
Volume 105, No. 1, October 2009
ISSN 0165-1765
ELSEVIER
economics letters
Editor:
ERIC MASKIN
Co-Editors:
STEPHEN DONALD
ANDREW SAMWICK
Advisory Editors:
JUSHAN BAI
STEVE BERRY
AMITABH CHANDRA
RICHARD H. CLARIDA
SUSAN M. COLLINS
GRAHAM ELLIOTT
BO HONORÉ
DAVID A. HSIEH
THOMAS KANE
MILES KIMBALL
NOBUHIRO KIYOTAKI
FRANK KLEIBERGEN
KALA KRISHNA
WOLFGANG MAYER
BENNETT McCALLUM
JAMES MIRRLEES
DEREK NEAL
WILLIAM S. NEILSON
MASAO OGAKI
ARIEL PAKES
DAVID SAPPINGTON
DAVID SCHARFSTEIN
DANIEL SLESNICK
NICHOLAS STERN

Economics Letters 105 (2009) 83–85

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Contemporaneous and long run canonical correlations in the linear IV model: Implications for instrument selection

Gunce Eryuruk [a], Alastair R. Hall [b,*], Kalidas Jana [c]

[a] ITAM, Mexico
[b] Economics, School of Social Sciences, University of Manchester, Manchester M13 9PL, UK
[c] University of Texas at Brownsville, United States

ARTICLE INFO

Article history:
Received 18 July 2008
Received in revised form 18 May 2009
Accepted 10 June 2009
Available online 21 June 2009

Keywords:
Contemporaneous canonical correlations
Long run canonical correlations
Instrument selection
Two-stage least squares

JEL classification:
C13
C15
C30

ABSTRACT

In the normal linear simultaneous equations model, we demonstrate a close relationship between two recently proposed methods of instrument selection by presenting a fundamental relationship between the two sets of canonical correlations upon which the methods are based.

1. Introduction

Ordinary Least Squares (OLS) estimation yields inconsistent estimators of the parameters of linear regression models when 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. In the past, such choices have been informal at best. Recently, a number of formal criteria have been proposed in the literature to remedy that problem. This paper relates to two among these proposed criteria, namely, the Canonical Correlations Information Criterion (CCIC) of (Hall and Peixe, 2003) and the Relevant Moments Selection Criterion (RMSC) of Hall et al. (2007). The objective of Hall and Peixe (2003) and Hall et al. (2007) 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 on certain canonical correlations: CCIC exploits explicitly the canonical correlations (CCs) between the regressors and instruments; RMSC exploits implicitly the long run canonical correlations (LRCCs) between the unknown true score vector and the product of the instrument vector and error.

In this paper, we establish an interesting relation between LRCCs and CCs in a linear simultaneous equations model that helps explain the connection between CCIC and RMSC in this model. We further use the aforementioned result to reveal an interesting structure to the information measure that underlies RMSC, and also to relate RMSC to CCIC.

2. Canonical correlations and information in IV estimation

It is noted in the Introduction that, in the linear model, RMSC implicitly exploits the information in the LRCCs between the score and product of the instrument and error. In this section, we derive an explicit representation for these LRCC's in the linear simultaneous equations model with normal errors. This representation turns out to involve the CCs between the regressors and instruments, and we explore its implications for the information metric upon which RMSC is based.

To begin, it is useful to formally define CCs and LRCCs.[1]

Definition 1. Canonical correlations

* Corresponding author. Tel.: +44 1612754875; fax: +44 1612754812.
E-mail address: alastair.hall@manchester.ac.uk (A.R. Hall).

0165-1765/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2009.06.009

[1] CCs are introduced by Hotelling (1935); LRCCs are introduced (to our knowledge) in Hall et al. (2007). Note that, for ease of presentation, it is taken for granted in Definitions 1 and 2 that all expectations and inverses exist.

Let $v_{i,t}$ be $k_i \times 1$ vectors for $i = 1,2$ and $m = \min(k_1, k_2)$. Let $v_t = (v_{1,t}', v_{2,t}')'$, $\{v_t; t = 1, 2, \ldots T\}$, be a covariance stationary process with $\Sigma_v = \mathrm{Var}[v_t]$ where

$$\Sigma_v = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}$$

in the obvious notation. The population canonical correlations between $v_{1,t}$ and $v_{2,t}$ are denoted by $\{r_i; i = 1,2,\ldots,m\}$, where by convention $r_i \geq 0$ for $i = 1,2\ldots,m$ and $r_i \geq r_{i+1}$ for $i = 1,2\ldots,m-1$, and have the following properties: (i)$\{r_i^2\}$ are the m largest solutions to the determinantal equation $|\sum_{1,2}\sum_{2,2}^{-1}\sum 2,1 - r^2\sum_{1,1}| = 0$; (ii) $r_i = a_i'\sum_{1,2}b_i$ where a_i and b_i satisfy $(\sum_{1,2}\sum_{2,2}^{-1}\sum_{2,1} - r_i^2\sum_{1,1})a_i = 0$ and $(\sum_{2,1}\sum_{1,1}^{-1}\sum_{1,2} - r_i^2\sum_{2,2})b_i = 0$ for $i = 1,2\ldots, m$.[2]

Definition 2. Long run canonical correlations

Let v_t be as in Definition 1 and define $\Omega_v = \lim_{T\to\infty}\mathrm{Var}[T^{-1/2}\sum_{t=1}^{T} v_t]$ where

$$\Omega_v = \begin{bmatrix} \Omega_{1,1} & \Omega_{1,2} \\ \Omega_{2,1} & \Omega_{2,2} \end{bmatrix}$$

in the obvious notation. The long run population canonical correlations between v_{1t} and v_{2t} are denoted by $\{\rho_i; i = 1,2,\ldots, m\}$ and satisfy the equations given in Definition 1(i) and (ii) when $\Sigma_{i,j}$ is replaced by $\Omega_{i,j}$ for all $i,j = 1,2$.

A comparison of Definitions 1 and 2 indicates that the LRCCs between $v_{1,t}$ and $v_{2,t}$ are the limiting CCs between $T^{-1/2}\sum_{t=1}^{T} v_{1,t}$ and $T^{-1/2}\sum_{t=1}^{T} v_{2,t}$.

For the rest of this paper, we consider the following linear simultaneous equations model

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

$$\underset{(p\times1)}{x_t} = \underset{(p\times q)}{\pi_0}\ \underset{(q\times1)}{z_t} + \underset{(p\times1)}{e_t}, \quad t = 1, 2, \ldots, T$$

where

$$\begin{pmatrix} u_t \\ e_t \end{pmatrix} |z_t \underset{\sim}{iid}\, N\Big(\underset{(p+1)\times1}{0}, \underset{(p+1)\times(p+1)}{\Sigma}\Big), \tag{2}$$

with

$$\underset{(p+1)\times(p+1)}{\sum} = \begin{pmatrix} \underset{(1\times1)}{\sigma_u^2} & \underset{(1\times p)}{\sum_{ue}} \\ \underset{(p\times1)}{\sum_{eu}} & \underset{(p\times p)}{\sum_{ee}} \end{pmatrix} \tag{3}$$

It is mentioned above that the LRCCs of interest involve the score function. Following (Hall et al., 2007) the relevant score is the one associated with the model in Eqs. (1)–(3) for the case in which the only unknown parameters are contained in θ_0. Accordingly, we define $w_t = (y_t, x_t')'$ and $p(w_t|z_t,\theta)$ to be the conditional probability density function (pdf) of w_t given z_t implied by Eqs. (1)–(3) assuming Σ and π_0 are known, and $s_t(\theta_0)$ to be the associated score with respect to θ evaluated at $\theta = \theta_0$, that is $s_t(\theta_0) = \{\partial \ln[p(w_t|z_t,\theta_0)] / \partial\theta\}|_{\theta=\theta_0}$. Let $\{\rho_i^2, i = 1, 2, \ldots, p\}$ denote the population squared LRCCs between the score vector $S_t(\theta_0)$ and the vector $z_t u_t(\theta_0)$.

In addition to the conditions above, we impose the following regularity conditions.

Assumption 1. (i) θ_0 is an interior point of the parameter space Θ; (ii) z_t is covariance stationary, weakly exogenous for estimation of θ_0, and $E[z_t z_t']$ exists, is finite and is positive definite; (iii) $\pi_0'\pi_0 < \infty$; (iv)Σ is finite and positive definite.

The following theorem specifies the form of the aforementioned LRCCs in this model.

Theorem 1. *Let the data be generated by Eqs. (1)–(3) and Assumption 1 hold. Then the population squared LRCCs between the score vector $S_t(\theta_0)$ and the vector $z_t u_t(\theta_0)$ are given by*

$$\rho_i^2 = (1 - R_{u,e}^2)\, r_i^2(z), \; i = 1, 2, \ldots, p, \tag{4}$$

where $R_{u,e}^2$ is the population squared multiple correlation coefficient between the structural equation error u_t and the reduced form equation error e_t, and $r_i^2(z)$ is the ith population squared CC between x_t and z_t.

The proof is available upon request. We now explore some interesting implications of Theorem 1.

Remark 1. Since $0 \leq (1 - R_{u,e}^2) \leq 1$, it follows that $\rho_i^2 \leq r_i^2(z)$, $i = 1, 2, \ldots, p$. In other words, the population squared LRCCs between the score vector $s_t(\theta_0)$ and the vector $z_t u_t(\theta_0)$ are bounded by the corresponding population squared canonical correlation between x_t and z_t. Furthermore, *ceteris paribus*, the LRCCs are positively related to the CCs between x_t and z_t but negatively related to the multiple correlation between u_t and e_t.

Remark 2. If $p = 1$—and hence x_t is a scalar—then $R_{u,e}^2$ is the squared correlation between the structural equation error, u_t, and the scalar reduced form error, e_t, and $1 - R_{u,e}^2$ can be interpreted as a measure of endogeneity of x_t. For, if $1 - R_{u,e}^2 = 1$, then u_t and e_t, are uncorrelated and hence x_t is exogenous; if $1 - R_{u,e}^2 < 1$, then u_t and e_t are correlated and hence x_t is endogenous. In this case, it can be shown that, *ceteris paribus*, the correlation between x_t and u_t is positively related to $R_{u,e}^2$. Therefore, the magnitudes of the LRCCs are inversely related to the degree of endogeneity of the regressor, *ceteris paribus*.[3]

Remark 3. Hall et al. (2007) argue that the entropy of the limiting distribution of the GMM—or in this case IV—estimator provides a natural metric for the information about the parameters in moment based estimation. They further show that this entropy only depends on the choice of moment condition via the LRCCs between the score and the function of the data and parameter vector appearing in the moment condition. Combining Hall et al. (2007) [Eq. (6)] with Eq. (4), it follows that, in the model considered here, the entropy of the limiting distribution of the IV estimator of θ_0, $ent_\theta(z)$ say, has the following structure

$$ent_\theta(z) = 0.5p[1 + ln(2\pi)] + 0.5ln[|\mathcal{I}_\theta^{-1}|] - 0.5pln[1 - R_{u,e}^2] \tag{5}$$
$$- 0.5\sum_{i=1}^{p} ln[r_i^2(z)].$$

Note that the first two terms on the right hand side of Eq. (5) represent the entropy of the MLE based on the true pdf $p(w_t|z_t,\theta)$, that is the GMM estimator based on the true score. This MLE is, of course, infeasible because in general $\sum$ and π_0 are unknown; however, this infeasible MLE can be regarded as optimal within the class of GMM estimators of θ_0.[4] Since squared correlations lie in the unit interval, the third and fourth terms are positive and so represent the efficiency loss relative to this aforementioned infeasible optimal estimator. Interestingly, Eq. (5) reveals that this efficiency loss depends positively on the multiple correlation between u_t and e_t and negatively on the CCs between x_t and z_t.

Remark 4. The results in Theorem 1 point to a relationship between two recently proposed methods of instruments selection, the CCIC of (Hall and Peixe, 2003) and the RMSC of Hall et al. (2007). Both criteria are designed to eliminate redundant moment conditions based on certain canonical correlations: CCIC exploits explicitly the canonical

[2] Recall that the linear combinations are chosen so as to normalize the variances to one, that is, $a_i'\sum_{1,1}a_i = b_i'\sum_{2,2}b_i = 1$.

[3] Note that $1 - R_{u,e}^2 = 0$ is ruled out by Assumption 1(iv).

[4] Optimal in the sense that it is consistent and asymptotically efficient.

correlations (CCs) between the regressors and instruments; RMSC exploits implicitly the long run canonical correlations (LRCCs) between the unknown true score vector and the product of the instrument vector and error. The results in Theorem 1 indicate that, for the model under consideration here, the two criteria are fundamentally linked as they are both driven by the canonical correlations between the regressors and instruments. Using $p=1$, that is, only a single endogenous regressor in models (1)–(3) considered above, this fundamental link can be easily seen as follows.

CCIC of (Hall and Peixe, 2003) is defined to be

$$CCIC(c) = \Xi_T(c) + P(T, |c|) \tag{6}$$

where the statistic

$$\Xi_T(c) = T\sum_{i=1}^{p} ln[1-r_{i,T}^2(c)] \tag{7}$$

captures the sample information; the $q\times 1$ selection vector c denotes, in the notation of (Andrews, 1999), which elements of the instrument vector z_t are included in a particular moment condition: if $c_j=1$ then the jth element of z_t is included, if $c_j=0$ then the jth element of z_t is excluded; $|c|=c'c$ equals the number of elements in the instrument vector $z_t(c)$ and $P(T,|c|)$ is a "penalty" term.

If the regressor x_t is a scalar, CCIC involves only one sample squared canonical correlation, r_T^2, which is equal to the squared multiple correlation coefficient, also commonly known as the coefficient of determination. Specializing the definitions in Eq. (6) to the case of a single endogenous regressor and the penalty term associated with Bayesian Information Criterion, CCIC takes the form

$$CCIC(c) = T\,ln[1-r_T^2(c)] + (|c|-1)\,lnT. \tag{8}$$

RMSC of (Hall et al., 2007) is given by

$$RMSC(c) = ln[|\hat{V}_{\theta,T}(c)|] + \kappa(|c|, T) \tag{9}$$

where $\hat{V}_{\theta,T}(c)$ denotes a consistent estimator of the asymptotic variance $V_\theta(c)$ of the GMM estimator $\hat{\theta}_T(c)$ and $\kappa(|c|,T)$ is a deterministic penalty function. Specializing Eq. (9) to our simple linear IV model yields RMSC criterion

$$RMSC(c) = ln\left(|\hat{\sigma}_u^2[\sum_{t=1}^{T} x_t z_t(c)'\{\sum_{t=1}^{T} z_t(c) z_t(c)'\}^{-1}\sum_{t=1}^{T} z_t(c) x_t(c)']^{-1}| \right) + \frac{(|c|-1)}{\sqrt{T}}\,ln\sqrt{T}$$

Using similar arguments to those behind Theorem 1, it can be shown that

$$RMSC(c) = b_T - ln[r_T^2(c)] + \frac{(|c|-1)}{\sqrt{T}}\,ln\sqrt{T} \tag{10}$$

where b_t is $O_p(1)$ and independent of c. It can be seen from Eq. (10) that instrument selection based on RMSC depends purely on the data via the squared canonical correlation which in this case is the squared multiple correlation coefficient, $r_T^2(c)$. Thus both CCIC and RMSC base instrument selection on $r_T^2(c)$ but this correlation is compared to a different deterministic function in each case. For example, suppose we consider the choice between two instrument vectors: $z_t(c_1)$ and $z_t(c_2)$ where the latter includes all the instruments in the former plus one more. The vector $z_t(c_2)$ is selected over $z_t(c_1)$ according to CCIC if

$$\frac{1-r_T^2(c_2)}{1-r_T^2(c_1)} < exp\{-[\ln T]/T\} \tag{11}$$

and $z_t(c_2)$ is selected over $z_t(c_1)$ according to RMSC if

$$\frac{r_T^2(c_2)}{r_T^2(c_1)} > exp\{0.5[lnT]\}/\sqrt{T}\}. \tag{12}$$

Calculations based on Eqs. (11) and (12) reveal the following about the relative properties of the two criteria in this setting for $30<T<1000$. If $r_T^2(c_1)\in\{0.01, 0.05\}$ then as $\eta = r_T^2(c_2)-r_T^2(c_1)$ increases from zero we can divide the range of η in three intervals: small values ($\eta\in[0,n_1)$) for which neither *CCIC* nor *RMSC* indicate the additional instrument should be included; then a range of values ($\eta\in[n_1,n_2)$) for which *RMSC* indicates that the additional instrument should be included but *CCIC* does not; and finally a range of values ($\eta>n_2$) for which both criteria indicate the additional instrument should be included. If $r_T^2 = 0.1$ then the range of η can again be divided into three intervals but this time the qualitative decision in the middle range depends on T: for $T<266$ the middle range involves values of η for which *RMSC* indicates inclusion of the additional instrument but *CCIC* does not, but for $T\geq 266$ this is reversed. For $r_T^2\in\{0.25, 0.5\}$ and $T>30$, the middle range involves values of η for which *CCIC* indicates inclusion but *RMSC* does not. These values of $\{n_i\}$ depend on both T and $r_T^2(c_1)$. We note that for $r_T^2(c_1)\in\{0.01, 0.05\}$, there are values of T greater than 1000 at which the decision in the middle range is reversed. However, for these samples sizes, the width of the middle range is very small. The latter reflects the consistency of both methods which implies that the first two intervals combined, $(0,n_2]$, become empty so that both methods indicate the additional instrument should be included for any $\eta>0$ and any $r_T^2(c_1)<1-\eta$ in the limit with probability one.

References

Andrews, D.W.K., 1999. Consistent moment selection procedures for generalized method of moments estimation. Econometrica 543–564.
Hall, A.R., Inoue, A., Jana, K., Shin, C., 2007. Information in generalized method of moments estimation and entropy based moment selection. Journal of Econometrics 488–512.
Hall, A.R., Peixe, F.P.M., 2003. A consistent method for the selection of relevant instruments. Econometric Reviews 3, 269–287.
Hotelling, H., 1935. The most predictable criterion. Journal of Educational Psychology 139–142.