

Contemporaneous and Long Run Canonical Correlations in the Linear IV Model: Implications for Instrument Selection

Gunce Eryuruk

North Carolina State University¹

Alastair R. Hall

University of Manchester²

and

Kalidas Jana

University of Texas at Brownsville³

July 18, 2008

¹Department of Economics, Box 8110, North Carolina State University, Raleigh, NC 27695, USA.
Email: geryuru@ncsu.edu

²Corresponding author: Economics, School of Social Studies, University of Manchester, Manchester M13 9PL, UK. E-mail: alastair.hall@manchester.ac.uk

³Department of Business Administration, University of Texas at Brownsville, 80 Fort Brown, Brownsville, TX 78520, USA. E-mail: kalidas.jana@utb.edu

Abstract

Hall, Inoue, Jana & Shin (2007) consider the problem of moment selection within the Generalized Method of Moments (GMM) framework. As part of this work, they show that the long run canonical correlations (LRCC's) between the moment condition and unknown score form a natural metric for information within the GMM framework. In their analysis, the LRCC's are only analyzed in abstract. In this paper, we consider the form of the LRCC's in the leading example of IV estimation in the linear simultaneous equations model with normal errors. We show that the LRCC's depend on the contemporaneous canonical correlation (CC's) between endogenous regressors and instruments and the degree of endogeneity of the endogenous regressors. This decomposition is used to shed light on the relationship between two recently proposed methods of instrument selection.

Key Words: Contemporaneous Canonical Correlations; Long Run Canonical Correlations; Instrument Selection; Two-Stage Least Squares.

JEL Classification: C13, C15, C30.

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 has 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 & Peixe (2003) and the Relevant Moments Selection Criterion (RMSC) of Hall, Inoue, Jana & Shin (2007). The objective of Hall & Peixe (2003) and Hall, Inoue, Jana & Shin (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 (CC's) between the regressors and instruments; RMSC exploits implicitly the long run canonical correlations (LRCC's) between the unknown true score vector and the product of the instrument vector and error.

In this paper, we establish an interesting relation between LRCC's and CC's 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 LRCC's 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 equation model with normal errors. This representation turns out to involve the CC's

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 CC's and LRCC's.¹

Definition 1 Canonical correlations

Let $v_{i,t}$ be $k_i \times 1$ vectors for $i = 1, 2$ and $m = \min(k_1, k_2)$. Let $v_t = (v'_{1,t}, v'_{2,t})'$, $\{v_t; t = 1, 2, \dots, T\}$ be a covariance stationary process with $\Sigma_v = \text{Var}[v_t]$ 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 $\{r_i; i = 1, 2, \dots, m\}$, where by convention $r_i \geq 0$ for $i = 1, 2, \dots, m$ and $r_i \geq r_{i+1}$ for $i = 1, 2, \dots, m-1$, and have the following properties: (i) $\{r_i^2\}$ are the m largest solutions to the determinantal equation $|\Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} - r^2\Sigma_{1,1}| = 0$; (ii) $r_i = a'_i\Sigma_{1,2}b_i$ where a_i and b_i satisfy $(\Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1} - r_i^2\Sigma_{1,1})a_i = 0$ and $(\Sigma_{2,1}\Sigma_{1,1}^{-1}\Sigma_{1,2} - r_i^2\Sigma_{2,2})b_i = 0$ for $i = 1, 2, \dots, m$.²

Definition 2 Long run canonical correlations

Let v_t be as in Definition 1 and define $\Omega_v = \lim_{T \rightarrow \infty} \text{Var}[T^{-1/2} \sum_{t=1}^T v_t]$ 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 $v_{1,t}$ and $v_{2,t}$ are denoted by $\{\rho_i; i = 1, 2, \dots, m\}$ and satisfy the equations given in Definition 1(i) and (ii) when $\Sigma_{i,j}$ is replaced by $\Omega_{i,j}$ for all $i, j = 1, 2$.

¹CC's are introduced by Hotelling (1935); LRCC's are introduced (to our knowledge) in Hall, Inoue, Jana & Shin (2007). Note that, for ease of presentation, it is taken for granted in Definitions 1 and 2 that all expectations and inverses exist.

²Recall that the linear combinations are chosen so as to normalize the variances to one, that is $a'_i\Sigma_{1,1}a_i = b'_i\Sigma_{2,2}b_i = 1$.

A comparison of Definitions 1 and 2 indicates that the LRCC's between $v_{1,t}$ and $v_{2,t}$ are the CC's 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

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

where

$$\begin{pmatrix} u_t \\ e_t \end{pmatrix} \Big| z_t \stackrel{iid}{\sim} N \left(\begin{pmatrix} 0 \\ \Sigma \end{pmatrix}_{(p+1) \times 1}, \begin{pmatrix} \Sigma \end{pmatrix}_{(p+1) \times (p+1)} \right), \quad (2)$$

with

$$\Sigma_{(p+1) \times (p+1)} = \begin{pmatrix} \sigma_u^2 & \Sigma_{ue} \\ \Sigma_{eu} & \Sigma_{ee} \end{pmatrix}_{\substack{(1 \times 1) & (1 \times p) \\ (p \times 1) & (p \times p)}}. \quad (3)$$

It is mentioned above that the LRCC's of interest involve the score function. Following Hall, Inoue, Jana & Shin (2007) the relevant score is the one associated with the model in (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 (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 $\{\rho_i^2, i = 1, 2, \dots, p\}$ denote the population squared LRCC's 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 LRCC's in this model.

Theorem 1 *Let the data be generated by (1)-(3) and Assumption 1 hold. Then the population squared LRCC's between the score vector $s_t(\theta_0)$ and the vector $z_t u_t(\theta_0)$ are given by*

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

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 i^{th} population squared CC between x_t and z_t .

A mathematical appendix with the proof of Theorem 1 is available upon request from the authors. We now explore some interesting implications of Theorem 1.

Remark 1: Since $0 \leq (1 - R_{u,e}^2) \leq 1$, it follows that $\rho_i^2 \leq r_i^2(z)$, $i = 1, 2, \dots, p$. In other words, the population squared LRCC's 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 LRCC's are positively related to the CC's 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 LRCC's are inversely related to the degree of endogeneity of the regressor, *ceteris paribus*.³

Remark 3: Hall, Inoue, Jana & Shin (2007) argue that the entropy of the limiting dis-

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

tribution 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 LRCC's between the score and the function of the data and parameter vector appearing in the moment condition. Combining Hall, Inoue, Jana & Shin (2007) [equation (6)] with (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

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

Note that the first two terms on the right hand side of (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 Σ 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, equation (5) reveals that this efficiency loss depends inversely on the multiple correlation between u_t and e_t and positively on the CC's 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 Canonical Correlations Information Criterion (CCIC) of Hall & Peixe (2003) and the Relevant Moment Selection Criterion of Hall, Inoue, Jana & Shin (2007). Both criteria are designed to eliminate redundant moment conditions based on certain canonical correlations: CCIC exploits explicitly the canonical correlations (CC's) between the regressors and instruments; RMSC exploits implicitly the long run canonical correlations (LRCC's) 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

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

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 the model (1)-(3) considered above, this fundamental link can be easily seen as follows.

CCIC of Hall & Peixe (2003) is defined to be

$$CCIC(c) = \Xi_T(c) + P(T, |c|) \quad (6)$$

where the statistic

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

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 j^{th} element of z_t is included, if $c_j = 0$ then the j^{th} 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 (6) to the case of a single endogenous regressor, CCIC takes the form

$$CCIC(c) = T \ln[1 - r_T^2(c)] + (|c| - 1) \ln T. \quad (8)$$

RMSC of Hall, Inoue, Jana & Shin (2007) is given by

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

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 (9) to our

simple linear IV model yields RMSC criterion

$$\begin{aligned} RMSC(c) &= \ln \left(\left| \hat{\sigma}_u^2 \left[\sum_{t=1}^T x_t z_t(c)' \left\{ \sum_{t=1}^T z_t(c) z_t(c)' \right\}^{-1} \sum_{t=1}^T z_t(c) x_t(c)' \right]^{-1} \right| \right) + \frac{(|c| - 1)}{\sqrt{T}} \ln \sqrt{T} \\ &= \ln \left(\frac{\hat{\sigma}_u^2}{\frac{x'x}{T} r_T^2(c)} \right) + \frac{(|c| - 1)}{\sqrt{T}} \ln \sqrt{T} \end{aligned} \quad (10)$$

$$= \ln \left(\frac{\hat{\sigma}_u^2}{\hat{\mu}_x^2 r_T^2(c)} \right) + \frac{(|c| - 1)}{\sqrt{T}} \ln \sqrt{T}, \quad [\text{where } \hat{\mu}_x^2 = \frac{x'x}{T}], \quad (11)$$

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

Thus, we observe that CCIC and RMSC are analytically connected through their dependence on canonical correlations between regressors and instruments.

References

- Andrews, D. W. K. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 67, 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*, 138, 488–512.
- Hall, A. R. & Peixe, F. P. M. (2003). A consistent method for the selection of relevant instruments. *Econometric Reviews*, 22(3), 269–287.
- Hotelling, H. (1935). The most predictable criterion. *Journal of Educational Psychology*, 26, 139–142.

NOT FOR PUBLICATION

Mathematical Appendix

Proof of Theorem 1

The log-likelihood function implied by the model is

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

where

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

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

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

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 any function of the conditioning variable e_t .

The population squared long run canonical correlations between the score vector $s_t(\theta_0)$ and the vector $z_t u_t(\theta_0)$ are solutions to the determinantal equation $|V_\theta^{-1} - \rho^2 I_\theta| = 0$ where

$$\begin{aligned}
V_\theta^{-1} &= \lim_{T \rightarrow \infty} E \left\{ \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right]' \right\} \left[\lim_{T \rightarrow \infty} E \left\{ \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right)' \right\} \right]^{-1} \\
&\quad \times \lim_{T \rightarrow \infty} E \left\{ \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right]' \right\} \\
&= P_1 P_2^{-1} P_1';
\end{aligned}$$

$$\begin{aligned}
P_1 &= \lim_{T \rightarrow \infty} E \left\{ \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right]' \right\}, \\
P_2 &= \lim_{T \rightarrow \infty} E \left\{ \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right)' \right\}
\end{aligned}$$

and

$$I_\theta = \lim_{T \rightarrow \infty} E \left\{ \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right)' \right\}.$$

Now

$$\begin{aligned} P_1 &= \lim_{T \rightarrow \infty} E \left\{ \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0) \right]' \right\} \\ &= \lim_{T \rightarrow \infty} E \left\{ \left[\frac{1}{\sqrt{T}} \frac{1}{\delta} \sum_{t=1}^T x_t \xi_t \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t \right]' \right\} \quad (\text{suppressing } \theta_0) \\ &= \frac{1}{\delta} \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T x_t \xi_t u'_s z'_s \right] \\ &= \frac{1}{\delta} \lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T x_t \xi_t (\Sigma_{ue} \Sigma_{ee}^{-1} e_s + \xi_s)' z'_s \right] \\ &= \frac{1}{\delta} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E[x_t \xi_t (e'_s \Sigma_{ee}^{-1} \Sigma_{eu} + \xi'_s) z'_s] \\ &= \frac{1}{\delta} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E[x_t \xi_t e'_s \Sigma_{ee}^{-1} \Sigma_{eu} z'_s + x_t \xi_t \xi'_s z'_s] \\ &= \frac{1}{\delta} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E[x_t \xi_t e'_s \Sigma_{ee}^{-1} \Sigma_{eu} z'_s] \\ &\quad + \frac{1}{\delta} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E[x_t \xi_t \xi'_s z'_s] \\ &= \frac{1}{\delta} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E[(\pi_0 z_t + e_t) \xi_t e'_s \Sigma_{ee}^{-1} \Sigma_{eu} z'_s] \\ &\quad + \frac{1}{\delta} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E[x_t \xi_t \xi'_s z'_s] \end{aligned}$$

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

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

$$\begin{aligned}
P_2 &= \lim_{T \rightarrow \infty} E[(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0))(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t u_t(\theta_0))'] \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} E \sum_{t=1}^T \sum_{s=1}^T z_t u_t u'_s z'_s \quad (\text{suppressing } \theta_0) \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E(z_t u_t u'_s z'_s) \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(u_t^2) E(z_t z'_t) \quad [z_t \text{ is independent of } u_t, \text{ and } E(u_t u'_s) = 0 \text{ for } s \neq t]
\end{aligned}$$

Again,

$$\begin{aligned}
E(u_t^2) &= E\{[E(u_t|e_t) + \xi_t]^2\} \\
&= E[\Sigma_{ue}\Sigma_{ee}^{-1}e_t + \xi_t]^2 \\
&= \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{ee}\Sigma_{ee}^{-1}\Sigma_{eu} + 2\Sigma_{ue}\Sigma_{ee}^{-1}E(e_t\xi_t) + E(\xi_t^2) \\
&= \Sigma_{ue}\Sigma_{ee}^{-1}\Sigma_{eu} + \sigma_\xi^2 \quad [\text{since } E(e_t\xi_t) = 0]
\end{aligned} \tag{12}$$

Therefore,

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

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

Finally,

$$\begin{aligned}
I_\theta &= \lim_{T \rightarrow \infty} E \left\{ \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T s_t(\theta_0) \right)' \right\} \\
&= \lim_{T \rightarrow \infty} E \left\{ \left(\frac{1}{\sqrt{T}} \frac{1}{\delta} \sum_{t=1}^T x_t \xi_t \right) \left(\frac{1}{\sqrt{T}} \frac{1}{\delta} \sum_{t=1}^T x_t \xi_t \right)' \right\} \quad (\text{suppressing } \theta_0) \\
&= \frac{1}{\delta^2} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E(x_t \xi_t \xi_s' x_s') \\
&= \frac{1}{\delta^2} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T E(\xi_t^2) E(x_t x_t') \quad [E(\xi_t \xi_s') = 0 \text{ for } t \neq s \text{ and } \xi_t \text{ is uncorrelated with } x_t] \\
&= \frac{1}{\delta^2} \lim_{T \rightarrow \infty} \frac{1}{T} T \sigma_\xi^2 \Sigma_{xx} \\
&= \frac{\sigma_\xi^2}{\delta^2} \Sigma_{xx}
\end{aligned}$$

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

Hence,

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

Therefore,

$$\begin{aligned} |V_\theta^{-1} - \rho^2 I_\theta| &= 0 \\ \Rightarrow |\Sigma_{xz} \Sigma_{zz}^{-1} \Sigma_{zx} - \rho^2 \frac{1}{k} \Sigma_{xx}| &= 0 \end{aligned}$$

where

$$k = \frac{\sigma_\xi^2}{\Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu} + \sigma_\xi^2}. \quad (13)$$

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 $s_t(\theta_0)$ and the vector $z_t u_t(\theta_0)$ are given by the solutions of the determinantal equation

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

Returning to k , we have

$$\begin{aligned} k &= \frac{\sigma_\xi^2}{\Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu} + \sigma_\xi^2} \\ &= \frac{\sigma_u^2 - \Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu}}{\Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu} + \sigma_u^2 - \Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu}} \quad [\text{using (12)}] \\ &= \frac{\sigma_u^2 - \Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu}}{\sigma_u^2} \\ &= 1 - \frac{\Sigma_{ue} \Sigma_{ee}^{-1} \Sigma_{eu}}{\sigma_u^2} \\ &= 1 - R_{u,e}^2 \end{aligned}$$

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 . Hence the proof is complete. ■