ABSTRACT

JANA, KALIDAS. Canonical Correlations and Instrument Selection in Econometrics. (Under the direction of Alastair R. Hall.)

This dissertation relates to three recent methods of instrument selection in econometrics, namely, the Canonical Correlations Information Criterion (CCIC), the Relevant Moments Selection Criterion (RMSC) and the approximate Mean Square Error Criterion (MSE). Usual canonical correlations measure the degree of association between two random vectors and provide the basis for construction of the CCIC. A new kind of canonical correlations called Long Run Canonical Correlations (LRCC) has recently emerged in econometrics and provides the basis for construction of the RMSC. Although the concept of LRCC has emerged in the literature, methods of their estimation and inference have not been developed. Developing these methods constitutes the first chapter of the dissertation. In addition, this chapter illustrates the usefulness of LRCC beyond their usefulness in relevant moments selection for GMM models in dynamic nonlinear settings. In particular, it demonstrates how LRCC can be used to develop econometric tests that play a role in (i) structural stability testing, and (ii) exogeneity testing of regressors in time series models where the regressors are nonstationary.

Although the properties of each of the above three methods of instrument selection have been explored by their proponents, there have been no comparative studies of these methods to date. The second chapter of this dissertation fills that gap.

The final and third chapter extends the statistical theory of the CCIC by considering the case where the number of instruments tends to infinity at an appropriate rate as the sample size tends to infinity. The importance of this extension stems from the fact that this can lead to a further gain in efficiency of the estimator by systematically capturing all relevant instruments from the growing candidate set.

CANONICAL CORRELATIONS AND INSTRUMENT SELECTION IN ECONOMETRICS

by

KALIDAS JANA

A dissertation submitted to the graduate faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

ECONOMICS

Raleigh, NC

2005

APPROVED BY:

Peter Bloomfield	David A. Dickey
Alastair R. Hall, Chair of Advisory Committee	Denis Pelletier

Dedication

To my late father, late mother and late sister.

Biography

I received M.A. in economics from the Southern Methodist University in the Spring of 1997 and transferred to the North Carolina State University in the Fall of 1997 to pursue doctoral program in economics. I defended my doctoral dissertation on December 17, 2004 and am currently a visiting instructor of economics at Trinity University, San Antonio, TX.

Acknowledgements

The journey of my academic pursuit has been long and has followed a path of a continuous random walk, sometimes with a drift and sometimes without a drift. The graduate work at NC State has given me the opportunity to meet with distinguished academicians who have helped me build a Brownian Bridge that has made it possible to eventually reach my goal.

I extend my sincerest gratitude to Professor Wally Thurman for offering me admission into the graduate program of economics and for providing me with every help that I needed to continue my pursuit.

I owe my deepest gratitude to Professor Alastair R. Hall for granting me the privilege of his invaluable guidance, support, and encouragement. I am eternally grateful to him for his kindness, patience, commitment and dedication to being instrumental in turning what was only my dream into a reality.

My deepest gratitude extends equally to Professors Peter Bloomfield, David A. Dickey, Atsushi Inoue and Denis Pelletier for their open door policy, ever welcoming persona, pleasant and kind manner, and unflinching guidance and help. I am eternally grateful to them for playing unique and vital roles in the metamorphosis of my dream.

Contents

List of Tables

Chapter 1

Introduction

Ordinary Least Squares (OLS) estimation yields inconsistent estimators of parameters of linear regression models where regressors are correlated with errors of the model. In such models a popular method for obtaining consistent estimators is application of the Instrumental Variables (IV) method. To implement the IV method in practice, the researcher must choose a set of instruments from a candidate set. Such choice has been informal at best and fails to satisfy many of the now recognized desirable properties of instrument selection. Recently, a number of instrument selection criteria has been proposed in the literature to help guide a practitioner in this regard.

This dissertation relates to three among these proposed criteria, namely, the Canonical Correlations Information Criterion (CCIC) of Hall and Peixe (2003), the Relevant Moments Selection Criterion (RMSC) of Hall and Inoue (2003) and the approximate Mean Square Error Criterion (MSE) of Donald and Newey (2001).

As the name suggests, at the heart of the Canonical Correlations Information Criterion (CCIC) are canonical correlations. In the section below, we present a formal description of these correlations.

1.1 Canonical Correlations

The standardized covariance, $\frac{Cov(x_t, z_t)}{\sqrt{Var(x_t)}\sqrt{(Var(z_t)}}$, known as the simple correlation coefficient proposed by Galton (1888) measures the degree of linear dependence between two random scalars x_t and z_t. The multiple correlation coefficient is a generalization of this idea and measures the degree of linear association between a random scalar x_t and a random vector $\mathbf{z_t}$. Canonical correlations proposed by Hotelling (1935, 1936) generalize this notion even further to measure the degree of linear dependence between two random vectors.

More specifically, the multiple correlation coefficient chooses a weight vector α such that the correlation between the scalar x_t and the linear combination $\alpha'\mathbf{z_t}$, $\frac{Cov(x_t, \alpha'\mathbf{z_t})}{\sqrt{Var(x_t)}\sqrt{(Var(\alpha'\mathbf{z_t})}}$, is maximum. Canonical correlations generalize this idea one step further by choosing weight vectors α and β such that the correlations between the linear combinations $\alpha'\mathbf{x_t}$ and $\beta'\mathbf{z_t}$, $\frac{Cov(\alpha'\mathbf{x_t}, \beta'\mathbf{z_t})}{\sqrt{Var(\alpha'\mathbf{x_t})}\sqrt{(Var(\beta'\mathbf{z_t})}}$ are maximum.

To elaborate the notions of canonical correlations a little further, let

$$v_t = \begin{pmatrix} \underset{p\times 1}{x_t} \\ \\ \underset{q\times 1}{z_t} \end{pmatrix} \sim N(0, \Sigma_v), \quad q \geq p, \quad \Sigma_v = \begin{pmatrix} Ex_tx_t' & Ex_tz_t' \\ \\ Ez_tx_t' & Ez_tz_t' \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}.$$

Then define a pair of variables: $\zeta_1 = \alpha_1' x_t, \omega_1 = \beta_1' z_t$ such that $\rho_1 = \underset{\alpha_1,\beta_1}{Max}\ Cov(\zeta_1,\ \omega_1)$ subject to $Var(\zeta_1) = Var(\omega_1) = 1$, that is, $\rho_1 = \underset{\alpha_1,\beta_1}{Max}\ Cov(\alpha_1' x_t,\ \beta_1' z_t)$ subject to $Var(\alpha_1' x_t) = Var(\beta_1' z_t) = 1$.

Now find another pair of variables: $\zeta_2 = \alpha_2' x_t, \omega_2 = \beta_2' z_t$, whose elements are uncorrelated with the corresponding elements of the first pair, such that $\rho_2 = \underset{\alpha_2,\beta_2}{Max}\ Cov(\zeta_2,\ \omega_2)$ subject to $Cov(\zeta_1, \zeta_2) = Cov(\omega_1, \omega_2) = Cov(\zeta_1, \omega_2) = Cov(\zeta_2, \omega_1) = 0$, and $Var(\zeta_2) = Var(\omega_2) = 1$, that is, $\rho_2 = \underset{\alpha_2,\beta_2}{Max}\ Cov(\alpha_2' x_t,\ \beta_2' z_t)$ subject to $Cov(\alpha_1' x_t, \alpha_2' x_t)$ $= Cov(\beta_1' z_t, \beta_2' z_t) = Cov(\alpha_1' x_t, \beta_2' z_t) = Cov(\alpha_2' x_t, \beta_1' z_t) = 0$, and $Var(\alpha_2' x_t) = Var(\beta_2' z_t) = 1$, and so on. Thus, it is possible to decompose random vectors x_t and z_t into mutually independent linear combinations of their elements displaying pairwise maximal correlations such that the quantities $\rho_i = \frac{Cov(\alpha_i' x_t,\ \beta_i' z_t)}{\sqrt{Var(\alpha_i' x_t)}\sqrt{Var(\beta_i' z_t)}}$, $i = 1, ..., p$, are maximized subject to the restrictions $Cov(\alpha_i' x_t, \alpha_j' x_t) = Cov(\beta_i' z_t, \beta_j' z_t) = Cov(\alpha_i' x_t, \beta_j' z_t) = 0 \; \forall \; i \neq j$, and the normalizing conditions $Var(\alpha_i' x_t) = Var(\beta_i' z_t) = 1, \; i = 1, ..., p$. The pairs $(\alpha_i' x_t, \beta_i' z_t), \; i = 1, ..., p$ are called canonical variables and the correlations $\rho_i, \; i = 1, ..., p$, are called canonical correlations. It turns out that the canonical correlations can be obtained as square roots of the ordered solutions of the generalized eigenvalue problem, that is, as square roots of the ordered eigenvalues of the determinantal equation $|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \lambda\Sigma_{xx}| = 0$, that is, $\rho_i = \sqrt{\lambda_i}, \;$ for $\lambda_1 \geq \lambda_2 \geq, ..., \geq \lambda_p$. Canonical correlations are widely used in multivariate statistics [Anderson (1984), ch.12; Dhrymes (1970), Ch. 2]. One of their principal applications is in testing independence of random vectors.

Early use of canonical correlations in econometrics traces back to Sargan (1958). He demonstrates that the asymptotic variance of the instrumental variables (IV) estimators can be written as a function of the population canonical correlations between regressors and instruments.

Recently, in the literature of Generalized Method of Moments (GMM) estimation (Hansen, 1982), canonical correlations have emerged to provide a natural metric for the purpose of moment selection (Hall and Inoue, 2003). The latter has introduced the concept of long run canonical correlations (LRCC).

To understand how canonical correlations arise in moment selection, we must first briefly summarize the recent literature on moment selection, to which we now turn.

1.2 Canonical Correlations and Moment Selection

In many applications of interest where GMM estimation is appropriate, the underlying economic/statistical model implies a candidate set of population moment conditions. In deciding which elements of this set to choose for estimation, it has been a standard practice to use the overidentifying restrictions tests. However, such practice suffers from repeated testing problem, namely, the type I error of a given test in a sequence of tests will differ from the significance level of the same test performed in isolation. This renders inference suspect.

For inference about the parameters of the model based on "conventional" asymptotic theory to hold, it is desirable for selected population moment conditions to satisfy four conditions which Hall and Peixe (2003) refer to as (i) the identification condition: they be satisfied at only one value in the parameter space; (ii) the orthogonality condition: this value be the "true" parameter value, θ_0 say, implying that the selected population moment conditions represent valid information; (iii) the efficiency condition: they minimize the asymptotic variance of the estimator; and (iv) the non-redundancy condition: none of the selected population moment conditions is redundant in the sense that the asymptotic variance of the estimator increases if any element of the selected vector is excluded. All these conditions impact directly on the asymptotic distribution theory of the estimator, $\hat{\theta}_T$ say, where T is the sample size, or on the adequacy of this theory as an approximation to the finite sample behavior of the estimator.

In practice, it is impossible to verify *a priori* which elements of the candidate set satisfy the four conditions enumerated above for any given data set. Instead, it is an empirical matter. Hall and Peixe (2003) note that once the selection process becomes data dependent, a fifth condition becomes important, a condition that the selection process must not contaminate the asymptotic distribution theory of the estimator. So they introduce an *inference condition* which requires that the asymptotic distribution of the estimator be the same as if the selected moment conditions had been picked *a priori*. As already noted above, the standard practice of applying overidentifying restrictions tests to choose from a candidate set of moment conditions fails to satisfy this inference condition.

To remedy this problem, Andrews (1999) proposes an information criterion approach that does not require repeated tests. He modifies the overidentifying restrictions test by adding to the overidentifying restrictions test a bonus term that is a function of the number of overidentifying restrictions. He provides a set of conditions under which his information criterion satisfies the inference condition. But while his procedure satisfies the inference condition, it suffers from a weakness. In terms of the other four conditions described above, it can be recognized that Andrews' (1999) method selects moment conditions on the basis of the orthogonality condition because the latter is the null hypothesis of the overidentifying restrictions test. But while this method weeds out any invalid moment conditions in the candidate set, it makes no distinction between redundant and non-redundant moment conditions. This can have a serious consequence. As demonstrated by Hall and Peixe (2003), the use of this method can lead to the inclusion of redundant moment conditions which in turn can cause asymptotic distribution theory to provide a poor approximation to the finite sample behavior of the estimator.

To rectify the above problem, Hall and Peixe (2003) propose a method of moments selection based on the combination of the efficiency and non-redundancy conditions. The authors refer to this combination as the *relevance condition.* They focus exclusively on the particular class of GMM models in which the population moment condition takes the form $E[z_t u_t(\theta_0)] = 0$ where $u_t(.)$ is a scalar, possibly nonlinear function of a set of dynamic random variables, z_t is a vector of instruments and the asymptotic variance of the GMM estimator is given by $\sigma_0^2\{E[(\partial u_t(\theta_0)/\partial\theta)z_t']E[z_t z_t']^{-1}E[z_t(\partial u_t(\theta_0)/\partial\theta)']\}^{-1}$. Because in most cases of interest, $u_t(\theta_0)$ is implied by the underlying economic model, the problem of moment selection reduces to one of choosing a vector of instruments z_t from a candidate set, $\mathcal{Z}$, say.

Developing a method for selecting instruments based on relevance requires a suitable metric for relevance. Hall and Peixe (2003) show that the population canonical correlations between $\partial u_t(\theta_0)/\partial\theta$ and z_t provide such a metric. This leads them to propose a canonical correlations information criterion (CCIC) which is the sum of two parts: (i) a function of the sample canonical correlations between $\partial u_t(\tilde{\theta}_T)/\partial\theta$ and z_t, where $\tilde{\theta}_T$ is a preliminary estimator, and (ii) a penalty term which is a function of the number of overidentifying restrictions. The selected instrument vector is then the choice which minimizes this criterion. They provide regularity conditions under which this procedure satisfies the inference condition stated above.

Just as Andrews' (1999) method focuses on orthogonality and ignores relevance, Hall and Peixe's (2003) method focuses on relevance and ignores orthogonality. Because both orthogonality and relevance are desirable properties of the selected instruments, intuition suggests that a combination of the two methods should have the strength of both without the weaknesses of either. Hall and Peixe (2003) verify this to be the case. While Hall and Peixe's results indicate that the sequence is inconsequential in terms of asymptotic properties, their simulation results suggest that it is preferable to select instruments first based on relevance and then based on orthogonality.

The relevance of the CCIC is limited to a particular class of GMM models as described above. Hall and Inoue (2003) generalize the CCIC for the relevant instruments selection in this special class of GMM models to the Relevant Moments Selection Criterion (RMSC) for the entire class of GMM models in nonlinear dynamic setting. From this work a concept of a new kind of canonical correlations has emerged. These canonical correlations have been called long run canonical correlations (LRCC). Hall and Inoue (2003) show that the asymptotic variance of the GMM estimator can be written in terms of the population LRCC between population moment conditions used in estimation and the unknown true score vector associated with the data. They exploit this result for the purpose of moment selection. Hall and Inoue (2003) show that the log of the determinant of the asymptotic variance matrix can be decomposed into two parts of which one depends on the LRCC and the other equals the log of the determinant of the information matrix. This decomposition of the asymptotic variance provides the basis for efficiency comparisons based on LRCC without actually calculating the individual canonical correlations.

1.3 Contributions of this Dissertation

Although the concept of LRCC has originated in Hall and Inoue (2003), methods of their estimation and inference have not been developed. Developing these methods constitutes the first chapter of the dissertation. In addition, this chapter shows that the usefulness of LRCC extends beyond moment selection. In particular, it demonstrates how LRCC can be used to develop econometric tests that play a role in (i) structural stability testing, and (ii) exogeneity testing of regressors in time series models where the regressors are nonstationary.

Hall and Peixe (2003), Hall and Inoue (2003), and Donald and Newey (2001) explore the properties of their proposed methods. However, to date, there have been no comparative studies of these methods. The second chapter of this dissertation fills that gap. To this end, it (i) establishes a relation between contemporaneous and long run canonical correlations in a linear simultaneous equations model, (ii) shows an analytical connection among the CCIC, the RMSC and the MSE in the context of a simple linear IV model, and (iii) assesses the relative performance of these three methods via a simulation study that investigates the finite sample behavior of the post selection estimator of this simple linear IV model by comparing median bias of the post selection estimator and coverage probability of 90% confidence interval under the three criteria.

The final and third chapter extends the statistical theory of the CCIC by considering the case in which the candidate set of instruments increases with the sample size. The importance of this extension stems from the fact that it can lead to a further gain in efficiency of the estimator by systematically capturing all relevant instruments from the growing candidate set.

Chapter 2

Long Run Canonical Correlations: Estimation, Inference and Usefulness in Econometric Analysis of Time Series

2.1 Introduction

In this chapter, we first present Hall and Inoue's (2003) definition of LRCC. Then we formally derive LRCC and present a lemma that is at the heart of Hall and Inoue's (2003) definition of redundancy of moment conditions. Next, we discuss methods of estimating LRCC. To this end, we first establish a link between the concept of LRCC and that of canonical coherence, which has previously been developed in the frequency domain analysis of time series. Finally, we close this chapter by illustrating the usefulness of LRCC in econometric analysis of time series, beyond their usefulness in moment selection.

2.2 Long Run Canonical Correlations

The short run, that is, contemporaneous canonical correlations between two random vectors x_t and z_t are correlations between certain linear combinations of x_t and z_t which are chosen to satisfy certain orthogonality and normalization constraints.

Hall and Inoue (2003) extend the concept of short run canonical correlations to capture the association between the standardized sums $X_T = T^{-1/2}\sum_{t=1}^{T} x_t$ and $Z_T = T^{-1/2}\sum_{t=1}^{T} z_t$. They refer to the resulting statistics as the "long run" canonical correlations between x_t and z_t. We present below their definition of LRCC.

Definition 2.2.1 *Let x_t and z_t be $p \times 1$ and $q \times 1$, respectively, where $q \geq p$. Suppose that $T^{-1/2}\sum_{t=1}^{T} v_t \xrightarrow{D} N(0, \Sigma_v)$ where $v_t = (x_t^{'}, z_t^{'})'$ and $\Sigma_v = lim_{T\to\infty} Var[T^{-1/2}\sum_{t=1}^{T} v_t]$ is a finite positive definite matrix. Partition Σ_v as follows:*

$$\Sigma_v = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{bmatrix}$$

using the obvious notation. The population long run canonical correlations between x_t and z_t are denoted by $\{\rho_i; \text{ı} = 1, 2, \ldots, p\}$, 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 $|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\Sigma_{xx}| = 0$; (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 $i = 1, 2, \ldots, p$. [1]

It can be recognized that the only difference between the short run canonical correlations and their long run counterpart lies in the form of the variance-covariance matrices in the determinantal equation. Specifically, the short run canonical correlations are calculated using the contemporaneous variance-covariance matrices while the long run canonical

[1] Recall that the linear combinations are chosen so as to normalize the variances to one, that is, $\alpha_i^{'}\Sigma_{xx}\alpha_i = \beta_i^{'}\Sigma_{zz}\beta_i = 1$.

correlations are calculated using the long run variance-covariance matrices.

Definition of long run canonical correlations is predicated on the existence of long run variance-covariance matrices. Following Andrews (1991), we present a lemma that guarantees the existence of such matrices. However, because the lemma uses the concept of α-mixing sequence, we define an α-mixing sequence first.

Definition 2.2.2 *Suppose $\{v_t\}$ is a sequence of vectors defined on a common probability space $(\Omega, \mathcal{F}, P)$. Denote for $j \geq 0$,*

$$\alpha(j) = sup_t sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+j}^{\infty}} |P(A \cap B) - P(A)P(B)|$$

where $\mathcal{F}_{-\infty}^t$ is the minimal σ-field generated by $v_t, v_{t-1}, \ldots$ and $\mathcal{F}_{t+j}^{\infty}$ is the minimal σ-field generated by $v_{t+j}, v_{t+j+1}, \ldots$. Then $\{v_t\}$ is called an α-mixing sequence with mixing coefficient $\alpha(j)$ if $lim_{j \to \infty} \alpha(j) = 0$.

Thus, in essence, α-mixing is a notion of asymptotic independence. The mixing coefficient $\alpha(j)$ measures the memory of the process. To be more precise, it measures how much dependence exists between events separated by at least j time periods. With this preamble on α-mixing, we are now ready for our desired lemma.

Lemma 2.2.1 *Suppose $\{v_t\}$ is a mean zero, fourth order stationary α-mixing sequence of random vectors. If $sup_{t \geq 1} E \parallel v_t \parallel^{4\nu} < \infty$ and $\sum_{j=1}^{\infty} j^2 \alpha(j)^{(\nu-1)/\nu} < \infty$ for some $\nu > 1$, then $\sum_{j=0}^{\infty} sup_{t \geq 1} E \parallel v_t v_{t+j}' \parallel < \infty$ and thus the long run variance-covariance matrix of $\{v_t\}$ exists and is uniformly bounded.*

The condition on the mixing numbers in Lemma 2.2.1 is satisfied if they are of size $-3\nu/(\nu-1)$, i.e., if $\alpha(j) = O(j^{-\epsilon-3\nu)/(\nu-1)})$ for some $\epsilon > 0$[2].

In the next section, we formally derive long run canonical correlations.

Derivation of Long Run Canonical Correlations:

The derivation of the long run canonical correlations is similar to that of the contemporaneous canonical correlations [Rao (1973), p. 583]. The only difference is that here one

[2] $\{v_t\}$ is an α-mixing sequence of size $-c_0$ if $\alpha(j) = O(j^{-c})$ for some $c > c_0$.

works with the vectors of standardized sums, X_T and Z_T, instead of the original vectors x_t and z_t. Thus, the long run canonical correlations and the long run canonical variates are obtained as solutions to the following maximization problem:

$$\underset{\alpha,\,\beta}{Max}\ \frac{Cov(\alpha' X_T, \beta' Z_T)}{\sqrt{Var(\alpha' X_T)}\sqrt{Var(\beta' Z_T)}} = \underset{\alpha,\,\beta}{Max} \frac{\alpha' \Sigma_{xz}\beta}{\sqrt{\alpha'\Sigma_{xx}\alpha}\sqrt{\beta'\Sigma_{zz}\beta}}$$

In words, the objective is to choose p and q dimensional vectors α and β, respectively, such that the correlation between the linear combinations $\alpha' X_T$ and $\beta' Z_T$ is maximum. The maximand clearly is homogeneous of degree zero in α and β. Thus if α_0 and β_0 are a solution to the above problem, then so are $c_1\alpha_0$ and $c_2\beta_0$ for arbitrary c_1 and c_2. To eliminate this scale indeterminacy and achieve a unique solution, one imposes the normalizing conditions:

$$Var(\alpha' X_T) = \alpha'\Sigma_{xx}\alpha = 1,$$

$$Var(\beta' Z_T) = \beta'\Sigma_{zz}\beta = 1.$$

Then the problem is to maximize $\alpha'\Sigma_{xz}\beta$ subject to $\alpha'\Sigma_{xx}\alpha = 1$ and $\beta'\Sigma_{zz}\beta = 1$. Hence the Lagrangean is:

$$\mathcal{L} = \alpha'\Sigma_{xz}\beta - \frac{\lambda_1}{2}(\alpha'\Sigma_{xx}\alpha - 1) - \frac{\lambda_2}{2}(\beta'\Sigma_{zz}\beta - 1)$$

where λ_1 and λ_2 are Lagrangean multipliers. Differentiating the Lagrangean with respect to α and β and setting them equal to zero yields

$$\Sigma_{xz}\beta - \lambda_1\Sigma_{xx}\alpha = 0, \tag{2.2.1}$$

$$\Sigma_{zx}\alpha - \lambda_2\Sigma_{zz}\beta = 0. \tag{2.2.2}$$

Premultiplying (2.2.1) by α' yields $\alpha'\Sigma_{xz}\beta = \lambda_1$ and premultiplying (2.2.2) by β' yields $\beta'\Sigma_{zx}\alpha = \lambda_2$. Again, $\beta'\Sigma_{zx}\alpha$ is a scalar and hence equal to its transpose. Thus $\beta'\Sigma_{zx}\alpha = \alpha'\Sigma_{xz}\beta = \lambda_2$. Therefore, $\lambda_1 = \lambda_2 = \rho$, say. Premultiplying (2.2.1) by $\Sigma_{zx}\Sigma_{xx}^{-1}$ and by adding it to to ρ times (2.2.2), gives

$$\Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz}\beta - \rho\Sigma_{zx}\alpha + \rho\Sigma_{zx}\alpha - \rho^2\Sigma_{zz}\beta = 0$$

or,

$$(\Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz} - \rho^2\Sigma_{zz})\beta = 0. \tag{2.2.3}$$

Thus ρ^2 is an eigenvalue and β is the corresponding eigenvector obtained from the determinantal equation

$$|\Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz} - \rho^2\Sigma_{zz}| = 0. \tag{2.2.4}$$

Let $\rho_1^2, \ldots, \rho_q^2$ be the eigenvalues and $\beta_1, \ldots, \beta_q$ the corresponding eigenvectors. Further, let $B = (\beta_1, \ldots, \beta_q)$. Then by the theorem of simultaneous diagonalization of square matrices,

$$\begin{aligned} B'\Sigma_{zz}B = I \quad &\text{or,} \quad \Sigma_{zz} = B'^{-1}B^{-1}; \\ B'\Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz}B = R_2 \quad &\text{or,} \quad \Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz} = B'^{-1}R_2B^{-1}, \end{aligned} \tag{2.2.5}$$

where R_2 is the diagonal matrix of eigenvalues $\rho_1^2, \ldots, \rho_q^2$. Similarly, we obtain the determinantal equation

$$|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\Sigma_{xx}| = 0 \tag{2.2.6}$$

with eigenvalues $\rho_1^2, \ldots, \rho_p^2$ and corresponding eigenvectors $\alpha_1, \ldots, \alpha_p$. The non-zero eigenvalues of (2.2.4) and (2.2.6) are the same. The multiplicity of zero roots is however different in the two cases. If A and R_1 correspond to B and R_2 in (2.2.5), then

$$\begin{aligned} A'\Sigma_{xx}A = I \quad &\text{or,} \quad \Sigma_{xx} = A'^{-1}A^{-1}; \\ A'\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx}A = R_1 \quad &\text{or,} \quad \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} = A'^{-1}R_1A^{-1}. \end{aligned} \tag{2.2.7}$$

The non-zero roots $\rho_1 \geq \rho_2 \geq, \ldots$ are the long run canonical correlations and the linear combinations

$$\alpha_1'X_T, \ldots, \alpha_p'X_T \quad \text{and} \quad \beta_1'Z_T, \ldots, \beta_q'Z_T$$

are the long run canonical variates. ■

Next, collecting above results we present a lemma that is at the heart of Hall and Inoue's (2003) construction of RMSC.

Lemma 2.2.2 *Let α_i be the i^{th} column of A and β_i be the i^{th} column of B, where A and*

B are, respectively, matrices of generalized eigenvectors corresponding to the determinantal equations (2.2.6) and (2.2.4) above. Then the following identities hold:

$$\begin{aligned} \Sigma_{xx} &= A'^{-1} A^{-1}, \\ \Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} &= A'^{-1} R^2 A^{-1} \end{aligned}$$

where $R = diag(\rho_1, \rho_2, \ldots, \rho_p)$.

Proof of Lemma 2.2.2 We begin by noting that the matrices A and B reduce the dispersion matrix of the transformed variables $(A'X_T, B'Z_T)$ to the simpler form:

$$\begin{pmatrix} I_p & R_{p\times q} \\ R'_{p\times q} & I_q \end{pmatrix}$$

where I_p and I_q are unit matrices of order p and q, respectively, and $R_{p\times q}$ is a $p \times q$ matrix with first $k = rank(\Sigma_{xz})$ diagonal elements as $\rho_1, \ldots, \rho_k$ and the rest of the elements zero. That the above simplification holds can be seen as follows. We know

$$Var\begin{pmatrix} A'X_T \\ B'Z_T \end{pmatrix} = \begin{pmatrix} A'\Sigma_{xx}A & A'\Sigma_{xz}B \\ B'\Sigma_{xz}A & B'\Sigma_{zz}B \end{pmatrix}.$$

Now the terms on the main diagonal are, respectively, $A'\Sigma_{xx}A = I_p$ from (2.2.7) and $B'\Sigma_{zz}B = I_q$ from (2.2.5). To determine the off-diagonal terms, consider the i-th equation of (2.2.1):

$$\Sigma_{xz}\beta_i = \rho_i \Sigma_{xx}\alpha_i.$$

Premultiplying by α_i', we have

$$\alpha_i'\Sigma_{xz}\beta_i = \rho_i \alpha_i'\Sigma_{xx}\alpha_i = \rho_i.$$

Premultiplying by α_j', we have

$$\alpha_j'\Sigma_{xz}\beta_i = \rho_i \alpha_j'\Sigma_{xx}\alpha_i = 0.$$

Therefore,

$$
\begin{aligned}
\underset{(p\times p)}{A}{}'\underset{(p\times q)}{\Sigma_{xz}}\underset{(q\times q)}{B} &= (\alpha_1, \alpha_2, \ldots, \alpha_p)'\,\Sigma_{xz}\,(\beta_1, \beta_2, \ldots, \beta_q) \\
&= \begin{pmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_p' \end{pmatrix} \Sigma_{xz}\,(\beta_1, \beta_2, \ldots, \beta_q) \\
&= \begin{pmatrix}
\alpha_1'\Sigma_{xz}\beta_1 & \alpha_1'\Sigma_{xz}\beta_2 & \alpha_1'\Sigma_{xz}\beta_3 & \cdots & \alpha_1'\Sigma_{xz}\beta_k & \alpha_1'\Sigma_{xz}\beta_{k+1} & \cdots & \alpha_1'\Sigma_{xz}\beta_q \\
\alpha_2'\Sigma_{xz}\beta_1 & \alpha_2'\Sigma_{xz}\beta_2 & \alpha_2'\Sigma_{xz}\beta_3 & \cdots & \alpha_2'\Sigma_{xz}\beta_k & \alpha_2'\Sigma_{xz}\beta_{k+1} & \cdots & \alpha_2'\Sigma_{xz}\beta_q \\
\alpha_3'\Sigma_{xz}\beta_1 & \alpha_3'\Sigma_{xz}\beta_2 & \alpha_3'\Sigma_{xz}\beta_3 & \cdots & \alpha_3'\Sigma_{xz}\beta_k & \alpha_3'\Sigma_{xz}\beta_{k+1} & \cdots & \alpha_3'\Sigma_{xz}\beta_q \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\alpha_k'\Sigma_{xz}\beta_1 & \alpha_k'\Sigma_{xz}\beta_2 & \alpha_k'\Sigma_{xz}\beta_3 & \cdots & \alpha_k'\Sigma_{xz}\beta_k & \alpha_k'\Sigma_{xz}\beta_{k+1} & \cdots & \alpha_k'\Sigma_{xz}\beta_q \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\alpha_p'\Sigma_{xz}\beta_1 & \alpha_p'\Sigma_{xz}\beta_2 & \alpha_p'\Sigma_{xz}\beta_3 & \cdots & \alpha_p'\Sigma_{xz}\beta_k & \alpha_p'\Sigma_{xz}\beta_{k+1} & \cdots & \alpha_p'\Sigma_{xz}\beta_q
\end{pmatrix} \\
&= \begin{pmatrix}
\rho_1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \rho_2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & 0 & \rho_3 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \rho_k & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix} \\
&= \underset{(p\times q)}{R} \qquad (2.2.8)
\end{aligned}
$$

Hence,

$$
Var\begin{pmatrix} A'X_T \\ B'Z_T \end{pmatrix} = \begin{pmatrix} A'\Sigma_{xx}A & A'\Sigma_{xz}B \\ B'\Sigma_{xz}A & B'\Sigma_{zz}B \end{pmatrix} = \begin{pmatrix} I_p & R_{p\times q} \\ R_{q\times p} & I_q \end{pmatrix}.
$$

From $A'\Sigma_{xx}A = I$ and $B'\Sigma_{zz}B = I$, it follows, respectively, that $\Sigma_{xx} = A'^{-1}A^{-1}$ and

$\Sigma_{zz} = B'^{-1}B^{-1}$. Again from $A'\Sigma_{xz}B = R$, it follows that $\Sigma_{xz} = (A')^{-1}R(B)^{-1}$. Therefore, we have

$$\begin{aligned}
\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} &= (A')^{-1}\ R\ (B^{-1})BB'(B^{-1\prime}R'A^{-1}) \\
&= A'^{-1}\ R\ (B^{-1}B)(B^{-1}B)'R'A^{-1}) \\
&= A'^{-1}RR'A^{-1} = A'^{-1}R^2A^{-1}. \quad \blacksquare
\end{aligned}$$

2.3 Estimation of Long Run Canonical Correlations

Because the long run canonical correlations are solutions to a generalized eigenvalue problem where the determinantal equation involves long run variance-covariance matrices, one obvious first step in estimating the LRCC is to obtain consistent estimators $\hat{\Sigma}_{xx}, \hat{\Sigma}_{zz}$ and $\hat{\Sigma}_{xz}$ of the long run variance-covariance matrices Σ_{xx}, Σ_{zz} and Σ_{xz}, then to solve the generalized eigenvalue problem by replacing the population variance-covariance matrices by their estimated values. Or, equivalently, as part (iii) of Definition 1 suggests, the LRCC can be estimated as $\hat{\rho}_i = \hat{\alpha}_i'\hat{\Sigma}_{xz}\hat{\beta}_i$, 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 satisfying $(\hat{\Sigma}_{xz}\hat{\Sigma}_{zz}^{-1}\hat{\Sigma}_{zx} - \rho^2\hat{\Sigma}_{xx})\hat{\alpha}_i = 0$ and $(\hat{\Sigma}_{zx}\hat{\Sigma}_{xx}^{-1}\hat{\Sigma}_{xz} - \rho^2\hat{\Sigma}_{zz})\hat{\beta}_i = 0$ for $i = 1, 2, \ldots, p$, and $\hat{\Sigma}_{xx}, \hat{\Sigma}_{zz}$ and $\hat{\Sigma}_{xz}$ are consistent estimators, respectively, of Σ_{xx}, Σ_{zz} and Σ_{xz}.

Which ever of the above two alternative methods one adopts, the first step involves estimation of the long run variance-covariance matrices $\Sigma_{ij}, (i, j = x, z)$ which are given by the sum of variance-covariance matrices at lags s ranging from $-\infty$ to $+\infty$:

$$\Sigma_{ij} = \sum_{s=-\infty}^{\infty} \Sigma_{ij}(s).$$

However, in practice, sample size T is finite. So it is instructive to estimate the sum by truncating it at a finite lag $s = T - 1$ since with T observations one can estimate at most $T - 1$ autocovariances. Thus one could obtain the estimated long run variance-covariance matrices as:

$$\hat{\Sigma}_{ij} = \sum_{s=-(T-1)}^{(T-1)} \hat{\Sigma}_{ij}(s),$$

where

$$\hat{\Sigma}_{ij}(s) = T^{-1} \sum_{t=s+1}^{T} v_t v_{t-s}^{'}$$

are the autocovariance matrices at lag s. Such estimators, however, have two undesirable features. One is that if true autocovariances at arbitrarily large lags are non-zero, in finite samples, such estimators need not be consistent. The reason being that the larger is the autocovariance lag s, the smaller is the number of observations available to estimate them. The second undesirable feature is that such estimators can fail, in finite samples, to be positive semidefinite.

In time series models, where the data exhibit heteroscedasticity and autocorrelation of unknown form, one popular class [3] that overcomes the above mentioned problems and delivers consistent estimators of the long run variance-covariance matrices is known as the Heteroscedasticity and Autocorrelation Consistent Covariance (HACC) estimator. The estimators belonging to this class are called kernel HACC estimators. The kernels, by assigning appropriately declining weights to autocovariances at distant lags, yield consistent and positive semidefinite estimators. These estimators take the form

$$\hat{\Sigma}_{ij} = \sum_{s=-(T-1)}^{(T-1)} k(\frac{s}{M_T})\ \hat{\Sigma}_{ij}(s).$$

where $k(.)$ is a "kernel" or "lag window generator" that belongs to the set $\mathcal{K}$ given by

$$\begin{aligned}\mathcal{K} =& \{k(.) : \mathbb{R} \mapsto [-1,1] \mid k(0) = 1,\ k(x) = k(-x)\ \forall x \in \mathbb{R},\\ & \int_{-\infty}^{\infty} |k(x)|\, dx < \infty,\ k(.) \text{ is continuous at 0 and at all}\\ & \text{but a finite number of other points}\}.\end{aligned}$$

The conditions $k(0) = 1$ and $k(.)$ is continuous at 0 ensure that the variance-covariance matrices estimated at lag zero receive weight equal to one and at lags near zero weights

[3](see Andrews, 1991)

close to one. Finally, M_T is a "window" or "bandwidth parameter" that depends on T and stretches or contracts the window, and hence is also known as a "scale parameter". If $k(x) = 0$ for $|x| > 1$ and $k(x) \neq 0$ for some $|x|$ arbitrarily close to 1, then M_T is also referred to as the "lag truncation parameter", because lags of order $s > M_T$ are assigned weight zero.

Two kernels, among others, that are widely used econometrics and are members of the above class $\mathcal{K}$, are the following:

$$\text{Bartlett:} \qquad k_{BT}(x) = \begin{cases} 1 - |x| & \text{for } |x| \leq 1, \\ 0 & \text{for } |x| > 1. \end{cases}$$

$$\text{Parzen:} \qquad k_{PR}(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| \leq 1/2, \\ 2(1 - |x|)^3 & \text{for } 1/2 \leq |x| \leq 1, \\ 0 & \text{for } |x| > 1. \end{cases}$$

The use of the Bartlett kernel was introduced by Newey and West (1987) and that of the Parzen kernel was introduced by by Gallant (1987, p. 533). Consistency of $\hat{\Sigma}_{ij}$ under these kernels requires the bandwidths go to infinity as a power of the sample size:

$$M_T \to \infty \text{ as } T \to \infty, \text{ and } M_T = o(T^{1/2}).$$

In the next section, we show a connection between LRCC and canonical coherences that have previously been developed in the frequency domain analysis of time series.

2.4 Link between LRCC and Canonical Coherences

Canonical coherences in the frequency domain are defined similarly as the canonical correlations in the time domain. The only difference is that they are the solutions of the determinantal equation where the variance-covariance matrices are replaced by their corresponding spectral- and cross-spectral density matrices. [4]

Proposition 2.4.1 *In addition to Assumption 1, let x_t and z_t be covaraince stationary time series with covariance function that satisfies $\sum_{-\infty}^{\infty} |h|\gamma(h)| = L < \infty$. Then canonical correlations between X_T and Z_T, that is, LRCC between x_t and z_t are equal to canonical coherences between x_t and z_t at frequency zero.*

Proof of Proposition 2.4.1:

We note that $\Sigma_v = \lim_{T\to\infty} Var\binom{X_T}{Z_T} = \lim_{T\to\infty} \begin{pmatrix} EX_TX_T' & EX_TZ_T' \\ EZ_TX_T' & EZ_TZ_T' \end{pmatrix}$.

Now,

$$EX_TX_T' = E \begin{pmatrix} X_{1T}^2 & X_{1T}X_{2T} & \cdots & X_{1T}X_{pT} \\ X_{2T}X_{1T} & X_{2T}^2 & \cdots & X_{2T}X_{pT} \\ \vdots & \vdots & \vdots & \vdots \\ X_{pT}X_{1T} & X_{pT}X_{2T} & \cdots & X_{pT}^2 \end{pmatrix}.$$

Again,

$EX_{1T}^2 = E\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{t1}\frac{1}{\sqrt{T}}\sum_{j=1}^{T} x_{j1} = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} Ex_{t1}x_{j1}$

$= \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\gamma_{x_1x_1}(t-j) = \sum_{h=-(T-1)}^{T-1}\frac{T-|h|}{T}\gamma_{x_1x_1}(h)$.

Then by Kronecker's Lemma (Lemma 3.1.4, Fuller, 1996, p. 126),

$\lim_{T\to\infty} EX_{1T}^2 = \sum_{h=-\infty}^{\infty}\gamma_{x_1x_1}(h)$.

Now following (4.1.2) of Fuller (1996, p. 144), the spectral density of x_1 at frequency ω is:

$$f_{x_1x_1}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-i\omega h}\gamma_{x_1x_1}(h).$$

This implies

$$f_{x_1x_1}(0) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_{x_1x_1}(h),$$

which, in turn, implies

$$\sum_{h=-\infty}^{\infty}\gamma_{x_1x_1}(h) = 2\pi f_{x_1x_1}(0).$$

[4] See Hannan (1970) [Theorem 14, p. 299]

Thus, $\lim_{T\to\infty} EX_{1T}^2 = 2\pi f_{x_1x_1}(0)$, where $f_{x_1x_1}(0)$ denotes spectral density of x_1 at frequency zero.

Similarly,

$EX_{1T}X_{2T} = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} Ex_{t1}x_{j2}$

$= \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}\gamma_{x_1x_2}(t-j) = \sum_{h=-(T-1)}^{T-1}\frac{T-|h|}{T}\gamma_{x_1x_2}(h)$

and by Kronecker's Lemma,

$\lim_{T\to\infty} EX_{1T}X_{2T} = \sum_{h=-\infty}^{\infty}\gamma_{x_1x_2}(h)$. Then from cross-spectral density at frequency ω:

$$f_{x_1x_2}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty} e^{-i\omega h}\gamma_{x_1x_2}(h)$$

we have

$$f_{x_1x_2}(0) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_{x_1x_2}(h)$$

from which it follows that

$$\sum_{h=-\infty}^{\infty}\gamma_{x_1x_2}(h) = 2\pi f_{x_1x_2}(0).$$

Thus, $\lim_{T\to\infty} EX_{1T}X_{2T} = 2\pi f_{x_1x_2}(0)$, where $f_{x_1x_2}(0)$ denotes cross-spectral density of x_1 and x_2 at frequency zero.

Proceeding in a like manner, we have

$$\lim_{T\to\infty} EX_TX_T^{'} = \Sigma_{xx} = 2\pi\begin{pmatrix} f_{x_1x_1}(0) & f_{x_1x_2}(0) & \cdots & f_{x_1x_p}(0) \\ f_{x_2x_1}(0) & f_{x_2x_2}(0) & \cdots & f_{x_2x_p}(0) \\ \vdots & \vdots & \vdots & \vdots \\ f_{x_px_1}(0) & f_{x_px_2}(0) & \cdots & f_{x_px_p}(0) \end{pmatrix} = 2\pi f_{xx}(0).$$

Similarly,

$$\lim_{T\to\infty} EX_TZ_T^{'} = \Sigma_{xz} = 2\pi f_{xz}(0);\ \lim_{T\to\infty} EZ_TX_T^{'} = \Sigma_{zx} = 2\pi f_{zx}(0);\ \textit{and}$$

$$\lim_{T\to\infty} EZ_TZ_T^{'} = \Sigma_{zz} = 2\pi f_{zz}(0).$$

Thus,

$$\Sigma_v = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix} = 2\pi\begin{pmatrix} f_{xx}(0) & f_{xz}(0) \\ f_{zx}(0) & f_{zz}(0) \end{pmatrix} = 2\pi f_v(0).$$

Therefore, the long run variance-covariance matrices are equal to a constant 2π times the corresponding spectral and cross-spectral density matrices at frequency zero. Hence the solutions of the determinantal equation $|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\ \Sigma_{xx}| = 0$ are identical to those of the determinantal equation $|f_{xz}(0)f_{zz}^{-1}(0)f_{zx}(0) - \rho^2 f_{xx}(0)| = 0$, that is, long run canonical correlations are equal to canonical coherences at frequency zero. ■

2.5 Usefulness of LRCC in Econometric Analysis of Time Series

Hall and Inoue (2003) show how LRCC provide a metric for the purpose of relevant moments selection in Generalized Method of Moments (GMM) models (Hansen, 1982) in dynamic nonlinear settings. Usefulness of this metric derives from the fact that moments selected on the basis of this metric yields the most efficient GMM estimator. The objective of this section is to illustrate further usefulness of LRCC in econometric analysis of time series.

2.5.1 Structural Stability Testing

One intriguing area of econometric testing is structural stability testing. Andrews and Fair (1988), Ghysels and Hall (1990a, b), Hall and Sen (1999), among others, develop such tests. These tests assume that subsamples on either side of the break are asymptotically independent. Here we show how this can be expressed as a hypothesis about the LRCC between the subsamples.

As an illustrative example, suppose that v_t is a $k \times 1$ vector of random variables and that we are interested in testing the null hypothesis

$$H_0 : E[v_t] = \mu_t = \mu, \qquad \forall t$$

against the alternative hypothesis

$$\begin{aligned} H_1 : E[v_t] &= \mu_1, \quad t \leq [0.5T] \\ &= \mu_2, \quad t > [0.5T] \end{aligned}$$

where $[0.5T]$ denotes the integer part of $0.5T$. Thus we are interested in testing constancy of the mean throughout the sample against a break midway through the sample. This is an example of structural stability testing where the break point is *fixed*.

Following Andrews and Fair (1988) this can be tested using the Wald statisitc

$$W_T = T(\hat{\mu}_{1,T} - \hat{\mu}_{2,T})' \left(\frac{\hat{\Omega}_{1,T}}{[0.5T]/T} + \frac{\hat{\Omega}_{2,T}}{1 - [0.5T]/T} \right)^{-1} (\hat{\mu}_{1,T} - \hat{\mu}_{2,T}) \tag{2.5.1}$$

where

$$\begin{aligned} \hat{\mu}_{1,T} &= \frac{1}{[0.5T]} \sum_{t=1}^{[0.5T]} v_t, \\ \hat{\mu}_{2,T} &= \frac{1}{[0.5T]} \sum_{t=[0.5T]+1}^{T} v_t, \\ \hat{\Omega}_{1,T} &= \hat{\Gamma}_{1,0} + \sum_{i=1}^{T-1} k(\frac{i}{T})\{\hat{\Gamma}_{1,i} + \hat{\Gamma}'_{1,i}\}, \\ \hat{\Omega}_{2,T} &= \hat{\Gamma}_{2,0} + \sum_{i=1}^{T-1} k(\frac{i}{T})\{\hat{\Gamma}_{2,i} + \hat{\Gamma}'_{2,i}\}, \\ \hat{\Gamma}_{1,i} &= \frac{1}{[0.5T]} \sum_{t=i+1}^{[0.5T]} (v_t - \hat{\mu}_{1,T})(v_{t-i} - \hat{\mu}_{1,T})', \\ \hat{\Gamma}_{2,i} &= \frac{1}{[0.5T]} \sum_{t=[0.5T]++i+1}^{T} (v_t - \hat{\mu}_{2,T})(v_{t-i} - \hat{\mu}_{2,T})', \end{aligned}$$

and $k(.)$ is a kernel giving desired weights to the autocovariance matrices $\hat{\Gamma}_{1,i}, \hat{\Gamma}'_{1,i}, \hat{\Gamma}_{2,i}$ and $\hat{\Gamma}'_{2,i}$.

Andrews and Fair (1988) show that under $H_0 : E[v_t] = \mu_t = \mu, \forall t$ the Wald statistic $W_T \xrightarrow{D} \chi^2_k$. Their conditions imply that

$$\begin{bmatrix} [0.5T]^{-1/2} \sum_{t=1}^{[0.5T]} v_t \\ [0.5T]^{-1/2} \sum_{t=[0.5T]+1}^{T} v_t \end{bmatrix} \equiv \begin{bmatrix} V_{1,T} \\ V_{2,T} \end{bmatrix} \xrightarrow{D} N \left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \frac{\Omega_{11}}{0.5} & 0 \\ 0 & \frac{\Omega_{22}}{0.5} \end{pmatrix} \right] \tag{2.5.2}$$

which, in turn, implies that

$$\underset{T\to\infty}{lim}\, Cov(V_{1,T}, V_{2,T}) = 0. \tag{2.5.3}$$

To see the above result, we modify the definition of LRCC to conform it to the structural stability testing framework. Let

$$\begin{aligned} \underset{p\times 1}{X_T(\pi)} &= T^{-1/2}\textstyle\sum_{t=1}^{[\pi T]} x_t, \\ \underset{q\times 1}{Z_T(\pi)} &= T^{-1/2}\textstyle\sum_{t=[\pi T]+1}^{T} z_t, \end{aligned}$$

and $m = min(p, q)$. Further, let

$$V_T(\pi) = \begin{bmatrix} X_T(\pi) \\ Z_T(\pi) \end{bmatrix},$$

and

$$\Sigma_v(\pi) = \lim_{T\to\infty} Var(V_T(\pi)) = \lim_{T\to\infty} \begin{pmatrix} \Sigma_{xx}(\pi) & \Sigma_{xz}(\pi) \\ \Sigma_{zx}(\pi) & \Sigma_{zz}(\pi) \end{pmatrix}.$$

Then LRCC's for fraction π, denoted $\rho_i(\pi),\ i = 1, 2, 3, \ldots, m$, where by convention $\rho_i(\pi) > 0,\ i = 1, 2, 3, \ldots, m$, and $\rho_i(\pi) \geq \rho_{i+1}(\pi)$ for $i = 1, 2, 3, \ldots, m-1$, are such that $\{\rho_i^2(\pi)\}$ are the m largest solutions to $|\Sigma_{xz}(\pi)\Sigma_{zz}^{-1}(\pi)\Sigma_{zx}(\pi) - \rho^2(\pi)\Sigma_{xx}(\pi)| = 0$.

Next, we show that the LRCC's must be zero if $T^{-1/2}\sum_{t=1}^{[\pi T]} x_t$ satisfies FCLT. To establish this result, set $x_t = z_t$ and assume that $X_T(\pi) \Rightarrow \Sigma_{xx}^{1/2} B_p(\pi)$, where (a) $\Sigma_{xx}^{1/2}\Sigma_{xx}^{1/2} = \Sigma_{xx}(1)$, and (b) $B_p(\pi)$ is a p-vector Brownian motion. Then, by the property of the vector Brownian motion [Hamilton (1994), p. 544], it follows that

$$\begin{aligned} \Sigma_{zx}(\pi) &= \underset{T\to\infty}{lim}\, Cov[X_T(\pi),\ X_T(1) - X_T(\pi)] \\ &= \underset{p\times p}{O}, \end{aligned}$$

implying that $\rho_i^2(\pi) = 0\ \forall i$.

Thus, using the definition of LRCC, it can be seen that (2.5.3) is equivalent to the restriction that $\rho_i = 0$ for $i = 1, 2, \cdots, k$, where $\{\rho_i\}_{i=1}^{k}$ are the LRCC between $\{v_t\}_{t=1}^{[0.5T]}$ and

$\{v_t\}_{t=[0.5T]+1}^{T}$. This suggests that asymptotic independence of subsamples can be tested by conducting a test of zero LRCC between subsamples. This, in turn, by virtue of Proposition 2.4.1, is equivalent to a test of zero canonical coherence at frequency zero between subsamples.

To simplify matters, in what follows, we set $k = 1$ and so v_t is now a scalar random variable. This reduces canonical coherences at frequency zero between two subsamples of random vectors to a simple coherence at frequency zero between two subsamples $\{v_t\}_{t=1}^{[0.5T]}$ and $\{v_t\}_{t=[0.5T]+1}^{T}$ of scalar random variables:

$$\rho_{12}(0) = \frac{f_{12}(0)}{\sqrt{f_{11}(0)}\sqrt{f_{22}(0)}}$$

where $\rho_{12}(.)$ denotes coherence between subsamples $\{v_t\}_{t=1}^{[0.5T]}$ and $\{v_t\}_{t=[0.5T]+1}^{T}$, $f_{12}(.)$ denotes cross-spectral density between subsamples $\{v_t\}_{t=1}^{[0.5T]}$ and $\{v_t\}_{t=[0.5T]+1}^{T}$, $f_{11}(.)$ denotes spectral density of subsample $\{v_t\}_{t=1}^{[0.5T]}$, and $f_{22}(.)$ denotes spectral density of subsample $\{v_t\}_{t=[0.5T]+1}^{T}$.

The population coherence $\rho_{12}(.)$ can be consistently estimated by the sample coherence:

$$\hat{\rho}_{12}(0) = \frac{\hat{f}_{12}(0)}{\sqrt{\hat{f}_{11}(0)}\sqrt{\hat{f}_{22}(0)}}$$

where the hats denote consistent estimates of the corresponding population quantities.

Exploiting conditions of Theorem 11 of Hannan [1970, pp. 288-289] and Proposition A.1. of Hall and Inoue (2003), it follows from Hannan [1970, p. 290] that, under $H_0 : \rho_{12}(0) = 0$,

$$\frac{\nu}{2}\, \hat{\rho}_{12}(0)^2 \xrightarrow{D} \chi^2_1$$

where ν, given in Hannan [1970, p. 281 and Table 1, p. 282], is the equivalent number of degrees of freedom of relevant spectral and cross-spectral density estimators. We call these tests Hannan tests.

We also consider LR test inspired by Hannan [1970, pp. 299-306]. Under $H_0 : \rho_{12}(0) = 0$,

$$-\frac{\nu}{2}\, ln[1 - \hat{\rho}_{12}(0)^2] \xrightarrow{D} \chi^2_1.$$

The formula for the equivalent degrees of freedom, ν, for the Bartlett kernel is: $\nu = 3T/M_T$, and for the Parzen kernel is: $\nu = 3.709T/M_T$.

In the following section, we conduct a simulation study to explore the finite sample performance of the three statistics: the Wald statistic W_T for structural stability testing; Hannan statistic $\frac{\nu}{2}\ \hat{\rho}_{12}(0)^2$ and LR statistic $-\frac{\nu}{2}\ ln[1-\hat{\rho}_{12}(0)^2]$ for pre-testing the assumption of the Wald structural stability testing, that is, for pre-testing the assumption of asymptotic independence of subsamples.

2.5.2 Simulation Design for Hannan and LR Tests of Persistence and Wald Test of Structural Stability

In this section, we simulate data from a time series model following AR(1) scheme:

$$v_t = \theta v_{t-1} + e_t, \qquad t = 1, 2, \cdots, T \tag{2.5.4}$$

where $e_t \sim N(0, 1)$ and the autoregressive parameter θ is generated as in Phillips (1987) by

$$\theta = exp(c_T/T), \quad (-\infty < c_T < \infty) \tag{2.5.5}$$

where T is the sample size and c_T is a noncentrality parameter. The noncentrallity parameter c_T measures deviations of the model from the unit root process. When $c_T = 0$, the model is unit root. When $c_T < 0$ and T is finite, $0 < \theta < 1$ and the model is stationary AR(1). When $c_T > 0$ and T is finite, $\theta > 1$ and the model is explosive. When $c_T \to 0$ and so $\theta \to 1$ as $T \to \infty$, the model is described as having a root that is local to unity.

We are interested in exploring the effectiveness of the Hannan and LR tests in screening out cases where the data are too persistent to allow asymptotic independence of subsamples, thereby rendering the Wald test of structural stability invalid. With this objective in mind, we choose values of noncentrality parameter c_T that generate the corresponding persistence parameter θ = 0.10, 0.25, 0.50, 0.75, 0.80, 0.85, 0.90, 0.95 and 1.00. We set the sample size T = 50, 100, 250, 500 and 1000. We use two kernels: Bartlett and Parzen. For the

Bartlett kernel, the values of the bandwidth chosen corresponding to the above values of T are, respectively, 3, 4, 6, 7 and 9. For the Parzen kernel, the corresponding values of the bandwidth chosen are, respectively, 4, 6, 8, 10 and 12.

We also conduct the above tests for the above five values of T for the near-integrated case. To this end, we set the values of the non-centrality parameter: c_T = -8.1259465, -5.2680258 and -2.5646647, that start us, respectively, at $\theta = exp(c_T/T)$ = 0.85, 0.90 and 0.95 for $T = 50$. Thus, given each value of c_T, as T gets bigger and bigger, θ gets closer and closer to one, thereby generating a near-integrated process.

The above tests are first conducted without prewhitening and recoloring. Then they are conducted with prewhitening and recoloring along the lines of Andrews and Monahan (1992). For this purpose, we first estimate the AR(1) model given by (2.5.4) to obtain $\hat{\theta}$ and then filter the series to give $w_t = v_t - \bar{\theta} v_{t-1}$, where

$$\bar{\theta} = \begin{cases} 0.97 & \text{if } \hat{\theta} \geq 0.97, \\ \hat{\theta} & \text{if } |\hat{\theta}| \leq 0.97, \\ -0.97 & \text{if } \hat{\theta} \leq 0.97. \end{cases}$$

We then estimate the long-run variance of the prewhitened series w_t, call this $\hat{\sigma}_w^2$, by an HACC estimator. Finally, we apply recoloring to this HACC estimate to obtain our desired estimate of the long-run variance as $\hat{\sigma}_w^2/(1-\bar{\theta})^2$.

We note that if $v_0 = 0$ and $\theta = 1$, then the data generating scheme is a unit root process and $v_t = \sum_{i=1}^{t} e_i$, and by Functional Central Limit Theorem, $T^{-1/2} \sum_{i=1}^{[\pi T]} e_i \Rightarrow B(\pi)$, where $B(\pi) \sim N(0, \pi)$, is a standardized Wiener or Brownian motion process for $0 \leq \pi \leq 1$.

Therefore, $T^{-1/2}v_t \Rightarrow B(\pi)$ for $[\pi T] = t$. Then

$$\begin{aligned}
\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{[0.5T]} v_t\right]\left[\frac{1}{\sqrt{T}}\sum_{t=[0.5T]+1}^{T} v_t\right] &= T^2\left[\frac{1}{T}\sum_{t=1}^{[0.5T]} T^{-1/2}v_t\right]\left[\frac{1}{T}\sum_{t=[0.5T]+1}^{T} T^{-1/2}v_t\right] \\
&= T^2\left[\frac{1}{T}\sum_{t=1}^{[0.5T]} T^{-1/2}v_t\right]\left[\frac{1}{T}\sum_{t=1}^{T} T^{-1/2}v_t - \frac{1}{T}\sum_{t=1}^{[0.5T]} T^{-1/2}v_t\right] \\
&\Rightarrow T^2\int_0^{0.5} B(r)dr\left[\int_0^1 B(r)dr - \int_0^{0.5} B(r)dr\right] \qquad (2.5.6)
\end{aligned}$$

where we have used Lemma 1(a) of Phillips (1986) to obtain $T^{-3/2}\sum_{t=1}^{[0.5T]} v_t \Rightarrow \int_0^{0.5} B(r)dr$ and $T^{-3/2}\sum_{t=1}^{T} v_t \Rightarrow \int_0^1 B(r)dr$ where " $\Rightarrow$ " denotes *Weak Convergence* of probability measures.

From (2.5.6) it follows that, with probability one, $\underset{T\to\infty}{lim}\, Cov\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{[0.5T]} v_t, \frac{1}{\sqrt{T}}\sum_{t=[0.5T]+1}^{T} v_t\right]$ diverges to ∞ at rate T^2. Thus, under unit root situations, as $T \to \infty$, the asymptotic covariance between subsamples grows without bound. Therefore, the assumption of asymptotic independence is invalid, and hence, use of the Wald test of structural stability is misleading.

2.5.3 Simulation Results of Hannan and LR Tests of Persistence and Wald Test of Structural Stability

The simulation results are presented in Tables 2.1-2.24. Tables 2.1-2.6 furnish results for AR(1) and unit root case without using prewhitening and recoloring. Tables 2.7-2.12 give results for the near-integrated case without prewhitening and recoloring. Tables 2.13-2.18 and Tables 2.19-2.24, respectively, provide results corresponding to the above cases but using prewhitening and recoloring. The results show the empirical size of the test when the nominal size is 5%. All results are computed using 10,000 simulations.

Inspection of results in Tables 2.1-2.2 reveals a clear pattern in the empirical size of the Hannan test: as the value of the persistence parameter θ increases, even when it stays below one, the empirical size increases, that is, the test becomes more and more oversized.

When the unit root in θ is reached, the time series process becomes nonstationary and the empirical size of the test increases significantly as we would expect since the null hypothesis of zero coherence between subsamples is now false.

The results in Tables 2.3-2.4 for the LR test show a pattern in the empirical size similar to that of the Hannan test. Such similarity in results, however, should come as no surprise since the LR statistic $-\frac{\nu}{2}\ ln[1-\hat{\rho}_{12}(0)^2]$ and the Hannan statistic $\frac{\nu}{2}\ \hat{\rho}_{12}(0)^2$ constitute a one-to-one mapping.

The results in Tables 2.5-2.6 clearly show that the Wald test rejects the null (truth) of no structural break too often: the empirical probability of rejecting the null increases significantly with the increase in the persistence in data. This corroborates our claim that inference based on Wald test of structural stability is misleading when data are persistent.

Corresponding results for the near-integrated case presented in Tables 2.7-2.12 lead to similar conclusions as above. Tables 2.7-2.10 clearly show that the Hannan and LR test of persistence perform nicely and Tables 2.11-2.12 provide evidence that the Wald test of structural stability is misleading when the time series is near-integrated.

However, as results in Tables 2.13-2.24 reveal, the picture changes dramatically when prewhitening and recoloring is used. Tables 2.13-2.18 show that the empirical size of all three tests, namely, the Hannan and LR tests of persistence as well as the Wald test of structural stability for AR(1) and unit root case is close to the nominal size of 5%, other than for large sample sizes in the unit root case. Similarly, the results for the near-integrated case using prewhitening and recoloring, furnished in Tables 2.19-2.24, yield empirical size of the test contrary to what we would expect and thus indicating a deterioration in the performance of the Hannan and LR test of persistence when prewhitening and recoloring is used.

2.5.4 Exogeneity Testing of Regressors in a Cointegration Model

In this section, we illustrate yet another use of LRCC. It is how testing exogeneity of integrated regessors can also be reduced to testing hypothesis about LRCC. Following Wooldridge (1994, Handbook of Econometrics, Vol. 4, p. 2713) consider the linear model

$$\underset{1\times 1}{y_t} = \underset{1\times 1}{\alpha_0} + \underset{1\times k}{x_t'}\underset{k\times 1}{\beta_0} + \underset{1\times 1}{u_t}, \qquad t = 1, 2, \cdots, \tag{2.5.7}$$

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

$$\underset{k\times 1}{x_t} = \underset{k\times 1}{x_{t-1}} + \underset{k\times 1}{v_t}, \tag{2.5.8}$$

where v_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 $\{w_t \equiv (u_t, v_t')'\}$ be a $(1+k)\times 1$ strictly stationary, weakly dependent stochastic process with zero mean and finite second moments. Define

$$\Sigma \equiv E(w_t w_t'), \qquad \Lambda \equiv \sum_{s=1}^{\infty} E(w_t w_{t-s}'),$$

and

$$\Omega \equiv \Sigma + \Lambda + \Lambda' \equiv \underset{T\to\infty}{lim}\ Var\left[T^{-1/2}\sum_{t=1}^{T} w_t\right] \equiv \begin{bmatrix} \Omega_{11} & \Omega_{21}' \\ \Omega_{21} & \Omega_{22} \end{bmatrix}.$$

Note that

$$\Sigma_{21} \equiv E(v_t u_t), \qquad \Lambda_{21} \equiv \sum_{s=1}^{\infty} E(v_t u_{t-s}),$$

and

$$\Delta_{21} \equiv \Sigma_{21} + \Lambda_{21}.$$

The model specified by (2.5.7) and (2.5.8) imply that y_t and x_t are cointegrated, and because the coefficient on y_t is normalized to unity, there is only one cointegrating vector. The parameters α_0 and β_0 can be estimated by OLS regression of

$$y_t \text{ on } 1, x_t, \qquad t = 1, \cdots, T.$$

Using Lemma 11.1 of Wooldridge 1994, Handbook Vol. 4, p. 2714, Park and Phillips (1988) derive the limiting distribution of $T(\hat{\beta}_T - \beta_0)$:

$$T(\hat{\beta}_T - \beta_0) \xrightarrow{D} \left[\int_0^1 \bar{B}_2(r)'\bar{B}_2(r)\ d(r)\right]^{-1} \left[\int_0^1 \bar{B}_2(r)'dB_1(r) + \Delta_{21}\right], \tag{2.5.9}$$

where $\bar{B}_2$ denotes the demeaned process B_2, that is, for each $0 \leq r \leq 1$,

$$\bar{B}_2(r) = B_2(r) - \int_0^1 B_2(s)ds \tag{2.5.10}$$

where $B_1(r)$ and $B_2(r)$ are independent Brownian motions of dimension 1 and k, respectively, with variance matrices Ω_{11} and Ω_{22}. The integral $\int_0^1 \bar{B}_2(r)'dB_1(r)$ is a vector of stochastic integrals with respect to the univariate Brownian motion $B_1(r)$ and the matrix $\int_0^1 \bar{B}_2(r)'\bar{B}_2(r)\ d(r)$ is a quadratic functional of the demeaned vector Brownian motion $\bar{B}_2(r)$ and is nonsingular with probability one.

The limiting distribution (2.5.9) depends, in an intractable way, on the nuisance parameters Ω_{21} and Δ_{21}. However, there is one case where this limiting distribution is independent of the nuisance parameters, namely, the case where the regressors are *strictly exogenous* in the sense that

$$E(\Delta x_t' u_s) = 0, \quad \forall t \text{ and } s. \tag{2.5.11}$$

The above condition implies that $\Delta_{21} = \Omega_{21} = 0$, so that the long run variance-covariance matrix of the $(1+k) \times 1$ vector $w_t \equiv (u_t, v_t')'$ becomes block-diagonal:

$$\Omega \equiv \Sigma + \Lambda + \Lambda' \equiv \lim_{T\to\infty} Var\left[T^{-1/2}\sum_{t=1}^{T} w_t\right] \equiv \begin{bmatrix} \Omega_{11} & \Omega_{21}' \\ \Omega_{21} & \Omega_{22} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{bmatrix}. \tag{2.5.12}$$

Using the definition of LRCC, it can be seen again that this condition is equivalent to the restriction that $\rho = 0$, where ρ is the LRCC between $\{u_t\}$ and $\{v_t\}$. Thus test of exogeneity is reduced to a test of zero LRCC between two error processes. This, in turn, again by the virtue of Proposition 1, is equivalent to a test of zero canonical coherence at frequency zero between the error processes. Because in the model under consideration, the long run correlation concerns correlation between a scalar and a vector, the test becomes a test of zero long run multiple correlation and as a corollary of Proposition 2.4.1, it becomes equivalent to a test of zero multiple coherence at frequency zero.

2.5.5 Simulation Design for Exogeneity Testing

In this section, we construct a simulation design where both the regressor x_t and the regressand y_t are scalar variables that are generated as follows:

$$x_t = x_{t-1} + v_t;\ v_t \sim N(0,1),$$

$$u_t = \gamma v_t + e_t;\ e_t \sim N(0,1),$$

$$y_t = \alpha_0 + x_t\beta_0 + u_t.$$

We set the values of the parameters: $\alpha_0 = 1.0$; $\beta_0 = 2.0$; $\gamma = 0.0, -0.8, -0.4, 0.4$ *and* 0.8. It is to be noted that $\gamma = 0.0$ corresponds to the case of exogeneity while $\gamma \neq 0.0$ corresponds to the case of endogeneity. Finally, the sample size: $T = 50, 100, 250, 500$ *and* 1000. We calculate the empirical size of the test 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. All results are computed using 10,000 simulations.

2.5.6 Simulation Results of Exogeneity Testing

Results pertaining to the empirical size of the test are shown in Tables 2.25 - 2.28. The first row in each of these tables corresponds to the endogeneity parameter $\gamma = 0.0$. Hence, the results on the first row correspond to case where the null hypothesis that the regressor x_t is exogenous, is true. As a consequence, we would expect the actual size of the test to be close to the nominal size of 5%, and that is precisely what we observe. Results on the remaining four rows in each of these tables correspond to the endogeneity parameter $\gamma \neq 0.0$. Hence, the null hypothesis that the regressor x_t is exogenous, is false. Therefore, in these four cases, we would expect the actual size of the test to be much larger than 5%, and that is exactly what the results show. In particular, the higher is the endogeneity and/or the sample size, the higher is the rejection rate. In addition, it is worth noting that the test performs highly satisfactorily even in moderate size samples.

Results relating to the empirical coverage probability of the 95% confidence interval of the estimator of β_0 are presented in Tables 2.29 - 2.30. As in the earlier four tables, the first row in each of these tables corresponds to the endogeneity parameter $\gamma = 0.0$. Hence, the null hypothesis is true. Hence, we would expect the empirical coverage probability to be close to the nominal coverage probability of 95%, and the results show precisely that. Results on the remaining four rows in each of these tables correspond to the endogeneity parameter $\gamma \neq 0.0$. Hence, the null hypothesis is false. So, in these four cases, we would expect to see distortions in actual coverage rates and the results bear testimony to that. Specifically, these results show that the higher the degree of endogeneity, the higher is the distortion.

Thus the above results clearly demonstrate the usefulness of long run canonical correlations in testing exogeneity of regressors in a cointegration model with a very high degree of accuracy.

2.6 Conclusions

The objective of this chapter has been to present the definition of LRCC, to formally derive LRCC and to prove a lemma that is the basis of Hall and Inoue's (2003) construction of RMSC. It also describes methods for estimating the LRCC and it establishes a link between LRCC and canonical coherence that has formerly been developed in frequency domain analysis. The final objective has been to present two additional uses of LRCC in econometric analysis of time series beyond their usefulness in moments selection for GMM estimation. One of them is development of the Hannan and LR tests of persistence designed to pre-test the assumption of asymptotic independence of subsamples underlying structural stability testing. The other additional use of LRCC show how they can be used for testing exogeneity of nonstationary regressors. We call these test of exogeneity the Hannan and LR tests of exogeneity.

This chapter also presents simulation studies of the above tests. It conducts the Hannan and LR tests of persistence and the Wald test of structural stability, both without prewhitening and recoloring and with prewhitening and recoloring. Simulation results without prewhitening and recoloring show that the Hannan and LR tests perform well and corroborate our claim that the use of the Wald test of structural stability is misleading when date are persistent because the empirical probability of rejecting the null (truth) of no structural break significantly increases with the increase in persistence. However, quality of performance of the Hannan and LR test of persistence deteriorates substantially when these tests are conducted using prewhitening and recoloring.

Finally, simulation results of the Hannan and LR tests of exogeneity indicate that their performance is highly satisfactory.

Table 2.1: Hannan Test of Persistence using Bartlett Kernel.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.075	0.061	0.054	0.060	0.057
0.25	0.096	0.080	0.070	0.064	0.068
0.50	0.143	0.121	0.100	0.090	0.081
0.75	0.257	0.219	0.165	0.150	0.131
0.80	0.292	0.263	0.199	0.178	0.159
0.85	0.335	0.305	0.254	0.218	0.196
0.90	0.384	0.370	0.309	0.282	0.266
0.95	0.453	0.460	0.424	0.413	0.393
1.00	0.511	0.565	0.638	0.697	0.776

Note: Nominal Size = 0.05.

Table 2.2: Hannan Test of Persistence using Parzen Kernel.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.052	0.049	0.046	0.051	0.051
0.25	0.068	0.057	0.054	0.051	0.056
0.50	0.096	0.084	0.074	0.066	0.059
0.75	0.180	0.148	0.125	0.111	0.092
0.80	0.210	0.184	0.154	0.129	0.113
0.85	0.243	0.219	0.201	0.164	0.145
0.90	0.287	0.266	0.254	0.217	0.202
0.95	0.349	0.360	0.370	0.348	0.320
1.00	0.403	0.466	0.590	0.653	0.734

Note: Nominal Size = 0.05.

Table 2.3: Likelihood Ratio Test of Persistence using Bartlett Kernel.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.094	0.072	0.062	0.064	0.060
0.25	0.115	0.094	0.078	0.068	0.071
0.50	0.171	0.138	0.109	0.095	0.083
0.75	0.285	0.240	0.176	0.156	0.135
0.80	0.321	0.282	0.212	0.185	0.163
0.85	0.361	0.327	0.267	0.224	0.200
0.90	0.414	0.392	0.321	0.290	0.270
0.95	0.477	0.476	0.436	0.421	0.397
1.00	0.537	0.579	0.646	0.701	0.778

Note: Nominal Size = 0.05.

Table 2.4: Likelihood Ratio Test of Persistence using Parzen Kernel.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.081	0.065	0.056	0.056	0.054
0.25	0.096	0.079	0.064	0.056	0.059
0.50	0.136	0.108	0.084	0.070	0.063
0.75	0.228	0.180	0.139	0.118	0.097
0.80	0.258	0.215	0.166	0.139	0.117
0.85	0.294	0.254	0.219	0.173	0.151
0.90	0.335	0.304	0.272	0.226	0.208
0.95	0.402	0.394	0.385	0.360	0.327
1.00	0.455	0.497	0.604	0.659	0.737

Note: Nominal Size = 0.05.

Table 2.5: Wald Test of Structural Stability using Bartlett Kernel.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.089	0.071	0.065	0.063	0.057
0.25	0.108	0.086	0.075	0.065	0.060
0.50	0.175	0.134	0.098	0.082	0.077
0.75	0.324	0.248	0.172	0.145	0.130
0.80	0.370	0.299	0.214	0.171	0.161
0.85	0.436	0.356	0.268	0.224	0.205
0.90	0.508	0.443	0.354	0.305	0.277
0.95	0.601	0.580	0.494	0.447	0.426
1.00	0.713	0.751	0.799	0.848	0.882

Note: Nominal Size = 0.05.

Table 2.6: Wald Test of Structural Stability using Parzen Kernel.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.104	0.075	0.063	0.062	0.055
0.25	0.107	0.080	0.067	0.058	0.055
0.50	0.143	0.107	0.079	0.064	0.059
0.75	0.255	0.194	0.138	0.108	0.090
0.80	0.296	0.241	0.176	0.128	0.114
0.85	0.353	0.292	0.227	0.172	0.145
0.90	0.429	0.374	0.310	0.242	0.202
0.95	0.526	0.520	0.457	0.385	0.347
1.00	0.652	0.711	0.782	0.827	0.859

Note: Nominal Size = 0.05.

Table 2.7: Hannan Test of Persistence using Bartlett Kernel; Near-Integrated Case.

Empirical Size of the test

$c \diagdown T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.344	0.407	0.490	0.581	0.667
-5.268026	0.380	0.448	0.528	0.626	0.695
-2.564665	0.446	0.508	0.582	0.664	0.729

Note: Nominal Size = 0.05.

Table 2.8: Hannan Test of Persistence using Parzen Kernel; Near-Integrated Case.

Empirical Size of the test

$c \diagdown T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.252	0.308	0.434	0.524	0.614
-5.268026	0.275	0.344	0.477	0.574	0.645
-2.564665	0.346	0.408	0.531	0.615	0.686

Note: Nominal Size = 0.05.

Table 2.9: Likelihood Ratio Test of Persistence using Bartlett Kernel; Near-Integrated Case.

Empirical Size of the test

$c \diagdown T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.371	0.427	0.501	0.587	0.670
-5.268026	0.405	0.466	0.538	0.631	0.698
-2.564665	0.472	0.528	0.591	0.669	0.733

Note: Nominal Size = 0.05.

Table 2.10: Likelihood Ratio Test of Persistence using Parzen Kernel; Near-Integrated Case.

Empirical Size of the test

$c \diagdown T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.299	0.342	0.449	0.532	0.618
-5.268026	0.330	0.379	0.494	0.582	0.649
-2.564665	0.396	0.440	0.546	0.623	0.690

Note: Nominal Size = 0.05.

Table 2.11: Wald Test of Structural Stability using Bartlett Kernel; Near-Integrated Case.

Empirical Size of the test

$c \setminus T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.434	0.506	0.573	0.669	0.736
-5.268026	0.510	0.563	0.652	0.719	0.790
-2.564665	0.605	0.657	0.719	0.775	0.824

Note: Nominal Size = 0.05.

Table 2.12: Wald Test of Structural Stability using Parzen Kernel; Near-Integrated Case.

Empirical Size of the test

$c \setminus T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.358	0.436	0.535	0.625	0.684
-5.268026	0.430	0.500	0.622	0.673	0.749
-2.564665	0.529	0.602	0.694	0.740	0.785

Note: Nominal Size = 0.05.

Table 2.13: Hannan Test of Persistence using Bartlett Kernel and Prewhitening and Recoloring.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.076	0.055	0.050	0.050	0.051
0.25	0.075	0.057	0.050	0.049	0.051
0.50	0.074	0.056	0.051	0.051	0.048
0.75	0.080	0.060	0.053	0.050	0.048
0.80	0.077	0.060	0.054	0.054	0.053
0.85	0.076	0.065	0.055	0.049	0.051
0.90	0.077	0.061	0.051	0.051	0.055
0.95	0.076	0.065	0.059	0.056	0.053
1.00	0.073	0.056	0.069	0.129	0.328

Note: Nominal Size = 0.05.

Table 2.14: Hannan Test of Persistence using Parzen Kernel and Prewhitening and Recoloring.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.051	0.047	0.045	0.049	0.050
0.25	0.055	0.049	0.048	0.047	0.051
0.50	0.054	0.048	0.049	0.049	0.048
0.75	0.055	0.050	0.051	0.049	0.047
0.80	0.057	0.051	0.051	0.052	0.052
0.85	0.055	0.057	0.054	0.047	0.050
0.90	0.055	0.055	0.049	0.049	0.054
0.95	0.057	0.056	0.056	0.054	0.053
1.00	0.053	0.048	0.069	0.128	0.327

Note: Nominal Size = 0.05.

Table 2.15: Likelihood Ratio Test of Persistence using Bartlett Kernel and Prewhitening and Recoloring.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.094	0.065	0.056	0.055	0.055
0.25	0.093	0.069	0.057	0.052	0.054
0.50	0.092	0.066	0.058	0.055	0.050
0.75	0.101	0.072	0.062	0.054	0.050
0.80	0.098	0.071	0.061	0.058	0.055
0.85	0.097	0.075	0.062	0.052	0.053
0.90	0.095	0.071	0.058	0.054	0.057
0.95	0.096	0.074	0.066	0.059	0.057
1.00	0.089	0.068	0.076	0.136	0.332

Note: Nominal Size = 0.05.

Table 2.16: Likelihood Ratio Test of Persistence using Parzen Kernel and Prewhitening and Recoloring.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.077	0.064	0.053	0.055	0.053
0.25	0.082	0.068	0.056	0.052	0.054
0.50	0.081	0.063	0.058	0.055	0.051
0.75	0.084	0.069	0.059	0.053	0.050
0.80	0.085	0.070	0.061	0.057	0.054
0.85	0.084	0.075	0.063	0.053	0.054
0.90	0.083	0.073	0.058	0.053	0.058
0.95	0.085	0.077	0.065	0.059	0.056
1.00	0.081	0.066	0.077	0.137	0.332

Note: Nominal Size = 0.05.

Table 2.17: Wald Test of Structural Stability using Bartlett Kernel and Prewhitening and Recoloring.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.069	0.062	0.059	0.059	0.051
0.25	0.060	0.060	0.060	0.054	0.052
0.50	0.049	0.055	0.056	0.049	0.051
0.75	0.026	0.039	0.046	0.051	0.048
0.80	0.020	0.033	0.046	0.049	0.050
0.85	0.013	0.021	0.038	0.048	0.046
0.90	0.009	0.010	0.030	0.041	0.045
0.95	0.005	0.006	0.014	0.030	0.040
1.00	0.005	0.050	0.374	0.634	0.784

Note: Nominal Size = 0.05.

Table 2.18: Wald Test of Structural Stability using Parzen Kernel and Prewhitening and Recoloring.

Empirical Size of the test

$\theta \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.10	0.096	0.073	0.062	0.062	0.055
0.25	0.088	0.071	0.064	0.056	0.054
0.50	0.079	0.067	0.058	0.052	0.054
0.75	0.055	0.052	0.050	0.054	0.050
0.80	0.044	0.048	0.051	0.052	0.053
0.85	0.034	0.033	0.041	0.051	0.048
0.90	0.023	0.020	0.035	0.044	0.047
0.95	0.017	0.012	0.018	0.034	0.043
1.00	0.013	0.059	0.373	0.625	0.771

Note: Nominal Size = 0.05.

Table 2.19: Hannan Test of Persistence using Bartlett Kernel and Prewhitening and Recoloring; Near-Integrated Case.

Empirical Size of the test

$c \diagdown T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.082	0.059	0.054	0.067	0.131
-5.268026	0.078	0.063	0.059	0.076	0.185
-2.564665	0.075	0.059	0.061	0.098	0.240

Note: Nominal Size = 0.05.

Table 2.20: Hannan Test of Persistence using Parzen Kernel and Prewhitening and Recoloring; Near-Integrated Case.

Empirical Size of the test

$c \diagdown T$	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.059	0.054	0.051	0.066	0.131
-5.268026	0.058	0.056	0.056	0.075	0.185
-2.564665	0.056	0.055	0.061	0.097	0.241

Note: Nominal Size = 0.05.

Table 2.21: Likelihood Ratio Test of Persistence using Bartlett Kernel and Prewhitening and Recoloring; Near-Integrated Case.

Empirical Size of the test

c ＼ *T*	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.100	0.069	0.060	0.071	0.134
-5.268026	0.094	0.074	0.066	0.080	0.189
-2.564665	0.092	0.070	0.069	0.102	0.244

Note: Nominal Size = 0.05.

Table 2.22: Likelihood Ratio Test of Persistence using Parzen Kernel and Prewhitening and Recoloring; Near-Integrated Case.

Empirical Size of the test

c ＼ *T*	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.088	0.074	0.061	0.071	0.136
-5.268026	0.086	0.075	0.066	0.082	0.190
-2.564665	0.083	0.074	0.070	0.104	0.247

Note: Nominal Size = 0.05.

Table 2.23: Wald Test of Structural Stability using Bartlett Kernel and Prewhitening and Recoloring; Near-Integrated Case.

Empirical Size of the test

c ╲ *T*	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.014	0.008	0.033	0.204	0.463
-5.268026	0.010	0.007	0.089	0.305	0.575
-2.564665	0.006	0.009	0.195	0.455	0.664

Note: Nominal Size = 0.05.

Table 2.24: Wald Test of Structural Stability using Parzen Kernel and Prewhitening and Recoloring; Near-Integrated Case.

Empirical Size of the test

c ╲ *T*	*50*	*100*	*250*	*500*	*1000*
-8.125946	0.034	0.016	0.037	0.201	0.444
-5.268026	0.026	0.013	0.093	0.300	0.557
-2.564665	0.019	0.017	0.194	0.445	0.647

Note: Nominal Size = 0.05.

Table 2.25: Hannan Test of Exogeneity using Bartlett Kernel.

Empirical Size of the test

$\gamma \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.0	0.050	0.049	0.051	0.046	0.051
-0.8	0.871	0.978	1.000	1.000	1.000
-0.4	0.407	0.599	0.833	0.973	0.999
0.4	0.402	0.589	0.830	0.976	0.998
0.8	0.870	0.975	0.999	1.000	1.000

Note: Nominal Size = 0.05.

Table 2.26: Hannan Test of Exogeneity using Parzen Kernel.

Empirical Size of the test

$\gamma \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.0	0.045	0.047	0.052	0.046	0.050
-0.8	0.851	0.946	0.999	1.000	1.000
-0.4	0.374	0.504	0.806	0.954	0.998
0.4	0.373	0.505	0.802	0.953	0.998
0.8	0.846	0.949	0.999	1.000	1.000

Note: Nominal Size = 0.05.

Table 2.27: Likelihood Ratio Test of Exogeneity using Bartlett Kernel.

Empirical Size of the test

$\gamma \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.0	0.060	0.055	0.054	0.048	0.052
-0.8	0.885	0.980	1.000	1.000	1.000
-0.4	0.437	0.619	0.841	0.974	0.999
0.4	0.436	0.611	0.838	0.976	0.998
0.8	0.884	0.978	0.999	1.000	1.000

Note: Nominal Size = 0.05.

Table 2.28: Likelihood Ratio Test of Exogeneity using Parzen Kernel.

Empirical Size of the test

$\gamma \diagdown T$	*50*	*100*	*250*	*500*	*1000*
0.0	0.057	0.055	0.056	0.048	0.052
-0.8	0.869	0.952	0.999	1.000	1.000
-0.4	0.411	0.534	0.815	0.957	0.998
0.4	0.411	0.530	0.811	0.955	0.998
0.8	0.867	0.954	0.999	1.000	1.000

Note: Nominal Size = 0.05.

Table 2.29: Hannan Test of Exogeneity using Bartlett Kernel.

Coverage Probability of 95% Confidence Interval for β_0

γ \ T	*50*	*100*	*250*	*500*	*1000*
0.0	0.926	0.932	0.936	0.944	0.943
-0.8	0.833	0.837	0.846	0.855	0.850
-0.4	0.886	0.895	0.907	0.910	0.909
0.4	0.887	0.901	0.906	0.911	0.916
0.8	0.832	0.836	0.845	0.845	0.856

Table 2.30: Hannan Test of Exogeneity using Parzen Kernel.

Coverage Probability of 95% Confidence Interval for β_0

γ \ T	*50*	*100*	*250*	*500*	*1000*
0.0	0.925	0.929	0.936	0.944	0.943
-0.8	0.833	0.834	0.845	0.854	0.850
-0.4	0.885	0.893	0.907	0.910	0.910
0.4	0.886	0.898	0.906	0.911	0.915
0.8	0.832	0.832	0.845	0.844	0.855

Chapter 3

A Comparative Study of the CCIC, the RMSC and the MSE

3.1 Introduction

The objective of this chapter is to do a comparative study of the three methods CCIC, RMSC and MSE in the context of a simple linear IV model:

$$y_t = x_t\theta_0 + u_t, \qquad (3.1.1)$$

$$x_t = {z_t}'\pi_0 + e_t, \quad t = 1, 2, \ldots, T \qquad (3.1.2)$$

where y_t is the dependent variable, x_t is a scalar regressor correlated with the regression error u_t which has a mean zero and which could exhibit heteroscedasticy and autocorrelation. A $q \times 1$ vector of valid instruments, z_t, is available so that $cov(x_t, z_t) \neq 0$ and satisfies the population moment conditions $E[z_t u_t(\theta_0)] = 0$. The last condition implies that we are considering instrument selection in models that are correctly specified.

The contribution of the paper to the emerging literature on instrument selection is that it brings the above three papers together, synthesizes and unifies them by showing through an analytical comparison the common thread that runs between them, and brings out their relative strengths and weaknesses via a simulation study.

Although the three methods under study are tied by the common goal of instrument selection, they are different in terms of their underlying objectives. Donald and Newey's (2001) objective is to achieve an improved finite sample risk property of the estimators. They attain this goal by minimizing the approximate MSE. The objective of Hall and Peixe (2003) and Hall and Inoue (2003), on the other hand, is to achieve an improved quality of asymptotic approximation to the finite sample behavior of the estimators. They gain this objective by eliminating the redundant moment conditions based, respectively, on SRCC and LRCC.

Even though each of the aforementioned papers explores the properties of its proposed method, there have been to date no comparative studies of these methods. What, if any, is the analytical connection among the three methods? Is there any systematic pattern in the finite sample properties of the post selection estimators obtained by the three methods so that a unique ranking of these methods emerge? If no systematic pattern emerges, can we still provide guidance to a practitioner as to which of these methods should be used in any particular real life application of interest?

In this chapter, we explore the above issues. We begin by briefly summarizing the CCIC, the RMSC and the approximate MSE, and by presenting the corresponding criteria in the context of a simple linear IV model specified in Section 1 and to be subsequently used in simulation. Then we establish a relation between contemporaneous and long run canonical correlations in a linear simultaneous equation model. Next we show an analytical connection among the three criteria. Then we follow it up by an assessment of their relative performance via a simulation study that investigates the finite sample behavior of the post selection estimator. To this end, we compare median bias of the post selection estimator and coverage probability of 90% confidence interval under the three criteria.

3.2 Summary of the three Criteria

3.2.1 Hall and Peixe's Canonical Correlations Information Criterion

The objective of Hall and Peixe (2003) is to develop a method of instrument selection from a *fixed* candidate set in the context of Generalized Method of Moments (GMM) models (Hansen, 1982) by satisfying five desirable conditions of moment selection, namely, (i) identification: the moment conditions be satisfied at only one value in the parameter space; (ii) orthogonality: this value be the "true" parameter value, θ_0 say, implying that they represent valid information; (iii) efficiency: they minimize the asymptotic variance of the estimator; (iv) non-redundancy: the asymptotic variance increases if any element of the selected vector is excluded, and (v) inference: the selection process must not contaminate the asymptotic distribution theory of the estimator. The authors define a $q_{max} \times 1$ selection vector c to index the instrument vector which they denote as $z_t(c)$. The value of q_{max} is fixed, independent of sample size, and the elements of c indicate which elements of the candidate set are included in $z_t(c)$ and which elements are excluded: $c_j = 1$ implies that the j^{th} element is included, and $c_j = 0$ implies that the j^{th} element is excluded. They focus exclusively on the particular class of GMM models in which the population moment conditions take the form

$$E[z_t(c)u_t(\theta_0)] = 0 \tag{3.2.1}$$

where $u_t(.)$ is a scalar possibly nonlinear function of a set of dynamic random variables, $z_t(c)$ is a vector of instruments selected from a candidate set that is *fixed* and the asymptotic variance of the GMM estimator is

$$V(c) = \sigma_0^2 \, \{E[d_t(\theta_0)z_t(c)^{'}]\{E[z_t(c)z_t(c)^{'}]\}^{-1}[z_t(c)d_t(\theta_0)^{'}]\}^{-1} \tag{3.2.2}$$

where σ_0^2 is the variance of $u_t(\theta_0)$ under the assumption of conditional homoscedasticity, that is, $\sigma_0^2 = E[u_t(\theta_0)^2|\Omega_t]$, Ω_t is the information set at date t, and $d_t(\theta_0) = \partial u_t(\theta_0)/\partial\theta$.

Validity of the "conventional" asymptotic distribution theory of the estimator used for inference about the parameters of the underlying model as well as for adequacy of this theory as an approximation to the finite sample behavior hinges crucially on simultaneous satisfaction of the above five conditions. Existing methods of instrument selection fail to meet this requirement. For example, while Andrews' (1999) proposed information criterion approach satisfies the conditions of orthogonality and inference, it need not satisfy the condition of non-redundancy. It depends on moment condition and whether or not there are any redundant moment conditions. Hall and Peixe (2003) show that the use of this method can lead to inclusion of redundant instruments which causes a deterioration in the quality of the conventional asymptotic approximation to the finite sample behavior of the estimator. This evidence motivates Hall and Peixe (2003) to consider the problem of instrument selection based on a combination of the efficiency and non-redundancy conditions which they refer to as the relevance condition.

Basing instrument selection on relevance condition requires a metric of relevance. Hall and Peixe (2003) show that certain canonical correlations provide a natural metric for relevance by adapting Sargan's (1958) arguments that the asymptotic variance $V(c)$ in (3.2.2) above can be rewritten as

$$V(c) = \sigma_0^2 A(c)\Lambda(c)^{-2}A(c)' \tag{3.2.3}$$

where $\Lambda(c) = diag(\rho_1(c)\ldots\rho_p(c))$, $\{\rho_i(c);\ i = 1, 2, \ldots p\}$ are canonical correlations between $d_t(\theta_0)$ and $z_t(c)$, and $A(c)$ is the $p \times p$ matrix whose i^{th} row contains the weights in the linear combinations associated with $d_t(\theta_0)$ in the i^{th} canonical correlation, that is,

$$Corr[a_i(c)' d_t(\theta_0), b_i(c)' z_t(c)] = \rho_i(c) \tag{3.2.4}$$

where $b_i(c)$ is used to denote the vector of weights associated with $z_t(c)$.

Equation (3.2.3) reveals that the asymptotic variance depends crucially on the population canonical correlations. Hall and Peixe (2003) exploit this result to show how these population canonical correlations can provide a suitable metric for redundancy. To this end, they define an additional set of instruments to be *redundant*[1] when the inclusion of this set has no impact on the population canonical correlations. This implies through equation (3.2.3) that the inclusion of this set should have no impact on the asymptotic variance of the estimator. Thus, formally,

Definition 3.2.1 *Let $c_i \in C$ for $i = 1, 2$ and satisfy $c_1'c_2 = 0$, then $z_t(c_2)$ is redundant given $z_t(c_1)$ if and only if $V(c_1 + c_2) = V(c_1)$ where $V(c)$ is given by (3.2.2).*

Conversely, they define an additional set of instruments to be *non-redundant* when the inclusion of this set increases[2] at least one of the population canonical correlations. This, in turn, implies through equation (3.2.3), that the exclusion of this set increases the asymptotic variance of the estimator, and hence, formally,

Definition 3.2.2 *Let $c_i \in C$ for $i = 1, 2$ and satisfy $c_1'c_2 = 0$, then $z_t(c_2)$ is redundant given $z_t(c_1)$ if and only if $V(c_1 + c_2) - V(c_1)$ is negative definite.*

Thus the above definitions of redundancy and non-redundancy show how the population canonical correlations can be used to deduce which instruments are redundant, and so form the natural basis for a method of instrument selection based on relevance. They characterize a set of instruments to be relevant if, given this set, the remaining elements of the candidate set are redundant and no element of this set is redundant given the other elements of this set. Their objective is to select from a valid candidate set this relevant instrument set with probability one in the limit and their proposed canonical correlations information criterion (CCIC) is defined to be

$$CCIC(c) = \; \Xi_T(c) \; + \; P(T, |c|) \tag{3.2.5}$$

where the statistic

$$\Xi_T(c) = \; T \sum_{i=1}^{p} ln[1 - r_{i,T}^2(c)] \tag{3.2.6}$$

[1]For other equivalent definitions of redundancy, see Breusch, Quian, Schmidt, and Wyhowski (1999).

[2]Note that the population canonical correlations can never decrease as a result of augmenting the instrument vector.

captures the sample information, $|c| = c'c$ equals the number of elements in $z_t(c)$ and $P(T, |c|)$ is a "penalty" term which satisfies the following conditions: *(i)* $P(T, |c|) = h(|c|)\mu_T$; *(ii) h(.) is non-negative and strictly increasing; (iii)* $\mu_T \to \infty$ *as* $T \to \infty$ *and* $\mu_T = o(T)$.

In our specified simple linear IV model given by equations (3.1.1) and (3.1.2) above, the regressor x_t is a scalar. Therefore, the CCIC involves only one sample canonical correlation r_T which is equal to the multiple correlation coefficient, also commonly known as the coefficient of determination. Thus our criterion takes the form

$$CCIC(c) = T\, ln[1 - r_T^2(c)] \; + \; P(T, |c|). \tag{3.2.7}$$

Possible choices of the penalty term $P(T, |c|)$ correspond to Akaike's (1974) criterion where: $h(|c|) = |c| - p,\ \mu_T = 2$; Schwarz's (1978) criterion where: $h(|c|) = |c| - p,\ \mu_T = lnT$; and Hannan and Quinn's (1979) criterion where: $h(|c|) = |c| - p,\ \mu_T = QlnlnT$, for some $Q > 2$.

Because in Akaike's (1974) criterion $\mu_T = 2$ does not tend to infinity, condition (iii) above for the penalty term is violated, and so we adopt the specifications corresponding to Schwarz (1978) and Hannan and Quinn (1979). Of these two, we have explored the case corresponding to Schwarz.

3.2.2 Hall and Inoue's Relevant Moments Selection Criterion

The objective of Hall and Inoue (2003) is to generalize the CCIC of Hall and Peixe (2003) to Relevant Moments Selection Criterion (RMSC) for GMM models in nonlinear dynamic setting. The RMSC is based on their result that the asymptotic variance of the GMM estimator can be written in terms of the population long run canonical correlations between the population moment condition used in the estimation and the unknown true score vector associated with the data.

Hall and Inoue (2003) assume that the data generating process is strictly stationary and ergodic. They consider the GMM estimator of the unknown parameter vector θ_0 based on a *finite* set of population moment conditions: $E[f(v_t, \theta_0)] = 0$ where $f : \mathcal{V} \times \Theta \rightarrow \Re^q$. This estimator is defined to be

$$\hat{\theta}_T(c) = Argmin_{\theta \in \Theta} g_T(\theta; c)' W_T g_T(\theta; c) \tag{3.2.8}$$

where $\hat{\theta}_T$ and g_T are indexed by the selection vector c to indicate which moments from the candidate set are included and which are excluded: $c_j = 1$ implies that the j^{th} moment is included, $c_j = 0$ implies that the j^{th} moment is excluded; $g_T(\theta; c) = T^{-1} \sum_{t=1}^{T} f(v_t, \theta; c)$; and W_T is a positive semi-definite weighting matrix which converges in probability to S^{-1}, where

$$S = \lim_{T \rightarrow \infty} Var[T^{1/2} g_T(\theta_0)]. \tag{3.2.9}$$

They premise their analysis on the assumption that the asymptotic distribution of $\hat{\theta}_T(c)$ takes the following form:

$$T^{1/2}(\hat{\theta}_T(c) - \theta_0) \xrightarrow{d} N(0, \Omega(c)) \tag{3.2.10}$$

where $\Omega(c) = [G_0'(c) S^{-1} G_0(c)]^{-1}, G_0(c) = E[\partial f(v_t, \theta_0; c)/\partial \theta']$, and show that under certain conditions, $\Omega(c)$ can be written as:

$$\Omega(c) = L(c) R(c)^{-2} L(c)' \tag{3.2.11}$$

where $R(c) = diag(\rho_1(c), \rho_2(c), \ldots, \rho_p(c)), \{\rho_i(c); i = 1, 2, \ldots, p\}$ are the set of population long run canonical correlations between the population moment vector $f(v_t, \theta_0; c)$ and the score vector $s_{\theta,t}(\psi_0; c)$, $L(c)$ is the $p \times p$ matrix with i^{th} column $l_i(c)$, $l_i(c)$ is the generalized eigenvector satisfying $[G_0(c)' S^{-1} G_0(c) - \rho_i(c)^2 \mathcal{I}_\theta(c)] l_i(c) = 0$ and $\mathcal{I}_\theta(c) = E[s_{\theta,t}(\psi_0; c) s_{\theta,t}(\psi_0; c)']$.

The expression for $\Omega(c)$ in (3.2.11) above shows that the variance of the GMM estimator depends crucially on the long run canonical correlations between the population moment and the score. Hall and Inoue (2003) exploit this result to deduce a condition for redundancy of moment conditions. They define a set of additional moment conditions as redundant if inclusion of this set has no impact on the long run canonical correlations between the population moment vector and the score vector. Using this concept of redundancy they develop a method for the selection of the relevant set of moment conditions from a valid candidate set. They characterize a set of moment conditions to be relevant if given this set, all other members of the candidate set are redundant and no member of this set is redundant given the remaining members of this set. The relevant set is asymptotically efficient within the class of moment conditions that can be constructed from the valid set. Thus the idea of Hall and Inoue (2003) is to construct an information criterion that involves estimated asymptotic variance. Their proposed criterion is

$$RMSC = ln[|\hat{\Omega}(c)|] + \mu(|c|, T) \tag{3.2.12}$$

where $\hat{\Omega}(c)$ is a consistent estimator of $\Omega(c)$ and $\mu(|c|, T) = \frac{(|c|-p)}{\sqrt{T}} ln\sqrt{T}$. [3]

Specializing these results to our model given by equations (3.1.1) and (3.1.2) yields the criterion:

$$RMSC(c) = ln\left(|\hat{\sigma}_u^2[x'z(c)\{z(c)'z(c)\}^{-1}z(c)'x]^{-1}|\right) + \frac{(|c|-1)}{\sqrt{T}}\, ln\sqrt{T}. \tag{3.2.13}$$

3.2.3 Donald and Newey's Approximate Mean-Square Error Criterion

While Hall and Peixe (2003) and Hall and Inoue(2003) keep the candidate set of instruments *fixed*, Donald and Newey (2001) allow the candidate set to *grow* with the sample size. Their objective is to develop a criterion for instrument selection based on minimizing Nagar (1959) type approximation of the mean square error (MSE) of the estimator. They show that this method can improve the finite sample properties of IV estimators, including among others, the two-stage least squares (2SLS) estimators. Choosing instruments to minimize MSE helps reduce misleading IV inferences that can occur with many instruments. For 2SLS, the MSE explicitly accounts for an important bias term [Bound, Jaeger, and Baker (1995)],

[3]This penalty is correct under (3.1.2), but not necessarily in all cases.

so choosing instruments to minimize MSE avoids cases where asymptotic inferences are poor due to the bias being large relative to the standard deviation.

In our notation, their model becomes

$$y = x\,\theta_0 + u; \tag{3.2.14}$$

$$x = f(z) + e \tag{3.2.15}$$

where they allow the function $f(.)$ in the first stage reduced form equation (3.2.15) to be of any unknown form that can be approximated arbitrarily closely by nonparametric methods.

Their method is based on higher-order asymptotics as follows. Expand the centered and scaled 2SLS estimator:

$$\begin{aligned}
\sqrt{T}(\hat{\theta} - \theta_0) &= (\frac{x'Px}{T})^{-1}\,\frac{x'Pu}{\sqrt{T}} \\
&= (\frac{\hat{f}'\hat{f}}{T})^{-1}\,\frac{\hat{f}'u}{\sqrt{T}} \\
&= \hat{H}^{-1}\hat{h} \\
&= [H - (H - \hat{H})]^{-1}[h + (\hat{h} - h)] \\
&= [\{I - (H - \hat{H})H^{-1}\}H]^{-1}[h + (\hat{h} - h)] \\
&= [H^{-1}\{I - (H - \hat{H})H^{-1}\}^{-1}][h + (\hat{h} - h)] \\
&= H^{-1}[I + (H - \hat{H})H^{-1} + (H - \hat{H})^2H^{-2} + \cdots][h + (\hat{h} - h)] \\
&= H^{-1}[h + \hat{h} - h + (H - \hat{H})H^{-1}h + (H - \hat{H})H^{-1}(\hat{h} - h) + (H - \hat{H})^2H^{-2}h + \cdots] \\
&= H^{-1}[\hat{h} + (H - \hat{H})H^{-1}h + (H - \hat{H})H^{-1}(\hat{h} - h) + (H - \hat{H})^2H^{-2}h + ...]
\end{aligned}$$

where P is the projection matrix of basis functions used to approximate the unknown function f, $H = \frac{f'f}{T}$, and $h = \frac{f'u}{\sqrt{T}}$. Now square the above expansion and take expectation of terms that are largest in probability and are functions of the number of instruments q. This yields the desired approximate mean square error criterion:

$$MSE(c) = \sigma_{eu}^2\frac{|c|^2}{T} + \sigma_u^2\,\frac{f'(I-P)f}{T}. \tag{3.2.16}$$

Minimizing the approximate MSE requires estimating the approximate MSE. This in turn requires preliminary estimates of the parameters of the model and a goodness of fit criterion $R(c)$ for estimation of the first stage reduced form using the instruments $z_t(c)$. For example, the preliminary estimator might be an IV estimator with only as many instruments as right-hand side variables, or it might be an IV estimator where the instruments are chosen to minimize one of the first stage goodness of fit criteria. Donald and Newey (2001) consider the cross-validation and Mallows' (1973) reduced form goodness of fit criteria.

They minimize the approximate MSE of a linear combination $\hat{\lambda}'\hat{\theta}_T(c)$ of the IV estimator, where $\hat{\lambda}$ is some vector of estimated linear combination coefficients. Because the structural parameter θ_0 in our model is a scalar, for our purposes, λ is equal to one.

Let $\tilde{\theta}_T(\tilde{c})$ and $\tilde{\pi}_T(\tilde{c})$ be some preliminary estimators of θ_0 and π_0, respectively. Then the estimated approximate MSE of the 2SLS estimator of our proposed model is

$$MSE(c) = \hat{\sigma}^2_{eu}\frac{|c|^2}{T} + \hat{\sigma}^2_u\left(\hat{R}(|c|) - \hat{\sigma}^2\frac{|c|}{T}\right) \tag{3.2.17}$$

where for any preliminary selection vector $\tilde{c}$, $\hat{\sigma}_{eu} = \tilde{e}'(\tilde{c})\tilde{u}(\tilde{c})/T$, $\tilde{e}(\tilde{c}) = \tilde{e}(c)[\tilde{\pi}'(\tilde{c})z'(\tilde{c})z(\tilde{c})$ $\tilde{\pi}(\tilde{c})/T]^{-1}, \tilde{e}(c) = [I - P(c)]x$ (Note, the vector of residuals $\tilde{e}(c)$ is calculated using all elements of the candidate set; $c_j = 1$ for $j = 1, 2, \ldots, q_{max}$), $P(c) = z(c)[z(c)'z(c)]^{-1}z(c)'$, $\tilde{u}(\tilde{c}) = y - x\tilde{\theta}_T(\tilde{c})$, $\hat{\sigma}^2_u(\tilde{c}) = \tilde{u}(\tilde{c})'\tilde{u}(\tilde{c})/T$, $\hat{\sigma}^2(\tilde{c}) = \tilde{e}(\tilde{c})'\tilde{e}(\tilde{c})/T$, and the cross-validation criterion is,

$$\begin{aligned}\hat{R}^{cv}(c) &= \frac{1}{T}\sum_{i=1}^{T}\frac{[\hat{e}_i(c)]^2}{[1-P_{ii}(c)]^2}\\ &= \frac{1}{T}\sum_{i=1}^{T}\left[\frac{\hat{e}_i(c)}{[1-P_{ii}(c)}\right]^2\end{aligned}$$

and the Mallows' criterion is,

$$\hat{R}^m(c) = \frac{\hat{e}(\tilde{c})'\hat{e}(\tilde{c})}{T} + \frac{\hat{e}(c)'\hat{e}(c)}{T}\,(2\,\frac{|\tilde{c}|}{T}).$$

3.3 Relationship between Contemporaneous and Long Run Canonical Correlations

In this section, we show the relationship between contemporaneous and long run canonical correlations in the following linear simultaneous equation model:

$$\begin{aligned} \underset{(1\times 1)}{y_t} &= \underset{(1\times p)}{x_t'}\underset{(p\times 1)}{\theta_0} + \underset{(1\times 1)}{u_t}, \\ \underset{(p\times 1)}{x_t} &= \underset{(p\times q)}{\pi_0}\underset{(q\times 1)}{z_t} + \underset{(p\times 1)}{e_t}, \quad t = 1, 2, \ldots, T \end{aligned} \tag{3.3.1}$$

where

$$\begin{pmatrix} u_t \\ e_t \end{pmatrix} \overset{iid}{\sim} N\left(\underset{(p+1)\times 1}{0}, \underset{(p+1)\times(p+1)}{\Sigma}\right),$$

and

$$\underset{(p+1)\times(p+1)}{\Sigma} = \begin{pmatrix} \underset{(1\times 1)}{\sigma_u^2} & \underset{(1\times p)}{\Sigma_{ue}} \\ \underset{(p\times 1)}{\Sigma_{eu}} & \underset{(p\times p)}{\Sigma_{ee}} \end{pmatrix}.$$

Proposition 3.3.1 *Let $v_t = (x_t', z_t')'$ be stationary and $\psi_0 = (\theta_0', \phi_0')'$. Then in a linear simultaneous equation model described by equation (3.3.1) under the assumption of identically and independently jointly normally distributed structural and reduced form errors, the population long run canonical correlations between the score vector $s_{\theta,t}(\psi_0)$ with respect to the parameter on the endogenous regressor x_t and the population moment vector $f(v_t, \theta_0)$ are proportional to the contemporaneous canonical correlations between the endogenous regressor vector x_t and the instrument vector z_t, where the constant of proportionality is given by $\sqrt{\frac{\sigma_\xi^2}{\Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu}+\sigma_\xi^2}}$.*

The proof is presented in the appendix.

3.4 Analytical Connection among the Three Methods

In this section, we investigate the analytical connection among the three criteria under the proposed simple linear IV model:

$$\underset{(1\times1)}{y_t} = \underset{(1\times1)}{x_t}\underset{(1\times1)}{\theta_0} + \underset{(1\times1)}{u_t},$$

$$\underset{(1\times1)}{x_t} = \underset{(1\times q)}{z_t{}'}\underset{(q\times1)}{\pi_0} + \underset{(1\times1)}{e_t}, \quad t = 1, 2, \ldots, T.$$

In matrix notation the model becomes

$$\underset{(T\times1)}{y} = \underset{(T\times1)}{x}\underset{(1\times1)}{\theta_0} + \underset{(T\times1)}{u},$$

$$\underset{(T\times1)}{x} = \underset{(T\times q)}{z}\underset{(q\times1)}{\pi_0} + \underset{(T\times1)}{e}.$$

We know from (3.2.7)) that for this model the CCIC is given by

$$CCIC(c) = T\ ln[1 - r_T^2(c)] + (|c| - 1)\ lnT.$$

We also know from equation (3.2.13) that for this model the RMSC given by

$$\begin{aligned} RMSC(c) &= ln\left(|\hat{\sigma}_u^2[\sum_{t=1}^{T} x_t z_t(c)'\{\sum_{t=1}^{T} z_t(c)z_t(c)'\}^{-1}\sum_{t=1}^{T} z_t(c)x_t(c)']^{-1}|\right) + \frac{(|c|-1)}{\sqrt{T}}\ ln\sqrt{T} \\ &= ln\left(\frac{\hat{\sigma}_u^2}{\frac{x'x}{T}r_T^2(c)}\right) + \frac{(|c|-1)}{\sqrt{T}}\ ln\sqrt{T} \\ &= ln\left(\frac{\hat{\sigma}_u^2}{\hat{\mu}_x^2 r_T^2(c)}\right) + \frac{(|c|-1)}{\sqrt{T}}\ ln\sqrt{T}, \quad [\text{where}\ \ \hat{\mu}_x^2 = \frac{x'x}{T}], \end{aligned}$$

which shows that RMSC is a function of the squared canonical correlation which in this case is the squared multiple correlation, $r_T^2(c)$.

To explore the link of Donald and Newey's (2001) approximate MSE criterion with the canonical correlations, it is convenient to first obtain in their notation the following form of their MSE. Substituting for $S(K)$ from [p. 1167] into equation (2) on [p. 1166] of Donald and Newey (2001), the N times population approximate MSE:

$$
\begin{aligned}
E[N(\hat{\delta}-\delta_0)(\hat{\delta}-\delta_0)'] &= \sigma_\epsilon^2 H^{-1} + H^{-1}[\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{K^2}{N} + \sigma_\epsilon^2\frac{f'(I-P)f}{N}]H^{-1} + o_p(1) \\
&= \sigma_\epsilon^2 H^{-1} + H^{-1}[\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{K^2}{N}]H^{-1} + H^{-1}[\sigma_\epsilon^2(\frac{f'f}{N} - \frac{f'Pf}{N})]H^{-1} + o_p(1) \\
&= \sigma_\epsilon^2 H^{-1} + H^{-1}[\sigma_{u\epsilon}\sigma_{u\epsilon}'\frac{K^2}{N}]H^{-1} + H^{-1}[\sigma_\epsilon^2(H - \frac{f'Pf}{N})]H^{-1} + o_p(1) \\
&= \frac{\sigma_\epsilon^2}{H} + \frac{\sigma_{u\epsilon}^2}{H^2}\frac{K^2}{N} + \frac{\sigma_\epsilon^2}{H} - \frac{\sigma_\epsilon^2}{H^2}\frac{f'Pf}{N} + o_p(1).
\end{aligned}
$$

Therefore, the population approximate MSE in our model is given by

$$
\frac{\sigma_u^2}{H} + \frac{\sigma_{eu}^2}{H^2}\frac{|c|^2}{T} + o_p(1)
$$

and hence the estimated approximate MSE is

$$
\begin{aligned}
MSE(c) &= \frac{\hat{\sigma}_u^2}{\hat{H}} + \frac{\hat{\sigma}_{eu}^2}{\hat{H}^2}\frac{|c|^2}{T} \\
&= \frac{\hat{\sigma}_u^2}{\frac{\hat{f}'\hat{f}}{T}} + \frac{\hat{\sigma}_{eu}^2}{(\frac{\hat{f}'\hat{f}}{T})^2}\frac{|c|^2}{T} \\
&= \frac{T\hat{\sigma}_u^2}{\hat{\pi}'z'z\hat{\pi}} + \frac{\hat{\sigma}_{eu}^2}{(\hat{\pi}'z'z\hat{\pi})^2}|c|^2 T \\
&= \frac{T\hat{\sigma}_u^2}{[(z'z)^{-1}z'x]'z'z[(z'z)^{-1}z'x]} + \frac{\hat{\sigma}_{eu}^2}{([(z'z)^{-1}z'x]'z'z[(z'z)^{-1}z'x])^2}|c|^2 T \\
&= \frac{T\hat{\sigma}_u^2}{x'z(z'z)^{-1}z'x} + \frac{\hat{\sigma}_{eu}^2}{[x'z'(z'z)^{-1}z'x]^2}|c|^2 T \\
&= \frac{\hat{\sigma}_u^2}{\frac{x'x}{T}r_T^2(c)} + \frac{\hat{\sigma}_{eu}^2}{[\frac{x'x}{T}r_T^2(c)]^2}|c|^2 \\
&= \frac{\hat{\sigma}_u^2}{\hat{\mu}_x^2 r_T^2(c)} + \frac{\hat{\sigma}_{eu}^2}{[\hat{\mu}_x^2 r_T^2(c)]^2}|c|^2,
\end{aligned}
$$

thereby showing that Donald and Newey's approximate MSE is also a function of canonical correlations between regressors and instruments.

Thus, we see all three criteria are analytically connected through their dependence on canonical correlations between regressors and instruments.

3.5 Simulation

3.5.1 Design

We follow Donald and Newey (2001) and Hall and Peixe (2003) to design a Monte-Carlo experiment using the simple linear IV model used in the previous sections:

$$\begin{aligned} y_t &= x_t\theta_0 + u_t, \\ x_t &= z_t{}'\pi_0 + e_t, \quad t = 1, 2, \dots, T. \end{aligned}$$

For a fixed value of $\theta_0 = 0.1$ and for different specifications of π_0, random samples are generated under the assumption that $v_t \sim N(0, \Sigma_v)$ where $v_t = [u_t, e_t, z_t']'$. The main diagonal of Σ_v are all set to unity; the only non-zero off diagonal elements are $cov(u_t, e_t) = \sigma_{ue}$, that is, $\Sigma_v(1,2)$ and $\Sigma_v(2,1)$. Hahn and Hausman (2002) show that this specification implies a theoretical first stage R-squared of the form

$$R_f^2 = \frac{\pi_0'\pi_0}{\pi_0'\pi_0 + 1}. \tag{3.5.1}$$

We consider three models that differ in the specification of the π_0 vector. Models 1 and 2 below are the models used by Donald and Newey (2001). Within the three models, each experiment consists of a specification of $(T, R_f^2, \sigma_{ue}, q_{max})$. In Model 1, for a given value of R_f^2, the ith element of π_0 is given by

$$\pi_{0,i}^{(1)} = c(q_{max})\left(1 - \frac{i}{q_{max}+1}\right)^4 \quad \text{for } i = 1, \dots, q_{max}, \tag{3.5.2}$$

where the constant $c(q_{max})$ is chosen so that $\pi_{0,i}^{(1)\prime}\pi_{0,i}^{(1)} = \frac{R_f^2}{(1-R_f^2)}$, while in Model 2 it is given by

$$\pi_{0,i}^{(2)} = \pi_0^{(2)} = \sqrt{\frac{R_f^2}{q_{max}(1-R_f^2)}} \quad \text{for } i = 1, \dots, q_{max}. \tag{3.5.3}$$

For each model, experiments are conducted with the following specifications:

$$T \in \{100, 500\}, \quad R_f^2 \in \{0.1, 0.5, \}, \quad \sigma_{ue} \in \{0.1, 0.5, 0.9\}, \quad \text{and} \quad q_{max} \in \{20, 30\}.$$

In Model 1, one has apriori knowledge that instruments are from the best quality to the worst, as given by (3.5.2). But a selection criterion does not say in which order the instruments should be selected, and so, one could be curious to know what might happen if selection proceeds from the worst quality instrument to the best instead of the other way around. Therefore, we follow two selection strategies, namely, selection strategy 1 and selection strategy 2. Under strategy 1, the selection vectors are given by $c_i = [\imath_i', 0_{q_{max}-i}']'$ and $\imath_i$ is a $i \times 1$ vector of ones and $0_{q_{max}-i}$ is a $(q_{max} - i) \times 1$ vector of zeros and thus selection proceeds from the best quality instrument to the worst. Under strategy 2, they are given by $c_i = [0_{i-1}', \imath_{q_{max}-i+1}']'$ where 0_{i-1} is a $(i - 1) \times 1$ vector of zeros and $\imath_{q_{max}-i+1}$ is a $(q_{max} - i + 1) \times 1$ vector of ones and thus selection proceeds from the worst quality instrument to the best. In Model 2, where no apriori information exists about the quality of the instruments, we use selection strategy 1 and thus selection proceeds from the first element in the candidate set to the last.

We note that in the above three cases, the instruments are selected in an increasing sequence of one, two, three, and so on, up to a total of q_{max}. Thus there are q_{max} selection vectors. Each time, the selection vector is augmented by retaining the previously included instruments and adding to the vector the next instrument from the candidate set. But it is also of interest to explore the consequences of including different combinations of redundant and non-redundant instruments. To investigate such consequences, we consider a third model by modifying Model 1 along the lines of Hall and Peixe (2003). Because the total number of possible combinations of instruments quickly grows large with the value of q_{max}, we set a small value to q_{max}, equal to 8. Thus, π_0 and z_t are 8×1 vectors. Even this small value of q_{max} generates 255 possible combinations of instruments. For a given value of R_f^2, the ith element of π_0 is generated by the scheme

$$\begin{aligned} \pi_{0,i} &= c(q_{max}) \left(1 - \frac{i}{q_{max}+1}\right)^4, \; i = 1, 2; \\ &= 0 \quad \text{for } \; i = 3, \ldots, q_{max}, \end{aligned}$$

where the constant $c(q_{max})$ is chosen so that $\pi_0'\pi_0 = \frac{R_f^2}{(1-R_f^2)}$. Each experiment consists of a specification of $(T, R_f^2, \sigma_{ue}, q_{max})$ from the following sets: $T \in \{100, 500\}$, $R_f^2 \in \{0.1, 0.5, \}$, $\sigma_{ue} \in \{0.1, 0.5, 0.9\}$, and $q_{max} \in \{8\}$. The selection vectors c_k are comprised of combinations of i ones and $8-i$ zeros, $k = 1, \ldots, 255$; $i = 1, \ldots, 8$. As in Hall and Peixe (2003), $(z_{t,3}, z_{t,4}, \ldots, z_{t,8})$ are redundant given $(z_{t,1}, z_{t,2})$ and so $(z_{t,1}, z_{t,2})$ constitute the "relevant" instruments. For $i > 2$, $z_t(c_k)$ contains $i-2$ redundant instruments.

Because the total number of possible selection vectors is large, following Hall and Peixe (2003) we group these possibilities into six cases: $1R, 2R, 1R/I, 2R/I^*, I$ and *all*, where $1R$ denotes the cases in which $c = (a, 0_6')'$ for $a \in \{(1,0), (0,1)\}$, implying that that the selection vector consists of only one of the relevant instruments; $2R$ denotes the case in which $c = (1, 1, 0_6')'$, indicating that the selection vector consists of only both relevant instruments; $1R/I$ denotes the cases in which $c = (a', b')'$ for a given above and $b \neq 0_6$, meaning that the selection vector consists of one relevant instrument and at least one redundant instrument; $2R/I^*$ denotes the cases in which $c = (1, 1, d')'$ and $d \neq 0_6$ *or* i_6, that is, the selection vector consists of both relevant and at least one but not all six redundant instruments; I denotes the cases in which $c = (0, 0, b')'$ for b given above, implying that the selection vector consists of only redundant instruments; and finally, *all* denotes the case in which $c = \imath_8'$, indicating that the selection vector contains all eight instruments, the two relevant instruments as well as the six redundant instruments.

To assess the relative performance of the CCIC, the RMSC and the approximate MSE, we compare median bias of the post selection estimator and coverage probability of 90% confidence intervals.

3.5.2 Results

The results are presented in Tables 3.1 through 3.14. All results correspond to 10,000 replications. First we look at the performance of the three criteria in terms of median bias.

We begin with Model 1 under strategy 1. Recall that this is a case where the practitioner has prior knowledge that all instruments in the candidate set are relevant but the relevance is of a declining one. The order of selection of instruments is from the best quality to the worst. The results for this case are shown in Tables 3.1 through 3.4. We find that for both sizes of the candidate set, that is, $q_{max} = 20$ and $q_{max} = 30$, with $T = 100$, when the first stage R-square, $R_f^2 = 0.1$ [Tables 3.1 and 3.3], for all values of endogeneity, σ_{ue}, a unique ranking emerges: CCIC ranks higher than approximate MSE which, in turn, ranks higher than RMSC. However, when $R_f^2 = 0.5$ [Tables 3.2 and 3.4], the median bias becomes very similar under each criterion and thus no unique ranking becomes possible. Also, with $T = 500$, no unique ranking emerges under either value of R_f^2.

Next, we look at Model 1 under strategy 2. The results of this case are presented in Tables 3.5 through 3.8. Recall that here the order of selection of instruments is reversed, that is, selection proceeds from the worst quality to the best. When $T = 100$, we find that a unique ranking emerges: approximate MSE performs better than CCIC which, in turn, performs better than RMSC, for all values of q_{max}, R_f^2 and σ_{ue}, except for the case where $q_{max} = 30, R_f^2 = 0.1, \sigma_{ue} = 0.1$ in Table 3.7. When sample size increases to $T = 500$, at $R_f^2 = 0.1$ a unique ranking still emerges and approximate MSE still does the best but CCIC and RMSC switch ranks. However, at $R_f^2 = 0.5$, no unique ranking of the three criteria becomes possible.

The results of Model 2, where the instruments are of equal relevance, are shown in Tables 3.9 through 3.12. Tables 3.9 and 3.11 indicate that at $R_f^2 = 0.1$, a unique ranking emerges for both sizes of the candidate set q_{max}, for both sample sizes T, and for all values of the endogeneity parameter σ_{ue}. The $CCIC$ ranks higher than MSE which, in turn, ranks higher than $RMSC$. Tables 3.10 and 3.12 reveal that at $R_f^2 = 0.5$, similar result holds. $CCIC$ tends to outrank MSE which, in turn, outranks $RMSC$.

In Model 3, where the candidate set consists of a combination of both relevant and irrelevant instruments, Table 3.13 shows that at $R_f^2 = 0.1$, a unique ranking of the three criteria emerges: CCIC ranks higher than RMSC which, in turn, ranks higher than approximate MSE. Table 3.14 shows that when the first stage R-square increases to $R_f^2 = 0.5$, a unique ranking of CCIC and RMSC no longer holds as they perform almost equally well, but the canonical correlation based criteria still rank higher than the approximate MSE.

Finally, we look at the coverage rate. Inspection of the results in Tables 3.1 through 3.14 reveals that no unique ranking of the three criteria in terms of coverage rate becomes possible except in only two cases where $R_f^2 = 0.5$ and $T = 500$. In Model 1 under strategy 1 where the selection is from the best quality to the worst [Table 3.2], MSE outranks $RMSC$ which, in turn, outranks $CCIC$. In Model 2 where instruments are equally relevant [Table 3.12], $RMSC$ ranks higher than MSE which, in turn, ranks higher than $CCIC$.

3.6 Conclusions

This chapter set out by raising three questions relating to three recent methods of instrument selection in econometrics. The first question concerns the analytical connection among them. Because CCIC and RMSC are based on canonical correlations, it is no wonder that they would be a function of canonical correlations. However, what was not clear is whether approximate MSE could also be a function of canonical correlations. Our analytical investigation shows that it is.

The second question under investigation was whether a unique ranking of the three methods is possible in terms of the finite sample behavior of the post selection estimator. Simulation results reveal that the answer to this question is a conditional one, in the sense that, a unique ranking emerges under certain parameter configurations while it does not under others.

The third question raised was what guidance we could provide a practitioner as to which of these three methods one should use in any practical application of interest. In light of the nature of the second answer, the answer to the third question is obviously a conditional one. For example, a practitioner would be better off implementing CCIC if the practitioner is aware that all the instruments in the candidate set are relevant but the relevance is of a declining one and the first stage R-square is small, or that all instruments are equally relevant and the first stage R-square is small, or if the practitioner is not sure about the composition of the candidate set and it is possible that the candidate set could contain a mix of both relevant and irrelevant instruments. On the other hand, if one is aware that all instruments are relevant but the relevance is of an increasing one and the the first stage R-square is small, one would be better off implementing approximate MSE. In other situations one would do about equally well using any of the three criteria. Finally, from our study, no clear guidance emerges in terms of coverage rate.

Mathematical Appendix

Proof of Proposition 3.3.1

The log-likelihood function implied by the model is

$$\mathcal{L}(\psi_0) = -\frac{Tp}{2}\ln(2\pi) + \frac{T}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\sum_{t=1}^{T}\{(y_t - x_t'\theta_0) \quad (x_t - \pi_0 z_t)'\}\Sigma^{-1}\{ \begin{matrix} (y_t - x_t'\theta_0) \\ (x_t - \pi_0 z_t) \end{matrix} \}$$

where

$$\begin{aligned}
&\left[(y_t - x_t'\theta_0) \quad (x_t - \pi_0 z_t)'\right]\Sigma^{-1}\begin{bmatrix} (y_t - x_t'\theta_0) \\ (x_t - \pi_0 z_t) \end{bmatrix} \\
= \ & \left[(y_t - x_t'\theta_0) \quad (x_t - \pi_0 z_t)'\right]\begin{bmatrix} (\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1} & -(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\Sigma_{ue}\Sigma_{ee}^{-1} \\ -\Sigma_{ee}^{-1}\Sigma_{eu}(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1} & (\Sigma_{ee} - \Sigma_{eu}(\sigma_u^2)^{-1}\Sigma_{ue})^{-1} \end{bmatrix} \\
& \times \begin{bmatrix} (y_t - x_t'\theta_0) \\ (x_t - \pi_0 z_t) \end{bmatrix} \\
= \ & (y_t - x_t'\theta_0)(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}(y_t - x_t'\theta_0) \\
& -(x_t - \pi_0 z_t)'\Sigma_{ee}^{-1}\Sigma_{eu}(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}(y_t - x_t'\theta_0) \\
& -(y_t - x_t'\theta_0)(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\Sigma_{ue}\Sigma_{ee}^{-1}(x_t - \pi_0 z_t) \\
& +(x_t - \pi_0 z_t)'(\Sigma_{ee} - \Sigma_{eu}(\sigma_u^2)^{-1}\Sigma_{ue})^{-1}(x_t - \pi_0 z_t) \\
= \ & (y_t - x_t'\theta_0)(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}(y_t - x_t'\theta_0) \\
& -2(y_t - x_t'\theta_0)(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\Sigma_{ue}\Sigma_{ee}^{-1}(x_t - \pi_0 z_t) \\
& +(x_t - \pi_0 z_t)'(\Sigma_{ee} - \Sigma_{eu}(\sigma_u^2)^{-1}\Sigma_{ue})^{-1}(x_t - \pi_0 z_t)
\end{aligned}$$

So, the score with respect to the parameter θ on the endogenous regressor x_t is given by

$$\begin{aligned}
\frac{\partial \mathcal{L}(\psi_0)}{\partial \theta} &= \sum_{t=1}^{T} s_{\theta,t}(\psi_0) \\
&= \sum_{t=1}^{T} x_t(y_t - x_t'\theta_0)(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1} - \sum_{t=1}^{T} x_t(\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\Sigma_{ue}\Sigma_{ee}^{-1}(x_t - \pi_0 z_t) \\
&= (\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\sum_{t=1}^{T}[x_t(y_t - x_t'\theta_0) - x_t\Sigma_{ue}\Sigma_{ee}^{-1}(x_t - \pi_0 z_t)] \\
&= (\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\sum_{t=1}^{T}[x_t u_t(\theta_0) - x_t\Sigma_{ue}\Sigma_{ee}^{-1}e_t] \\
&= (\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu})^{-1}\sum_{t=1}^{T} x_t[u_t(\theta_0) - \Sigma_{ue}\Sigma_{ee}^{-1}e_t] \\
&= \delta^{-1}\sum_{t=1}^{T} x_t[u_t(\theta_0) - \Sigma_{ue}\Sigma_{ee}^{-1}e_t], \quad \text{where } \delta = (\sigma_u^2 - \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu}).
\end{aligned}$$

Note that

$$\begin{aligned}
u_t &= \Sigma_{ue}\Sigma_{ee}^{-1}e_t + \xi_t \\
&= E(u_t|e_t) + \xi_t.
\end{aligned}$$

Because $E(u_t|e_t) = \Sigma_{ue}\Sigma_{ee}^{-1}e_t$ is a linear projection of u_t on e_t, the prediction error, $u_t - E(u_t|e_t) = \xi_t$, is uncorrelated with the conditioning variable $e_t : E\xi_t e_t' = 0$.

The population long run canonical correlations between the score $s_{\theta,t}(\psi_0)$ and the population moment $f(v_t, \theta_0)$ are solutions to the determinantal equation $|V_\theta^{-1} - \rho^2 I_\theta| = 0$ where

$$\begin{aligned}
V_\theta^{-1} &= \lim_{T\to\infty} E[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{\theta,t}(\psi_0)][\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0)]'[\lim_{T\to\infty} E(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0))(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0))']^{-1} \\
&\quad \times \lim_{T\to\infty} E[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0)][\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{\theta,t}(\psi_0)]' \\
&= P_1 P_2^{-1} P_3;
\end{aligned}$$

$$P_1 = \lim_{T\to\infty} E[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{\theta,t}(\psi_0)][\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0)]',$$

$$P_2 = \lim_{T\to\infty} E(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0))(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0))',$$

$$P_3 = \lim_{T\to\infty} E[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0)][\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{\theta,t}(\psi_0)]',$$

and

$$I_\theta = \lim_{T\to\infty} E(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{\theta,t}(\psi_0))(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{\theta,t}(\psi_0))'.$$

Now

$$\begin{aligned}
P_1 &= \lim_{T\to\infty} E[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s_{\theta,t}(\psi_0)][\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f(v_t, \theta_0)]' \\
&= \lim_{T\to\infty} E[\frac{1}{\sqrt{T}} \frac{1}{\delta} \sum_{t=1}^{T} x_t \xi_t][\frac{1}{\sqrt{T}} \sum_{t=1}^{T} z_t u_t]' \\
&= \frac{1}{\delta} \lim_{T\to\infty} E[\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} x_t \xi_t u_s' z_s'] \\
&= \frac{1}{\delta} \lim_{T\to\infty} E[\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} x_t \xi_t (\Sigma_{ue} \Sigma_{ee}^{-1} e_s + \xi_s)' z_s'] \\
&= \frac{1}{\delta} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[x_t \xi_t (e_s' \Sigma_{ee}^{-1} \Sigma_{eu} + \xi_s') z_s'] \\
&= \frac{1}{\delta} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[x_t \xi_t e_s' \Sigma_{ee}^{-1} \Sigma_{eu} z_s' + x_t \xi_t \xi_s' z_s'] \\
&= \frac{1}{\delta} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[x_t \xi_t e_s' \Sigma_{ee}^{-1} \Sigma_{eu} z_s'] \\
&\quad + \frac{1}{\delta} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[x_t \xi_t \xi_s' z_s'] \\
&= \frac{1}{\delta} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[(\pi_0 z_t + e_t) \xi_t e_s' \Sigma_{ee}^{-1} \Sigma_{eu} z_s'] \\
&\quad + \frac{1}{\delta} \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[x_t \xi_t \xi_s' z_s']
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\pi_0 z_t\xi_t e_s'\Sigma_{ee}^{-1}\Sigma_{eu}z_s'] + \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[e_t\xi_t e_s'\Sigma_{ee}^{-1}\Sigma_{eu}z_s'] \\
&\quad + \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[x_t\xi_t\xi_s'z_s'] \\
&= \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\xi_t e_s'\Sigma_{ee}^{-1}\Sigma_{eu}\pi_0 z_t z_s'] + \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[e_t\xi_t e_s'\Sigma_{ee}^{-1}\Sigma_{eu}z_s'] \\
&\quad + \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[x_t\xi_t\xi_s'z_s'] \\
&= \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[E(\xi_t e_s'\Sigma_{ee}^{-1}\Sigma_{eu}\pi_0 z_t z_s')|z_s, z_t] + \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[e_t\xi_t e_s'\Sigma_{ee}^{-1}\Sigma_{eu}z_s'] \\
&\quad + \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[x_t\xi_t\xi_s'z_s'] \\
&= \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[x_t\xi_t\xi_s'z_s'] \quad \text{since } E\xi_t e_s' = 0 \text{ for all } s \text{ and } t \\
&= \frac{1}{\delta}\lim_{T\to\infty}\frac{1}{T}(T\sigma_\xi^2\Sigma_{xz}) \quad \text{since } E\xi_t\xi_s' = 0 \text{ for all } s \neq t \\
&= \frac{\sigma_\xi^2}{\delta}\Sigma_{xz},
\end{aligned}
$$

where $\Sigma_{xz} = E(x_t z_t')$.

$$
\begin{aligned}
P_2 &= \lim_{T\to\infty} E(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0))(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f(v_t,\theta_0))' \\
&= \lim_{T\to\infty}\frac{1}{T}E[(\sum_{t=1}^{T} z_t u_t)(\sum_{t=1}^{T} z_t u_t)'] \\
&= \lim_{T\to\infty}\frac{1}{T}E\sum_{t=1}^{T}\sum_{s=1}^{T} z_t u_t u_s' z_s' \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E(z_t u_t u_s' z_s') \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E(u_t^2)E(z_t z_t') \quad [z_t \text{ is independent of } u_t, \text{ and } E(u_t u_s') = 0 \text{ for } s \neq t]
\end{aligned}
$$

Again,

$$\begin{aligned}
E(u_t^2) &= E[E(u_t|e_t) + \xi_t]^2 \\
&= E[\Sigma_{ue}\Sigma_{ee}^{-1}e_t + \xi_t]^2 \\
&= \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{ee}\Sigma_{ee}^{-1}\Sigma_{eu} + 2\Sigma_{ue}\Sigma_{ee}^{-1}E(e_t\xi_t) + E(\xi_t^2) \\
&= \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2 \quad [\text{since } E(e_t\xi_t) = 0]
\end{aligned}$$

Therefore,

$$\begin{aligned}
P_2 &= \lim_{T\to\infty} \frac{1}{T} T(\Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2)\Sigma_{zz} \\
&= (\Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2)\Sigma_{zz}
\end{aligned}$$

where $\Sigma_{zz} = E(z_t z_t')$.

$$\begin{aligned}
P_3 &= (P_1)' \\
&= (\frac{\sigma_\xi^2}{\delta}\Sigma_{xz})' \\
&= \frac{\sigma_\xi^2}{\delta}\Sigma_{zx}
\end{aligned}$$

where $\Sigma_{zx} = E(z_t x_t')$.

Finally,

$$\begin{aligned}
I_\theta &= \lim_{T\to\infty} E(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{\theta,t}(\psi_0))(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} s_{\theta,t}(\psi_0))' \\
&= \lim_{T\to\infty} E(\frac{1}{\sqrt{T}}\frac{1}{\delta}\sum_{t=1}^{T} x_t\xi_t)(\frac{1}{\sqrt{T}}\frac{1}{\delta}\sum_{t=1}^{T} x_t\xi_t)' \\
&= \frac{1}{\delta^2}\lim_{T\to\infty}\sum_{t=1}^{T}\sum_{s=1}^{T} E(x_t\xi_t\xi_s' x_s') \\
&= \frac{1}{\delta^2}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E(\xi_t^2)E(x_t x_t') \quad [E(\xi_t\xi_s') = 0 \text{ for } t \neq s \text{ and } \xi_t \text{ is uncorrelated with } x_t] \\
&= \frac{1}{\delta^2}\lim_{T\to\infty}\frac{1}{T}T\sigma_\xi^2\Sigma_{xx} \\
&= \frac{\sigma_\xi^2}{\delta^2}\Sigma_{xx}
\end{aligned}$$

where $\Sigma_{xx} = E(x_t x_t')$.

Hence,

$$\begin{aligned} V_\theta^{-1} - \rho^2 I_\theta &= \frac{\sigma_\xi^2}{\delta}\Sigma_{xz}[(\Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2)\Sigma_{zz}]^{-1}\frac{\sigma_\xi^2}{\delta}\Sigma_{zx} - \rho^2\frac{\sigma_\xi^2}{\delta^2}\Sigma_{xx} \\ &= \frac{\sigma_\xi^2}{\delta^2}[\frac{\sigma_\xi^2}{\Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2}\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\Sigma_{xx}] \end{aligned}$$

Therefore,

$$|V_\theta^{-1} - \rho^2 I_\theta| = 0$$

$$\Rightarrow |\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \rho^2\frac{1}{k}\Sigma_{xx}| = 0$$

where

$$k = \frac{\sigma_\xi^2}{\Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2}.$$

Thus we note that while the contemporaneous canonical correlations between the regressor vector x_t and the instrument vector z_t are given by the solutions of the determinantal equation

$$|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - r^2\Sigma_{xx}| = 0,$$

the long run canonical correlations between the score vector $s_{\theta,t}(\psi_0)$ and the population moment vector $f(v_t, \theta_0)$ are given by the solutions of the determinantal equation

$$|\Sigma_{xz}\Sigma_{zz}^{-1}\Sigma_{zx} - \frac{\rho^2}{k}\Sigma_{xx}| = 0.$$

Hence the proof is complete. ■

Table 3.1: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c).*

Model 1 under strategy 1: Best to Worst

q_{max} = *20*	*T=100*			*T=500*		
	R_f^2 = *0.1;* σ_{ue} = *0.1.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.032	0.021	0.041	0.012	0.008	0.012
90% Nom. Cov Rate	0.922	0.957	0.947	0.898	0.913	0.924
	R_f^2 = *0.1;* σ_{ue} = *0.5.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.163	0.113	0.188	0.056	0.051	0.061
Cov Rate	0.756	0.872	0.818	0.844	0.874	0.868
	R_f^2 = *0.1;* σ_{ue} = *0.9.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.229	0.191	0.341	0.079	0.087	0.105
Cov Rate	0.651	0.724	0.470	0.787	0.787	0.762

Notes: Nominal Coverage Rate = 90%; *Best to Worst* indicates that the order of selection of instruments is from the best quality to the worst.

Table 3.2: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 1 under strategy 1: Best to Worst

$q_{max} = 20$		*T=100*			*T=500*	
	$R_f^2 = 0.5;\ \sigma_{ue} = 0.1.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.008	0.007	0.006	0.002	0.002	0.003
Cov Rate	0.898	0.904	0.914	0.901	0.898	0.900
	$R_f^2 = 0.5;\ \sigma_{ue} = 0.5.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.033	0.031	0.034	0.010	0.010	0.010
Cov Rate	0.862	0.875	0.893	0.898	0.893	0.896
	$R_f^2 = 0.5;\ \sigma_{ue} = 0.9.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.053	0.059	0.059	0.013	0.015	0.015
Cov Rate	0.809	0.798	0.831	0.883	0.873	0.881

Notes: Nominal Coverage Rate = 90%; *Best to Worst* indicates that the order of selection of instruments is from the best quality to the worst.

Table 3.3: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c).*

Model 1 under strategy 1: Best to Worst

q_{max} = *30*	*T=100*			*T=500*		
			R_f^2 = *0.1;* σ_{ue} = *0.1.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.042	0.029	0.046	0.016	0.010	0.013
Cov Rate	0.921	0.965	0.951	0.890	0.911	0.924
			R_f^2 = *0.1;* σ_{ue} = *0.5.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.191	0.133	0.213	0.072	0.065	0.077
Cov Rate	0.723	0.880	0.798	0.817	0.859	0.851
			R_f^2 = *0.1;* σ_{ue} = *0.9.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.285	0.235	0.384	0.104	0.117	0.132
Cov Rate	0.595	0.700	0.401	0.749	0.735	0.701

Notes: Nominal Coverage Rate = 90%; *Best to Worst* indicates that the order of selection of instruments is from the best quality to the worst.

Table 3.4: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 1 under strategy 1: Best to Worst

q_{max} = *30*	*T=100*			*T=500*		
	R_f^2 = *0.5;* σ_{ue} = *0.1.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.011	0.009	0.009	0.003	0.002	0.003
Cov Rate	0.897	0.904	0.921	0.903	0.901	0.908
	R_f^2 = *0.5;* σ_{ue} = *0.5.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.048	0.045	0.046	0.012	0.012	0.013
Cov Rate	0.843	0.858	0.885	0.889	0.885	0.890
	R_f^2 = *0.5;* σ_{ue} = *0.9.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.067	0.078	0.075	0.019	0.022	0.021
Cov Rate	0.780	0.753	0.810	0.868	0.844	0.862

Notes: Nominal Coverage Rate = 90%; *Best to Worst* indicates that the order of selection of instruments is from the best quality to the worst.

Table 3.5: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 1 under strategy 2: Worst to Best

q_{max} = *20*	*T=100*			*T=500*		
			R^2_f = *0.1;* σ_{ue} = *0.1.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.022	0.098	0.099	0.024	0.086	0.034
Cov Rate	0.944	0.998	0.967	0.902	0.997	0.915
			R^2_f = *0.1;* σ_{ue} = *0.5.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.091	0.429	0.449	0.110	0.461	0.201
Cov Rate	0.804	0.952	0.669	0.648	0.960	0.564
			R^2_f = *0.1;* σ_{ue} = *0.9.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.057	0.802	0.807	0.001	0.799	0.741
Cov Rate	0.818	0.725	0.046	0.814	0.740	0.018

Notes: Nominal Coverage Rate = 90%; *Worst to Best* indicates that the order of selection of instruments is from the worst quality to the best.

Table 3.6: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c).*

Model 1 under strategy 2: Worst to Best

q_{max} = *20*	*T=100*			*T=500*		
R_f^2 = *0.5;* σ_{ue} = *0.1.*						
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.015	0.020	0.047	0.003	0.003	0.088
Cov Rate	0.895	0.939	0.969	0.903	0.899	0.966
R_f^2 = *0.5;* σ_{ue} = *0.5.*						
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.078	0.102	0.235	0.019	0.019	0.019
Cov Rate	0.746	0.822	0.811	0.876	0.867	0.867
R_f^2 = *0.5;* σ_{ue} = *0.9.*						
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.011	0.188	0.439	0.033	0.033	0.033
Cov Rate	0.821	0.550	0.366	0.812	0.792	0.788

Notes: Nominal Coverage Rate = 90%; *Worst to Best* indicates that the order of selection of instruments is from the worst quality to the best.

Table 3.7: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c).*

Model 1 under strategy 2: Worst to Best

q_{max} = *30*	*T=100*			*T=500*		
	R^2_f = *0.1;* σ_{ue} = *0.1.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.017	0.100	0.088	0.034	0.101	0.089
Cov Rate	0.954	0.995	0.966	0.896	0.998	0.958
	R^2_f = *0.1;* σ_{ue} = *0.5.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.084	0.432	0.449	0.053	0.463	0.440
Cov Rate	0.856	0.949	0.665	0.727	0.964	0.524
	R^2_f = *0.1;* σ_{ue} = *0.9.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.112	0.804	0.809	0.007	0.814	0.804
Cov Rate	0.812	0.716	0.043	0.794	0.738	0.005

Notes: Nominal Coverage Rate = 90%; *Worst to Best* indicates that the order of selection of instruments is from the worst quality to the best.

Table 3.8: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c).*

Model 1 under strategy 2: Worst to Best

q_{max} = *30*	*T=100*			*T=500*		
			R_f^2 = *0.5;* σ_{ue} = *0.1.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.024	0.042	0.052	0.005	0.005	0.009
Cov Rate	0.894	0.991	0.971	0.906	0.896	0.913
			R_f^2 = *0.5;* σ_{ue} = *0.5.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.093	0.216	0.255	0.027	0.027	0.048
Cov Rate	0.657	0.954	0.817	0.842	0.826	0.782
			R_f^2 = *0.5;* σ_{ue} = *0.9.*			
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.007	0.414	0.450	0.005	0.050	0.293
Cov Rate	0.840	0.847	0.395	0.839	0.661	0.338

Notes: Nominal Coverage Rate = 90%; *Worst to Best* indicates that the order of selection of instruments is from the worst quality to the best.

Table 3.9: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 2: Equally Relevant Instruments

q_{max} = *20*	*T=100*			*T=500*		
	R_f^2 = *0.1;* σ_{ue} = *0.1.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.058	0.053	0.067	0.025	0.024	0.032
Cov Rate	0.945	0.988	0.954	0.901	0.968	0.935
	R_f^2 = *0.1;* σ_{ue} = *0.5.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.303	0.290	0.338	0.167	0.133	0.176
Cov Rate	0.676	0.925	0.708	0.645	0.873	0.721
	R_f^2 = *0.1;* σ_{ue} = *0.9.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.548	0.500	0.602	0.263	0.233	0.301
Cov Rate	0.489	0.719	0.144	0.628	0.693	0.322

Notes: Nominal Coverage Rate = 90%.

Table 3.10: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 2: Equally Relevant Instruments

q_{max} = *20*	*T=100*			*T=500*		
	R_f^2 = *0.5;* σ_{ue} = *0.1.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.015	0.022	0.023	0.004	0.004	0.008
Cov Rate	0.899	0.921	0.943	0.909	0.903	0.913
	R_f^2 = *0.5;* σ_{ue} = *0.5.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.092	0.094	0.101	0.018	0.018	0.038
Cov Rate	0.738	0.778	0.857	0.876	0.861	0.855
	R_f^2 = *0.5;* σ_{ue} = *0.9.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.168	0.170	0.183	0.046	0.033	0.063
Cov Rate	0.641	0.495	0.642	0.748	0.792	0.748

Notes: Nominal Coverage Rate = 90%.

Table 3.11: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 2: Equally Relevant Instruments

q_{max} = *30*	*T=100*			*T=500*		
	R_f^2 = *0.1;* σ_{ue} = *0.1.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.065	0.064	0.075	0.035	0.026	0.042
Cov Rate	0.949	0.992	0.964	0.895	0.980	0.938
	R_f^2 = *0.1;* σ_{ue} = *0.5.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.330	0.329	0.363	0.215	0.163	0.218
Cov Rate	0.695	0.937	0.696	0.601	0.907	0.693
	R_f^2 = *0.1;* σ_{ue} = *0.9.*					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.611	0.578	0.656	0.323	0.258	0.376
Cov Rate	0.486	0.717	0.105	0.585	0.744	0.227

Notes: Nominal Coverage Rate = 90%.

Table 3.12: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 2: Equally Relevant Instruments

q_{max} = *30*	*T=100*			*T=500*		
	$R_f^2 = 0.5;\ \sigma_{ue} = 0.1.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.023	0.020	0.026	0.006	0.006	0.011
Cov Rate	0.892	0.960	0.946	0.905	0.898	0.920
	$R_f^2 = 0.5;\ \sigma_{ue} = 0.5.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.125	0.112	0.128	0.031	0.029	0.049
Cov Rate	0.642	0.869	0.848	0.832	0.824	0.851
	$R_f^2 = 0.5;\ \sigma_{ue} = 0.9.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.205	0.199	0.223	0.068	0.050	0.085
Cov Rate	0.609	0.705	0.597	0.696	0.656	0.705

Notes: Nominal Coverage Rate = 90%.

Table 3.13: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 3: Combinations of Relevant & Irrelevant Instruments

$q_{max} = 8$	T=100			T=500		
	$R_f^2 = 0.1;\ \sigma_{ue} = 0.1.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.071	0.027	0.042	0.030	0.004	0.005
Cov Rate	0.984	0.935	0.947	0.958	0.913	0.919
	$R_f^2 = 0.1;\ \sigma_{ue} = 0.5.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.380	0.130	0.210	0.369	0.019	0.033
Cov Rate	0.740	0.841	0.804	0.733	0.901	0.897
	$R_f^2 = 0.1;\ \sigma_{ue} = 0.9.$					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.799	0.214	0.477	0.813	0.030	0.084
Cov Rate	0.294	0.645	0.253	0.377	0.875	0.766

Notes: Nominal Coverage Rate = 90%.

Table 3.14: Properties of *MSE(c)*, *CCIC(c)* & *RMSC(c)*.

Model 3: Combinations of Relevant & Irrelevant Instruments

q_{max} = 8	T=100			T=500		
	R_f^2 = 0.5; σ_{ue} = 0.1.					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.006	0.003	0.003	0.002	0.001	0.001
Cov Rate	0.911	0.903	0.906	0.905	0.902	0.903
	R_f^2 = 0.5; σ_{ue} = 0.5.					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.031	0.012	0.017	0.006	0.002	0.001
Cov Rate	0.884	0.901	0.903	0.904	0.908	0.908
	R_f^2 = 0.5; σ_{ue} = 0.9.					
	MSE	*CCIC*	*RMSC*	*MSE*	*CCIC*	*RMSC*
Med Bias	0.124	0.019	0.027	0.024	0.002	0.002
Cov Rate	0.720	0.877	0.876	0.860	0.900	0.898

Notes: Nominal Coverage Rate = 90%.

Chapter 4

Selection of Instruments when the Number of Instruments Tends to Infinity

4.1 Introduction

While Hall and Peixe (2003) and Hall and Inoue (2003) keep the candidate set of instruments z_t and the candidate set of moments $f(.)$, respectively, fixed, when sample size T increases, Donald and Newey (2001) allow the candidate set of instruments z_t to increase with the sample size T. In this chapter, our objective is to extend the statistical theory of the CCIC by considering the case in which the candidate set of instruments increases with the sample size. We first define canonical correlations where the regressor x_t is a vector and then extend the statistical theory of the CCIC by limiting our focus to the case where x_t is a scalar.

With this objective in mind we let $v_{t,q_T} = (x_t^{'}, z_{t,q_T}^{'})'$ where: (i) x_t is $p \times 1$, z_{t,q_T} is $q_T \times 1$ and $q_T \geq p$; (ii) for every T, $\{v_{t,q_T} \in \mathcal{V};\ t = 1, 2, \ldots, T;\ \mathcal{V} \subseteq \Re^{p+q_T}\}$ is a sequence of covariance stationary random vectors with mean vector zero and variance-covariance matrix

$$
\begin{aligned}
\underset{(p\times p)}{\Omega_v} &= E(\underset{(p+q_T)\times 1}{v_{t,q_T}} \underset{1\times(p+q_T)}{v_{t,q_T}{}'}) = \begin{pmatrix} Ex_t x_t^{'} & Ex_t z_{t,q_T}^{'} \\ Ez_{t,q_T} x_t{}' & Ez_{t,q_T} z_{t,q_T}^{'} \end{pmatrix} \\
&= \begin{bmatrix} \underset{(p\times p)}{\Omega_{xx}} & \underset{(p\times q_T)}{\Omega_{xz}} \\ \underset{(q_T\times p)}{\Omega_{zx}} & \underset{(q_T\times q_T)}{\Omega_{zz}} \end{bmatrix}
\end{aligned} \tag{4.1.1}
$$

where the second subscript q_T on the instrument vector z_t indicates that the size, q, of the instrument vector z_t, now depends on the sample size T. So $\{z_{t,q_T}\}$ constitutes a double array of random variables.

4.2 Definition and Existence of Canonical Correlations

We define the squared population canonical correlations between x_t and z_{t,q_T} denoted by $\{\rho^2_{i,q_T}; \mathrm{i} = 1, 2, \ldots, p\}$, where $\rho_{i,q_T} \geq \rho_{i+1,q_T}$ for $i = 1, 2, \ldots, p-1$, as solutions to the determinantal equation

$$
|\underset{(p\times q_T)}{\Omega_{xz}} \underset{(q_T\times q_T)}{\Omega_{zz}^{-1}} \underset{(q_T\times p)}{\Omega_{zx}} - \rho^2_{q_T} \underset{(p\times p)}{\Omega_{xx}}| = 0. \tag{4.2.1}
$$

However, because Ω_{zz} becomes infinite dimensional when $q_T \to \infty$ as $T \to \infty$, for these solutions to exist, certain structure on z_t and Ω_{zz} needs to be imposed.

Case: $p = 1$.

To explore conditions that guarantee the existence, we begin with the case where $p = 1$, that is, regressor x_t is a scalar.

To fix the idea, we begin with a simple linear IV model:

$$
\underset{(1\times 1)}{y_t} = \underset{(1\times 1)}{x_t} \underset{(1\times 1)}{\theta_0} + \underset{(1\times 1)}{u_t} \tag{4.2.2}
$$

$$
\underset{(1\times 1)}{x_t} = \underset{(1\times q_T)}{\gamma_{q_T}{}'} \underset{(q_T\times 1)}{z_{t,q_T}} + \underset{(1\times 1)}{\varepsilon_t} \tag{4.2.3}
$$

$$
z_{t,q_\infty} = lim_{T\to\infty} z_{t,q_T} \tag{4.2.4}
$$

The regression (4.2.2) is the object of interest. The relation (4.2.3) states that the regressor x_t is stochastically related to q_T instruments z_{t,q_T}. Finally, (4.2.4) indicates that the number of instruments tends to infinity as the sample size T tends to infinity.

The population moment condition for IV estimation implied by the model is:

$$E[\underset{(q_T\times 1)}{z_{t,q_T}}\underset{(1\times 1)}{(y_t - x_t\theta_0)}] = \underset{(q_T\times 1)}{0}. \tag{4.2.5}$$

and the resulting IV estimator is:

$$\begin{aligned}\hat{\theta}_T &= \{\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'}\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t{'}\}^{-1} \\ &\quad \frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'}\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} y_t \\ &= \{X_T' Z_T (Z_T' Z_T)^{-1} Z_T' X_T\}^{-1} X_T' Z_T (Z_T' Z_T)^{-1} Z_T' Y_T\end{aligned} \tag{4.2.6}$$

We restrict attention to the ordered sequence of candidate sets $\{z_{t,q_T}\}$ and consider the following specification of the γ_∞ vector.

Specification of γ_∞:

$$\sum_{i=1}^{\infty}\gamma_i^2 < \infty, \text{ and there does not exist finite } s \text{ such that } \gamma_i = 0 \text{ for } i > s. \tag{4.2.7}$$

In other words, this is a case where every additional instrument is useful in approximating x_t.

For x_t scalar, the determinantal equation (4.2.1) reduces to

$$\underset{(1\times q_T)}{\Omega_{xz}}\underset{(q_T\times q_T)}{\Omega_{zz}^{-1}}\underset{(q_T\times 1)}{\Omega_{zx}} - \rho_{q_T}^2 \underset{(1\times 1)}{\Omega_{xx}} = 0 \tag{4.2.8}$$

or,

$$E x_t z_{t,q_T}^{'}\left(E z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} E z_{t,q_T} x_t - \rho_{q_T}^2 E x_t x_t^{'} = 0. \tag{4.2.9}$$

Thus the eigenvalue, which is the squared population canonical correlation, is given by

$$\rho_{q_T}^2 = \frac{\underset{(1\times q_T)}{\Omega_{xz}}\underset{(q_T\times q_T)}{\Omega_{zz}^{-1}}\underset{(q_T\times 1)}{\Omega_{zx}}}{\underset{(1\times 1)}{\Omega_{xx}}} = \frac{E x_t z_{t,q_T}^{'}\left(E z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} E z_{t,q_T} x_t}{E x_t^2} \tag{4.2.10}$$

Therefore, for x_t scalar, the squared population canonical correlation is equal to the squared population multiple correlation coefficient, commonly denoted R^2.

The corresponding squared sample canonical correlation $r^2_{q_T}$ is given by the solution of

$$\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t{'} - r^2_{q_T}\frac{1}{T}\sum_{t=1}^{T} x_t x_t^{'} = 0 \qquad (4.2.11)$$

Thus,

$$r^2_{q_T} = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} \qquad (4.2.12)$$

Consistency and asymptotic normality of the linear IV estimator, and existence of the inverses $\left(E z_{t,q_T} z_{t,q_T}^{'}\right)^{-1}$ and $\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1}$ and hence that of $\rho^2_{q_T}$ and $r^2_{q_T}$ take place under certain conditions. Portnoy (1984, 1985) shows that such conditions place restrictions on the design matrix Z_T. They require that the empirical distribution of the vectors $\{z_{t,q_T}\}$ be near a distribution in $\Re^{q_T}$ with an appropriately smooth density. He shows that these conditions will hold in probability whenever the distribution of the sample $\{z_{t,q_T}\}$ is not too concentrated in any fixed direction, that is, the distribution of $\{a' z_{t,q_T}\}$ does not depend too strongly on the direction, $a/||a||$. Equivalently, the directions $\{z_{t,q_T}/||z_{t,q_T}||\}$ should be at least somewhat smoothly distributed over the unit sphere. Such conditions rule out cases where there may be different rates of information accumulating along different directions, such as the case when there are both trending and non-trending regressors.

Following Portnoy (1984, 1985), we formally impose the conditions referred to above on the design matrix Z_t.

Conditions on Z_T: Let

$$I(w,c) = \{t = 1, 2, ..., T : |z_{t,q_T}' w| \leq c\}$$

and let S be the ball in $\Re^{q_T}$ of radius δ and S^* be the sphere of radius 1.

(i) For any $c > 0$ and $\varepsilon > 0$, there are constants $\delta' > 0$ and $C > 0$ such that for all

$\gamma \in S, w \in S^*$ and T=1, 2, ...

$$\sum_{t \in J} (z'_{t,q_T} w)^2 \leq \varepsilon T \text{ where } J = I(\gamma, c) \cap I(w, C).$$

(ii) There exist constants b and B with $0 < b \leq B < \infty$ such that

$$\lambda_{max}(Z'_T Z_T) \leq BT \ \ a.s., \quad \lambda_{min}(Z'_T Z_T) \geq bT \ \ a.s.$$

Condition (i) above [Portnoy (1984, p. 1300)] is designed for the situation where $Cov(z_{t,q_T}) = Ez_{t,q_T} z^{'}_{t,q_T} = I$. It is innocuous because if it does not hold, then the transformation $\tilde{z}_{t,q_T} = \Omega_{zz}^{-1/2} z_{t,q_T}, \ \beta = \Omega_{zz}^{1/2} \gamma_{q_T}$ yields an equivalent problem with $Cov(\tilde{z}_{t,q_T}) = I$.

Then under the assumption that in the population, $\{z_{t,q_T}\}$ constitutes a sequence of orthonormal vectors, and thus $Ez_{t,q_T} z^{'}_{t,q_T} = I$, the inverse $(Ez_{t,q_T} z^{'}_{t,q_T})^{-1}$ always exist in the population. Again, by virtue of condition (ii), the inverse $\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z^{'}_{t,q_T}\right)^{-1}$ always exist in the sample. So, the existence of the population squared canonical correlation

$$\rho^2_{q_T} = \frac{Ex_t z^{'}_{t,q_T} Ez_{t,q_T} x_t}{Ex_t^2} \tag{4.2.13}$$

and of the sample squared canonical correlation

$$r^2_{q_T} = \frac{\frac{1}{T}\sum_{t=1}^{T} x_t z^{'}_{t,q_T} \left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z^{'}_{t,q_T}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t}{\frac{1}{T}\sum_{t=1}^{T} x_t^2} \tag{4.2.14}$$

is guaranteed.

4.3 Consistency of the Sample Canonical Correlation

Having guaranteed the existence of $\rho^2_{q_T}$ and $r^2_{q_T}$, we now proceed to establish the consistency of $r^2_{q_T}$ for $\rho^2_{q_T}$. To this end, let

$$Ex_t z^{'}_{t,q_\infty} Ez_{t,q_\infty} x_t = \xi < \infty. \tag{4.3.1}$$

Then we have

Proposition 4.3.1 *Assume that (i) $\{z_{t,q_T}\}$ is a sequence of $q_T \times 1$ i.i.d. random vectors* *with* $E\{z_{t,q_T}\}=0$ *and* $Cov\{z_{t,q_T}\}=I$, *(ii)* $\{\varepsilon_t\}$ *is a sequence of i.i.d. random variables with mean zero and variance* σ^2 *and is uncorrelated with* $\{z_{t,q_T}\}$, *(iii)* $\sup_T \max_{t=1,\ldots,T} ||Ex_t z'_{t,q_T}|| \le k$, *a finite constant, and (iv)* $\frac{{q_T}^2}{T} \to 0$. *Then* $r^2_{q_T} \xrightarrow{p} \rho^2_{q_T}$.

Proof of Proposition 4.3.1:

By the triangle inequality,

$$
\begin{aligned}
& |\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z'_{t,q_T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t-\xi| \\
\le\quad & |\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z'_{t,q_T}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t - Ex_t z'_{t,q_T} E z_{t,q_T} x_t| \\
& +|Ex_t z'_{t,q_T} E z_{t,q_T} x_t-\xi| \\
=\quad & |\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\left[\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z'_{t,q_T}\right)^{-1}-I\right]\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t| \\
& +|\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t - Ex_t z'_{t,q_T} E z_{t,q_T} x_t| \\
& +|Ex_t z'_{t,q_T} E z_{t,q_T} x_t-\xi| \\
=\quad & A+B+C, \text{ say,}
\end{aligned}
\tag{4.3.2}
$$

where

$$A=|\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\left[\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z'_{t,q_T}\right)^{-1}-I\right]\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t|, \tag{4.3.3}$$

$$B=|\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t - Ex_t z'_{t,q_T} E z_{t,q_T} x_t|, \tag{4.3.4}$$

$$C=|Ex_t z'_{t,q_T} E z_{t,q_T} x_t-\xi|. \tag{4.3.5}$$

Then,

$$
\begin{aligned}
B &= |\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}x_t - Ex_t z'_{t,q_T} E z_{t,q_T} x_t| \\
\le\quad & ||\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T} - Ex_t z'_{t,q_T}||^2 \\
& +2\,||Ex_t z'_{t,q_T}||\,||\frac{1}{T}\sum_{t=1}^{T} x_t z'_{t,q_T} - Ex_t z'_{t,q_T}||
\end{aligned}
\tag{4.3.6}
$$

where

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}x_t z_{t,q_T}^{'} - Ex_t z_{t,q_T}^{'} &= \frac{1}{T}\sum_{t=1}^{T}\left(\gamma_{q_T}{}'z_{t,q_T}+\varepsilon_t\right)z_{t,q_T}^{'} - E[\left(\gamma_{q_T}{}'z_{t,q_T}+\varepsilon_t\right)z_{t,q_T}^{'}] \\ &= \frac{1}{T}\sum_{t=1}^{T}\gamma_{q_T}{}'z_{t,q_T}z_{t,q_T}^{'} + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}^{'} - E(\gamma_{q_T}{}'z_{t,q_T}z_{t,q_T}^{'}) \\ &\qquad [\text{since } E(\varepsilon_t z_{t,q_T}^{'})=0] \\ &= \gamma_{q_T}{}'[\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}z_{t,q_T}^{'} - Ez_{t,q_T}z_{t,q_T}^{'}] + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}^{'}.\end{aligned} \tag{4.3.7}$$

Now, from our imposed condition on the sequence of γ's it directly follows that, $||\gamma_{q_T}|| = O(1)$.

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'} - E z_{t,q_T} z_{t,q_T}^{'} &= \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'} - I \\
&= \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} z_{t,1} \\ z_{t,2} \\ \vdots \\ z_{t,q_T} \end{pmatrix} \begin{pmatrix} z_{t,1} & z_{t,2} & \dots & z_{t,q_T} \end{pmatrix} - I \\
&= \frac{1}{T}\sum_{t=1}^{T} \begin{pmatrix} {z_{t,1}}^2 & z_{t,1} z_{t,2} & \dots & z_{t,1} z_{t,q_T} \\ z_{t,2} z_{t,1} & {z_{t,2}}^2 & \dots & z_{t,2} z_{t,q_T} \\ \dots & & & \\ z_{t,q_T} z_{t,1} & z_{t,q_T} z_{t,2} & \dots & {z_{t,q_T}}^2 \end{pmatrix} - I \\
&= \begin{pmatrix} \underbrace{\frac{1}{T}\sum_{t=1}^{T} {z_{t,1}}^2 - 1}_{O_p(\frac{1}{\sqrt{T}})} & \underbrace{\frac{1}{T}\sum_{t=1}^{T} z_{t,1} z_{t,2}}_{O_p(\frac{1}{\sqrt{T}})} & \dots & \underbrace{\frac{1}{T}\sum_{t=1}^{T} z_{t,1} z_{t,q_T}}_{O_p(\frac{1}{\sqrt{T}})} \\ \underbrace{\frac{1}{T}\sum_{t=1}^{T} z_{t,2} z_{t,1}}_{O_p(\frac{1}{\sqrt{T}})} & \underbrace{\frac{1}{T}\sum_{t=1}^{T} {z_{t,1}}^2 - 1}_{O_p(\frac{1}{\sqrt{T}})} & \dots & \underbrace{\frac{1}{T}\sum_{t=1}^{T} z_{t,2} z_{t,q_T}}_{O_p(\frac{1}{\sqrt{T}})} \\ \vdots & \vdots & \vdots & \vdots \\ \underbrace{\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,1}}_{O_p(\frac{1}{\sqrt{T}})} & \underbrace{\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,2}}_{O_p(\frac{1}{\sqrt{T}})} & \dots & \underbrace{\frac{1}{T}\sum_{t=1}^{T} {z_{t,q_T}}^2 - 1}_{O_p(\frac{1}{\sqrt{T}})} \end{pmatrix}
\end{aligned}
\tag{4.3.8}
$$

Therefore, $||\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'} - E z_{t,q_T} z_{t,q_T}^{'}|| = O_p(\frac{q_T}{\sqrt{T}})$ [where we use the Frobenius norm: $||A|| = \sqrt{tr(A'A)}$].

Next we note that, $\varepsilon_1 z_{1,q_T}, \varepsilon_2 z_{2,q_T}, \dots, \varepsilon_T z_{T,q_T}$ are a sequence of $q_T \times 1$ i.i.d. random vectors with $E(\varepsilon_t z_{t,q_T}) = 0$ by independence of ε_t and z_{t,q_T}, and $Cov(\varepsilon_t z_{t,q_T}) = E(\varepsilon_t z_{t,q_T} {z_{t,q_T}}' \varepsilon_t) = E(\varepsilon_t^2) E(z_{t,q_T} {z_{t,q_T}}') = \sigma^2 I$ by $E(\varepsilon_t^2) = \sigma^2$ and $E(z_{t,q_T} {z_{t,q_T}}') = I$.

Therefore, by Theorem 1.1 of Portnoy (1986) [p. 572], $||\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}^{'}|| = O_p(\frac{1}{\sqrt{T}})$. Hence,

$$\begin{aligned}
B &= |\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t - E x_t z_{t,q_T}^{'} E z_{t,q_T} x_t| \\
&\leq ||\gamma_{q_T}{}'[\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'} - E z_{t,q_T} z_{t,q_T}^{'}] + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}^{'}||^2 \\
&+ 2\,||E x_t z_{t,q_T}^{'}||\,||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} - E x_t z_{t,q_T}^{'}|| \\
&\leq ||\gamma_{q_T}{}'[\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'} - E z_{t,q_T} z_{t,q_T}^{'}]||^2 + 2||\gamma_{q_T}{}'[\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'} - E z_{t,q_T} z_{t,q_T}^{'}]||\,||\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}^{'}|| \\
&+ ||\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}^{'}||^2 + 2\,||E x_t z_{t,q_T}^{'}||\,||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} - E x_t z_{t,q_T}^{'}|| \\
&\leq O(1) O_p(\frac{{q_T}^2}{T}) + 2\,O(1) O_p(\frac{q_T}{\sqrt{T}}) O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{T}) + 2\,O(1) O_p(\frac{q_T}{\sqrt{T}}) \\
&= O_p(\frac{q_T}{\sqrt{T}}) \\
&= o_p(1) \text{ since} \frac{q_T}{\sqrt{T}} \to 0.
\end{aligned} \tag{4.3.9}$$

Again,

$$A = |\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} I^{-1}(I - \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'})(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'})^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t|$$

[Using $E^{-1} - F^{-1} = F^{-1}(F - E)E^{-1}$]

$$\leq ||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'}||\,||I - \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}||\,||\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t|| = abc \tag{4.3.10}$$

where

$$a = ||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'}||, \tag{4.3.11}$$

$$b = ||I - \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}||, \tag{4.3.12}$$

$$c = ||\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t||. \tag{4.3.13}$$

Again,

$$
\begin{aligned}
a = ||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'}|| &= ||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} - Ex_t z_{t,q_T}^{'} + Ex_t z_{t,q_T}^{'}|| \\
&\leq ||\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} - Ex_t z_{t,q_T}^{'}|| + ||Ex_t z_{t,q_T}^{'}||, \\
&= O_p\left(\frac{q_T}{\sqrt{T}}\right) + O(1) \\
&= O(1).
\end{aligned} \tag{4.3.14}
$$

By our earlier result we know, $b = ||I - \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}|| = O_p(\frac{q_T}{\sqrt{T}})$.

Next, $\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t = \hat{\gamma}_{q_T,OLS} = \hat{\gamma}_{q_T,OLS} - \gamma_{q_T} + \gamma_{q_T}$

$\leq \underbrace{||\hat{\gamma}_{q_T,OLS} - \gamma_{q_T}||}_{o_p(1)\text{under the condition} \frac{q_T}{\sqrt{T}} \to 0 \text{ by Portnoy (1984)}} + \underbrace{||\gamma_{q_T}||}_{O(1)} = O(1).$

and thus, $c = ||\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t|| = O(1)$.

Hence, $A = abc = O(1)\ O_p(\frac{q_T}{\sqrt{T}})\ O(1) = o_p(1),$ since $\frac{q_T}{\sqrt{T}} \to 0$.

Also, $C = |Ex_t z_{t,q_T}^{'} E z_{t,q_T} x_t - \xi| \to 0$ by the definition of ξ.

So it only remains to show that $\frac{1}{T}\sum_{t=1}^{T} x_t^2 \xrightarrow{p} Ex_t^2$, where

$$
\begin{aligned}
Ex_t^2 &= E[(\gamma_{q_T}{}' z_{t,q_T} + \varepsilon_t)\,(z_{t,q_T}{}'\gamma_{q_T} + \varepsilon_t)] \\
&= \gamma_{q_T}{}' E(z_{t,q_T} z_{t,q_T}{}')\gamma_{q_T} + \gamma_{q_T}{}' E(z_{t,q_T}\varepsilon_t) + E(\varepsilon_t z_{t,q_T}{}')\gamma_{q_T} + E\varepsilon_t^2 \\
&= \gamma_{q_T}{}'\gamma_{q_T} + \sigma^2 \ \ [\text{since } E(z_{t,q_T} z_{t,q_T}{}') = I,\ E(z_{t,q_T}\varepsilon_t) = 0 \text{ and } E\varepsilon_t^2 = \sigma^2].
\end{aligned} \tag{4.3.15}
$$

Now,

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} x_t^2 &= \frac{1}{T}\sum_{t=1}^{T}[(\gamma_{q_T}{}' z_{t,q_T} + \varepsilon_t)\,(z_{t,q_T}{}'\gamma_{q_T} + \varepsilon_t)] \\
&= \gamma_{q_T}{}'\gamma_{q_T} + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}{}'\gamma_{q_T} + \gamma_{q_T}{}'\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2.
\end{aligned} \tag{4.3.16}
$$

Let $S_T = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}{}'\gamma_{q_T}$ and note that S_T is a scalar. Also,

$$
E(S_T) = \frac{1}{T}\sum_{t=1}^{T}\{E(\varepsilon_t)\}\{E(z_{t,q_T}{}'\gamma_{q_T})\} = 0 \ \ [\text{by independence of } \{\varepsilon_t\} \text{ and} \{z_{t,q_T}\}] \tag{4.3.17}
$$

and

$$
\begin{aligned}
Var(S_T) = E(S_T^2) = E(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t z_{t,q_T}{}'\gamma_{q_T})^2 &= \frac{1}{T^2}\sum_{t=1}^{T}\{E(\varepsilon_t^2)\}\{E(z_{t,q_T}{}'\gamma_{q_T})^2\} \\
= \frac{\sigma^2}{T^2}\sum_{t=1}^{T}\gamma_{q_T}{}'\gamma_{q_T} = \frac{\sigma^2}{T^2}\,T\sum_{i=1}^{q_T}\gamma_i{}^2 &= \frac{\sigma^2}{T}\sum_{i=1}^{q_T}\gamma_i{}^2.
\end{aligned} \tag{4.3.18}
$$

Therefore, by Chebychev's inequality, $P(|S_T| \geq \varepsilon) \leq \frac{Var(S_T)}{\varepsilon^2} = O(\frac{1}{T}) \to 0$ as $T \to \infty$. Also, by WLLN, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} E\varepsilon_t^2 = \sigma^2$. Thus, $\frac{1}{T}\sum_{t=1}^{T}x_t^2 \xrightarrow{p} [\gamma_{q_T}{}'\gamma_{q_T} + \sigma^2] = Ex_t^2 < \infty$. Hence, $r_{q_T}^2 \xrightarrow{p} \rho_{q_T}^2$. ■

4.4 Consistency and Asymptotic Normality of the Linear IV Estimator

Proposition 4.4.1 *Assume conditions (i) and (ii) on the design matrix Z_T and assumptions of Proposition 1 above hold. Then*

(a) $\hat{\theta}_T \xrightarrow{P} \theta_0$, *and*

(b) $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{D} N(0,\ \sigma^2[\Omega_{xz}^{\infty}\underbrace{(\Omega_{zz}^{\infty})^{-1}}_{=I}\Omega_{zx}^{\infty})]^{-1}) \equiv N(0,\ \sigma^2[\Omega_{xz}^{\infty}\Omega_{zx}^{\infty}]^{-1})$.

Proof of Proposition 4.4.1:

(a) From (4.2.6) we have

$$
\begin{aligned}
\hat{\theta}_T = \theta_0 + \{&\frac{1}{T}\sum_{t=1}^{T}x_t z_{t,q_T}^{'}\left(\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}x_t{}'\}^{-1} \\
&\frac{1}{T}\sum_{t=1}^{T}x_t z_{t,q_T}^{'}\left(\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}\varepsilon_t \\
= \theta_0 + M_{q_T}&\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}\varepsilon_t
\end{aligned} \tag{4.4.1}
$$

where

$$
\begin{aligned}
M_{q_T} = \{&\frac{1}{T}\sum_{t=1}^{T}x_t z_{t,q_T}^{'}\left(\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}z_{t,q_T}^{'}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}x_t{}'\}^{-1} \\
&\frac{1}{T}\sum_{t=1}^{T}x_t z_{t,q_T}^{'}\left(\frac{1}{T}\sum_{t=1}^{T}z_{t,q_T}z_{t,q_T}^{'}\right)^{-1}.
\end{aligned} \tag{4.4.2}
$$

By WLLN, $\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \xrightarrow{P} E x_t z_{t,q_T}^{'}$, $\left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right) \xrightarrow{P} E z_{t,q_T} z_{t,q_T}{}'$.

Then, by Continuous Mapping Theorem, $M_{q_T} \xrightarrow{P} M$, where

$M = \{E x_t z_{t,q_T}^{'} (E z_{t,q_T} z_{t,q_T}{}')^{-1} E z_{t,q_T} x_t{}'\}^{-1} E x_t z_{t,q_T}^{'} (E z_{t,q_T} z_{t,q_T}{}')^{-1}$.

Again by WLLN, $\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}\varepsilon_t \xrightarrow{P} E z_{t,q_T}\varepsilon_t = 0$. Therefore,

$plim\ \hat{\theta}_T = \theta_0 + plim\ M_{t,q_T}\ plim\ \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}\varepsilon_t = \theta_0 + M \times 0 = \theta_0$. ■

(b) From (4.4.1) we have

$$\begin{aligned} T^{1/2}(\hat{\theta}_T - \theta_0) &= \{\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} x_t{}'\}^{-1} \\ &\quad \frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \left(\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T} z_{t,q_T}^{'}\right)^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{t,q_T}\varepsilon_t \\ &= M_{q_T} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{t,q_T}\varepsilon_t \end{aligned} \tag{4.4.3}$$

As noted in the proof of Proposition 4.3.1, $z_{1,q_T}\varepsilon_1, z_{2,q_T}\varepsilon_2, \ldots, z_{T,q_T}\varepsilon_T$ are a sequence of $q_T \times 1$ i.i.d. random vectors with $E\left(z_{t,q_T}\varepsilon_t\right) = 0$ by independence of z_{t,q_T} and ε_t, and $Cov\left(z_{t,q_T}\varepsilon_t\right) = E\left(z_{t,q_T}\varepsilon_t\varepsilon_t z_{t,q_T}{}'\right) = E\left(\varepsilon_t^2\right) E\left(z_{t,q_T} z_{t,q_T}{}'\right) = \sigma^2 I$ by $E\left(\varepsilon_t^2\right) = \sigma^2$ and $E\left(z_{t,q_T} z_{t,q_T}{}'\right) = I$. Therefore, by Theorem 1.1 of Portnoy (1986) [p. 572],

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{t,q_T}\varepsilon_t \xrightarrow{D} N(0,\ \sigma^2 I). \tag{4.4.4}$$

Assume that a Lindberg-type condition holds for the elements of the vector $x_t z_{t.q_T}^{'}$, that is,

$$lim_{T\to\infty} max_{t\le T} max_{q\le q_T} x_t^2 z_{t,q}^2 / \sum_{t=1}^{T}\sum_{q=1}^{q_T} x_t^2 z_{t,q}^2 = 0. \tag{4.4.5}$$

Then, by Central Limit Theorem, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t z_{t,q_T}^{'} - E[x_t z_{t,q_T}^{'}]) \xrightarrow{D} N(0,\ V)$ where $V = lim_{T\to\infty} Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t z_{t,q_T}^{'})$, and it follows that $\frac{1}{T}\sum_{t=1}^{T} x_t z_{t,q_T}^{'} \xrightarrow{P} E(x_t z_{t,q_T}^{'}) = \Omega_{xz}^{\infty}$. Similarly, assume that a Lindberg-type condition holds for the elements of the matrix $z_{t,q_T} z_{t,q_T}^{'}$, that is,

$$lim_{T\to\infty} max_{t\le T} max_{q\le q_T} z_{t,q}^2 / \sum_{t=1}^{T}\sum_{q=1}^{q_T} z_{t,q}^2 = 0. \tag{4.4.6}$$

Then, by Central Limit Theorem, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_{t,q_T}z_{t,q_T}^{'} - E[z_{t,q_T}z_{t,q_T}^{'}]) \xrightarrow{D} N(0,\ \Gamma)$ where $\Gamma = lim_{T\to\infty}Var(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{t,q_T}z_{t,q_T}^{'})$, and it follows that $\frac{1}{T}\sum_{t=1}^{T} z_{t,q_T}z_{t,q_T}^{'} \xrightarrow{P} E(z_{t,q_T}z_{t,q_T}^{'}) = \Omega_{zz}^{\infty} = I$. Then by Continuous Mapping Theorem, $M_{q_T} \xrightarrow{P} (\Omega_{xz}^{\infty}(\Omega_{zz}^{\infty})^{-1}\Omega_{zx}^{\infty})^{-1}\Omega_{xz}^{\infty}(\Omega_{zz}^{\infty})^{-1} = (\Omega_{xz}^{\infty}\Omega_{zx}^{\infty})^{-1}\Omega_{xz}^{\infty}$. Finally, applying Slutsky's Theorem completes the proof. ■

4.5 Canonical Correlations Information Criterion

Hall and Peixe's (2003) canonical correlations information criterion is given by:

$$CCIC(c) = \ \Xi_T(c) \ + \ P(T, |c|) \tag{4.5.1}$$

where the statistic

$$\Xi_T(c) = \ T\sum_{i=1}^{p} ln[1 - r_{i,T}^2(c)] \tag{4.5.2}$$

captures the sample information, and $P(T, |c|)$ is the penalty term which satisfies the following conditions: (i) $P(T, |c|) = h(|c|)\mu_T$; (ii) h(.) is non-negative and strictly increasing; (iii) $\mu_T \to \infty$ as $T \to \infty$ and $\mu_T = o(T)$.

The functional form of $\Xi_T(c)$ is motivated by the form of a likelihood ratio statistic when the dimension of the parameter vector underlying the model of interest is fixed. Suppose that x_t is generated by the following model

$$\underset{1\times 1}{x_t} \ = \underset{1\times q}{\gamma'} \ \underset{q\times 1}{z_t(c)} + \underset{1\times 1}{e_t} \tag{4.5.3}$$

$$= \underset{1\times q_1}{\gamma_1'} \ \underset{q_1\times 1}{z_t(c_1)} + \underset{1\times q_2}{\gamma_2'} \ \underset{q_2\times 1}{z_t(c_2)} + \underset{1\times 1}{e_t} \tag{4.5.4}$$

where the number of instruments, $q = (q_1 + q_2)$ is fixed, $z_t(c) = [z_t(c_1)', z_t(c_2)']'$, $c_1'c_2 = 0$ and $E[e_t|z_t] = 0$. The statement that $z_t(c_2)$ is redundant given $z_t(c_1)$ is equivalent to the null hypothesis $H_0 : \gamma_2 = 0$. If $e_t \sim IN(0, \Omega_0)$, then the likelihood ratio statistic is

$$LR_T = \Xi_T^0(c_1) - \Xi_T^0(c) \tag{4.5.5}$$

where $\Xi_T^0(c) = \ T\sum_{i=1}^{p} ln[1 - r_{i,T}^2(c)]$ and $r_{i,T}(c)$ is the ith sample canonical correlation between $x_t(\theta_0)$ and $z_t(c)$.

It can be seen that the two components of CCIC(c) (4.5.1) move in opposite directions in response to the inclusion of an additional instrument: the first term $\Xi_T(c)$ either stays the same or decreases, while the second term $P(T, |c|)$ increases. Hence, if the selection vector is chosen to minimize CCIC(c) then the resulting instrument vector retains only those instruments whose inclusion reduces $\Xi_T(c)$ sufficiently to offset the increase in the penalty term $P(T, |c|)$.

In contrast to model (4.5.3) above, our model becomes

$$\underset{1\times 1}{x_t} = \underset{1\times q_T}{\gamma'_{q_T}} \underset{q_T\times 1}{z_{t,q_T}(c)} + \underset{1\times 1}{e_t} \tag{4.5.6}$$

$$= \underset{1\times q_{1T}}{\gamma'_{q_{1T}}} \underset{q_{1T}\times 1}{z_{t,q_T}(c_1)} + \underset{1\times q_{2T}}{\gamma'_{q_{2T}}} \underset{q_{2T}\times 1}{z_{t,q_T}(c_2)} + \underset{1\times 1}{e_t} \tag{4.5.7}$$

where the number of instruments, $q_T = (q_{1T} + q_{2T}) \to \infty$ as $T \to \infty$, $z_{t,q_T}(c) = [z_{t,q_T}(c_1)',$ $z_{t,q_T}(c_2)']'$, $c_1'c_2 = 0$ and $E[e_t|z_t] = 0$. Similarly to the Hall and Peixe (2003) setting above, the statement that $z_{t,q_T}(c_2)$ is redundant given $z_{t,q_T}(c_1)$ is equivalent to the null hypothesis $H_0 : \gamma_{q_{2T}} = 0$. Although the dimension of $\gamma_{q_{2T}} \to \infty$ as $T \to \infty$, given our specification of the γ_∞ vector in (4.2.7), it follows from the consistency of sample canonical correlation given by Proposition 1 above, that the CCIC(c) of Hall and Peixe (2003) will lead to the selection of all relevant instruments with probability one.

Theorem 4.5.1 *Assume that (i) the specification of the γ_∞ vector in (4.2.7) holds; (ii) $\tilde{c}_T = \underset{c\in C}{argminCCIC}(c)$ where C is the set of all possible choices of c; (iii) the penalty function $P(T, |c|) = h(|c|)\mu_T$, satisfies the following conditions: (a) h(.) is non-negative and strictly increasing; (b) $\mu_T \to \infty$ as $T \to \infty$ and $\mu_T = o(T)$. Then $\tilde{c}_T \xrightarrow{P} c_{r,\infty}$, where $c_{r,\infty}$ is the set of all relevant instruments.*

Proof of Theorem 4.5.1 As in Hall and Peixe (2003), the proof relies on considering the limiting behavior of CCIC(c) when the instrument vector $z_{t,q_T}(c_1)$ is augmented by including the vector $z_{t,q_T}(c_2)$. Let $\hat{\rho}_T$ be the sample canonical correlation between x_t and $z_{t,q_T}(c)$, r_T be the sample canonical correlation between x_t and $z_{t,q_T}(c_1)$.

Then,

$$
\begin{aligned}
\Delta &= CCIC(z_{t,q_T}(c)) - CCIC(z_{t,q_T}(c_1)) \\
&= Tln[1 - \hat{\rho}_T^2[x_t : (z_{t,q_T}(c_1), z_{t,q_T}(c_2))]] + h(|c|)\mu_T \\
&\quad - Tln[1 - r_T^2[x_t : (z_{t,q_T}(c_1)]] - h(|c_1|)\mu_T \\
&= Tln\left\{\frac{1 - \hat{\rho}_T^2[x_t : (z_{t,q_T}(c_1), z_{t,q_T}(c_2))]}{1 - r_T^2[x_t : (z_{t,q_T}(c_1)]}\right\} + [h(|c|) - h(|c_1|)]\mu_T.
\end{aligned}
$$

Hence,

$$
\frac{\Delta}{T} = ln\left\{\frac{1 - \hat{\rho}_T^2[x_t : (z_{t,q_T}(c_1), z_{t,q_T}(c_2))]}{1 - r_T^2[x_t : (z_{t,q_T}(c_1)]}\right\} + [h(|c|) - h(|c_1|)]\frac{\mu_T}{T}.
$$

Now, from the specification of the γ_∞ vector, it follows that $z_{t,q_T}(c_2)$ is not redundant given $z_{t,q_T}(c_1)$ and that inclusion of the additional instruments $z_{t,q_T}(c_2)$ will yield the result that

$$
\hat{\rho}_T[x_t : (z_{t,q_T}(c_1), z_{t,q_T}(c_2))] > r_T[x_t : (z_{t,q_T}(c_1)].
$$

Therefore, because sample canonical correlations are consistent estimators of the population canonical correlations, as $T \to \infty$,

$$
ln\left\{\frac{1 - \hat{\rho}_T^2[x_t : (z_{t,q_T}(c_1), z_{t,q_T}(c_2))]}{1 - r_T^2[x_t : (z_{t,q_T}(c_1)]}\right\} \xrightarrow{P} k < 0,
$$

where k is a constant. Again, as a consequence of $|c| > |c_1|$ and of condition $(ii)(a)$, $[h(|c|) - h(|c_1|)] > 0$. Finally, by condition $(ii)(b)$, as $T \to \infty$, $\mu_T/T \to 0$. Thus, $\Delta < 0$ with probability one in the limit, implying that implementation of the CCIC(c) of Hall and Peixe (2003) will continue to decrease the value of the criterion and thus inclusion of the relevant instruments until the vector $c_{r,\infty}$ is selected in the limit. ■

4.6 Conclusions

The objective of this chapter has been to extend the statistical theory of the CCIC of Hall and Peixe (2003) to the case in which the the candidate set of instruments increases with the sample size. It focuses on the case where the regressor x_t is a scalar. It establishes consistency of the sample canonical correlation and also proves consistency and asymptotic normality of the linear IV estimator. Finally, this chapter shows that for the specific instrument generating scheme considered, implementation of the CCIC of Hall and Peixe (2003) will lead to the selection of all relevant instruments from the growing candidate set with probability one in the limit.

Chapter 5

Summary and Directions of Future Work

In this dissertation we develop methods of estimation of a new kind of canonical correlations called Long Run Canonical Correlations (LRCC) that has recently emerged in econometrics to provide a metric of relevance for moments used in GMM estimation. In addition, we show further usefulness of LRCC beyond their usefulness in moment selection. In particular, we show how LRCC can be used to develop econometric tests that can play a role in (i) structural stability testing, and (ii) exogeneity testing of regressors in a cointegration model. To this end, we establish a link between LRCC and canonical coherence at frequency zero. By exploiting this result, we develop what we call the Hannan and LR tests of persistence and the Hannan and LR tests of exogeneity. The importance of tests of persistence is that they can be used to pre-test data to screen out cases where subjecting the data to the Wald test of structural stability would be misleading. The importance of tests of exogeneity of regressors in a cointegration model is that in the case of strict exogeneity, the limiting distribution of the estimator of the slope vector becomes nuisance parameter free. Simulation results of the Hannan and LR tests of persistence are mixed. It is seen that if the tests are conducted without prewhitening and recoloring, then they perform well. However, if they are implemented with prewhitening and recoloring, then the quality deteriorates. Simulation results of the Hannan and LR tests of exogeneity, on the other hand, indicate that the tests are highly satisfactory.

Next, we conduct a comparative study of three recent methods of instrument selection in econometrics, namely, the CCIC of Hall and Peixe (2003), the RMSC of Hall and Inoue (2003), and the approximate MSE of Donald and Newey (2001). In this context, we explore three questions: (i) What, if any, is the analytical connection among the three methods? (ii) Is a unique ranking of the three methods possible in terms of the finite sample behavior of the post selection estimator? (iii) What guidance can we provide a practitioner as to which of these three methods one should use in any practical application of interest? The answer to the first question reveals that all three methods are functions of canonical correlations. The answer to the second is a conditional one, in the sense that, a unique ranking emerges under certain parameter configurations while it does not under others. In light of the nature of the second answer, the answer to the third question is obviously found to be a conditional one as well.

Finally, we extend the statistical theory of the CCIC of Hall and Peixe (2003) to the case where the number of instruments goes to infinity with the sample size. Here, we limit our focus to the case where there is only one endogenous regressor and show that, for the particular kind of instrument generation scheme considered, implementation of the CCIC of Hall and Peixe (2003) will lead to the selection of all relevant instruments from the growing candidate set with probability one in the limit.

So far we discussed works that have been done in this dissertation. Now we turn to directions of future works that emerge as a natural outgrowth of this research.

Although the methods of estimation of LRCC have been developed, methods of their inference have not been developed. It is of interest to know how many population LRCC are different from zero; that is, how many long run canonical variates are needed to explain the correlations between X_T and Z_T. The number of nonzero LRCC is equal to the rank of Σ_{xz}. One could be interested in testing the null hypothesis that all LRCC are zero: $\rho_1 = \rho_2 = \ldots = \rho_p = 0$, i.e., $\Sigma_{xz} = 0$. If this is accepted, then there are clearly no useful long run canonical variates. If it is rejected, then it is possible that $\rho_1 > \rho_2 = \ldots = \rho_p = 0$, and so, $rank(\Sigma_{xz}) = 1$, in which case only the first long run canonical variates are useful. If this is tested and rejected, we can test whether the smallest p - 2 LRCC are zero, and so on. Thus, we have a sequence of hypotheses H_k: $\rho_{k+1} = \ldots = \rho_p = 0$ for $k = 0, 1, \ldots, p-1$.

By exploiting the result of Proposition 2.4.1 of Chapter 2 of this dissertation, we can conduct a likelihood ratio test in the frequency domain by adapting Theorem 15 of Hannan [(1970), p. 300]. It can be implemented through Wilk's lambda statistic: [1]

$$\begin{aligned}\hat{\Lambda} &= \frac{|\hat{f}_{xx}(0) - \hat{f}_{xz}(0)\hat{f}_{zz}^{-1}(0)\hat{f}_{zx}(0)|}{|\hat{f}_{xx}(0)|} \\ &= \prod_{i=k+1}^{p} (1 - \hat{\rho}_i^2)\end{aligned}$$

Deriving the asymptotic distribution of the above lambda statistic under the null hypothesis:

$$H_k : \rho_{k+1} = \ldots = \rho_p = 0$$

for $k = 0, 1, \ldots, p-1$, is one of the areas of future work.

Other areas of future research relate to: generalizing result (2.5.6) to near-integrated processes along the lines of Phillips (1987); modifications of CCIC(c) for other specifications of the γ_∞ vector; extension of analysis of Chapter 4 to the case where $p > 1$, that is, where the regressor x_t is a vector, and comparing the performance of $CCIC(c)$ in expanding parameter case with that of Donald and Newey's (2001) $MSE(c)$.

[1] Hannan [(1970), p. 301]

Bibliography

[1] Akaike, H. (1974). 'A New Look at the Statistical Model Identification', *IEEE Transactions on Automatic Control*, AC-19(6): 716-723.

[2] Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York, Wiley.

[3] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York, U.S.A., 2nd edn.

[4] Andrews, D. W. K. (1991). 'Heteroscedasticity and Autocorrelation Consistent Covaraince Matrix Estimation', *Econometrica*, 57: 817-858.

[5] Andrews, D. W. K. (1999). 'Consistent Moment Selection Procedures for Generalized Method of Moments Estimation', *Econometrica*, 67: 543-564.

[6] Andrews, D. W. K., and Ray C. Fair (1988). 'Inference in Nonlinear Econometric Models with Structural Change', *Review of Economic Studies*, 55: 615-640.

[7] Andrews, D. W. K., and J. Christopher Monahan (1992). 'An Improved Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimator', *Econometrica*, 60: 953-666.

[8] Bound, J., D. A. Jaeger and R. Baker (1995). 'Problems with Instrumental Variables Estimation when the Correlations between the Instruments and the Endogenous Ex-

planatory Variables is Weak', *Journal of the American Statistical Association*, 90: 443-450.

[9] Breusch, T., Qian, H., Schmidt, P., and Wyhowski, D. (1999). 'Redundancy of Moment Conditions', *Journal of Econometrics*, 91: 89-111.

[10] Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* New York: Holden-Day.

[11] Donald, S. G., and Newey, W. K. (2001). 'Choosing the Number of Instruments', *Econometrica*, 69: 1161-1192.

[12] Dhrymes, P. H. (1974). *Econometrics: Statistical Foundations and Applications.* Springer-Verlag, New York, U.S.A.

[13] Fuller, W. A. (1996). *Introduction to Statistical Time Series.* Wiley, New York, U.S.A., 2nd edn.

[14] Gallant, A. R. (1987). *Nonlinear Statistical Models.* Wiley, New York, U.S.A.

[15] Galton, F. (1888). 'Co-relations and their measurement, chiefly from anthropometric data', *Proceedings of the Royal Society*, 35: 135-145.

[16] Ghysels, E. and Hall A. (1990a). 'A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator', *International Economic Review*, 31: 355-364.

[17] Ghysels, E. and Hall A. (1990b). 'Are Consumption-based Intertemporal Capital Asset Pricing Models Structural?', *Journal of Econometrics*, 45: 121-139.

[18] Hahn, J., and J. Hausman (2002). 'A New Specification Test for the Validity of Instrumental Variables', *Econometrica*, 70: 163-189.

[19] Hall, A. R., and Inoue, A. (2003). 'Information in Empirical Likelihood and Generalized Method of Moments Estimation', *Manuscript*, Department of Economics, North Carolina State University, Raleigh, NC.

[20] Hall, A. R., and Peixe, F. P. M. (2003). 'A Consistent Method for the Selection of Relevant Instruments', *Econometric Reviews*, Vol. 22, No. 3, 269-287.

[21] Hall, A. R. and Sen, A. (1999). 'Structural Stability Testing Based on Generalized Method of Moments Estimators', *Journal of Business and Economic Statistics*, July 1999, 335-348.

[22] Hannan, E. J. (1970). *Multiple Time Series.* Wiley, New York, NY, U.S.A.

[23] Hannan, E. J., and Quinn, B. G. (1979). 'The Determination of the Order of an Autoregression', *Journal of the Royal Statistical Society, Series B, 41(2):*190-195.

[24] Hansen, L. P. (1982). 'Large Sample Properties of Generalized Method of Moments Estimators', *Econometrica*, 50: 1029-1054.

[25] Hotelling, H. (1935). 'The Most Predictable Criterion', *Journal of Educational Psychology*, 26: 139-142.

[26] Hotelling, H. (1936). 'Relations between two Sets of Variates', *Biometrika*, 28: 321-377.

[27] Lehmann, E. L. (1983). *Theory of Point Estimation.* Wiley, New York, NY, U.S.A.

[28] Mallows, C. L. (1973). 'Some Comments on C_p', *Technometrics,* 15: 661-675

[29] Nagar, A. L. (1959). 'The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations', *Econometrica*, 27: 575-595.

[30] Newey, W. K., and West, K. D. (1987). 'A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica*, 55: 703-708.

[31] Phillips, P. C. B. (1986). 'Understanding Spurious Regressions in Econometrics', *Journal of Econometrics*, 33: 311-340.

[32] Phillips, P. C. B. (1987). 'Towards a Unified Asymptotic Theory for Autoregression', *Biometrika*, Vol. 74, No.3, 535-547.

[33] Portnoy, S. (1984). 'Asymptotic Behavior of M-Estimators of p Regression Parameters when $\frac{p^2}{n}$ is Large. I. Consistency', *The Annals of Statistics*, Vol. 12, No. 4, 1298-1309.

[34] Portnoy, S. (1985). 'Asymptotic Behavior of M-Estimators of p Regression Parameters when $\frac{p^2}{n}$ is Large. II. Normal Approximation', *The Annals of Statistics*, Vol. 13, No. 4, 1403-1417.

[35] Portnoy, S. (1986). 'On the Central Limit Theorem in R^p when $p \to \infty$', *Probability Theoryand Related Fields*, 73: 571-583.

[36] Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York, U.S.A., 2nd edn.

[37] Sargan, J. D. (1958). 'The Estimation of Economic Relationships using Instrumental Variables', *Econometrica*, 26: 393-415.

[38] Schwarz, G. (1978). 'Estimating the Dimension of a Model', *The Annals of Statistics*, 6: 461-464.

[39] Wooldridge, J. (1994). 'Estimation and Inference for Dependent processes', in R. F. Engle and D. McFadden, eds., *Handbook of Econometrics.* Vol. 4, North-Holland: New York.