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

Ivana Komunjer* Yinchu Zhu†

December 27, 2016

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 properties of the likelihood ratio (LR) test in linear Gaussian state space models. Our contributions

*Corresponding Author: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. Email: komunjer@ucsd.edu

†Rady School of Management, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093. Email: Yinchu.Zhu@rady.ucsd.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 comments and suggestions.

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 (Hazewinkel, 1979; Hannan and Deistler, 1988): that is, they can be parameterized by a lower-dimensional "canonical" parameter which is by construction identified. 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.

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. This distribution is, however, difficult to compute, which would make their test difficult to operationalize in our context. Andrews and Mikusheva (2015) derive the asymptotic distribution of the Lagrange multiplier and LR test statistics under weak identification. They assume however that in the limit the model is identified. Working in the frequency domain, Qu (2014) derives the asymptotic distribution of the score (i.e. Lagrange multiplier) test statistic. While the distribution is the usual chi-squared, its number of degrees of freedom has to be estimated. 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. 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 the problem of specification testing in DSGE models. Over the last two decades, these models have become the workhorse of modern macroeconomic analysis. Yet, there is growing consensus among macroeconomists today that DSGE models are misspecified in various aspects (see, e.g., Schorfheide, 2013, for a recent survey). 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 a more recent example). 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. There are other aspects of the DSGE model that can be compared to a finite lag VAR. Such limited information procedures are found, for example, in Smith (1993), Canova (1994), Schorfheide (2000), Del Negro, Schorfheide, Smets, and Wouters (2007), Le, Meenagh, Minford, and Wickens (2011), and Dufour, Khalaf, and Kichian (2013).[1] It is worth pointing out that all of these methods maintain correct specification of the large model (finite lag VAR). If the DSGE model is correct, then the coefficients of the large model satisfy some equality restrictions. As a result, the null hypothesis of correct specification is naturally nested in the maintained hypothesis. In general, however, DSGE models only admit VARMA (or linear state space) representations, which makes the finite lag VAR assumption untenable.

[1] 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).

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. The previous LR test approaches proposed in Sargent (1978) and Christiano (2007), for example, embed the DSGE model within a finite lag VAR model. Since the VAR parameters are identifiable, there are no identification issues in the unrestricted model. A more recent proposal by Guerron-Quintana, Inoue, and Kilian (2013) is to instead embed the DSGE model within a finite order state-space model but to also assume the latter is identified. This choice of unrestricted model is better suited for the analysis of DSGE models, for which finite order VAR representations exist only in special cases (see, e.g., Ravenna, 2007). Our testing results depart from these earlier papers in two important ways: First, we work in a framework in which the unrestricted model (a finite order state-space model) is not identified. Lack of identification will of course affect the asymptotic distribution of the LR test statistic. 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. Second, we are going to provide a small sample correction, namely the Bartlett correction, to the first order asymptotic distribution of the likelihood ratio statistics. We view this as a very useful result that allows researchers to apply our method to small sample sizes.

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. 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. Final section 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.[2] 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] It is the unique $n \times m$ matrix A^+ satisfying: (i) $AA^+A = A$; (ii) $A^+AA^+ = A^+$; (iii) $(AA^+)' = AA^+$; and (iv) $(A^+A)' = A^+A$.

2 Setup

2.1 Model and Assumptions

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

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

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 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).[3] Second is the requirement that the disturbances be Gaussian. The normality assumption is common in both classical (see, e.g., Altug, 1989; Ireland, 2004) 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 state-space models such as (1).[4]

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.*

[3]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 (1989).

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

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:

$$\Gamma(j) = \begin{cases} D\Sigma D' + CP_X C', & j = 0 \\ CA^{j-1}\left(AP_X C' + B\Sigma D'\right), & j > 0 \end{cases}, \tag{2}$$

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). 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 (2). Letting L be the Cholesky factor of $\Gamma(0)$, i.e. $\Gamma(0) = LL'$, and $N \equiv AP_X C' + B\Sigma D'$, 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

$$\begin{aligned} S_{t+1} &= AS_t + NU_{t+1} \\ V_{t+1} &= CS_t + \Gamma(0)U_{t+1}, \end{aligned} \tag{3}$$

in which both the inputs U_t and the outputs V_t are observed. Indeed, the Markov parameters in $V_t = H(L)U_t$, $H(L) = \sum_{j=0}^{\infty} h(j)L^j$, 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 (3) 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* $N = AP_X C' + B\Sigma D'$ *and* $P_X = AP_X A' + B\Sigma B'$. *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 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 nonsigularity of the process $\{Y_t\}_{t\in\mathbb{Z}}$.

Lemma 1. $\Omega(e^{i\omega}) > 0$ *for all* $\omega \in [-\pi, \pi]$ *if and only if 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, positive definiteness of the spectral density everywhere on the unit circle 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:*

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

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

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

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

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

The rank condition in Assumption 3 is then automatically satisfied whenever $A = \widetilde{A}$ is stable, as required by Assumption 1.

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.*

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

$$\Sigma_{a,t} = CP_{t|t}C' + D\Sigma D' \tag{4}$$

$$K_t = \left[AP_{t|t}C' + B\Sigma D'\right]\Sigma_{a,t}^{-1} \tag{5}$$

$$\widehat{X}_{t+1|t+1} = A\widehat{X}_{t|t} + K_t[Y_{t+1} - C\widehat{X}_{t|t}] \tag{6}$$

$$P_{t+1|t+1} = AP_{t|t}A' + B\Sigma B' - K_t\Sigma_{a,t}K_t'. \tag{7}$$

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 (4).

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

$$P = APA' + B\Sigma B' - \left[APC' + B\Sigma D'\right]\left[CPC' + D\Sigma D'\right]^{-1}\left[CPA' + D\Sigma B'\right]. \tag{8}$$

It then follows that $P_{t|t} = P$ for all $t \geq 1$ and the recursions in (4)-(7) yield the so-called innovations representation of the state-space system in (1):

$$\begin{aligned} \widehat{X}_{t+1|t+1} &= A\widehat{X}_{t|t} + Ka_{t+1} \\ Y_{t+1} &= C\widehat{X}_{t|t} + a_{t+1} \end{aligned} \tag{9}$$

where P is a solution to the Riccati equation in (8), 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* (8) *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 (9). 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 (9), the parameters appearing in the likelihood of the state space system (1) are the elements of

$$\pi \equiv \left((\text{vec}A)', (\text{vec}K)', (\text{vec}C)', (\text{vech}\Sigma_a)'\right)'. \tag{10}$$

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

$$\ln L_T(\pi) = -\sum_{t=1}^{T} \left[\frac{n_Y}{2} \ln(2\pi) + \frac{1}{2} \ln \det \Sigma_a + \frac{1}{2} a_t' \Sigma_a^{-1} a_t\right], \tag{11}$$

where a_t and Σ_a are determined through (9).

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 (9).

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:

$$\Pi \equiv \left\{\pi : A \text{ stable}, \ (A, K) \text{ controllable}, \ (A, C) \text{ observable}, \ A - KC \text{ stable}, \ \Sigma_a > 0\right\}. \tag{12}$$

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

3 Likelihood Ratio Test

3.1 Simple Null Hypothesis

We start our analysis of likelihood based inference 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 (9) 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 (12). We can assess the veracity of H_1 by examining the behavior of the log-likelihood ratio statistics:

$$LR_{1T} \equiv 2\left(\sup_{\pi\in\Pi} \ln L_T(\pi) - \ln L_T(\pi_0)\right). \tag{13}$$

In regular cases, 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 (9) 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 (9).

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,

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

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 "free components" in π. Since π is not identified, certain functional relations exist between its elements. This implies in particular that the number of "free components" 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., Hazewinkel, 1979; Hannan and Deistler, 1988).

The result of Theorem 1 provides an asymptotic approximation to the distribution of LR_{1T}. It can be restated by saying that

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

that is, the distribution of LR_{1T} 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}, letting

$$LR^*_{1T} \equiv \left(1 + \frac{b_{1T}}{T}\right)^{-1} LR_{1T}, \tag{14}$$

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

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

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} 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:

$$\begin{aligned}\frac{\partial \ln L_T(\pi)}{\partial \pi} &= -\frac{1}{2}\sum_{t=1}^{T}\left[\mathrm{vec}\left(\frac{\partial \ln\det\Sigma_a}{\partial\pi}\right)'\left(\frac{\partial \mathrm{vec}\Sigma_a}{\partial \pi}\right) + \frac{\partial \mathrm{tr}(a_t'\Sigma_a^{-1}a_t)}{\partial\pi}\right] \\ &= -\frac{1}{2}\sum_{t=1}^{T}\left[\mathrm{vec}\left[\Sigma_a^{-1}\left(\mathrm{Id}_{n_Y} - a_t a_t'\Sigma_a^{-1}\right)\right]'\left(\frac{\partial \mathrm{vec}\Sigma_a}{\partial\pi}\right) + 2a_t'\Sigma_a^{-1}\left(\frac{\partial a_t}{\partial \pi}\right)\right]\end{aligned} \tag{15}$$

The computation requires (i) $\partial\mathrm{vec}\Sigma_a/\partial\pi$, 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 $\partial a_t/\partial\partial\pi$ 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}, let $I(\pi)$ denote the Fisher information matrix:

$$
\begin{aligned}
I(\pi) &\equiv T^{-1}E\left[-\frac{\partial^2 \ln L_T(\pi)}{\partial\pi\partial\pi'}\right] \\
&= T^{-1}\sum_{t=1}^{T}\left[\frac{1}{2}\left(\frac{\partial \mathrm{vec}\Sigma_a}{\partial\pi}\right)'\left(\Sigma_a^{-1}\otimes\Sigma_a^{-1}\right)\left(\frac{\partial \mathrm{vec}\Sigma_a}{\partial\pi}\right) + E\left\{\left(\frac{\partial a_t}{\partial\pi}\right)'\Sigma_a^{-1}\left(\frac{\partial a_t}{\partial\pi}\right)\right\}\right]. \qquad (16)
\end{aligned}
$$

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} in (14).

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:*

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

and let

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

Then, the Bartlett adjustment b_{1T} *in* (14) *can be computed as:*

$$b_{1T} = b_{1T}(\pi_0).$$

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} need to be made: first, it is the number d of the "free" 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 nonsingular, the Bartlett adjustment now depends on its pseudo-inverse rather than its inverse.

3.2 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

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

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. The likelihood ratio test statistic now takes the form:

$$LR_{2T} \equiv 2\left(\sup_{\pi\in\Pi} \ln L_T(\pi) - \sup_{\theta\in\Theta} \ln L_T(\pi(\theta))\right), \tag{17}$$

where Θ is the parameter space for the deep parameter θ, and the parameter space Π for π is as defined in (12).

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 2. *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 2 is non-trivial. Put in words, Theorem 2 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 "free components" in π, and the dimension of the "free components" in θ. Since θ is assumed to be identified, the number of its "free components" simply equals its dimension.

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

$$LR^*_{2T} \equiv \left(1 + \frac{b_{2T}}{T}\right)^{-1} LR_{2T}, \tag{18}$$

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 2 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 (18) 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:

$$X_{t+1} = \underbrace{\Pi}_{A(\theta)} X_t + \underbrace{\begin{pmatrix} W & 0 \end{pmatrix}}_{B(\theta)} \epsilon_{t+1}, \quad Y_{t+1} = \underbrace{U\Pi}_{C(\theta)} X_t + \underbrace{\begin{pmatrix} UW & \mathrm{Id} \end{pmatrix}}_{D(\theta)} \epsilon_{t+1}, \tag{19}$$

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:

$$\beta_0 = 0.95, \quad \vartheta_0 = 2, \quad 1 - \alpha_0 = 0.36, \quad \gamma_0 = 1.0041, \quad \delta_0 = 0.025$$

$$a_0 = 2.7818, \quad \rho_0 = 0.95, \quad \sigma_0 = 0.04, \quad \Sigma_0 = v\text{Id} \;\text{ with }\; v = 0.02$$

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: $\theta = (\beta, \rho, \sigma, v_{11}, v_{12}, v_{13})$. All other parameters are kept fixed. We generate time series of length $T = 250$, 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 size ($T = 250$). Table 1 summarizes our results.

Table 1: Empirical Size

Null hypothesis	Unadjusted	Adjusted
H_0	0.1168	0.1098
H_1	0.1263	0.1115
H_2	0.1240	0.1035

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

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%.

4.3 Power Experiment

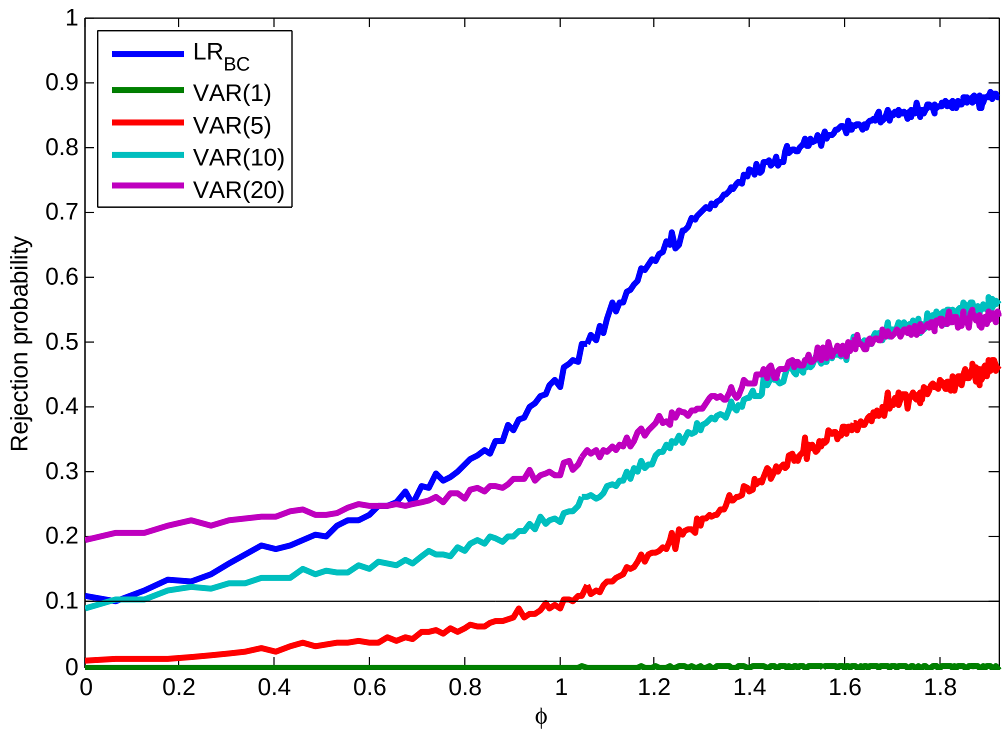
To evaluate the empirical power properties of our LR specification test (i.e. our LR test of the null hypothesis H_2), we generate the data under alternative RBC model specifications. These specifications suppose that within period preferences of any household are:

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

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 $u(C_t, L_t) = \ln C_t + \vartheta L_t$ 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:

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

Figure 1: Rejection probability of LR tests. Nominal size: 10%.



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.

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.

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.

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.

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.

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.

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.

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. (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.

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.

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.

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.

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

Qu, Z. (2014): “Inference in dynamic stochastic general equilibrium models with possible weak identification,” *Quantitative Economics*, 5(2), 457–494.

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. (1989): "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.

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.

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

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

$$\begin{aligned}
\frac{\gamma}{c_t} &= \beta E_t \left[\frac{1}{c_{t+1}} \left((1-\alpha) \frac{q_{t+1}}{k_{t+1}} + (1-\delta) \right) \right] \\
\vartheta c_t (1-h_t)^{-\varphi} &= \alpha \frac{q_t}{h_t} \\
\gamma k_{t+1} &= q_t + (1-\delta) k_t - c_t \\
q_t &= a_t k_t^{1-\alpha} h_t^{\alpha} \\
\ln a_t &= (1-\rho) \ln a + \rho \ln a_{t-1} + \epsilon_t
\end{aligned} \tag{20}$$

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

$$\begin{aligned}
q &= A^{1/\alpha}\left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta}\right]^{(1-\alpha)/\alpha} h \\
k &= q\left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta}\right] \\
c &= q\left[1 - \frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta}\right] \\
\frac{(1-h)^{\varphi}}{h} &= \frac{\vartheta}{\alpha}\left[1 - \frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta}\right] \\
a &= A
\end{aligned} \tag{21}$$

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 (21).

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 (20) around the steady state $(q_t, k_t, c_t, h_t, a_t) = (q, k, c, h, a)$ leads to the following equations:

$$\begin{aligned}
(\gamma/\beta)E_t[\bar{c}_{t+1}] - (\gamma/\beta)\bar{c}_t &= (\gamma/\beta - 1 + \delta)\, E_t[\bar{q}_{t+1}] - (\gamma/\beta - 1 + \delta)\bar{k}_{t+1} \\
\bar{c}_t + \left[1 + \varphi\frac{h}{1-h}\right]\bar{h}_t &= \bar{q}_t \\
\gamma\left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta}\right]\bar{k}_{t+1} &= \bar{q}_t + (1-\delta)\left[\frac{1-\alpha}{\gamma/\beta - 1 + \delta}\right]\bar{k}_t - \left[1 - \frac{(1-\alpha)(\gamma - 1 + \delta)}{\gamma/\beta - 1 + \delta}\right]\bar{c}_t \\
\bar{q}_t &= \bar{a}_t + (1-\alpha)\bar{k}_t + \alpha\bar{h}_t \\
\bar{a}_t &= \rho\bar{a}_{t-1} + \epsilon_t
\end{aligned} \tag{22}$$

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 (22) 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 (22) 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 (22) 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:

$$(D + FA^{-1}B)E_t[s_{t+1}] = (G + HA^{-1}B)s_t + (HA^{-1}C - \rho FA^{-1}C)\bar{a}_t, \tag{23}$$

where we have used the fact that $E_t[\bar{a}_{t+1}] = \rho \bar{a}_t$, as implied by the last equality in (22).

A.4 The Blanchard-Kahn Procedure

Looking back at (23), 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

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

Then let

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

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 $\det(K - \mathrm{Id}) < 0$. Now,

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

so

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

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, $\det(K - \mathrm{Id}) < 0$ 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 first 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 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

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

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

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

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 (8) 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 (9). Using the innovations representation (9), 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 (9). 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 (9), we have $\widehat{X}_{t|t} = (\mathrm{Id} - AL)^{-1}Ka_t$. 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 (11), 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

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

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 $\text{rank}\widetilde{I}(\eta_0) = d$. By Theorem 4.2.1 of Hannan and Deistler (1988), the maximum likelihood estimator for $(k_0(\cdot), \text{vech}\Sigma_{a0})$ is consistent. Hence, under the null hypothesis of $\pi = \pi_0$, we have

$$P\left(LR_{1T} = \widetilde{LR}_{1T}\right) \to 1, \tag{24}$$

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). It can be verified by tedious algebra that $A_T(\cdot)$ is continuous. 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 $\text{rank}\widetilde{I}(\eta_0) = d$, implies that

$$\widehat{\eta}_T - \eta_0 = \left[\left(\widetilde{I}(\eta_0)\right)^{-1} + o_p(1)\right] s_T(\eta_0). \tag{25}$$

Applying Taylor's theorem, we have

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

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 (25) 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

$$\widetilde{LR}_{1T} = 2\left(\ln \widetilde{L}_T(\widehat{\eta}_T) - \ln \widetilde{L}_T(\eta_0)\right) = T s_T(\eta_0)'\left[\widetilde{I}(\eta_0)\right]^{-1} s_T(\eta_0) + o_p(1) \overset{d}{\to} \chi_d^2. \tag{26}$$

The desired result follows by (24). □

Proof of Lemma 5. First, let us consider the partial derivatives of $\text{vec}\Sigma_a$ with respect to π. Recall that π contains $\text{vech}\Sigma_a$, 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, $\text{vec}(S) = \mathcal{G}_n \text{vech}(S)$. Then,

$$\frac{\partial \text{vec}\Sigma_a}{\partial \pi} = \begin{pmatrix} 0_{n_Y^2 \times n_X(n_X+2n_Y)} & \mathcal{G}_{n_Y} \end{pmatrix}. \tag{27}$$

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 (9) as:

$$\widehat{X}_{t+1|t+1} = F\widehat{X}_{t|t} + KY_{t+1} \tag{28}$$

$$a_{t+1} = Y_{t+1} - C\widehat{X}_{t|t}. \tag{29}$$

Then, from (29)

$$\frac{\partial a_{t+1}}{\partial \pi} = -\frac{\partial (C\widehat{X}_{t|t})}{\partial \pi}. \tag{30}$$

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 (30)

$$\begin{aligned} \frac{\partial a_{t+1}}{\partial \pi} &= -(\widehat{X}'_{t|t} \otimes \text{Id}_{n_Y})\frac{\partial \text{vec}C}{\partial \pi} - C\frac{\partial \widehat{X}_{t|t}}{\partial \pi} \\ &= \begin{pmatrix} 0_{n_Y \times n_X^2} & 0_{n_Y \times n_X n_Y} & -(\widehat{X}'_{t|t} \otimes \text{Id}_{n_Y}) & 0_{n_Y \times n_Y(n_Y+1)/2} \end{pmatrix} - C\frac{\partial \widehat{X}_{t|t}}{\partial \pi}. \end{aligned} \tag{31}$$

To compute the second term, we use (28), 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:

$$\frac{\partial \widehat{X}_{t+1|t+1}}{\partial \pi} = \begin{pmatrix} (\widehat{X}'_{t|t} \otimes \text{Id}_{n_X}) & (a'_{t+1} \otimes \text{Id}_{n_X}) & -(\widehat{X}'_{t|t} \otimes K) & 0_{n_X \times n_Y(n_Y+1)/2} \end{pmatrix} + F\frac{\partial \widehat{X}_{t|t}}{\partial \pi}, \tag{32}$$

with initial condition

$$\frac{\partial \widehat{X}_{0|0}}{\partial \pi} = 0. \tag{33}$$

Equations (31), (32) and (33) 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

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

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

$$b_{1T} = \frac{1}{12}\left(3\rho_{13}^2 + 2\rho_{23}^2 - 3\rho_4\right) + \frac{1}{4d}\left[\frac{2}{T}E\left(V^{ij}_{ij}\right) - \frac{1}{T}\text{var}\left(V_{ij}v^{i,j}\right) - 2\text{cov}\left(\frac{1}{T}V_iV_jv^{i,j}, \frac{1}{\sqrt{T}}V_{ij}v^{i,j}\right)\right], \tag{34}$$

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 (34) 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 $\frac{1}{T} \mathrm{var}\left(\bar{V}_{ij} \bar{v}^{i,j}\right)$ is the same as $\frac{1}{T} \mathrm{var}\left(V_{ij} v^{i,j}\right)$.

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* $\bar{V}_r \bar{V}_s \bar{v}^{r,s} = V_i V_j v^{i,j}$.

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 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

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

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. $\square$

Proof of Theorem 2. 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 $\text{rank}\Psi(\theta_0) = d_\theta$.

Recall from (26) in the proof of Theorem 1, we have that

$$2\left(\ln \widetilde{L}_T(\widehat{\eta}_T) - \ln \widetilde{L}_T(\eta_0)\right) = T s_T(\eta_0)' \left[\widetilde{I}(\eta_0)\right]^{-1} s_T(\eta_0) + o_p(1). \tag{35}$$

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

$$2\left(\ln \widetilde{L}_T(\psi(\widehat{\theta}_T)) - \ln \widetilde{L}_T(\psi(\theta_0))\right) = T s_T(\eta_0)'\Psi(\theta_0) \left[\Psi(\theta_0)'\widetilde{I}(\eta_0)\Psi(\theta_0)\right]^{-1} \Psi(\theta_0)' s_T(\eta_0) + o_p(1). \tag{36}$$

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 (35) and (36) 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 $\text{rank}W_0 = \text{rank}\Psi(\theta_0) = d_\theta$. 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. $\square$

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. □