ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Assessing the Magnitude of the Concentration Parameter in a Simultaneous Equations Model

D S Poskitt and C L Skeels

December 2004

Working Paper 29/04

Assessing the Magnitude of the Concentration Parameter in a Simultaneous Equations Model

D S Poskitt
Department of Econometrics
and Business Statistics
Monash University
Vic 3800, Australia

Don.Poskitt@buseco.monash.edu.au

C L Skeels
Department of Economics
The University of Melbourne
Vic 3010, Australia

Chris.Skeels@unimelb.edu.au

JEL: C12,C39,C52

Proposed running head:

Assessing the Concentration Parameter

Corresponding author:

Dr CL Skeels
Department of Economics
The University of Melbourne
Vic 3010, Australia

email: Chris.Skeels@unimelb.edu.au
telephone: +61-3-8344 3783
facsimile: +61-3-8344 6899

Abstract:

Poskitt and Skeels (2003) provide a new approximation to the sampling distribution of the IV estimator in a simultaneous equations model. This approximation is appropriate when the concentration parameter associated with the reduced form model is small and a basic purpose of this paper is to provide the practitioner with a method of ascertaining when the concentration parameter is small, and hence when the use of the Poskitt and Skeels (2003) approximation is appropriate. Existing procedures tend to focus on the notion of correlation and hypothesis testing. Approaching the problem from a different perspective leads us to advocate a different statistic for use in this problem. We provide exact and approximate distribution theory for the proposed statistic and show that it satisfies various optimality criteria not satisfied by some of its competitors. Rather than adopting a testing approach we suggest the use of p-values as a calibration device.

Key Words: Concentration parameter, simultaneous equations model, alienation coefficient, Wilks-lambda distribution, admissible invariant test.

1 Introduction

In a recent paper Poskitt and Skeels (2003) present a new approximation to the exact sampling distribution of the instrumental variables (IV) estimator of the coefficients on the endogenous regressors in a single equation from a linear system of simultaneous equations. More specifically, they examine the properties of the two-stage least squares estimator and show that when the non-centrality, or concentration, parameter associated with the reduced form model is small then certain functions of the IV estimator can be closely approximated by various t-distributions. These distributions are different, in general, from those that have previously appeared in the literature; see, for example, Phillips (1980, p. 870). A feature of the approximation is that it proves to be remarkably accurate for the situations for which it is designed, despite the simplicity of its functional form. Thus the approximation is easy for practitioners to implement and potentially useful for empirical work. A basic purpose of this paper is to provide the practitioner with a method of ascertaining when the concentration parameter is small and hence when the use of the Poskitt and Skeels (2003) approximation is appropriate.

An interesting feature of the Poskitt and Skeels (2003) approximation is its ability to capture many of the stylized facts that have been obtained under the different paradigms used to analyze weak identification and the related issue of weak instruments.[1] Although there appears to be no universally agreed definition of what exactly constitutes weakness, the consensus that emerges from the literature is that weakness manifests itself in the concentration parameter — or more correctly, some function of the concentration parameter — being small, and that this has a deleterious effect on many standard techniques of inference that cannot be ignored.

In the case of a single endogenous regressor some find it appealing to re-scale the concentration parameter by the degree of over-identification, for when the IV estimate is viewed as a two-stage least squares procedure such a re-scaling invites analogy to an F-statistic computed from the first stage regression; see Bound, Jaeger, and Baker (1995). An almost inevitable consequence of this analogy is that discussions of IV estimation and instrument relevance have focused around the perception that variables used as instruments should be highly correlated with the variables that they replace. Two measures that have been developed in this vane, and which have found common acceptance in the literature, are the partial R^2 statistics proposed by Bound et al. (1995) and Shea (1997), and other statistics that may be thought of in the same light are those explored by Cragg and Donald (1993) and Hall, Rudebusch, and Wilcox (1996).

In this paper we adopt a rather different perspective. We are motivated by two notions: First, the Poskitt and Skeels (2003) approximation to the distribution of the IV estimator is designed to work well when the concentration parameter is small and hence is applicable under circumstances that differ significantly from those for which the standard asymptotic normal approximation and Edgeworth type expansions of the distribution of the IV estimator, see Rothenberg (1984). Second, the practitioner will be faced with given endogenous and exogenous variables, dictated by the underlying economic model, and may have little control over the instrument set available. Any inference based on the IV estimate that the applied worker conducts will therefore have to be tailored to the structure of the model and the data set at hand. Hence we seek a reliable statistical measure that will characterize the magnitude of the concentration parameter, and that can be used to guide subsequent inference, and we propose the use of a *partial coefficient of alienation*, denoted $\mathcal{A}^2$, a partial version of the vector alienation coefficient introduced by Hotelling (1936).

The structure of the remainder of the paper is as follows. In the next section we outline the model and present our basic notation and assumptions. In Section 3 we provide the definition of $\mathcal{A}^2$ in the context of models containing arbitrary numbers of endogenous regressors and we also show that it is possible to provide an analytical exploration of its sampling distribution. This enables us to analyze the relationship of $\mathcal{A}^2$ to the magnitude of the concentration parameter and in Subsection 4.1 we develop appropriate inferential procedures and construct a probabilistic calibration device. Methods of approximating the sampling distribution of $\mathcal{A}^2$ based on standard distributions that facilitate implementation using commonly available software are presented in Subsection 4.2. In Section 5 we provide a multivariate version of the partial R^2 statistics proposed by Bound et al. (1995) and Shea (1997) which is applicable when there is more than one endogenous regressor in the equation of interest. We denote this measure by $\mathcal{R}^2$ and establish the finite sample distribution of $\mathcal{R}^2$ under the assumptions of this paper. Section 6 discusses the relationships between $\mathcal{A}^2$, $\mathcal{R}^2$ and canonical correlations. It also presents a comparison with the statistics of Cragg and Donald (1993) and Hall et al. (1996). Section 7 develops a likelihood ratio interpretation of $\mathcal{A}^2$ and shows that $\mathcal{A}^2$ possess desirable optimality properties. Section 8 presents a brief conclusion.

2 The Model, Notation and Assumptions

Consider the classical structural equation model

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{u}\,, \qquad \mathbf{u} \sim N(\mathbf{0}, \sigma_{\mathrm{u}}^2 \mathbf{I}_T) \tag{1}$$

where the endogenous matrix variables $\mathbf{y}$ and $\mathbf{Y}$ are $T \times 1$ and $T \times n$, respectively, the matrix of exogenous variables $\mathbf{X}$ is $T \times k$, and $\mathbf{u}$ denotes a $T \times 1$ vector of stochastic disturbances. The vectors of structural coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are $n \times 1$ and $k \times 1$, respectively. If we define $[\mathbf{X}\ \mathbf{Z}]$ to be the $T \times K$ instrument set, where $\mathbf{Z}$ denotes a $T \times \nu$ matrix of instruments —

exogenous regressors not appearing in equation (1) — and $K = k + \nu$, then we are interested in making inferences about $\boldsymbol{\beta}$ using the IV estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{P}\mathbf{y}, \tag{2}$$

where $\mathbf{P} = \mathbf{P}_{[\mathrm{X\ Z}]} - \mathbf{P}_X = \mathbf{R}_{\mathrm{X}} - \mathbf{R}_{[\mathrm{X\ Z}]}$ and, for any $N \times q$ matrix $\mathbf{A}$ of full column rank, $\mathbf{P}_{\mathrm{A}}$ denotes the idempotent, symmetric matrix $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ and $\mathbf{R}_{\mathrm{A}} = \mathbf{I}_N - \mathbf{P}_{\mathrm{A}}$. $\mathbf{P}_{\mathrm{A}}$ is the $N \times N$ (prediction) operator of rank q that projects on to the space spanned by the columns of $\mathbf{A}$ and $\mathbf{R}_{\mathrm{A}}$ is the associated (residual) operator of rank $N - q$ which projects on to the orthogonal complement of that space. In our case we can assume, without loss of generality, that the exogenous regressors and the instruments contain no redundancies, so that $[\mathbf{X}\ \mathbf{Z}]$ has full column rank, $\rho\{[\mathbf{X}\ \mathbf{Z}]\} = K$, almost surely, and

$$\mathbf{P} = \mathbf{R}_{\mathrm{X}}\mathbf{Z}(\mathbf{Z}'\mathbf{R}_{\mathrm{X}}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_{\mathrm{X}}$$

is a $T \times T$ matrix of rank $\nu \geq n$.

The corresponding reduced form model is

$$[\mathbf{y}\ \mathbf{Y}] = [\mathbf{X}\ \mathbf{Z}] \begin{bmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\Pi}_1 \\ \boldsymbol{\pi}_2 & \boldsymbol{\Pi}_2 \end{bmatrix} + [\mathbf{v}\ \mathbf{V}]\,. \tag{3}$$

Here the rows of the $T \times (n+1)$ matrix $[\mathbf{v}\ \mathbf{V}]$ are independent normal vectors with zero mean and common $(n+1) \times (n+1)$ covariance matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_{12} \\ \boldsymbol{\omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}, \tag{4}$$

ω_{11} scalar, so that $[\mathbf{v}\ \mathbf{V}] \sim N(\mathbf{0}, \boldsymbol{\Omega} \otimes \mathbf{I}_T)$, where $[\mathbf{v}\ \mathbf{V}]$ is partitioned conformably with $[\mathbf{y}\ \mathbf{Y}]$.[2] The components of the reduced form coefficient matrix $\boldsymbol{\Pi}$ — namely $\boldsymbol{\pi}_1$, $\boldsymbol{\Pi}_1$, $\boldsymbol{\pi}_2$ and $\boldsymbol{\Pi}_2$ — are of dimension $k \times 1$, $k \times n$, $\nu \times 1$ and $\nu \times n$, respectively. Note that, by implication, the structural variance $\sigma_{\mathrm{u}}^2 = [1, -\boldsymbol{\beta}']\boldsymbol{\Omega}[1, -\boldsymbol{\beta}']'$ and

$$[\mathbf{y}\ \mathbf{Y}] \sim N([\mathbf{X}\ \mathbf{Z}]\boldsymbol{\Pi}, \boldsymbol{\Omega} \otimes \mathbf{I}_T)\,. \tag{5}$$

It is easily shown that (5) implies that

$$\mathbf{S} = [\mathbf{y}\ \mathbf{Y}]'\mathbf{P}[\mathbf{y}\ \mathbf{Y}] \sim \mathrm{W}_{n+1}(\nu, \boldsymbol{\Omega}, \nu\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-\frac{1}{2}})\,, \tag{6}$$

where

$$\boldsymbol{\Delta} = \nu^{-1}[\boldsymbol{\pi}_2'\ \boldsymbol{\Pi}_2]'\mathbf{Z}'\mathbf{R}_X\mathbf{Z}[\boldsymbol{\pi}_2\ \boldsymbol{\Pi}_2] \tag{7}$$

and $\boldsymbol{\Omega} = \boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Omega}^{\frac{1}{2}}$, with $\boldsymbol{\Omega}^{\frac{1}{2}}$ the symmetric square root of $\boldsymbol{\Omega}$.[3] That is, $\mathbf{S}$ has a non-central Wishart distribution with ν degrees of freedom, covariance matrix $\boldsymbol{\Omega}$ and non-centrality parameter $\nu\boldsymbol{\Omega}^{-\frac{1}{2}}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-\frac{1}{2}}$. Since we are assuming that the usual compatibility condition

$$\boldsymbol{\pi}_2 = \boldsymbol{\Pi}_2\boldsymbol{\beta}$$

holds, it follows that

$$\mathbf{\Delta} = \begin{bmatrix} \delta_{11} & \boldsymbol{\delta}_{12} \\ \boldsymbol{\delta}_{21} & \mathbf{\Delta}_{22} \end{bmatrix} = \nu^{-1}[\boldsymbol{\beta}, \mathbf{I}_n]'\mathbf{\Pi}_2'\mathbf{Z}'\mathbf{R}_{\mathrm{X}}\mathbf{Z}\mathbf{\Pi}_2[\boldsymbol{\beta}, \mathbf{I}_n]\,, \tag{8}$$

where the partition of $\mathbf{\Delta}$ occurs after the first row and column.

Following standard practice we refer to $\mathbf{\Gamma}_{22} = \nu\mathbf{\Omega}_{22}^{-\frac{1}{2}}\mathbf{\Delta}_{22}(\mathbf{\Omega}_{22}^{-\frac{1}{2}})'$ as the concentration parameter. The importance of the magnitude of $\mathbf{\Gamma}_{22}$ for the sampling behaviour of $\widehat{\boldsymbol{\beta}}$ has been well documented in the literature — see, *inter alia*, Mariano (1982, Sections 3 and 4) and Phillips (1983, Section 3.6) — and, using the result in (6), Poskitt and Skeels (2003) show that if $\mathbf{\Delta}$ is small then the distribution of $\widehat{\boldsymbol{\beta}}$ can be closely approximated by an n-variate t-distribution with $\nu - n + 1$ degrees of freedom, mean vector

$$\boldsymbol{\beta} + (\mathbf{\Omega}_{22}^{-\frac{1}{2}})'(\mathbf{I}_n + \nu^{-1}\mathbf{\Gamma}_{22})^{-1}\boldsymbol{\rho}\sigma_{\mathrm{u}}$$

and precision

$$(\nu - n - 1)\frac{\mathbf{\Omega}_{22}^{\frac{1}{2}}(\mathbf{I}_n + \nu^{-1}\mathbf{\Gamma}_{22})(\mathbf{\Omega}_{22}^{\frac{1}{2}})'}{\sigma_{\mathrm{u}}^2(1 - \boldsymbol{\rho}'(\mathbf{I}_n + \nu^{-1}\mathbf{\Gamma}_{22})^{-1}\boldsymbol{\rho})}\,,$$

where $\boldsymbol{\rho} = \mathbf{\Omega}_{22}^{-\frac{1}{2}}(\boldsymbol{\omega}_{21} - \mathbf{\Omega}_{22}\boldsymbol{\beta})/\sigma_{\mathrm{u}}$. Clearly this approximation to the distribution of the IV estimator is no more difficult to implement than is the Normal approximation that arises in standard asymptotic analysis, but the practitioner needs to be appraised of the likely value of $\mathbf{\Delta}$ before it is employed. To relate the magnitude of $\mathbf{\Delta}$ to the concentration parameter note from expression (8) that

$$\|\mathbf{\Delta}\| \leq (\|\boldsymbol{\beta}\|^2 + n)\|\mathbf{\Delta}_{22}\|$$

where, for any matrix $\mathbf{A}$, we have used $\|\mathbf{A}\| = \sqrt{\mathrm{tr}\{\mathbf{A}'\mathbf{A}\}}$ to denote the Euclidean norm. If we assume that $0 < \mathbf{\Omega} < \infty$, meaning $0 < \lambda_{\min}(\mathbf{\Omega}) \leq \lambda_{\max}(\mathbf{\Omega}) < \infty$, it is clear that small values of $\mathbf{\Gamma}_{22}$ and $\mathbf{\Delta}_{22}$, and hence $\mathbf{\Delta}$, are equivalent.

3 A Formal Measure of $\mathbf{\Gamma}_{22}$ and Its Distribution

To construct a measure of the magnitude of $\mathbf{\Gamma}_{22}$, first consider the reduced form for the endogenous regressors in the equation of interest, namely

$$\mathbf{Y} = \mathbf{X}\mathbf{\Pi}_1 + \mathbf{Z}\mathbf{\Pi}_2 + \mathbf{V}\,. \tag{9}$$

Pre-multiplying by $\mathbf{R}_{\mathrm{X}}$, taking the Euclidean norm and evaluating the expectation, gives us the result that

$$\mathsf{E}[\|\mathbf{R}_{\mathrm{X}}\mathbf{Y}\|^2] = \|\mathbf{R}_{\mathrm{X}}\mathbf{Z}\mathbf{\Pi}_2\|^2 + (T - k)\mathrm{tr}\mathbf{\Omega}_{22}\,.$$

Thus $\|\mathbf{R}_X\mathbf{Z}\mathbf{\Pi}_2\|^2$ corresponds to the regression mean square in the regression of $\mathbf{R}_X\mathbf{Y}$ on $\mathbf{R}_X\mathbf{Z}$. Moreover, the inequalities

$$\|\mathbf{\Pi}_2'\mathbf{Z}'\mathbf{R}_X\mathbf{Z}\mathbf{\Pi}_2\| \leq \|\mathbf{\Omega}_{22}^{\frac{1}{2}}\|^2\|\mathbf{\Gamma}_{22}\|$$

and

$$\|\mathbf{\Gamma}_{22}\| \leq \|\mathbf{\Omega}_{22}^{-\frac{1}{2}}\|^2\|\mathbf{R}_X\mathbf{Z}\mathbf{\Pi}_2\|^2\,,$$

imply that the proximity of $\|\mathbf{\Gamma}_{22}\|$ and $\|\mathbf{R}_X\mathbf{Z}\mathbf{\Pi}_2\|$ to zero is equivalent and we therefore begin by seeking a statistical measure of $\|\mathbf{R}_X\mathbf{Z}\mathbf{\Pi}_2\|^2$.

Consider then the ratio of determinants

$$\mathcal{A}^2 = \frac{\det[\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}]}{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]}$$

where $\widetilde{\mathbf{Y}} = \mathbf{R}_X\mathbf{Y}$ and $\widetilde{\mathbf{Z}} = \mathbf{R}_X\mathbf{Z}$. The statistic $\mathcal{A}^2$ equals a partial version of the vector alienation coefficient introduced by Hotelling (1936) in the context of studying the relationships between two sets of variables, hence our notation. To interpret $\mathcal{A}^2$ observe from the equality

$$\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}} + \widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$$

that $\mathcal{A}^2 = 1$ when $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are orthogonal and $\mathcal{A}^2 = 0$ if there exists a matrix $\mathbf{D}$ of full column rank such that $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Z}}\mathbf{D}$. Thus $\mathcal{A}^2$ can be viewed as a measure of the perpendicularity between $\mathbf{Y}$ and $\mathbf{Z}$ having adjusted for the effects of $\mathbf{X}$. From the expression

$$\mathcal{A}^2 = \det[\mathbf{I}_n - (\mathbf{Y}'\mathbf{R}_X\mathbf{Y})^{-\frac{1}{2}}(\mathbf{Y}'\mathbf{P}\mathbf{Y})(\mathbf{Y}'\mathbf{R}_X\mathbf{Y})^{-\frac{1}{2}}]$$

we see that $\mathcal{A}^2$ is a sample counterpart to the population relative measure

$$\det[\mathbf{I}_n - (\mathbf{Y}'\mathbf{R}_X\mathbf{Y})^{-\frac{1}{2}}(\mathbf{\Pi}_2'\mathbf{Z}'\mathbf{R}_X\mathbf{Z}\mathbf{\Pi}_2)(\mathbf{Y}'\mathbf{R}_X\mathbf{Y})^{-\frac{1}{2}}]$$

and represents the proportion of the generalized variance of $\widetilde{\mathbf{Y}}$ that remains once the regression mean square in the multivariate regression of $\widetilde{\mathbf{Y}}$ on $\widetilde{\mathbf{Z}}$ has been accounted for.

In order to derive the distribution of $\mathcal{A}^2$, consider once again the reduced form equation (9). Pre-multiplying by $\mathbf{R}_X$ we see that

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Z}}\mathbf{\Pi}_2 + \mathbf{E}\,,$$

where $\mathbf{E} = \mathbf{R}_X\mathbf{V}$ and $\mathbf{E}|[\mathbf{X}\ \mathbf{Z}] \sim N(\mathbf{0}, \mathbf{\Omega}_{22} \otimes \mathbf{R}_X)$. Post-multiplying by the constant vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$ we obtain the equation

$$\widetilde{\mathbf{Y}}\boldsymbol{\alpha} = \widetilde{\mathbf{Z}}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{10}$$

where now $\boldsymbol{\gamma} = \mathbf{\Pi}_2\boldsymbol{\alpha}$ and $\boldsymbol{\eta}|[\mathbf{X}\ \mathbf{Z}] \sim N(\mathbf{0}, \sigma_\alpha^2\mathbf{R}_X)$, a singular normal distribution (see, for example, Rao, 1973, §8a) with $\sigma_\alpha^2 = \boldsymbol{\alpha}'\mathbf{\Omega}_{22}\boldsymbol{\alpha}$.

Now $\mathbf{P}_{\widetilde{Z}}$ and $\mathbf{R}_{\widetilde{Z}}$ are idempotent with ranks ρ and $T-\rho$, respectively, where $\rho = \rho\{\widetilde{\mathbf{Z}}\}$. Since $\widetilde{\mathbf{Z}} = \mathbf{R}_X\mathbf{Z}$ it is easily shown that $\mathbf{P}_{\widetilde{Z}}\mathbf{R}_X\mathbf{R}_{\widetilde{Z}} = \mathbf{0}$. Moreover, given that we have assumed that $[\mathbf{X}\ \mathbf{Z}]$ has full column rank, it follows that $\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}} = \mathbf{Z}'\mathbf{R}_X\mathbf{Z} > 0$ and $\rho\{\widetilde{\mathbf{Z}}\} = \rho\{\mathbf{Z}\} = \nu$ almost surely. Consequently $\rho\{\widetilde{\mathbf{Z}}\}$ is known with probability one, it is simply the number of instruments used in addition to $\mathbf{X}$, supposing that $\mathbf{X}$ is employed as its own instrument.

We can therefore conclude that, for given $\mathbf{Z}$ and $\mathbf{X}$, the quadratic forms $\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ are independently distributed as $\sigma_\alpha^2 \cdot \chi^2(\nu,\mu)$, $\mu = \boldsymbol{\gamma}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma}/\sigma_\alpha^2$, and $\sigma_\alpha^2 \cdot \chi^2(T-\nu)$ random variables, respectively. Since $\boldsymbol{\alpha}$ is arbitrary we therefore have from Rao (1973, §8b.2 (ii) & (iii)) that the matrices $\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ will have independent Wishart distributions:

$$\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}} \sim \mathcal{W}_n(\nu, \boldsymbol{\Omega}_{22}, \boldsymbol{\Gamma}_{22}) \qquad \text{and} \qquad \widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}} \sim \mathcal{W}_n(T-\nu, \boldsymbol{\Omega}_{22})\,.$$

In directions $\boldsymbol{\theta} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|$ such that $\boldsymbol{\theta}'\boldsymbol{\Pi}_2'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\Pi}_2\boldsymbol{\theta} = 0$ the non-centrality parameter $\mu = 0$ and both $\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ will have central Wishart distributions. Writing $\mathcal{A}^2$ as the ratio of $\det[\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}]$ to $\det[\widetilde{\mathbf{Y}}'(\mathbf{R}_{\widetilde{Z}}+\mathbf{P}_{\widetilde{Z}})\widetilde{\mathbf{Y}}]$ it follows that in any direction such that $\mu = 0$ the statistic $\mathcal{A}^2$ will possess Wilks'-Λ distribution

$$\Lambda(n, T-\nu, \nu) \sim \begin{cases} \prod_{i=1}^{n} \mathcal{B}(\frac{T-\nu+1-i}{2}, \frac{\nu}{2}) & \nu \geq n, \\ \prod_{i=1}^{\nu} \mathcal{B}(\frac{T-\nu-n+i}{2}, \frac{n}{2}) & \text{otherwise}, \end{cases}$$

the product of independent Beta random variables, see Wilks (1962, §18.5.1).

4 Probabilistic Calibration and Calculations

4.1 Probabilistic Calibration

We are now interested in determining the influence of $\boldsymbol{\Gamma}_{22}$ on the properties of $\mathcal{A}^2$ with a view to using $\mathcal{A}^2$ to assess the likely size of $\boldsymbol{\Gamma}_{22}$ on an appropriate (probabilistic) scale. By way of background, note that if $\boldsymbol{\theta}$ is a given unit vector and we adopt a hypothesis testing perspective, then under the null hypothesis $\mathcal{H}_0 : \boldsymbol{\Pi}_2\boldsymbol{\theta} = \mathbf{0}$ the distribution of

$$F_\theta = \left(\frac{T-\nu}{\nu}\right)\left(\frac{1-A_\theta^2}{A_\theta^2}\right)$$

where

$$A_\theta^2 = \frac{\boldsymbol{\theta}'\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}\boldsymbol{\theta}}{\boldsymbol{\theta}'\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}\boldsymbol{\theta}}$$

will be central $\mathcal{F}$ with degrees of freedom ν and $T-\nu$, whilst under the alternative hypothesis $\mathcal{H}_1 : \mathbf{\Pi}_2\boldsymbol{\theta} \neq \mathbf{0}$ the distribution of F_θ will be non-central $\mathcal{F}$ with degrees of freedom ν and $T-\nu$, and non-centrality parameter

$$\lambda_\theta = \frac{\boldsymbol{\theta}'\mathbf{\Pi}_2'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{\Pi}_2\boldsymbol{\theta}}{\boldsymbol{\theta}'\mathbf{\Omega}_{22}\boldsymbol{\theta}} = \frac{\|\widetilde{\mathbf{Z}}\mathbf{\Pi}_2\boldsymbol{\theta}\|^2}{\boldsymbol{\theta}'\mathbf{\Omega}_{22}\boldsymbol{\theta}}\,. \tag{11}$$

If $\mathbb{P}(\lim_{T\to\infty} T^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}} > 0) = 1$ then it follows via a standard argument that

$$\mathcal{CR}\{A_\theta^2, \alpha\} = \{A_\theta^2 : F_\theta > \mathcal{F}_{(1-\alpha)}\{\nu, T-\nu\}\}\,,$$

where $\mathcal{F}_{(1-\alpha)}\{\nu, T-\nu\}$ denotes the $(1-\alpha)100\%$ percentile point of the $\mathcal{F}\{\nu, T-\nu\}$ distribution, defines a strongly consistent critical region of size α for testing $\mathcal{H}_0$ against $\mathcal{H}_1$.

Now consider calculating A_θ^2 and then computing the associated p-value $p_\theta = \mathbb{P}(\mathcal{F}\{\nu, T-\nu\} > F_\theta)$ as a means of assessing whether λ_θ is small or large. Equation (11) indicates that when λ_θ is large the signal-to-noise ratio in the implied model is high and we can expect $\boldsymbol{\theta}'\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}\boldsymbol{\theta}$ to be close to its upper bound of $\boldsymbol{\theta}'\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}\boldsymbol{\theta}$. Thus large values of λ_θ will correspond to situations where we can expect that $A_\theta^2 \approx 0$ and $p_\theta \ll \alpha$. Similarly, the case where λ_θ is close to zero corresponds to situations where A_θ^2 will approximately equal one with high probability and $p_\theta \gg \alpha$. We are not directly concerned with drawing sharp distinctions between data sets where $p_\theta \ll \alpha$ and $p_\theta \gg \alpha$ in order to make explicit decisions about the acceptance or rejection of $\mathcal{H}_0$ *vis-a-vis* $\mathcal{H}_1$. Nevertheless, it is clear that p_θ yields a probability scale that differentiates realizations that are indicative of directions in which λ_θ is large from those that suggest that λ_θ is small.

To relate the previous argument to $\mathcal{A}^2$ note from the implicit model (10) that the presumption that $\mathcal{H}_0$ obtains *for all* $\boldsymbol{\theta}$ is equivalent to the statement that $\widetilde{\mathbf{Y}} \perp \widetilde{\mathbf{Z}}$, wherein we have employed the notation $\widetilde{\mathbf{Y}} \perp \widetilde{\mathbf{Z}}$ as a 'shorthand' for $\mathbb{P}(\lim_{T\to\infty} \|T^{-1}\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Z}}\| > 0) = 0$, meaning that $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are (asymptotically) orthogonal. This follows since $\mathsf{E}[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Z}}] = \mathbf{0}$ implies $\boldsymbol{\gamma} = \mathbf{\Pi}_2\boldsymbol{\alpha} = \|\boldsymbol{\alpha}\|\mathbf{\Pi}_2\boldsymbol{\theta} = \mathbf{0}$ and, conversely, $\mathsf{E}[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Z}}] = \mathsf{E}[\boldsymbol{\eta}'\widetilde{\mathbf{Z}}]$ when $\boldsymbol{\gamma} = \mathbf{0}$ and $\mathsf{E}[\boldsymbol{\eta}'\widetilde{\mathbf{Z}}] = \mathbf{0}$. Thus Wilks'-$\Lambda$ distribution can be used to assess the significance of departures of the measure $\mathcal{A}^2$ from unity under the presumption that $\widetilde{\mathbf{Y}} \perp \widetilde{\mathbf{Z}}$, significantly small values of $\mathcal{A}^2$ being taken as being indicative of statistically significant departures from (asymptotic) orthogonality.

Now let $\mathbf{u}_i$, $i = 1, \ldots, n$, denote a set of orthonormal characteristic vectors of $(\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-\frac{1}{2}}\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}(\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-\frac{1}{2}}$. Then it is a relatively simple exercise in linear algebra to show that

$$\mathcal{A}^2 = \prod_{i=1}^{n} A_{\mathrm{u}_i}^2\,. \tag{12}$$

Applying the Rayleigh-Ritz theorem to $\mathbf{\Gamma}_{22}$ we can also deduce that

$$\lambda_{\min}(\mathbf{\Gamma}_{22}) \leq \frac{\boldsymbol{\theta}'\mathbf{\Pi}_2'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{\Pi}_2\boldsymbol{\theta}}{\boldsymbol{\theta}'\mathbf{\Omega}_{22}\boldsymbol{\theta}} \leq \lambda_{\max}(\mathbf{\Gamma}_{22})\,. \tag{13}$$

Suppose then that $\lambda_{\max}(\mathbf{\Gamma}_{22})$ is small. Then it is obvious from the inequality $\|\mathbf{\Gamma}_{22}\| \leq \sqrt{n} \cdot \lambda_{\max}(\mathbf{\Gamma}_{22})$ that $\mathbf{\Gamma}_{22}$ is small and (13) implies that λ_θ must also be small in all possible directions $\boldsymbol{\theta}$. It follows that we can expect that $A^2_{\mathrm{u}_i} \approx 1$ for all $i = 1, \ldots, n$ and hence that $\mathcal{A}^2 \approx 1$. On the other hand, if $\lambda_{\min}(\mathbf{\Gamma}_{22})$ is large then $\mathbf{\Gamma}_{22}$ is large, λ_θ will be large in all possible directions $\boldsymbol{\theta}$, $A^2_{\mathrm{u}_i} \approx 0$ for all $i = 1, \ldots, n$ with high probability and hence $\mathcal{A}^2 \approx 0$.

The previous analysis suggests that we consider calculating $\mathcal{A}^2$ and then computing the p-value

$$p = \mathbb{P}(\Lambda(n, T - \nu, \nu) < \mathcal{A}^2)$$

as a means of assessing whether $\mathbf{\Gamma}_{22}$ is small or large. We can anticipate that small values of p will be indicative of situations where the concentration parameter $\mathbf{\Gamma}_{22}$ is bounded away from zero and significantly large, indicating that the use of the Poskitt and Skeels (2003) approximation is likely to be inappropriate, whereas large values of p will provide evidence that $\mathbf{\Gamma}_{22}$ is near zero and use of the Poskitt and Skeels (2003) approximation is legitimate.

In order to use p to calibrate subsequent inferential statements note that the Poskitt and Skeels (2003) approximation may be viewed as providing the (approximate) probability distribution of $\widehat{\boldsymbol{\beta}}$ conditional on $\mathbf{\Gamma}_{22}$ being small, whilst p may be interpreted as giving the probability that $\mathbf{\Gamma}_{22}$ is small. To construct a $(1-\alpha)100\%$ confidence interval or test a hypothesis about $\boldsymbol{\beta}$ at the $(1-\alpha)100\%$ level of significance, therefore, the practitioner should use the $(1-\alpha')100\%$ percentile points where $\alpha' = \alpha \times p$. This will have the effect, for example, of widening the confidence interval at any given level of significance in order to reflect the uncertainty as to the size of $\mathbf{\Gamma}_{22}$. As p gets smaller and the appropriateness of the approximation is called into question, the confidence interval at any pre-assigned level of significance becomes wider.

4.2 Probability Calculations

To implement the above calibration the practitioner will need to calculate $p = \mathbb{P}(\Lambda(n, T - \nu, \nu) < \mathcal{A}^2)$. Box (1949) provides a series expansion for Wilks'-Λ distribution in terms of Chi-squared distributions and Banerjee (1958) uses Mellin transforms to construct an exact expression for the distribution of $\Lambda(n, T - \nu, \nu)$, involving sums, products and ratios of Gamma functions, that depends on whether n and ν are even or odd. Schatzoff (1966a) gives exact closed form representations applicable when n or ν is even and supplies tables of correction factors that can be used to convert Chi-squared percentile points to percentile points of $\Lambda(n, T-\nu, \nu)$ for n or ν even and $n\nu \leq 70$. In the current situation $n \geq 1$ and $\nu \geq n$. In the special case where $n = 1$ we have

$$\left(\frac{T-\nu}{\nu}\right)\left(\frac{1-\Lambda(1, T-\nu, \nu)}{\Lambda(1, T-\nu, \nu)}\right) \sim \mathcal{F}\{\nu, T-\nu\},$$

and when $n = 2$ we can use the exact result that

$$F = \left(\frac{T-\nu-1}{\nu}\right)\left(\frac{1-\mathcal{A}}{\mathcal{A}}\right) \sim \mathcal{F}\{2\nu, 2(T-\nu-1)\}$$

for any ν to calculate p. In general, however, Wilks'-Λ distribution is sufficiently complicated to make an appropriate approximation that can be easily implemented using standard software worth pursuing.

One such approximation is due to Bartlett (1947). Bartlett's results imply that

$$-m\ln(\mathcal{A}^2),$$

where

$$m = T - \frac{n+\nu+1}{2},$$

will converge in distribution to $\chi^2(n\nu)$ as $T \to \infty$. A closer asymptotic approximation correct up to terms of order $O(T^{-3})$ can be constructed using a second order version of Box's expansion and Box (1949) also presents an $\mathcal{F}$ approximation for $-m\ln(\mathcal{A}^2)$ that has a remainder term $O(T^{-3})$. Box found that the latter gives close agreement with the exact distribution even when the sample size is small, $10 \leq T \leq 20$ say.

An even more precise $\mathcal{F}$ approximation is given by Rao (1951). Rao's approximation implies that

$$F_{\mathcal{A}} = \left(\frac{ms-2q}{n\nu}\right)\left(\frac{1-\mathcal{A}^{2/s}}{\mathcal{A}^{2/s}}\right),$$

where

$$s = \sqrt{\frac{(n\nu)^2-4}{n^2+\nu^2-5}} \quad \text{and} \quad q = \frac{n\nu-2}{4},$$

may be treated as an $\mathcal{F}\{n\nu, ms-2q\}$ random variate. For practical purposes the integer part of $ms-2q$ may be taken as the denominator degrees of freedom. Not only does this approximation yield an error of order $O(T^{-4})$ but the structure of the approximation also has a certain appeal in the current circumstances since the statistic $F_{\mathcal{A}}$ can be employed to evaluate p via $\mathcal{F}\{n\nu, ms-2q\}$ and thereby calibrate $\mathcal{A}^2$ in much the same way that the $\mathcal{F}$ distribution is used in simple regression and analysis of variance models.

5 Multivariate Partial R^2

The arguments presented in Hotelling (1936) suggest that an appropriate generalization of the univariate partial R^2 of Bound et al. (1995) to the case where we are interested in studying the relationships between all the endogenous regressors in $\mathbf{Y}$ and the instruments $\mathbf{Z}$ is given by

$$\mathcal{R}^2 = \frac{\det[\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}]}{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]}, \tag{14}$$

a partial version of Hotelling's coefficient of vector correlation. Writing $\mathcal{R}^2$ as the ratio of $\det[\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}]$ to $\det[\widetilde{\mathbf{Y}}'(\mathbf{R}_{\widetilde{Z}} + \mathbf{P}_{\widetilde{Z}})\widetilde{\mathbf{Y}}]$ and recalling that the derivation surrounding equation (10) shows that when the non-centrality parameter $\mu = \boldsymbol{\alpha}'\boldsymbol{\Pi}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\Pi}\boldsymbol{\alpha}$ is zero $\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ are independently distributed as $\mathcal{W}_n(\nu, \boldsymbol{\Omega})$ and $\mathcal{W}_n(T-\nu, \boldsymbol{\Omega})$ random variables, respectively, leads to the conclusion that the statistic $\mathcal{R}^2$ will possess Wilks'-$\Lambda(n, \nu, T-\nu)$ distribution.

Wilks'-Λ distribution can therefore be used to calibrate the measure $\mathcal{R}^2$ in much the same way it is used to calibrate $\mathcal{A}^2$. If we employ Rao's $\mathcal{F}$ approximation to $\Lambda(n, \nu, T-\nu)$ then it will be large values of the statistic

$$F_{\mathcal{R}} = \left(\frac{ms - 2q}{n(T-\nu)}\right)\left(\frac{1 - \mathcal{R}^{2/s}}{\mathcal{R}^{2/s}}\right),$$

where

$$m = \frac{T - n + \nu - 1}{2}, \quad s = \sqrt{\frac{(n(T-\nu))^2 - 4}{n^2 + (T-\nu)^2 - 5}} \quad \text{and} \quad q = \frac{n(T-\nu) - 2}{4},$$

that will lend support to the hypothesis that $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are orthogonal. Large values of $\mathbb{P}(\mathcal{F}\{n(T-\nu), ms - 2q\} \geq F_{\mathcal{R}})$ will suggest that the regressors in $\widetilde{\mathbf{Z}}$ contain components that are sufficiently correlated with $\widetilde{\mathbf{Y}}$ to make $\mathcal{R}^2$ close to one, indicating that the regression mean square $\|\widetilde{\mathbf{Z}}\boldsymbol{\Pi}_2\|^2$ is in some sense large.

It is important to observe that in general $\mathcal{R}^2 \neq 1 - \mathcal{A}^2$ and so probability calculations based on $\mathcal{A}^2$ and $\mathcal{R}^2$ will not be identical. This raises the question of which measure is most appropriate for our current needs, an issue to which we will return in the following section. First we wish to consider the relationship of the statistic proposed by Shea (1997) to those considered here.

Shea (1997) motivates his statistic as being proportional to the ratio of the variance of the OLS estimate to that of the IV estimate. Let us therefore define the multivariate version of Shea's measure, which we will denote by $\mathcal{S}^2$, as the ratio of the generalized variances of the OLS and IV estimators of $\boldsymbol{\beta}$. This gives us the analogous expression

$$\mathcal{S}^2 = \frac{\det[\widehat{\mathbf{Y}}'\mathbf{R}_{\widehat{\mathbf{X}}}\widehat{\mathbf{Y}}]}{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]}, \tag{15}$$

where $\widehat{\mathbf{Y}} = \mathbf{P}_{[\mathrm{X\ Z}]}\mathbf{Y}$ and $\widehat{\mathbf{X}} = \mathbf{P}_{[\mathrm{X\ Z}]}\mathbf{X}$. In this guise the numerator of $\mathcal{S}^2$ has a form that corresponds to that of the statistic $\mathcal{A}^2$ in that it is structured in terms of the residual operator $\mathbf{R}$. The numerator of $\mathcal{S}^2$ is based on the projection of $\mathbf{Y}$ and $\mathbf{X}$ on to the space spanned by the instruments, however, rather than the residual from the projection of $\mathbf{Y}$ and $\mathbf{Z}$ onto the space spanned by $\mathbf{X}$, as is the case with $\mathcal{A}^2$. In the light of Shea's claim (Shea, 1997, §II) that his statistic is equivalent to partial R^2, it is natural at this point to enquire into the relationship between $\mathcal{A}^2$, $\mathcal{R}^2$ and $\mathcal{S}^2$.

From the decomposition $\mathbf{P}_{[\mathrm{X\ Z}]} = \mathbf{P}_{\mathrm{X}} + \mathbf{R}_{\mathrm{X}}\mathbf{Z}(\mathbf{Z}'\mathbf{R}_{\mathrm{X}}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_{\mathrm{X}}$ it follows that $\widehat{\mathbf{X}} = \mathbf{X}$ and hence that $\widehat{\mathbf{Y}}'\mathbf{R}_{\widehat{\mathbf{X}}}\widehat{\mathbf{Y}} = \widehat{\mathbf{Y}}'\mathbf{R}_{\mathrm{X}}\widehat{\mathbf{Y}}$. We can also deduce from the equality $\mathbf{R}_{\mathrm{X}}\mathbf{P}_{\mathrm{X}} = \mathbf{P}_{\mathrm{X}}\mathbf{R}_{\mathrm{X}} = \mathbf{0}$ that

$$\mathbf{P}_{[\mathrm{X\ Z}]}\mathbf{R}_{\mathrm{X}}\mathbf{P}_{[\mathrm{X\ Z}]} = \mathbf{R}_{\mathrm{X}}\mathbf{Z}(\mathbf{Z}'\mathbf{R}_{\mathrm{X}}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_{\mathrm{X}},$$

giving us the result that $\widehat{\mathbf{Y}}'\mathbf{R}_{\mathrm{X}}\widehat{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}}$. We are therefore lead to the conclusion that $\widehat{\mathbf{Y}}'\mathbf{R}_{\widehat{\mathbf{X}}}\widehat{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}}$. Thus although at first it might appear that $\mathcal{S}^2$ will behave in a manner similar to that of an alienation coefficient, the opposite will in fact be the case because $\mathcal{S}^2 = \mathcal{R}^2$.

6 Alienation, Partial R^2 and Canonical Correlation

In the light of the interpretation of $\mathcal{A}^2$ as a partial version of Hotelling's vector alienation coefficient and given that Hotelling (1936) was also the father of canonical correlation analysis it is not surprising to observe that factorizing $\det[\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}}] = \det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}}]$ into the product of $\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]$ and $\det[\mathbf{I}_n - (\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-\frac{1}{2}}\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}}(\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-\frac{1}{2}}]$ shows that

$$\mathcal{A}^2 = \prod_{i=1}^{n}(1 - r_i^2)\,, \tag{16}$$

where $r_1^2 \geq \ldots \geq r_n^2$ lists in descending order the partial canonical correlations between $\mathbf{Y}$ and $\mathbf{Z}$ having adjusted for the effects of $\mathbf{X}$.

Let us now address the question of which multivariate measure, $\mathcal{A}^2$ or $\mathcal{R}^2$, appears to be best suited our needs. Assume, for the sake of argument, that exact correlation between $\widetilde{\mathbf{Y}}$ on $\widetilde{\mathbf{Z}}$ is characterised by all the partial canonical correlations being equal to one, whereas orthogonality between $\widetilde{\mathbf{Y}}$ on $\widetilde{\mathbf{Z}}$ implies that $r_1^2 = \ldots = r_n^2 = 0$. From the expression in (16) and the corresponding representation of $\mathcal{R}^2$, namely

$$\mathcal{R}^2 = \prod_{i=1}^{n} r_i^2\,,$$

it follows that it is only necessary for the largest (smallest) partial canonical correlation to deviate substantially from zero (one) for $\mathcal{A}^2$ ($\mathcal{R}^2$) to deviate significantly from unity. Thus, whereas $\mathcal{A}^2$ will be sensitive to departures from orthogonality $\mathcal{R}^2$ is designed to detect exact correlation. Now recall that the use of Wilks'-Λ distribution as a calibration device is contingent on the non-centrality parameter being equal to zero, which we have already observed is equivalent to the hypothesis that $\widetilde{\mathbf{Y}} \perp \widetilde{\mathbf{Z}}$ and hence that $r_1^2 = \ldots = r_n^2 = 0$. It appears therefore that $\mathcal{A}^2$ is more in accord with the basic assumption underlying the application of Wilks'-Λ distribution than is $\mathcal{R}^2$. Given this feature, and given that we are not seeking to detect exact or perfect correlation but rather in assessing the proximity of $\mathbf{\Gamma}_{22}$ to zero, the measure $\mathcal{A}^2$ appears to be far more suited to our purpose.

The use of canonical correlations in the context of IV estimation and simultaneous equations has, of course, a long history dating back to the seminal works of Sargan (1958) and Hooper (1959). To relate the canonical correlations to the concepts underlying the developments in this paper let us form the linear combinations $\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and $\widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ from the adjusted variables $\mathbf{R}_X\mathbf{Y}$ and $\mathbf{R}_X\mathbf{Z}$. Given $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, the squared partial correlation

$$R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) = \frac{(\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma})^2}{(\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}\boldsymbol{\alpha})(\boldsymbol{\gamma}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma})}$$

estimates the corresponding population partial correlation coefficient and the region

$$\left\{R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) > R_c^2\right\}\,, \tag{17}$$

where R_c^2 is an appropriate critical value, denotes those values of $R^2(\boldsymbol{\alpha},\boldsymbol{\gamma})$ indicating the presence of components in $\mathbf{Y}$ and $\mathbf{Z}$ that induce a significant partial correlation. The intersection

$$\mathcal{RR} = \bigcap_{\alpha\ \gamma} \left\{R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) > R_c^2\right\}$$

of all regions of the type given in (17) across all non-null vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ corresponds to the statement that all partial correlations between $\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and $\widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ are in some sense significant.

Using the Union-Intersection principle of Roy (1957) we see that $\mathcal{RR}$ can serve as a critical region for testing the hypothesis that there is at least one pair of non-null vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ for which $\widetilde{\mathbf{Y}}\boldsymbol{\alpha} \perp \widetilde{\mathbf{Z}}\boldsymbol{\gamma}$. But the region $\mathcal{RR}$ is equivalent to that specified by

$$\min_{\alpha,\gamma\neq 0} R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) > R_c^2\,,$$

for if the smallest partial correlation between $\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and $\widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ lies in (17) then all partial correlations of such linear combinations must do so. Similarly, the region

$$\max_{\alpha,\gamma} R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) \leq R_c^2\,,$$

provides evidence against the hypothesis that there is at least one pair of non-null vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ for which the partial correlation between $\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and

$\widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ is non-zero and in favour of the hypothesis that $\widetilde{\mathbf{Y}}\boldsymbol{\alpha} \perp \widetilde{\mathbf{Z}}\boldsymbol{\gamma}$. It is a standard exercise to show that $\max_{\alpha\gamma} R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) = r_1^2$ and $\min_{\alpha\gamma} R^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) = r_n^2$. It is now natural to consider handling the intermediate extremes of $R_p^2(\boldsymbol{\alpha},\boldsymbol{\gamma})$ in a similar manner. From the Courant-Fischer theorem the extremes are equal to $r_1^2 \geq r_2^2 \geq \ldots \geq r_{n-1}^2 \geq r_n^2$.

In their discussion of identification tests Cragg and Donald (1993) point out that the coefficients in the equation of interest will be identified if and only if the rank of the coefficient matrix in the reduced form $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Z}}\mathbf{\Pi}_2 + \mathbf{E}$ equals n. A version of their procedure for testing the rank of $\mathbf{\Pi}_2$ that is "concerned with whether $X_2(\mathbf{Z})$ can serve as instruments for $Y_2(\mathbf{Y})$ in the sense that there is enough correlation" is given by (in the notation of this paper) the smallest eigenvalue of $\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}$ in the metric of $\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}$. See hypothesis H_I^0 and Theorem 3 of Cragg and Donald (1993). Using the relationship $\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}} + \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}$ gives us the expression

$$\begin{aligned}\det[\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}} - \lambda\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}] = &\\ \det[(1+\lambda)\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]\times \det[(\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-\frac{1}{2}}\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}(\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-\frac{1}{2}} - \frac{\lambda}{1+\lambda}\mathbf{I}_n]\,. \end{aligned} \tag{18}$$

From (18) we can conclude that $\lambda/(1+\lambda) = r^2$ and hence that this version of Cragg and Donald (1993)'s statistic is equivalent to testing the significance of the smallest canonical correlation.

Hall et al. (1996) have also advocated using the smallest canonical correlations between $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ to assess the relevance of the instruments for the estimation of $\boldsymbol{\beta}$. They argue that if the smallest canonical correlations are not significantly different from zero then the first stage estimates are likely to be ill-conditioned (rank deficient) and IV estimation will perform poorly, see also Bowden and Turkington (1984, §2.3). In particular Hall et al. (1996) suggest testing the smallest canonical correlations using an hypothesis testing procedure based on an application of the likelihood principle and asymptotic distribution theory.

If we are interested in looking for evidence that $\mathbf{\Gamma}_{22}$ is small then our previous discussion suggests that we should examine those linear combinations $\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and $\widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ that yield evidence in favour of the hypothesis that $\widetilde{\mathbf{Y}} \perp \widetilde{\mathbf{Z}}$. Roy's Union-Intersection principle indicates that this would ultimately lead to a procedure akin to those considered by Cragg and Donald (1993) and Hall et al. (1996), except that we would examine the size of the largest rather than the smallest partial canonical correlation. It is of interest to note that Theorem 8.10.4 of Anderson (2003) implies that whereas Roy's maximum root test with an acceptance region of the form $\{r_1^2 : r_1^2 \leq \kappa_\alpha\}$ is admissible, in the sense that it cannot be improved upon by reducing the probability of Type I and/or Type II errors, the minimum root test with acceptance region $\{r_n^2 : r_n^2 \leq \kappa'_\alpha\}$ is not.

7 Optimality Properties

From the previous analysis it is apparent that although $\mathcal{A}^2$ has been derived from a rather different perspective it uses some of the same building blocks as the partial R^2 statistics of Bound et al. (1995) and Shea (1997), and the asymptotic test procedures considered by Cragg and Donald (1993) and Hall et al. (1996). Continuing the analogy, we therefore wish to ascertain if the constructions proposed in this paper exhibit any desirable properties.

We have already noted that if IV estimation is thought of as a two-stage procedure then some of the statistics discussed above can be interpreted as arising from an examination of the properties of the first stage regression. In the same vane, let us consider the augmented reduced form equation

$$[\mathbf{Y}\ \mathbf{Z}] = \mathbf{X}[(\mathbf{\Pi}_1 + \mathbf{\Pi}_3\mathbf{\Pi}_2)\ \mathbf{\Pi}_3] + [\mathbf{U}_1\ \mathbf{U}_2] \tag{19}$$

where the conditional distribution of $\mathbf{U} = [\mathbf{U}_1\ \mathbf{U}_2] = [(\mathbf{V} + \mathbf{U}_2\mathbf{\Pi}_2)\ \mathbf{U}_2]$ given $\mathbf{X}$ is Gaussian with mean zero and variance-covariance $\mathbf{\Sigma}\otimes\mathbf{I}_T$, $\mathbf{U}|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma}\otimes\mathbf{I}_T)$. If we regard (19) as a specification for the joint distribution of $[\mathbf{Y}\ \mathbf{Z}]$ conditional on $\mathbf{X}$ we can contemplate testing that the instruments are orthogonal to the endogenous regressors by testing the hypothesis that $\mathbf{U}_1 \perp \mathbf{U}_2$, i.e. that $\mathbf{\Sigma}_{12} = \mathbf{0}$.

To construct the likelihood ratio statistic, LR_T, we first concentrate the likelihood with respect to the parameters in $\mathbf{\Sigma}$ to give a maximized value for the log likelihood of

$$-\frac{T}{2}\ln\det\left[[\mathbf{Y}\ \mathbf{Z}]'\mathbf{R}_{\mathrm{X}}[\mathbf{Y}\ \mathbf{Z}]\right] - \frac{T}{2}(n+\nu-k)(1+\ln 2\pi)$$

in the unrestricted parameter space and

$$-\frac{T}{2}(\ln\det[\mathbf{Y}'\mathbf{R}_{\mathrm{X}}\mathbf{Y}] + \ln\det[\mathbf{Z}'\mathbf{R}_{\mathrm{X}}\mathbf{Z}]) - \frac{T}{2}(n+\nu-k)(1+\ln 2\pi)$$

when subjected to the restriction that $\mathbf{\Sigma}_{12} = \mathbf{0}$. Hence we find that

$$-2\ln LR_T = T\{\ln\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}] + \ln\det[\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}] - \ln\det\left[[\widetilde{\mathbf{Y}}\ \widetilde{\mathbf{Z}}]'[\widetilde{\mathbf{Y}}\ \widetilde{\mathbf{Z}}]\right]$$

and

$$LR_T^{2/T} = \det\begin{bmatrix}\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} & \widetilde{\mathbf{Y}}'\widetilde{\mathbf{Z}}\\ \widetilde{\mathbf{Z}}'\widetilde{\mathbf{Y}} & \widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\end{bmatrix} \div \left\{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]\cdot\det[\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}]\right\}. \tag{20}$$

From (20) we can readily deduce that $LR_T^{2/T} = \mathcal{A}^2$.

Thus $\mathcal{A}^2$ may be interpreted as arising out of a likelihood ratio test of multivariate orthogonality between the instruments and the endogenous regressors. That $\mathcal{A}^2$ depends on the relative magnitudes of the generalized variances of these two sets of variables is obvious. It seems reasonable to suppose therefore that $\mathcal{A}^2$ will reflect the internal variance-covariance structure of the instruments and the endogenous regressors and will provide a precise measure of the orthogonality between $\mathbf{Z}$ and $\mathbf{Y}$, after having adjusted for the effects of $\mathbf{X}$. Indeed, $\mathcal{A}^2$ yields an admissible, invariant test that possesses a power function that is monotonically increasing in each ρ_i^2, $i = 1, \ldots, n$, where ρ_i^2 are the (population) canonical correlations, the characteristic roots of $\mathbf{\Sigma}_{11}^{-\frac{1}{2}}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-\frac{1}{2}}$.

To prove the last statement, first note that the problem of testing the hypothesis that $\mathbf{\Sigma}_{12} = \mathbf{0}$ or, equivalently, $\rho_i^2 = 0$, $i = 1, \ldots, n$, is invariant under the group of non-singular linear transformations. It is well known that the canonical correlations are maximal invariants under this group of transformations and so $\mathcal{A}^2$ is an invariant test statistic. Admissibility follows by writing the acceptance region of the test as

$$\begin{aligned} \mathcal{AR}\{\mathcal{A}^2, \alpha\} &= \{\mathcal{A}^2 : \prod_{i=1}^{n}(1 - r_i^2)^{-1} \leq \kappa_\alpha\} \\ &= \{\mathcal{A}^2 : \prod_{i=1}^{n}(1 + \lambda_i) \leq \kappa_\alpha\}\,, \end{aligned} \tag{21}$$

where the $\lambda_i = r_i^2/(1 - r_i^2)$ coincide with the non-zero characteristic roots of $\mathbf{P}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}}(\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{\mathrm{Z}}}\widetilde{\mathbf{Y}})^{-1}\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{\mathrm{Z}}}$, and applying Corollary 8.10.2 of Anderson (2003).

Now, since $\mathbf{R}_{\mathrm{X}}$ is idempotent of rank $T - k$ there exists a $T \times (T - k)$ column orthonormal matrix $\mathbf{Q}_{\mathrm{X}}$, $\mathbf{Q}_{\mathrm{X}}'\mathbf{Q}_{\mathrm{X}} = \mathbf{I}_{T-k}$, such that $\mathbf{R}_{\mathrm{X}} = \mathbf{Q}_{\mathrm{X}}\mathbf{Q}_{\mathrm{X}}'$ and $\mathbf{Q}_{\mathrm{X}}'[\mathbf{Y}\ \mathbf{Z}] = \mathbf{Q}_{\mathrm{X}}'[\mathbf{U}_1\ \mathbf{U}_2]$. There also exists two non-singular matrices $\mathbf{A}$ and $\mathbf{G}$ that map $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, to the canonical variates so that $[\mathbf{V}_1\ \mathbf{V}_2] = \mathbf{Q}_{\mathrm{X}}'[\mathbf{U}_1\mathbf{A}\ \mathbf{U}_2\mathbf{G}]$ is distributed

$$N\left([\mathbf{0}\ \mathbf{0}], \begin{bmatrix} \mathbf{I}_n & [\mathbf{\Lambda}\ \mathbf{0}] \\ [\mathbf{\Lambda}\ \mathbf{0}]' & \mathbf{I}_\nu \end{bmatrix} \otimes \mathbf{I}_{T-k}\right)$$

where $\mathbf{\Lambda} = \mathrm{diag}[\rho_1, \ldots, \rho_n]$. Given the instruments, the conditional distribution of $\mathbf{V}_1$ is $N(\mathbf{V}_2[\mathbf{\Lambda}\ \mathbf{0}]', (\mathbf{I}_n - \mathbf{\Lambda}^2) \otimes \mathbf{I}_{T-k})$ and $\mathbf{W}_1 = \mathbf{V}_1(\mathbf{I}_n - \mathbf{\Lambda}^2)^{-\frac{1}{2}} \sim N(\mathbf{V}_2\mathbf{M}', \mathbf{I}_n \otimes \mathbf{I}_{T-k})$ where $\mathbf{M} = [\mathrm{diag}[\delta_1, \ldots, \delta_n]\ \mathbf{0}]$, $\delta_i = \rho_i/(1 - \rho_i^2)^{\frac{1}{2}}$.

Define $\mathbf{H}$ as the $(T - k) \times (T - k)$ orthogonal matrix

$$\mathbf{H} = \begin{bmatrix} (\mathbf{V}_2'\mathbf{V}_2)^{-\frac{1}{2}}\mathbf{V}_2' \\ (\mathbf{W}_1'\mathbf{R}_{\mathrm{V}_2}\mathbf{W}_1)^{-\frac{1}{2}}\mathbf{W}_1'\mathbf{R}_{\mathrm{V}_2} \\ \mathbf{H}_3 \end{bmatrix}$$

where $\mathbf{H}_3$ is a $(T - k - \nu - n) \times (T - k)$ matrix that makes $\mathbf{H}$ orthogonal. Then

$$\mathbf{H}\mathbf{W}_1 = \begin{bmatrix} (\mathbf{V}_2'\mathbf{V}_2)^{-\frac{1}{2}}\mathbf{V}_2'\mathbf{W}_1 \\ (\mathbf{W}_1'\mathbf{R}_{\mathrm{V}_2}\mathbf{W}_1)^{\frac{1}{2}} \\ \mathbf{H}_3\mathbf{W}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{11} \\ \mathbf{W}_{12} \\ \mathbf{0} \end{bmatrix}$$

and $\mathbf{W}_{11} \sim N((\mathbf{V}_2'\mathbf{V}_2)^{\frac{1}{2}}\mathbf{M}', \mathbf{I}_n \otimes \mathbf{I}_\nu)$ is distributed independently of $\mathbf{W}_{12} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_n)$. Moreover, by construction, the λ_i, $i = 1, \ldots, n$, of expression (21) are the non-zero characteristic roots of $\mathbf{W}_{11}(\mathbf{W}_{12}'\mathbf{W}_{12})^{-1}\mathbf{W}_{11}'$. Applying the same argument as that used by Anderson (2003, pp. 368–369) it follows that $\mathcal{AR}\{\mathcal{A}^2, \alpha\}$ is convex in each row of $\mathbf{W}_{11}$ given $\mathbf{W}_{12}$ and the other rows of $\mathbf{W}_{11}$ and hence by Theorem 8.10.6 of Anderson (2003) the conditional power of $\mathcal{A}^2$, given the instruments, is monotonically increasing in the characteristic roots of $\mathbf{M}\mathbf{V}_2'\mathbf{V}_2\mathbf{M}'$. But the characteristic roots of $\mathbf{M}\mathbf{V}_2'\mathbf{V}_2\mathbf{M}'$ are all monotonically increasing in ρ_i, $i = 1, \ldots, n$, by Lemma 9.10.2 of Anderson (2003). Taking the unconditional power, recognizing that the marginal distributions of $\mathbf{W}_{12}'\mathbf{W}_{12} \sim \mathcal{W}_n(n, \mathbf{I}_n)$ and $\mathbf{V}_2'\mathbf{V}_2 \sim \mathcal{W}_\nu(T-k, \mathbf{I}_\nu)$ do not depend on the ρ_i, $i = 1, \ldots, n$, gives us the result that for all possible sets of the instruments the power of $\mathcal{A}^2$ is monotonically increasing in each ρ_i, $i = 1, \ldots, n$.

Returning briefly to Roy's maximum root test, note that it is easily shown that

$$\frac{\mathbf{u}_i\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{\mathbf{Z}}}\widetilde{\mathbf{Y}}\mathbf{u}_i}{\mathbf{u}_i'\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}\mathbf{u}_i} = 1 - r_i^2, \qquad i = 1, \ldots, n,$$

and hence that

$$\max_{\|\theta\|=1} F_\theta = \left(\frac{T-\nu}{\nu}\right)\left(\frac{r_1^2}{1-r_1^2}\right).$$

This prompts consideration of a test based on the largest canonical correlation with acceptance region

$$\mathcal{AR}\{r_1^2, \alpha\} = \left\{r_1^2 : r_1^2 \leq \kappa_\alpha = \frac{(T-\nu)\mathcal{F}_{(1-\alpha)}\{\nu, T-\nu\}}{\nu + (T-\nu)\mathcal{F}_{(1-\alpha)}\{\nu, T-\nu\}}\right\}.$$

Such a test is invariant and, as we have already seen, admissible. The fact that in the model induced by $\boldsymbol{\theta}$ it can be shown that $\mathcal{CR}\{A_\theta^2, \alpha\}$ determines a uniformly most powerful critical region of size α suggests that a test based on $\mathcal{AR}\{r_1^2, \alpha\}$ might have reasonable power properties. Results presented in Schatzoff (1966b) indicate, however, that although Roy's maximum root test will have good power in alternative directions where $\rho_1^2 > 0, \rho_2^2 = 0, \ldots, \rho_n^2 = 0$, its performance will be inferior to that of $\mathcal{A}^2$ more generally. An obvious advantage of using $\mathcal{A}^2$ is that it does not focus on a particular canonical correlation but summarizes the simultaneous impact of all ρ_i^2, $i = 1, \ldots, n$, suggesting that $\mathcal{A}^2$ will be sensitive to deviations of $\mathbf{\Gamma}_{22}$ from zero in all possible directions.

8 Conclusion

The main contribution of this paper has been to introduce a new multivariate measure of the magnitude of the concentration parameter in a simultaneous equations model. The underlying motivation is to provide the practitioner with a method of ascertaining when the concentration parameter $\mathbf{\Gamma}_{22}$ is small and hence when the use of the Poskitt and Skeels (2003) approximation to the exact sampling distribution of the IV estimator is appropriate. This is achieved by adopting a perspective very different from that employed in the existing literature on weak identification and weak instruments, using notions of alienation rather than correlation. As $\mathcal{A}^2$ is a measure of the magnitude of $\mathbf{\Gamma}_{22}$ it is clearly applicable to models with weak instruments, but it was not designed to detect instrument weakness *per se* and is by no means limited to that case.

The second contribution of this paper was to develop the exact finite sample distribution theory for $\mathcal{A}^2$. That said, we are somewhat uncomfortable about the use of traditional inferential techniques associated with hypothesis testing for assessing the magnitude of $\mathbf{\Gamma}_{22}$ and favour an approach based on the use of p-values as a calibration devise.

The third contribution of the paper was to generalize the partial R^2 measures proposed by Bound et al. (1995) and Shea (1997) to a multivariate measure $\mathcal{R}^2$ and to develop the exact finite sample distribution theory for this generalization. Unfortunately the lack of complementarity between alienation and correlation in multivariate settings results in potentially different inferences when using the multivariate measures $\mathcal{A}^2$ and $\mathcal{R}^2$. Only $\mathcal{A}^2$, however, is in accord with the basic *desideratum* of being sensitive to departures from asymptotic orthogonality and hence of being able to detect the proximity of $\mathbf{\Gamma}_{22}$ to zero. The paper also explores the relationships that exist between the measures considered here and other statistics that have been advanced elsewhere in the literature.

Our fourth contribution has been to show that $\mathcal{A}^2$ admits an interpretation as a likelihood ratio statistic and that it inherits several desirable properties of that statistic. Indeed, the optimality properties of $\mathcal{A}^2$ suggest that it can be expected to yield a reliable guide to the magnitude of the concentration parameter.

Notes

1. For a comprehensive survey see Stock, Wright, and Yogo (2002).

2. The notation $\mathbf{V} \sim N(\mathbf{M}, \mathbf{\Omega})$ should be read as $\text{vec}(\mathbf{V}) \sim N(\text{vec}(\mathbf{M}), \mathbf{\Omega})$. The assumption of normality is simply one of convenience. What is of fundamental importance here is the sampling distribution of the quadratic form $\mathbf{S}$ defined in (6). To the extent that Wishartness holds only approximately, which it will under reasonably general conditions, the subsequent results will also hold approximately. As such matters are not germane to the ideas that we are seeking to convey, we will not pursue them further.

3. If the spectral decomposition of $\mathbf{\Omega}$ is $\mathbf{H}'\mathbf{\Omega}\mathbf{H} = \mathbf{D}$, where $\mathbf{H}$ is an orthogonal matrix of characteristic vectors of $\mathbf{\Omega}$ and $\mathbf{D} = \text{diag}[\lambda_1(\mathbf{\Omega}), \ldots, \lambda_{n+1}(\mathbf{\Omega})]$ is the diagonal matrix of characteristic roots, then $\mathbf{\Omega}^{\frac{1}{2}} = \mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{H}'$ where $\mathbf{D}^{\frac{1}{2}} = \text{diag}[\lambda_1(\mathbf{\Omega})^{\frac{1}{2}}, \ldots, \lambda_{n+1}(\mathbf{\Omega})^{\frac{1}{2}}]$; see, for example, Searle (1982, Section 11.6).

References

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

Banerjee, D.P. (1958) On the exact distribution of a test in multivariate analysis. *Journal of the Royal Statistical Society* B 20, 108–110.

Bartlett, M.S. (1947) The general canonical correlation distribution. *Annals of Mathematical Statistics* 18, 1–17.

Bound, J., D.A. Jaeger, & R.M. Baker (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443–450.

Bowden, R.J. & D.A. Turkington (1984) *Instrumental Variables*. Cambridge University Press.

Box, G.E.P. (1949) A general distribution theory for a class of likelihood criteria. *Biometrika* 36, 317–346.

Cragg, J.G. & S.G. Donald (1993) Testing identifiability and specification in instrument variable models. *Econometric Theory* 9, 222–240.

Hall, A.R., G.D. Rudebusch, & D.W. Wilcox (1996) Judging instrument relevance in instrumental variables estimation. *International Economic Review* 37, 283–298.

Hooper, J.W. (1959) Simultaneous equations and canonical correlation theory. *Econometrica* 27, 245–256.

Hotelling, H. (1936) Relations between two sets of variables. *Biometrika* 28, 321–377.

Mariano, R.S. (1982) Analytical small-sample distribution theory in econometrics: The simultaneous-equations case. *International Economic Review* 23, 503–533.

Phillips, P.C.B. (1980) The exact distribution of instrumental variable estimators in an equation containing n+1 endogenous variables. *Econometrica* 48, 861–878.

Phillips, P.C.B. (1983) Exact small sample theory in the simultaneous equations model. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 1, 449–516. North-Holland.

Poskitt, D.S. & C.L. Skeels (2003) Approximating the distribution of the instrumental variables estimator when the concentration parameter is small. Paper presented at the 14th EC^2 Conference, London.

Rao, C.R. (1951) An asymptotic expansion of the distribution of Wilk's Λ-criterion. *Bulletin of The International Statistical Institute* 33, 117–180.

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

Rothenberg, T.J. (1984) Approximating the distributions of econometric estimators and test statistics. In Z.Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 2, 881–935. North-Holland.

Roy, S.N. (1957) *Some Aspects of Multivariate Analysis*. Wiley.

Sargan, J.D. (1958) The estimation of economic relationships using instrumental variables. *Econometrica* 26(3), 393–415.

Schatzoff, M. (1966a) Exact distributions of Wilk's likelihood ratio criterion. *Biometrika* 53, 347–358, and correction, (1967), 54, 688.

Schatzoff, M. (1966b) Sensitivity comparisons among tests of the general linear hypothesis. *Journal of the American Statistical Association* 61, 415–435.

Searle, S.R. (1982) *Matrix Algebra Useful For Statistics*. Wiley.

Shea, J. (1997) Instrument relevance in multivariate linear models: A simple measure. *The Review of Economics and Statistics* LXXIX, 348–352.

Stock, J.H., J. Wright, & M. Yogo (2002) A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20, 518–529.

Wilks, S. S. (1962) *Mathematical Statistics*. Wiley.