstrap inference. We introduce our main bootstrap algorithm which defines the bootstrap parameter estimator, $\hat{\theta}_T^*$, and then discuss first-order validity of the bootstrap and consistency of the related bootstrap standard errors. The use of the bootstrap as a diagnostic tool is considered separately in Section 4.

Consider the innovation form representation in (3)-(4). As in the previous section, $\hat{\theta}_T$ denotes the QML estimator from (12), obtained on the original sample $\{y_1, y_2, ..., y_T\}$. The bootstrap analog of $\hat{\theta}_T$, $\hat{\theta}_T^*$, is defined through the following nonparametric algorithm, adapted from Stoffer and Wall (1991, 2004); see also Berkowitz and Kilian (2000). Henceforth, with '*' we denote bootstrap analogs of estimators and test statistics.[7]

Algorithm 1 (nonparametric bootstrap)

1. Given the innovation residuals $\hat{\epsilon}_t := y_t - C(\hat{\theta}_T)\hat{Z}_{t-1|t-1}$ and the estimated covariance matrices $\hat{\Sigma}_{\epsilon,t} = \Sigma_{\epsilon,t}(\hat{\theta}_T)$, construct the standardized innovations as
$$\hat{e}_t := \hat{\Sigma}_{\epsilon,t}^{-1/2}\hat{\epsilon}_t^c, \;\; t = 2, ..., T, \tag{13}$$
where $\hat{\epsilon}_t^c$, $t = 2, ..., T$, are the centered residuals $\hat{\epsilon}_t^c := \hat{\epsilon}_t - (T-1)^{-1}\sum_{t=2}^T \hat{\epsilon}_t$;[8]

2. Sample, with replacement, $T-1$ times from $\{\hat{e}_2, ..., \hat{e}_T\}$ to obtain the bootstrap standardized innovations $\{e_2^*, ..., e_T^*\}$;

3. Mimicking the innovation form representation in (3)-(4), the bootstrap

[7] Matlab codes for the computation of the bootstrap estimator are available upon request.
[8] We strictly follow Stoffer and Wall (1991) with this standardization.

sample $\{y_1^*, y_2^*, ..., y_T^*\}$ is generated recursively, for $t = 1, ..., T-1$, as

$$\begin{pmatrix} \hat{Z}^*_{t+1|t+1} \\ y^*_{t+1} \end{pmatrix} = \begin{pmatrix} A(\hat{\theta}_T) & 0_{n_z \times n_y} \\ C(\hat{\theta}_T) & 0_{n_y \times n_y} \end{pmatrix} \begin{pmatrix} \hat{Z}^*_{t|t} \\ y^*_t \end{pmatrix} + \begin{pmatrix} K_t(\hat{\theta}_T) \\ I_{n_y} \end{pmatrix} \hat{\Sigma}^{1/2}_{\epsilon,t+1} e^*_{t+1} \tag{14}$$

with initial condition $\hat{Z}^*_{1|1} = \hat{Z}_{1|1}$ and $y_1^* = y_1$;

4. Using the bootstrap sample $\{y_1^*, y_2^*, ..., y_T^*\}$, the bootstrap estimator $\hat{\theta}^*_T$ of the parameters of the DSGE model is given by

$$\hat{\theta}^*_T := \arg\max_{\theta \in \mathcal{T}} L^*_T(\theta) \tag{15}$$

where $L^*_T(\theta)$ is the bootstrap analog of $L_T(\theta)$, defined by $L^*_T(\theta) := \sum_{t=1}^T \ell^*_t(\theta)$ where, for $t = 1, ..., T$,

$$\ell^*_t(\theta) := -\{\log\det(\Sigma_{\epsilon,t}(\theta)) + \epsilon^{*\prime}_t \Sigma_{\epsilon,t}(\theta)^{-1} \epsilon^*_t\}, \tag{16}$$

and $\epsilon^*_t := \epsilon^*_t(\theta) := y^*_t - C(\theta)\hat{Z}^*_{t-1|t-1}$.

With $\hat{\theta}^*_T$ as defined in Algorithm 1, the distribution of $W^*_T := T^{1/2}(\hat{\theta}^*_T - \hat{\theta}_T)$ conditional on the data, say $G^*_T(\cdot)$, is used to approximate the (unknown) distribution of $W_T := T^{1/2}(\hat{\theta}_T - \theta_0)$, say $G_T(\cdot)$.

Although the conditional cumulative distribution function (CDF) $G^*_T(\cdot)$ is unknown, in practice, as is standard, it can be approximated by repeating Steps 2–4 an arbitrarily large number of times, say N, such that a set of independent and identically distributed bootstrap realizations of $\hat{\theta}_T$, say $\{\hat{\theta}^*_{T:1}, \hat{\theta}^*_{T:2}, ..., \hat{\theta}^*_{T:N}\}$, is obtained. Then, $G^*_T(\cdot)$ is approximated by the empirical distribution function

$$G^*_{T:N}(x) := \frac{1}{N}\sum_{b=1}^N \mathbb{I}\left(W^*_{T;b} \leq x\right) \tag{17}$$

where $\mathbb{I}(\cdot)$ is the indicator function and $W^*_{T:b} := T^{1/2}(\hat{\theta}^*_{T:b} - \hat{\theta}_T)$. By the Glivenko-Cantelli theorem, $\sup_{x \in R} |G^*_{T:N}(x) - G^*_T(x)| \to 0$ a.s. as $N \to \infty$. The bootstrap misspecification test we discuss in Section 4 will be based on $B < N$ bootstrap realizations $\{\hat{\theta}^*_{T:1}, \hat{\theta}^*_{T:2}, ..., \hat{\theta}^*_{T:B}\}$ (or $\{W^*_{T:1}, ..., W^*_{T:B}\}$).

Remark 3.1 The algorithm is a nonparametric, or i.i.d., bootstrap scheme, in the sense that in step 1 the bootstrap innovations are obtained as random draws from the standardized residuals $\hat{e}_t$, $t = 2, ...T$. However, if the normality hypothesis holds true, one may alternatively employ a parametric version of the bootstrap algorithm, which simply requires ignoring steps 1 and 2 and starting from the step 3, with the e_t^*'s now taken as independent random draws from the $N(0_{n_y \times 1}, I_{n_y})$ distribution.

Remark 3.2 An alternative algorithm to our i.i.d. bootstrap is the wild bootstrap, which allows to mimic possible (conditional and unconditional) heteroskedasticity patterns present in the original data (e.g. when the conditional variance $E(\epsilon_t \epsilon_t' \mid \mathcal{F}_{t-1,1}^y)$ in Assumption A4(i) might change over time); we conjecture that results in Gonçalves and Kilian (2004) for the case of stationary univariate autoregressions carry over the state space framework. The wild bootstrap shocks would be generated in Step 3 as

$$e_{w,t}^* := \hat{\epsilon}_t^c w_t^*,\ t = 2, ..., T$$

where w_t^* is an i.i.d. scalar sequence with 0 mean, unit variance, and finite fourth order moments. Using the wild bootstrap, the standardization of the residuals in (13) (Step 1) is no longer necessary (since, conditionally on the original data, $E^*(e_{w,t}^* e_{w,t}^{*\prime}) = \hat{\epsilon}_t^c \hat{\epsilon}_t^{c\prime}$) and, consequently, the recursion in (14) can be replaced by the simpler recursion

$$\begin{pmatrix} \hat{Z}_{t+1|t+1}^* \\ y_{t+1}^* \end{pmatrix} = \begin{pmatrix} A(\hat{\theta}_T) & 0_{n_z \times n_y} \\ C(\hat{\theta}_T) & 0_{n_y \times n_y} \end{pmatrix} \begin{pmatrix} \hat{Z}_{t|t}^* \\ y_t^* \end{pmatrix} + \begin{pmatrix} K_t(\hat{\theta}_T) \\ I_{n_y} \end{pmatrix} e_{w,t+1}^*$$

for $t = 1, ..., T-1$, again initialized at $\hat{Z}_{1|1}^* = \hat{Z}_{1|1}$, $y_1^* = y_1$.

The bootstrap implemented as in Algorithm 1 above is first-order valid. Specifically, we have that under regularity conditions the distribution of $W_T^* := T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$, conditionally on the original data, converges in probability to the asymptotic distribution of $W_T := T^{1/2}(\hat{\theta}_T - \theta_0)$; hence, the bootstrap replicates the asymptotic distribution of the original estimator. Similarly, the bootstrap standard errors converge to the QML standard errors. The technical result is provided in the following Proposition, whose proof may be found in SM. However, in order to establish bootstrap consistency we also need the innovations ϵ_t to possess finite eigtht-order moments. We strengthen Assumption A4 accordingly.

Assumption A4'. Assumption A4 holds for some $\nu \geq 8$.

Proposition 2 *Consider the state space model in (1)-(2), with fixed initial conditions $\hat{Z}_{1|1}$. With $\hat{\theta}_T$ as defined in (12) and its bootstrap analog $\hat{\theta}_T^*$ as defined in Algorithm 1, under Assumptions A1–A3, A4', A5, as $T \to \infty$:*

$$\hat{\theta}_T^* - \hat{\theta}_T \xrightarrow{p^*}_p 0_{n_\theta \times 1} \tag{18}$$

$$R_T^* := \hat{\Omega}_T^{-1/2} W_T^* \xrightarrow{d^*}_p N\left(0_{n_\theta \times 1}, I_{n_\theta}\right). \quad (19)$$

Moreover,

$$\hat{\Omega}_T^* - \hat{\Omega}_T \xrightarrow{p^*}_p 0_{n_\theta \times n_\theta} \quad (20)$$

where $\hat{\Omega}_T^* := \hat{\mathcal{A}}_T^* \hat{\mathcal{B}}_T^{*\,-1} \hat{\mathcal{A}}_T^*$, $\hat{\mathcal{B}}_T^* := T^{-1}\sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T) \times \nabla_\theta \ell_t^*(\hat{\theta}_T)'$ *and* $\hat{\mathcal{A}}_T^* := -T^{-1}\sum_{t=1}^T \nabla_{\theta\theta}^2 \ell_t^*(\hat{\theta}_T)$.

Proposition 2 is novel and generalizes Stoffer and Wall's (1991) main result on state space models by formalizing the consistency of the bootstrap under different conditions relative to those in Ljung and Caines (1979).[9]

A remark on the construction of boostrap standard errors is in order.

Remark 3.3 Given the N bootstrap realizations $\{\hat{\theta}_{T:1}^*, \hat{\theta}_{T:2}^*, ..., \hat{\theta}_{T:N}^*\}$, studentized bootstrap confidence intervals for the parameters can be constructed in the usual way. Let θ_i be the i-th element of θ, $i = 1, ..., n_\theta$, $\hat{\theta}_{iT}$ the corresponding QML estimate with associated standard error $s(\hat{\theta}_{iT})$, and $\hat{\theta}_{iT:b}^*$ its bootstrap QML analog with associated standard error $s(\hat{\theta}_{iT:b}^*)$ obtained from the b-th bootstrap sample, $b = 1, ..., N$, see the matrix $\hat{\Omega}_T^*$ in Proposition 2. Compute the bootstrap t-statistic as $t_{i:b}^* := s(\hat{\theta}_{iT:b}^*)^{-1}(\hat{\theta}_{iT:b}^* - \hat{\theta}_{iT})$; then the studentized bootstrap confidence interval for θ_{0i} is given by $[\hat{\theta}_{iT} - c_{1-\eta/2}^* s(\hat{\theta}_{iT}),\ \hat{\theta}_{iT} - c_{\eta/2}^* s(\hat{\theta}_{iT})]$, where $c_{\eta/2}^*$ and $c_{1-\eta/2}^*$ are the $\eta/2$ and $1-\eta/2$ quantiles of $\{t_{i:1}^*, ..., t_{i:N}^*\}$. Likewise, the percentile bootstrap confidence interval for θ_{0i} is given by the $\eta/2$ and $1-\eta/2$ quantiles of $\{\hat{\theta}_{iT:1}^*, \hat{\theta}_{iT:2}^*, ..., \hat{\theta}_{iT:N}^*\}$, denoted $\hat{\theta}_{i,\eta/2}^*$ and $\hat{\theta}_{i,1-\eta/2}^*$, respectively. Finally, the basic bootstrap confidence interval for θ_{0i} is given by $[2\hat{\theta}_{iT} - \hat{\theta}_{i,1-\eta/2}^*,\ 2\hat{\theta}_{iT} - \hat{\theta}_{i,\eta/2}^*]$.

4 Bootstrap diagnostic test

In the previous section we have established the validity of bootstrap inference in state space models under a set of regularity conditions as stated in Proposi-

[9] For sufficiently large N, one can always obtain an arbitrarily accurate estimate of $\mathrm{Var}^*(\hat{\theta}_T^*)$ from the bootstrap realizations $\hat{\theta}_{T:1}^*$, $\hat{\theta}_{T:2}^*$, ..., $\hat{\theta}_{T:N}^*$, by computing

$$\widehat{\mathrm{Var}^*}(\hat{\theta}_T^*) = \frac{1}{N}\sum_{b=1}^N (\hat{\theta}_{T:b}^* - \overline{\hat{\theta}}_T^*)(\hat{\theta}_{T:b}^* - \overline{\hat{\theta}}_T^*)', \quad \overline{\hat{\theta}}_T^* := \frac{1}{N}\sum_{b=1}^N \hat{\theta}_{T:b}^*.$$

Squared bootstrap standard errors correspond to the elements on the main diagonal of $\widehat{\mathrm{Var}^*}(\hat{\theta}_T^*)$, hence practitioners can skip the direct evaluation of the Hessian. Our simulation esperiments (available upon request) show that under the conditions of Proposition 2, the standard errors obtained from $\widehat{\mathrm{Var}^*}(\hat{\theta}_T^*)$ approximate fairly well the standard errors obtained from $\hat{\Omega}_T^*$ (hence the analytic ones obtained from $\hat{\Omega}_T$).

tion 2. One key result is that, under such conditions, the bootstrap estimator is asymptotically normal. As a consequence, lack of normality of the bootstrap estimator, in large samples, may indicate that some of the regularity conditions are violated. In DSGE models this mostly happens when the structural parameters are weakly identified; see e.g. Guerron-Quintana *et al.* (2013), Andrews and Mikusheva (2014, 2015) and references therein.

The null hypothesis we have in mind is that the regularity conditions under which Proposition 2 holds are valid; more precisely, that the statistic R_T^* in (19) is asymptotically Gaussian, conditional on the data. We want to show that $B < N$ bootstrap repetitions out of the N from Algorithm 1 can indeed be used to form a diagnostic test which evaluates model misspecifications along directions that make R_T^* asymtotically non-Gaussian. In particular, our idea is to assess whether the deviations of R_T^* from normality are large enough to reject the null.

To fix ideas, let $G_T^*(x) := P^*(R_T^* \leq x) = P(R_T^* \leq x \mid \text{data})$ denote the CDF (conditional on the original data) of the normalized bootstrap estimator $R_T^* := \hat{\Omega}_T^{-1/2} T^{1/2} (\hat{\theta}_T^* - \hat{\theta}_T)$, see (19), which without loss of generality we assume to be scalar (like in the case where all structural parameters of the state space model but one have been calibrated). Under the conditions of Proposition 2, R_T^* converges to a standard normal random variable. That is, $G_T^*(\cdot)$ satisfies

$$\sup_{x \in \mathbb{R}} |G_T^*(x) - \Phi_Z(x)| \to_p 0$$

as $T \to \infty$. Since this is an asymptotic result, for T fixed the distribution G_T^* (conditional on the original data) will in general deviate from the normal even in cases where Proposition 2 holds. Therefore, our idea is to evaluate the significance of such deviations.

It would be tempting to build a test based on $G_T^*(x) - \Phi_Z(x)$; however, the distribution of this quantity is unknown even in cases where the bootstrap admits an Edgeworth expansion of $G_T^*(x) - \Phi_Z(x)$ of order $O_p\left(T^{-1/2}\right)$. Hence, we take an alternative route based on the bootstrap realizations, i.e. on B i.i.d. draws of (conditional on the data).

Let $R_{T:1}^*, ..., R_{T:B}^*$ denote an i.i.d. sample of B bootstrap realizations of R_T^*. Since the distribution $G_T^*(x)$ of R_T^* is unknown, it is customary to estimate it from $R_{T:1}^*, ..., R_{T:B}^*$ using $G_{T,B}^*(x) := B^{-1} \sum_{b=1}^{B} \mathbb{I}\{R_{T:b}^* \leq x\}$, see (17). For any x, deviation of $G_{T,B}^*(x)$ from the normal distribution can be evaluated by considering

$$G_{T,B}^*(x) - \Phi_Z(x). \tag{21}$$

To derive a proper normalization for (21), notice that the (conditional) independence of $R^*_{T:1}, ..., R^*_{T:B}$ implies that, as $B \to \infty$ (keeping T fixed)

$$B^{1/2}(G^*_{T,B}(x) - G^*_T(x)) \xrightarrow{d} N\left(0, V_T(x)\right) \tag{22}$$

where $V_T(x) := G^*_T(x)\left(1 - G^*_T(x)\right)$. Therefore, we may consider the statistic

$$d_{T,B}(x) := B^{1/2}\hat{V}_T(x)^{-1/2}(G^*_{T,B}(x) - \Phi_Z(x)), \tag{23}$$

where $\hat{V}_T(x)$ is a consistent estimator of $V_T(x)$ (for instance, one may consider $\hat{V}_T(x) := G^*_{T,N}(x)(1 - G^*_{T,N}(x))$ for an arbitrary large value of N, or can be simply set to its theoretical value under normality, i.e. $\hat{V}_T(x) := \Phi_Z(x)\left(1 - \Phi_Z(x)\right)$).

Statistic (23) captures the (normalized) distance between the estimated (over B repetitions) bootstrap distribution $G^*_{T,B}(x)$ and the normal distribution. Its asymptotic distribution under the assumptions of Proposition 2 can be investigated by noticing that $d_{T,B}(x)$ can be decomposed as:

$$\begin{aligned} d_{T,B}(x) &= B^{1/2}\hat{V}_T(x)^{-1/2}(G^*_{T,B}(x) - G^*_T(x)) \\ &\quad + B^{1/2}\hat{V}_T(x)^{-1/2}(G^*_T(x) - \Phi_Z(x)). \end{aligned} \tag{24}$$

For T fixed, by the CLT in (22) the first term on the right-hand side of (24) converges, as $B \to \infty$, to a $N(0,1)$ variate regardless of the validity of the assumptions underlying Proposition 2. Second, suppose that $G^*_T(x) - \Phi_Z(x)$ admits a standard Edgeworth expansion such that $G^*_T(x) - \Phi_Z(x) = O_p\left(T^{-1/2}\right)$, see e.g. Bose (1988) and Kilian (1998). Horowitz (2001, p. 3171) notices that an Edgeworth expansion such that $G^*_T(x) - \Phi_Z(x) = O_p\left(T^{-1/2}\right)$ is the typical case in the presence of asymptotically normal statistics. This would imply that the second term in (24) is of $O_p\left(B^{1/2}T^{-1/2}\right)$ and hence converges to zero in probability provided $B = o(T)$ as both $B \to \infty$ and $T \to \infty$. Summing up, under the convergence facts in Proposition 2 and $G^*_T(x) - \Phi_Z(x) = O_p\left(T^{-1/2}\right)$, $d_{T,B}(x)$ is expected to be asymptotically $N(0,1)$ provided

$$T, B \to \infty \text{ jointly and } BT^{-1} = o(1). \tag{25}$$

Conversely, if (19) in Proposition 2 does not hold, then $G^*_T(x) - \Phi_Z(x)$ does not converge (in probability) to 0, the second term on the right hand side of (24) does not vanishes asymptotically and hence $d_{T,B}(x)$ diverges at the rate of $B^{1/2}$ as $B, T \to \infty$. This is e.g. what we expect to happen when $W_T := T^{1/2}(\hat{\theta}_T - \theta_0)$ is not asymptotically Gaussian, which includes e.g. the case where θ is unidentified or weakly identified, the case where θ_0 lies on the boundaries of Θ, the case of non stationary variables, etc. This result is formalized in the next theorem.

Proposition 3 (Bootstrap diagnostics) *Consider the state space model in (1)-(2) under Assumptions A1–A3, A4', A5, with θ one-dimensional, and assume that for some $\alpha > 0$,*

$$G_T^*(x) - \Phi_Z(x) = O_p\left(T^{-\alpha}\right). \tag{26}$$

Furthermore, for $x \in \mathbb{R}$ and some positive B, let $G_{T,B}^*(x) := B^{-1}\sum_{b=1}^{B} \mathbb{I}\{R_{T:b}^* \leq x\}$, where $R_{T:b}^*$, $b = 1, ..., B$, are i.i.d. draws from $G_T^*(x)$, the distribution of R_T^* conditional on the original sample. Finally, let

$$d_{T,B}(x) := B^{1/2}\hat{V}_T(x)^{-1/2}(G_{T,B}^*(x) - \Phi_Z(x))$$

where $\hat{V}_T(x) := \Phi_Z(x)(1 - \Phi_Z(x))$. Then, for $T, B \to \infty$ jointly and

$$BT^{-2\alpha} = o(1), \tag{27}$$

it holds the convergence

$$d_{T,B}(x) \xrightarrow{d^*}_p N(0,1).$$

Conversely, if $G_T^*(x) \to_p G_\infty^*(x) \neq \Phi_Z(x)$, $d_{T,B}(x)$ diverges at the rate of $B^{1/2}$.

Notice that a test based on $d_{T,B}(x)$ is simply a normality test based on the B bootstrap realizations of the QML estimator of θ. Such B realizations are usually available from the N used e.g. to compute bootstrap standard errors for the structural parameters, critical regions or p-values, see Section 3, hence no extra computational effort is required to compute the test.

Few remarks are in order.

Remark 4.1 In (26) the notation $O_p\left(T^{-\alpha}\right)$ is consistent with the fact that $G_T^*(x)$ is by definition a function of the original data only. In the standard case where the bootstrap admits an Edgeworth expansion such that α in (26) equals 1/2, the number of bootstrap repetitions used to compute the test should not be large compared to T, i.e. BT^{-1} should be of order $o(1)$. In general, if the ratio $BT^{-\alpha}$ does not converge to zero, the normalized distance $d_{T,B}(x)$ in Proposition 3 does not converge to the Gaussian distribution, even when the bootstrap is consistent. This means that in practice the ratio B/T must be selected carefully in finite samples in order to reduce the risk of false rejections. In the next sections and SM we provide some practical examples which show that even in relatively small samples the suggested test detects violations from asymptotic normality with reasonable finite sample power.

Remark 4.2 Proposition 3 covers the case of a simple test based on the CDF of the normal distribution at a given point x. Different normality tests could be employed as well, following the same principle; see the next sections where different (univariate and multivariate) normality tests will be considered and applied to small scale monetary DSGE models.

Remark 4.3 When θ and R_T^* are $n_\theta \times 1$ ($n_\theta > 1$) it is possible to associate a quantity like $d_{T,B}(x)$ in (23) to any component of the vector R_T^*, hence our diagnostic test can be computed by considering multivariate normality tests for R_T^* as well as univariate normality tests on the single components. Our suggestion is to look at the univariate normality tests conditional on the outcome of the multivariate normality tests.

We explore the empirical performance of our bootstrap approach in Sections 5 and 6 below.

5 Monte Carlo study

In this section we investigate the empirical performance of our diagnostic test on simulated data. The reference model is the state space form associated with Guerron-Quintana *et al.* (2013)'s DSGE model. Section 5.1 describes the design and Sections 5.2-5.3 summarize the results obtained with two versions of the model, one where the estimated parameter is strongly identified, and the other where the estimated parameters are, according to the literature, suspected to be weakly identified.[10] Additional Monte Carlo results based on the ARMA(1,1) model are confined in SM.

5.1 Design

The state space model is taken from the first-order equilibrium of the small-scale DSGE model analyzed in Guerron-Quintana *et al.* (2013), simulation design 1; see also Guerron-Quintana *et al.* (2017). The structural equations are given by:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t \tag{28}$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\phi_\pi \pi_t + (1-\rho_r)\phi_x x_t + \sigma_r \varepsilon_{r,t} \tag{29}$$

$$x_t = E_t x_{t+1} - \sigma(r_t - E_t \pi_{t+1} - z_t) \tag{30}$$

[10]DSGE models have been extensively used in the econometric literature to study identification issues, see among others, Komunjer and Ng (2011), Qu and Tkachenko (2012), Qu (2014) and Castelnuovo and Fanelli (2015).

$$z_t = \rho_z z_{t-1} + \sigma_z \varepsilon_{z,t}. \tag{31}$$

Equation (28) is a purely forward-looking New-Keynesian Phillips Curve with slope $\kappa := \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}\frac{(\omega+\sigma)}{\sigma(\omega+\theta)}$, π_t is the inflation rate and x_t the output gap; (29) is the monetary policy rule, r_t is the policy rate and $\varepsilon_{r,t}$ is monetary policy shock assumed to be i.i.d. with unit variance; (30) is a forward-looking output-gap equation; finally, (31) maintains that the aggregate demand disturbance (z_t) is an autoregressive process driven by i.i.d. shock $\varepsilon_{z,t}$ with unit variance. The whole vector of structural parameters is $(\beta, \alpha, \omega, \sigma, \theta, \rho_r, \phi_\pi, \phi_x, \rho_z, \sigma_r^2, \sigma_z^2)'$.

The structural model (28)-(31) can be solved for rational expectations and the implied equilibrium can be represented in the state space form (1)-(2) with associated AKC form as in (3)-(4). Our Monte Carlo experiments consider two versions of this model, denoted GQ-DGP1 and GQ-DGP2, respectively. In GQ-DGP1 the estimated structural parameter is the probability of not adjusting prices for firms, $\theta_1 := (\alpha)$ ($n_{\theta_1} = 1$), and all the other parameters are calibrated at their DGP values as in Guerron-Quintana *et al.* (2017). In GQ-DGP2, the estimated structural parameters are all policy rule coefficients $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ ($n_{\theta_2} = 3$), and all the other structural parameters are calibrated at their DGP values as in Guerron-Quintana *et al.* (2017).

For both DGPs we generate samples of length $T = 100$ and 500 from the AKC form (3)-(4) $M = 2000$ times, assuming non-Gaussian shocks. More precisely, for each structural shock of the model we use independent Student-t distributions with 5 degrees of freedom. The initial condition $Z_{1|1}$ is fixed to zero. For each replication, a sample of $T + 200$ observations is actually generated and the first 200 observations are then discarded. Estimation is carried out by combining the Kalman filter with the 'BFGS' likelihood maximization algorithm, imposing bounds on the permissible parameter values and determinacy.[11] Bootstrap estimation follows the algorithm described in Section 3. Bootstrap confidence intervals are computed as explained in the Remark 3.3, using $N = 499$ bootstrap replications.

Assumptions A1-A5 are satisfied for GQ-DGP1. As concerns GQ-DGP2, based on the empirical evidence reported in Mavroeidis (2010), Qu and Tkachenko (2012), Qu (2014) and Castelnuovo and Fanelli (2015), we expect some of the policy rules parameters in $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ to be weakly identified with possible consequence on the validity of standard asymptotic inference. It is worth

[11] The condition $\phi_\pi > 1$ sufficies to ensure determinacy in this DSGE model. The determinacy condition ensures that the rational expectations solution of the structural model (28)-(31) can be represented in the form (1)-(2) without involving 'additional' parameters other those considered in the analysis that follows, or 'additional' shocks other $(\varepsilon_{r,t}, \varepsilon_{z,t})'$, see e.g. Castelnuovo and Fanelli (2015) for details.

observing that in this state space model it is not possible to directly tie the strength of identification to a local-to-zero embedding as we can easily do e.g. for the class of ARMA(1,1) models analyzed in SM. While this fact represents a challenge to the reliability of our diagnostic test, we do not need to take a stand a-priori on the directions of near identification or unidentification.

5.2 GQ-DGP1

Results obtained from GQ-DGP1 are summarized in Tables 1-2, and in the left-panel of Figure 1. Table 1 reports estimation results and Table 2 the diagnostic tests. Figure 1 reports the fan chart of the empirical cumulative density functions (CDFs) of the bootstrap realizations of the QML estimator (across Monte Carlo simulations) of the parameter $\theta_1{:=}(\alpha)$ used to compute the tests, see Cavaliere and Georgiev (2020) for details.

From Table 1 we observe a situation in which the bootstrap QML estimates and standard errors of $\theta_1{:=}(\alpha)$ tend to closely replicate their non-bootstrap analogs in line with the results (18) and (20) in Proposition 2. The lengths of 90% bootstrap confidence intervals for α track closely their non-bootstrap counterparts. The empirical coverage probabilities of 90% bootstrap confidence intervals tend to nominal size as the sample length increases.

For $T = 100$, the empirical coverage probability of the 90% bootstrap confidence interval for $\theta_1{:=}(\alpha)$ is comparable with the empirical coverage probabilities reported for the strongly identified α in Table 2 of Guerron-Guintana *et al.* (2017), who generate the data under Gaussian shocks. In Guerron-Guintana *et al.* (2017), empirical rejection probabilities vary with the lag order of the VAR model used to approximate the observable variables, the maximum horizon for the impulse response functions used in the impulse response matching estimation exercise, and the choice of the weighting matrix (diagonal or optimal). Our approach does not require any VAR specification, lag order and weighting matrix. It can be noticed from Table 1 that the empirical coverage probabilities of the 90% bootstrap confidence intervals are not inferior, on average, to the coverages reported in Table 2 of Guerron-Guintana *et al.* (2017).

Table 2 refers to computationally straightforward versions of our tests of model misspecification. The tests are designed to verify the asymptotic normality of the sequences $\{\hat{\theta}^*_{1,T:1}, ..., \hat{\theta}^*_{1,T:B}\}$, where $B < N$ is selected as detailed below. The table reports the empirical rejection frequencies of Doornik and Hansen's (2008) multivariate normality test (DH),[12] Jarque and Bera's (1987)

[12] Since in GQ-DGP1 $n_{\theta_1} = 1$, the DH multivariate test of normality boils down to a univariate test.

univariate normality test (JB)[13] and Shapiro and Wilk's (1965) univariate normality test (SW). All tests are computed at the 5% nominal level of significance. Recall that B, the number of bootstrap realizations used to compute the misspecification tests, must be chosen to satisfy the condition (25) (see Remark 4.1) and, therefore, should be small relative to T. We select B by the practical rule: B :=int$[(1/i)T^{4/5}]$, $i = 2, 3$ which, as it will be shown below, provides a reasonable compromise between size control and power increasing with samples length. In practice, for e.g. $T = 100$, the tests are computed by using $B = 19$ and $B = 13$ bootstrap replications (out of the $N = 499$ used to compute confidence intervals) of the QML estimator of the parameters, respectively.

The left-panel of Figure 1 plots the percentiles of the empirical CDF of the sequences $\{\hat{\theta}^*_{1,T:1}, ..., \hat{\theta}^*_{1,T:B}\}$ generated across the Monte Carlo simulations. The graph confirms that the CDFs tend to normality as the sample size T increases (and B/T remaining 'small').

Summing up, the results in Table 2 and the left-panel of Figure 1 appear consistent with the result (19) of Proposition 2, i.e. they support the convergence of the bootstrap QML estimator of the parameter θ_1:=(α) to the Gaussian distribution. With the chosen rules for B, the empirical size of the normality tests tend to fluctuate around the 5% nominal level of significance. In general, size appears under control for all tests. Overall, the validity of standard asymptotic inference can safely be considered valid for this state space model.

5.3 GQ-DGP2

Results obtained from GQ-DGP2 are also summarized in Tables 1-2 and in the right-panel of Figure 1.

Based on the estimates in Table 1, we notice that regardless of the sample size, the bootstrap standard errors of the QML estimator of the policy rule parameters $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ tends to depart from the analytic standard errors (i.e those taken from the diagonal of the matrix $\hat{\Omega}_T$; see Propositions 1 and 2). The same phenomenon can be observed by inspecting the lengths of the bootstrap and non-bootstrap 90% confidence intervals, especially for the parameter which captures the policy response to inflation, ϕ_π. We interpret these facts as prima facie evidence of discrepancy of the bootstrap distribution of the QML estimator of θ_2 from its non-bootstrap counterpart.

The results of the asymptotic normality tests in Table 2, considered jointly

[13] Throughout the paper we apply a version of the JB test which does not incorporate Kilian and Demiroglu's (2000) correction.

at the multivariate and individual levels, lead us to challenge the idea that inference is standard in this state space model. This is seen from the rejection frequencies of the normality tests which tends to increase with the sample size. In general, the tests suggest deviations from the Gaussian distribution with rejection frequencies of the DH multivariate normality test which lie in the range 0.25-0.37 for $T = 100$, and in the range 0.40-0.51 for $T = 500$. The rejection frequencies of the asymptotic univariate normality tests JB and SW tend to approach one for $T = 500$.

The CDFs of the bootstrap distributions plotted in the right-panel of Figure 1 show that deviations from the Gaussian distribution are substantial and persist when T increases. The graphs point out that deviations from normality are particularly marked for the parameters capturing the central bank's response to inflation and the output gap.

Overall, the simulation experiments based on model GQ-DGP2 confirms that the bootstrap distribution of the QML estimator is very informative and useful in this model. The policy rule parameters in equation (29) are weakly identified so that in this estimated state space model inference on the structural parameters cannot be conducted and interpreted 'in the usual way'. Importantly, the test displays power against weakly identified parameters even in relatively short samples ($T = 100$), which means that practitioners can robustify their inference by moving to the identification-robust methods discussed e.g. in Dufour *et al.* (2013), Guerron-Quintana *et al.* (2013), Qu (2014) and Andrews and Mikusheva (2015).

The simulation experiments discussed in this and in the previous section suggest some simple practical rules-of-thumb that practitioners can follow to interpret the outcomes of the normality tests. For instance, practitioners should interpret the normality tests with caution when the rejection/non-rejection of normality is associated with the observation of mild/sharp differences between Hessian-based and bootstrap-based standard errors and confidence intervals. Second, simulation results stress that one of the advantages of our approach is that we do not need to take a stand on the directions of (near) identification failures. In this respect, we suggest to primarily assess the multivariate normality of $\hat{\theta}^*_T$ and, in case or rejection, move to the univariate normality tests in order to envisage the possible directions of (near) identification failure.[14]

[14] Obviously, in order to assess asymptotic multivariate normality, alternative strategies could be implemented based on the univariate asymptotic normality tests. In these cases, while we know that there exists multivariate non-Gaussian distributions that have normal univariate marginals, rejection of normality of a single component of $\hat{\theta}^*_T$ would suffice to reject multivariate normality. However, strategies based on univariate normality tests entail multiple testing problems that can be addressed through Bonferroni-type procedures. We

6 Empirical illustration

In this section we take the state space form of Guerron-Quintana *et al.* (2013)'s DSGE model (28)-(31) already analyzed in the Monte Carlo section to U.S. data. The objective is to test the reliability of standard asymptotic inference in two distinct versions of this model. We employ quarterly observations relative to the 'Great Moderation' sample 1984Q2-2008Q3 (T=98). The starting date, 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 episodes.

Based on an extensive literature, we can assume that the monetary DSGE models is determinate on the Great Moderation period 1984Q2-2008Q3, see e.g. Lubik and Schorfheide (2004), Castelnuovo and Fanelli (2015) and references therein. This ensures that QML estimation based on the specification (1)-(2) does not omit propagation mechanisms, 'additional' parameters unrelated to θ and additional shocks unrelated to the fundamental shocks that would arise in the presence of multiple solutions. However, to check the performance of our test in situations in which the estimated state space model might be misspecified because of unaccounted shocks or unaccounted changes of regimes, in the SM we estimate the model also on the 'Great Inflation' sample, 1954Q3-1984Q1, and on the full sample, 1954Q3-2008Q3, respectively.

The two observable variables in $y_t := (\pi_t, r_t)'$ are measured as follows. The inflation rate, π_t, is the quarterly growth rate of the GDP deflator. The short-term nominal interest rate r_t is proxied by the effective Federal funds rate expressed in quarterly terms (averages of monthly values). Data are collected from the Federal Reserve Bank of St. Louis' web site.

As in the Monte Carlo study, we consider two estimable versions of this model. In the former, denoted GQ-M1, the estimated structural parameter is the probability of not adjusting prices $\theta_1 := (\alpha)$ (n_{θ_1}=1), and all other parameters are calibrated as in the Monte Carlo exercise. Based on Monte Carlo results we expect standard asymptotic inference to hold in this model. In the second, denoted GQ-M2, the estimated parameters are the policy rule parameters $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ (n_{θ_2}=3) and, again, all other parameters are calibrated as in the Monte Carlo exercise. The Monte Carlo results in the previous section and the available empirical evidence on U.S. quarterly data suggest that the policy rule parameters might be poorly identified on 'Great

refer to e.g. Looney (1995) for a practical treatment of how multivariate normality can be assessed based on univariate normality tests.

Moderation' samples, see e.g. Mavroeidis (2010) and Castelnuovo and Fanelli (2015). Hence, we expect standard asymptotic inference to be problematic in model GQ-M2.

Estimation is performed by combining the Kalman filter with the 'BFGS' likelihood maximization algorithm maintaining a Gaussian distribution. In all models, the bootstrap standard errors associated with the QML estimates of $\theta_1 := (\alpha)$ and $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ are computed by using $N := 1999$ bootstrap replications and the algorithm summarized in Section 3. The bounds on the permissible parameter values are specified as in the Monte Carlo experiments.[15] Driven by the results of the Monte Carlo experiments, the rule used to select the number of bootstrap replications in the diagnostic tests is $B = \text{int}[(1/i)T^{4/5}]$, $i = 2, 3$; since $T = 98$, in practice this means using $B = 19$ and $B = 13$ replications of the QML estimator, respectively. The normality tests are the same used in the Monte Carlo experiments (DM, JB and SW) and are computed at the 5% nominal level of significance. Empirical results are summarized in Table 3.

As regards model GQ-M1, we notice that bootstrap standard errors and the associated 90% bootstrap confidence intervals for $\theta_1 := (\alpha)$ are numerically similar to the Hessian-based standard errors and the 90% asymptotic confidence intervals, respectively. Our diagnostic tests indicate that asymptotic normality is strongly supported at the 5% level. The left panel of Figure 2 plots the CDF and the empirical probability distribution function (PDF) (bottom panel) of the sequence $\{\hat{\theta}^*_{1,T:1}, ..., \hat{\theta}^*_{1,T:B}\}$ (with $B = 19$) of the bootstrap estimates of $\theta_1 := (\alpha)$, contrasted with the Gaussian distribution. The graphical inspection seems to confirm the results of the tests. We can conclude that in the estimated GQ-M1 model the conditions for standard asymptotic inference are at work on the Great Moderation period. It turns out that the reported asymptotic and bootstrap standard errors and asymptotic and bootstrap 90% confidence intervals for the probability parameter of not adjusting prices for firms can be considered highly reliable.

As regards model GQ-M2, from Table 3 we observe a substantial discrepancy between the bootstrap QML estimates, bootstrap standard errors and their non-bootstrap counterparts. This is particularly true for the policy parameters ϕ_π and ϕ_x. For this model, the combination of the outcomes of the multivariate and univariate normality tests point towards the rejection that

[15] Actually, in order to avoid computational issues in the likelihood maximization, the interval of permissible parameter values for the policy parameter ϕ_π is (0.5, 5) rather than (1, 5) (that would ensure determinacy). Estimation results on the Great Inflation sample, however, show that the QML point estimate of ϕ_π lies in the determinacy region even in the absence of any restriction. As our diagnostic test will show, the main issue with the parameter ϕ_π is weak identification.

that standard asymptotic inference holds. The multivariate DH test rejects normality at the 5% level of significance. The univariate tests: (i) seem to support normality for the estimator of the inertia parameter ρ_r; (ii) are more controversial for the estimator of the response to output gap ϕ_x; (iii) firmly reject asymptotic normality for the estimator of the response to inflation parameter ϕ_π. These evidences are consistent with the graphs reported in the right-panel of Figure 2, where the empirical CDF and empirical PDF of $B = 19$ bootstrap replications of the QML estimator of $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ are plotted against the Gaussian distribution. The graphs show that departure from asymptotic normality is substantial for ϕ_π.

Overall, and in line with the results of the Monte Carlo experiments, our empirical analysis suggests that the validity of standard asymptotic inference should not be taken for granted in model GQ-M2 on the Great Moderation sample. Thus, reported asymptotic and bootstrap standard errors and asymptotic and bootstrap 90% confidence intervals should be interpreted with caution. In this model the inference on the policy parameters in θ_2 can be robustified by resorting to the methods discussed in e.g. Mavroeidis (2010), Guerron-Quintana *et al.* (2013), Qu (2014) and Andrews and Mikusheva (2015).

7 Conclusions

We have proposed a novel approach for state space models where the bootstrap is used as a diagnostic tool. Using the state space form associated with a small-scale monetary DSGE model, we have investigated how our approach works on simulated and actual data in the presence of weakly identified parameters. Finite sample results suggest that the bootstrap distribution of the QML estimator of the parameters of interest is informative about the strength of identification and the quality of the inference in an estimated state space model.

For a proper choice of the ratio B/T, the suggested test controls size and has power against forms of misspecification of the state space model which lead to deviations from asymptotic normality. When the null is not rejected, the bootstrap can be used in the 'conventional' way to improve finite sample inference. Conversely, practitioners should interpret bootstrap (and asymptotic) standard errors and p-values of tests with caution when the bootstrap distribution deviates asymptotically from the Gaussian. In these cases the inference can be robustified along the lines suggested by e.g. Guerron-Quintana *et al.* (2013), Dufour *et al.* (2013), Qu (2014) and Andrews and Mikusheva (2015).

ACKNOWLEDGMENTS

A previous version of this paper circulated under the title 'Bootstrapping DSGE models'. We are grateful to the Editor, Barbara Rossi, and three anonymous referees for the many constructive comments and suggestions. We also thank for their comments Majid Al-Sadoom, Guido Ascari, Gunnar Bårdsen, Efrem Castelnuovo, Athanas Christev, Riccardo Lucchetti, Katerina Petrova, Anders Rahbek, Denis Tkachenko and seminar and conference participants at Norwegian University of Science and Technology, National University of Singapore, Durham University, Herriot-Watt University, Vrije Universiteit Amsterdam, as well as participants to the 2016 Latin American Meeting of the Econometric Society (Medellín, 10-12 November 2016), the IAAE 2016 Annual Meeting (Milan, 22-25 June 2016), the 7th ICEEE (Messina, 25-27 January, 2017), the 5th Padova Macro Talks (Padova, July 2019), the 2nd IWEEE (Venice, 23-24 January, 2020). We gratefully acknowledge financial support from MIUR (PRIN 2017, Grant 2017TA7TYC) and the University of Bologna (ALMA IDEA 2017 grants).

References

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

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

Andrews, I. and A. Mikusheva (2014), Weak identification in maximum likelihood: A question of information, *American Economic Review: Papers & Proceedings* 104, 195-199.

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

Angelini G., G. Cavaliere and L. Fanelli (2021), An identification strategy for proxy-SVARs with weak proxies, *Working Paper*, University of Bologna.

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

Bårdsen, G. and L. Fanelli (2015b), Frequentist evaluation of small DSGE models, *Journal of Business and Economic Statistics*, Supplementary Material.

Beran, R. (1997), Diagnosing bootstrap success, *Annals of the Institute of Statistical Mathematics* 49, 1-24.

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

Bose, A. (1988), Edgeworth correction by bootstrap in Autoregressions, *Annals of Statistics* 16, 1709–1722.

Caines, P.E. (1988), *Linear stochastic systems*, Wiley, New York.

Canova F. and L. Sala (2009), Back to square one: identification issues in DSGE models, *Journal of Monetary Economics* 56, 431-449.

Castelnuovo, E. and L. Fanelli (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. and I. Georgiev (2020), Inference under random limits bootstrap measures, *Econometrica* 88, 2547-2574.

Cavaliere, G., Nielsen, H.B. and A. Rahbek (2017), On the consistency of bootstrap testing for a parameter on the boundary of the parameter space, *Journal of Time Series Analysis* 38,513-534.

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

Commandeur, J.J.F and S.J. Koopman (2007), *An introduction to state space time series analysis*, Oxford University Press, Oxford.

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

Dovonon, P. and S. Gonçalves (2017), Bootstrapping the GMM overidentification test under first-order underidentification, *Journal of Econometrics* 201, 43-71.

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

Durbin, J. and S.J. Koopman (2001), *Time series analysis by state space methods*, Oxford Statistical Science Series, Oxford University Press, Oxford.

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

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

Gonçalves, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89-120.

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

Guerron-Quintana P, Inoue A. and L. Kilian (2017), Impulse response matching estimators for DSGE models, *Journal of Econometrics* 196, 144-155.

Hall, A. R., Inoue, A., Nason, J.M. and B. Rossi (2012), Information criteria for impulse response function matching estimation of DSGE models, *Journal of Econometrics* 170, 499-518.

Hannan, E.J. and M. Deistler (1988), *The statistical theory of linear systems*, SIAM, Wiley, New York.

Hansen, L. P. and T. Sargent (2005), *Recursive methods of dynamic linear economies*, Unpublished Manuscript, University of Chicago. Available at http://home.uchicago.edu/~lhansen/mbook2.pdf.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge University Press, Cambridge MA.

Horowitz, J.L. (2001), The bootstrap, *Handbook of Econometrics* Vol. 5, Heckman, J.J. and E. Leamer (eds.), Ch. 52.

Inoue, A. and B. Rossi (2011), Testing for weak identification in possibly nonlinear models, *Journal of Econometrics* 161, 246-261.

Jarque, C. M. and A. K. Bera (1987), A test for normality of observations and regression residuals, *International Statistical Review* 55, 163-172.

Khalaf L., Lin Z. and A. Reza (2019), Identification and persistence-robust exact inference in DSGE models, *Working Paper,* Department of Economics, Carleton University.

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

Kilian, L. and U. Demiroglu (2000), Residual-based tests for normality in autoregressions. Asymptotic theory and simuation evidence, *Journal of Business and Economic Statistics* 18, 40-50.

Kleibergen. F. and S. Mavroeidis (2009), Weak instrument robust tests and the new Keynesian Phillips curve, *Journal of Business and Economic Statistics* 27, 293-311.

Kocięcki, A. and M. Kolasa (2018), Global identification of linearized DSGE models, *QuantitiativeEconomics* 9, 1243-1263.

Komunjer, I. and S. Ng (2011), Dynamic identification of dynamic stochastic general equilibrium models, *Econometrica* 79, 1995-2032.

Le, V.P.M, Meenagh, D., Minford, P. and M. Wickens (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.

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

Looney, S.W. (1995), How to use tests for univariate normality to assess multivariate normality, *The American Statistician* 49, 64-70.

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

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

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

McConnell, M. M., and G. Perez-Quiros (2000), Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review* 90, 1464-1476.

Mertens, K. and Ravn, M. (2013), The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* 103, 1212-1247.

Moreira, M., Porter, J.R. and G.A. Suarez (2004), Bootstrap and higher-order expansions validity when instruments may be weak, *NBER Working Paper* 302.

Morris, S.D. (2017), DSGE pileups, *Journal of Economic Dynamics and Control*, 74, 56-96.

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

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

Qu, Z. and D. Tkachenko (2017), Global identification in DSGE models allowing for indeterminacy, *Review of Economic Studies* 84, 1306-1345.

Shapiro, S.S., and M. B. Wilk (1965), An analysis of variance test for normality (complete samples), *Biometrika* 52, 591-611.

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

Stock, J.H., and M.W. Watson, (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal* 128, 917-948.

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

Stoffer, D.S. and K. Wall (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 K. Wall (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. Shephard, Chap. 9, Cambridge University Press, Cambridge.

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

Zhan, Z. (2018), Detecting weak identification by bootstrap, *mimeo.*

TABLE 1. Monte Carlo experiment: estimates.

θ (a)	$Bounds$ (b)	$\hat{\theta}_T$ (c)	$s(\hat{\theta}_T)$ (d)	90%-CI (e)	$\hat{\theta}^*_T$ (f)	$s(\hat{\theta}^*_T)$ (g)	90%-BSCI (h)	$Asymp.$ (i)	Bootstrap $Stud.$ (l)	 $Perc.$ (j)	 $Basic$ (k)
$T = 100$											
					GQ-DGP1						
$\alpha = 0.75$	$(0, 1)$	0.75	0.02	$0.05, [0.73, 0.78]$	0.75	0.07	$0.05, [0.73, 0.79]$	0.85	0.82	0.82	0.81
					GQ-DGP2						
$\rho_r = 0.75$	$(0, 1)$	0.74	1.59	$0.38, [0.53, 0.94]$	0.75	0.89	$0.18, [0.65, 0.86]$	0.96	0.87	0.86	0.78
$\phi_\pi = 1.50$	$(1, 5)$	1.69	31.21	$4.00, (1.00, 5.00)$	1.69	17.39	$1.75, (1.00, 2.79]$	0.97	0.89	0.87	0.61
$\phi_x = 0.13$	$(0, 1.5)$	0.11	6.27	$0.12, (0.00, 0.80]$	0.12	3.78	$0.35, (0.00, 0.36]$	0.96	0.91	0.86	0.43
$T = 500$											
					GQ-DGP1						
$\alpha = 0.75$	$(0, 1)$	0.75	0.01	$0.02, [0.74, 0.76]$	0.75	0.03	$0.02, [0.74, 0.77]$	0.86	0.84	0.82	0.84
					GQ-DGP2						
$\rho_r = 0.75$	$(0, 1)$	0.74	0.90	$0.15, [0.65, 0.81]$	0.75	0.43	$0.10, [0.69, 0.80]$	0.96	0.92	0.86	0.68
$\phi_\pi = 1.50$	$(1, 5)$	1.63	17.21	$2.57, (1.00, 3.58]$	1.62	7.19	$1.46, (1.00, 2.48]$	0.96	0.93	0.85	0.62
$\phi_x = 0.13$	$(0, 1.5)$	0.11	3.63	$0.34, (0.00, 0.35]$	0.11	1.58	$0.30, (0.00, 0.31]$	0.95	0.93	0.85	0.62

Notes: Results are based on $M = 2000$ Monte Carlo simulations and $N = 499$ bootstrap replications of the QML estimator. Column (a): true parameter values. Column (b): bounds for permissible parameter values. Column (c): QML estimates (average across Monte Carlo simulations). Column (d): standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}_T$, see Remark 2.1, average across Monte Carlo simulations). Column (e): lengths of 90% confidence intervals (in each cell, the first value is the median length of the intervals and the remaining two values are the medians of their lower and upper limits across Monte Carlo simulations. Column (f): bootstrap QML estimates (average across Monte Carlo simulations of the $\bar{\hat{\theta}}^*_T$s, see footnote 8). Column (g): bootstrap standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}^*_T$, see Proposition 2, average across Monte Carlo simulations). Column (h): lengths of 90% bootstrap confidence intervals (computed as explained in the Remark 3.3; the first value is the median length of the intervals and the remaining two values are the medians of their lower and upper limits across Monte Carlo simulations). Column (i): empirical probability coverages, i.e. the frequencies that the asymptotic confidence intervals contain the true parameter values. Column $(l$-$k)$: bootstrap coverages.

TABLE 2. Monte Carlo experiment: misspecification tests for *GQ-DGP1* and *GQ-DGP2*.

	$B=\frac{1}{3}T^{4/5}$	$B=\frac{1}{2}T^{4/5}$	$B=\frac{1}{3}T^{4/5}$	$B=\frac{1}{2}T^{4/5}$	$B=\frac{1}{3}T^{4/5}$	$B=\frac{1}{2}T^{4/5}$
	DH		*JB*		*SW*	
(a)	(b)		(c)		(d)	
$T=100$						
			GQ-DGP1			
$\alpha=0.75$	0.05	0.06	0.04	0.05	0.07	0.07
			GQ-DGP2			
$\rho_r=0.75$			0.07	0.08	0.08	0.10
$\phi_\pi=1.50$	0.25	0.37	0.34	0.41	0.67	0.83
$\phi_x=0.13$			0.36	0.43	0.72	0.83
$T=500$						
			GQ-DGP1			
$\alpha=0.75$	0.04	0.05	0.05	0.05	0.07	0.06
			GQ-DGP2			
$\rho_r=0.75$			0.15	0.21	0.25	0.33
$\phi_\pi=1.50$	0.40	0.51	0.77	0.94	0.99	1.00
$\phi_x=0.13$			0.79	0.93	0.97	0.98

Notes: Empirical rejection frequencies based on $M=2000$ Monte Carlo simulations and B bootstrap samples selected out of $N=499$ bootstrap replications. Column (b): rejection frequency of Doornik and Hansen's (2008) multivariate normality test. Column (c): rejection frequency of Jarque and Bera's (1987) univariate normality test. Column (d) rejection frequency of Shapiro and Wilk's (1965) univariate normality test. All tests are computed at the 5% nominal significance level.

TABLE 3. Estimated DSGE model on U.S. quarterly data.

Sample 1984:Q2-2008:Q3, $T = 98$, $N = 1999$													
Estimates								Misspecification tests					
								$B = \frac{1}{3}T^{4/5}$	$\frac{1}{2}T^{4/5}$	$\frac{1}{3}T^{4/5}$	$\frac{1}{2}T^{4/5}$	$\frac{1}{3}T^{4/5}$	$\frac{1}{2}T^{4/5}$
θ (a)	*Bounds* (b)	$\hat{\theta}_T$ (c)	$s(\hat{\theta}_T)$ (d)	90%-CI (e)	$\bar{\hat{\theta}}^*_T$ (f)	$s(\hat{\theta}^*_T)$ (g)	90%-BSCI (h)	*DH* (i)		*JB* (l)		*SW* (j)	
							GQ-M1						
α	$(0, 1)$	0.81	0.01	$[0.79, 0.82]$	0.81	0.01	$[0.80, 0.83]$	0.72	0.88	0.47	0.62	0.54	0.72
							GQ-M2						
ρ_r	$(0, 1)$	0.89	0.02	$[0.85, 0.92]$	0.87	0.48	$[0.87, 0.88]$			0.55	0.06	0.52	0.11
ϕ_π	$(1, 5)$	1.46	0.27	$[1.01, 1.91]$	1.18	15.52	$[1.16, 1.46]$	0.01	0.00	0.07	0.04	0.01	0.00
ϕ_x	$(0, 1.5)$	0.44	0.31	$(0.00, 0.96]$	0.51	5.72	$[0.44, 0.64]$			0.00	0.00	0.00	0.00

Notes: Column (a): estimated parameters. Column (b): bounds for permissible parameter values. Column (c): QML estimates. Column (d): standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}_T$, see Remark 2.1). Column (e): 90% confidence intervals (lower and upper limits). Column (f): bootstrap QML estimates ($\bar{\hat{\theta}}^*_T$s, see footnote 8). Column (g): bootstrap standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}^*_T$, see Proposition 2). Column (h): 90% bootstrap confidence intervals (lower and upper limits). Column (i): p-values of Doornik and Hansen's (2008) multivariate normality test. Column (j): p-values of Jarque and Bera's (1987) univariate normality test. Column (j) p-values of Shapiro and Wilk's (1965) univariate normality test.

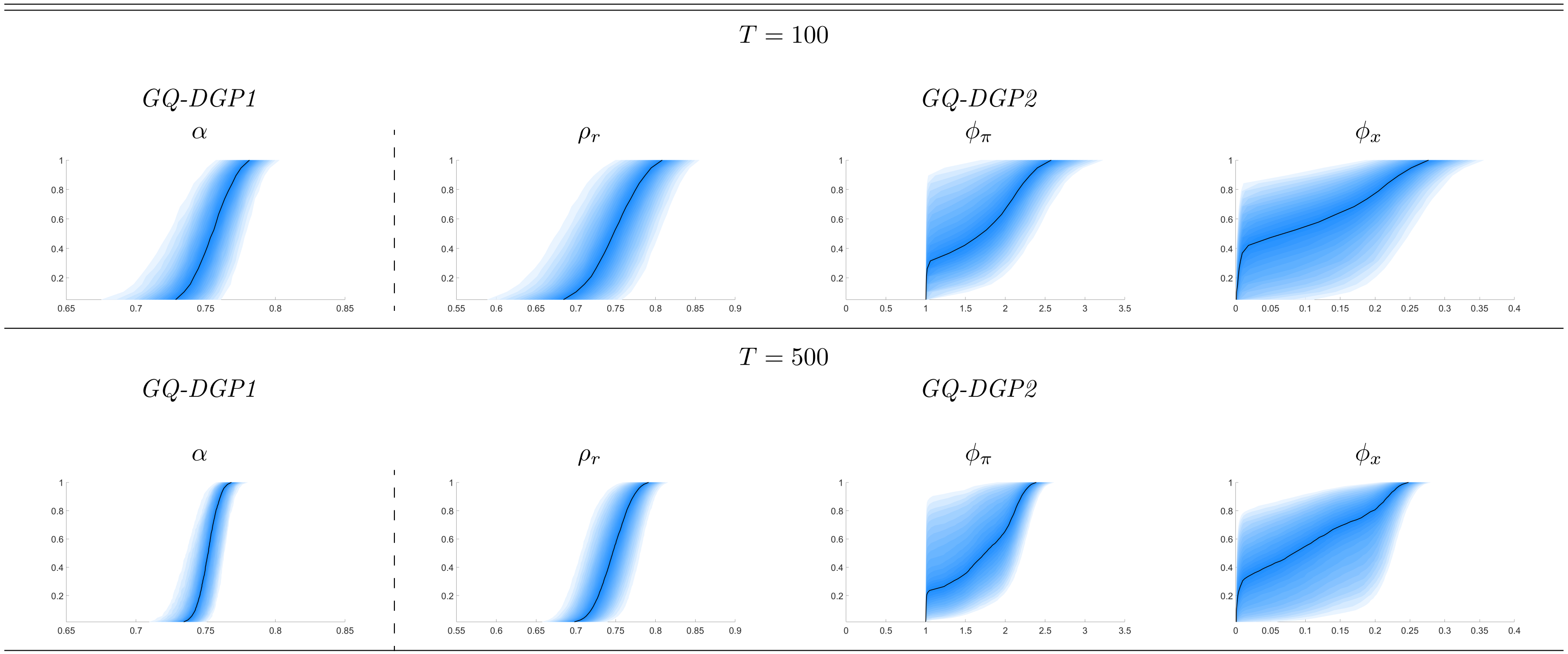

FIGURE 1: *Percentiles of the CDFs of $B = \frac{1}{2}T^{4/5}$ bootstrap replications of the QML estimator of the parameters across Monte Carlo simulations. The black lines indicate the median.*

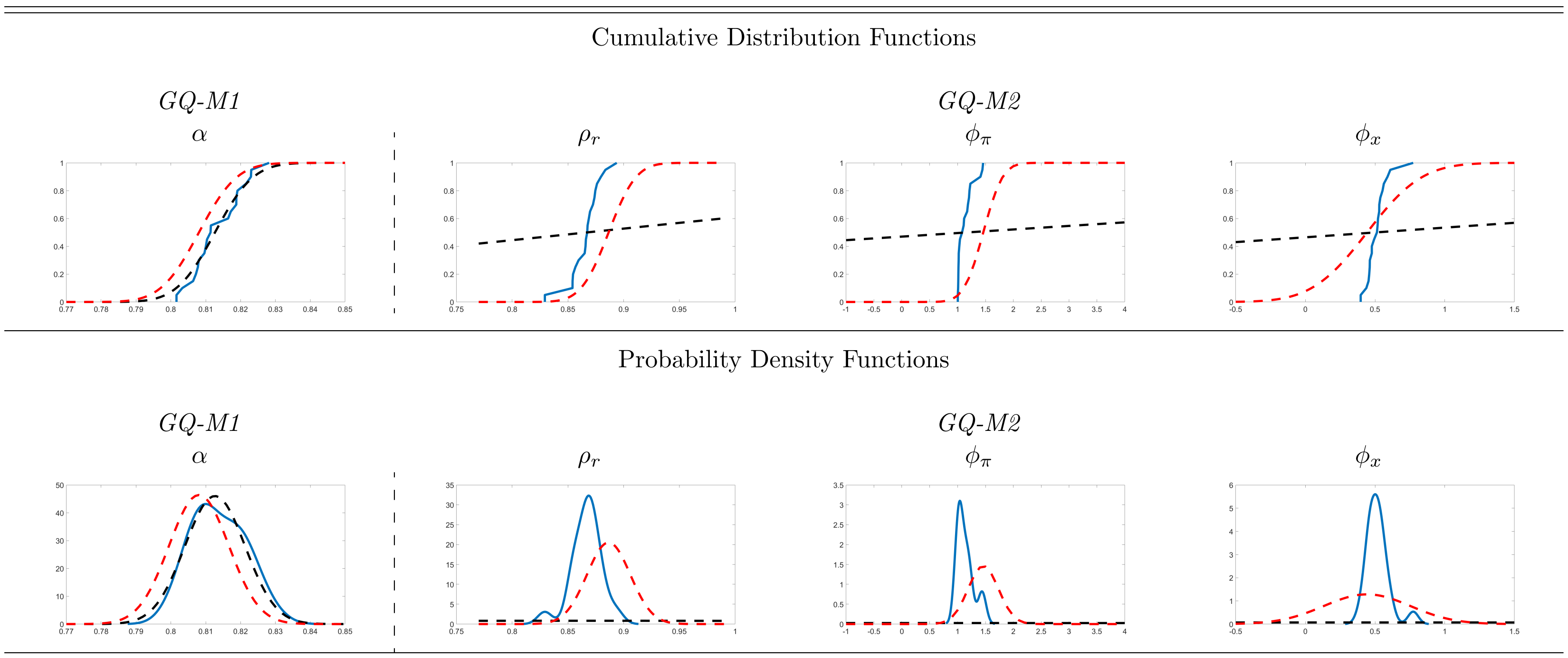

FIGURE 2: *CDFs (top panel) and PDFs (bottom panel) of the bootstrap realizations of the QML estimators used in the normality tests in Table 2, contrasted with the Gaussian. Blue lines indicate the empirical CDFs and the empirical PDFs. Red dashed lines report the CDFs and the PDFs of the Gaussian* $\mathcal{N}(\hat{\theta}_T, s(\hat{\theta}_T))$. *Black dashed lines plot the CDFs and the PDFs of the Gaussian* $\mathcal{N}(\overline{\hat{\theta}}^*_T, s(\hat{\theta}^*_T))$. *The values* $\hat{\theta}_T, s(\hat{\theta}_T), \overline{\hat{\theta}}^*_T$, *and* $s(\hat{\theta}^*_T)$ *are reported in Table 3.*

SUPPLEMENT TO 'BOOTSTRAP INFERENCE AND DIAGNOSTICS IN STATE SPACE MODELS: WITH APPLICATIONS TO DYNAMIC MACRO MODELS'

Giovanni Angelini[a], Giuseppe Cavaliere[a,b*]
and Luca Fanelli[a]

First draft: March 2020. First revision: November 2020; This version: April 2021.

S.1 Content and structure

This supplement material to Angelini, Cavaliere and Fanelli (2020), henceforth ACF, provides: (i) a short preliminary section which introduces the notation used throughout (Section S.2); (ii) the proofs of our main results and auxiliary lemmas (Section S.3), (iii) further Monte Carlo results based on the ARMA(1,1) model (Section S.4), (iv) further empirical results (Section S.5), (v) a summary of computation time (Section S.6).

S.2 Notation

With P denoting the probability measure for the data, we use $E(\cdot)$ and $Var(\cdot)$ to denote expectations and variance computed under P, respectively. We use P^* to denote the probability measure induced by the bootstrap, i.e. conditional on the original sample. Expectation and variance computed under P^* are denoted by $E^*(\cdot)$ and $\mathrm{Var}^*(\cdot)$, respectively.

Define, for $\delta > 0$, $p_T^*(\delta) := P^*(||\hat{\theta}_T^* - \hat{\theta}_T|| > \delta)$, where $\hat{\theta}_T^*$ is the bootstrap analog of the QML estimator $\hat{\theta}_T$, and $\|\cdot\|$ is the Euclidean norm. With the notation '$\hat{\theta}_T^* - \hat{\theta}_T \xrightarrow{p^*}_p 0$', which reads '$\hat{\theta}_T^* - \hat{\theta}_T$ convergences in p*-probability to 0, in probability', we mean that the (stochastic) sequence $\{p_T^*(\delta)\}$ converges in probability to zero ($p_T^*(\delta) \xrightarrow{p} 0$) for any δ .

[a]Department of Economics, University of Bologna, Italy. [b] Department of Economics, University of Exter Business School, UK.

Likewise, for $\delta > 0$, define $s_T^*(\delta) := P^*(\sup_{\theta\in\Theta} |\hat{Q}_T^*(\theta) - Q_0(\theta)| > \delta)$, where $\hat{Q}_T(\theta)$ is a criterion functions defined on the set Θ and $\hat{Q}_T^*(\theta)$ is its bootstrap analog. With '$\hat{Q}_T^*(\theta) - \hat{Q}_T(\theta)$ converges uniformly in p*-probability to 0, in probability' we mean that the stochastic sequence $\{s_T^*(\delta)\}$ converges to zero in probability ($s_T^*(\delta) \xrightarrow{p} 0$) for any$\delta$, i.e. that $\sup_{\theta\in\Theta} |\hat{Q}_T^*(\theta) - \hat{Q}_T(\theta)| \xrightarrow{p^*}_p 0$.

Finally, consider a random variable X, with associated CDF denoted $G_X(x) := P(X \leq x)$, and let $\{X_T^*\}$ be a sequence of bootstrap counterparts of X, with associated CDF (conditional on the data) $G^*_{X_T^*}(x) := P^*(X_T^* \leq x)$. We say that X_T^* 'converges in conditional distribution to X, in probability', denoted by '$X_T^* \xrightarrow{d^*}_p X$', if $G^*_{X_T^*}(x) \to_p G_X(x)$ at all continuity points of G_X. **[let** $G_T^*(x) := P^*(R_T^* \leq x) = P(R_T^* \leq x \mid \textbf{data})$**]**

S.3 Proofs

S.3.1 Proof of Proposition 1

Preliminaries. Consider the linear filter $H(z,\theta) := \sum_{j=0}^{\infty} h_j(\theta)z^j$ which defines the VMA(∞) representation for y_t in terms of ϵ_t under the stated assumptions. Let $\partial h(z,\theta)/\partial\theta$ be a shortcut for the derivative of the vector field $h(z,\cdot)$: $\mathbb{R}^{n_\theta} \to \mathbb{R}^{n_y}$. Then, for all $\theta \in \Theta$ and all z in the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ it holds that $h(z,\theta) = \sum_{j=0}^{\infty} h_j(\theta)z^j$ is bounded and bounded away from zero and $h_0(\theta) = 1$. Moreover, the function $\psi(z,\theta) := \sum_{j=0}^{\infty} \psi_j(\theta)z^j = H(z,\theta)^{-1}$ is well-defined by its power series expansion for $|z| \leq 1 + \epsilon$ for some $\epsilon > 0$, and is also bounded and bounded away from zero on the complex unit disk and $\psi_0(\theta) = 1$. The coefficients $h_j(\theta)$, $\psi_j(\theta)$, $\dot{h}_j(\theta) := \partial h_j(\theta)/\partial\theta$, and $\dot{\psi}_j(\theta) := \partial\psi_j(\theta)/\partial\theta$ are exponentially decaying, and satisfy

$$\begin{aligned} |a_j(\psi)| &= O(j^{-2-\zeta}),\ |b_j(\psi)| = O(j^{-2-\zeta}) \\ ||\dot{a}_j(\psi)|| &= O(j^{-2-\zeta}),\ ||\dot{b}_j(\psi)|| = O(j^{-2-\zeta}) \end{aligned}$$

for all $\zeta > 0$, uniformly in $\theta \in \Theta$; see Zygmund (2003, pp. 46 and 71). In the following, we will also assume that (i) $h(e^{\mathrm{i}\lambda},\theta)$ is twice differentiable in λ with second derivative in $Lip(\zeta)$ for $\zeta > 0$ and (ii) the function $\dot{h}(z,\theta) := \frac{\partial h(z,\theta)}{\partial\theta} = \sum_{j=0}^{\infty} \dot{h}_j(\theta)z^j$ exists and $\dot{h}(e^{\mathrm{i}\lambda},\theta)$ is differentiable in λ with derivative in $Lip(\zeta)$ for $\zeta > 0$.

Finally, notice that for all $\theta \in N_\delta(\theta_0)$ it holds that $H(z,\theta) \neq H(z,\theta_0)$ on a subset of $\{z \in \mathbb{C} : |z| = 1\}$ of positive Lebesgue measure.

Part (i) (Consistency). According to Theorem 2.1 in Newey and McFadden

(1994) (see also Theorem A1 in Wooldridge, 1994), if there are functions $Q_0(\theta)$ and $\hat{Q}_T(\theta)$ defined in the parameter space Θ such that: (dd.1) Θ is compact; (dd.2) θ_0 is the unique maximizer of $Q_0(\theta)$ in Θ; (dd.3) $Q_0(\theta)$ is continuous in θ; (dd.4) $\hat{Q}_T(\theta)$ converges uniformly in probability to $Q_0(\theta)$ in Θ, then

$$\arg\max_{\theta\in\Theta} \hat{Q}_T(\theta) =: \hat{\theta}_T \to_p \theta_0 := \arg\max_{\theta\in\Theta} Q_0(\theta).$$

We show how to verify that these conditions hold in a neighborhood $N_\delta(\theta_0)$. Notice that $\hat{Q}_T(\theta) := T^{-1}\sum_{t=1}^T \ell_t(\theta)$, with $\ell_t(\theta)$ given in (11) of ACF, and

$$Q_0(\theta) := E(\hat{Q}_T(\theta)) = E(T^{-1}\sum\nolimits_{t=1}^T \ell_t(\theta)) = T^{-1}\sum\nolimits_{t=1}^T E(\ell_t(\theta)) = E(\ell_t(\theta)),$$

where the last equality holds because of the weakly stationarity and ergodicity of $\{y_t\}$ in Assumption A1. First, dd.1 holds by assumption. Second, in case θ_0 is globally identified, dd.2 follows from Assumptions A2-A3, while under local identification dd.2 holds in the neighborhood $N_\delta(\theta_0)$. In the latter case, θ_0 is a local unique maximizer (cf. Definition 3 in Qu and Tkachenko, 2012) of $Q_0(\theta)$. Third, dd.3 follows from Assumption A2 and the postulated normal distribution used to construct the QML estimator.

Finally, to verify dd.4 observe first that pointwise convergence of $\hat{Q}_T(\theta)$ to $Q_0(\theta)$ holds for any $\theta \in \Theta$ as discussed e.g. in Stoffer and Wall (1991). This result can be strengthened to uniform convergence in probability by showing that $Q_T(\theta)$ is stochastically equicontinuous. From Newey (1991, Corollary 2.2), this holds if the derivative of $Q_T(\theta)$ is dominated uniformly in θ by a random variable $U_T = O_p(1)$. To prove this, first notice that, as in Watson (1989, p.79), under Assumption A1 we can neglect the initial values and define Q_T as the average (log-)likelihood associated with the steady state solution to the model, see (6)-(7) in ACF; that is,

$$Q_T(\theta) := -\frac{1}{2T}\sum_{t=1}^T \log|\Sigma_\epsilon(\theta)| - \frac{1}{2T}\sum_{t=1}^T \epsilon_t(\theta)' \Sigma_\epsilon(\theta)^{-1}\epsilon_t(\theta).$$

Then, the i-th element of the score (see Watson, 1989) is given by

$$\frac{\partial Q_T(\theta)}{\partial\theta_i} = \frac{1}{T}\sum_{t=1}^T L1_{i,t} + \frac{1}{T}\sum_{t=1}^T L2_{i,t}, \tag{S.1}$$

where

$$L1_{i,t} := -\frac{1}{2} tr\{M_i(\theta)(I_{n_y} - \Sigma_\epsilon(\theta)^{-1}\epsilon_t(\theta)\epsilon_t(\theta)')\},$$

$$M_i(\theta) := \Sigma_\epsilon(\theta)^{-1}\,\partial\Sigma_\epsilon(\theta)/\partial\theta_i$$
$$L2_{i,t} := -\left(\frac{\partial\epsilon_t(\theta)}{\partial\theta_i}\right)'\Sigma_\epsilon(\theta)^{-1}\epsilon_t(\theta).$$

To see that (S.1) is bounded uniformly in θ, notice that

$$\begin{aligned}\epsilon_t(\theta) &= y_t - \sum_{j=1}^{\infty}\psi_i(\theta)\,y_{t-1} = \epsilon_t + \sum_{j=1}^{\infty}(\psi_i - \psi_i(\theta))\,y_{t-1} \\ \frac{\partial\epsilon_t(\theta)}{\partial\theta} &= -\sum_{j=1}^{\infty}\dot{\psi}_j(\theta)\,y_{t-1}\end{aligned}$$

with $\sup_\theta \sum_{j=1}^{\infty}|\psi_i - \psi_i(\theta)| < \infty$. Then, simple algebra as in Lemma B.3 of Cavaliere, Nielsen and Taylor (2017) shows that this fact, together with $\inf_\theta|\Sigma_\epsilon(\theta)| > 0$ and the moment Assumption A4, implies $E|\sup_\theta \partial Q_T(\theta)/\partial\theta| < \infty$. The desired result is thus obtained.

Part (ii) (Asymptotic normality). We now refer to Theorem 3.1 in Newey and McFadden (1994) (see also Theorem A2 in Wooldridge, 1994), which states that if there are functions $Q_0(\theta)$ and $\hat{Q}_T(\theta)$ as defined before and such that $\hat{\theta}_T \to_p \theta_0$, and if: (dd.5) $\theta_0 \in$int(Θ), the interior of Θ; (dd.6) $\hat{Q}_T(\theta)$ is twice continuously differentiable in a neighborhood $\mathcal{N}_{\theta_0}$ of θ_0; (dd.7) $T^{1/2}\nabla_\theta\hat{Q}_T(\theta_0) \to_d N(0,V)$, with V nonsingular; (dd.8) there is $\Psi(\theta)$ that is continuous at θ_0 such that $\sup_{\theta\in\mathcal{N}_{\theta_0}}\left\|\nabla^2_{\theta\theta}\hat{Q}_T(\theta) - \Psi(\theta)\right\| \xrightarrow{p} 0$; (dd.9) $\Psi := \Psi(\theta_0)$ is nonsingular; then

$$T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N\left(0_{n_\theta\times 1}, \Psi^{-1}V\Psi^{-1}\right).$$

In our case, (dd.5) is assumed and (dd.6) follows from the postulated normal distribution for the innovation errors. (dd.7) holds under Assumption A5 with $V = \mathcal{B}_0 := \lim_{T\to\infty} Var(T^{-1/2}\sum_{t=1}^{T}\nabla_\theta\ell_t(\theta_0))$, see Stoffer and Wall (1991). Consider now (dd.8). The second derivative $\nabla^2_{\theta\theta}\hat{Q}_T(\theta)$ is tight (stochastically equicontinuous) by Newey (1991, Corollary 2.2) if its derivative is dominated uniformly in a neighborhood of θ_0, i.e. for $\theta \in \mathcal{N}_\delta(\theta_0)$, by a random variable $U_T = O_p(1)$. Again, this condition can be verified under Assumption A4 as e.g. in Lemma B.3 of Cavaliere, Nielsen and Taylor (2017). Tightness, together with the result $\hat{\theta}_T - \theta_0 \xrightarrow{p} 0$ from part (i), implies that the second derivative can be evaluated at the true value only, see Lemma A.3 of Johansen and Nielsen (2010). This evaluation is done in Stoffer and Wall (1991), where it is shown that $\left.\nabla^2_{\theta\theta}\hat{Q}_T(\theta)\right|_{\theta=\theta_0} - \Psi(\theta_0) \xrightarrow{p} 0$ for $\Psi(\theta_0) = \Psi = \mathcal{A}_0$, where $\mathcal{A}_0 := \mathcal{A}_0(\theta_0)$

is as stated in the Proposition. Finally, (dd.9) follows from Assumption A3. This proves part (ii). ■

S.3.2 Proof of Proposition 2

Let $W_T^* := T^{1/2}(\hat{\theta}_T^* - \hat{\theta}_T)$ be as in Section 3 and let $\hat{Q}_T^*(\theta) := T^{-1}L_T^*(\theta) = T^{-1}\sum_{t=1}^T \ell_t^*(\theta)$. Moreover, let the variance (conditional on the original sample) of the bootstrap score be denoted by $\mathcal{B}_T^*(\hat{\theta}_T) := \mathrm{Var}^*(T^{-1/2}\sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T))$.

We first provide a lemma which characterizes the properties of the bootstrap log-likelihood function and its derivatives.

Lemma S.1 *Under Assumptions A1-A3, A4' and A5:*
(i) with $Q_0^*(\theta) := E^*(\hat{Q}_T^*(\theta))$*, it holds that* $\nabla_\theta^{(h)} Q_0^*(\theta) = \nabla_\theta^{(h)} \hat{Q}_T(\theta) + o_p\left(T^{-1/2}\right)$ *for all* $\theta \in \Theta$*, and* $h = 0, 1, 2$*;*

(ii) $\mathcal{B}_T^*(\hat{\theta}_T) - \mathcal{B}_T(\hat{\theta}_T) \xrightarrow{p^*}_p 0_{n_\theta \times n_\theta}$*;*

(iii) $T^{1/2}\nabla_\theta \hat{Q}_T^*(\hat{\theta}_T) \xrightarrow{d^*}_p N\left(0_{n_\theta \times 1}, \mathcal{B}^*\right)$*,* $\mathcal{B}^* := plim_{T\to\infty}\mathcal{B}_T^*(\hat{\theta}_T)$*.*

Proof. The proof of part (i) mimics the proof of Lemma 1 in Stoffer and Wall (1991). The $o_p(1)$ term in (i) is erroneously missing in Lemma 1 of Stoffer and Wall (1991) and is due to the fact that the initial value $\hat{F}_{1|0}^* = \hat{F}_{1|0}$ is not necessarily zero.

Part (ii) follows from Lemma 2 in Stoffer and Wall (1991).

To prove part (iii), we first show that the bootstrap score, evaluated at the bootstrap pseudo-true parameter $\hat{\theta}_T$ – that is, $T^{-1/2}\nabla_\theta L_T^*(\hat{\theta}_T) = T^{-1/2}\sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T)$ – satisfies a CLT. As for the proof of Proposition 1, we can neglect the initial values and define L_T^* as the (log-)likelihood associated with the steady state solution to the state space model. At the bootstrap pseudo-true values, the i-th element of the associated score can be written as (see also Stoffer and Wall, 1991, proofs of Lemmas 1 and 2)

$$\left.\frac{\partial L_T^*(\theta)}{\partial \theta_i}\right|_{\theta=\hat{\theta}_T} = \sum_{t=1}^T \left.\frac{\partial \ell_t^*(\theta)}{\partial \theta_i}\right|_{\theta=\hat{\theta}_T} = \sum_{t=1}^T L1_{i,t}^* + \sum_{t=1}^T L2_{i,t}^*,$$

where

$$L1_{i,t}^* := -\frac{1}{2}tr\{M_i(\hat{\theta}_T)(I_{n_y} - \Sigma_\epsilon(\hat{\theta}_T)^{-1}\epsilon_t^*(\hat{\theta}_T)\epsilon_t^*(\hat{\theta}_T)')\}$$

with $M_i(\theta) := \Sigma_\epsilon(\theta)^{-1}\partial\Sigma_\epsilon(\theta)/\partial\theta_i$, and

$$L2_{i,t}^* := -\left(\left.\frac{\partial \epsilon_t^*(\theta)}{\partial \theta_i}\right|_{\theta=\hat{\theta}_T}\right)' \Sigma_\epsilon(\hat{\theta}_T)^{-1}\epsilon_t^*(\hat{\theta}_T),$$

$i = 1, ..., n_\theta$. Since $\epsilon_t^*(\hat{\theta}_T) = \Sigma_\epsilon(\hat{\theta}_T)^{1/2} e_t^*$, where e_t^* is (conditionally on the original data) i.i.d., the vector $\mathcal{L}_t^* := (L1_{1,t}^*, ..., L1_{n_\theta,t}^*, L2_{1,t}^*, ..., L2_{n_\theta,t}^*)'$ is a Martingale Difference Array (MDA). We can therefore make use of a bootstrap CLT for MDAs, see e.g. Gonçalves and Kilian (2004, Theorem A.1, and the proof of their Lemma A.3). To do so, it suffices to prove that (a) the variance of $T^{-1/2}\sum_{t=1}^T \nabla_\theta \ell_t^*(\hat{\theta}_T)$ (conditional on the original sample), $\mathcal{B}_T^*(\hat{\theta}_T)$, converge in probability and that (b) moments of higher order exist, see below. Regarding (a), let $\hat{\mathcal{B}}_T(\hat{\theta}_T) := T^{-1}\sum_{t=1}^T \nabla_\theta \ell_t(\hat{\theta}_T)\nabla_\theta \ell_t(\hat{\theta}_T)'$ which, under Assumptions A1–A4 is such that $\hat{\mathcal{B}}_T(\hat{\theta}_T) \to_p \mathcal{B}_0$. This result, together with part (ii) of the lemma, implies that $\mathcal{B}_T^*(\hat{\theta}_T) \to_p \mathcal{B}^*$, as requested. For part (b), Theorem A.1 in Gonçalves and Kilian (2004) and a standard application of the Cramér-Wold device applied to $\mathcal{L}_t^*$ require that, for all $\lambda \in \mathbb{R}^{n_\theta}$ and some $r > 0$, $T^{-r}\sum_{t=1}^T E^*|\sum_{i=1}^{n_\theta}\lambda_i(L1_{i,t}^* + L2_{i,t}^*)|^{2r} \to_p 0$. Taking $r = 2$, by the c_r inequality we have (with K_c denoting a generic constant)

$$T^{-2}\sum_{t=1}^T E^*\left|\sum_{i=1}^{n_\theta}\lambda_i(L1_{i,t}^* + L2_{i,t}^*)\right|^4 \leq K_c T^{-2}\sum_{t=1}^T\sum_{i=1}^{n_\theta}(E^*\left|L1_{i,t}^*\right|^4 + E^*\left|L2_{i,t}^*\right|^4) \quad \text{(S.2)}$$

Consider $L1_{i,t}^*$ first. Since

$$tr\{M_i(\hat{\theta}_T)(I_{n_y} - \Sigma_\epsilon(\hat{\theta}_T)^{-1}\epsilon_t^*(\hat{\theta}_T)\epsilon_t^*(\hat{\theta}_T)')\} =$$

$$= tr\{M_i(\hat{\theta}_T)(I_{n_y} - \Sigma_\epsilon(\hat{\theta}_T)^{-1}\Sigma_\epsilon(\hat{\theta}_T)^{1/2}e_t^* e_t^{*\prime}\Sigma_\epsilon(\hat{\theta}_T)^{1/2})\}$$
$$= tr\{\Sigma_\epsilon(\hat{\theta}_T)^{1/2}M_i(\hat{\theta}_T)\Sigma_\epsilon(\hat{\theta}_T)^{-1/2}(I_{n_y} - e_t^* e_t^{*\prime})\}$$

and, under the stated assumption, $\sup_\theta ||\Sigma_\epsilon(\theta)^{1/2}M_i(\theta)\Sigma_\epsilon(\theta)^{-1/2}|| \leq K_c < \infty$, we have that, for all i and t,

$$E^*\left|L1_{i,t}^*\right|^4 \leq K_c E^*\left|\sum_{k,k'=1}^{n_y}|1 - e_{kt}^* e_{k't}^*|\right|^4 \leq K_c \sum_{k,k'=1}^{n_y} E^*(1 - e_{kt}^* e_{k't}^*)^4 =$$
$$= K_c \sum_{k,k'=1}^{n_y}\frac{1}{T}\sum_{t=1}^T(\hat{\epsilon}_{kt}^c\hat{\epsilon}_{k't}^c - 1)^4 = O_p(1)$$

under the assumption of 8^{th} order moments, see also Gonçalves and Kilian (2004). Consider now $L2_{i,t}^*$. As in Watson (1989, pp. 87-88), up to an $o_p^*(1)$

term (in probability) we can write

$$L2_{i,t}^{*} = \left(\sum_{j=1}^{t-1} M_{i,j}(\hat{\theta}_T)\epsilon_{t-j}^{*}(\hat{\theta}_T) \right)' \Sigma_{\epsilon}(\theta)^{-1} \epsilon_t^{*}(\hat{\theta}_T),$$

which, using the fact that $\sup_{\theta} \sum_{j=1}^{\infty} |M_{i,j}(\theta)| < \infty$, is of $O_p^{*}(1)$ under the assumption of finite 8^{th} order moments. This finally implies that (S.2) is of $o_p^{*}(1)$, in probability, hence proving (iii). ■

We can now turn to the proof of our main Proposition 2. The proof mimics the one given for the original statistic in Proposition 1 using the additional results provided in Lemma S.1.

For the consistency part, we need to show that the condition given in Theorem 2.1 of Newey and McFadden (1994) holds conditionally on the original data with probability tending to one in a neighborhood $N_{\delta}(\hat{\theta}_T)$ of the bootstrap true value $\hat{\theta}_T$, see the proof of Proposition 1. Conditions dd.1 and dd.3 are obviously satisfied. Condition dd.2 holds since, by Lemma S.1(i), $E^{*}\hat{Q}_T^{*}(\theta) = \hat{Q}_T(\theta) + o_p\left(T^{-1/2}\right)$, which for T large enough has the unique maximizer $\hat{\theta}_T$.

We now focus on dd.4. Again, pointwise convergence of $\hat{Q}_T^{*}(\theta)$ to $\hat{Q}_T(\theta)$ for any $\theta \in \Theta$ holds as discussed e.g. in Stoffer and Wall (1991) and stochastic equicontinuity can be shown by proving that the derivative of $\hat{Q}_T^{*}(\theta)$ is dominated uniformly in θ by a random variable B_T^{*} which is of order $O_p^{*}(1)$, in probability. To prove this we can proceed as in the proof of Lemma S.1(iii) and evaluate the average bootstrap (log-)likelihood associated to the steady state solution to the model, which lead to the equation

$$\frac{\partial L_T^{*}(\theta)}{\partial \theta_i} = \sum_{t=1}^{T} \frac{\partial \ell_t^{*}(\theta)}{\partial \theta_i} = \sum_{t=1}^{T} L1_{i,t}^{*}(\theta) + \sum_{t=1}^{T} L2_{i,t}^{*}(\theta),$$

with

$$\begin{aligned} L1_{i,t}^{*}(\theta) &:= -\frac{1}{2} tr\{M_i(\theta)(I_{n_y} - \Sigma_{\epsilon}(\theta)^{-1}\epsilon_t^{*}(\theta)\epsilon_t^{*}(\theta)')\}, \\ M_i(\theta) &:= \Sigma_{\epsilon}(\theta)^{-1} \partial \Sigma_{\epsilon}(\theta) / \partial \theta_i \\ L2_{i,t}^{*}(\theta) &:= -\left(\frac{\partial \epsilon_t^{*}(\theta)}{\partial \theta_i} \right)' \Sigma_{\epsilon}(\theta)^{-1} \epsilon_t^{*}(\theta). \end{aligned}$$

Here for $\hat{\psi}_i := \psi_i(\hat{\theta}_T)$, we have

$$\begin{aligned}
\epsilon_t^*(\theta) &= y_t^* - \sum_{j=1}^{\infty} \psi_i(\theta)\, y_{t-1}^* = \epsilon_t^* + \sum_{j=1}^{\infty} \left(\hat{\psi}_i - \psi_i(\theta)\right) y_{t-1}^* \\
\frac{\partial \epsilon_t^*(\theta)}{\partial \theta} &= -\sum_{j=1}^{\infty} \dot{\psi}_j(\theta)\, y_{t-1}^*
\end{aligned}$$

with $\sup_\theta \sum_{j=1}^{\infty} |\hat{\psi}_i - \psi_i(\theta)| < \infty$. Again, as in the proof of Lemma B.3 of Cavaliere, Nielsen and Taylor (2017), we can prove that this fact, together with $\inf_\theta |\Sigma_\epsilon(\theta)| > 0$ and the 8^{th} order moment Assumption A4 implies that $E^*|\sup_\theta \partial \hat{Q}_T^*(\theta)/\partial\theta| < \infty$, as required.

For the asymptotic normality and the consistency of the bootstrap standard errors we refer again to Theorem 3.1 in Newey and McFadden (1994), whose conditions dd.5, dd.6 and dd.9 trivially hold while dd.7 holds in probability as T diverges with $V = \mathcal{B}^* := \lim_{T\to\infty} \mathrm{Var}^*(T^{-1/2} \sum_{t=1}^{T} \nabla_\theta \ell_t^*(\hat{\theta}_T))$ as demonstrated in Stoffer and Wall (1991). For dd.8, it is sufficient to show that the second derivative $\nabla_{\theta\theta}^2 \hat{Q}_T^*(\theta)$ is tight if its derivative is dominated uniformly in a neighborhood of $\hat{\theta}_T$, i.e. for $\theta \in \mathcal{N}_\delta(\hat{\theta}_T)$, by a random variable $B_T^* = O_p^*(1)$, in probability. Again, this condition can be verified under Assumption A4 as done for the non-bootstrap proposition. The rest of the proof mimics the one given for Proposition 1. ■

S.3.3 Proof of Proposition 3

Consider the decomposition

$$d_{T,B}(x) = (\kappa_T^*)^{1/2} S_{T,B}^* + B^{1/2} \hat{m}_T,$$

with

$$\begin{aligned}
\kappa_T^* &:= \frac{G_T^*(x)\,(1 - G_T^*(x))}{\Phi_Z(x)(1 - \Phi_Z(x))} \\
S_{T,B}^* &:= \frac{1}{B^{1/2}} \sum_{b=1}^{B} Z_{T:b}^*,\ Z_{T:b}^* := \frac{\mathbb{I}(R_{T:b}^* \leq x) - G_T^*(x)}{(G_T^*(x)\,(1 - G_T^*(x))^{1/2}}
\end{aligned}$$

and

$$\hat{m}_T := \hat{V}_T(x)^{-1/2}(G_T^*(x) - \Phi_Z(x)),$$

see (24) in ACF. Under (26) in ACF, $\kappa_T^* \to_p 1$ as $T \to \infty$; moreover, $\hat{m}_T = O_p(T^{-\alpha})$ and hence $B^{1/2}\hat{m}_T$ is of order $O_p\left(B^{1/2}T^{-\alpha}\right)$, which converges in probability to zero under the assumption in (27) in ACF. Finally, consider $S_{T,B}^*$.

Conditional on the original sample, the $Z^*_{T:b}$'s are i.i.d. with zero mean, unit variance (provided T is sufficiently large, such that $G^*_T(x)(1 - G^*_T(x)) > 0$) and a.s. bounded third order moment. Hence, by the Berry-Esseen bound, for some constant c,

$$\sup_{z\in\mathbb{R}} \left| P^* \left(\frac{1}{B^{1/2}} \sum_{b=1}^{B} Z^*_{T:b} \leq z \right) - \Phi_Z(z) \right| \leq cB^{-1/2}$$

This implies that, for T large enough, $S^*_{T,B} \xrightarrow{d^*}_p N(0,1)$ for $B \to \infty$, which completes the proof. ■

S.4 Further Monte Carlo results

In this section we provide additional Monte Carlo results other the ones reported in Section 5 of ACF. Section S.4.1 investigates the performance of the suggested bootstrap approach by considering the state space form associated with an ARMA(1,1) model. Relative to the DSGE model analyzed in ACF, with the ARMA(1,1) we can easily control the strength of identification of the associated state space model.

S.4.1 The ARMA(1,1) model

The ARMA(1,1) model reads as an interesting case study for our bootstrap approach because it is particularly suited to characterize the case of weakly and strongly identified parameters.

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

$$y_t = (\pi + \beta)y_{t-1} + \omega_t - \pi\omega_{t-1} \quad , \quad \omega_t \sim iidN(0,1), \quad t = 1, ..., T \tag{S.3}$$

where y_0 and ω_0 are given, and the vector of parameters is $\phi := (\phi_1, \phi_2)' = (\pi, \pi + \beta)'$.[1] With this parameterization, $\beta = \phi_2 - \phi_1$ can be interpreted as the difference between the autoregressive and moving average parameters. It is seen that in the special case where $\beta = 0$ ($\phi_2 = \phi_1$), the model collapses to

$$y_t = \omega_t$$

and the moving average parameter π is unidentified. More precisely, Assumptions A2-A3 in ACF are violated and the ML estimator of $\phi_1 = \pi$ is not consistent and is not asymptotically Gaussian. Indeed, when β gets close to zero

[1] In this section we consider Gaussian ω_ts to simplify computations. Results obtained with non-normal ω_ts do not change significantly and are available upon request to the authors.

the likelihood function becomes flat in the direction of π and the identification of this parameter π from the data deteriorates.[2]

For $\theta := (\theta_1, \theta_2)' = (\phi_1, \phi_2 - \phi_1)' = (\pi, \beta)'$, the associated (minimal) state-space representation of the ARMA(1,1) model in (S.3) is given by

$$Z_t = \underbrace{\begin{pmatrix} \pi+\beta & 1 \\ 0 & 0 \end{pmatrix}}_{A(\theta)} Z_{t-1} + \underbrace{\begin{pmatrix} 1 \\ -\pi \end{pmatrix}}_{B(\theta)} \omega_t \tag{S.4}$$

$$y_t = (1, 0)\ Z_t \tag{S.5}$$

where the parameters satisfy the conditions $-1 < \pi < 1$, $-1 < \pi + \beta < 1$, which ensure stationarity and invertibility. The alternative representations discussed in Section 2 of ACF can be easily obtained from (S.4)-(S.5).

We generate $M = 2000$ samples of length $T = 100, 500$ from this model and, as in Andrews and Cheng (2012), we select the parameter $\theta_{2,0} = \beta_0$ from the set $\{-0.76, 0.5/T^{1/2}\}$. The data generating process for the case of strongly identified parameters is obtained with $\beta_0(= \theta_{2,0}) = -0.76$ and $\pi_0(= \theta_{1,0}) = 0.40$, and is denoted 'ARMA-DGP1' (see Tables SM.1-SM.2); the data generating process corresponding to the case of weakly identified parameters is obtained with the local-to-zero embedding $\beta_0(= \theta_{2,0}) = 0.5/T^{1/2}$ (keeping $\pi_0(= \theta_{1,0}) = 0.40$), and is denoted 'ARMA-DGP2' (see Tables SM.1-SM.2). ARMA-DGP2 features a 'near cancelling roots' scenario in the sense that the AR and MA roots tend to coincide (and cancel) as T increases. Notably, under this setup only the MA parameter π is weakly identified (unidentified asymptotically), while the parameter β is strongly identified. Andrews and Cheng (2012) prove formally that the ML estimator of π is not consistent in ARMA-DGP2.

The log-likelihood of the model is maximized by combining the Kalman filter with the 'BFGS' method by imposing that the optimization parameter spaces for the MA and AR coefficients are constrained to $[-0.90, 0.90]$ and $[-0.90, 0.90]$, respectively. Bootstrap estimation follows the algorithm described in ACF. The Steps 2-4 of the bootstrap algorithm are repeated $N := 499$ times. Bootstrap confidence intervals are computed as explained in Remark 3.3 in ACF.

We consider computationally straightforward versions of our test of model misspecification. The tests are designed to verify the asymptotic normality of

[2] Andrews and Mikusheva (2015) show that if the parameter β is defined through the embedding $\beta_T = C/T^{1/2}$ for some constant C, then suitably normalized versions of the measures of information $\mathcal{B}_{0,T}$ and $\mathcal{A}_{0,T}$ of the state space model (see Section 3 of ACF) are no longer interchangeable measures of information even if White's (1982) information matrix equality is still valid.

the sequences $\{\hat{\theta}^*_{T:1}, ..., \hat{\theta}^*_{T:B}\}$ of the ML bootstrap estimator, where $B < N$ is selected by using the practical rules suggested used in ACF. In particular, we set $B = \frac{1}{3}T^{4/5}$ and $B = \frac{1}{2}T^{4/5}$, respectively. The computed normality tests are the same as in the Monte Carlo experiments in ACF and are run at the 5% nominal significance level. Results for the two DGPs are discussed separately.

S.4.1.1 ARMA-DGP1: Strong identification

From Table SM.1 we observe that the ML estimates of π and β are substantially similar to their bootstrap counterparts and tend to converge to their true population values as T increases. Hessian-based standard errors associated with $\hat{\pi}_T$ and $\hat{\beta}_T$ tend to be similar to the bootstrap standard errors. The coverage of the 90% bootstrap confidence intervals for the two parameters converges to the nominal 90% level as T increases and, in general, do not perform worse than the corresponding asymptotic 90% confidence intervals. The tests of model misspecification in Table SM.2 display rejection frequencies that approach the nominal 5% level as T increases.

The left-panel of Figure SM.1 reports the fan chart of the empirical cumulative density functions (CDFs) of the bootstrap realizations of the ML estimator of $\theta := (\pi, \beta)'$ used in the normality tests (see Cavaliere and Georgiev, 2020). The graphs clearly show that the bootstrap distributions converge to the Gaussian as T increases (and $B/T \to 0$).

Overall, the results in Tables SM.1-SM.2 and in Figure SM.1 show that the bootstrap works in the expected direction in ARMA models with strongly identified parameters. Bootstrap and asymptotic inference are highly reliable in this model.

S.4.1.2 ARMA-DGP2: Weak identification

From Table SM.1 we observe that, regardless of the sample size, the ML estimator of the MA parameter $\pi(=\theta_1)$ and its bootstrap analog deviate markedly from the true parameter value, while the mismatch between Hessian-based and bootstrap standard errors seems to increase with T. Instead, the ML estimator of $\beta(=\theta_2)$ and its bootstrap analog tend to converge to the true population value as T increases, consistently with Andrews and Mikusheva's (2015) findings. Interestingly, for this state space model, the coverages of 90% bootstrap confidence intervals perform generally better than 90% asymptotic confidence intervals for both parameters.

The empirical rejection frequencies of the tests of model misspecification in Table SM.2 suggest that for large T the tests detects non-normality of $\hat{\theta}^*_T := (\hat{\theta}^*_{1,T}, \hat{\theta}^*_{1,T})'$ quite convincingly. As it should be the case, focusing on the univariate normality tests, departures from normality characterize the QML estimator $\hat{\pi}_T^*$ $(=\hat{\theta}_{1,T}^*)$, not the QML estimator $\hat{\beta}_T^*(=\hat{\theta}_{2,T}^*)$ (recall that β is strongly identified). Admittedly, for $T = 100$ the power of the test tends to be low.

The left-panel of Figure SM.1 plot the empirical CDFs of the B bootstrap realizations (across Monte Carlo simulations) of the ML estimator of $\theta := (\pi, \beta)'$ used in the tests of model misspecification. The graphs confirm what already observed in Table SM.2.

By combining all these evidences, we can conclude that in the estimated ARMA(1,1) model with 'near cancelling roots', our bootstrap diagnostic test tends to inform a practitioner that standard inference does not hold for the MA parameter π.

S.5 Further empirical results

In this section we turn on the monetary DSGE models estimated in Section 6 of ACF on the Great Moderation sample, and repeat our empirical analyses by estimating and testing the model on different samples. In particular, we estimate models GQ-M1 and GQ-M2 on the 'Great Inflation' sample, 1954Q3-1984Q1, and the full sample 1954Q3-2008Q3, respectively. We do so to check the reliability of the suggested bootstrap-based diagnostic tests in situations in which the estimated state space model might be misspecified along dimensions that do not necessarily affect the asymptotic normality of the QML estimator. On both samples estimation is carried out by imposing the determinacy condition.

Great Inflation sample

Assuming determinacy on the Great Moderation sample, a monetary DSGE model like the one estimated in Section 6 of ACF can be expected to be misspecified on the Great Inflation period 1954Q3-1984Q1 (especially if determinacy is imposed in estimation). This is so because of the omission of indeterminacy parameters unrelated to θ and, possibly, because of the omission of shocks unrelated to the fundamental shocks (e.g. Lubik and Schorfheide 2004; Castelnuovo and Fanelli, 2015, and references therein). Results, reported in Table SM.3 show that asymptotic normality is not rejected by the normality tests in model GQ-M1. This evidence is somehow expected because $\theta_1 := (\alpha)$ is a strongly identified parameter and the misspecification that possibly characterizes model GQ-M1 on the Great Inflation sample is expected to affect the consistency of the QML estimator $\hat{\theta}_{1,T} = (\hat{\alpha}_T)$, not its asymptotic normality.

Conversely, asymptotic normality is rejected in model GQ-M2 which involves the estimation of the monetary policy rule parameters $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$. In particular, for the parameter ϕ_π we observe a boundary estimation issue (due to the fact that we impose determinacy in estimation). Overall, empirical results on the Great Inflation period 1954Q3-1984Q1 confirm that our test solely captures misspecifications of the state space model that affect the asymptotic normality of the QML estimator.

Full sample

The estimation of the monetary DSGE model on the full sample 1954Q3-2008Q3 does not possibly account for a potential change in the conduct of monetary policy (e.g. Lubik and Schorfheide 2004; Castelnuovo and Fanelli, 2015, and references therein). In model GQ-M1, the fact that $\theta_1 := (\alpha)$ is a strongly identified parameter should not be altered by the occurrence of a break in the slope of the Phillips curve around the mid-eighties. Empirical results in Table SM.3 confirm this fact.

Conversely, in model GQ-M2, the QML estimates of $\theta_2 := (\rho_r, \phi_\pi, \phi_x)'$ obtained by ignoring a possible structural break in ϕ_π and ϕ_x around the mid-eighties should not improve their identifiability. Empirical results in Table SM.3 confirm that Gaussian asymptotic inference on ϕ_π and ϕ_x remains problematic also in the full sample where determinacy is imposed and an important structural break is not accounted for.

Overall, estimation and testing results discussed in this section confirm that our test tends to capture misspecifications of the state space model that affect the asymptotic normality of the QML estimator.

S.6 Computation time

Table SM.4 summarizes computation time (in seconds) employed to run the Monte Carlo experiments and the estimations on US quarterly data discussed both in ACF and in this supplement.

References

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

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

Angelini G., G. Cavaliere and L. Fanelli (2020), Bootstrap inference and diagnostics in state space models: with applications to dynamic macro models.

Cavaliere, G. and I. Georgiev (2020), Inference under random limits bootstrap measures, *Econometrica* 88, 2547-2574.

Cavaliere, G., Nielsen, M.Ø. and A.M.R. Taylor (2017), Quasi-maximum likelihood estimation and inference in fractional time series models with heteroskedasticity of unknown form, *Journal of Econometrics* 198, 165-188.

Castelnuovo, E. and L. Fanelli (2015), Monetary policy indeterminacy and identification failures in the U.S.: Results from a robust test, *Journal of Applied Econometrics* 30,857-1010.

Gonçalves, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89-120.

Johansen, S. and M. Ø. Nielsen (2010), Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics* 158, 51-66.

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

Newey, W.K. (1991), Uniform convergence in probability and stochastic equicontinuity, *Econometrica* 59, 1161-1167.

Newey, W.K and D. McFadden (1994), Large sample estimation and hypothesis testing, *Handbook of Econometrics*, R.F. Engle and D.L. McFadden (eds.), Vol. IV, Chap. 36

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

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

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

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

Wooldridge, J.M. (1994), Estimation and inference for dependent processes, *Handbook of Econometrics*, R.F. Engle and D.L. McFadden (eds.), Vol. IV, Chap. 45.

Zygmund, A. (2003), *Trigonometric series*, (third ed.), Cambridge University Press, Cambridge.

TABLE SM.1. Monte Carlo experiment, ARMA(1,1) model: estimates.

θ (a)	$\hat{\theta}_T$ (b)	$s(\hat{\theta}_T)$ (c)	90%-CI (d)	$\hat{\theta}^*_T$ (e)	$s(\hat{\theta}^*_T)$ (f)	90%-BSCI (g)	*Asymp.* (h)	Bootstrap *Stud.* (i)	*Perc.* (l)	*Basic* (j)
$T = 100$										
				ARMA-DGP1						
$\pi = 0.40$	0.41	0.14	$0.45, [0.19, 0.65]$	0.42	0.14	$0.48, [0.21, 0.71]$	0.87	0.84	0.88	0.86
$\beta = -0.76$	-0.76	0.10	$0.33, [-0.93, -0.59]$	-0.76	0.10	$0.32, [-0.92, -0.60]$	0.89	0.87	0.88	0.88
				ARMA-DGP2						
$\pi = 0.40$	0.22	0.36	$1.01, [-0.58, 0.90]$	0.15	0.32	$1.80, [-0.90, 0.90]$	0.67	0.86	0.96	0.74
$\beta = \frac{-0.5}{\sqrt{T}} = -0.05$	-0.06	0.09	$0.31, [-0.23, 0.07]$	-0.07	0.10	$0.34, [-0.25, 0.11]$	0.81	0.78	0.78	0.87
$T = 500$										
				ARMA-DGP1						
$\pi = 0.40$	0.40	0.06	$0.20, [0.30, 0.50]$	0.40	0.06	$0.20, [0.31, 0.51]$	0.90	0.89	0.90	0.90
$\beta = -0.76$	-0.76	0.04	$0.15, [-0.83, -0.68]$	-0.76	0.04	$0.14, [-0.83, -0.69]$	0.91	0.91	0.91	0.91
				ARMA-DGP2						
$\pi = 0.40$	0.17	0.35	$1.02, [-0.65, 0.90]$	0.10	0.31	$1.80, [-0.90, 0.90]$	0.66	0.85	0.96	0.73
$\beta = \frac{-0.5}{\sqrt{T}} = -0.02$	-0.03	0.04	$0.14, [-0.10, 0.04]$	-0.03	0.04	$0.15, [-0.10, 0.06]$	0.82	0.84	0.81	0.88

Notes: Results are based on $M = 2000$ Monte Carlo simulations and $N = 499$ bootstrap replications of the ML estimator. Column (a): true parameter values. Column (b): ML estimates (average across Monte Carlo simulations). Column (c): standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}_T$, see Remark 2.1, average across Monte Carlo simulations). Column (c): lengths of 90% confidence intervals (in each cell, the first value is the median length of the intervals and the remaining two values are the medians of their lower and upper limits across Monte Carlo simulations. Column (e): bootstrap ML estimates (average across Monte Carlo simulations of the $\bar{\hat{\theta}}^*_T$s, see footnote 8). Column (f): bootstrap standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}^*_T$, see Proposition 2, average across Monte Carlo simulations). Column (g): lengths of 90% bootstrap confidence intervals (computed as explained in the Remark 3.3; the first value is the median length of the intervals and the remaining two values are the medians of their lower and upper limits across Monte Carlo simulations). Column (h): empirical probability coverages, i.e. the frequencies that the asymptotic confidence intervals contain the true parameter values. Column $(i$-$j)$: bootstrap coverages.

TABLE SM.2. Monte Carlo experiment, ARMA(1,1) model: misspecification tests.

θ (a)	DH (b) $B=\frac{1}{3}T^{4/5}$	DH (b) $B=\frac{1}{2}T^{4/5}$	JB (c) $B=\frac{1}{3}T^{4/5}$	JB (c) $B=\frac{1}{2}T^{4/5}$	SW (d) $B=\frac{1}{3}T^{4/5}$	SW (d) $B=\frac{1}{2}T^{4/5}$
$T=100$						
			ARMA-DGP1			
$\pi=0.40$	0.06	0.08	0.09	0.11	0.09	0.12
$\beta=-0.76$			0.05	0.06	0.05	0.06
			ARMA-DGP2			
$\pi=0.40$	0.19	0.31	0.20	0.27	0.34	0.48
$\beta=\frac{-0.5}{\sqrt{T}}=-0.05$			0.09	0.12	0.11	0.15
$T=500$						
			ARMA-DGP1			
$\pi=0.40$	0.06	0.06	0.07	0.07	0.08	0.08
$\beta=-0.76$			0.05	0.06	0.07	0.06
			ARMA-DGP2			
$\pi=0.40$	0.64	0.78	0.42	0.62	0.77	0.88
$\beta=\frac{-0.5}{\sqrt{T}}=-0.02$			0.17	0.22	0.26	0.35

Notes: Empirical rejection frequencies based on $M = 2000$ Monte Carlo simulations and B bootstrap samples selected out of $N = 499$ bootstrap replications. Column (a): parameters. Column (b): rejection frequency of Doornik and Hansen's (2008) multivariate normality test. Column (c): rejection frequency of Jarque and Bera's (1987) univariate normality test. Column (d) rejection frequency of Shapiro and Wilk's (1965) univariate normality test. All test are computed at the 5% nominal significance level.

TABLE SM.3. Estimated DSGE model on U.S. quarterly data.

								Misspecification tests					
Estimates on different samples													
								$B=\frac{1}{3}T^{4/5}$	$\frac{1}{2}T^{4/5}$	$\frac{1}{3}T^{4/5}$	$\frac{1}{2}T^{4/5}$	$\frac{1}{3}T^{4/5}$	$\frac{1}{2}T^{4/5}$
θ (a)	*Bounds* (b)	$\hat{\theta}_T$ (c)	$s(\hat{\theta}_T)$ (d)	90%-CI (e)	$\overline{\hat{\theta}}^*_T$ (f)	$s(\hat{\theta}^*_T)$ (g)	90%-BSCI (h)	DH (i)		JB (l)		SW (j)	
							GQ-M1: GI - 1954:Q3-1984:Q1						
α	$(0,1)$	0.71	0.01	$[0.69,0.73]$	0.73	0.01	$[0.71,0.76]$	0.79	0.99	0.83	0.74	0.89	0.76
							GQ-M1: ALL - 1954:Q3-2008:Q3						
α	$(0,1)$	0.74	0.01	$[0.73,0.76]$	0.77	0.01	$[0.75,0.79]$	0.69	0.93	0.94	0.86	0.97	0.98
							GQ-M2: GI - 1954:Q3-1984:Q1						
ρ_r	$(0,1)$	0.82	1.01	$(0.00,1.00)$	0.81	0.22	$[0.13,1.00)$			0.95	0.75	0.99	0.86
ϕ_π	$(1,5)$	1.00	43.82	$(1,00,5.00)$	1.01	7.34	$(1.00,1.28]$	0.07	0.00	0.09	0.00	0.01	0.00
ϕ_x	$(0,1.5)$	0.14	6.13	$(0.00,1.50)$	0.17	1.45	$[0.14,1.36]$			0.11	0.10	0.08	0.06
							GQ-M2: ALL - 1954:Q3-2008:Q3						
ρ_r	$(0,1)$	0.85	0.060	$(0.00,1.00)$	0.84	0.08	$[0.33,0.96]$			0.88	0.87	0.47	0.86
ϕ_π	$(1,5)$	1.00	37.71	$(1.00,5.00)$	1.01	4.27	$(1.00,1.29]$	0.02	0.00	0.00	0.00	0.00	0.00
ϕ_x	$(0,1.5)$	0.20	6.50	$(0.00,1.50)$	0.26	0.83	$[0.22,1.5)$			0.57	0.31	0.57	0.23

Notes: Column (a): parameters. Column (b): bounds for permissible parameter values. Column (c): QML estimates. Column (d): standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}_T$, see Remark 2.1). Column (e): 90% confidence intervals (lower and upper limits). Column (f): bootstrap QML estimates ($\overline{\hat{\theta}}^*_T$s, see footnote 8). Column (g): bootstrap standard errors (computed from the main diagonal of the covariance matrix $\hat{\Omega}^*_T$, see Proposition 2). Column (h): 90% bootstrap confidence intervals (lower and upper limits). Column (i): p-values of Doornik and Hansen's (2008) multivariate normality test. Column (j): p-values of Jarque and Bera's (1987) univariate normality test. Column (j) p-values of Shapiro and Wilk's (1965) univariate normality test.

TABLE SM.4. Computation time.

	Monte Carlo # est. param.	$T = 100$	$T = 500$
GQ-DGP1 ($N = 499$, $M = 2000$)	1	72442[8]	297125[8]
GQ-DGP2 ($N = 499$, $M = 2000$)	3	191972[8]	745785[8]
ARMA-DGP1 ($N = 499$, $M = 2000$)	2	11016	19584
ARMA-DGP2 ($N = 499$, $M = 2000$)	2	16776	30600
	Empirical # est. param.	$T = 98$ (baseline sample)	
GQ-M1 ($N = 1999$)	1	648	
GQ-M2 ($N = 1999$)	3	1620	

Notes: Computation times, seconds. The results are based on a Processor Intel(R) Xeon(R) Gold 5220 CPU 2.20GHz 2.19 GHz, 256GB RAM and the software used is Matlab R2020a. N is the number of bootstrap replications, M the number of Monte Carlo simulations and T is the sample size. In brackets the number of cores used in parallel.

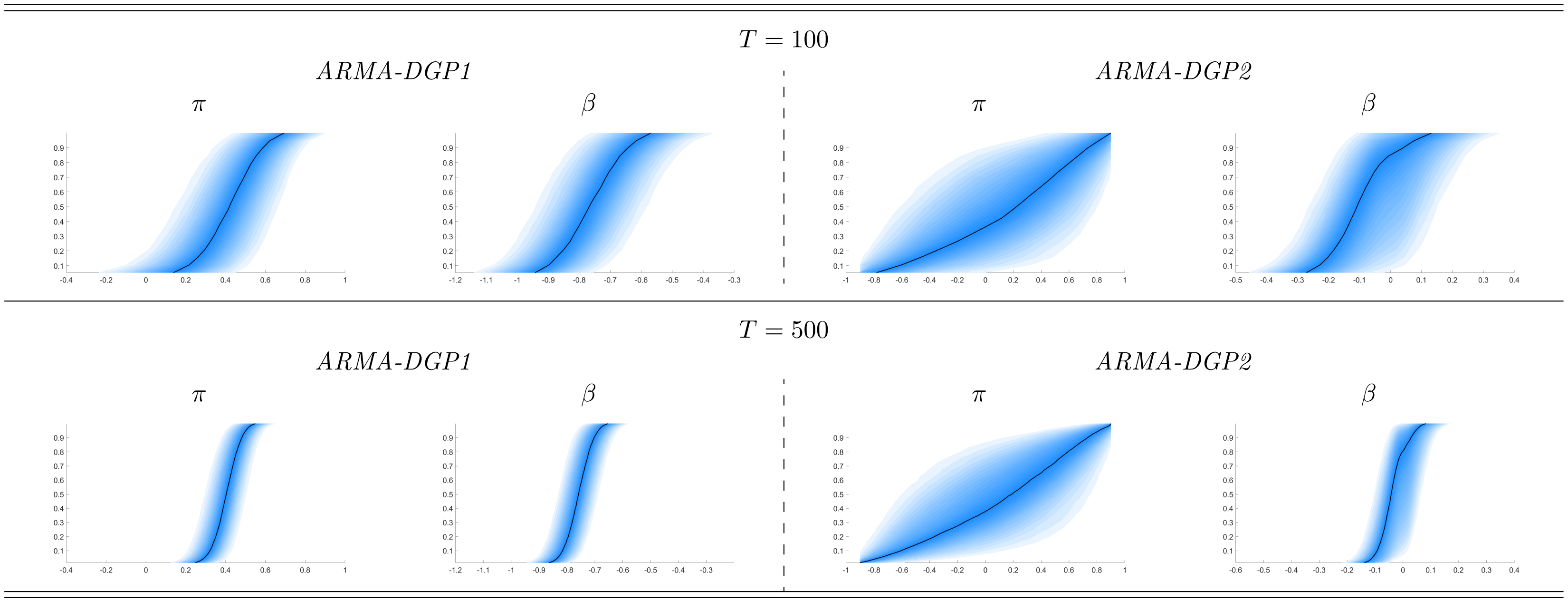

FIGURE SM.1: *Percentiles of the CDFs of $B = \frac{1}{2}T^{4/5}$ bootstrap replications of the QML estimator of the parameters across Monte Carlo simulations. The black lines indicate the median.*