Inference About

Long Run Canonical Correlations*

Prosper Dovonon

Concordia University†

Alastair R. Hall

University of Manchester‡

and

Kalidas Jana

University of Texas at Brownsville§

October 31, 2011

*The authors are grateful to Tata Subba Rao for useful comments on aspects of this work. An earlier version of the paper entitled "Long Run Canonical Correlations: Estimation, Inference and Usefulness in Econometric Analysis of Time Series" was presented at the European Meetings of the Econometric Society in Barcelona, August 23-27, 2009. A version of this paper was also presented at the Canadian Economics Association conference in Ottawa, June 2-5, 2011 under the title "Long Run Canonical Correlations: Estimation and Inference".

†Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, H3G 1M8 Canada. E-mail: prosper.dovonon@concordia.ca.

‡Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK. E-mail: alastair.hall@manchester.ac.uk.

§Department of of Finance & Economics, School of Business, University of Texas at Brownsville, 80 Fort Brown, Brownsville, TX 78520-4956, USA. E-mail: kalidas.jana@utb.edu.

Abstract

This paper proposes methods for testing the null hypothesis that a number of so-called long run canonical correlations (LRCCs) are zero. Two test statistics are proposed and their limiting distributions are derived under the null hypothesis. The finite sample properties of the tests are illustrated via a simulation study that reveals the asymptotic theory provides a good guidance to behaviour in moderate or large sized samples. It is shown that the statistics provide a natural way for testing the asymptotic independence of two standardized sums. The usefulness of the tests is illustrated via the following examples: inference about cointegrating vector in a particular cointegration model; inference about break points in a cointegration model; moment estimation; parameter estimation in Generalized Method of Moments estimation.
Key Words: Long run canonical correlations; Canonical coherences; Asymptotic independence of standardized sums; Cointegration; Inference about and based on moment conditions.

JEL Classification: C12, C13, C22, C32.

1 Introduction

Contemporaneous canonical correlations (CCCs) are used extensively in multivariate statistics[1]. They measure the degree of association between linear combinations of two random vectors which are chosen to satisfy certain orthogonality and normalizing conditions. Recently, a new kind of canonical correlations, called long run canonical correlations (LRCCs), has emerged from the work of Hall, Inoue, Jana & Shin (2007) in their analysis of the information content of moment based econometric estimators. As we show in this paper, LRCCs also arise in other situations of interest in econometrics. While the concept of LRCCs has been introduced into the econometrics literature, methods for inference about LRCCs have not been developed. The objective of this paper is to fill this gap. Specifically, we propose statistics for testing the null hypothesis that a number of LRCCs are zero. Our asymptotic analysis exploits a connection between LRCCs and canonical coherences in the frequency domain literature. It is shown that the statistics provide a natural way for testing the asymptotic independence of two standardized sums. The usefulness of the tests is illustrated via examples involving: inference in a cointegration model; inference about break points in a cointegration model; moment estimation; parameter estimation in Generalized Method of Moments estimation. In each case, the properties of the statistic in question are shown to depend on whether or not certain LRCCs are zero.

The rest of the paper is organized as follows. Section 2 presents the definition of LRCCs, establishes certain useful properties including the connection to canonical coherences, and illustrates how they are naturally related to the property of asymptotic independence between two standardized sums. Section 3 discusses consistent estimation of LRCCs, proposes two statistics for testing certain LRCCs are zero and derives the limiting distributions of these statistics under the null. Section 4 illustrates the finite sample performance of the statistics. Section 5 concludes. A mathematical appendix contains both proofs of the main results and some extensions.

[1] See Anderson (2003) [Chapter 12].

2 Long Run Canonical Correlations

2.1 Definitions

To present a formal definition of LRCCs we first introduce the following notations and assumptions. Let x_t and z_t be $p\times 1$ and $q\times 1$, respectively, where $q \geq p$, and set $v_t = (x_t', z_t')'$; also set $X_T \equiv T^{-1/2}\sum_{t=1}^{T} x_t$, $Z_T \equiv T^{-1/2}\sum_{t=1}^{T} z_t$ and $V_T \equiv T^{-1/2}\sum_{t=1}^{T} v_t$. The long run variance of v_t is denoted $\lim_{T\to\infty} Var[V_T] = \Sigma_{vv}$ where Σ_{vv} is partitioned conformably with v_t,

$$\Sigma_{vv} = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}. \tag{1}$$

In our analysis, we require Σ_{vv} to be finite and positive definite. We make the following assumption.

Assumption 1(l) *$\{v_t\}$ is a mean zero, stationary up to order $2l$, strongly mixing with cumulant functions of order 2 to $2l$ that are absolutely summable.*[2]

Let $\Gamma_{vv}(h) = Cov[v_t, v_{t-h}]$ be the autocovariance function of $\{v_t\}$. The absolute summability of the second order cumulants amounts to $\sum_{h=-\infty}^{+\infty} \|\Gamma_{vv}(h)\| < \infty$[3] and guarantees the existence of the long run variance Σ_{vv}. It is worth mentioning that the stationarity condition in Assumption 1(l) is not necessary for the long run variance Σ_{vv} to exist. If $\{v_t\}$ is a strong mixing process but not stationary, Lemma 1 of Andrews (1991) proposes some restrictions on the rate of decay of the mixing coefficients that guarantee the existence of Σ_{vv} as well as the absolute summability of the fourth order cumulants.

With this background, the LRCCs are defined as follows.

Definition 1 *Let v_t satisfy Assumption 1(1). The population long run canonical correlations between x_t and z_t are denoted by $\{\rho_i; i = 1, 2, \ldots, p\}$, where by convention $\rho_i \geq 0$ for $i = 1, ..., p$, and $\rho_i \geq \rho_{i+1}$ for $i = 1, 2, \ldots, p-1$, and have the following properties:*

(i) $\{\rho_i^2\}$ are the solutions to the determinantal equation $|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\Sigma_{xx}| = 0$;

[2] Note that the quantity in parentheses in the Assumption number indicates (half) the order up to which the process is assumed stationary and its cumulants are assumed absolutely summable.

[3] Throughout this paper, $\|A\| \equiv (trace(AA'))^{1/2}$ for any matrix A.

(ii) $\{\rho_i^2\}$ *are the* p *largest solutions to the determinantal equation* $|\Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz} - \rho^2\Sigma_{zz}| = 0$*; and*

(iii) $\rho_i = \alpha_i'\Sigma_{xz}\beta_i$ *where* α_i *and* β_i *satisfy* $(\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\Sigma_{xx})\alpha_i = 0$ *and* $(\Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz} - \rho^2\Sigma_{zz})\beta_i = 0$ *for* $i = 1, 2, \ldots, p$.

Remark 1: A comparison with the definition of the CCCs (*e.g.* see Rao (1973)[p. 582-3]) indicates that the only difference between CCCs and LRCCs is that the CCCs are defined via determinantal equations involving contemporaneous variances and covariances and LR-CCs are defined via determinantal equations involving long run variances and covariances. As a consequence, the LRCCs between x_t and z_t can be equivalently interpreted as the (limiting) canonical correlations between X_T and Z_T.

Remark 2: As a consequence of Remark 1, ρ_i can be equivalently defined via sequential constrained optimization in which α_i and β_i are chosen to maximize the correlation between $\alpha_i'X_T$ and $\beta_i'Z_T$ subject to the constraints that $\alpha_i'\Sigma_{xx}\alpha_i = 1$, $\alpha_j'\Sigma_{xx}\alpha_i = 0$, $\beta_i'\Sigma_{zz}\beta_i = 1$, $\beta_j'\Sigma_{zz}\beta_i = 0$, $i \neq j$; see Rao (1973)[p. 582-3] for a description of the sequential derivation of CCCs.

Remark 3: From the definition, it follows that LRCCs can be interpreted as canonical coherences at frequency zero. Hannan (1970)[p. 298] defines the canonical coherences between x_t and z_t at frequency λ to be $\{\rho_i(\lambda)\}_{i=1}^p$, the (positive) solutions to the determinantal equation

$$|f_{xz}(\lambda)[f_{zz}(\lambda)]^{-1}f_{zx}(\lambda) - \rho(\lambda)^2 f_{xx}(\lambda)| = 0,$$

where $f_{xx}(\cdot)$ and $f_{zz}(\cdot)$ are the spectral density matrices of x and z respectively and $f_{xz}(\cdot)$ is the cross-spectral density matrix between x and z.[4] The equivalence then follows directly from the definitions of the two quantities upon noting that $\Sigma_{vv} = 2\pi f_{vv}(0)$.[5]

[4] The spectral density matrix of v at frequency λ is defined to be $f_{vv}(\lambda) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-i\lambda h}\Gamma_{vv}(h)$ and $f_{vv}(\cdot)$ is partitioned into $f_{xx}(\cdot)$, $f_{zz}(\cdot)$, $f_{xz}(\cdot)$ and $f_{zx}(\cdot)$ conformably with the partition of Σ_{vv} in (1).

[5] See Hannan (1970)[Corollary 4, p. 208].

2.2 Examples

Define $v_t = (x_t', z_t')'$ and $V_T(r) = [X_T(r)', Z_T(r)']' = T^{-1/2} \sum_{t=1}^{[Tr]} v_t$ where $r \in [0,1]$ and $[Tr]$ denotes the integer part of Tr. Phillips & Durlauf (1986) provide conditions under which $V_T(r) \Rightarrow B_n(r)$ where $B_n(r)$ denotes a multivariate Brownian motion with variance Σ_{vv}, $n = p + q$ and $\Rightarrow$ denotes weak convergence. From these results, it follows that $X_T(r)$ and $Z_T(r)$ are asymptotically independent if $\Sigma_{xz} = 0_{p\times q}$, the $p \times q$ null matrix. From Definition 1, it follows that this condition for asymptotic independence can be equivalently stated as $\rho_i = 0$ for $i = 1, 2, \ldots, p$.

Remark 4: We note that the hypothesis of asymptotic independence of X_T and Z_T is different from the hypothesis that x_t and z_t are independent. The latter hypothesis involves the restriction that $Cov[x_t, z_s] = 0$ for all t, s. Clearly the latter restriction is sufficient but not necessary for $\Sigma_{xz} = 0$.[6]

We now present 4 examples of where the concept of LRCCs arises in models of interest in econometrics. In examples 1 and 2, it is the asymptotic independence of two multivariate Brownian motions that is the critical issue for the inference described. Examples 3 and 4 involve cases where the properties of various estimators depend on whether or not a certain long run covariance is zero.

Example 1: Testing for cointegration

Consider the following cointegration model,

$$\underset{1\times 1}{y_t} = \underset{1\times 1}{\alpha_0} + \underset{1\times k}{x_t'}\underset{k\times 1}{\beta_0} + \underset{1\times 1}{u_t}, \qquad t = 1, 2, \cdots, \tag{2}$$

where u_t is an $I(0)$, zero-mean process, and the regressor vector x_t is an $I(1)$ process:

$$x_t = x_{t-1} + w_t, \tag{3}$$

where w_t is an $I(0)$, zero-mean process, and there are no cointegrating relations among the x_t (x_0 is an arbitrary random vector). Let $\{v_t \equiv (u_t, w_t')'\}$ be a strictly stationary, weakly

[6]Tests for the independence of two time series have been proposed by Haugh (1976), El Himdi & Roy (1997), Hong (1996) and Bouhaddioui & Roy (2004).

dependent stochastic process with finite second moments. For what follows, it is convenient to decompose the long run variance of v_t, Σ_{vv} as

$$\Sigma_{vv} \;=\; \Gamma_{vv}(0) \,+\, \Lambda_{vv} \,+\, \Lambda'_{vv}$$

where $\Lambda_{vv} = \sum_{h=1}^{\infty} \Gamma_{vv}(h)$. Assume $V_T(r) \Rightarrow [B_u(r), B_w(r)']'$, a $(k+1) \times 1$ standard Brownian motion where $B_u(r)$ is a scalar. Let $\hat{\alpha}_T, \hat{\beta}_T$ be the OLS estimators of α_0 and β_0 from the regression of y_t on 1, x_t for $t = 1, 2 \cdots, T$. Park & Phillips (1988) derive the limiting distribution of $T(\hat{\beta}_T - \beta_0)$:

$$T(\hat{\beta}_T - \beta_0) \Rightarrow \left[\int_0^1 \bar{B}_w(r)'\bar{B}_w(r)\, d(r)\right]^{-1} \left[\int_0^1 \bar{B}_w(r)' dB_u(r) + \Delta_{uw}\right], \tag{4}$$

where $\bar{B}_w$ denotes the demeaned process B_w and $\Delta_{uw} = \Gamma_{uw}(0) + \Lambda_{uw}$ and all matrices in the latter equation are partitioned conformably with Σ_{vv}.[7]

As noted by Park & Phillips (1988), the limiting distribution (4) depends, in an intractable way, on the nuisance parameters Σ_{uw} and Δ_{uw}. However, if $\Sigma_{uw} = 0$ then the distribution becomes much simpler because B_u and B_w are independent and it can be shown that

$$\hat{\Sigma}_{uu}^{-1/2} A_T^{1/2} T(\hat{\beta}_T - \beta_0) - \hat{\Sigma}_{uu}^{-1/2} A_T^{-1/2} \hat{\Delta}_{uw} \Rightarrow \mathcal{N}(0,\, I_k) \tag{5}$$

where $A_T = T^{-2} \sum_{t=1}^{T} (x_t - \bar{x}_T)(x_t - \bar{x}_T)'$, $\bar{x}_T = T^{-1} \sum_{t=1}^{T} x_t$, $\hat{\Sigma}_{uu} \xrightarrow{p} \Sigma_{uu}$ and $\hat{\Delta}_{uw} \xrightarrow{p} \Delta_{uw}$. Inference can be performed about β_0 based on (5) in a straightforward fashion. Thus, in this example, the condition that $\Sigma_{uw} = 0_{k \times 1}$ - or equivalently that the LRCC between u_t and w_t is zero - is of practical relevance.[8] $\diamond$

Example 2: Test for structural break dates in cointegrated regression models

Kejriwal & Perron (2008) have recently proposed an inference in cointegrated models with multiple structural changes allowing both stationary and integrated regressors. In particular, they propose an inference for multiple structural break dates, $T_j, j = 1, \ldots, m+1$, in

[7] $\bar{B}_w(r) = B_w(r) - \int_0^1 B_w(s)ds$.

[8] Park & Phillips (1988) note that if x_t is strictly exogenous, in the sense that $E(\Delta x_t' u_s) = 0$, $\forall t$ and s, then $\Sigma_{uw} = \Delta_{uw} = 0_{1 \times k}$ and the second term on the left hand side of (5) can be omitted. Notice this is sufficient but not necessary for $\Sigma_{uw} = 0_{1 \times k}$.

models such as

$$\begin{aligned} y_t &= \beta_{0j} + \beta_{1j}' x_t + \beta_{2j}' z_t + u_t \quad (T_{j-1} < t \leq T_j) \\ z_t &= z_{t-1} + w_t, \end{aligned} \tag{6}$$

where x_t is $I(0)$, $T_0 = 0$ and $T_{m+1} = T$. Kejriwal & Perron (2008) consider the limiting behaviour of the break points obtained by minimizing an OLS criterion based on (6). They show that the limiting distribution of these OLS break points depends crucially on whether the long-run covariance Σ_{uw} between u_t and w_t is equal to 0, and present this distribution for the case in which $\Sigma_{uw} = 0$. Therefore, it is of importance to assess whether the LRCC between u_t and w_t are all zero in this context.[9] $\diamond$

Example 3: GEL weighting in moments estimation

Consider that we are interested in estimating $\mu = Eg(x_t) \in \mathbb{R}^r$ and we have available an overidentifying moment condition model of the form

$$E[f(x_t, \theta_0)] = 0 \tag{7}$$

that describes x_t. It is well known, at least since Back & Brown (1993), that such an overidentifying moment condition model is also informative about the distribution of the underlying random time-dependent process $\{x_t : t = 1, \dots, T\}$. Any moment of x_t such as μ can therefore be more efficiently estimated than by the usual sample mean using this extra-information through the so-called implied probabilities. We show next that the condition of this superior efficiency of the estimation of μ can be expressed in terms of LRCCs.

From the smoothed maximum empirical likelihood estimation theory introduced by Kitamura (1997) and extended by Smith (2004) to the smoothed Generalized Empirical Likelihood (GEL) settings, the implied probabilities $\hat{p}_i : i = 1, \dots, Q$ are defined as an optimal empirical distribution of a smoothed version of the estimating function $f(x_t, \theta)$.

[9] When $\Sigma_{uw} \neq 0$, Kejriwal & Perron (2008) propose the use of dynamic-OLS (DOLS) and the asymptotic distribution of the structural break dates remains valid if the leads and lags orders are chosen appropriately. In samples of reasonable size for empirical applications, the proposed inference with OLS outperforms the DOLS when Σ_{uw} is actually equal to 0. For this reason, it is in the interest of applied researchers to choose between OLS and DOLS after investigating whether $\Sigma_{uw} = 0$ or not.

The induced estimator of μ is $\hat{\mu} = \sum_{i=1}^{Q} \hat{p}_i G_i$, where G_i are the smoothed version of $\{g(x_t) : t = 1, \ldots, T\}$. We can show that the asymptotic variance of $\hat{\mu}$ is $V = \Sigma_{gg} - \Sigma_{gf}\Sigma_{ff}^{-1/2}\left(Id_r - \Sigma_{ff}^{-1/2}\Gamma\left(\Gamma'\Sigma_{ff}^{-1}\Gamma\right)^{-1}\Gamma'\Sigma_{ff}^{-1/2}\right)\Sigma_{ff}^{-1/2}\Sigma_{fg}$, with $\Gamma = E\left[\partial f(x_t,\theta)/\partial\theta'\right]|_{\theta=\theta_0}$, $\Sigma \equiv \lim_{T\to\infty} \text{Var}\left(T^{-1/2}\sum_{t=1}^{T} h_t\right)$; $h_t = (f'(x_t,\theta_0), g'(x_t))'$; and Σ is partitioned conformably with h_t into Σ_{ff}, Σ_{fg}, Σ_{gf}, Σ_{gg}. (see Appendix for the details on the derivation of V.)

From the expression of V, the superior efficiency of $\hat{\mu}$ over the sample mean vanishes if the long run covariance $\Sigma_{fg} = 0$, or equivalently, the LRCCs between $f(x_t,\theta_0)$ and $g(x_t)$ are all zero. $\diamond$

Example 4: System estimation using GMM

Consider the case in which two non-overlapping parameter vectors γ_0 and δ_0 are estimated via GMM based on the information in the population moment conditions, $E[g(c_t,\gamma_0)] = 0$ and $E[h(d_t,\delta_0)] = 0$ respectively, where c_t and d_t are two data vectors that may include common variables. The parameters can be estimated via individual estimations that is, based on separate GMM estimations using the appropriate moment condition, or from a system estimation that obtains estimates from a single GMM estimation based on combining the two moment conditions. Intuition suggests correctly that the system estimation can never yield less efficient estimators than those obtained from the individual estimations. However, system estimation is not guaranteed to provide any efficiency gains, and so, given the increased computational complexity of system estimation, it is not clear that system estimation is preferable from a practical viewpoint. We now demonstrate that the condition for no gains from system estimation can be expressed in terms of LRCCs.

To this end, let $\theta = (\gamma', \delta')'$, $f(\cdot) = [g(\cdot)'\ h(\cdot)']'$, e_t contain the distinct elements of c_t and d_t, and $\Sigma = lim_{T\to\infty} Var[T^{-1/2}\sum_{t=1}^{T} f(e_t,\theta_0)]$; also define $F = E[\partial f(e_t,\theta)/\partial\theta']|_{\theta=\theta_0}$. Partition Σ into Σ_{gg}, Σ_{hh}, Σ_{gh}, and Σ_{hg} conformably with $f(\cdot)$; assume that $\hat{\Sigma}$ is a consistent estimator of Σ and $\hat{\Sigma}_{aa}$ is a consistent estimator of Σ_{aa} for $a = g, h$. Define $G = E[\partial g(c_t,\gamma)/\partial\gamma']|_{\gamma=\gamma_0}$ and $H = E[\partial h(d_t,\delta)/\partial\delta']|_{\delta=\delta_0}$. Let $\hat{\theta}_T$ be the GMM estimator based on individual estimations that is, the GMM based on $E[f(e_t,\theta_0)] = 0$ with weighting matrix $W_T^{(1)} = diag[\hat{\Sigma}_{gg}^{-1}, \hat{\Sigma}_{hh}^{-1}]$, and $\tilde{\theta}_T$ be the system GMM estimator, that is the GMM estimator based on $E[f(e_t, \theta_0)] = 0$ with weighting matrix $W_T^{(2)} = \hat{\Sigma}^{-1}$.

Using standard first order arguments[10] it can be shown that $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, V^{(1)})$ and $T^{1/2}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, V^{(2)})$ where $V^{(1)} = diag[(G'\Sigma_{gg}^{-1}G)^{-1}, (H'\Sigma_{hh}^{-1}H)^{-1}]$ and $V^{(2)} = (F'\Sigma^{-1}F)^{-1}$. Using the partitioned inversion formula (*e.g.* Magnus & Neudecker (1991)[p.11]) and taking account of the structure of $f(.)$, it follows that $V^{(1)} = V^{(2)}$ if $\Sigma_{gh} = 0$ or, equivalently, if the LRCCs between $g(c_t, \gamma_0)$ and $h(d_t, \delta_0)$ are all zero. $\diamond$

3 Inference

We consider estimation of LRCCs based on solving the equations analogous to Definition 1 *(i)-(iii)* only with the population long run variances and covariances replaced by consistent estimators.

Definition 2 *Let $\hat{\Sigma}_{xx} \xrightarrow{p} \Sigma_{xx}$, $\hat{\Sigma}_{zz} \xrightarrow{p} \Sigma_{zz}$ and $\hat{\Sigma}_{xz} \xrightarrow{p} \Sigma_{xz}$. The sample long run canonical correlations between x_t and z_t are denoted by $\{r_i; i = 1, 2, \ldots, p\}$, where by convention $r_i \geq 0$ for $i = 1, \ldots, p$, and $r_i \geq r_{i+1}$ for $i = 1, 2, \ldots, p-1$, and have the following properties:*

(i) $\{r_i^2\}$ *are the solutions to the determinantal equation* $|\hat{\Sigma}_{xz}\hat{\Sigma}_{zz}^{-1}\hat{\Sigma}_{zx} - r^2\hat{\Sigma}_{xx}| = 0$;

(ii) $\{r_i^2\}$ *are the p largest solutions to the determinantal equation* $|\hat{\Sigma}_{zx}\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xz} - r^2\hat{\Sigma}_{zz}| = 0$;

(iii) $r_i = \hat{\alpha}_i{}'\hat{\Sigma}_{xz}\hat{\beta}_i$, *which is the positive square root of the i-th generalized eigenvalue, where* $\hat{\alpha}_i$ *and* $\hat{\beta}_i$ *are the corresponding i-th generalized eigenvectors associated with* r_i^2 *in (i) and (ii), respectively.*

We consider the class of Heteroscedasticity and Autocorrelation Consistent Covariance (HAC) estimators $\hat{\Sigma}_{vv}$ of Σ_{vv}; see Andrews (1991). By definition, $\hat{\Sigma}_{vv} = 2\pi\hat{f}_{vv}(0)$ where $\hat{f}_{vv}(0)$ is the kernel estimator of the spectral density matrix of v_t at frequency zero as

[10] For example, Hall (2005)[Chap. 3.4 & 3.6].

introduced by Parzen (1957), that is

$$\hat{f}_{vv}(\lambda) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} k(B_T h) \hat{\Gamma}_{vv}(h) e^{-ih\lambda}, \quad -\infty < \lambda < \infty, \tag{8}$$

where $\hat{\Gamma}_{vv}(h)$ is the sample autocovariance function,

$$\hat{\Gamma}_{vv}(h) = \frac{1}{T} \sum_{t=h+1}^{T} v_t v'_{t-h}, \quad \text{for} \quad h \geq 0, \quad \text{and} \quad \hat{\Gamma}_{vv}(-h) = \hat{\Gamma}'_{vv}(h), \tag{9}$$

$k(\cdot)$ is the covariance averaging kernel or the lag window generator and the bandwidth parameter B_T is a sequence of constants tending to 0, as $T \to \infty$ in such a way that $B_T T \to \infty$. $k(\cdot)$ satisfies the following assumption.

Assumption 2 *$k(\cdot) : \mathbb{R} \to \mathbb{R}$ is piecewise continuous, continuous at zero with $k(0) = 1$, symmetric about zero, absolutely integrable and is such that $xk(x)$ is bounded.*

The class of kernels described by Assumption 2 is the one considered by Rosenblatt (1984).[11] The boundedness of $xk(x)$ is not particularly restrictive as it includes most popular choices of the kernels used in practice.

To derive the properties of the sample LRCCs and also the inference procedures discussed below, we need to characterize the large sample behaviour of $\hat{f}_{vv}(0)$.

Lemma 1 *If Assumptions 1(4) and 2 hold, $k(\cdot)$ has $r > 0$ as characteristic exponent, $\|f^{(r)}_{vv}(0)\| < \infty$ and $B_T = o\left(T^{-\frac{1}{1+2r}}\right)$, then:*

$$\sqrt{\frac{\nu}{2}} \left(\hat{f}_{vv}(0) - f_{vv}(0) \right) \xrightarrow{d} \mathcal{N}(0, V(0)), \tag{10}$$

where $\nu = \frac{2B_T T}{\int_{-\infty}^{\infty} k^2(x)dx}$, $V_{kl,k'l'}(\lambda) = \{f_{vv}(0)\}_{kk'}\{f_{vv}(0)\}_{ll'} + \{f_{vv}(0)\}_{kl'}\{f_{vv}(0)\}_{k'l}$, $1 \leq k, k', l, l' \leq p + q$, and $\{M\}_{ij}$ denotes the (i, j)-th element of the matrix M and $f^{(r)}_{vv}$ is the generalized r-th spectral derivative of f_{vv}.

The value for ν is given in Hannan (1970)[Table 1, p.282].[12] The terms $V_{kl,kl}(0)$ are given by Rosenblatt (1984) (see in particular his comments on p. 1178) while one can refer

[11] This class is essentially included in the class of kernels $\mathcal{K}_1$ considered by Andrews (1991). Note that the class $\mathcal{K}_1$ defined by Andrews (1991) requires the absolute integrability of $k(\cdot)$ instead of the square-integrability; see footnote of Andrews & Monahan (1992)[p. 955].

[12] Also see the Supplementary Appendix available from the authors upon request.

to Hannan (1970)[Ch. V, Th. 9] for the off-diagonal terms of $V(0)$. Lemma 1 follows from Rosenblatt (1984)[Corollary 3] and the formulae for the bias of the spectral density estimator in Parzen (1957).

Politis & Romano (1992) refer to the choice of bandwidth B_T suggested by Lemma 1 as *undersmoothing*. The minimization of the MSE of $\hat{f}_{vv}(0)$ requires $B_T \propto T^{-\frac{1}{1+2r}}$. Therefore, the choice of B_T defined by Lemma 1 would give in general larger weights to the autocovariances of order close to 0. By doing so, the resulting spectral density estimator would not smooth the sample autocovariances enough to be optimal. However, they also recognize the necessity of such undersmoothing to obtain the type of central limit result in Lemma 1.

Since $\Sigma_{vv} = 2\pi f_{vv}(0)$, it follows that under the conditions of Lemma 1 we have that $2\pi \hat{f}_{vv}(0) \xrightarrow{p} \Sigma_{vv}$. Using these results, we can establish the consistency of the sample LRCCs.

Proposition 1 *Let $\{r_i^2\}$ be the sample LRCCs defined in Definition 2 with $\hat{\Sigma}_{vv} = 2\pi \hat{f}_{vv}(0)$. If the conditions of Lemma 1 hold then $r_i^2 \xrightarrow{p} \rho_i^2$ for $i = 1, 2, \ldots, p$ where $\{\rho_i\}$ are the LRCCs between x_t and z_t.*

Proposition 1 follows via the Continuous Mapping Theorem from the consistency of the long run variance/covariance matrix estimators and the continuity of the eigenvalues as a function of the underlying matrices; *e.g.* see Hiriart-Urruty & Ye (1995).

We now consider inference about the LRCCs. Given the examples in Section 2, it is desirable to derive a test for the hypothesis that the smallest k LRCCs are all zero. To our knowledge, such an appropriate test statistic has not been presented in the literature. Brillinger (1981) derives the limiting distribution of the sample canonical coherences when their population analogs are distinct which is different from our desired hypothesis if $k > 1$. Hannan (1970) considers testing our hypothesis of interest but only develops a test statistic for the case in which $p = 1$. To our knowledge, the extension to $p > 1$ has not been considered in the literature.

Let $r_{p-k+1} \geq r_{p-k+2} \geq \cdots \geq r_p$ be the k smallest estimated long run canonical correlations between x_t and z_t derived from $\hat{\Sigma}_{vv} = 2\pi \hat{f}_{vv}(0)$ in Definition 2. We consider two test statistics:

$$LR_T = -\frac{\nu}{2}\sum_{j=1}^{k}\ln\left(1 - r_{p-k+j}^2\right), \tag{11}$$

and

$$H_T = \frac{\nu}{2}\sum_{j=1}^{k} r_{p-k+j}^2. \tag{12}$$

The functional forms are chosen because LR_T mimics the likelihood ratio statistic for testing contemporaneous independence between two normal vectors in multivariate statistics (see Anderson (2003)). In fact, as we establish as a step of the proof of Theorem 1 below, $\hat{f}_{vv}(0)$ is asymptotically distributed as the sample covariance of $\nu/2$ independent Gaussian random variables with mean 0 and variance $f_{vv}(0)$. This justifies LR_T as the likelihood ratio test statistic for exactly k canonical correlations equal 0. H_T is a generalization of the statistic proposed by Hannan (1970)[p. 290] for the case in which $p = 1$.

The large sample behaviour of these statistics is given in the following theorem, the proof of which is relegated to the appendix.

Theorem 1 *If the conditions of Lemma 1 hold and Σ_{vv} is positive definite, then under $H_0 : \rho_{p-k+1} = \rho_{p-k+2} = \ldots = \rho_p = 0$, LR_T and H_T are asymptotically distributed as a $\chi^2_{k(q-p+k)}$ as T grows to infinity.*

Remark 5: The degrees of freedom agree with the conventional result from multivariate statistics for testing the independence of two normal vectors; see Anderson (2003)[Ch. 9].

If v_t depends on a parameter θ_0 of finite dimension which is estimated by $\hat{\theta}_T$ (as would be the case in Example 4 above, say), Theorem 2 below ensures that one can base the test for k LRCCs equal 0 on $v_t(\hat{\theta}_T)$ and still rely on the test statistics and the asymptotic distributions given by Theorem 1. We make the following assumptions.

Assumption 3 *(i) The conditions of Lemma 1 hold with v_t replaced by*

$$\left(v_t(\theta_0)', vec\left(\frac{\partial}{\partial\theta'}v_t(\theta_0) - E\frac{\partial}{\partial\theta'}v_t(\theta_0)\right)'\right)',$$

(ii) $\sup_{t\geq 1} E\sup_{\theta\in\Theta}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}v_{at}(\theta)\right\|^2 < \infty$, *$\forall a = 1,\ldots,p+q$ where* $v_t(\theta) = (v_{1t}(\theta),\ldots,v_{p+q,t}(\theta))'$, *(iii)* $\sqrt{T}(\hat{\theta}_T - \theta_0) = O_p(1)$, *(iv)* $\sup_{t\geq 1} E\|v_t(\theta_0)\|^2 < \infty$, *(v)* $\sup_{t\geq 1} E\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta'}v_t(\theta)\right\|^2 < \infty$.

Assumption 3 is quite standard in the literature of long run variance estimation. It ensures the asymptotic equivalence of the long run variance estimators obtained using the parameter estimate $\hat{\theta}_T$ and the true parameter value θ_0. (See Andrews (1991).) The validity of the test of exactly k null LRCCs is given by the following theorem.

Let $LR_T(\hat{\theta}_T)$ and $H_T(\hat{\theta}_T)$ denote the likelihood ratio and the generalized Hannan test statistics based on $\hat{f}_{\hat{v}\hat{v}}(0)$, with $\hat{v} = v_t(\hat{\theta}_T)$. We have:

Theorem 2 *If Assumption 3 holds and* $H_0 : \rho_{p-k+1} = \rho_{p-k+2} = \ldots = \rho_p = 0$ *(where* ρ_i *is the* i^{th} *LRCC between* $x_t(\theta_0)$ *and* $z_t(\theta_0)$*),* $LR_T(\hat{\theta}_T)$ *and* $H_T(\hat{\theta}_T)$ *are asymptotically distributed as* $\chi^2_{k(q-p+k)}$ *as* T *grows to infinity.*

4 Simulation Study

In this section we report results from simulation studies designed to shed light on the finite sample properties of the test statistics introduced in the previous section. In Section 4.1, the design involves an application of the statistics to testing whether the long run covariance of two stationary random vectors is zero. Sections 4.2 and 4.3 involve applications of the tests in the context of the cointegration model in Example 1 (in Section 2.2) and the system estimation using GMM setting in Example 4, respectively.

4.1 Application to testing zero long run covariance of two stationary random vectors (Experiment 1)

We consider the random process $x_t \in \mathbb{R}^4$ with

$$x_t = R_1 x_{t-1} + R_2 \varepsilon_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim NID(0, \Omega)$.

We set $R_1 = 0.5Id_4$ and $R_2 = 0$ in Design 1 corresponding to an AR(1) dynamics for each component of x_t. Design 2 corresponds to an MA(1) dynamics for each component with $R_1 = 0$ and $R_2 = 0.5Id_4$ while $R_1 = R_2 = 0.5Id_4$ in Design 3, corresponding to an ARMA(1,1) dynamics for each component of x_t.

Let $y_t = (x_{1t}, x_{2t})'$ and $z_t = (x_{3t}, x_{4t})'$ and ρ_1, ρ_2 ($\rho_1 \leq \rho_2$) the LRCCs between y_t and z_t. We generate samples of x_t of size $T = 50$, 100, 200 and 1,000 and perform the tests for:

$$H_{01} \;:\; \rho_1 = 0 \quad \text{and} \quad \rho_2 \neq 0$$

and

$$H_{02} \;:\; \rho_1 = \rho_2 = 0.$$

Under H_{01}, LR_T and H_T are asymptotically distributed as χ^2_1 and under H_{02}, LR_T and H_T are asymptotically distributed as χ^2_4.

The true values of ρ_1 and ρ_2 are set up by the choice of Ω. We consider 4 cases, denoted *Case i*, $i = 1, 2, 3, 4$, for which $\Omega = \Omega_i$ with :

$$\Omega_1 = \begin{pmatrix} 1 & 0.6 & 0 & 0 \\ 0.6 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.6 \\ 0 & 0 & 0.6 & 1 \end{pmatrix}, \quad \Omega_2 = \begin{pmatrix} 1 & 0.6 & 0.6 & 0 \\ 0.6 & 1 & 0.6 & 0 \\ 0.6 & 0.6 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\Omega_3 = \begin{pmatrix} 1 & 0.6 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 & 0.6 \\ 0.6 & 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 0.6 & 1 \end{pmatrix}, \quad \text{and} \quad \Omega_4 = \begin{pmatrix} 1 & 0.6 & 0.6 & 0.1 \\ 0.6 & 1 & 0.6 & 0.3 \\ 0.6 & 0.6 & 1 & 0.2 \\ 0.1 & 0.3 & 0.2 & 1 \end{pmatrix}.$$

For any of the three designs, we have that: H_{02} is true for Case 1; H_{01} is true for Cases 2 and 3; H_{01} and H_{02} are both false for Case 4.

We evaluate the performance of the LRCC tests through 10,000 Monte Carlo replications. The simulated rejection rates for the tests are displayed by Tables 1 through 3. We obtain the estimated LRCCs using the usual HAC estimator of the long run variance with the Bartlett kernel. We also report the results based on the prewhitening and recolouring Bartlett kernel estimates of the long run variance. (See Andrews & Monahan (1992).) We rely on the AR(1) adjustment for each component of x_t for the prewhitening step.

From Tables 1-3, it can be seen that the rejection rates are more sensitive to the choice of kernel than to the choice of test statistic. The use of the Bartlett kernel alone yields tests with approximately correct size for Design 2, the VMA(1), but yields oversized tests for both Designs 1 and 3 in which there is a VAR component present in the data. In contrast, the use of the Bartlett kernel with prewhitening and recolouring yields tests that have approximately the correct size for samples with $T \geq 250$. In terms of power, the two tests perform comparably.

4.2 Application to cointegration regressions (Experiment 2)

In this section we investigate the performance of LR_T and H_T in the context of the cointegration example in Section 2. The data are generated via (2)-(3) with $k = 1$, $\alpha_0 = 1.0$; $\beta_0 = 2.0$. We consider three cases for the error process $[u_t, w_t]$ all of which fit within the following framework

$$\begin{aligned} u_t &= \gamma w_t + a_t + \theta\varepsilon_{t-1} - \theta\varepsilon_{t-2} \\ w_t &= b_t + \theta\varepsilon_{t-1} + \theta\varepsilon_{t-2} \end{aligned}$$

$$(a_t, b_t, \varepsilon_t)' \overset{i.i.d}{\sim} \mathcal{N}(0_3, I_3).$$

- *Case 1:* $\gamma = \theta = 0$; under these conditions x_t is strictly exogenous and so inference about β_0 can be legitimately based on

$$\hat{\Sigma}_{uu}^{-1/2} A_T^{1/2} T(\hat{\beta}_T - \beta_0) \Rightarrow \mathcal{N}(0,\, 1). \tag{13}$$

- *Case 2:* $\gamma = 0$ and $\theta \neq 0$; under these conditions it can be shown that $\Sigma_{uw} = 0$ and $B_u(r)$ and $B_w(r)$ are independent, and so inference can be performed about β_0 based on (5). We choose $\theta = 0.5$.

- *Case 3:* $\gamma \neq 0$; under this condition $\Sigma_{uw} \neq 0$. We choose values $\gamma = \pm 0.4, \pm 0.8$, and for simplicity set $\theta = 0$.

Finally, the sample size: $T = 50, 100, 250, 500,$ and 1000. We calculate the empirical size of LR_T and H_T when the nominal size is 5%, and also compute the empirical coverage probability of the 95% confidence interval of the estimator of the slope parameter β_0 based on (5) and (13). All results are computed using 10,000 simulations.

Table 4 presents the simulated rejection rates of LR_T and H_T for *Cases 1-3.* As can be seen, both tests have rejection rates close to the nominal size of the test for both *Case 1* and *Case 2*, that is, for the cases where their null hypothesis is satisfied. In *Case 3*, the null is false and under both tests, rejection rates are substantially higher than the nominal size and are clearly converging to one as the sample size increases. This evidence suggests that both LR_T and H_T perform well. The evidence also suggests that neither LR_T nor H_T dominates the other.

Table 5 presents the simulated coverage rates of 95% confidence intervals for β_0 based on the results in (13) and (5). Recall that (13) is only valid in *Case 1*, and it can be seen from Table 5 that it is only in this case that the simulated coverage rate converges to 0.95 as T increases. In contrast, (5) is valid in both *Case 1* and *Case 2*, and the simulated coverage rates converge to 0.95 for both these cases as T increases. Neither confidence interval is valid for *Case 3*, and this is reflected in the simulated coverage rates. Interestingly, the coverage rate is closer to the nominal level for the interval based on (5) but the coverage rate is nevertheless too low by 0.1 even in the largest sample considered here.

Taken together, the results in Tables 4 and 5 indicate the practical importance of the restriction that $\Sigma_{wu} = 0$ in this model, and that LR_T and H_T offer a method for determining when this restriction is satisfied.

4.3 Application to system GMM estimation (Experiment 3)

In this section, we provide an illustration by Monte Carlo experiments of the possibility of efficiency gain when the LRCCs between the initial estimating function and the additional ones are non-zeros in the context of moment condition models. We consider the data generating process:

$$x_t = Rx_{t-1} + \varepsilon_t,$$

with $x_t = (y_t, z_t)'$ and $\varepsilon_t \sim NID(0, (1, \omega, 1))$ and $R = (r_1|r_2)$; $r_1 = (0.8, 0)'$ and $r_2 = (0, 0.5)'$.

The moment restrictions that we consider are:

$$E\left(y_t - m_1\right) = 0, \quad E\left(y_t^3 - m_3\right) = 0, \quad E\left(y_t^5 - m_5\right) = 0, \tag{14}$$

$$E\left(z_t - \mu\right) = 0, \quad E\left(z_t - \mu\right)^3 = 0, \quad E\left(z_t - \mu\right)^5 = 0 \tag{15}$$

in which the parameters of interest are m_1, m_3 and m_5. We estimate these parameters using the just identifying moment restrictions in (14) that we label *Model 1* and we also estimate m_1, m_3 and m_5 along with μ using *Model 2* corresponding to the moment conditions (14) and (15), jointly. The correlation coefficient ω controls the LRCCs between the estimating function in (14) and (15). If $\omega = 0$, these LRCCs are all equal to 0 and they increase as ω increases. As (15) is overidentifying for μ, we expect m_3 and m_5 to be more efficiently estimated by *Model 2*, for $\omega > 0$. Note that we do not expect any such improvement in the estimation of m_1 because within our design the moment condition for m_1 is proportional to the score vector for estimation of the mean of y.

Table 6 displays the simulated rejection rates of the LRCC tests of $H_0 : \rho_1 = \rho_2 = \rho_3 = 0$. In terms of size properties, $\omega = 0$, it can be seen that the tests are slightly undersized for $T \geq 100$ and do not exhibit size equal to the nominal level at even $T = 1000$. This contrasts with our findings in Sections 4.1 and 4.2. We attribute this difference to the fact that the previous two designs involve linear models and the one here involves a moment condition that depends on polynomial moments. We conjecture that in this context the prewhitening/recolouring tends to underreject in our experiments except when we prewhiten with the true AR(1) structure as is the case in Experiment 1. Nevertheless, the tests can clearly discriminate between zero and non-zero LRCCs in this context as well, albeit with conservative size.

Table 7 reports the simulated bias and RMSE of these estimates as well as the simulated-average asymptotic variance of the corresponding estimators. The results are obtained from 10,000 Monte Carlo replications. We consider the sample sizes $T = 50$, 100, 250, 500 and $1,000$ and $\omega = 0$, $\omega = 0.6$ and $\omega = 0.8$. For the reason stated above we treat discussion of the estimation of m_1 and (m_3, m_5) separately. Consider first the estimators of (m_3, m_5). It can be seen that the bias is smaller in Model 1 but the variance is smaller in Model

2. However, if $\omega = 0.0$ - and the LRCCs are zero - then the MSE is smaller for Model 1 although the differences in all three statistics across models 1 and 2 are negligible in large samples. Whereas if $\omega > 0.0$ - and the LRCCs non-zero - then the MSE is smaller for Model 2 and this comparative advantage persists even in large samples. Now consider m_1: as anticipated above, Model 2 shows no gains over Model 1.

Taken together, the results show that both the comparative large sample properties of the system GMM and individual GMM estimators are sensitive to whether or not the LR-CCs between the two sets of moment are zero, and that the $LR(\hat{\theta}_T)$ and $H_T(\hat{\theta}_T)$ estimator can distinguish between these two states of the world. Although this is a relatively simple design, in more complicated nonlinear models, system GMM estimation may involve a considerable increase in computational burden over the individual GMM estimations and so it may be considered desirable to test *a priori* if there is a potential gain from system estimation. Our results suggest $LR(\hat{\theta}_T)$ and $H_T(\hat{\theta}_T)$ offer a convenient method for performing this type of inference.

5 Concluding Remarks

In this paper, we propose two statistics for testing whether a number of LRCCs are zero, and also derive the limiting distributions of these statistics under the null. We show that the hypothesis of asymptotic independence between two standardized sums can be expressed in terms of LRCCs and thus that our test statistics can be used to test this hypothesis. Interest in this type of hypothesis is illustrated via a number of examples. We evaluate the finite sample performance of our tests statistics in a number of settings via simulation studies which collectively suggest the limiting distribution under the null hypothesis is a reasonable approximation to behaviour in moderate and large sized samples.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: Wiley.

Andrews, D. W. K. (1991). Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *57*, 817–858.

Andrews, D. W. K. & Monahan, C. J. (1992). An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, *60*, 953–666.

Antoine, B., Bonnal, H., & Renault, E. (2007). On the effcient use of the informational content of estimating equations: Implied probabilities and euclidean empirical likelihood. *Journal of Econometrics*, *138*, 461–487.

Back, K. & Brown, D. P. (1993). Implied probabilities in gmm estimators. *Econometrica*, *61*, 971–975.

Bouhaddioui, C. & Roy, R. (2004). A generalized portmanteau test for independence of two infinite-order vector autoregressive series. *Journal of Time Series Analysis*, *27*, 505–544.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory* (republished by SIAM (2001) ed.). San Francisco: Holden-Day.

Chu, K.-W. E. (1990). On multiple eigenvalues of matrices depending on several parameters. *SIAM Journal on Numerical Analysis*, *27*, 1368–1385.

Constantine, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Annals of Mathematical Statistics*, *34*, 1270–1285.

El Himdi, K. & Roy, R. (1997). Tests for noncorrelation of two multivariate arma time series. *Canadian Journal of Statistics*, *25*, 233–256.

Hall, A. R. (2005). *Generalized Method of Moments*. New York: Oxford University Press.

Hall, A. R., Inoue, A., Jana, K., & Shin, C. (2007). Information in generalized method of moments estimation and entropy based moment selection. *Journal of Econometrics*, *138*, 488–512.

Hall, A. R., Rudebusch, G. D., & Wilcox, D. W. (2003). Judging instrument relevance in instrumental variables estimation. *International Economic Review*, *37*(2), 283–298.

Hannan, E. J. (1970). *Multiple Time Series.* New York: Wiley.

Haugh, L. D. (1976). Checking the independence of two covariance-stationary time series: a univariate residual cross correlation approach. *Journal of the American Statistical Association*, *71*, 376–385.

Hiriart-Urruty, J.-B. & Ye, D. (1995). Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numer. Math.*, *70*, 45–72.

Hong, Y. (1996). Testing for independence between two covariance stationary time series. *Biometrika*, *83*, 615–625.

Hsu, P. L. (1941). On the limiting distribution of the canonical correlations. *Biometrika*, *32*, 38–45.

Kato, T. (1966). *Perturbation Theory for Linear Operators.* New York: Springer-Verlag.

Kejriwal, M. & Perron, P. (2008). The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics*, *146*, 59–73.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Annals of Statistics*, *25*, 2084–2102.

Magnus, J. R. & Neudecker, H. (1991). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* New York, NY.

Park, J. Y. & Phillips, P. C. B. (1988). Statistical inference in regressions with integrated processes, part 1. *Econometric Theory*, *4*, 468–497.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics*, *28*, 329–348.

Phillips, P. C. B. & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, *53*, 473–495.

Politis, P. N. & Romano, J. P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics*, *20*, 1985–2007.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley.

Rosenblatt, M. (1984). Asymptotic normality, strong mixing and spectral density estimates. *Annals of Probability*, *12*, 1167–1180.

Smith, R. J. (2004). Gel criteria for moment condition models. *http://www.cemmap.ac.uk/wps/cwp0419.pdf.*

Appendix

This appendix is divided into **three** sections. Section A briefly describes some results from the linear algebra literature that are exploited in our analysis. Section B presents the proofs of the main results in the text and Section C contains the tables.

A Some properties of the LRCCs as an implicit function

The long run canonical correlations (LRCCs) between x_t and z_t are given by Definition 1*(i)* and can equivalently be considered as square root of the eigenvalues of $\Sigma_{xx}^{-1/2}\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\Sigma_{xx}^{-1/2}$ where, for any symmetric positive definite matrix A, $A^{1/2}$ is the unique symmetric positive definite matrix B satisfying $B^2 = A$.

Let $\mathbb{U}$ be the open subset of the $(p+q)\times(p+q)$ symmetric positive definite matrices, $\mathbb{A}$ be the subset of the the $p \times p$ symmetric positive semidefinite matrices.

For any $\Sigma \in \mathbb{U}$, we write $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ where the blocks are analogue to those of Σ_{vv}. Let us define the LRCCs function r by:

$$r : \mathbb{U} \to \mathbb{R}^p_+, \quad \forall \Sigma \in \mathbb{U},\ r(\Sigma) = (r_1(\Sigma), \cdots, r_p(\Sigma))',$$

where $r_1(\Sigma) \geq r_2(\Sigma) \geq \ldots \geq r_p(\Sigma)$ are the LRCCs between x_t and z_t derived from Σ. We consider the following decomposition of r:

$$r = s \circ l \circ h,$$

where

$$\mathbb{U} \xrightarrow{h} \mathbb{A} \xrightarrow{l} \mathbb{R}^p_+ \xrightarrow{s} \mathbb{R}^p_+,$$

$h(\Sigma) = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}$, $s(u)$ is the square root of u component-wise and, for any $A \in \mathbb{A}$, $l(A)$ is the vector of the p eigenvalues of A arranged from the largest to the smallest.

This eigenvalue function l has some interesting properties that one can get from the literature on matrix perturbation theory. (See Kato (1966), Chu (1990), Hiriart-Urruty & Ye (1995).) We will partially rely on those results to set up some useful properties of the canonical correlation function r. These properties are presented for completeness in Lemma A.2 below. Before doing so it is useful to recall that a function $g : \mathcal{O}_1 \to \mathcal{O}_2$ ($\mathcal{O}_1$ and $\mathcal{O}_2$, two normed spaces) is said to be locally Hölder continuous on $\mathcal{O}_1$ with exponent $\alpha > 0$ if for any $a \in \mathcal{O}_1$, there exists a neighborhood $\mathcal{V}_a$ of a and a constant $\delta_a > 0$ such that, for any $b, c \in \mathcal{V}_a$,

$$\|g(b) - g(c)\| \leq \delta_a \|b - c\|^{\alpha}.$$

Lemma A.2 *(i) $s \circ l \circ h$ is locally Hölder continuous of exponent $\alpha = 1/2$ on $\mathbb{U}$.*

(ii) Let $A \in \mathbb{A}$ and η a multiple eigenvalue of A of order m. Let $l_{i_1},\ l_{i_2}, \ldots,\ l_{i_m}$ be the m components of l satisfying $l_{i_1}(A) = l_{i_2}(A) = \cdots = l_{i_m}(A) = \eta$. Then the function

$$S \equiv l_{i_1} + l_{i_2} + \ldots + l_{i_m}$$

is indefinitely differentiable in a neighborhood of A.

(iii) Let $\Sigma_0 \in \mathbb{U}$ such that $h(\Sigma_0)$ has multiple eigenvalue as described by the conditions of (ii). Then, $S \circ h$ is indefinitely differentiable in a neighborhood of Σ_0.

Proof. (i) As a continuously differentiable function on $\mathbb{U}$, h is also locally Hölder continuous on $\mathbb{U}$ with exponent 1 (i.e. locally Lipschitz continuous). Also, from Hiriart-Urruty & Ye (1995)[Th. 4.1], each component of l is locally Lipschitz continuous as difference of two locally Lipschitz continuous functions. Therefore, l is also locally Lipschitz continuous and so is $l \circ h$.

Hence, for any $\Sigma \in \mathbb{U}$, there exists a neighborhood $\mathcal{V}$ of Σ and a constant $\gamma_1 > 0$ such that, for any $\Sigma_1, \Sigma_2 \in \mathcal{V}$,

$$\|l \circ h(\Sigma_1) - l \circ h(\Sigma_2)\| \leq \gamma_1 \|\Sigma_1 - \Sigma_2\|. \tag{A.1}$$

Now, we show that for any $a, b \in \mathbb{R}_+^p$,

$$\|s(a) - s(b)\| \leq \gamma_2 \|a - b\|^{1/2},$$

for some $\gamma_2 > 0$. Clearly, $\|s(a) - s(b)\|^2 = \sum_{i=1}^p \left(\sqrt{a_i} - \sqrt{b_i}\right)^2$. But, it is easy to check that for any $x, y \in \mathbb{R}_+$, $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|}$. Thus

$$\|s(a) - s(b)\|^2 \leq \sum_{i=1}^p |a_i - b_i| = p\left(\frac{1}{p}\sum_{i=1}^p |a_i - b_i|\right) \leq p\left(\frac{1}{p}\sum_{i=1}^p (a_i - b_i)^2\right)^{1/2} = \sqrt{p}\|a - b\|;$$

the last inequality follows from the Jensen's inequality. Hence $\|s(a) - s(b)\| \leq \gamma_2 \|a - b\|^{1/2}$, $\gamma_2 = p^{1/4}$. Using (A.1), we have

$$\|s(l \circ h(\Sigma_1)) - s(l \circ h(\Sigma_2))\| \leq \gamma_2 \|l \circ h(\Sigma_1) - l \circ h(\Sigma_2)\|^{1/2} \leq \gamma_2 \gamma_1^{1/2} \|\Sigma_1 - \Sigma_2\|^{1/2}$$

which establishes *(i)*.

(ii) This is a straightforward consequence of Hiriart-Urruty & Ye (1995)[Corollary 4.3]. One can also refer to Chu (1990)[p. 1375].

(iii) Obvious, thanks to *(ii)* and the fact that h is indefinitely differentiable on $\mathbb{U}$. $\diamond$

Remark 6: It is worthwhile to recall that the eigenvalue function l is not in general differentiable at matrices A_0 in $\mathbb{A}$ having multiple eigenvalues. (See Chu (1990)[p. 1375].) However, as stated by Lemma A.2*(ii)*, the sums of components of l returning the same eigenvalues at A_0 are smooth functions in a neighborhood of A_0. This observation will play an essential role in our approach to establish the main results of this paper.

B Proofs of the main results

Proof of Proposition 1.

Locally Hölder continuous functions are also continuous and the continuity of r is therefore guaranteed by Lemma A.2*(i)*. Since $\hat{\Sigma}_{vv} \xrightarrow{p} \Sigma_{vv}$, the Continuous Mapping Theorem implies that

$r(\hat{\Sigma}_{vv}) \xrightarrow{p} r(\Sigma_{vv})$. $\diamond$

Proof of Theorem 1.

Let $Z_s : s = 1, \ldots, m$ be $p + q$-vector valued random variables independently and identically distributed with $Z_1 \sim \mathcal{N}(0, f_{vv}(0))$.

Let $W = \sum_{s=1}^{m} Z_s Z_s'$. W is distributed as $W_{p+q}(m, f_{vv}(0))$, where W_{p+q} denotes the Wishart distribution. For m large, by the Central Limit Theorem,

$$\sqrt{m}\left(\frac{1}{m}W - f_{vv}(0)\right) \xrightarrow{p} \mathcal{N}(0, V_1), \tag{A.2}$$

where we make a similar abuse of notation as in Lemma A.3. The components of V_1 are

$$V_{1,kl,k'l'} = Cov(Z_{1k}Z_{1l}, Z_{1k'}Z_{1l'}) = \{f_{vv}(0)\}_{kk'}\{f_{vv}(0)\}_{ll'} + \{f_{vv}(0)\}_{kl'}\{f_{vv}(0)\}_{k'l},$$

$1 \leq k, l, k', l' \leq p + q$. Noting that $V_1 = V(0)$, we can deduce that

$$\sqrt{m}\left(\frac{1}{m}W - f_{vv}(0)\right) \quad \text{and} \quad \sqrt{\frac{\nu}{2}}\left(\hat{f}_{vv}(0) - f_{vv}(0)\right)$$

have the same asymptotic distribution. We set the correspondence

$$m = \frac{\nu}{2}, \quad \nu = \frac{2TB_T}{\int_{-\infty}^{+\infty} k^2(x)dx}.$$

Let $\hat{r}^{(b)} = r(W)$ be the LRCCs calculated from the infeasible estimator $\frac{1}{m}W$ of $f_{vv}(0)$. Since W is distributed as a Wishart, the distribution of $\hat{r}^{(b)}$ is given by Constantine (1963).

The distribution of $\hat{r}^{(b)}$ for large values of m is given by Hsu (1941), see also Anderson (2003)[p. 505]. Under the data configuration giving W, the likelihood ratio test statistic for exactly k null long run canonical correlations is given for large m by

$$LR_T^{(b)} = -m \sum_{j=1}^{k} \ln\left(1 - r_{p-k+j}^{(b)^2}\right).$$

(See Anderson (2003)[Sec. 12.4] and Hall, Rudebusch & Wilcox (2003).)

We recall that under the null hypothesis, $r_{p-k+j}^{(b)} = O_p(1/\sqrt{m})$, $j = 1, \ldots, k$. Since $\log(1 - x) = -x + O(x^2)$ in the neighborhood of 0, $LR_T^{(b)}$ is asymptotically equivalent to

$$H_T^{(b)} = m \sum_{j=1}^{k} r_{p-k+j}^{(b)^2}$$

and both are asymptotically distributed as $\chi^2_{k(q-p+k)}$ for large m.

Having set up this benchmark and turning back to the feasible statistics of the theorem, we remark that $H_T = \frac{\nu}{2} S \circ h(\hat{f}_{vv}(0))$ and $H_T^{(b)} = mS \circ h(\frac{1}{m}W)$ where S is the sum of the k smallest components of l. These are exactly the k components of the eigenvalue function l taking the value 0 at $h(f_{vv}(0))$ under the null.

From Lemma A.2*(iii)*, $S \circ h$ is indefinitely differentiable in a neighborhood of $f_{vv}(0)$. Since $\frac{1}{m}W$ and $\hat{f}_{vv}(0)$ converge in probability to $f_{vv}(0)$, by a second order Taylor expansion, we can write

$$\begin{aligned} S \circ h(\tfrac{1}{m}W) &= S \circ h(f_{vv}(0)) + \tfrac{\partial(S\circ h)}{\partial vech'(\Sigma)}(f_{vv}(0)) vech\left(\tfrac{1}{m}W - f_{vv}(0)\right) \\ &\quad + \tfrac{1}{2} vech'\left(\tfrac{1}{m}W - f_{vv}(0)\right) \tfrac{\partial^2(S\circ h)(f_{vv}(0))}{\partial vech(\Sigma)\partial vech'(\Sigma)} vech\left(\tfrac{1}{m}W - f_{vv}(0)\right) \\ &\quad + O_P(\|\tfrac{1}{m}W - f_{vv}(0)\|^3) \end{aligned} \quad \text{(A.3)}$$

and

$$\begin{aligned} S \circ h(\hat{f}_{vv}(0)) &= S \circ h(f_{vv}(0)) + \tfrac{\partial(S\circ h)}{\partial vech'(\Sigma)}(f_{vv}(0)) vech\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \\ &\quad + \tfrac{1}{2} vech'\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \tfrac{\partial^2(S\circ h)}{\partial vech(\Sigma)\partial vech'(\Sigma)}(f_{vv}(0)) vech\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \\ &\quad + O_P(\|\hat{f}_{vv}(0) - f_{vv}(0)\|^3), \end{aligned} \quad \text{(A.4)}$$

where $vech$ is the usual matrix operator that transforms a symmetric matrix into vector stacking its lower triangular elements.

Under the null hypothesis, $S \circ h(f_{vv}(0)) = 0$ as the sum of the square of the k canonical correlations which are equal to 0. By multiplying (A.3) and (A.4) by m and $\nu/2$, respectively, we obtain

$$\begin{aligned} H_T^{(b)} &= \sqrt{m} \tfrac{\partial(S\circ h)}{\partial vech'(\Sigma)}(f_{vv}(0)) \sqrt{m}\, vech\left(\tfrac{1}{m}W - f_{vv}(0)\right) \\ &\quad + \tfrac{1}{2}\sqrt{m}\, vech'\left(\tfrac{1}{m}W - f_{vv}(0)\right) \tfrac{\partial^2(S\circ h)(f_{vv}(0))}{\partial vech(\Sigma)\partial vech'(\Sigma)} \sqrt{m}\, vech\left(\tfrac{1}{m}W - f_{vv}(0)\right) \\ &\quad + O_P(m^{-1/2}) \end{aligned} \quad \text{(A.5)}$$

$$\begin{aligned} H_T &= \sqrt{\tfrac{\nu}{2}} \tfrac{\partial(S\circ h)}{\partial vech'(\Sigma)}(f_{vv}(0)) \sqrt{\tfrac{\nu}{2}}\, vech\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \\ &\quad + \tfrac{1}{2}\sqrt{\tfrac{\nu}{2}}\, vech'\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \tfrac{\partial^2(S\circ h)}{\partial vech(\Sigma)\partial vech'(\Sigma)}(f_{vv}(0)) \sqrt{\tfrac{\nu}{2}}\, vech\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \\ &\quad + O_P(\nu^{-1/2}). \end{aligned} \quad \text{(A.6)}$$

Since $\sqrt{m}\, vech\left(\frac{1}{m}W - f_{vv}(0)\right)$ and $\sqrt{\frac{\nu}{2}}\, vech\left(\hat{f}_{vv}(0) - f_{vv}(0)\right)$ have the same asymptotic distribution and $m = \nu/2$, we can deduce from (A.5) and (A.6) that $H_T^{(b)}$ and H_T have the same asymptotic distribution. This shows that

$$H_T \overset{a}{\sim} \chi^2_{k(q-p+k)}.$$

Turning to LR_T, we remark that, since $H_T = O_P(1)$, $r^2_{p-k+j} = O_P(\nu^{-1})$, $\forall j = 1, \ldots, k$. Hence, $\ln(1 - r^2_{p-k+j}) = -r^2_{p-k+j} + O_P(\nu^{-2})$, $\forall j = 1, \ldots, k$. Multiplying this by $\nu/2$ and summing over j, we have

$$LR_T = -\frac{\nu}{2}\sum_{j=1}^{k}\ln(1-r^2_{p-k+j}) = \frac{\nu}{2}\sum_{j=1}^{k} r^2_{p-k+j} + O_P(\nu^{-1}) = H_T + O_P(\nu^{-1}).$$

This implies that $LR_T = H_T + o_P(1)$ and we can deduce that LR_T and H_T have the same asymptotic distribution,

$$LR_T \overset{a}{\sim} \chi^2_{k(q-p+k)}. \qquad \diamond$$

Remark 7: A closer examination of (A.5) and (A.6) shows that the first term in each right hand side seems to explode while the left hand sides and the second terms of the right hand side are all asymptotically bounded in probability. Hence, the only way for (A.5) and (A.6) to hold is to have

$$\frac{\partial(S\circ h)}{\partial vech'(\Sigma)}(f_{vv}(0)) = 0 \tag{A.7}$$

under the null hypothesis. This is an additional information that one can extract from these expansions.

While a direct proof of (A.7) may be tedious, this equation is easy to verify if all of the long run canonical correlations between x_t and z_t are equal to 0 so that $\Sigma_{xz} = 0$. In this case, for any $j = 1,\ldots,p+q$, and $i = j,\ldots,p+q$,

$$\frac{\partial(S\circ h)}{\partial\Sigma_{i,j}}(f_{vv}(0)) = \frac{\partial S}{\partial vech'(A)}(h(f_{vv}(0)))\cdot\frac{\partial vech(h)}{\partial\Sigma_{i,j}}(f_{vv}(0)).$$

But,

$$\begin{aligned}\tfrac{\partial h}{\partial\Sigma_{i,j}}(f_{vv}(0)) = \; & \tfrac{\partial\Sigma_{11}^{-1/2}}{\partial\Sigma_{i,j}}\Big|_{f_{vv}(0)}\cdot\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\Sigma_{xx}^{-1/2} + \Sigma_{xx}^{-1/2}\,\tfrac{\partial\Sigma_{12}}{\partial\Sigma_{i,j}}\Big|_{f_{vv}(0)}\cdot\Sigma_{zz}^{-1}\Sigma_{zx}\Sigma_{xx}^{-1/2}\\ & +\Sigma_{xx}^{-1/2}\Sigma_{xz}\,\tfrac{\partial\Sigma_{22}^{-1}}{\partial\Sigma_{i,j}}\Big|_{f_{vv}(0)}\cdot\Sigma_{zx}\Sigma_{xx}^{-1/2} + \Sigma_{xx}^{-1/2}\Sigma_{xz}\Sigma_{zz}^{-1}\,\tfrac{\partial\Sigma_{21}}{\partial\Sigma_{i,j}}\Big|_{f_{vv}(0)}\cdot\Sigma_{xx}^{-1/2}\\ & +\Sigma_{xx}^{-1/2}\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}\,\tfrac{\partial\Sigma_{11}^{-1/2}}{\partial\Sigma_{i,j}}\Big|_{f_{vv}(0)} = 0 \quad \text{since } \Sigma_{xz} = 0.\end{aligned}$$

Therefore, $\frac{\partial(S\circ h)}{\partial vech'(\Sigma)}(f_{vv}(0)) = 0.$ $\diamond$

Proof of Theorem 2.

Let $\hat{f}_{vv}(0)$ and $\hat{f}_{\hat{v}\hat{v}}(0)$ be the estimators of the spectral density at frequency 0, $f_{vv}(0)$, calculated from $v_t(\theta_0)$ and $v_t(\hat{\theta}_T)$, respectively. From Andrews (1991)[Th. 1(b)],

$$\sqrt{TB_T}(\hat{f}_{vv}(0) - \hat{f}_{\hat{v}\hat{v}}(0)) \xrightarrow{p} 0.$$

As a result,

$$\sqrt{\frac{\nu}{2}}\left(\hat{f}_{vv}(0) - f_{vv}(0)\right) \quad \text{and} \quad \sqrt{\frac{\nu}{2}}\left(\hat{f}_{\hat{v}\hat{v}}(0) - f_{vv}(0)\right), \quad \nu = \frac{2TB_T}{\int_{-\infty}^{+\infty}k^2(x)dx}$$

have the same asymptotic distribution. The asymptotic distribution of $LR_T(\hat{\theta}_T)$ and $H_T(\hat{\theta}_T)$ are derived readily along the lines of the proof of Theorem 1. $\diamond$

Derivation of the variance V in Example 3.

The smoothed implied probabilities are the solutions of

$$\{\hat{\theta}, \hat{p}_i : i = 1, \ldots, Q\} = \arg\max_{\theta\in\Theta} \sup_{p_i : i=1,\ldots,Q} \left\{ \prod_{i=1}^{Q} p_i | p_i > 0, \sum_{i=1}^{Q} p_i = 1, \sum_{i=1}^{Q} p_i F_i(\theta) = 0 \right\},$$

where $\{f(x_t, \theta) : t = 1, \ldots, T\}$ is smoothed by some moving average resulting in $\{F_i(\theta) : i = 1, \ldots, Q\}$ and $\hat{\theta}$ is the smoothed maximum empirical likelihood (SEL) estimator of θ_0. We derive the asymptotic variance V of the efficient estimator

$$\hat{\mu} = \sum_{i=1}^{Q} \hat{p}_i G_i,$$

where G_i are the smoothed version of $\{g(x_t) : t = 1, \ldots, T\}$.

We, actually, show that

$$\sqrt{T}(\hat{\mu} - \mu_0) \xrightarrow{d} N(0, V).$$

As showed by Theorem 3.3 of Antoine, Bonnal & Renault (2007) for the cross-sectional *i.i.d* case, we can establish along the same lines that $\hat{\mu}$ is numerically equal to the corresponding SEL estimate of $(\theta', \mu')'$ from the augmented moment condition:

$$E \begin{pmatrix} f(x_t, \theta) \\ g(x_t) - \mu \end{pmatrix} = 0.$$

We maintain the conditions of Theorem 1 of Kitamura (1997) for this augmented model. From this same Theorem, we know that:

$$\sqrt{T} \left(\begin{pmatrix} \hat{\theta} \\ \hat{\mu} \end{pmatrix} - \begin{pmatrix} \theta_0 \\ \mu_0 \end{pmatrix} \right) \xrightarrow{d} N\left(0, (D'\Sigma^{-1}D)^{-1}\right),$$

where $D = \begin{pmatrix} \Gamma & 0 \\ 0 & -Id_r \end{pmatrix}$. From the usual matrix block-inverse formula, we have

$$\Sigma^{-1} = \begin{pmatrix} K_1^{-1} & -K_1^{-1}\Sigma_{fg}\Sigma_{gg}^{-1} \\ -K_2^{-1}\Sigma_{gf}\Sigma_{ff}^{-1} & K_2^{-1} \end{pmatrix}; \; K_1 = \Sigma_{ff} - \Sigma_{fg}\Sigma_{gg}^{-1}\Sigma_{gf}, \; K_2 = \Sigma_{gg} - \Sigma_{gf}\Sigma_{ff}^{-1}\Sigma_{fg}$$

and

$$D'\Sigma^{-1}D = \begin{pmatrix} \Gamma' K_1^{-1}\Gamma & \Gamma' K_1^{-1}\Sigma_{fg}\Sigma_{gg}^{-1} \\ K_2^{-1}\Sigma_{gf}\Sigma_{ff}^{-1}\Gamma & K_2^{-1} \end{pmatrix}.$$

The asymptotic variance of $\hat{\mu}$ is given by the south-east block of $(D'\Sigma^{-1}D)^{-1}$ which is, using again the block-inverse formula:

$$V = \left(K_2^{-1} - K_2^{-1}\Sigma_{gf}\Sigma_{ff}^{-1}\Gamma(\Gamma' K_1^{-1}\Gamma)^{-1}\Gamma' K_1^{-1}\Sigma_{fg}\Sigma_{gg}^{-1}\right)^{-1}.$$

Since $D'\Sigma^{-1}D$ is symmetric,

$$\Gamma' K_1^{-1}\Sigma_{fg}\Sigma_{gg}^{-1} = \Gamma'\Sigma_{ff}^{-1}\Sigma_{fg}K_2^{-1} \text{ and } V = \left(K_2^{-1} - K_2^{-1}\Sigma_{gf}\Sigma_{ff}^{-1}\Gamma(\Gamma' K_1^{-1}\Gamma)^{-1}\Gamma'\Sigma_{ff}^{-1}\Sigma_{fg}K_2^{-1}\right)^{-1}.$$

Next, we use the matrix algebra result stating that: If A, B and $C^{-1} + DA^{-1}B$ are non-singular square matrices, $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$.

Applying this formula with $A = K_2$, $B = \Sigma_{gf}\Sigma_{ff}^{-1}\Gamma$, $D = B'$ and $C = \left(\Gamma'\Sigma_{ff}^{-1}\Gamma\right)^{-1}$, it is straightforward to see that $V = \left(A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}\right)^{-1}$. Actually,

$$\begin{aligned} C^{-1} + DA^{-1}B &= (\Gamma'\Sigma_{ff}^{-1}\Gamma) + \Gamma'\Sigma_{ff}^{-1}\Sigma_{fg}K_2^{-1}\Sigma_{gf}\Sigma_{ff}^{-1}\Gamma &= (\Gamma'\Sigma_{ff}^{-1}\Gamma) + \Gamma' K_1^{-1}\Sigma_{fg}\Sigma_{gg}^{-1}\Sigma_{gf}\Sigma_{ff}^{-1}\Gamma \\ &= (\Gamma'\Sigma_{ff}^{-1}\Gamma) + \Gamma' K_1^{-1}(\Sigma_{ff} - K_1)\Sigma_{ff}^{-1}\Gamma &= \Gamma' K_1^{-1}\Gamma. \end{aligned}$$

The second equality uses again the equality $\Gamma'\Sigma_{ff}^{-1}\Sigma_{fg}K_2^{-1} = \Gamma' K_1^{-1}\Sigma_{fg}\Sigma_{gg}^{-1}$ and from the definition of K_1, $\Sigma_{fg}\Sigma_{gg}^{-1}\Sigma_{gf} = \Sigma_{ff} - K_1$ and the third equality follows. Thus

$$\begin{aligned} V &= A + BCD = K_2 + \Sigma_{gf}\Sigma_{ff}^{-1}\Gamma\left(\Gamma'\Sigma_{ff}^{-1}\Gamma\right)^{-1}\Gamma'\Sigma_{ff}^{-1}\Sigma_{fg} \\ &= \Sigma_{gg} - \Sigma_{gf}\Sigma_{ff}^{-1}\Sigma_{fg} + \Sigma_{gf}\Sigma_{ff}^{-1}\Gamma\left(\Gamma'\Sigma_{ff}^{-1}\Gamma\right)^{-1}\Gamma'\Sigma_{ff}^{-1}\Sigma_{fg} \\ &= \Sigma_{gg} - \Sigma_{gf}\Sigma_{ff}^{-1/2}\left(Id_r - \Sigma_{ff}^{-1/2}\Gamma\left(\Gamma'\Sigma_{ff}^{-1}\Gamma\right)^{-1}\Gamma'\Sigma_{ff}^{-1/2}\right)\Sigma_{ff}^{-1/2}\Sigma_{fg} \qquad \diamond \end{aligned}$$

C Tables

Table 1: (Experiment 1) Simulated rejection rates of LR_T and H_T with nominal size, $\alpha = 0.05$; Design 1: VAR(1)

	Bartlett HAC					Bartlett HAC with prewhitening/recoloring				
T	50	100	250	500	1000	50	100	250	500	1000
					Case 1					
H_{01}										
LR_T	0.030	0.019	0.010	0.011	0.008	0.007	0.004	0.003	0.004	0.004
H_T	0.024	0.016	0.008	0.010	0.008	0.005	0.004	0.003	0.003	0.004
H_{02}										
LR_T	0.260	0.207	0.149	0.129	0.108	0.086	0.069	0.057	0.052	0.050
H_T	0.174	0.156	0.123	0.115	0.101	0.044	0.043	0.044	0.045	0.045
					Case 2					
H_{01}										
LR_T	0.125	0.111	0.094	0.089	0.079	0.061	0.059	0.053	0.052	0.049
H_T	0.110	0.102	0.088	0.085	0.078	0.050	0.052	0.049	0.049	0.047
H_{02}										
LR_T	0.917	0.980	0.999	1.000	1.000	0.915	0.985	1.000	1.000	1.000
H_T	0.870	0.970	0.999	1.000	1.000	0.850	0.975	0.999	1.000	1.000
					Case 3					
H_{01}										
LR_T	0.133	0.119	0.094	0.085	0.081	0.066	0.063	0.055	0.051	0.049
H_T	0.116	0.111	0.089	0.082	0.079	0.055	0.058	0.051	0.049	0.048
H_{02}										
LR_T	0.977	0.997	1.000	1.000	1.000	0.982	0.999	1.000	1.000	1.000
H_T	0.957	0.996	1.000	1.000	1.000	0.959	0.997	1.000	1.000	1.000
					Case 4					
H_{01}										
LR_T	0.248	0.311	0.433	0.634	0.809	0.191	0.271	0.421	0.652	0.835
H_T	0.224	0.294	0.423	0.629	0.807	0.170	0.254	0.407	0.646	0.832
H_{02}										
LR_T	0.944	0.989	1.000	1.000	1.000	0.941	0.993	1.000	1.000	1.000
H_T	0.905	0.982	1.000	1.000	1.000	0.892	0.987	1.000	1.000	1.000

Table 2: (Experiment 1) Simulated rejection rates of LR_T and H_T with nominal size, $\alpha = 0.05$; Design 2: VMA(1)

	Bartlett HAC					Bartlett HAC with prewhitening/recoloring				
T	50	100	250	500	1000	50	100	250	500	1000
					Case 1					
H_{01}										
LR_T	0.013	0.008	0.005	0.006	0.005	0.005	0.003	0.002	0.002	0.002
H_T	0.009	0.007	0.004	0.006	0.005	0.003	0.002	0.002	0.002	0.002
H_{02}										
LR_T	0.152	0.120	0.090	0.078	0.070	0.072	0.047	0.036	0.032	0.033
H_T	0.085	0.084	0.070	0.068	0.064	0.034	0.027	0.026	0.026	0.029
					Case 2					
H_{01}										
LR_T	0.088	0.081	0.069	0.066	0.060	0.055	0.047	0.039	0.038	0.037
H_T	0.075	0.072	0.063	0.064	0.058	0.044	0.041	0.036	0.037	0.036
H_{02}										
LR_T	0.918	0.983	0.999	1.000	1.000	0.917	0.988	1.000	1.000	1.000
H_T	0.863	0.975	0.999	1.000	1.000	0.850	0.979	1.000	1.000	1.000
					Case 3					
H_{01}										
LR_T	0.089	0.086	0.072	0.064	0.061	0.059	0.050	0.041	0.039	0.038
H_T	0.077	0.079	0.067	0.062	0.059	0.049	0.044	0.038	0.037	0.037
H_{02}										
LR_T	0.980	0.999	1.000	1.000	1.000	0.985	0.999	1.000	1.000	1.000
H_T	0.961	0.997	1.000	1.000	1.000	0.962	0.999	1.000	1.000	1.000
					Case 4					
H_{01}										
LR_T	0.219	0.294	0.429	0.643	0.826	0.193	0.265	0.419	0.660	0.850
H_T	0.197	0.279	0.419	0.637	0.823	0.169	0.246	0.405	0.654	0.847
H_{02}										
LR_T	0.947	0.992	1.000	1.000	1.000	0.945	0.994	1.000	1.000	1.000
H_T	0.905	0.985	1.000	1.000	1.000	0.895	0.989	1.000	1.000	1.000

Table 3: (Experiment 1) Simulated rejection rates of LR_T and H_T with nominal size, $\alpha = 0.05$; Design 3: VARMA(1)

	Bartlett HAC					Bartlett HAC with prewhitening/recoloring				
T	50	100	250	500	1000	50	100	250	500	1000
					Case 1					
H_{01}										
LR_T	0.039	0.025	0.012	0.013	0.010	0.007	0.003	0.002	0.001	0.002
H_T	0.032	0.021	0.010	0.012	0.009	0.005	0.003	0.001	0.001	0.002
H_{02}										
LR_T	0.325	0.244	0.169	0.143	0.117	0.099	0.059	0.031	0.022	0.018
H_T	0.228	0.187	0.140	0.129	0.110	0.052	0.038	0.023	0.018	0.015
					Case 2					
H_{01}										
LR_T	0.144	0.122	0.102	0.092	0.084	0.066	0.052	0.035	0.032	0.027
H_T	0.128	0.114	0.096	0.091	0.082	0.055	0.046	0.031	0.030	0.027
H_{02}										
LR_T	0.920	0.978	0.999	1.000	1.000	0.914	0.987	1.000	1.000	1.000
H_T	0.874	0.968	0.999	1.000	1.000	0.850	0.976	1.000	1.000	1.000
					Case 3					
H_{01}										
LR_T	0.157	0.131	0.102	0.092	0.086	0.071	0.057	0.037	0.032	0.027
H_T	0.139	0.122	0.097	0.089	0.084	0.060	0.050	0.035	0.030	0.026
H_{02}										
LR_T	0.976	0.997	1.000	1.000	1.000	0.982	0.999	1.000	1.000	1.000
H_T	0.956	0.995	1.000	1.000	1.000	0.959	0.998	1.000	1.000	1.000
					Case 4					
H_{01}										
LR_T	0.259	0.317	0.434	0.632	0.808	0.199	0.270	0.415	0.664	0.863
H_T	0.237	0.302	0.424	0.626	0.805	0.176	0.250	0.403	0.657	0.860
H_{02}										
LR_T	0.944	0.988	1.000	1.000	1.000	0.942	0.994	1.000	1.000	1.000
H_T	0.906	0.980	0.999	1.000	1.000	0.891	0.988	1.000	1.000	1.000

Table 4: (Experiment 2) Simulated rejection rates of LR_T and H_T with nominal size, $\alpha = 0.05$

	LR_T					H_T				
T	50	100	250	500	1000	50	100	250	500	1000
						Case 1:				
BT	0.059	0.057	0.053	0.051	0.052	0.050	0.051	0.049	0.049	0.050
PZ	0.054	0.057	0.054	0.050	0.051	0.043	0.048	0.050	0.049	0.050
						Case 2: $\theta = 0.5$				
BT	0.056	0.050	0.045	0.047	0.046	0.046	0.043	0.042	0.044	0.045
PZ	0.053	0.052	0.048	0.049	0.048	0.043	0.043	0.043	0.046	0.046
						Case 3: $\gamma = -0.4$				
BT	0.458	0.638	0.859	0.981	0.999	0.428	0.618	0.851	0.979	0.999
PZ	0.431	0.558	0.833	0.962	0.998	0.397	0.531	0.825	0.960	0.998
						Case 3: $\gamma = 0.4$				
BT	0.458	0.627	0.854	0.980	0.998	0.425	0.608	0.847	0.979	0.998
PZ	0.428	0.548	0.826	0.961	0.998	0.394	0.522	0.817	0.959	0.997
						Case 3: $\gamma = -0.8$				
BT	0.912	0.988	1.000	1.000	1.000	0.896	0.986	1.000	1.000	1.000
PZ	0.896	0.970	1.000	1.000	1.000	0.878	0.965	1.000	1.000	1.000
						Case 3: $\gamma = 0.8$				
BT	0.910	0.989	1.000	1.000	1.000	0.897	0.988	1.000	1.000	1.000
PZ	0.897	0.971	1.000	1.000	1.000	0.878	0.965	0.999	1.000	1.000

Notes: BT, PZ stand, respectively, for tests constructed using the Bartlett and Parzen kernel. In Case 1, $\gamma = \theta = 0$; in Case 2, $\gamma = 0$; in Case 3 $\theta = 0$.

Table 5: (Experiment 2) Simulated coverage probability of 95% confidence interval for β_0

	Inference via Eq. (5)*: Bias corr. int.*					*Inference via Eq.* (13)*: Naive int.*				
T	50	100	250	500	1000	50	100	250	500	1000
					Case 1:					
BT:	0.883	0.914	0.928	0.938	0.948	0.912	0.936	0.940	0.945	0.949
PZ:	0.883	0.910	0.926	0.937	0.947	0.912	0.933	0.940	0.944	0.949
					Case 2: $\theta = 0.5$					
BT:	0.899	0.928	0.938	0.947	0.950	0.900	0.916	0.918	0.917	0.916
PZ:	0.894	0.918	0.932	0.943	0.945	0.895	0.904	0.910	0.910	0.910
					Case 3: $\gamma = -0.4$					
BT:	0.885	0.904	0.915	0.913	0.915	0.884	0.903	0.907	0.909	0.915
PZ:	0.885	0.899	0.914	0.911	0.914	0.884	0.900	0.906	0.908	0.915
					Case 3: $\gamma = 0.4$					
BT:	0.885	0.903	0.912	0.915	0.915	0.883	0.899	0.910	0.914	0.917
PZ:	0.883	0.900	0.911	0.915	0.914	0.881	0.896	0.910	0.913	0.916
					Case 3: $\gamma = -0.8$					
BT:	0.890	0.884	0.877	0.870	0.866	0.813	0.837	0.851	0.848	0.856
PZ:	0.888	0.884	0.876	0.871	0.866	0.811	0.833	0.850	0.848	0.855
					Case 3: $\gamma = 0.8$					
BT:	0.882	0.886	0.875	0.869	0.865	0.813	0.838	0.845	0.850	0.850
PZ:	0.882	0.886	0.874	0.870	0.864	0.813	0.834	0.843	0.849	0.850

Notes: BT, PZ refer, respectively, to the use of the Bartlett and Parzen kernel for the estimation of Σ_{uu} and Δ_{wu}. In Case 1, $\gamma = \theta = 0$; in Case 2, $\gamma = 0$; in Case 3 $\theta = 0$.

Table 6: (Experiment 3) Simulated rejection rates of the likelihood ratio and Hannan's tests for zero LRCCs between $Y_t = (y_t - m_1, y_t^3 - m_3, y_t^5 - m_5)'$ and $Z_t = (z_t - \mu, (z_t - \mu)^3, (z_t - \mu)^5)'$. The long run variance is estimated using the Bartlett kernel with prewhitening/recolouring. Univariate AR(2) are considered for the prewhitening step. The tests are performed at the nominal level of $\alpha = 5\%$. Under the null hypothesis, LR_T and H_T are asymptotically distributed as χ^2_9. All the LRCCs are null for $\omega = 0$ and different from zero otherwise.

T	50	100	250	500	1000
			$\omega = 0$		
LR_T	0.104	0.066	0.045	0.040	0.036
H_T	0.044	0.038	0.034	0.032	0.032
			$\omega = 0.6$		
LR_T	0.662	0.831	0.976	1.000	1.000
H_T	0.426	0.705	0.956	0.999	1.000
			$\omega = 0.8$		
LR_T	0.969	0.998	1.000	1.000	1.000
H_T	0.876	0.993	1.000	1.000	1.000

Table 7: (Experiment 3) Simulated bias, RMSE and average asymptotic variance of the estimates of m_1, m_2 and m_3 from Model 1 and Model 2

	$\omega = 0.0$						$\omega = 0.6$						$\omega = 0.8$					
	Model 1			Model 2			Model 1			Model 2			Model 1			Model 2		
T	m_1	m_3	m_5	m_1	m_3	m_5	m_1	m_3	m_5	m_1	m_3	m_5	m_1	m_3	m_5	m_1	m_3	m_5
50																		
Bias	-0.01	-0.04	0.21	-0.01	-0.05	0.17	0.00	-0.03	-0.47	0.00	-0.02	0.10	0.00	-0.01	-0.33	0.00	0.01	0.30
$\sqrt{T}\times$Rmse	4.75	43.46	760.86	4.82	44.12	779.25	4.75	43.73	819.39	4.87	43.64	799.31	4.72	43.39	830.25	4.89	43.00	786.53
ave. Avar$\times 10^{-3}$	0.01	0.55	233.56	0.00	0.49	209.20	0.01	0.57	267.84	0.00	0.43	195.96	0.01	0.57	279.92	0.00	0.38	176.75
100																		
Bias	0.00	0.00	-0.55	0.00	-0.01	-0.75	0.00	-0.01	-0.30	0.00	-0.02	-0.62	0.00	-0.01	-0.32	0.00	-0.02	-0.36
$\sqrt{T}\times$Rmse	4.83	44.50	796.13	4.90	45.26	811.10	4.86	45.24	842.51	4.98	44.92	807.58	4.87	45.18	829.71	5.02	43.86	741.81
ave. Avar$\times 10^{-3}$	0.01	0.75	300.81	0.01	0.69	280.69	0.01	0.76	339.14	0.01	0.63	272.90	0.01	0.76	333.93	0.01	0.58	235.04
250																		
Bias	0.00	0.02	0.48	0.00	0.03	0.58	0.00	0.00	0.13	0.00	0.01	0.19	0.00	0.00	0.05	0.00	0.00	0.07
$\sqrt{T}\times$Rmse	4.96	45.71	832.47	5.01	46.13	836.87	4.94	45.67	818.34	5.03	45.31	780.97	4.94	45.66	819.73	5.05	44.48	733.18
ave. Avar$\times 10^{-3}$	0.01	1.03	416.00	0.01	0.99	400.87	0.01	1.03	400.37	0.01	0.92	343.95	0.01	1.03	403.72	0.01	0.87	301.72
500																		
Bias	0.00	0.01	0.28	0.00	0.01	0.29	0.00	0.04	0.78	0.00	0.04	0.79	0.00	0.04	0.70	0.00	0.04	0.70
$\sqrt{T}\times$Rmse	4.96	45.79	825.25	4.98	45.90	826.38	4.97	45.97	834.42	5.01	45.50	804.02	4.98	46.06	828.78	5.04	44.94	759.13
ave. Avar$\times 10^{-3}$	0.01	1.15	444.90	0.01	1.12	435.24	0.01	1.15	451.01	0.01	1.08	405.63	0.01	1.15	443.60	0.01	1.03	354.05
1000																		
Bias	0.00	0.03	0.39	0.00	0.03	0.39	0.00	0.02	0.26	0.00	0.02	0.24	0.00	0.01	0.27	0.00	0.01	0.17
$\sqrt{T}\times$Rmse	4.93	45.91	839.86	4.94	45.90	839.20	4.99	46.06	834.38	5.01	45.73	813.45	5.02	46.25	831.15	5.04	45.16	768.82
ave. Avar$\times 10^{-3}$	0.01	1.32	498.40	0.01	1.30	491.14	0.01	1.32	500.48	0.01	1.26	462.69	0.01	1.32	493.94	0.01	1.21	409.94

Supplementary Appendix - not for publication

In this supplementary appendix we present background and derivations of certain results used in the paper. This material is not for publication but is available from the authors upon request.

Proof of Lemma 1 From Rosenblatt (1984)[Corollary 3], we have the following result.

Lemma A.3 *If Assumptions 1(4) and 2 hold then:*[13]

$$\sqrt{\frac{\nu}{2}}\left(\hat{f}_{vv}(0) - E(\hat{f}_{vv}(0))\right) \xrightarrow{d} \mathcal{N}(0, V(0)). \tag{A.8}$$

The central limit result given by Lemma A.3 characterizes the distribution of $\hat{f}_{vv}(0)$ around its mean. This formulation is different from the usual ones which rather characterize the estimation error and therefore are more suitable for testing for restrictions on the parameter of interest. The form of the central limit result in (A.8) is induced by the asymptotic order of magnitude of the bias of $\hat{f}_{vv}(0)$. Except for some appropriate choices of bandwidth B_T, the square of $\hat{f}_{vv}(0)$'s bias has the same asymptotic order of magnitude as the variance of $\hat{f}_{vv}(0)$. In this case, only the deviations of the spectral density estimate about its mean can be expressed by the central limit theorem.

The bias of the spectral density estimator in (8) is derived by Parzen (1957). At frequency 0, this bias involves the smoothness of the spectral density at 0 as well as the local behaviour of the kernel function. Let $r > 0$ be the largest real number such that

$$k^{(r)} = \lim_{u \to 0} \frac{1 - k(u)}{|u|^r}$$

exist and is nonzero. r is known as the *characteristic exponent* of the kernel $k(\cdot)$. If $k^{(r)}$ exists for every positive number r, the characteristic exponent is set to ∞. The generalized rth spectral derivative $f^{(r)}(\lambda)$ is defined by

$$f_{vv}^{(r)}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\lambda} |h|^r \Gamma_{vv}(h),$$

if this series converges. If $\|f_{vv}^{(r)}(0)\| < \infty$, the asymptotic order of magnitude of the bias is

$$E\left(\hat{f}_{vv}(0)\right) - f_{vv}(0) = O(B_T^r).$$

In this case, by choosing $B_T = o\left(T^{-\frac{1}{1+2r}}\right)$, we have

$$\sqrt{TB_T}\left(E\left(\hat{f}_{vv}(0)\right) - f_{vv}(0)\right) = o(1)$$

which guarantees the result stated in Lemma 1.

[13]Our notations here are abusive since $\hat{f}_{vv}(0)$ and $f_{vv}(0)$ are both matrices; in this context, this notation refers to the *vech* of the corresponding matrix.

Table 8: Characteristic exponent (r) and equivalent degree of freedom (ν) of some commonly used spectral density estimators

Estimate	Lag window generator $k(x)$	Characteristic exponent r	Degree of freedom ν
The truncated kernel window	$\begin{cases} 1, & \lvert x\rvert \le 1, \\ 0, & \lvert x\rvert > 1. \end{cases}$	∞	$B_T T$
The Bartlett window	$\begin{cases} 1-\lvert x\rvert, & \lvert x\rvert \le 1, \\ 0, & \lvert x\rvert > 1. \end{cases}$	1	$3B_T T$
The Parzen window	$\begin{cases} 1-6x^2+6\lvert x\rvert^3, & \lvert x\rvert \le 0.5, \\ 2(1-\lvert x\rvert)^3, & 0.5 \le \lvert x\rvert \le 1 \\ 0, & \lvert x\rvert > 1. \end{cases}$	2	$\frac{560}{151} B_T T$
The general Tukey window $(0 < a \le 1/4)$	$\begin{cases} 1-2a+2a\cos(\pi x), & \lvert x\rvert \le 1, \\ 0, & \lvert x\rvert > 1. \end{cases}$	2	$\frac{B_T T}{1-4a+6a^2}$
The Bartlett-Priestley window	$\frac{3}{(\pi x)^2}\left(\frac{\sin \pi x}{\pi x} - \cos \pi x\right)$	2	$\frac{5}{3} B_T T$
The Daniell window	$\frac{\sin(\pi x)}{\pi x}$	2	$2B_T T$
The finite Fourier transform	$\frac{\sin\left(\frac{1}{2}\pi x\right)}{\frac{1}{2}\pi x}$	2	$B_T T$