Small Sample Properties of Likelihood Ratio Tests in Linear State Space Models: An Application to DSGE Models

Ivana Komunjer* Yinchu Zhu†

December 17, 2017

Abstract

This paper considers the problem of hypothesis testing in linear Gaussian state space models. We consider two hypotheses of interest: a simple null and a hypothesis of explicit parameter restrictions. We derive the asymptotic distributions of the corresponding likelihood ratio test statistics and compute the Bartlett adjustments. The results are non-trivial because the unrestricted state space model is not (even locally) identified. We apply our analysis to test the validity of the Dynamic Stochastic General Equilibrium (DSGE) models. A Monte Carlo exercise illustrates our findings and confirms the importance of Bartlett corrections at sample sizes typically encountered in macroeconomics.

Keywords: linear Gaussian state space models; likelihood ratio test; Bartlett adjustment.

JEL Codes: C12; C32.

1 Introduction

Linear Gaussian state space models are commonly used across a variety of fields in Economics: in macroeconomics, they are used to analyze first-order solutions to Dynamic Stochastic General Equilibrium (DSGE) models; in finance, they appear in the context of affine term structure models. The econometric analysis of likelihood based inference in these models has, however, remained scarce. The purpose of our paper is to fill this literature gap by formally studying the asymptotic

*Corresponding Author: Department of Economics, Georgetown University, 37th St NW & O St NW, Washington, DC 20007. Email: Ivana.Komunjer@georgetown.edu

†Lundquist College of Business, University of Oregon, 1208 University St, Eugene, OR 97403. Email: yzhu6@uoregon.edu.

This paper was previously circulated under the title "Testing the Validity of DSGE Models." We would like to thank the participants of the 2014 "Workshop on Identification" at the National Bank of Poland, 2015 Vienna Workshop on High Dimensional Time Series in Macroeconomic and Finance, and 2016 CIREQ Econometrics Conference in Honor of Jean-Marie Dufour, for their feedback. The authors are grateful to the co-editors, Marine Carrasco, Marcelo J. Moreira, Benoit Perron, Vicky Zinde-Walsh, and three anonymous referees for their comments and suggestions which greatly improved the paper.

properties of the likelihood ratio (LR) test in linear Gaussian state space models. Our contributions are twofold. First, we derive the asymptotic distributions of the LR tests under two hypotheses of interest: a simple null and a null of explicit parameter restrictions. Second, we propose Bartlett adjustments to both LR test statistics that for given samples of size T reduce level-error from order T^{-1} to order T^{-2}. Such adjustments appear important at simple sizes typically encountered in macroeconomics. As an application of our results, we propose a likelihood based test for the validity of DSGE models.

The literature on identification and estimation in linear Gaussian state space models is well established and complete; see, e.g., the books by Hannan and Deistler (1988) in econometrics, and Kailath, Sayed, and Hassibi (2000) in control theory. Considerably less is known about inference procedures in these models. This is primarily due to the fact that without additional restrictions, linear Gaussian state space models are not locally identified. It is well known (see, e.g. Komunjer and Ng, 2011) that similarity transforms rotating the latent variables of the state-space model leave unchanged the second-order properties of the observables. Thus, the information contained in the autocovariances of the observables alone does not suffice to identify the state-space model parameters. Without identification, standard regularity conditions needed for likelihood based inference are not satisfied, which makes the study of the asymptotic properties of the LR test non-trivial.

The starting point of our approach is the observation that even though they are not identified, linear Gaussian state space models have a manifold structure (see, e.g., Theorem 2.6.3 in Hannan and Deistler, 1988): that is, they can be parameterized locally by a lower-dimensional "canonical" parameter which is by construction identified.[1] Since the likelihood of the model is invariant to re-parameterizations, one can study the asymptotic behavior of the LR test in terms of the "canonical" parameter (for a detailed analysis of invariance see, e.g., Dagenais and Dufour, 1991). Its asymptotic distribution will have a standard chi-squared form, whose number of degrees of freedom will depend on the "canonical" parameter dimension. The latter is easy to compute and does not require constructing the "canonical" parameter itself, issue which has pre-occupied much of the estimation literature (see, e.g., Hannan and Deistler, 1988, for an overview). The construction of a Bartlett adjustment to the LR statistic proceeds following a similar argument. An additional difficulty here is that because of non-identification, Fisher information matrices become singular, which requires the use of pseudo-inverses in the construction of the Bartlett factor.[2]

[1] Global parameterizations do not exist in general; see, e.g., Section 3.5 in Hazewinkel (1977).

[2] It is, to the best of our knowledge, still an open question whether without identification an appropriate bootstrap likelihood ratio test à la Beran (1988) achieves the same Bartlett correction.

We should point out several important related papers that address the problem of likelihood based inference under nonidentification. Liu and Shao (2003) derive the asymptotic distribution of the LR test statistic for testing simple null hypotheses in parametric models more general than ours. Their limiting distribution is, however, characterized as a supremum of a function of a Gaussian process over a subset in $\mathcal{L}_2$, which is difficult to compute.[3] Andrews and Cheng (2012) provide comprehensive estimation and inference results under strong, semi-strong, weak and nonidentification. Similar to Liu and Shao (2003), the limiting distributions derived in Andrews and Cheng (2012) are generally non-trivial to compute when identification fails. Andrews and Guggenberger (2015) provide a number of identification-robust testing procedures for moment condition models, which in likelihood scenarios like ours are based on the score function.[4] While the distribution is the usual chi-squared, its number of degrees of freedom has to be estimated. Additional works that explicitly consider DSGE models include: Kleibergen and Mavroeidis (2009), Dufour, Khalaf, and Kichian (2013), Guerron-Quintana, Inoue, and Kilian (2013), Qu (2014), and Andrews and Mikusheva (2015), and we further discuss them below. Our contributions relative to these papers are threefold: first, in addition to simple nulls we also consider a composite null hypothesis of explicit parameter restrictions.[5] Second, without assuming identification in the limit, we obtain "chi-squared" limiting distributions with known degrees of freedom. And third, we calculate Bartlett adjustments which make the asymptotic chi-squared approximations more accurate.

To illustrate the usefulness of our results, we consider Dynamic Stochastic General Equilibrium (DSGE) models and ask the following two questions: first, how to construct confidence sets for the deep parameters θ of DSGE models that are fully robust to identification failure? and second, how to test the validity of the DSGE model specification? Both questions are of paramount importance to the applied researchers, as these models have become the workhorse of modern macroeconomic analysis. Yet, the issue of weak or nonidentfication of the DSGE model parameters has been largely documented (see, e.g., Komunjer and Ng, 2011; Qu and Tkachenko, 2012), and there is growing

[3]While the test of Liu and Shao (2003) seems difficult to operationalize in general, we conjecture that in the context of linear state space models one could show equality between our limiting distribution and theirs. The required calculations are however highly non-trivial, and not considered here.

[4]It is worth pointing out that while in theory a score (or Lagrange Multiplier) test can be derived and shown to be asymptotically equivalent to the LR test in our setup, there are important differences between the two tests when it comes to Bartlett corrections (see, e.g., Section 3.2 in Cordeiro and Cribari-Neto, 2014). First, the two tests have different higher-order properties: the error of the LR test before any correction is generally $O(T^{-1})$ and becomes $O(T^{-2})$ after the correction; the score test, on the other hand, has error $O(T^{-1})$ even after the correction. Second, the Bartlett correction is considerably simpler to calculate for the LR test than for the score test. This is why we focus on the LR test in this paper.

[5]To the best of our knowledge, the only papers that consider subvector tests, which are an example of a composite null, are Kleibergen and Mavroeidis (2009) and Andrews and Mikusheva (2015). Kleibergen and Mavroeidis (2009) only provide bounds on the limiting distributions of their test statistics. Andrews and Mikusheva (2015) require that the nuisance parameter be strongly identified.

consensus among macroeconomists today that DSGE models are misspecified in various aspects (see, e.g., Schorfheide, 2013, for a recent survey).

Our confidence sets for θ are obtained by inverting the LR test (with and without Bartlett adjustments) of the simple null hypothesis that the DSGE model reduced form parameter π takes value $\pi(\overline{\theta})$ with $\overline{\theta}$ varying over the deep parameter space Θ. More formally, our confidence sets for θ contain all the values $\overline{\theta} \in \Theta$ for which the unadjusted and Bartlett adjusted likelihood ratio tests fail to reject the point null hypothesis $H_1 : \pi = \pi(\overline{\theta})$ against the alternative that that π is unrestricted. It is worth emphasizing that our test is fully robust to identification failure and does not require either the deep parameter θ nor the reduced form parameter π to be (either weakly or strongly) identified. While the idea of inverting a test statistic to obtain confidence sets for the deep parameters is not new, there are important differences between our result and those available in the DSGE literature. Adapting results from the weak IV literature, Kleibergen and Mavroeidis (2009) study the asymptotic behavior of several test statistics under nonidentification of θ. However, they require an identifiable reduced form, and only provide bounds on the limiting distributions. The inversion method in Dufour, Khalaf, and Kichian (2013) requires that the observables admit finite-order VAR representations. Our approach on the other hand allows the observables to follow a linear state space model, i.e. have a VARMA representation. Guerron-Quintana, Inoue, and Kilian (2013) also invert a LR test but there are three key differences between their confidence sets and ours: first, they require that the reduced form parameter be identified, which is not the case in our setup. Second, because of their identification requirement for π, their confidence sets do not have uniform coverage. We allow π to be unidentified, and show that our confidence sets do. Third, in order to construct their confidence sets, Guerron-Quintana, Inoue, and Kilian (2013) need to estimate the rank of the transformation from θ to π. In our setup, the confidence sets are constructed by inverting LR test statistics whose limiting distributions are "chi-squared" with known degrees of freedom, thus no rank estimation is necessary. Working in the frequency domain, Qu (2014) proposes confidence sets for θ that do not require the existence of an identified reduced form parameter. The key differences are, first, that Qu (2014) inverts a score test while we invert the LR test. As already mentioned, this is likely to lead to better higher-order properties of our procedure. Second, the limiting "chi-squared" distribution derived in Qu (2014) depends on a rank of a matrix that needs to be estimated, while ours is known and thus requires no estimation. Andrews and Mikusheva (2015) derive the asymptotic distribution of the Lagrange multiplier and LR test statistics under weak identification. As the authors point out on page 132 therein, their method rules out models that are locally nonidentified.

Our specification test for the DSGE model is a test of explicit parameter restrictions on the parameter π of the linear state space model. Specifically, if the DSGE model is correctly specified, then π is a known function of a lower dimensional deep parameter θ. The validity of the DSGE model can then be tested by testing the composite null hypothesis $H_2 : \pi \in \{\pi(\theta), \theta \in \Theta\}$ against the alternative that π is unrestricted. Several approaches to assessing the accuracy of DSGE models have emerged in the literature. Some of the earliest methods (e.g., Sargent, 1977, 1978; Hansen and Sargent, 1980) propose using the theory of classical tests (see also Christiano, 2007, for more recent examples). The idea is to nest the DSGE model under consideration in a larger family of models (typically a finite lag VAR) and to examine whether the restrictions imposed by the DSGE structure are acceptable within this larger family. Since analysis is typically conducted in a parametric framework, namely Gaussian, testing the DSGE model restrictions is possible using any of the classical likelihood based tests.[6] It is important to stress that all the existing full information methods maintain correct specification of the large model: finite lag VAR. In general, however, DSGE models only admit VARMA (or linear state space) representations, which makes the finite lag VAR assumption restrictive (see, e.g., Ravenna, 2007). While the idea of using the LR test to formally test the restrictions imposed by the DSGE model is not new, the form of our LR test statistic presents difficulties not seen in the earlier literature. As already pointed out, those difficulties stem from the non-identification of the unrestricted state space parameters π. Lack of identification will of course affect the asymptotic distribution of the LR test statistic (see, e.g., Lindsay, 1995). The more surprising result is that nonidentification does not affect the form of the limiting distribution ("chi-squared") but only its number of degrees of freedom. In order to ensure that the latter is known, we will impose in our treatment of the null H_2 the additional restriction that the deep parameter θ is identified. This is an improvement over the procedures proposed by Sargent (1977), Sargent (1978), Hansen and Sargent (1980) or Christiano (2007) that assume both π and θ to be identified.

It is worth pointing out that there is an alternative fully nonidentification robust specification testing procedure based on confidence sets. Say one constructs a confidence set for the DSGE model parameter θ by using one of the procedures described in either Dufour, Khalaf, and Kichian (2013), Guerron-Quintana, Inoue, and Kilian (2013), or our paper. If the constructed confidence set is empty, then one can conclude that the DSGE model is misspecified. The advantage of

[6] Various limited information specification testing procedures are found, for example, in Smith (1993), Canova (1994), Schorfheide (2000), Del Negro, Schorfheide, Smets, and Wouters (2007), and Le, Meenagh, Minford, and Wickens (2011). A yet different strand of literature proposes other metrics that compare the closeness of the DSGE model to the data. These include: R^2-like measures (Watson, 1993), visual spectrum based checks (Diebold, Ohanian, and Berkowitz, 1998), prior-predictive checks (Canova, 1995), posterior-predictive checks (An and Schorfheide, 2007), or Bayes factors (Fernández-Villaverde and Rubio-Ramírez, 2004; Rabanal and Rubio-Ramírez, 2005; An and Schorfheide, 2007).

such a specification testing procedure is that it is fully robust to nonidentification of either θ or π, if our confidence sets are used.[7] The disadvantage of such a procedure, however, is that it is computationally more demanding than our LR test of explicit parameter restrictions. Finally, we should emphasize that previously discussed testing methods such as Andrews and Mikusheva (2015) and Qu (2014) cannot be used to test the validity of the DSGE model as both maintain correct specification under both null and alternative hypothesis.

The remainder of this paper is organized as follows. In Section 2, we describe the setup needed for likelihood based analysis. In Section 3, we derive the asymptotic properties of the LR test for two hypothesis of interest: a simple null, and a null of explicit parameter restrictions. As a byproduct of the first result, we propose confidence sets that are fully robust to identification failure. The same section derives expressions for the Bartlett adjustments. Section 4 presents a Monte Carlo experiment, in which we apply our results to a simple Real Business Cycle (RBC) model. Section 5 concludes. Proofs as well as additional details regarding the RBC model are relegated to an Appendix.

As a matter of notation, for any $m \times n$ real matrix A, A^+ denotes the pseudo-inverse (or Moore-Penrose inverse) of A.[8] If v is a vector, then v_i denotes the ith component of v, while if v is a matrix, then $v_{i,j}$ denotes the ith row and jth column element of v. For notational brevity, we use summation convention that implies summation over repeated indices not otherwise defined; e.g. $c_{i,j} = v_{i,k}v_{j,k} \equiv \sum_k v_{i,k}v_{j,k}$. Summation over the time index t is always indicated explicitly.

2 Setup

2.1 Model and Assumptions

We are concerned with the linear Gaussian state-space models that take the form:

$$\begin{array}{c} X_{t+1} = AX_t + B\epsilon_{t+1} \\ Y_{t+1} = CX_t + D\epsilon_{t+1} \end{array}, \qquad \{\epsilon_t\}_{t\in\mathbb{Z}} \sim iid\, N(0, \Sigma) \tag{1}$$

with $X_t \in \mathbb{R}^{n_X}$, $\epsilon_t \in \mathbb{R}^{n_\epsilon}$, $Y_t \in \mathbb{R}^{n_Y}$, and the dimensions of the matrices A, B, C, D conform with those of the variables. While the econometrician is assumed to observe all the components of Y_t, the state vector X_t and the vector of disturbances ϵ_t are allowed to remain unobserved. Though unable

[7] As already pointed out, the procedure proposed by Dufour, Khalaf, and Kichian (2013) only works when the DSGE model has a finite-lag VAR reduced form. Similarly, Guerron-Quintana, Inoue, and Kilian (2013) require working with an identified reduced form parameter. Thus, a specification test based on either Dufour, Khalaf, and Kichian (2013) or Guerron-Quintana, Inoue, and Kilian (2013) confidence sets is not robust to nonidentification of π.

[8] It is the unique $n \times m$ matrix A^+ satisfying: (i) $AA^+A = A$; (ii) $A^+AA^+ = A^+$; (iii) $(AA^+)' = AA^+$; and (iv) $(A^+A)' = A^+A$.

to directly observe X_t or ϵ_t, we shall assume that the econometrician knows their dimensions. In particular, the dimension n_X of the state X_t, also called the *order* of the state space system (1), is known. One can think of ϵ_t as containing both the structural shocks as well as the measurement errors in the model. Although latent, we shall assume in this paper that ϵ_t is known to be randomly drawn (i.e. independent and identically distributed or iid) from a Gaussian distribution with mean zero and covariance matrix Σ that is positive definite, $\Sigma > 0$. There are two parts to this restriction: first is the requirement that ϵ_t be iid. We argue that this condition is not too restrictive as models with serially correlated disturbances can easily be transformed to fit into our setup. Provided the measurement error process has a finite dimensional state vector, the latter can be included in X_t in the standard way thus permitting the system to be represented in the state-space form (1).[9] Second is the requirement that the disturbances be Gaussian. The normality assumption is common in the examples of (1) found in the literature, which we discuss below.

Example 1: DSGE Models Consider a generic nonlinear discrete time Dynamic Stochastic General Equilibrium (DSGE) model in which the parameter of interest (or "deep" parameter), denoted by θ, belongs to a set $\Theta \subseteq \mathbb{R}^{d_\theta}$. Log-linearizing the DSGE model solution around a unique steady-state leads to a recursive equilibrium law of motion given by:

$$\begin{array}{l} X_t = \widetilde{A}(\theta)X_{t-1} + \widetilde{B}(\theta)u_t \\ Y_t = \widetilde{C}(\theta)X_t + v_t \end{array}, \qquad \{(u_t', v_t')'\}_{t\in\mathbb{Z}} \sim iid\ N(0, \Sigma(\theta)), \tag{2}$$

where Y_t is an n_Y-vector of endogenous variables observed by the econometrician, X_t is an n_X-vector of latent states, and u_t is an n_u-vector of structural shocks, and v_t is an n_Y-vector of measurement errors. The matrices $\widetilde{A}, \widetilde{B}, \widetilde{C}$ and Σ in (2) are functions θ. Their expressions are available either analytically—in the case of simple DSGE models that one can solve by hand—or else numerically (see the algorithms of Anderson and Moore, 1985; Uhlig, 1999; Sims, 2002; Klein, 2000; King and Watson, 2002, for e.g.). Note that the model (2) is a special case of (1), obtained by collecting the disturbances into $\epsilon_t \equiv (u_t', v_t')'$ and letting

$$A \equiv \widetilde{A}(\theta), \quad B \equiv \begin{pmatrix} \widetilde{B}(\theta) & 0 \end{pmatrix}, \quad C \equiv \widetilde{C}(\theta)\widetilde{A}(\theta), \quad D \equiv \begin{pmatrix} \widetilde{C}(\theta)\widetilde{B}(\theta) & \mathrm{Id} \end{pmatrix}.$$

[9] The hybrid models in Ireland (2004); Khalaf, Lin, and Reza (2014), for example, take the form: $\bar{X}_t = \bar{A}\bar{X}_{t-1} + \bar{B}e_t$, $Y_t = \bar{C}\bar{X}_t + u_t$, $u_t = \bar{D}u_{t-1} + \xi_t$. Letting $X_t \equiv (\bar{X}_t', u_t')'$ and $\epsilon_t \equiv (e_t', \xi_t')'$ be the new state and disturbance, respectively, gives the state-space model in (1) with

$$A \equiv \begin{pmatrix} \bar{A} & 0 \\ 0 & \bar{D} \end{pmatrix}, \quad B \equiv \begin{pmatrix} \bar{B} & 0 \\ 0 & \mathrm{Id} \end{pmatrix}, \quad C \equiv \begin{pmatrix} \bar{C}\bar{A} & \bar{D} \end{pmatrix}, \quad D \equiv \begin{pmatrix} \bar{C}\bar{B} & \mathrm{Id} \end{pmatrix}.$$

For an alternative approach leading to smaller system matrices A, B, C, D but with Y_t that is not directly observable see, e.g., Sargent (1989a).

The Gaussian assumption on ϵ_t is commonly found in both classical (see, e.g., Altug, 1989; Ireland, 2004; Qu, 2014) as well as Bayesian (see, e.g., Schorfheide, 2000; Del Negro, Schorfheide, Smets, and Wouters, 2007; An and Schorfheide, 2007) likelihood-based analysis of the DSGE models such as (2).[10] □

Example 2: Affine Term Structure Models Affine models rely on a low-dimensional state vector X_t to describe what drives the yield curve. With τ denoting the maturity, affine models are state space systems with an observation equation that links observable yields $Y_t^{(\tau)}$ to the state vector X_t and a state equation that describes the dynamics of the state:

$$\begin{array}{l} Y_t^{(\tau)} = A(\tau) + B(\tau)X_t + \epsilon_t^{(\tau)} \\ X_t = c + \rho X_{t-1} + u_t, \end{array} \quad , \qquad \{(\epsilon_t^{(\tau)\prime}, u_t')'\}_{t\in\mathbb{Z}} \sim iid\ N(0,\Sigma). \tag{3}$$

The coefficients $A(\tau)$ and $B(\tau)$ can either be computed in closed form (see, e.g. Vasicek, 1977; Cox, Ingersoll, and Ross, 1985), or numerically by solving a system of ordinary differential equations (see, e.g. Ch 3.3 in Piazzesi, 2010). Hamilton and Wu (2012) consider the problems of identification and estimation of the model (3). Once demeaned, the variable $\widetilde{Y}_t^{(\tau)} = Y_t^{(\tau)} - E(Y_t^{(\tau)})$, follows the observation equation $\widetilde{Y}_t^{(\tau)} = B(\tau)\widetilde{X}_t + \epsilon_t^{(\tau)}$, with the state $\widetilde{X}_t = X_t - E(X_t)$ which evolves according to $\widetilde{X}_t = \rho\widetilde{X}_{t-1} + u_t$. □

Example 3: Dynamic Factor Models Dynamic factor models have a long history in macroeconomic data modeling, see, e.g., Geweke (1977); Sargent, Sims, et al. (1977); Engle and Watson (1981); Stock and Watson (1989); Sargent (1989b). Following Stock and Watson (2011), these models can be formulated as:

$$\begin{array}{l} Y_t = \lambda(L)f_t + e_t, \\ f_t = \Psi(L)f_{t-1} + \eta_t, \end{array} \quad , \quad \{(e_t', \eta_t')'\}_{t\in\mathbb{Z}} \sim iid\ N(0,\Sigma)$$

where f_t is the vector representing the underlying unobserved factors and $\Psi(L)$ and $\lambda(L)$ are matrix polynomials of the lag operator L. We can enlarge the state vector to include enough lags of f_t and let $X_t \equiv (f_t', f_{t-1}', \ldots, f_{t-p}')'$ with p being the maximal lags in $\Psi(L)$ and $\lambda(L)$. Then the dynamic factor models can be written as:

$$\begin{array}{l} Y_t = \widetilde{\Lambda}X_t + e_t \\ X_t = \widetilde{\Psi}X_t + \widetilde{\eta}_t, \end{array} \quad , \quad \{(e_t', \eta_t')'\}_{t\in\mathbb{Z}} \sim iid\ N(0,\Sigma)$$

where $\widetilde{\eta}_t = (\eta_t', 0, \ldots, 0)'$ and $\widetilde{\Lambda}$ and $\widetilde{\Psi}$ are matrices. □

[10] For likelihood-based estimation in nonlinear and/or non-Gaussian state-space models see, e.g., Rubio-Ramírez and Fernandez-Villaverde (2007).

Example 4: VARMA Models Another popular model is the Gaussian Vector Autoregressive Moving Average (VARMA) process. We ignore the mean for simplicity. The general form of a Gaussian VARMA(p, q) model is:

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t + M_1 u_{t-1} + \cdots + M_q u_{t-q}, \quad \{u_t\}_{t\in\mathbb{Z}} \sim iid\ N(0, \Sigma),$$

where u_t is the random noise, and $A_1, \ldots, A_p$ and $M_1, \ldots, M_q$ are matrices. It is well known that VARMA models can be cast in the form of linear state-space models (see, e.g., Hannan and Deistler, 1988; Lütkepohl, 2005). For example, consider VARMA$(1,1)$ model: $Y_t = AY_{t-1} + u_t + Mu_{t-1}$. We can define $X_t = (Y_t', u_t')'$ and $\widetilde{u}_t = (u_t', u_{t-1}')'$ and rewrite this model as

$$\begin{aligned} X_t &= \widetilde{A} X_{t-1} + \widetilde{u}_t \\ Y_t &= \widetilde{C} X_t \end{aligned} \qquad \text{where} \quad \widetilde{A} = \begin{pmatrix} A & M \\ 0 & 0 \end{pmatrix}, \ \widetilde{C} = (\mathrm{Id}, 0).$$

This is a special case of the state-space model (1) with $A \equiv \widetilde{A}$, $\widetilde{B} \equiv \mathrm{Id}$, $C \equiv \widetilde{C}\widetilde{A}$, and $D \equiv \widetilde{C}$. □

We now turn to our assumptions. To ensure that the sequence of observed variables $\{Y_t\}_{t\in\mathbb{Z}}$ is stationary, we impose stability of the transition matrix A in (1). This condition subsumes that all the necessary variable transformations have been performed so that all the variables of the DSGE model are stationary.

Assumption 1. *A is stable, i.e. all eigenvalues of A are inside the unit circle.*

Under Assumption 1, $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary Gaussian process with zero mean and autocovariances $\Gamma(j) \equiv E(Y_{t+j} Y_t')$ given by:

$$\Gamma(j) = \begin{cases} D\Sigma D' + CP_X C', & j = 0 \\ CA^{j-1}\left(AP_X C' + B\Sigma D'\right), & j > 0 \end{cases}, \tag{4}$$

where $P_X \equiv E(X_t X_t')$ is the unique positive semi-definite solution to the Lyapunov equation: $P_X = AP_X A' + B\Sigma B'$. Uniqueness obtains under Assumption 1; positive semi-definiteness uses $\Sigma > 0$. Note that the values of $\Gamma(j)$ when $j < 0$ can be obtained from $\Gamma(-j) = E(Y_{t-j} Y_t') = \Gamma(j)'$. It is clear from the expression of $\Gamma(j)$ above that alternatives quintuples (A, B, C, D, Σ) can give rise to processes $\{Y_t\}_{t\in\mathbb{Z}}$ with identical autocovariances. In fact, there may even exist state vectors $\widetilde{X}_t$ of dimension $\widetilde{n}_X \neq n_X$, such that the state space system $\widetilde{X}_{t+1} = \widetilde{A}\widetilde{X}_t + \widetilde{B}\epsilon_{t+1}$, $Y_{t+1} = \widetilde{C}\widetilde{X}_{t+1} + \widetilde{D}\epsilon_{t+1}$, has the same autocovariance structure as the one in (1). To eliminate such possibilities, we shall hereafter assume that the system in (1) is *autocovariance minimal* in a sense defined below.

Definition 1. *A state space system* (1) *of order n_X is autocovariance minimal if there exists no other state space system of order $\widetilde{n}_X \leqslant n_X$ that has the same autocovariance structure.*

Though similar to the usual notion of minimality of the transfer function (see, e.g., Hannan and Deistler, 1988), the autocovariance minimality does not require the econometrician to observe the "inputs" ϵ_t in (1).[11] Thus the notion is suited for the analysis of systems in which only the autocovariances of the "outputs" Y_t are observed by the econometrician. To state the primitive conditions for autocovariance minimality the following notions will be useful.

Definition 2. *The matrix pair (A, B) is said to be controllable if the $n_X \times n_X n_\epsilon$ controllability matrix $\begin{pmatrix} B & AB & \cdots & A^{n_X-1}B \end{pmatrix}$ has rank n_X. The pair (A, C) is called observable if (A', C') is controllable.*

Now consider again the autocovariances $\Gamma(j)$ in (4). Letting L be the Cholesky factor of $\Gamma(0)$, i.e. $\Gamma(0) = LL'$, and $N \equiv AP_XC' + B\Sigma D'$, the autocovariances can be written as:

$$\Gamma(j) = \begin{cases} \Gamma(0), & j = 0 \\ CA^{j-1}N, & j > 0 \end{cases} .$$

Note that the structure above is similar to that of a transfer function of an "artificial" system

$$\begin{aligned} S_{t+1} &= AS_t + NU_{t+1} \\ V_{t+1} &= CS_t + \Gamma(0)U_{t+1}, \end{aligned} \tag{5}$$

in which both the inputs U_t and the outputs V_t are observed. Indeed, the Markov parameters in $V_t = H(L)U_t$, $H(z) = \sum_{j=0}^{\infty} h(j)z^j$ ($z \in \mathbb{C}$), have precisely the form:

$$h(j) = \begin{cases} \Gamma(0), & j = 0 \\ CA^{j-1}N, & j > 0 \end{cases} .$$

In particular, the state space system (1) is autocovariance minimal if and only if the "artificial" system (5) is transfer function minimal. The latter is easy to characterize using the standard conditions: (A, N) controllable and (A, C) observable. This leads to the following necessary and sufficient condition for autocovariance minimality of the system (1).

Assumption 2. *Let $N = AP_XC' + B\Sigma D'$ and $P_X = AP_XA' + B\Sigma B'$. Then: (i) (A, N) controllable; (ii) (A, C) observable.*

Since our testing procedure is based on likelihood, certain nonsingularity restrictions are needed. In models with iid variables $\{Y_t\}_{t\in\mathbb{Z}}$, standard regularity conditions require that the distribution of Y_1 be absolutely continuous with respect to some reference measure (typically Lebesgue measure

[11] Following the usual terminology (see, e.g., Ch 1 in Hannan and Deistler, 1988), the input-output representation of a linear causal time-invariant system is given by: $Y_t = \sum_{j=0}^{\infty} h(j)\epsilon_{t-j}$, where Y_t are the outputs, ϵ_t are the inputs, and $h(j) \in \mathbb{R}^{n_Y \times n_\epsilon}$ are the coefficients of the transfer function $H(z) = \sum_{j=0}^{\infty} h(j)z^j$, $z \in \mathbb{C}$. $(h(j), 0 \leqslant j < \infty)$ is also called the impulse response function, and $h(j)$ the Markov parameters.

on $\mathbb{R}^{n_Y}$) with support that does not vary across the parameter space (see, e.g., Rothenberg, 1971). When Y_1 is Gaussian, this simply means that the covariance matrix of Y_1 needs to be nonsingular. Nonsingularity ensures that different components of Y_1 are not collinear. This simple condition can be generalized to dynamic models such as ours, provided, however, we now eliminate possible collinearity among the components of the entire process $\{Y_t\}_{t\in\mathbb{Z}}$. The condition will thus need to be stated in terms of the entire autocovariance generating function of $\{Y_t\}_{t\in\mathbb{Z}}$ or its *spectral density.*

Since the covariance function of $\{Y_t\}_{t\in\mathbb{Z}}$ is exponentially decaying, we can define its z-spectrum

$$\Omega(z) \equiv \sum_{j=-\infty}^{+\infty} \Gamma(j) z^{-j},$$

which is well defined in an annulus in the complex plan that contains the unit circle, $z = e^{i\omega}$ ($i = \sqrt{-1}$, $\omega \in [-\pi, \pi]$). In particular, the spectral density of $\{Y_t\}_{t\in\mathbb{Z}}$, $\Omega(e^{i\omega}) = \sum_{j=-\infty}^{+\infty} \Gamma(j) e^{-ij\omega}$, is well-defined for all $\omega \in [-\pi, \pi]$. It has the property that $\Omega'(e^{-i\omega}) = \Omega(e^{i\omega})$ (Hermitian), and $\Omega(e^{i\omega}) \geq 0$ for all $\omega \in [-\pi, \pi]$. The following result formally establishes the link between positive definiteness of the spectral density everywhere on the unit circle, and nonsingularity of the process $\{Y_t\}_{t\in\mathbb{Z}}$.

Lemma 1. *If $\Omega(e^{i\omega}) > 0$ for all $\omega \in [-\pi, \pi]$ then for every $T \geq 1$, the covariance matrix of $(Y_1', \ldots, Y_T')$ is full rank.*

Since $\{Y_t\}_{t\in\mathbb{Z}}$ is Gaussian, nonsingularity of the covariance matrix of $(Y_1', \ldots, Y_T')$ for every $T \geq 1$ is a necessary and sufficient condition for the existence of the joint density of $(Y_1', \ldots, Y_T')$ for every $T \geq 1$, and thus of the likelihood function. This raises the question of finding primitive conditions on the system matrices A, B, C, D in (1) that would ensure $\Omega(e^{i\omega}) > 0$, for every $\omega \in [-\pi, \pi]$. For this, we impose the following:

Assumption 3. *The system matrices A, B, C, D are such that:*

$$\mathit{rank} \begin{pmatrix} e^{i\omega} Id - A & B \\ -C & D \end{pmatrix} = n_X + n_Y, \quad \textit{for every } \omega \in [-\pi, \pi].$$

Notice that the above matrix is of dimensions $(n_X + n_Y) \times (n_X + n_\epsilon)$. Thus, a necessary condition for Assumption 3 is that $n_Y \leqslant n_\epsilon$. This is as we would expect, since it is well-known that the DSGE models with fewer disturbances ϵ_t than observables Y_t are stochastically singular. As already pointed out, the role of Assumption 3 is to ensure that the spectral density of $\{Y_t\}_{t\in\mathbb{Z}}$ is everywhere positive definite. This equivalence is formally established in the following lemma.

Lemma 2. *$\Omega(e^{i\omega}) > 0$ for every $\omega \in [-\pi, \pi]$ if and only if Assumption 3 holds.*

The rank requirement in Assumption 3 is particularly easy to check in state-space models with measurement errors that can be written as:

$$X_t = \widetilde{A} X_{t-1} + \widetilde{B} u_t,$$
$$Y_t = \widetilde{C} X_t + v_t,$$

where u_t is the vector of structural shocks, and v_t the vector of measurement errors. Note that the above model is a special case of (1), obtained by collecting the disturbances into $\epsilon_t \equiv (u_t', v_t')'$ and letting

$$A \equiv \widetilde{A}, \quad B \equiv \begin{pmatrix} \widetilde{B} & 0 \end{pmatrix}, \quad C \equiv \widetilde{C}\widetilde{A}, \quad D \equiv \begin{pmatrix} \widetilde{C}\widetilde{B} & \mathrm{Id} \end{pmatrix}.$$

The rank condition in Assumption 3 is then automatically satisfied whenever $A = \widetilde{A}$ is stable, as required by Assumption 1. This result is not surprising: recall from Lemma 2 that the requirement in Assumption 3 is equivalent to the requirement that the spectral density of $\{Y_t\}_{t \in \mathbb{Z}}$ be everywhere positive definite. In state-space models with measurement errors the additive white noise measurement error v_t effectively ensures that the spectral density of the observables is everywhere full rank, irrespective of the value of $\widetilde{C}$. Thus, $\widetilde{C}$ can be any $n_Y \times n_X$ matrix.

2.2 Innovations Representation

Our analysis to follow requires the computation of the likelihood of the state space model in (1). This construction typically uses the prediction error decomposition, which for any $T \geq 1$ consists in writing the joint distribution of $(Y_1, \ldots, Y_T)$ as a product of the conditional distributions of Y_{t+1} given the past $Y^t = (Y_t, \ldots, Y_1)$. Since $(Y_1, \ldots, Y_T)$ is multivariate Gaussian, all of the conditional distributions of Y_{t+1} given its past Y^t are Gaussian. The mean and variance of this distribution are typically computed through the Kalman filtering equations, which start with the initial conditions $\widehat{X}_0$ and $P_{0|0}$, then for $t \geq 1$ recursively compute $\widehat{X}_{t|t} = E[X_t | Y^t, \widehat{X}_0]$, and $P_{t|t} = E[(X_t - \widehat{X}_{t|t})(X_t - \widehat{X}_{t|t})' | Y^t, \widehat{X}_0]$ through

$$\Sigma_{a,t} = C P_{t|t} C' + D \Sigma D' \tag{6}$$
$$K_t = \left[A P_{t|t} C' + B \Sigma D' \right] \Sigma_{a,t}^{-1} \tag{7}$$
$$\widehat{X}_{t+1|t+1} = A \widehat{X}_{t|t} + K_t [Y_{t+1} - C \widehat{X}_{t|t}] \tag{8}$$
$$P_{t+1|t+1} = A P_{t|t} A' + B \Sigma B' - K_t \Sigma_{a,t} K_t'. \tag{9}$$

The prediction error $a_{t+1} \equiv Y_{t+1} - \widehat{Y}_{t+1|t} = Y_{t+1} - C \widehat{X}_{t|t}$ is conditionally normal with mean zero and variance $\Sigma_{a,t}$ given in (6).

Now, take $\widehat{X}_0 = 0$ and $P_{0|0} = P$ where P is a solution to the discrete time Riccati equation

$$P = APA' + B\Sigma B' - \left[APC' + B\Sigma D'\right]\left[CPC' + D\Sigma D'\right]^{-1}\left[CPA' + D\Sigma B'\right]. \tag{10}$$

It then follows that $P_{t|t} = P$ for all $t \geq 1$ and the recursions in (6)-(9) yield the so-called innovations representation of the state-space system in (1):

$$\begin{aligned} \widehat{X}_{t+1|t+1} &= A\widehat{X}_{t|t} + Ka_{t+1} \\ Y_{t+1} &= C\widehat{X}_{t|t} + a_{t+1} \end{aligned} \tag{11}$$

where P is a solution to the Riccati equation in (10), and

$$\Sigma_a = CPC' + D\Sigma D' \quad \text{and} \quad K = \left[APC' + B\Sigma D'\right]\Sigma_a^{-1}.$$

The prediction errors $a_{t+1} = Y_{t+1} - C\widehat{X}_{t|t}$ are now iid Gaussian with mean zero and variance Σ_a. In order to further construct the likelihood it is first necessary to establish that $\Sigma_a > 0$. For this, we have the following result.

Lemma 3. *Let Assumptions 1 and 2(ii) hold. Then there exists a unique positive semi-definite solution P to the Riccati equation* (10) *for which $A - KC$ is stable and $\Sigma_a > 0$, if and only if Assumption 3 holds.*

Put in words, when the process $\{Y_t\}_{t\in\mathbb{Z}}$ has a strictly positive spectral density, then all the conditional densities of Y_{t+1} given the past Y^t exist. This result is intuitive, though essential: without the existence of densities, a likelihood based approach would not be feasible.

2.3 Likelihood

The starting point in the construction of the likelihood of the state space model (1) is the prediction error decomposition (see, e.g., Harvey, 1989), which for any $T \geq 1$ writes the joint density of $(Y_1, \ldots, Y_T)$ as $p(Y_1, \ldots, Y_t) = \prod_{t=0}^{T-1} p(Y_{t+1} \mid Y^t)$. Each of the conditional densities $p(Y_{t+1} \mid Y^t)$ is Gaussian with mean $C\widehat{X}_{t|t}$ and variance Σ_a obtained from (11). The likelihood is therefore a function of the prediction errors a_{t+1} and their variance Σ_a. This has an important implication in terms of the parameters that enter the likelihood. Since a_{t+1} depends on the parameters A, B, C, D and Σ of the state space system only through the matrices A, K, C and Σ_a of the innovations representation in (11), the parameters appearing in the likelihood of the state space system (1) are the elements of

$$\pi \equiv \left((\text{vec}A)', (\text{vec}K)', (\text{vec}C)', (\text{vech}\Sigma_a)'\right)'. \tag{12}$$

For any $T \geq 1$, let $L_T(\pi)$ denote the Gaussian likelihood of the model, $L_T(\pi) \equiv p(Y_1, \dots, Y_T; \pi)$. Then, the log-likelihood $\ln L_T(\pi)$ of the state-space system in (1) is given by:

$$\ln L_T(\pi) = -\sum_{t=1}^{T} \left[\frac{n_Y}{2} \ln(2\pi) + \frac{1}{2} \ln \det \Sigma_a + \frac{1}{2} a_t' \Sigma_a^{-1} a_t \right], \tag{13}$$

where a_t and Σ_a are determined through (11).

As already pointed out, the existence of the likelihood $L_T(\pi)$ requires several restrictions on the innovations representation parameter π. This raises the question of what is the appropriate parameter space Π? To answer this question, we need to reconsider all our assumptions initially made on the A, B, C, D and Σ matrices of the state space system (1), and state them in terms of the matrices A, K, C and Σ_a of the innovations representation (11).

First, note that our stability Assumption 1 and our observability Assumption 2(ii) are directly stated in terms of the likelihood parameter π. Second, using the results of Lemma 3, our full rank Assumption 3, is equivalent to the restrictions that $A - KC$ be stable, and $\Sigma_a > 0$. Lastly, this leaves the question regarding the observability Assumption 2(i). For this, the following lemma is useful.

Lemma 4. *Let Assumptions 1 and 2(ii) hold, and moreover assume that $A - KC$ is stable and $\Sigma_a > 0$. Assumption 2(i) holds if and only if (A, K) is controllable.*

Using all of the above, we can now define the parameter space Π of the likelihood parameter $\pi = ((\text{vec}A)', (\text{vec}K)', (\text{vec}C)', (\text{vech}\Sigma_a)')'$ to be the following set:

$$\Pi \equiv \left\{ \pi : A \text{ stable}, \ (A, K) \text{ controllable}, \ (A, C) \text{ observable}, \ A - KC \text{ stable}, \ \Sigma_a > 0 \right\}. \tag{14}$$

The parameter space Π is an open subset of $\mathbb{R}^{d_\pi}$ with dimension $d_\pi \equiv n_X^2 + 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}$, therefore it is a smooth manifold of the same dimension. The latter is an important regularity condition needed for likelihood based inference on π. This is the problem to which we turn next.

3 Likelihood Ratio Inference

We now consider the problems of hypothesis testing and confidence set construction. It is worth emphasizing that both problems are non-standard because the parameter π in (12) is neither globally nor locally identified (see, e.g., Komunjer and Ng, 2011).

3.1 Hypothesis Testing: Simple Null

We start our analysis of likelihood based hypothesis testing with a case in which the hypothesis of interest takes the form

$$H_1: \quad \pi = \pi_0, \quad \pi_0 \in \Pi.$$

Put in words, we are interested in testing a simple null hypothesis that the system matrices A, K, C, Σ_a in the innovations representation (11) take particular values $A_0, K_0, C_0, \Sigma_{a0}$, at which $\pi_0 \equiv ((\text{vec}A_0)', (\text{vec}K_0)', (\text{vec}C_0)', (\text{vech}\Sigma_{a0})')'$ satisfies all the restrictions of the parameter set Π defined in (14). We can assess the veracity of H_1 by examining the behavior of the log-likelihood ratio statistics:

$$LR_{1T}(\pi_0) \equiv 2\left(\sup_{\pi\in\Pi} \ln L_T(\pi) - \ln L_T(\pi_0)\right). \tag{15}$$

In regular cases (see, e.g. Ch 17 in, Gourieroux and Monfort, 1995), LR_{1T} has an asymptotic chi-squared distribution with d_π degrees of freedom, where d_π is the dimension of π. Unfortunately, the model in (11) does not satisfy the needed regularity conditions: it is neither globally nor locally identified (see, e.g., Komunjer and Ng, 2011). The following result derives the asymptotic distribution of the LR test in our non-identified model (11).

Theorem 1. *Let Assumptions 1 to 3 hold. Define*

$$d \equiv 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}.$$

Then under H_1,

$$LR_{1T}(\pi_0) \xrightarrow{d} \chi^2_d.$$

Since the unrestricted model parameters π are neither globally nor locally identified, the result of Theorem 1 is non-trivial. Put in words, Theorem 1 states that the number of degrees of freedom in the asymptotic chi-squared distribution of the likelihood ratio statistic depends on the dimension d of the identifiable components in π, which is smaller than the dimension d_π of π. That the latter equals d follows from a classical manifold result in control theory of linear state space systems (see, e.g., Theorem 2.6.3 in Hannan and Deistler, 1988).

It is worth pointing out that while under standard regularity conditions (which include that of model identification), the likelihood ratio, score and Wald test statistics are well-known to be asymptotically equivalent, the same does not occur here. In particular, in order to obtain an asymptotic chi-square distribution for the Wald test statistic, one would need to establish that $\sqrt{T}(\widehat{\pi}-\pi_0)$ is asymptotically normal with possibly singular though positive semi-definite covariance matrix (see, e.g., Andrews, 1987). Since in our setup π is unidentified, it is not consistently estimable and the previous condition fails. The reason why the likelihood ratio test statistics is robust to identification failure of π is that it is based on the supremum of the likelihood function itself which is always identified.

Finally, note that the result of Theorem 1 crucially depends on the Gaussianity of the errors in (1). There are two reasons for this: it is only under Gaussianity that the information about the observed process $\{Y_t\}_{t\in\mathbb{Z}}$ reduces to the covariance function (or spectral density) of $\{Y_t\}_{t\in\mathbb{Z}}$. Thus, one can use the classical manifold result in control theory of linear state space systems to calculate the dimension of the unrestricted likelihood manifold only if the likelihood is Gaussian. Second, if the errors in (1) fail to be Gaussian, then the function $L_T(\pi)$ in (13) is no longer the log-likelihood but rather a pseudo-log-likelihood of the model. Since the model is then misspecified, the limit χ^2 distribution of Theorem 1 no longer obtains.

The result of Theorem 1 provides an asymptotic approximation to the distribution of $LR_{1T}(\pi_0)$ under the null hypothesis H_1. It can be restated by saying that

$$P_{\pi_0}\left(LR_{1T}(\pi_0) \leqslant r\right) = \Pr\left(\chi^2_d \leqslant r\right) + O(T^{-1}),$$

that is, the distribution of $LR_{1T}(\pi_0)$ is generally order T^{-1} away from that of χ^2_d. A simple multiplicative correction to the likelihood ratio statistic can further improve the quality of this approximation. Specifically, for an appropriate choice of a constant $b_{1T}(\pi_0)$, letting

$$LR^*_{1T}(\pi_0) \equiv \left(1 + \frac{b_{1T}(\pi_0)}{T}\right)^{-1} LR_{1T}(\pi_0), \tag{16}$$

results in a corrected likelihood ratio statistic whose distribution is order T^{-2} away from that of χ^2_d,

$$P_{\pi_0}\left(LR^*_{1T}(\pi_0) \leqslant r\right) = \Pr\left(\chi^2_d \leqslant r\right) + O(T^{-2}).$$

The idea for such a correction originated in Bartlett (1937), and the computation and efficacy of the adjustment have been discussed by Lawley (1956), McCullagh and Cox (1986), and Barndorff-Nielsen and Hall (1988), among others.

In general, the expression of $b_{1T}(\pi_0)$ depends on the higher order cumulants of the score function $\partial \ln L_T(\pi)/\partial\pi$ evaluated at $\pi = \pi_0$. In our setup, the scores can be computed as follows:

$$\begin{aligned}\frac{\partial \ln L_T(\pi)}{\partial \pi} &= -\frac{1}{2}\sum_{t=1}^{T}\left[\mathrm{vec}\left(\frac{\partial \ln\det\Sigma_a}{\partial \mathrm{vec}\Sigma_a}\right)'\left(\frac{\partial \mathrm{vec}\Sigma_a}{\partial\pi}\right) + \frac{\partial \mathrm{tr}(a_t'\Sigma_a^{-1}a_t)}{\partial\pi}\right]\\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[\mathrm{vec}\left[\Sigma_a^{-1}\left(\mathrm{Id}_{n_Y} - a_t a_t'\Sigma_a^{-1}\right)\right]'\left(\frac{\partial \mathrm{vec}\Sigma_a}{\partial\pi}\right) + 2a_t'\Sigma_a^{-1}\left(\frac{\partial a_t}{\partial\pi}\right)\right]\end{aligned} \tag{17}$$

The computation requires (i) $\partial \text{vec}\Sigma_a/\partial\pi$, and (ii) $\partial a_t/\partial\pi$. Both are available from the lemma below.

Lemma 5. *Let all the assumptions of Theorem 1 hold. Then,*

$$\frac{\partial vec\Sigma_a}{\partial\pi} = \begin{pmatrix} 0_{n_Y^2\times n_X(n_X+2n_Y)} & \mathcal{G}_{n_Y} \end{pmatrix},$$

where $\mathcal{G}_{n_Y}$ is an $n_Y^2 \times n_Y(n_Y+1)/2$ "duplication" matrix consisting of 0s and 1s, with a single 1 in each row, such that for any $n_Y \times n_Y$ symmetric matrix S, $vec(S) = \mathcal{G}_{n_Y} vech(S)$. Moreover, for any $t \geq 1$, the partial derivatives $\partial a_t/\partial\pi$ can be computed recursively from:

$$\begin{aligned}
\frac{\partial a_t}{\partial\pi} &= \begin{pmatrix} 0_{n_Y\times n_X^2} & 0_{n_Y\times n_X n_Y} & -(\widehat{X}'_{t-1|t-1}\otimes Id_{n_Y}) & 0_{n_Y\times n_Y(n_Y+1)/2} \end{pmatrix} - C\frac{\partial\widehat{X}_{t-1|t-1}}{\partial\pi} \\
\frac{\partial\widehat{X}_{t|t}}{\partial\pi} &= \begin{pmatrix} (\widehat{X}'_{t-1|t-1}\otimes Id_{n_X}) & (a'_t\otimes Id_{n_X}) & -(\widehat{X}'_{t-1|t-1}\otimes K) & 0_{n_X\times n_Y(n_Y+1)/2} \end{pmatrix} + (A-KC)\frac{\partial\widehat{X}_{t-1|t-1}}{\partial\pi} \\
\widehat{X}_{0|0} &= 0.
\end{aligned}$$

To derive the Bartlett adjustment $b_{1T}(\pi_0)$, let $I(\pi)$ denote the Fisher information matrix:

$$\begin{aligned}
I(\pi) &\equiv T^{-1}E\left[-\frac{\partial^2 \ln L_T(\pi)}{\partial\pi\partial\pi'}\right] \\
&= T^{-1}\sum_{t=1}^{T}\left[\frac{1}{2}\left(\frac{\partial\text{vec}\Sigma_a}{\partial\pi}\right)'\left(\Sigma_a^{-1}\otimes\Sigma_a^{-1}\right)\left(\frac{\partial\text{vec}\Sigma_a}{\partial\pi}\right) + E\left\{\left(\frac{\partial a_t}{\partial\pi}\right)'\Sigma_a^{-1}\left(\frac{\partial a_t}{\partial\pi}\right)\right\}\right]. \qquad (18)
\end{aligned}$$

The above expression has been derived in Klein and Neudecker (2000) and Klein, Mélard, and Zahaf (2000), for example. To simplify notation, let $\bar{v}_{i,j}$ denote the (i,j) entry of the Fisher information matrix $I(\pi)$. $\bar{v}^{i,j}$ denotes the (i,j) entry of its pseudo-inverse $I(\pi)^+$. In addition, let π_i denote the ith component of the parameter vector π, and define the following quantities:

$$\begin{aligned}
\bar{v}_{r,s,t} &\equiv T^{-1}E\left[\frac{\partial\ln L_T(\pi)}{\partial\pi_r}\frac{\partial\ln L_T(\pi)}{\partial\pi_s}\frac{\partial\ln L_T(\pi)}{\partial\pi_t}\right] \\
\bar{v}_{r,s,t,u} &\equiv T^{-1}E\left[\frac{\partial\ln L_T(\pi)}{\partial\pi_r}\frac{\partial\ln L_T(\pi)}{\partial\pi_s}\frac{\partial\ln L_T(\pi)}{\partial\pi_t}\frac{\partial\ln L_T(\pi)}{\partial\pi_u}\right] \\
\bar{\kappa}_{rs,i} &\equiv T^{-1}E\left[\frac{\partial^2\ln L_T(\pi)}{\partial\pi_r\partial\pi_s}\frac{\partial\ln L_T(\pi)}{\partial\pi_i}\right] \\
\bar{v}^{r,s,t} &\equiv \bar{v}_{i,j,k}\bar{v}^{i,r}\bar{v}^{j,s}\bar{v}^{k,t} \\
\bar{v}^{r,s,t,u} &\equiv \bar{v}_{i,j,k,l}\bar{v}^{i,r}\bar{v}^{j,s}\bar{v}^{k,t}\bar{v}^{l,u} \\
\bar{V}_r &\equiv \frac{\partial\ln L_T(\pi)}{\partial\pi_r} \\
\bar{V}_{rs} &\equiv \frac{\partial^2\ln L_T(\pi)}{\partial\pi_r\partial\pi_s} - \bar{v}^{i,j}\bar{\kappa}_{rs,j}\bar{V}_i \\
\bar{V}^{rs} &\equiv \bar{V}_{ij}\bar{v}^{i,r}\bar{v}^{j,s} \\
\bar{V}^{ij}_{kl} &\equiv \left[\bar{V}^{ij}-E(\bar{V}^{ij})\right]\left[\bar{V}_{kl}-E(\bar{V}_{kl})\right]
\end{aligned}$$

We are now ready to state the expression for the Bartlett adjustment $b_{1T}(\pi_0)$ in (16).

Corollary 1. *Let all the assumptions of Theorem 1 hold with* $d = 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}$ *defined as before. Moreover, define the constants:*

$$\begin{aligned}\bar{\rho}_{13}^2 &= d^{-1}\bar{v}^{i,j,k}\bar{v}^{l,m,n}\bar{v}_{i,j}\bar{v}_{k,l}\bar{v}_{m,n}\\ \bar{\rho}_{23}^2 &= d^{-1}\bar{v}^{i,j,k}\bar{v}^{l,m,n}\bar{v}_{i,l}\bar{v}_{j,m}\bar{v}_{k,n}\\ \bar{\rho}_4 &= d^{-1}\bar{v}^{i,j,k,l}\bar{v}_{i,j}\bar{v}_{k,l}.\end{aligned}$$

Then, the Bartlett adjustment $b_{1T}(\pi_0)$ *in* (16) *can be computed from:*

$$b_{1T}(\pi) \equiv \frac{1}{12}\left(3\bar{\rho}_{13}^2 + 2\bar{\rho}_{23}^2 - 3\bar{\rho}_4\right) + \frac{1}{4d}\left[\frac{2}{T}E\left(\bar{V}_{ij}^{ij}\right) - \frac{1}{T}\mathrm{var}\left(\bar{V}_{ij}\bar{v}^{i,j}\right) - 2\mathrm{cov}\left(\frac{1}{T}\bar{V}_i\bar{V}_j\bar{v}^{i,j}, \frac{1}{\sqrt{T}}\bar{V}_{ij}\bar{v}^{i,j}\right)\right].$$

The expression for $b_{1T}(\pi)$ defined above depends on the population moments which are difficult to evaluate analytically. In practice, we can use the sample analogues to obtain a consistent estimate of $b_{1T}(\pi_0)$.

Similar to Theorem 1, the result of Corollary 1 is non-trivial. This again is due to the lack of local identification of the parameter π. Compared to the usual result (e.g., McCullagh and Cox, 1986), two modifications to $b_{1T}(\pi)$ need to be made: first, it is the number d of the identifiable components in the parameter π that enters the formula for the adjustment, and not the dimension d_π of π; and second, since the information matrix $I(\pi_0)$ is singular, the Bartlett adjustment now depends on its pseudo-inverse rather than its inverse.

3.2 Confidence Sets

We now use the results of Theorem 1 and Corollary 1 to construct confidence sets for the deep parameter θ of a DSGE model considered in Example 1. The key property of the proposed sets is that they are fully robust to identification failure of either θ or π.

The starting point of our confidence set construction is the observation that the log-linearized DSGE model solutions take the form of the model in (1) with matrices A, B, C, D and Σ that are known functions of the DSGE deep model parameter θ. Specifically, with θ denoting the deep parameter of the DSGE model, let

$$\pi(\theta) \equiv \left((\mathrm{vec}A(\theta))', (\mathrm{vec}K(\theta))', (\mathrm{vec}C(\theta))', (\mathrm{vech}\Sigma_a(\theta))'\right)'$$

be the parameters of the innovations representation corresponding to the log-linearized DSGE model solution. Note that θ affects the likelihood only through its effect on the state-space parameter π.

We shall hereafter maintain that the deep parameter space Θ is such that the image by $\pi(\cdot)$ of Θ, $\pi(\Theta)$, is a subset of Π, i.e. $\pi(\Theta) \subseteq \Pi$.

We are interested in constructing subsets of the deep parameter space $\Theta \subseteq \mathbb{R}^{d_\theta}$ that cover the true value θ_0 of θ with probability $\geq 1-\alpha$. Our confidence set $CS_{1-\alpha,T}$ is obtained by inverting the likelihood ratio test in the previous section. Specifically, let

$$CS_{1-\alpha,T} \equiv \left\{ \bar{\theta} \in \Theta : \overline{LR}_{1T}(\bar{\theta}) \leqslant \chi^2_{d,1-\alpha} \right\}, \tag{19}$$

where

$$\overline{LR}_{1T}(\bar{\theta}) \equiv 2\left(\sup_{\pi\in\Pi} \ln L_T(\pi) - \ln L_T(\bar{\pi}) \right), \quad \bar{\pi} \equiv \pi(\bar{\theta}),$$

and where $\chi^2_{d,1-\alpha}$ denotes the $1-\alpha$ quantile of a χ^2_d random variable. Put in words, the confidence set $CS_{1-\alpha,T}$ contains all the values $\bar{\theta}$ of the deep parameter θ for which the LR test developed in Section 3.1 fails to reject the null hypothesis $\pi = \bar{\pi}$. The following theorem establishes that, asymptotically, our identification robust confidence set $CS_{1-\alpha,T}$ has confidence level $1-\alpha$.

Theorem 2. *Let Assumptions 1 to 3 hold. Let* $\widetilde{\Theta}$ *be a subset of* Θ *such that* $\pi(\widetilde{\Theta}) \subseteq \widetilde{\Pi}$ *which is a compact subset of* Π*. Then,*

$$\liminf_{T\to\infty} \inf_{\theta\in\widetilde{\Theta}} P_\theta\left(\theta \in CS_{1-\alpha,T}\right) \geq 1-\alpha.$$

Put in words, Theorem 2 shows that, asymptotically, our confidence set $CS_{1-\alpha,T}$ has confidence level $1-\alpha$ uniformly over appropriate subsets of the deep parameter space. If the mapping $\pi(\cdot)$ from the deep parameters to the reduced form parameters is known to be continuous, then any compact subset $\widetilde{\Theta}$ of Θ satisfies the requirement in Theorem 2. Similar to Theorem 1, this result is fully robust to non-identification of either θ or π, and is thus non-trivial.

It is useful to compare the results of Theorem 2 to the results available in the literature. Similar to here, Guerron-Quintana, Inoue, and Kilian (2013) propose a confidence set for the deep parameter θ by inverting a likelihood ratio test. There are however several important differences. First, the setup of Guerron-Quintana, Inoue, and Kilian (2013) requires the existence of an identified reduced form parameter. In our setup, the reduced form parameter is π is that parameter is neither locally nor globally identified. Second, in order to construct the confidence set of Guerron-Quintana, Inoue, and Kilian (2013), one needs to estimate the rank of the transformation that links the deep parameter θ with the reduced form parameter. In our setup, the confidence set is constructed by inverting a likelihood ratio statistics whose limiting distributions is "chi-squared" with known degrees of freedom. Lastly, the confidence set of Guerron-Quintana, Inoue, and Kilian (2013) does not have uniform coverage.

Finally, it is possible to improve the finite sample coverage by constructing a Bartlett corrected confidence set $CS^*_{1-\alpha,T}$ given by:

$$CS^*_{1-\alpha,T} \equiv \left\{\bar{\theta} \in \Theta : \overline{LR^*}_{1T}(\bar{\theta}) \leqslant \chi^2_{d,1-\alpha}\right\},$$

with

$$\overline{LR^*}_{1T}(\bar{\theta}) \equiv \left(1 + \frac{b_{1T}(\pi(\bar{\theta}))}{T}\right)^{-1} LR_{1T}(\bar{\theta}),$$

with Bartlett correction b_{1T} as previously computed in Corollary 1.

3.3 Explicit Parameter Restrictions

When a DSGE model is correctly specified, its first-order solution takes the form in (1) with matrices A, B, C, D and Σ that are known functions of the DSGE deep model parameter θ. We are now interested in testing the validity of these restrictions. The idea of using the LR to test the restrictions imposed by the theory can be traced back to Sargent (1977, 1978); see also Christiano (2007) for a more recent example. The key idea is simple: embed the DSGE model in a larger model and use the LR test to test if the parameters of that larger model satisfy the restrictions predicted by the DSGE theory. Specifically, with θ denoting the deep parameter of the DSGE model, let

$$\pi(\theta) \equiv \big((\text{vec}A(\theta))', (\text{vec}K(\theta))', (\text{vec}C(\theta))', (\text{vech}\Sigma_a(\theta))'\big)'$$

be the parameters of the innovations representation corresponding to the log-linearized DSGE model solution. Note that θ affects the likelihood only through its effect on the state-space parameter π. The likelihood ratio test statistic now takes the form:

$$LR_{2T} \equiv 2\left(\sup_{\pi\in\Pi} \ln L_T(\pi) - \sup_{\theta\in\Theta} \ln L_T(\pi(\theta))\right), \tag{20}$$

where Θ is the parameter space for the deep parameter θ, and the parameter space Π for π is as defined in (14).

Formally, the null hypothesis of correct DSGE model specification takes the form of explicit parameter restrictions,

$$H_2: \quad \pi \in \Pi_0, \quad \Pi_0 \equiv \{\pi \in \Pi : \pi = \pi(\theta), \theta \in \Theta\}.$$

As before, the unrestricted parameter space Π consists of all matrices (A, K, C, Σ_a) such that A and $A - KC$ are stable, (A, K) is controllable, and (A, C) is observable; it is a subset of $\mathbb{R}^{d_\pi}$. The deep parameter space Θ is a subset of $\mathbb{R}^{d_\theta}$. We shall work under the additional assumptions that the mapping from θ to π is smooth, and that the restricted model is identified. In what follows, let θ_0 denote the true value of θ under the null hypothesis H_2.

Assumption 4. *(i) the mapping $\pi : \theta \mapsto \pi(\theta)$ is twice continuously differentiable on Θ; (ii) θ_0 is identified with $\bar{I}(\theta_0)$ full rank, where $\bar{I}(\theta) \equiv T^{-1} E\left[-\frac{\partial^2 \ln L_T(\pi(\theta))}{\partial\theta\partial\theta'}\right]$.*

When the analytic forms of the system matrices $(A(\theta), K(\theta), C(\theta), \Sigma_a(\theta))$ are available, then the smoothness Assumption 4(i) can be checked analytically. It is worth pointing out that this condition puts additional restrictions on Θ, typically restricting the deep parameter space to be an open set (so that θ is never on the boundary of that set). Assumption 4(ii) requires that the restricted model Π_0 be identified. This requirement in particular guarantees that the regularity conditions for the maximum likelihood estimation of the restricted model parameter θ are met.

Theorem 3. *Let Assumptions 1 to 4 hold. As before, $d_\theta = \dim(\theta)$ and $d = 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}$. Then under H_2,*

$$LR_{2T} \xrightarrow{d} \chi^2_{d-d_\theta}.$$

Since the unrestricted model parameters π are neither globally nor locally identified, the result of Theorem 3 is non-trivial. Put in words, Theorem 3 states that the number of degrees of freedom in the asymptotic χ^2 distribution of the likelihood ratio statistic depends on the difference between the number d of the identifiable components in π, and the dimension of the identifiable components in θ. Since θ is assumed to be identified, the number of its identifiable components simply equals its dimension.

Similar to previously, the accuracy of the asymptotic chi-squared approximation in Theorem 3 can be improved by adjusting the statistic LR_{2T}. For an appropriately defined constant b_{2T}, letting

$$LR^*_{2T} \equiv \left(1 + \frac{b_{2T}}{T}\right)^{-1} LR_{2T}, \tag{21}$$

results in a corrected likelihood ratio statistic whose distribution is order T^{-2} away from that of $\chi^2_{d-d_\theta}$,

$$\Pr\left(LR^*_{2T} \leqslant r\right) = \Pr\left(\chi^2_{d-d_\theta} \leqslant r\right) + O(T^{-2}).$$

Corollary 2. *Let all the assumptions of Theorem 3 hold with $d = 2n_X n_Y + \frac{n_Y(n_Y+1)}{2}$ defined as before. Moreover, define $b_{\theta,1T}(\theta)$ exactly as $b_{1T}(\pi)$ in Corollary (1) with $(L_T(\pi), \pi, d)$ replaced by*

$(L_T(\pi(\theta)), \theta, d_\theta)$. *Let* $\pi_0 = \pi(\theta_0)$. *Then, the Bartlett adjustment* b_{2T} *in* (21) *can be computed as:*

$$b_{2T} = b_{1T}(\pi_0) - b_{\theta,1T}(\theta_0).$$

As before, the computation of the Bartlett adjustment requires replacing various population moments by their sample counterparts. In addition, θ_0 and $\pi_0 = \pi(\theta_0)$ now need to be replaced by their consistent estimates $\widehat{\theta}_T = \text{argmax}_{\theta \in \Theta} L_T(\pi(\theta))$ and $\pi(\widehat{\theta}_T)$. That the latter are consistent follows by the standard arguments given that the restricted model Π_0 is assumed to be identified.

4 Monte Carlo Experiment

4.1 Simple RBC Model

We consider a widely used example from the RBC theory: Hansen's (1985) indivisible-labor model. Versions of this model have been estimated by numerous authors using a variety of techniques. See, for example, Christiano and Eichenbaum (1992), Burnside, Eichenbaum, and Rebelo (1993), Ireland (2004), and Ruge-Murcia (2007), among others. In the model, there is a continuum of identical infinitely lived households who conditional on the information available at time $t = 0$ maximize the expected value of $\sum_{t=0}^{\infty} \beta^t u(C_t, L_t)$, where $0 < \beta < 1$ is the discount factor, C_t denotes the time t consumption, L_t is the time t leisure, and the within period preferences are given by:

$$u(C_t, L_t) = \ln C_t + \vartheta L_t, \qquad \vartheta > 0.$$

Output at time t, denoted by Q_t, is produced by a single firm via a Cobb-Douglas production function

$$Q_t = A_t K_t^{1-\alpha} (\gamma^t H_t)^\alpha,$$

where $0 < \alpha < 1$, K_t is the capital stock at the beginning of period t, $H_t = 1 - L_t$ are the hours worked (the representative agent's time endowment being normalized to one), $\gamma \geq 1$ is the constant unconditional growth rate of technology, and A_t is an aggregate shock to technology which is assumed to follow a first-order autoregressive process

$$\ln A_t = (1 - \rho) \ln a + \rho \ln A_{t-1} + \varepsilon_t,$$

with $|\rho| < 1$, $a > 0$, and iid innovations $\varepsilon_t \sim N(0, \sigma^2)$. Output, which is produced by the firm and sold to the households, can either be consumed (C_t) or invested (I_t), which yields the resource constraint $Q_t = C_t + I_t$. The law of motion for the capital stock is given by

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

where $0 < \delta < 1$ governs the depreciation rate on capital.

By making appropriate substitutions, one can solve the above model as a dynamic optimization problem with decision variables C_t, H_t, and I_t, and state variables K_t and A_t. That is, given the beginning of period capital stock and the current technology shock, households choose consumption, labor and investment. In the nonstochastic steady state of this economy, Q_t, K_t, and C_t all grow at rate γ, while H_t is constant. Writing the equilibrium conditions for the detrended variables $q_t = Q_t/\gamma^t$, $k_t = K_t/\gamma^t$, $c_t = C_t/\gamma^t$, $h_t = H_t$ and $a_t = A_t$, then log-linearizing around the steady state $(q_t, k_t, c_t, h_t, a_t) = (q, k, c, h, a)$, and solving the resulting system using the Blanchard-Kahn procedure leads to the following representation:

$$\begin{pmatrix} \bar{k}_{t+1} \\ \bar{a}_{t+1} \end{pmatrix} = \Pi \begin{pmatrix} \bar{k}_t \\ \bar{a}_t \end{pmatrix} + W\varepsilon_{t+1}, \qquad \begin{pmatrix} \bar{q}_t \\ \bar{c}_t \\ \bar{h}_t \end{pmatrix} = U \begin{pmatrix} \bar{k}_t \\ \bar{a}_t \end{pmatrix}$$

in log-deviation variables $\bar{q}_t = \ln(q_t/q)$, $\bar{k}_t = \ln(k_t/k)$, $\bar{c}_t = \ln(c_t/c)$, $\bar{h}_t = \ln(h_t/h)$, and $\bar{a}_t = \ln(a_t/a)$. The analytic expression of the matrices Π, W and U, expressed as functions of the model parameters, can be found, for example, in Ireland (2004). Now say that the econometrician observes realizations of log-deviated output, consumption and hours subject to an additive iid measurement error $v_t \sim N(0, \Sigma)$. Then, letting θ denote the vector of deep parameters, $\theta \equiv (\beta, \vartheta, \alpha, \gamma, \delta, a, \rho, \sigma, \Sigma)$, the empirical model can be written as:

$$X_{t+1} = \underbrace{\Pi}_{A(\theta)} X_t + \underbrace{\begin{pmatrix} W & 0 \end{pmatrix}}_{B(\theta)} \epsilon_{t+1}, \quad Y_{t+1} = \underbrace{U\Pi}_{C(\theta)} X_t + \underbrace{\begin{pmatrix} UW & \mathrm{Id} \end{pmatrix}}_{D(\theta)} \epsilon_{t+1}, \tag{22}$$

where $X_t \equiv \left(\bar{k}_t, \bar{a}_t\right)'$ is the state vector, $Y_t \equiv \left(\bar{q}_t^{obs}, \bar{c}_t^{obs}, \bar{h}_t^{obs}\right)'$ are the observables, and $\epsilon_t \equiv (\varepsilon_t, v_t')' \sim N(0, \mathrm{diag}(\sigma^2, \Sigma))$ is the disturbance in the model.

4.2 Size Experiment

We use the above model to examine the small sample properties of our test. To simulate data, we choose true parameter values in line with those originally suggested by Hansen (1985). We set β to $\beta_0 = 0.95$; the parameter ϑ in the utility function is set to $\vartheta_0 = 2$; the parameter α is set to $\alpha_0 = 0.64$ which corresponds to capital's share in production $1 - \alpha$ of 0.36; the technology growth rate parameter γ is set to $\gamma_0 = 1.0041$ in line with the values found in Eichenbaum (1991); the rate of depreciation of capital is set to $\delta_0 = 0.025$; finally, the technology shock parameters a, ρ, and σ are set to $a_0 = 2.7818$, $\rho_0 = 0.95$ and $\sigma_0 = 0.04$ in line with the values found in Burnside, Eichenbaum, and Rebelo (1993). These values result in the steady state hours worked of $h = 0.367$ (which matches the observation that individuals spend 1/3 of their time engaged in market activities). The covariance matrix Σ is set to be diagonal, with standard deviation $v = 0.02$ of the measurement errors in output, consumption and hours. The true values of the data generating process are summarized below:

$$\beta_0 = 0.95, \quad \vartheta_0 = 2, \quad 1 - \alpha_0 = 0.36, \quad \gamma_0 = 1.0041, \quad \delta_0 = 0.025$$
$$a_0 = 2.7818, \quad \rho_0 = 0.95, \quad \sigma_0 = 0.04, \quad \Sigma_0 = v^2 \mathrm{Id} \text{ with } v = 0.02$$

In all the simulations, we only use observed values of output $\bar{q}_t^{obs}$ and consumption $\bar{c}_t^{obs}$ in order to estimate the parameters of the model. The main reason why we only use two observables is to keep the number of parameters in the unrestricted model not too large. For the restricted model, the estimated deep parameter is: $\theta = (\beta, \rho, \sigma, v_{11}, v_{12}, v_{22})$, where v_{11}, v_{12}, v_{22} are the parameters of the Cholesky decomposition of Σ. All other parameters are kept fixed. We generate time series of length $T = 150$ and $T = 250$, and draw $N = 4000$ Monte Carlo samples. We construct the LR test statistics for the following three null hypotheses:

$$H_0: \quad \theta = \theta_0 \qquad \text{vs.} \qquad \theta \text{ unrestricted}$$
$$H_1: \quad \pi = \pi_0 \qquad \text{vs.} \qquad \pi \text{ unrestricted}$$
$$H_2: \quad \pi = \pi(\theta) \qquad \text{vs.} \qquad \pi \text{ unrestricted}$$

The value π_0 in H_1 is set to $\pi_0 = \pi(\theta_0)$. We consider the original LR statistic as well as the Bartlett adjusted one to correct for the relatively small sample sizes ($T = 150$ and $T = 250$). Table 1 summarizes our results.

Table 1: Empirical Size

Null	Unadjusted, $T = 150$	Adjusted, $T = 150$	Unadjusted, $T = 250$	Adjusted, $T = 250$
H_0	0.1258	0.1085	0.1168	0.1098
H_1	0.1685	0.1283	0.1263	0.1115
H_2	0.1663	0.1298	0.1240	0.1035

Empirical sizes of our LR tests with ("Adjusted") and without ("Unadjusted") Bartlett adjustments. Nominal size: 10%.

Table 1 shows that the unadjusted LR tests of the three null hypotheses are oversized, with size distortions that are more serious in smaller sample sizes ($T = 150$). The Bartlett correction provides an improvement in all cases, though the magnitude of the improvements is better for larger sample sizes ($T = 250$).

We also compare the results of our LR specification test (H_2) with those of the LR tests based on finite order VARs. That is, we also consider testing H_2 by assuming that the unrestricted model is a finite order VAR instead of the state space model (i.e. a VARMA). Table 2 summarizes the findings obtained for various values of the maximum VAR lag. We only report the empirical sizes of LR tests obtained without Bartlett adjustment. Since the unrestricted model is misspecified (a finite lag VAR instead of a VARMA), we would expect serious size distortions. Table 2 confirms this. The findings suggest that one would need to work with a VAR(11) or a VAR(12) to obtain acceptable rejection probabilities. This comes at important computational costs which are due to the number of parameters in the unrestricted model: 47 for a VAR(11) and 51 for a VAR(12).

Table 2: Empirical Sizes of VAR based tests

VAR lag	Size	# Parameters
1	0.0000	7
2	0.0000	11
3	0.0003	15
4	0.0003	19
5	0.0070	23
6	0.0160	27
7	0.0330	31
8	0.0532	35
9	0.0685	39
10	0.0835	43
11	0.0975	47
12	0.1105	51
13	0.1243	55
14	0.1365	59
15	0.1472	63

Empirical sizes of the LR tests for H_2 that assume a finite lag VAR. "VAR lag" refers to the maximum lag, "# Parameters" is the number of parameters in the unrestricted model. Nominal size: 10%. $T = 250$.

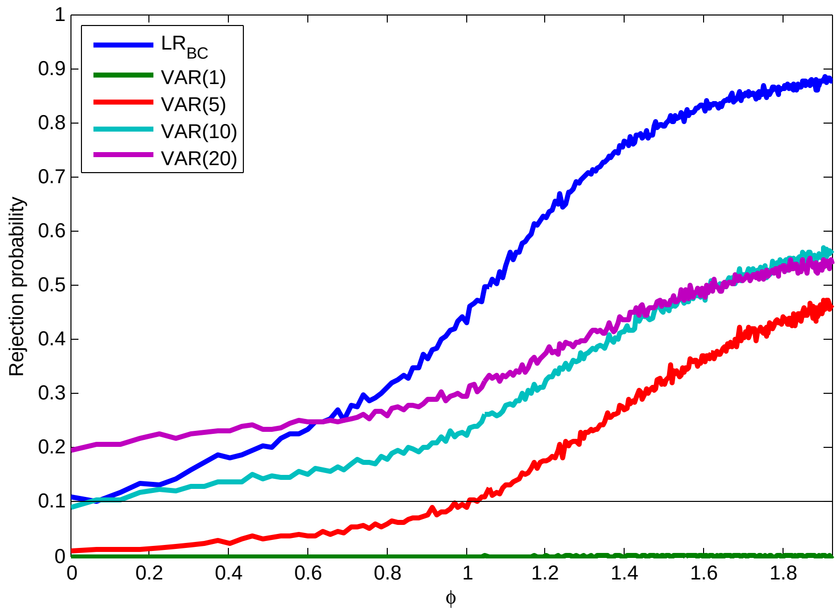
4.3 Power Experiment

To evaluate the empirical power properties of our LR specification test (i.e. our LR test of the null hypothesis H_2), we generate the data under alternative RBC model specifications. These specifications suppose that within period preferences of any household are:

$$u(C_t, L_t) = \ln C_t + \vartheta \frac{L_t^{1-\varphi} - 1}{1-\varphi}, \qquad \vartheta > 0, \quad \varphi \geq 0.$$

These preferences are a special case of the separable preferences considered in King, Plosser, and Rebelo (1988). They nest Hansen's (1985) indivisible labor specification $u(C_t, L_t) = \ln C_t + \vartheta L_t$

Figure 1: Rejection probability of LR tests. Nominal size: 10%. $T = 250$.



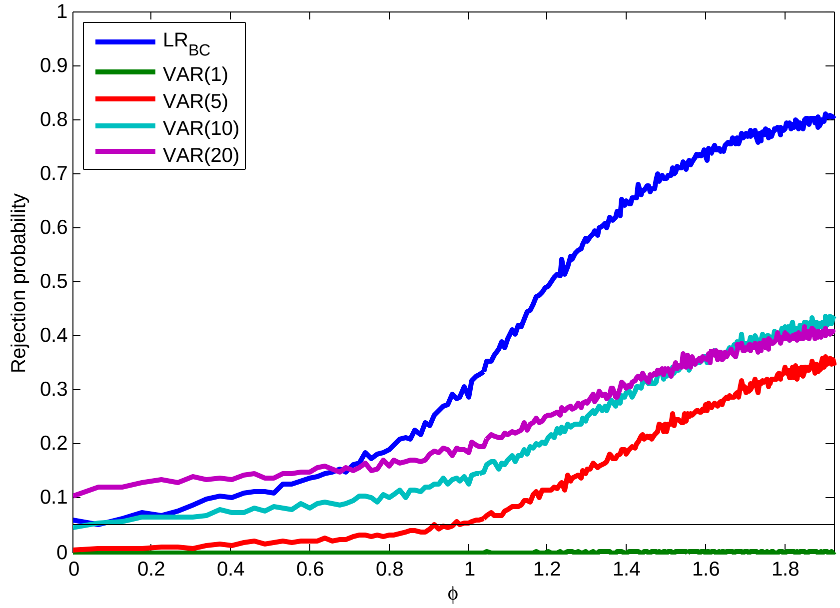
obtained when $\varphi = 0$. The case $\varphi = 1$ corresponds by l'Hôpital rule to $u(C_t, L_t) = \ln C_t + \vartheta \ln L_t$, which is the divisible labor model specification of Hansen (1985). All the other components of the model are as before. For general values of $\varphi \geq 0$, the model can no longer be solved analytically. We instead use the Blanchard-Kahn procedure whose details are reported in Appendix A. Steady state hours h now enter the dynamics through a parameter:

$$\phi = \varphi \frac{h}{1-h}.$$

Thus, the deviations from the null hypothesis can be measured by the deviations of ϕ from 0. Figure 1 plots the rejection probability of our Bartlett adjusted LR specification test (i.e. our LR test of the null hypothesis H_2). We also compare our test to the tests based on finite lag VARs.

As expected, our LR test has good power that increases to 1 as ϕ gets large. For comparison, we report the power of LR tests based on finite lag VAR unrestricted models: a VAR(10) based test that has good size (0.0825 empirical size for a test with 10% nominal size), has bad power properties with power increasing to around 50% as ϕ gets large. Thus, the misspecification of the unrestricted model (a finite lag VAR instead of a VARMA) affects not only the size but also the power of the LR test. Similar behavior occurs with tests that have 5% nominal size.

Figure 2: Rejection probability of LR tests. Nominal size: 5%. $T = 250$.



5 Conclusion

This papers considers the problem of likelihood based inference in linear Gaussian state space models. We derive the asymptotic distribution of the LR test statistic for two types of null hypotheses: a simple null and a null of explicit parameter restrictions. To address the issue of small sample sizes typically encountered in macroeconomic applications, we also derive the Bartlett adjustment factors to the LR test statistics. The key features of our results are: (i) we take into account the non-identification of the unrestricted model; (ii) the asymptotic distributions are chi-squared with the number of degrees of freedom which are known and need not be estimated; (iii) the Bartlett adjustments can be computed as usual, provided pseudo-inverses and correct dimensions of "free" components in the parameter vectors are used. A Monte Carlo examination of the small sample properties of our test in the context of DSGE models suggests that the Bartlett adjustments are useful at sample sizes typically encountered in macroeconomics.

References

Altug, S. (1989): "Time to Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30:4, 889–920.

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26: 2-4, 113–172.

Anderson, G., and G. Moore (1985): "A linear algebraic procedure for solving linear perfect foresight models," *Economic Letters*, 17(3), 247–252.

Andrews, D. W. K. (1987): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

Andrews, D. W. K., and X. Cheng (2012): "Estimation and Inference With Weak, Semi-Strong, and Strong Identification," *Econometrica*, 80(5), 2153–2211.

Andrews, D. W. K., X. Cheng, and P. Guggenberger (2017): "Generic Results for Establishing the Asymptotic Size of Confidence Sets and Tests," unpublshed.

Andrews, D. W. K., and P. Guggenberger (2015): "Identification- and Sigularity-Robust Inference for Moment Condition Models," unpublished manuscript, Yale University and Penn State University.

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

Barndorff-Nielsen, O. E., and P. Hall (1988): "On the Level-Error After Bartlett Adjustment of the Likelihood Ratio Statistic," *Biometrika*, 75(2), 374–378.

Bartlett, M. S. (1937): "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 160(901), 268–282.

Beran, R. (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83(403), 687–697.

Burnside, C., M. Eichenbaum, and S. Rebelo (1993): "Labor Hoarding and the Business Cycle," *Journal of Political Economy*, 101(2), pp. 245–273.

Canova, F. (1994): "Statistical inference in calibrated models. (In: Special Issue on Calibration Techniques and Econometrics," *Journal of Applied Econometrics*, 9(Supplement), S123–S144.

——— (1995): "Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies," *International Economic Review*, 36(2), pp. 477–501.

Christiano, L., and M. Eichenbaum (1992): "Liquidity Effects and the Monetary Transmissions Mechanisms," *American Economic Review*, 82:2, 346–353.

Christiano, L. J. (2007): "[On the Fit of New Keynesian Models]: Comment," *Journal of Business & Economic Statistics*, 25(2), pp. 143–151.

Cordeiro, G. M., and F. Cribari-Neto (2014): *An Introduction to Bartlett Correction and Bias Reduction*. Springer.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A theory of the term structure of interest rates," 53, 385407.

Dagenais, M. G., and J.-M. Dufour (1991): "Invariance, Nonlinear Models, and Asymptotic Tests," *Econometrica*, 59(6), pp. 1601–1615.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business & Economic Statistics*, 25(2), pp. 123–143.

Diebold, F. X., L. E. Ohanian, and J. Berkowitz (1998): "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *The Review of Economic Studies*, 65(3), pp. 433–451.

Dufour, J.-M., L. Khalaf, and M. Kichian (2013): "Identification-robust analysis of {DSGE} and structural macroeconomic models," *Journal of Monetary Economics*, 60(3), 340 – 350.

Eichenbaum, M. (1991): "Real business-cycle theory: Wisdom or whimsy?," *Journal of Economic Dynamics and Control*, 15(4), 607 – 626.

Engle, R., and M. Watson (1981): "A one-factor multivariate time series model of metropolitan wage rates," *Journal of the American Statistical Association*, 76(376), 774–781.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2004): "Comparing dynamic equilibrium models to data: a Bayesian approach," *Journal of Econometrics*, 123(1), 153 – 187.

Geweke, J. (1977): "The dynamic factor analysis of economic time series," *Latent variables in socio-economic models*.

Gourieroux, C., and A. Monfort (1995): *Statistics and econometric models*, vol. 1. Cambridge University Press.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist inference in weakly identified dynamic stochastic general equilibrium models," *Quantitative Economics*, 4(2), 197–229.

Hamilton, J. D., and J. C. Wu (2012): "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168(2), 315 – 331.

Hannan, E. E. J., and M. Deistler (1988): *The statistical theory of linear systems*, vol. 70. SIAM.

Hansen, G. D. (1985): "Indivisible labor and the business cycle," *Journal of Monetary Economics*, 16(3), 309 – 327.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, 2(0), 7 – 46.

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

Hazewinkel, M. (1977): "Moduli and Canonical Forms for Linear Dynamical Systems II: The Topological Case," *Mathematical Systems Theory*, 10, 363–385.

Hazewinkel, M. (1979): *On identification and the geometry of the space of linear systems*. Springer.

Ireland, P. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics & Control*, 28, 1205–1226.

Kailath, T., A. H. Sayed, and B. Hassibi (2000): *Linear estimation*, vol. 1.

Khalaf, L., Z. Lin, and A. Reza (2014): "Exact Inference in DSGE Models," unpublished manuscript, Carleton University and Bank of Canada.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988): "Production, growth and business cycles: I. The basic neoclassical model," *Journal of Monetary Economics*, 21(23), 195 – 232.

King, R. G., and M. W. Watson (2002): "System Reduction and Solution Algorithms for Singular Linear Difference Systems under Rational Expectations," *Computational Economics*, 20, pp. 57–86.

Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business & Economic Statistics*, 27(3), 293–311.

Klein, A., G. Mélard, and T. Zahaf (2000): "Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules," *Linear Algebra and its Applications*, 321, 209 – 232, Eighth Special Issue on Linear Algebra and Statistics.

Klein, A., and H. Neudecker (2000): "A Direct Derivation of the Exact Fisher Information Matrix of Gaussian Vector State Space Models," *Linear Algebra and its Applications*, 321, 233–238.

Klein, P. (2000): "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Jornal of Economic Dynamics and Control*, 4:2, 257–271.

Komunjer, I., and S. Ng (2011): "Dynamic identification of dynamic stochastic general equilibrium models," *Econometrica*, 79(6), 1995–2032.

Lawley, D. N. (1956): "A General Method for Approximating to the Distribution of Likelihood Ratio Criteria," *Biometrika*, 43(3/4), 295–303.

Le, V. P. M., D. Meenagh, P. Minford, 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(12), 2078 – 2104, Frontiers in Structural Macroeconomic Modeling.

Lee, J. (2012): *Introduction to smooth manifolds*, vol. 218. Springer.

Lehmann, E. L., and J. P. Romano (2006): *Testing statistical hypotheses*. springer.

Lindsay, B. G. (1995): "Mixture Models: Theory, Geometry and Applications," *NSF-CBMS Regional Conference Series in Probability and Statistics*, 5, i–163.

Liu, X., and Y. Shao (2003): "Asymptotics for likelihood ratio tests under loss of identifiability," *Ann. Statist.*, 31(3), 807–832.

Lütkepohl, H. (2005): *New introduction to multiple time series analysis*. Cambridge Univ Press.

Lutkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin.

McCullagh, P., and D. Cox (1986): "Invariants and Likelihood Ratio Statistics," *The Annals of Statistics*, 14(4), 1419–1430.

Piazzesi, M. (2010): "Affine Term Structure Models," in *Handbook of Financial Econometrics: Tools and Techniques*, ed. by Y. Aït-Sahalia, and L. P. Hansen, vol. 1, pp. 691 – 766. North-Holland.

Qu, Z. (2014): "Inference in dynamic stochastic general equilibrium models with possible weak identification," *Quantitative Economics*, 5(2), 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(1), 95–132.

Rabanal, P., and J. F. Rubio-Ramírez (2005): "Comparing New Keynesian models of the business cycle: A Bayesian approach," *Journal of Monetary Economics*, 52(6), 1151 – 1166, The econometrics of the New Keynesian price equation.

Ravenna, F. (2007): "Vector Autoregressions and Reduced Form Representations of DSGE Models," *Jornal of Monetary Economics*, 54:7, 2048:2064.

Rothenberg, T. (1971): "Identification in Parametric Models," *Econometrica*, 39:3, 577–591.

Rubio-Ramírez, J. F., and J. Fernandez-Villaverde (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059–1087.

Ruge-Murcia, F. J. (2007): "Methods to Estimate Dynamic Stochastic General Equilibrium Models," *Journal of Economic Dynamics and Control*, 31, 2599–2636.

Sargent, T. (1989a): "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, 97:2.

Sargent, T. J. (1977): "The Demand for Money during Hyperinflations under Rational Expectations: I," *International Economic Review*, 18(1), pp. 59–82.

——— (1978): "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," *Journal of Political Economy*, 86(6), pp. 1009–1044.

Sargent, T. J. (1989b): "Two models of measurements and the investment accelerator," *Journal of Political Economy*, 97(2), 251–287.

Sargent, T. J., C. A. Sims, et al. (1977): "Business cycle modeling without pretending to have too much a priori economic theory," *New methods in business cycle research*, 1, 145–168.

Schorfheide, F. (2000): "Loss function-based evaluation of DSGE models," *Journal of Applied Econometrics*, 15(6), 645–670, cited By (since 1996)148.

Schorfheide, F. (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," in *Advances in Economics and Econometrics*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, vol. 3, pp. 184–230. Cambridge University Press, Cambridge Books Online.

Sims, C. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, pp. 1–20.

Smith, A. (1993): "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8:S1, S63–S84.

Stock, J. H., and M. Watson (2011): "Dynamic factor models," *Oxford Handbook on Economic Forecasting.*

Stock, J. H., and M. W. Watson (1989): "New indexes of coincident and leading economic indicators," *NBER macroeconomics annual*, 4, 351–394.

Uhlig, H. (1999): "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily," in *Computational Methods for the Study of Dynamic Economies*, ed. by R. Marimon, and A. Scott, pp. 30–61. Oxford University Press.

Vasicek, O. (1977): "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5, 177188.

Watson, M. (1993): "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101:6, 1011–1041.

A RBC Extension

Suppose that within period preferences of any household are:

$$u(C_t, L_t) = \ln C_t + \vartheta \frac{L_t^{1-\varphi} - 1}{1-\varphi},$$

where $\varphi \geq 0$. These preferences are a special case of the separable preferences considered in King, Plosser, and Rebelo (1988). They nest Hansen's (1985) indivisible labor specification $u(C_t, L_t) = \ln C_t + \vartheta L_t$ obtained up to a constant when $\varphi = 0$. The case $\varphi = 1$ corresponds by l'Hôpital rule to $u(C_t, L_t) = \ln C_t + \vartheta \ln L_t$, which is the divisible labor model specification of Hansen (1985). All the other components of the model are as before.

A.1 Equilibrium Conditions

The new equilibrium conditions describing this economy are:

$$\begin{aligned}
\frac{1}{C_t} &= \beta E_t \left[\frac{1}{C_{t+1}} \left((1-\alpha)\frac{Q_{t+1}}{K_{t+1}} + (1-\delta) \right) \right] \\
\vartheta C_t (1-H_t)^{-\varphi} &= \alpha \frac{Q_t}{H_t} \\
K_{t+1} &= Q_t + (1-\delta)K_t - C_t \\
Q_t &= A_t K_t^{1-\alpha} (\gamma^t H_t)^\alpha
\end{aligned}$$

In the nonstochastic steady state of this economy, Q_t, K_t, and C_t all grow at rate γ, while H_t is constant. Using lowercase letters to denote detrended variables (e.g., $q_t = Q_t/\gamma^t$), the equilibrium variables $q_t = Q_t/\gamma^t$, $k_t = K_t/\gamma^t$, $c_t = C_t/\gamma^t$, $h_t = H_t$ and $a_t = A_t$ solve the system of equations

$$\begin{aligned}
\frac{\gamma}{c_t} &= \beta E_t \left[\frac{1}{c_{t+1}} \left((1-\alpha)\frac{q_{t+1}}{k_{t+1}} + (1-\delta) \right) \right] \\
\vartheta c_t (1-h_t)^{-\varphi} &= \alpha \frac{q_t}{h_t} \\
\gamma k_{t+1} &= q_t + (1-\delta)k_t - c_t \\
q_t &= a_t k_t^{1-\alpha} h_t^\alpha \\
\ln a_t &= (1-\rho)\ln a + \rho \ln a_{t-1} + \epsilon_t
\end{aligned} \tag{23}$$

A.2 Steady State

Let (q,k,c,h,a) denote the the steady-state values of (q_t,k_t,c_t,h_t,a_t). We have

$$\begin{aligned}
q &= A^{1/\alpha} \left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta} \right]^{(1-\alpha)/\alpha} h \\
k &= q \left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta} \right] \\
c &= q \left[1 - \frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta} \right] \\
\frac{(1-h)^\varphi}{h} &= \frac{\vartheta}{\alpha} \left[1 - \frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta} \right] \\
a &= A
\end{aligned} \tag{24}$$

Note that unlike in the indivisible labor model, the stead-state value h can no longer be solved for analytically. There is however always a unique solution $h \in (0,1)$ to the above equation since the function $(1-h)^\varphi/h$ is strictly decreasing on $(0,1)$ and onto $(0,+\infty)$. Once h is solved for, the values for (q,k,c) follow immediately from the first three equations in (24).

A.3 Log-linearized Equations

Let $\bar{q}_t = \ln(q_t/q)$, $\bar{k}_t = \ln(k_t/k)$, $\bar{c}_t = \ln(c_t/c)$, $\bar{h}_t = \ln(h_t/h)$, and $\bar{a}_t = \ln(a_t/a)$. Log-linearizing the equilibrium equations (23) around the steady state $(q_t, k_t, c_t, h_t, a_t) = (q, k, c, h, a)$ leads to the following equations:

$$
\begin{aligned}
(\gamma/\beta)E_t[\bar{c}_{t+1}] - (\gamma/\beta)\bar{c}_t &= (\gamma/\beta - 1 + \delta)\, E_t[\bar{q}_{t+1}] - (\gamma/\beta - 1 + \delta)\bar{k}_{t+1} \\
\bar{c}_t + \left[1 + \varphi\frac{h}{1-h}\right]\bar{h}_t &= \bar{q}_t \\
\gamma\left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta}\right]\bar{k}_{t+1} &= \bar{q}_t + (1-\delta)\left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta}\right]\bar{k}_t - \left[1 - \frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta}\right]\bar{c}_t \\
\bar{q}_t &= \bar{a}_t + (1-\alpha)\bar{k}_t + \alpha\bar{h}_t \\
\bar{a}_t &= \rho\bar{a}_{t-1} + \epsilon_t
\end{aligned}
\tag{25}
$$

Note that whenever $\varphi \neq 0$ the second equation depends on the steady state hours h. Thus, unlike in the case of indivisible labor, the dynamics of the model now depend on the utility parameter ϑ. To put the log-linearized equations (25) in matrix form, let

$$
\kappa \equiv \gamma/\beta - 1 + \delta > 0, \quad \lambda \equiv \gamma - 1 + \delta > 0, \quad \text{and} \quad \psi \equiv \varphi h/(1-h) \geq 0,
$$

and define

$$
s_t \equiv (\bar{k}_t, \bar{c}_t)' \quad \text{and} \quad f_t \equiv (\bar{q}_t, \bar{h}_t)'.
$$

Then the second and fourth equations in (25) can be written as:

$$
\underbrace{\begin{pmatrix} 1 & -(1+\psi) \\ 1 & -\alpha \end{pmatrix}}_{A} f_t = \underbrace{\begin{pmatrix} 0 & 1 \\ 1-\alpha & 0 \end{pmatrix}}_{B} s_t + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{C} \bar{a}_t,
$$

with $\det A = 1 - \alpha + \psi > 0$ for any $0 < \alpha < 1$, $\varphi \geq 0$, and $0 < h < 1$. Moreover, the first and third equations in (25) can be written as:

$$
\underbrace{\begin{pmatrix} \kappa & \gamma/\beta \\ \gamma(1-\alpha)/\kappa & 0 \end{pmatrix}}_{D} E_t[s_{t+1}] + \underbrace{\begin{pmatrix} -\kappa & 0 \\ 0 & 0 \end{pmatrix}}_{F} E_t[f_{t+1}] = \underbrace{\begin{pmatrix} 0 & \gamma/\beta \\ (1-\delta)(1-\alpha)/\kappa & -(1-(1-\alpha)\lambda/\kappa) \end{pmatrix}}_{G} s_t + \underbrace{\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}}_{H} f_t,
$$

so combining everything we get:

$$
(D + FA^{-1}B)E_t[s_{t+1}] = (G + HA^{-1}B)s_t + (HA^{-1}C - \rho FA^{-1}C)\bar{a}_t, \tag{26}
$$

where we have used the fact that $E_t[\bar{a}_{t+1}] = \rho\bar{a}_t$, as implied by the last equality in (25).

A.4 The Blanchard-Kahn Procedure

Looking back at (26), the matrix

$$D + FA^{-1}B = \begin{pmatrix} \kappa\frac{\alpha\psi}{1-\alpha+\psi} & \frac{\gamma}{\beta} + \kappa\frac{\alpha}{1-\alpha+\psi} \\ \frac{\gamma(1-\alpha)}{\kappa} & 0 \end{pmatrix},$$

is invertible, because

$$\det(D + FA^{-1}B) = -\frac{\gamma(1-\alpha)}{\beta\kappa(1-\alpha+\psi)}\left[\underbrace{\gamma(1+\psi)}_{\geq 1} - \underbrace{\beta(1-\delta)\alpha}_{<1}\right] < 0.$$

Then let

$$K \equiv (D + FA^{-1}B)^{-1}(G + HA^{-1}B)$$
$$L \equiv (G + HA^{-1}B)^{-1}(HA^{-1}C - \rho FA^{-1}C).$$

Next, we show that the matrix K has one eigenvalue outside the unit circle and one eigenvalue inside the unit circle, implying that the system has a unique solution. The solution can then be obtained following, for example, the steps in Ireland (2004). For this, it suffices to show that $\det(K - \text{Id}) < 0$. Now,

$$K - \text{Id} = (D + FA^{-1}B)^{-1}[G - D + (H - F)A^{-1}B],$$

so

$$\det(K - \text{Id}) = \frac{\det(G - D + (H - F)A^{-1}B)}{\det(D + FA^{-1}B)}.$$

Since we have already shown that the denominator is strictly negative, it suffices to show that the numerator is strictly positive. Now,

$$G - D + (H - F)A^{-1}B = \begin{pmatrix} -\kappa\frac{\alpha\phi}{1-\alpha+\phi} & -\kappa\frac{\alpha}{1-\alpha+\phi} \\ (1-\alpha)\left(\frac{(1-\gamma-\delta)}{\kappa} + \frac{1}{1-\alpha+\phi}\right) & -\left(1 - \frac{(1-\alpha)\lambda}{\kappa}\right) - \frac{\alpha}{1-\alpha+\phi} \end{pmatrix},$$

and

$$\det(G - D + (H - F)A^{-1}B) = \alpha\left[\gamma\left(1/\beta - 1\right) + \alpha\lambda\right](1 + \phi) > 0.$$

Thus, $\det(K - \text{Id}) < 0$ which implies that K has one eigenvalue outside the unit circle and one eigenvalue inside the unit circle.

B Proofs

Proof of Lemma 1. We show that $\Omega(e^{i\omega}) > 0$ for all $\omega \in [-\pi, \pi]$ implies that for every $T \geq 1$, the covariance matrix of $(Y_1', \ldots, Y_T')$ is full rank. The latter is true if and only if the the components of $(Y_0', ..., Y_T')'$ are linearly independent, i.e. for any $(n_Y T)$-vector $\alpha = (\alpha_0', \ldots, \alpha_{T-1}')$, $\alpha_t = (\alpha_{1t}, \ldots, \alpha_{n_Y t})'$,

$$\sum_{t=0}^{T-1} \sum_{k=1}^{n_Y} \alpha_{kt} Y_{k,T-t} = 0 \quad a.s. \qquad \text{implies} \qquad \alpha = 0.$$

Now, take any $T \geq 1$ and assume there exists a $(n_Y T)$-vector α such that $\sum_{t=0}^{T-1} \sum_{k=1}^{n_Y} \alpha_{kt} Y_{k,T-t} = 0$ a.s., i.e. such that $\sum_{t=0}^{T-1} \alpha_t' Y_{T-t} = 0$ a.s. Written in terms of the spectral densities, this implies

$$\left[\sum_{t=0}^{T-1} \alpha_t' e^{-it\omega} \right] \Omega(e^{i\omega}) \left[\sum_{t=0}^{T-1} \alpha_t e^{it\omega} \right] = 0 \quad \text{for a.e. } \omega \in [-\pi, \pi].$$

Since $\Omega(e^{i\omega})$ is everywhere nonsingular, the above implies that $\sum_{t=0}^{T-1} \alpha_t' e^{-it\omega} = 0$ for a.e. $\omega \in [-\pi, \pi]$, i.e. for every $1 \leqslant k \leqslant n_Y$,

$$\sum_{t=0}^{T-1} \alpha_{kt} e^{-it\omega} = s(e^{i\omega}) \quad \text{with} \quad s(e^{i\omega}) = 0 \text{ for a.e. } \omega \in [-\pi, \pi].$$

Using the inverse discrete-time Fourier transform, it then follows that for every $1 \leqslant k \leqslant n_Y$ and every $0 \leqslant t \leqslant T-1$,

$$\alpha_{kt} = \frac{1}{2\pi} \int_{-\pi}^{\pi} s(e^{i\omega}) e^{it\omega} d\omega = 0.$$

□

Proof of Lemma 2. We now show that Assumption 3 is equivalent to requiring that $\Omega(e^{i\omega}) > 0$, for every $\omega \in [-\pi, \pi]$. For this, write

$$\Omega(z) = \left[C(z\text{Id} - A)^{-1} B + D \right] \Sigma \left[B'(z^{-1}\text{Id} - A')^{-1} C' + D' \right]$$

(see, e.g., Komunjer and Ng, 2011, for details). Now, since $\Sigma > 0$, it is clear that $\Omega(e^{i\omega}) \geq 0$ for all $\omega \in [-\pi, \pi]$. However, $\Omega(e^{i\omega})$ will drop rank at some point on the unit circle if and only if there exists a non-zero n_Y-vector v and $\lambda \in [-\pi, \pi]$ such that

$$v' \left[C(e^{i\lambda}\text{Id} - A)^{-1} B + D \right] = 0,$$

which is equivalent to

$$\begin{pmatrix} v'C(e^{i\lambda}\text{Id} - A)^{-1} & v' \end{pmatrix} \begin{pmatrix} e^{i\lambda}\text{Id} - A & B \\ -C & D \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix},$$

that is

$$\operatorname{rank}\begin{pmatrix} e^{i\lambda}\mathrm{Id}-A & B \\ -C & D \end{pmatrix} < n_X + n_Y.$$

Thus, $\Omega(e^{i\omega}) > 0$ for all $\omega \in [-\pi, \pi]$ if and only if Assumption 3 holds. □

Proof of Lemma 3. To establish the result, we use Lemma 8.C.1 in Kailath, Sayed, and Hassibi (2000). For this, we need to check that A does not have unit-circle eigenvalues, which is ensured by Assumption 1; and that (A, C) is detectable, which is implied by the stronger observability requirement in Assumption 2(ii). Now applying Lemma 8.C.1 in Kailath, Sayed, and Hassibi (2000), we have that $\Omega(e^{i\omega}) > 0$ for all $\omega \in [-\pi, \pi]$ if and only if there exists a unique positive semi-definite solution P to the Riccati equation (10) for which $A - KC$ is stable and $\Sigma_a > 0$. To establish the result of Lemma 3, combine the above with Lemma 2. □

Proof of Lemma 4. Recall that the conditions in Assumption 2 are equivalent to the autocovariance minimality of the state space system (1). To establish Lemma 4, we first re-express autocovariance minimality in terms of the innovations representation (11). Using the innovations representation (11), we have $\Gamma(j) = CA^{j-1}\widetilde{N}$ for $j > 0$, where $\widetilde{N} = A\widetilde{P}_X C' + K\Sigma_a$, $\widetilde{P}_X = E[\widehat{X}_{t|t}\widehat{X}_{t|t}']$ is the solution to the Lyapunov equation $\widetilde{P}_X = A\widetilde{P}_X A' + K\Sigma_a K'$, and K and Σ_a are as defined in (11). Therefore, the system is autocovariance minimal if and only if (A, C) is observable and $(A, \widetilde{N})$ is controllable. We now show that the last condition is equivalent to (A, K) controllable. For this, we use an equivalent definition (on p.762 of Kailath, Sayed, and Hassibi (2000)): (A, K) is controllable if and only if $x'A = \lambda x'$ with $x \neq 0$ implies $x'K \neq 0$. Suppose that (A, K) is controllable but $(A, \widetilde{N})$ is not. Then there exists (x, λ) with $x'A = \lambda x'$ and $x \neq 0$ such that $x'\widetilde{N} = 0$. This means that $x'K\Sigma_a = -\lambda x'\widetilde{P}_X C'$. Since $\widetilde{P}_X = A\widetilde{P}_X A' + K\Sigma_a K'$, we have $\widetilde{P}_X x = \lambda (A - KC)\widetilde{P}_X x$. If $\lambda = 0$, then $x'\widetilde{P}_X x = 0$. Since $x'\widetilde{P}_X x = x'A\widetilde{P}_X A'x + (x'K\Sigma_a)\Sigma_a^{-1}(x'K\Sigma_a)'$ and $x'K\Sigma_a \neq 0$, $x'\widetilde{P}_X x = 0$ is not possible. Then $\lambda \neq 0$. Thus, $\left[\lambda^{-1}I - (A - KC)\right]\widetilde{P}_X x = 0$. In order to say that λ^{-1} is an eigenvalue of $A - KC$, we need to show that $\widetilde{P}_X x \neq 0$. By the Lyapunov equation and $x'A = \lambda x'$, we have $x'\widetilde{P}_X x = \lambda^2 x'\widetilde{P}_X x + x'K\Sigma_a K'x$. This means that $(1 - \lambda^2)x'\widetilde{P}_X x = x'K\Sigma_a K'x$. Since $x'K \neq 0$ and $\Sigma_a > 0$, we have $(1 - \lambda^2)x'\widetilde{P}_X x = x'K\Sigma_a K'x > 0$. It follows, by $|\lambda| < 1$, that $x'\widetilde{P}_X x > 0$. Hence, $\widetilde{P}_X x \neq 0$. We can now conclude that λ^{-1} is an eigenvalue of $A - KC$. By the stability of A, $\left|\lambda^{-1}\right| > 1$. Therefore, $A - KC$ has an eigenvalue outside the unit circle. This contradicts the stability of $A - KC$. □

Proof of Theorem 1. The proof proceeds in two steps. First, we show that the likelihood can be locally parameterized by a d-dimensional parameter that is identified. Second, we use classical arguments to derive the distribution of the reparameterized LR test statistic.

Step 1: reparameterize the likelihood. Let L be the lag operator. From the innovations representation (11), we have $\widehat{X}_{t|t} = (\mathrm{Id} - AL)^{-1} K a_t$. Since $Y_{t+1} = C\widehat{X}_{t|t} + a_{t+1}$, we have that $Y_{t+1} = k(L)a_{t+1}$, where $k(z) = C(\mathrm{Id} - Az)^{-1}Kz + \mathrm{Id}$. It is straightforward to verify that $k(z) = \mathrm{Id} + \sum_{j=1}^{\infty} CA^{j-1}Kz^j$. Hence, once we fix $k(\cdot)$, a_t is determined by $a_t = [k(L)]^{-1}Y_t$. By (13), the likelihood is determined by $\mathrm{vech}\Sigma_a$ and the sequence a_t. It follows that the likelihood can be parameterized by $(k(\cdot), \mathrm{vech}\Sigma_a)$.

Let $\mathcal{K} = \left\{k(\cdot) \mid k(z) = \mathrm{Id} + \sum_{j=1}^{\infty} CA^{j-1}Kz^j,\ A \text{ stable},\ (A, K) \text{ controllable},\ A - KC \text{ stable}\right\}$. By Theorems 2.6.2 and 2.6.3 of Hannan and Deistler (1988), there exist a finite set $\mathcal{A}$, sets $\{U_\alpha \mid \alpha \in \mathcal{A}\}$ and functions $\{\phi_\alpha \mid \alpha \in \mathcal{A}\}$ such that $\mathcal{K} = \bigcup_{\alpha \in \mathcal{A}} U_\alpha$ and ϕ_α is a homeomorphism from U_α to an open set in $\mathbb{R}^{2n_X n_Y}$ for any $\alpha \in \mathcal{A}$.

Let $(k_0(\cdot), \mathrm{vech}\Sigma_{a0})$ denote the parameters corresponding to the true likelihood. Fix $\alpha_0 \in \mathcal{A}$ such that $k_0(\cdot) \in U_{\alpha_0}$. Define

$$\eta_0 = (\phi_{\alpha_0}(k_0), \mathrm{vech}\Sigma_{a0}).$$

Define: (1) $\mathcal{V} = \left\{\mathrm{vech}\Sigma \mid a^{-1} \geq \lambda_{\max}(\Sigma)\lambda_{\min}(\Sigma) \geq a\right\}$ for some small constant $a > 0$ such that $\Sigma_{a0} \in \mathrm{interior}(\mathcal{V})$, and (2) $\mathcal{Q} \subset U_{\alpha 0}$ is a compact closed set with $k_0 \in \mathrm{interior}(\mathcal{Q})$. Notice that $\mathcal{D} = \phi_{\alpha_0}(\mathcal{Q}) \times \mathcal{V}$ is a compact set in $\mathbb{R}^d$ and $\eta_0 \in \mathrm{interior}(\mathcal{D})$. Let $f_t(\eta)$ denote the log likelihood of Y_t given Y^s evaluated at $\eta \in \mathcal{D}$. Hence, for $\eta \in \mathcal{D}$, the log likelihood can be written as $\ln \widetilde{L}_T(\eta) = \sum_{t=1}^{T} f_t(\eta)$. Since the model is correctly specified, $\{v_t(\eta_0)\}_{t=1}^{T}$ is a martingale difference sequence, where $v_t(\eta) = \partial f_t(\eta)/\partial\eta$; see e.g. Andrews and Mikusheva (2015). The analytical form of $v_t(\eta)$ can be obtained from Theorem 2.6.2(iii) of Hannan and Deistler (1988) and Lemma 5. The information matrix $\widetilde{I}(\eta)$ can be computed using Equation (18.4.7) of Lütkepohl (2005). After some algebra, it can be verified that $\mathrm{rank}\widetilde{I}(\eta_0) = d$. By Theorem 4.2.1 of Hannan and Deistler (1988), the maximum likelihood estimator for $(k_0(\cdot), \mathrm{vech}\Sigma_{a0})$ is consistent. Hence, under the null hypothesis of $\pi = \pi_0$, we have

$$P\left(LR_{1T} = \widetilde{LR}_{1T}\right) \to 1, \tag{27}$$

where $\widetilde{LR}_{1T} = 2\left(\ln \widetilde{L}_T(\widehat{\eta}_T) - \ln \widetilde{L}_T(\eta_0)\right)$ and $\widehat{\eta}_T = \arg\max_{\eta \in \mathcal{D}} \ln \widetilde{L}_T(\eta)$.

Step 2: show the desired result. Let $s_T(\eta) = T^{-1}\partial \ln \widetilde{L}_T(\eta)/\partial\eta$ and $A_T(\eta) = T^{-1}\partial^2 \ln \widetilde{L}_T(\eta)/\partial\eta\partial\eta'$. By construction, we have $s_T(\widehat{\eta}_T) = 0$. By the integral form of Taylor's theorem (Theorem C.15 of Lee (2012)), it follows that

$$-s_T(\eta_0) = s_T(\widehat{\eta}_T) - s_T(\eta_0) = \widetilde{A}_T(\widehat{\eta}_T - \eta_0),$$

where $\widetilde{A}_T = \int_0^1 A_T\left(\eta_0 + z(\widehat{\eta}_T - \eta_0)\right) dz$. As mentioned before, the consistency of $\widehat{\eta}_T$ follows by Theorem 4.2.1 of Hannan and Deistler (1988). The continuity of $A_T(\cdot)$ can be verified by using e.g.

the expression given in Equation (18.4.7) in Lutkepohl (2005):

$$A_T(\eta) = -\frac{1}{2T}\sum_{t=1}^{T}\left(\frac{\partial(\mathrm{vec}\Sigma_a)'}{\partial\eta}(\Sigma_a^{-1}\otimes\Sigma_a^{-1})\frac{\partial\mathrm{vec}\Sigma_a}{\partial\eta'} + 2\frac{\partial a_t'}{\partial\eta}\Sigma_a^{-1}\frac{\partial a_t}{\partial\eta'}\right).$$

By Theorem 2.6.2 in Hannan and Deistler (1988), η can be chosen as a subvector of $(A, K, C, \mathrm{vech}\Sigma_e)$. The term $\partial a_t/\partial\eta$ is continuous since its derivative can be derived by the implicit function theorem. Thus, $A_T(\eta)$ is continuous in η. Hence, the correct specification of the model implies that $\widetilde{A}_T = -\widetilde{I}(\eta_0) + o_p(1)$. This, together the above display and rank$\widetilde{I}(\eta_0) = d$, implies that

$$\widehat{\eta}_T - \eta_0 = \left[\left(\widetilde{I}(\eta_0)\right)^{-1} + o_p(1)\right] s_T(\eta_0). \tag{28}$$

Applying Taylor's theorem, we have

$$\begin{aligned}
\ln\widetilde{L}_T(\eta_0) - \ln\widetilde{L}_T(\widehat{\eta}_T) &= Ts_T(\widehat{\eta}_T)'(\eta_0-\widehat{\eta}_T) + \frac{T}{2}(\eta_0-\widehat{\eta}_T)'\left[\int_0^1 A_T\left(\widehat{\eta}_T + z(\eta_0-\widehat{\eta}_T)\right)dz\right](\eta_0-\widehat{\eta}_T)\\
&\overset{(i)}{=} \frac{T}{2}(\eta_0-\widehat{\eta}_T)'\left[\int_0^1 A_T\left(\widehat{\eta}_T + z(\eta_0-\widehat{\eta}_T)\right)dz\right](\eta_0-\widehat{\eta}_T)\\
&\overset{(ii)}{=} \frac{T}{2}(\eta_0-\widehat{\eta}_T)'\widetilde{A}_T(\eta_0-\widehat{\eta}_T)\\
&\overset{(iii)}{=} \frac{T}{2}s_T(\eta_0)'\left[\left(\widetilde{I}(\eta_0)\right)^{-1} + o_p(1)\right]\left[-\widetilde{I}(\eta_0) + o_p(1)\right]\left[\left(\widetilde{I}(\eta_0)\right)^{-1} + o_p(1)\right]s_T(\eta_0),
\end{aligned}$$

where (i) holds by $s_T(\widehat{\eta}_T) = 0$, (ii) holds by observing that $\int_0^1 A_T\left(\widehat{\eta}_T + z(\eta_0-\widehat{\eta}_T)\right)dz = \widetilde{A}_T$ and (iii) follows by (28) and $\widetilde{A}_T = -\widetilde{I}(\eta_0)+o_p(1)$. Since $T^{-1/2}s_T(\eta_0) \to^d N(0, \widetilde{I}(\eta_0))$, the above display implies that

$$\widetilde{LR}_{1T} = 2\left(\ln\widetilde{L}_T(\widehat{\eta}_T) - \ln\widetilde{L}_T(\eta_0)\right) = Ts_T(\eta_0)'\left[\widetilde{I}(\eta_0)\right]^{-1}s_T(\eta_0) + o_p(1) \xrightarrow{d} \chi^2_d. \tag{29}$$

The desired result follows by (27). □

Proof of Lemma 5. First, let us consider the partial derivatives of vecΣ_a with respect to π. Recall that π contains vechΣ_a, and that the vech operator performs column-wise vectorization with the upper portion excluded. In order to 'invert' the vech operator applied to any $n \times n$ symmetric matrix, we use an $n^2 \times n(n+1)/2$ duplication matrix $\mathcal{G}_n$ which is a matrix of 0s and 1s, with a single 1 in each row. Thus for any $n \times n$ symmetric matrix S, vec$(S) = \mathcal{G}_n$vech(S). Then,

$$\frac{\partial\mathrm{vec}\Sigma_a}{\partial\pi} = \begin{pmatrix} 0_{n_Y^2\times n_X(n_X+2n_Y)} & \mathcal{G}_{n_Y}\end{pmatrix}. \tag{30}$$

Next, we consider the computation of the partial derivatives of the innovations a_t with respect to Λ. For this, let $F \equiv A - KC$ and rewrite the innovation representation equations (11) as:

$$\widehat{X}_{t+1|t+1} = F\widehat{X}_{t|t} + KY_{t+1} \tag{31}$$

$$a_{t+1} = Y_{t+1} - C\widehat{X}_{t|t}. \tag{32}$$

Then, from (32)

$$\frac{\partial a_{t+1}}{\partial \pi} = -\frac{\partial (C\widehat{X}_{t|t})}{\partial \pi}. \tag{33}$$

The above can be computed using the product rule: if $M(\pi)$ and $N(\pi)$ are, respectively, $m \times p$ and $p \times q$ matrices of differentiable functions with respect to π, then

$$\frac{\partial \text{vec}(M(\pi)N(\pi))}{\partial \pi} = (N(\pi)' \otimes \text{Id}_m)\frac{\partial \text{vec}(M(\pi))}{\partial \pi} + (\text{Id}_q \otimes M(\pi))\frac{\partial \text{vec}(N(\pi))}{\partial \pi}.$$

So, from (33)

$$\begin{aligned}\frac{\partial a_{t+1}}{\partial \pi} &= -(\widehat{X}'_{t|t} \otimes \text{Id}_{n_Y})\frac{\partial \text{vec}C}{\partial \pi} - C\frac{\partial \widehat{X}_{t|t}}{\partial \pi} \\ &= \begin{pmatrix} 0_{n_Y \times n_X^2} & 0_{n_Y \times n_X n_Y} & -(\widehat{X}'_{t|t} \otimes \text{Id}_{n_Y}) & 0_{n_Y \times n_Y(n_Y+1)/2} \end{pmatrix} - C\frac{\partial \widehat{X}_{t|t}}{\partial \pi}. \end{aligned} \tag{34}$$

To compute the second term, we use (31), which, combined with the product rule gives:

$$\begin{aligned}\frac{\partial \widehat{X}_{t+1|t+1}}{\partial \pi} &= \frac{\partial (F\widehat{X}_{t|t})}{\partial \pi} + \frac{\partial (KY_{t+1})}{\partial \pi} \\ &= (\widehat{X}'_{t|t} \otimes \text{Id}_{n_X})\frac{\partial \text{vec}F}{\partial \pi} + F\frac{\partial \widehat{X}_{t|t}}{\partial \pi} + (Y'_{t+1} \otimes \text{Id}_{n_X})\frac{\partial \text{vec}K}{\partial \pi}.\end{aligned}$$

Since $F = A - KC$, we have

$$\begin{aligned}\frac{\partial \text{vec}F}{\partial \pi} &= \frac{\partial \text{vec}A}{\partial \pi} - (C' \otimes \text{Id}_{n_X})\frac{\partial \text{vec}K}{\partial \pi} - (\text{Id}_{n_X} \otimes K)\frac{\partial \text{vec}C}{\partial \pi} \\ &= \begin{pmatrix} \text{Id}_{n_X^2} & -(C' \otimes \text{Id}_{n_X}) & -(\text{Id}_{n_X} \otimes K) & 0_{n_X^2 \times n_Y(n_Y+1)/2} \end{pmatrix}.\end{aligned}$$

Combining the above then gives:

$$\frac{\partial \widehat{X}_{t+1|t+1}}{\partial \pi} = \begin{pmatrix} (\widehat{X}'_{t|t} \otimes \text{Id}_{n_X}) & (a'_{t+1} \otimes \text{Id}_{n_X}) & -(\widehat{X}'_{t|t} \otimes K) & 0_{n_X \times n_Y(n_Y+1)/2} \end{pmatrix} + F\frac{\partial \widehat{X}_{t|t}}{\partial \pi}, \tag{35}$$

with initial condition

$$\frac{\partial \widehat{X}_{0|0}}{\partial \pi} = 0. \tag{36}$$

Equations (34), (35) and (36) allow us to recursively compute $\partial a_t / \partial \pi$ for any $t \geq 1$. □

Proof of Corollary 1. In order to establish the expression for the Bartlett adjustment b_{1T} we need to establish the invariance of b_{1T} under re-parameterizations of the likelihood which are not necessarily one-to-one, such as the mapping from π to the "canonical" parameter, call it η, which is by construction identified. Specifically, let τ denote the mapping $\pi \mapsto \eta$, $\eta \in \mathbb{R}^d$, and let $\widetilde{L}_T(\eta) = L_T(\pi)$ so that $\widetilde{L}_T$ denotes the likelihood as a function of the "canonical" parameter η.

The idea of the proof is to use the known expressions for b_{1T} (see, e.g., McCullagh and Cox, 1986) written as functions of the scores with respect to η, $\partial \ln \widetilde{L}_T(\eta)/\partial\eta$, then show invariance when we express the latter in terms of π. This result is important because as already pointed out, the analytic expression for the "canonical" parameter η is hard to obtain. The dimension of η is known to be d (see, e.g., Hazewinkel, 1979; Hannan and Deistler, 1988).

Hereafter, let $v_{i,j}$ denote the (i,j) entry of the Fisher information matrix $\widetilde{I}(\eta)$. $v^{i,j}$ denotes the (i,j) entry of $\widetilde{I}(\eta)^{-1}$. Let

$$\begin{aligned}
v_{r,s,t} &\equiv T^{-1}E\left[\frac{\partial \ln \widetilde{L}_T}{\partial \eta_r}\frac{\partial \ln \widetilde{L}_T}{\partial \eta_s}\frac{\partial \ln \widetilde{L}_T}{\partial \eta_t}\right] \\
v_{r,s,t,u} &\equiv T^{-1}E\left[\frac{\partial \ln \widetilde{L}_T}{\partial \eta_r}\frac{\partial \ln \widetilde{L}_T}{\partial \eta_s}\frac{\partial \ln \widetilde{L}_T}{\partial \eta_t}\frac{\partial \ln \widetilde{L}_T}{\partial \eta_u}\right] \\
\kappa_{rs,i} &= T^{-1}E\left[\frac{\partial^2 \ln \widetilde{L}_T}{\partial \eta_r \partial \eta_s}\frac{\partial \ln \widetilde{L}_T}{\partial \eta_i}\right] \\
v^{r,s,t} &\equiv v_{i,j,k}v^{i,r}v^{j,s}v^{k,t} \\
v^{r,s,t,u} &\equiv v_{i,j,k,l}v^{i,r}v^{j,s}v^{k,t}v^{l,u} \\
V_r &\equiv \frac{\partial \ln \widetilde{L}_T}{\partial \eta_r} \\
V_{rs} &\equiv \frac{\partial^2 \ln \widetilde{L}_T}{\partial \eta_r \partial \eta_s} - v^{i,j}\kappa_{rs,j}V_i \\
V^{rs} &\equiv V_{ij}v^{i,r}v^{j,s} \\
V_{kl}^{ij} &\equiv \left[V^{ij} - E(V^{ij})\right]\left[V_{kl} - E(V_{kl})\right]
\end{aligned}$$

By McCullagh and Cox (1986), we can write the Bartlett correction term b_{1T} in terms of η:

$$b_{1T} = \frac{1}{12}\left(3\rho_{13}^2 + 2\rho_{23}^2 - 3\rho_4\right) + \frac{1}{4d}\left[\frac{2}{T}E\left(V_{ij}^{ij}\right) - \frac{1}{T}\mathrm{var}\left(V_{ij}v^{i,j}\right) - 2\mathrm{cov}\left(\frac{1}{T}V_iV_jv^{i,j}, \frac{1}{\sqrt{T}}V_{ij}v^{i,j}\right)\right], \tag{37}$$

where $\rho_{13}^2 = d^{-1}v^{i,j,k}v^{l,m,n}v_{i,j}v_{k,l}v_{m,n}$, $\rho_{23}^2 = d^{-1}v^{i,j,k}v^{l,m,n}v_{i,l}v_{j,m}v_{k,n}$ and $\rho_4 = d^{-1}v^{i,j,k,l}v_{i,j}v_{k,l}$.

The result of Corollary 1 follows from Lemmas 8, 9 and 10 below. To state the lemmas, let $A_{r,s} = \partial\eta_r/\partial\pi_s$ and $A_{r,st} = \partial^2\eta_r/\partial\pi_s\partial\pi_t$. Let $A \in \mathbb{R}^{d\times d_\pi}$ be the matrix whose (i,j) component is $A_{i,j}$. Define $D_{ij} = \partial^2 \ln \widetilde{L}_T/\partial\eta_i\partial\eta_j$ and $\bar{D}_{rs} = \partial^2 \ln L_T/\partial\pi_r\partial\pi_s$. Without loss of generality, we choose a coordinate system η such that the Fisher information matrix is Id_d. In other words, $v_{i,j} = v^{i,j} = \delta_{i,j}$, where $\delta_{i,j} = \mathbf{1}\{i = j\}$. By the invariance results in McCullagh and Cox (1986), the terms in (37) does not depend on how we choose η.

Lemma 6. *Let $A^{i,j}$ denote the (i,j) entry of A^+. Then the following hold:*

(1) $A_{i,s}A^{s,j} = \delta_{i,j}$.

(2) $\bar{V}_i = V_r A_{r,i}$

(3) $\bar{D}_{rs} = V_j A_{j,rs} + D_{ij} A_{i,r} A_{j,s}$.

(4) $\bar{\kappa}_{rs,u} = A_{j,rs} A_{j,u} + \kappa_{ij,q} A_{i,r} A_{j,s} A_{q,u}$

(5) $\bar{v}_{i,j} = A_{t,i} A_{t,j}$

(6) $\bar{v}^{i,j} = A^{i,q} A^{j,q}$.

(7) $\bar{v}_{i,j,k} = v_{r,s,t} A_{r,i} A_{s,j} A_{t,k}$.

(8) $\bar{v}^{l,m,n} = v_{r,s,t} A^{l,r} A^{m,s} A^{n,t}$.

The next result formulates the key observation that even though $\bar{V}_{rs}$ is a function of the second-order derivative of the log likelihood, it only depends on the first-order derivative of τ.

Lemma 7. $\bar{V}_{rs} = V_{ij} A_{i,r} A_{j,r}$.

The following result says that the term $\frac{1}{T}\text{var}\left(\bar{V}_{ij}\bar{v}^{i,j}\right)$ is the same as $\frac{1}{T}\text{var}\left(V_{ij}v^{i,j}\right)$.

Lemma 8. $\bar{V}_{rs}\bar{v}^{r,s} = V_{ij}v^{i,j}$.

Computations similar to those in the proof of Lemma 8 yield the following result. We omit the details for simplicity.

Lemma 9. $\bar{V}_{rs}^{rs} = V_{ij}^{ij}$ *and* $\bar{V}_r\bar{V}_s\bar{v}^{r,s} = V_iV_jv^{i,j}$.

The final lemma follows.

Lemma 10. $\bar{\rho}_{13}^2 = \rho_{13}^2$, $\bar{\rho}_{23}^2 = \rho_{23}^2$ *and* $\rho_4 = \bar{\rho}_4$.

□

Proof of Lemma 6. We first show part (1). Since $\text{rank}A = d$, we have $A^+ = A'(AA')^{-1}$ and thus $AA^+ = \text{Id}_d$. Part (1) follows.

Parts (2) and (3) follow by the chain rule of differentiation.

Part (4) follows from

$$\bar{\kappa}_{rs,u} = T^{-1}E\bar{D}_{rs}\bar{V}_u \overset{(i)}{=} T^{-1}E\left[\left(V_jA_{j,rs} + D_{ij}A_{i,r}A_{j,s}\right)V_qA_{q,u}\right] \overset{(ii)}{=} A_{j,rs}A_{j,u} + \kappa_{ij,q}A_{i,r}A_{j,s}A_{q,u},$$

where (i) holds by parts (2) and (3) and (ii) holds by $T^{-1}E(V_jV_q) = v_{j,q} = \delta_{j,q}$ and $T^{-1}E(D_{ij}V_q) = \kappa_{ij,q}$.

Part (5) follows by

$$\bar{v}_{i,j} = T^{-1}E\left(\bar{V}_i\bar{V}_j\right) \overset{(i)}{=} T^{-1}E\left(V_rV_s\right)A_{r,i}A_{s,j} \overset{(ii)}{=} v_{r,s}A_{r,i}A_{s,j} = A_{r,i}A_{r,j},$$

where (i) holds by part (2) and (ii) holds by $v_{r,s} = \delta_{r,s}$.

We now show part (6). Since the Fisher information matrix with respect to η is Id_d, we have that the Fisher information matrix with respect to π is $A'A$. It is not hard to verify that the pseudo-inverse of $A'A$ is $A^+(A^+)'$, whose (i,j) entry is $A^{i,q}A^{j,q}$.

For part (7), notice that:

$$\begin{aligned}\bar{v}_{i,j,k} = T^{-1}E\left(\bar{V}_i\bar{V}_j\bar{V}_k\right) \overset{(i)}{=} T^{-1}E\left(V_rA_{r,i}V_sA_{s,j}V_tA_{t,k}\right) &= T^{-1}E\left(V_rV_sV_t\right)A_{r,i}A_{s,j}A_{t,k} \\ &= v_{r,s,t}A_{r,i}A_{s,j}A_{t,k},\end{aligned}$$

where (i) holds by part (2).

Finally, we verify part (8). Notice that

$$\begin{aligned}\bar{v}^{l,m,n} = \bar{v}_{a,b,c}\bar{v}^{a,l}\bar{v}^{b,m}\bar{v}^{c,n} &\overset{(i)}{=} v_{r,s,t}A_{r,a}A_{s,b}A_{t,c}\bar{v}^{a,l}\bar{v}^{b,m}\bar{v}^{c,n} \\ &\overset{(ii)}{=} v_{r,s,t}A_{r,a}A_{s,b}A_{t,c}A^{a,q_1}A^{l,q_1}A^{b,q_2}A^{m,q_2}A^{c,q_3}A^{n,q_3} \\ &\overset{(iii)}{=} v_{r,s,t}\delta_{r,q_1}\delta_{s,q_2}\delta_{t,q_3}A^{l,q_1}A^{m,q_2}A^{n,q_3} \\ &= v_{r,s,t}A^{l,r}A^{m,s}A^{n,t},\end{aligned}$$

where (i) holds by part (7), (ii) holds by part (6) and (iii) holds by part (1). □

Proof of Lemma 7. Notice that

$$\begin{aligned}\bar{V}_{rs} = \bar{D}_{rs} - \bar{v}^{i,u}\bar{\kappa}_{rs,u}\bar{V}_i &\overset{(i)}{=} \bar{D}_{rs} - A^{i,q}A^{u,q}\bar{\kappa}_{rs,u}V_tA_{t,i} \\ &\overset{(ii)}{=} \bar{D}_{rs} - A^{u,q}\bar{\kappa}_{rs,u}V_t\delta_{t,q} \\ &= \bar{D}_{rs} - A^{u,t}\bar{\kappa}_{rs,u}V_t \\ &\overset{(iii)}{=} \bar{D}_{rs} - A^{u,t}V_t\left(A_{j,rs}A_{j,u} + \kappa_{ij,q}A_{i,r}A_{j,s}A_{q,u}\right) \\ &\overset{(iv)}{=} \bar{D}_{rs} - V_t\left(A_{j,rs}\delta_{j,t} + \kappa_{ij,q}A_{i,r}A_{j,s}\delta_{q,t}\right) \\ &= \bar{D}_{rs} - V_t\left(A_{t,rs} + \kappa_{ij,t}A_{i,r}A_{j,s}\right) \\ &\overset{(v)}{=} \left(V_jA_{j,rs} + D_{ij}A_{i,r}A_{j,s}\right) - V_t\left(A_{t,rs} + \kappa_{ij,t}A_{i,r}A_{j,s}\right) \\ &= \left(D_{ij} - V_t\kappa_{ij,t}\right)A_{i,r}A_{j,s},\end{aligned}$$

where (i) holds by Lemma 6(2) and (6), (ii) holds by Lemma 6(1), (iii) holds by Lemma 6(4), (iv) holds by Lemma 6(1) and (v) holds by Lemma 6(3).

The desired result follows from

$$V_{ij} = D_{ij} - v^{t,u}\kappa_{it,u}V_t \overset{(i)}{=} D_{ij} - \kappa_{ij,t}V_t,$$

where (i) holds by $v^{t,u} = \delta_{t,u}$. □

Proof of Lemma 8. Notice that

$$\bar{V}_{rs}\bar{v}^{r,s} \overset{(i)}{=} V_{ij}A_{i,r}A_{j,s}A^{r,q}A^{s,q} \overset{(ii)}{=} V_{ij}\delta_{i,q}\delta_{j,q} = V_{qq},$$

where (i) holds by Lemmas 6(6) and 7 and (ii) holds by Lemma 6(1). Since $v^{i,j} = \delta_{i,j}$, the desired result follows by $V_{ij}v^{i,j} = V_{ii}$. □

Proof of Lemma 10. Notice that

$$\begin{aligned}
\bar{v}^{i,j,k}\bar{v}^{l,m,n}\bar{v}_{i,j}\bar{v}_{k,l}\bar{v}_{m,n} &\overset{(i)}{=} v_{a,b,c}A^{i,a}A^{j,b}A^{k,c}v_{r,s,t}A^{l,r}A^{m,s}A^{n,t} \\
&\overset{(ii)}{=} v_{a,b,c}v_{r,s,t}A^{i,a}A^{j,b}A^{k,c}A^{l,r}A^{m,s}A^{n,t}A_{t_1,i}A_{t_1,j}A_{t_2,k}A_{t_2,l}A_{t_3,m}A_{t_3,n} \\
&\overset{(iii)}{=} v_{a,b,c}v_{r,s,t}\delta_{t,t_3}\delta_{t_2,r}\delta_{t_3,s}\delta_{t_1,a}\delta_{t_1,b}\delta_{t_2,c} \\
&= v_{a,a,r}v_{r,t,t},
\end{aligned}$$

where (i) holds by Lemma 6(8) (also applied to $\bar{v}^{i,j,k}$), (ii) holds by Lemma 6(5) and (iii) holds by Lemma 6(1).

Since $v^{i,r} = \delta_{i,r}$, we have $v^{r,s,t} = v_{r,s,t}$. It follows that

$$v^{i,j,k}v^{l,m,n}v_{i,j}v_{k,l}v_{m,n} = v_{i,j,k}v_{l,m,n}\delta_{i,j}\delta_{k,l}\delta_{m,n} = v_{i,i,k}v_{k,m,m}.$$

The above two displays imply that $\bar{\rho}_{13}^2 = \rho_{13}^2$. The proofs for $\bar{\rho}_{23}^2 = \rho_{23}^2$ and $\rho_4 = \bar{\rho}_4$ follow similar computations. □

Proof of Theorem 2. By the equivalence between confidence set construction and hypothesis testing (see, e.g., Theorem 3.5.1 of Lehmann and Romano, 2006), the coverage probability of our confidence set $CS_{1-\alpha,T}$ satisfies

$$\begin{aligned}
\liminf_{T\to\infty}\inf_{\theta\in\widetilde{\Theta}} P_\theta\left(\theta \in CS_{1-\alpha,T}\right) &= 1 - \limsup_{T\to\infty}\sup_{\theta\in\widetilde{\Theta}} P_\theta\left(\overline{LR}_{1T}(\theta) > \chi^2_{d,1-\alpha}\right) \\
&\geq 1 - \limsup_{T\to\infty}\sup_{\pi\in\widetilde{\Pi}} P_\pi\left(LR_{1T}(\pi) > \chi^2_{d,1-\alpha}\right).
\end{aligned}$$

It then suffices to show that

$$\limsup_{T\to\infty}\sup_{\pi\in\widetilde{\Pi}} P_\pi(LR_{1T}(\pi) > \chi^2_{d,1-\alpha}) \leqslant \alpha. \tag{38}$$

To establish (38), we invoke Corollary 2.1(c) of Andrews, Cheng, and Guggenberger (2017). By that result, it suffices to verify that for each sequence $\pi = \pi_T \in \widetilde{\Pi}$, $LR_{1T}(\pi_T)$ converges in distribution under P_{π_T} to χ^2_d.

We proceed by contradiction. Suppose that there exist a sequence $\pi_T \in \widetilde{\Pi}$ and constants $z_1, z_2 \in (0,1)$ such that $z_1 \neq z_2$ and $P_{\pi_T}(LR_{1T}(\pi_T) > \chi^2_{d,1-z_1}) \to z_2$.

The strategy is to adapt the arguments in the proof of Theorem 1. From Step 1 in in the proof of Theorem 1, recall that there exist a finite set $\mathcal{A}$, sets $\{U_\alpha \mid \alpha \in \mathcal{A}\}$ and functions $\{\phi_\alpha \mid \alpha \in \mathcal{A}\}$ such that $\mathcal{K} = \bigcup_{\alpha\in\mathcal{A}} U_\alpha$ and ϕ_α is a homeomorphism from U_α to an open set in $\mathbb{R}^{2n_X n_Y}$ for any $\alpha \in \mathcal{A}$, where the likelihood can be re-parameterized by elements in $\mathcal{K}$.

Let $(k_{0,T}(\cdot), \text{vech}\Sigma_{a0,T})$ with $k_{0,T} \in \mathcal{K}$ denote the parameters corresponding to π_T. Since $\mathcal{A}$ is finite, there exists $\alpha_0 \in \mathcal{A}$ such that U_{α_0} contains infinitely many members of the sequence $\{k_{0,T}(\cdot)\}$. Hence, we can take a subsequence π_{j_T} such that $k_{0,j_T} \in U_{\alpha_0}$. Let η_{0,j_T} denote the re-parameterized parameter, i.e., $\eta_{0,j_T} = (\phi_{\alpha_0}(k_{0,j_T}), \text{vech}\Sigma_{a0,j_T})$.

As in the proof of Theorem 1, we also define: (1) $\mathcal{V} = \left\{\text{vech}\Sigma \mid a^{-1} \geq \lambda_{\max}(\Sigma)\lambda_{\min}(\Sigma) \geq a\right\}$ for some small constant $a > 0$ such that $\Sigma_{a0,j_T} \in \text{interior}(\mathcal{V})$, and (2) $\mathcal{Q} \subset U_{\alpha 0}$ is a compact closed set with $k_{0,j_T} \in \text{interior}(\mathcal{Q})$. This is possible because $\widetilde{\Pi} \subset \text{interior}(\Pi)$ is compact and Π is compact. Thus, $\mathcal{D} = \phi_{\alpha_0}(\mathcal{Q}) \times \mathcal{V}$ is a compact set in $\mathbb{R}^d$ and $\eta_{0,j_T} \in \text{interior}(\mathcal{D})$. Let $\widetilde{L}(\eta)$ denote the likelihood parameterized by η.

By essentially the same argument as in Step 1 of the proof of Theorem 1, we have

$$P_{\pi_{j_T}}\left(LR_{1T}(\pi_{j_T}) = \widetilde{LR}_{1T}(\eta_{0,j_T})\right) \to 1, \tag{39}$$

where

$$\widetilde{LR}_{1T}(\eta_{0,j_T}) = 2\left(\sup_{\eta\in\mathcal{D}} \ln \widetilde{L}_T(\eta) - \ln \widetilde{L}_T(\eta_{0,j_T})\right).$$

With essentially the same argument as in Step 2 of the proof of Theorem 1, we can derive that

$$\widetilde{LR}_{1T}(\eta_{0,j_T}) \xrightarrow{d} \chi^2_d \qquad \text{under } P_{\pi_{j_T}}. \tag{40}$$

Combining (39) and (40), we have that $P_{\pi_{j_T}}(LR_{1T}(\pi_{j_T}) > \chi^2_{d,1-z_1}) \to z_1$. This contradicts $P_{\pi_T}(LR_{1T}(\pi_T) > \chi^2_{d,1-z_1}) \to z_2$ since $z_1 \neq z_2$! Therefore, $LR_{1T}(\pi_T)$ converges in distribution under P_{π_T} to χ^2_d. The proof is complete. □

Proof of Theorem 3. We adopt all the notations introduced in the proof of Theorem 1 and assume that the null hypothesis holds. Let ψ denote the mapping $\theta \mapsto \eta$. By Theorem 2.6.2 of Hannan and Deistler (1988), η parametrization is simply certain entries of π after normalizing other entries. Hence, Assumption 4 implies that ψ is twice continuously differentiable.

Notice that $L_T(\pi(\theta)) = \widetilde{L}_T(\psi(\theta))$. Let $\Psi(\theta) = \partial\psi(\theta)/\partial\theta' \in \mathbb{R}^{d\times d_\theta}$ and $s_{\theta,T}(\theta) = T^{-1}\partial \ln \widetilde{L}_T(\psi(\theta))/\partial\theta$. Hence, $s_{\theta,T}(\theta) = \Psi(\theta)' s_T(\psi(\theta))$ and $\bar{I}(\theta) = \Psi(\theta)'\widetilde{I}(\psi(\theta))\Psi(\theta)$. Since both $\bar{I}(\theta_0)$ and $\widetilde{I}(\psi(\theta_0)) = \widetilde{I}(\eta_0)$ have full rank and $d > d_\theta$, we have rank$\Psi(\theta_0) = d_\theta$.

Recall from (29) in the proof of Theorem 1, we have that

$$2\left(\ln \widetilde{L}_T(\widehat{\eta}_T) - \ln \widetilde{L}_T(\eta_0)\right) = T s_T(\eta_0)' \left[\widetilde{I}(\eta_0)\right]^{-1} s_T(\eta_0) + o_p(1). \tag{41}$$

Let $\widehat{\theta}_T = \arg\max_{\theta\in\Theta} L_T(\pi(\theta)) = \arg\max_{\theta\in\Theta} \widetilde{L}_T(\psi(\theta))$. Since θ_0 is identified with full rank Fisher information matrix, we can follow classical arguments (or Step 2 in the proof of Theorem 1 with $(\eta_0, \widehat{\eta}_T)$ replaced by $(\theta_0, \widehat{\theta}_T)$) and obtain

$$2\left(\ln \widetilde{L}_T(\psi(\widehat{\theta}_T)) - \ln \widetilde{L}_T(\psi(\theta_0))\right) = T s_{\theta,T}(\eta_0)' \left[\bar{I}(\eta_0)\right]^{-1} s_{\theta,T}(\eta_0) + o_p(1).$$

By $s_{\theta,T}(\theta) = \Psi(\theta)' s_T(\psi(\theta))$ and $\bar{I}(\theta) = \Psi(\theta)'\widetilde{I}(\psi(\theta))\Psi(\theta)$, the above display implies that

$$2\left(\ln \widetilde{L}_T(\psi(\widehat{\theta}_T)) - \ln \widetilde{L}_T(\psi(\theta_0))\right) = T s_T(\eta_0)'\Psi(\theta_0) \left[\Psi(\theta_0)'\widetilde{I}(\eta_0)\Psi(\theta_0)\right]^{-1} \Psi(\theta_0)' s_T(\eta_0) + o_p(1). \tag{42}$$

Let $Z_T = \sqrt{T}\left[\widetilde{I}(\eta_0)\right]^{-1/2} s_T(\eta_0)$ and $W_0 = \left[\widetilde{I}(\eta_0)\right]^{1/2}\Psi(\theta_0)$. Since $\eta_0 = \psi(\theta_0)$, it follows by (41) and (42) that

$$\begin{aligned} LR_{2T} = 2\left(\ln \widetilde{L}_T(\widehat{\eta}_T) - \ln \widetilde{L}_T(\psi(\widehat{\theta}_T))\right) &= Z_T' Z_T - Z_T' W_0 (W_0' W_0)^{-1} W_0 Z_T + o_p(1) \\ &= Z_T' \left[I_d - W_0(W_0'W_0)^{-1}W_0\right] Z_T + o_p(1). \end{aligned}$$

Notice that rankW_0 = rank$\Psi(\theta_0) = d_\theta$. Hence, $I_d - W_0(W_0'W_0)^{-1}W_0$ is a projection matrix with rank $d - d_\theta$. Since $Z_T \to^d N(0, I_d)$, we have that

$$2\left(\ln \widetilde{L}_T(\widehat{\eta}_T) - \ln \widetilde{L}_T(\psi(\widehat{\theta}_T))\right) \xrightarrow{d} \chi^2_{d-d_\theta}.$$

The proof is complete. □

Proof of Corollary 2. For the purpose of this proof, we view of the Bartlett correction as a correction of the mean of the likelihood ratio statistic, see e.g., Equation (2) in McCullagh and Cox (1986) or Equation (1.1) in Barndorff-Nielsen and Hall (1988). Hence,

$$E\left(LR_{1T}\right) - d = T^{-1} b_{1T}(\pi) + o(T^{-1}).$$

Similarly, when we view the model as a parametric model on Θ, we have

$$E\left(\overline{LR}_{\theta,1T}\right) - d_\theta = T^{-1} b_{\theta,1T}(\theta) + o(T^{-1}),$$

where $\overline{LR}_{\theta,1T} = 2\left(\sup_{\theta\in\Theta} \ln L_T(\pi(\theta)) - \ln L_T(\pi(\theta_0))\right)$. Notice that under the null hypothesis, $\pi_0 = \pi(\theta_0)$ and thus $LR_{1T} - \overline{LR}_{\theta,1T} = LR_{2T}$. Therefore, the above two display implies that

$$E\left(LR_{2T}\right) - (d - d_\theta) = \left(b_{1T}(\pi_0) - b_{\theta,1T}(\theta_0)\right) T^{-1} + o(T^{-1}).$$

As mentioned before, the Bartlett correction term is the same as the correction term of $E\left(LR_{2T}\right)$. The desired result follows. $\square$