Assessing Instrumental Variable Relevance: An Alternative Measure and Some Exact Finite Sample Theory*

DS Poskitt
Department of Econometrics
and Business Statistics
Monash University
Victoria 3800, Australia

CL Skeels
Department of Economics
The University of Melbourne
Victoria 3010, Australia

November 25, 2002

Abstract

Existing measures for the relevance of instrumental variables tend to focus on the ability of the instrument set to predict a single endogenous regressor, even if there is more than one endogenous regressor in the equation of interest. We propose new measures of instrument relevance in the presence of multiple endogenous regressors, taking both univariate and multivariate perspectives, and develop the accompanying exact finite sample distribution theory in each case. In passing, the paper also explores relationships that exist between the measures proposed here and other statistics that have been proposed elsewhere in the literature. These explorations highlight the close connection between notions of instrument relevance, identification and specification testing in simultaneous equations models.

Keywords: Instrumental variables, weak instruments, relevance, alienation, Wilks' Lambda.

JEL Codes: C120, C130, C160, C300, C510.

*The authors would like to thank seminar participants at University of Melbourne, QUT, and conference delegates at ESAM02 and the Econometric Study Group Conference in Bristol 2002. We are most grateful to Farshid Vahid for insightful comments and to Neville Norman for helpful suggestions on the presentation of this paper.

1 Introduction

Inference based on instrumental variable (IV) methods is contingent on the availability of valid instruments. Validity of instruments, as defined by the standard assumptions accompanying IV estimators, is comprised of two distinct attributes: (i) exogeneity and (ii) relevance.[1] In this paper we are solely concerned with issues of instrument relevance.

The recent past has seen much discussion of the consequences of weak, or irrelevant, instruments for inference based on IV methods; see, for example, Phillips (1989), Nelson and Startz (1990a,b), Choi and Phillips (1992), Bound, Jaeger, and Baker (1995), Hall, Rudebusch, and Wilcox (1996), Staiger and Stock (1997), Shea (1997), Dufour (1997), Zivot, Startz, and Nelson (1998), Wang and Zivot (1998), Startz, Nelson, and Zivot (2000), Woglom (2001) and Hahn and Hausman (2002). There is clear consensus that the deleterious effect of weak instruments on many standard techniques of inference cannot be ignored.

There is, however, no generally agreed definition of what exactly constitutes 'weakness'. For the most part weakness of instruments corresponds to a lack of identification and is characterized by a non-centrality parameter being, in some sense, close to zero. In the case of a single endogenous regressor this non-centrality parameter is often scaled by the degree of over-identification of the model. This scaling invites analogy to an F-statistic of the first stage regression when the IV estimate is viewed as a two-stage least squares procedure; see Bound et al. (1995) and Stock, Wright, and Yogo (2002). 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.[2] 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).[3] In general, these two statistics are distinct and have different characteristics. The former, which we will denote as R_p^2, is designed for models which contain just one endogenous regressor. The latter is available in models with several endogenous regressors but it only provides a 'rule of thumb' by which to gauge 'relevance' as there is no associated distribution theory upon which more formal inference can be based.

[1] Within a set of valid instruments one might also be concerned about issues of efficiency or optimal instrument choice. We shall abstract from such considerations in this paper.

[2] It is worth noting that such arguments do not necessarily lead to 'optimal' IV estimators; see, for example, Forchini and Hillier (1999) and Hahn and Hausman (2002).

[3] Other measures of instrument relevance exist, such as those explored by Hall et al. (1996) and Staiger and Stock (1997). However, if IV estimation is thought of as a two-stage procedure, all of these measures can be thought of as examining the goodness of fit of the first stage regression. The partial R^2 measures discussed here are the simplest ways of doing this and are the procedures most commonly encountered in practice.

In this paper we adopt a rather different perspective. We are motivated by the idea that the instrument set should be independent of, or at least asymptotically uncorrelated with, the stochastic disturbances in the system. Thus the instruments cannot be perfectly correlated with the endogenous variables, yet at the same time they are unlikely to be of much use if they are completely uncorrelated with the regressors. Hence we seek a reliable measure that will characterise complete lack of correlation, rather than one designed to detect high or perfect correlation. We therefore propose the use of a *partial coefficient of alienation*, denoted A_p^2, as a measure of instrument relevance. We generalize A_p^2 to models containing arbitrary numbers of endogenous regressors (subject to the usual identification conditions holding). Unlike the partial R^2 of Shea (1997), we also show that it is possible to provide an analytical exploration of the sampling distribution of A_p^2 and, consequently, to develop appropriate inferential procedures.

Coincidentally, we also generalize R_p^2 to models containing arbitrary numbers of endogenous regressors and derive it's sampling distribution and associated inferential procedures. Although A_p^2 is related to R_p^2 in the special case of a model with a single endogenous regressor it is seen that the relationship does not hold in more general models. Consequently we note that the generalizations of A_p^2 and R_p^2 may yield distinct outcomes.

The structure of the remainder of the paper is as follows. In the next section we shall outline the model and explore some existing measures of instrument relevance. In Section 3 we provide the definition of A_p^2 and explore its sampling properties. This enables us to provide a formal hypothesis test of instrument relevance although we advocate its use as a calibration device. Section 4 explores the relationship between instrument relevance, specification testing, and over-identification, in the context of the measures and tests discussed in the earlier sections. In Section 5 we provide multivariate versions of A_p^2 and R_p^2 which are applicable when there is more than one endogenous regressor in the equation of interest. These measures are denoted $\mathcal{A}_p^2$ and $\mathcal{R}_p^2$, respectively, and Section 6 discusses the relationships of $\mathcal{A}_p^2$ and $\mathcal{R}_p^2$ to alienation, canonical correlations and a likelihood ratio statistic.

2 Measures of Instrument Relevance

2.1 The Model and Notation

Suppose that

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ $(N \times 1)$ and $\mathbf{X}$ $(N \times k)$ represent the observations on an equation of interest that is to be estimated using the values of the instruments in $\mathbf{Z}$ $(N \times n)$, $N > n \geq k$. We shall assume, without loss of generality, that $\mathbf{Z}$

has rank equal to n. Finally, we shall partition $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$, where $\mathbf{X}_1$ is $(N \times 1)$, and $\boldsymbol{\beta} = (\beta_1\ \boldsymbol{\beta}_2')'$.

Before proceeding, we shall first establish some notational conventions. For any $(N \times m)$ matrix $\mathbf{W}$ let $\mathbf{P}_W = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}'$ and $\mathbf{R}_W = \mathbf{I}_N - \mathbf{P}_W$, where $(\mathbf{W}'\mathbf{W})^{-}$ denotes the Moore-Penrose generalized-inverse of $\mathbf{W}'\mathbf{W}$. Then $\mathbf{P}_W$ is the $(N \times N)$ idempotent symmetric (prediction) operator that projects on to the space spanned by the columns of $\mathbf{W}$, where this space is denoted by $Sp\{\mathbf{W}\}$, and $\mathbf{R}_W$ is the associated (residual) operator which projects on to the orthogonal complement of that space, $Sp\{\mathbf{W}^{\perp}\}$. Following Shea (1997) we shall adopt a special notation to denote the most frequently encountered projections. Specifically, a circumflex will denote projections onto $Sp\{\mathbf{Z}\}$ and a tilde will denote projections onto $Sp\{\mathbf{X}_2^{\perp}\}$, e.g. $\widehat{\mathbf{X}}_1 = \mathbf{P}_Z\mathbf{X}_1$ and $\widetilde{\mathbf{X}}_1 = \mathbf{R}_{X_2}\mathbf{X}_1$. Finally, we shall assume that all variables are measured in terms of deviations from their average so that $\mathbf{y} = \mathbf{R}_\imath\mathbf{y}$, $\mathbf{X} = \mathbf{R}_\imath\mathbf{X}$ and $\mathbf{Z} = \mathbf{R}_\imath\mathbf{Z}$, where $\boldsymbol{\imath}' = (1, \ldots, 1)$ and $\mathbf{R}_\imath = (\mathbf{I}_N - \boldsymbol{\imath}\boldsymbol{\imath}'/N)$.

2.2 $\mathbf{R}^2$ Measures

Shea (1997, p. 348) argues, not unreasonably, that 'in a multivariate context relevance requires that $\mathbf{Z}$ have components important to $\mathbf{X}_1$ that are linearly independent of those important to $\mathbf{X}_2$.' If we evaluate importance in terms of the contribution to mean squared error then the components of $\mathbf{Z}$ important to $\mathbf{X}_2$ cannot lie in the space of $\widetilde{\mathbf{Z}} = \mathbf{R}_{X_2}\mathbf{Z}$, since $\widetilde{\mathbf{Z}}'\mathbf{X}_2 = \mathbf{0}$. Consequently, it seems desirable that the components of $\mathbf{Z}$ important to $\mathbf{X}_1$ should lie in $Sp\{\widetilde{\mathbf{Z}}\}$, for then they will clearly be linearly independent of those important to $\mathbf{X}_2$. A statistic that appears to measure instrument relevance in the sense just defined is therefore the squared correlation between $\mathbf{X}_1$ and $\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1$,

$$R^2 = \frac{(\mathbf{X}_1'\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1)^2}{(\mathbf{X}_1'\mathbf{X}_1)(\mathbf{X}_1'\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1)} = \frac{\mathbf{X}_1'\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1}{\mathbf{X}_1'\mathbf{X}_1}.$$

Note that the measure R^2 does not adjust $\mathbf{X}_1$ for the influence of $\mathbf{X}_2$. An analogous statistic that does make this adjustment is, of course, the squared correlation between $\widetilde{\mathbf{X}}_1$ and $\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$,

$$R_p^2 = \frac{(\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1)^2}{(\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1)(\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1)} = \frac{\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1}{\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1}.$$

Now it is trivially true that $\mathbf{X}_1 = \mathbf{P}_{X_2}\mathbf{X}_1 + \widetilde{\mathbf{X}}_1$ and therefore $\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1 = \mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$ since, by definition, $\mathbf{P}_{\widetilde{Z}} = \mathbf{R}_{X_2}\mathbf{Z}(\mathbf{Z}'\mathbf{R}_{X_2}\mathbf{Z})^{-}\mathbf{Z}'\mathbf{R}_{X_2}$ and $\mathbf{R}_{X_2}\mathbf{P}_{X_2} = \mathbf{0}$. Hence $\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1 = \mathbf{X}_1'\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1$, so that R^2 and R_p^2 have the same numerator but different denominators, and $R^2 < R_p^2$, unless $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal, in which case $\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1 = \mathbf{X}_1'\mathbf{R}_{X_2}\mathbf{X}_1 = \mathbf{X}_1'\mathbf{X}_1$ and $R^2 = R_p^2$.

It is R_p^2 that accords with the conventional notion of partial R^2. Bound et al. (1995, p. 444), for example, define partial R^2 as the (in our notation) 'R^2 from the regression of $\mathbf{X}_1$ on $\mathbf{Z}$ once the common exogenous variables have been partialled out of both $\mathbf{X}_1$ and $\mathbf{Z}$.' Their stipulation on *common exogenous* variables in this definition simply reflects the fact that their basic model contains only one endogenous regressor, so that $\mathbf{X}_2$ is necessarily exogenous and is presumed to act as its own instrument. Here we explicitly partial out the effects of $\mathbf{X}_2$ from both $\mathbf{X}_1$ and $\mathbf{Z}$, and we allow $\mathbf{X}_2$ to contain both endogenous and exogenous variables. Moreover, $\mathbf{X}_2$ and $\mathbf{Z}$ need not contain any variables in common.

2.3 Shea's Statistic

Shea (1997) recognised that R_p^2, as proposed by Bound et al. (1995), is not applicable in models containing more than one endogenous regressor. He suggested that a measure of the relevance of the instrumental variables in $\mathbf{Z}$ for the estimation of the coefficient β_1 that is applicable is given by the correlation between (a) the component of $\mathbf{X}_1$ orthogonal to $\mathbf{X}_2$, and (b) the component of $\mathbf{X}_1$'s projection on to $\mathbf{Z}$ that is orthogonal to $\mathbf{X}_2$'s projection on to $\mathbf{Z}$. If, as above, $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$ then Shea's statistic is given by r^2, where

$$r = \frac{\widetilde{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1}{\sqrt{(\widetilde{\mathbf{X}}_1' \widetilde{\mathbf{X}}_1)(\widehat{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1)}},$$

the correlation between $\widetilde{\mathbf{X}}_1$ and $\mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1$.

Dubbed a partial R^2 by Shea (1997), r^2 can be shown to be proportional to the ratio of the variances of the ordinary least squares (OLS) and the IV estimator of β_1. To see this observe, following Godfrey (1999), that

$$\widetilde{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1 = \mathbf{X}_1' \mathbf{R}_{X_2} (\mathbf{I} - \mathbf{P}_{\widehat{X}_2}) \mathbf{P}_Z \mathbf{X}_1 = \mathbf{X}_1' (\mathbf{P}_Z - \mathbf{P}_{\widehat{X}_2}) \mathbf{X}_1 = \widehat{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1,$$

where the second equality uses Lemma A.2 to show that $\mathbf{P}_{X_2}(\mathbf{P}_Z - \mathbf{P}_{\widehat{X}_2}) = 0$. This allows us to simplify the expression for r by noting that $\widetilde{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1 = \widehat{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1$, yielding the result

$$r^2 = \frac{\widehat{\mathbf{X}}_1' \mathbf{R}_{\widehat{X}_2} \widehat{\mathbf{X}}_1}{\widetilde{\mathbf{X}}_1' \widetilde{\mathbf{X}}_1}. \tag{2.1}$$

When expressed in this manner the statistic appears as proportional to the ratio of the variances of the OLS and IV estimators, and it is in this simplified form that the statistic is motivated by Shea (1997) and subsequently discussed in Godfrey (1999).

Given the similarities between r^2 and R_p^2, and 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 the two measures.

2.4 The Relationship Between Measures

Given that they have the same denominators, the relationship between r^2 and R_p^2 is determined by the relationship between the numerators of the two statistics, i.e. between

$$N_1 = \widehat{\mathbf{X}}_1'\mathbf{R}_{\widehat{X}_2}\widehat{\mathbf{X}}_1 = \mathbf{X}_1'[\mathbf{P}_Z - \mathbf{P}_Z\mathbf{X}_2(\mathbf{X}_2'\mathbf{P}_Z\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{P}_Z]\mathbf{X}_1 \tag{2.2}$$

and

$$N_2 = \mathbf{X}_1'\mathbf{P}_{\widetilde{Z}}\mathbf{X}_1, \tag{2.3}$$

respectively. The question of immediate interest is whether $N_1 = N_2$. Starting with the algebraic relationship $\mathbf{R}_{R_Z} = \mathbf{I} - \mathbf{R}_Z(\mathbf{R}_Z'\mathbf{R}_Z)^{-}\mathbf{R}_Z = \mathbf{I} - \mathbf{R}_Z = \mathbf{P}_Z$, we can replace $\mathbf{P}_Z$ in equation (2.2) by $\mathbf{R}_{R_Z}$ and, using equation (A.1b) of Lemma A.1, we obtain

$$N_1 = \mathbf{X}_1'[\mathbf{R}_{R_Z} - \mathbf{R}_{R_Z}\mathbf{X}_2(\mathbf{X}_2'\mathbf{R}_{R_Z}\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{R}_{R_Z}]\mathbf{X}_1 = \mathbf{X}_1\mathbf{R}_{[X_2\ R_Z]}\mathbf{X}_1. \tag{2.4}$$

Applying equation (A.1a) of Lemma A.1 yields

$$\begin{aligned} N_1 &= \mathbf{X}_1'[\mathbf{R}_{X_2} - \mathbf{R}_{X_2}\mathbf{R}_Z(\mathbf{R}_Z\mathbf{R}_{X_2}\mathbf{R}_Z)^{-}\mathbf{R}_Z\mathbf{R}_{X_2}]\mathbf{X}_1 \\ &= \mathbf{X}_1'[\mathbf{R}_{X_2} - \mathbf{P}_{R_{X_2}R_Z}]\mathbf{X}_1. \end{aligned} \tag{2.5}$$

If $N_1 = N_2$ then

$$\mathbf{X}_1'\mathbf{R}_{X_2}\mathbf{X}_1 = \mathbf{X}_1'[\mathbf{P}_{\widetilde{Z}} + \mathbf{P}_{R_{X_2}R_Z}]\mathbf{X}_1. \tag{2.6}$$

According to Rao and Mitra (1971, Theorem 5.1.2), equation (2.6) is satisfied if and only if

$$\mathbf{P}_{\widetilde{Z}}\mathbf{P}_{R_{X_2}R_Z} = \mathbf{0}. \tag{2.7}$$

From Lemma A.2, $\mathbf{Z}'\mathbf{P}_{\widetilde{Z}} = \mathbf{Z}'\mathbf{R}_{X_2}$ and $\mathbf{R}_Z'\mathbf{P}_{R_{X_2}R_Z} = \mathbf{R}_{X_2}\mathbf{R}_Z$. So, on pre- and post-multiplying equation (2.7) by $\mathbf{Z}'$ and $\mathbf{R}_Z$, respectively, it follows that an equivalent requirement is that $\mathbf{Z}'\mathbf{R}_{X_2}\mathbf{R}_Z = \mathbf{0}$.

Partition $\mathbf{Z} = [\mathbf{Z}_1\ \mathbf{Z}_2]$ where, without loss of generality, we may suppose that $Sp\{\mathbf{Z}_2\} \subseteq Sp\{\mathbf{X}_2\}$, so that $\mathbf{Z}_2 = \mathbf{X}_2\mathbf{D}_2$ for some $\mathbf{D}_2$. The obvious example of such a partitioning is where $\mathbf{X}_2$ contains exogenous (predetermined) variables which are included in $\mathbf{Z}$, as would typically be the case in practice. Note that, because $n > k-1$ by assumption, $Sp\{\mathbf{Z}\} \not\subseteq Sp\{\mathbf{X}_2\}$, hence $\mathbf{R}_{X_2}\mathbf{Z} \neq \mathbf{0}$ and $\mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1 > 0$. Observe that, by construction,

$$\mathbf{R}_{X_2}\mathbf{Z} = [\mathbf{R}_{X_2}\mathbf{Z}_1\ \mathbf{R}_{X_2}\mathbf{Z}_2] = [\mathbf{R}_{X_2}\mathbf{Z}_1\ \mathbf{0}],$$

as $\mathbf{Z}_1 \notin Sp\{\mathbf{X}_2\}$ and $\mathbf{Z}_2 \in Sp\{\mathbf{X}_2\}$ by definition, and $\mathbf{R}_{X_2}\mathbf{R}_{Z_2} = \mathbf{R}_{X_2}$. Applying Lemma A.1 we obtain $\mathbf{R}_Z = \mathbf{R}_{Z_2} - \mathbf{R}_{Z_2}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1)^{-}\mathbf{Z}_1'\mathbf{R}_{Z_2}$ and hence $\mathbf{R}_{X_2} - \mathbf{R}_{X_2}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1)^{-}\mathbf{Z}_1'\mathbf{R}_{Z_2} = \mathbf{R}_{X_2}\mathbf{R}_Z$, from which we deduce that

$$\mathbf{Z}'\mathbf{R}_{X_2}\mathbf{R}_Z = \begin{bmatrix} \mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{R}_Z \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1'(\mathbf{R}_{X_2} - \mathbf{R}_{X_2}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1)^{-}\mathbf{Z}_1'\mathbf{R}_{Z_2}) \\ \mathbf{0} \end{bmatrix}.$$

Therefore $\mathbf{Z}'\mathbf{R}_{X_2}\mathbf{R}_Z = \mathbf{0}$ if, and only if,

$$\mathbf{Z}_1'\mathbf{R}_{X_2} = \mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1)^{-}\mathbf{Z}_1'\mathbf{R}_{Z_2}.$$

On post-multiplying by $\mathbf{R}_{X_2}\mathbf{Z}_1$, the latter requirement becomes equivalent to

$$\mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1 = \mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1(\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1)^{-}\mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1.$$

As $\mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1 > \mathbf{0}$, this equality reduces to $\mathbf{I} = (\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1)^{-}\mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1$, or simply $\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1 = \mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1$. Noting that $Sp\{\mathbf{X}_2\} = Sp\{\mathbf{Z}_2\} \oplus Sp\{\mathbf{R}_{Z_2}\mathbf{X}_2\}$ leads to the conclusion that $\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{Z}_1 \geq \mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1$, with equality if and only if $\mathbf{Z}_1'\mathbf{R}_{Z_2}\mathbf{X}_2 = \mathbf{0}$ so that $\mathbf{Z}_1$ provides no explanation of $\mathbf{X}_2$ beyond that provided by $\mathbf{Z}_2$. Therefore we see that r^2, Shea's statistic, will equal R_p^2, partial R^2 as conventionally defined, when $\mathbf{X}_2$ can act as, and is indeed used as, its own instrument. Conversely, whenever $\mathbf{X}_2$ contains endogenous regressors r^2 and R_p^2 will diverge.

The preceding results have implications for the practical use of these measures in models more general than those for which they appear to have been originally proposed. Recall that Bound et al. (1995) only have a single endogenous regressor and so r^2 will equal R_p^2 for their model, since $\mathbf{X}_2$ will be a 'common exogenous' regressor, although care must be taken with such an interpretation because their equation of interest contains no explicit $\mathbf{X}_2$ variables, exogenous or otherwise.

3 An Alternative Measure and Formal Test Procedure

To begin, note that the numerator of R_p^2, $\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$, equals the explained sum of squares in the *OLS* regression of $\widetilde{\mathbf{X}}_1$ on $\widetilde{\mathbf{Z}}$. That is, $N_2 = \widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1 = \mathbf{g}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{g}$, where

$$\widetilde{\mathbf{X}}_1 = \widetilde{\mathbf{Z}}\mathbf{g} + \mathbf{u} \tag{3.1}$$

with $\mathbf{g} = (\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{X}}_1$ and

$$\mathbf{u}'\mathbf{u} = \widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1. \tag{3.2}$$

The operator $\mathbf{R}_{\widetilde{Z}}$ is, of course, the natural complement to $\mathbf{P}_{\widetilde{Z}}$ since $\mathbf{P}_{\widetilde{Z}}\mathbf{R}_{\widetilde{Z}} = \mathbf{0}$ and $\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1 = \widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1 + \widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$, *c.f.* (2.6). It follows that in the current setting the statistic

$$A_p^2 = \frac{\widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1}{\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1}$$

forms a natural complement to R_p^2. Indeed, it is a simple exercise to show that A_p^2 equals the squared correlation between $\widetilde{\mathbf{X}}_1$ and $\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$. In particular A_p^2 takes the value one when $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Z}}$ are orthogonal and zero if $\widetilde{\mathbf{X}}_1$ lies in $Sp\{\widetilde{\mathbf{Z}}\}$. In terms of the original variables of the model, A_p^2 can be viewed as a measure of the perpendicularity between $\mathbf{X}_1$ and $\mathbf{Z}$ having adjusted for the effects of $\mathbf{X}_2$, and as such it may be interpreted as a measure of instrument relevance of the type that we seek.

In order to construct a formal test of instrument relevance based on the measure A_p^2, consider once again the *OLS* regression equation (3.1). The reduced form equation for $\mathbf{X}_1$,

$$\mathbf{X}_1 = \mathbf{Z}\boldsymbol{\Pi}_1 + \boldsymbol{\varepsilon}_1 \ ,$$

is derived from the reduced form $\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{E}$ where, to use an obvious notation, $\mathbf{Y} = [\mathbf{X}_1 \ \mathbf{Y}_2]$ and $\mathbf{Y}_2$ contains any additional endogenous variables that appear in $\mathbf{X}_2$, $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_1 \ \boldsymbol{\Pi}_2]$ and the conditional distribution of the stochastic disturbance $\mathbf{E} = [\boldsymbol{\varepsilon}_1 \ \mathbf{E}_2]$ given the instruments $\mathbf{Z}$ is Gaussian with mean zero and variance-covariance $\boldsymbol{\Sigma}\otimes\mathbf{I}$, $\langle \mathrm{vec}(\mathbf{E})|\mathbf{Z}\rangle \sim N(\mathbf{0}, \boldsymbol{\Sigma}\otimes\mathbf{I})$. Multiplying the reduced form for $\mathbf{X}_1$ by $\mathbf{R}_{X_2}$ we find that the implicit model underlying (3.1) is

$$\widetilde{\mathbf{X}}_1 = \widetilde{\mathbf{Z}}\boldsymbol{\gamma} + \boldsymbol{\eta} \tag{3.3}$$

where $\boldsymbol{\gamma} = \boldsymbol{\Pi}_1$ and $\langle \boldsymbol{\eta}|[\mathbf{Z}\ \mathbf{X}_2]\rangle \sim N(\mathbf{0}, \sigma_{1\cdot 2}^2\mathbf{R}_{X_2})$, a singular normal distribution (see, for example, Rao, 1973, §8a) with $\sigma_{1\cdot 2}^2 = \sigma_{11}^2 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-}\boldsymbol{\Sigma}_{21}$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Now consider testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$ in (3.3) via the variance ratio

$$VR = \frac{\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1}{\widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1}$$

computed from (3.1). The matrices $\mathbf{P}_{\widetilde{Z}}$ and $\mathbf{R}_{\widetilde{Z}}$ are both idempotent with ranks $\rho = \rho\{\widetilde{\mathbf{Z}}\}$ and $N - \rho$, respectively. Because $\mathbf{P}_{\widetilde{Z}}\mathbf{R}_{X_2} = \mathbf{P}_{\widetilde{Z}}$ and

$\mathbf{R}_{\widetilde{Z}}\mathbf{R}_{X_2} = \mathbf{R}_{X_2} - \mathbf{P}_{\widetilde{Z}}$ are also both idempotent, it follows from standard results on the distribution of quadratic forms in Gaussian variates that, given $\mathbf{Z}$ and $\mathbf{X}_2$, $\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$ will be distributed as $\sigma_{1\cdot 2}^2$ times Chi-squared variates with ρ and $N-\rho$ degrees of freedom, respectively, i.e. $\sigma_{1\cdot 2}^2 \cdot \chi^2(\rho)$ and $\sigma_{1\cdot 2}^2 \cdot \chi^2(N-\rho)$. Moreover, because $(\mathbf{R}_{X_2} - \mathbf{P}_{\widetilde{Z}})\mathbf{R}_{X_2}\mathbf{P}_{\widetilde{Z}} = \mathbf{0}$, $\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1$ are independent. Hence the re-scaled ratio

$$F = \frac{\widetilde{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1/\rho}{\widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1/(N-\rho)} = \left(\frac{N-\rho}{\rho}\right) VR$$

will possess an $\mathcal{F}$ distribution with ρ and $N-\rho$ degrees of freedom. See, for example, Rao (1973, §3.b and Complements and Problems 1–2.4) or Rao and Mitra (1971, Theorems 9.2.1 and 9.4.2).

From the previous development, however, we know that

$$VR = \frac{1-A_p^2}{A_p^2}.$$

It follows that conditional on $\mathbf{X}_2$ and $\mathbf{Z}$ the statistic

$$F = \left(\frac{N-\rho}{\rho}\right)\left(\frac{1-A_p^2}{A_p^2}\right) \sim \mathcal{F}\{\rho, N-\rho\} \tag{3.4}$$

under $\mathcal{H}_0$. Recall that $\widetilde{\mathbf{Z}} = [\mathbf{R}_{X_2}\mathbf{Z}_1\ \mathbf{0}] = [\widetilde{\mathbf{Z}}_1\ \mathbf{0}]$ and that $\widetilde{\mathbf{Z}}_1'\widetilde{\mathbf{Z}}_1 = \mathbf{Z}_1'\mathbf{R}_{X_2}\mathbf{Z}_1 > 0$ which implies that $\widetilde{\mathbf{Z}}_1$ has full column rank which, in turn, implies that $\rho = \rho\{\widetilde{\mathbf{Z}}_1\} = \rho\{\mathbf{Z}_1\}$. We can always write $\rho = n - n_2$, where $0 \leq n_2 = \rho\{\mathbf{Z}_2\} \leq k-1$, because $\mathbf{Z}$ has full column rank by assumption. Now suppose that we partition $\mathbf{X}_2$ into endogenous ($\mathbf{X}_{21}$) and exogenous ($\mathbf{X}_{22}$) variables, so that $\mathbf{X}_2 = [\mathbf{X}_{21}\ \mathbf{X}_{22}]$ and, without loss of generality, $\mathbf{Z} = [\mathbf{Z}_1\ \mathbf{Z}_2]$ where $\mathbf{Z}_2 = \mathbf{X}_{22}$. Then

$$\mathbb{P}[\mathbf{Z}_1 = \mathbf{X}_2\mathbf{D}] = 0$$

for any matrix $\mathbf{D}$. Consequently, $\rho = \text{rank}(\mathbf{Z}_1)$ is known with probability one, and

$$F \underset{as}{\overset{\mathcal{H}_0}{\sim}} \mathcal{F}\{\rho, N-\rho\}.$$

In particular, ρ is simply the number of instruments used in addition to $\mathbf{X}_{22}$, making the result (3.4) of considerable practical importance.

From the implicit model (3.3) it is clear that the hypothesis $\mathcal{H}_0 : \boldsymbol{\gamma} = \mathbf{0}$ is equivalent to the statement that $\widetilde{\mathbf{X}}_1$ and $\widetilde{\mathbf{Z}}$ are (asymptotically) orthogonal since $\mathsf{E}[\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{Z}}] = \mathsf{E}[\boldsymbol{\eta}'\widetilde{\mathbf{Z}}]$ under $\mathcal{H}_0$ and $\mathsf{E}[\boldsymbol{\eta}|[\mathbf{Z}\ \mathbf{X}_2]] = \mathbf{0}$ implies that $\mathsf{E}[\boldsymbol{\eta}'\widetilde{\mathbf{Z}}] = \mathbf{0}$. Thus the statistic F can be used to assess the significance of the departure of the measure A_p^2 from unity under the null hypothesis $\mathcal{H}_0 : \widetilde{\mathbf{X}}_1 \perp \widetilde{\mathbf{Z}}$, wherein we have employed the notation $\widetilde{\mathbf{X}}_1 \perp \widetilde{\mathbf{Z}}$ as a 'shorthand' for $\mathbb{P}(\lim_{N\to\infty} \|N^{-1}\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{Z}}\| > 0) = 0$. Values of F that exceed conventional significance levels may be taken as being indicative of statistically significant departures from (asymptotic) orthogonality and such values will thereby provide evidence of instrument relevance.

Given that decisions concerning instrument relevance will be contingent upon the behaviour of the statistic F it is of interest to observe that under the alternative hypothesis $\mathcal{H}_1 : \boldsymbol{\gamma} \neq \mathbf{0}$ the conditional distribution of F given $\mathbf{X}_2$ and $\mathbf{Z}$ is non-central $\mathcal{F}$ with degrees of freedom ρ and $N-\rho$ and non-centrality parameter $\lambda = \boldsymbol{\gamma}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma}/\sigma^2_{1\cdot 2}$. This result follows because $\widehat{\mathbf{X}}_1'\mathbf{P}_{\widetilde{Z}}\widehat{\mathbf{X}}_1/\sigma^2_{1\cdot 2}$ has a non-central chi-squared distribution with ρ degrees of freedom and non-centrality parameter λ, $\chi^2(\rho, \lambda)$, under $\mathcal{H}_1$. Let

$$\mathcal{CRA}_N(\alpha) = \{\{\mathbf{X}, \mathbf{Z}\} : F > \mathcal{F}_{(1-\alpha)}\{\rho, N-\rho\}\}$$

where $\mathcal{F}_{(1-\alpha)}\{\rho, N-\rho\}$ denotes the $(1-\alpha)100\%$ percentile point of the $\mathcal{F}\{\rho, N-\rho\}$ distribution. If $\mathbb{P}(\lim_{N\to\infty} N^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}} > 0) = 1$ then from a hypothesis testing perspective it follows that $\mathcal{CRA}_N(\alpha)$ defines a strongly consistent critical region of size α. In general it is apparent that finite sample power will be large in directions $\boldsymbol{\theta} = \boldsymbol{\gamma}/\|\boldsymbol{\gamma}\|$ such that

$$\lambda = \frac{\|\boldsymbol{\gamma}\|^2}{\sigma^2_{1\cdot 2}}\, \boldsymbol{\theta}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\theta}$$

is 'large', but the probability of rejection will be close to the size of the test in directions $\boldsymbol{\theta}$ for which λ is close to zero.

We are not so much concerned with making explicit decisions about the acceptance or rejection of $\mathcal{H}_0$, however, but rather in assessing instrument relevance on an appropriate scale. Given $\mathbf{X}$ and $\mathbf{Z}$ we can calculate F and then compute $p_{\text{obs}} = \mathbb{P}(\mathcal{F}\{\rho, N-\rho\} > F)$. In terms of the measure A_p^2, the case where λ is large corresponds to situations where $\mathbf{g}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{g}$ will be close to its upper bound of $\widetilde{\mathbf{X}}_1'\widetilde{\mathbf{X}}_1$ so $A_p^2 \approx 0$. Instrument relevance will be high and $p_{\text{obs}} \ll \alpha$. The case where λ is close to zero corresponds to situations where $\mathbf{g}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{g} \approx 0$. The measure A_p^2 will approximately equal one and instrument relevance will be low with $p_{\text{obs}} \gg \alpha$. Although we are not especially interested in making sharp distinctions between data sets where $p_{\text{obs}} \leq \alpha$ and data sets such that $p_{\text{obs}} > \alpha$, it is clear that p_{obs} yields a probability scale that delineates realizations of $\mathbf{X}$ and $\mathbf{Z}$ that are indicative of directions in which A_p^2 will be small(large) and the regressors in $\widetilde{\mathbf{Z}}$ will contain components that are important(unimportant) to $\widetilde{\mathbf{X}}_1$.

4 Over-identification, Relevance and Specification Tests

In order to gain some additional insight into the structure and interpretation of the results described in the previous section it is useful to consider other test procedures employed in the context of *IV* estimation.

First let us examine an *IV* test of whether the model

$$\mathbf{X}_1 = \mathbf{X}_2\boldsymbol{\delta} + \boldsymbol{\eta} \tag{4.1}$$

is the correct specification for $\mathbf{X}_1$, $\bar{\mathcal{H}}_0$ say, against the alternative $\bar{\mathcal{H}}_1$ that $\mathbf{Z}$ contains additional columns that do not lie in $Sp\{\mathbf{X}_2\}$ that should be included in the model, which is characterized by the specification

$$\mathbf{X}_1 = \mathbf{X}_2\boldsymbol{\delta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\eta} \ , \tag{4.2}$$

where $\langle\boldsymbol{\eta}\rangle \sim N(\mathbf{0}, \sigma_{11}^2\mathbf{I})$. Such tests, which date back to the work of Basmann (1960), are referred to as tests for over-identifying restrictions since to identify (4.1) as many columns of $\mathbf{Z}$ must be excluded from $\mathbf{X}_2$ as there are endogenous regressors but more columns may have been excluded than is necessary and a test of (4.1) against (4.2) can be viewed as a test of the latter possibility. Following Davidson and MacKinnon (1993, §7.8) and noting that upon elimination of redundant variables (4.2) is just identified, one such test leads to a consideration of the ratio

$$\frac{N(\mathbf{X}_1'\mathbf{P}_Z\mathbf{X}_1 - \mathbf{X}_1'\mathbf{P}_Z\mathbf{X}_2(\mathbf{X}_2'\mathbf{P}_Z\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{P}_Z\mathbf{X}_1)}{(\mathbf{X}_1 - \mathbf{X}_2\mathbf{d})'(\mathbf{X}_1 - \mathbf{X}_2\mathbf{d})}$$

where $\mathbf{d} = (\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_2)^{-}\widehat{\mathbf{X}}_2'\widehat{\mathbf{X}}_1$, see Davidson and MacKinnon (1993, Equations 7.56 & 7.57, p. 236). This ratio is precisely r^2 multiplied by $N(\mathbf{X}_1'\mathbf{R}_{X_2}\mathbf{X}_1) \div (\|\mathbf{X}_1 - \mathbf{X}_2\mathbf{d}\|^2)$. Thus Shea's statistic can be easily modified to produce an *IV* test of an artificial regression in which the endogenous regressor $\mathbf{X}_1$ appears as the regressand, the modification arising from the fact that $N^{-1}(\mathbf{X}_1\mathbf{R}_{X_2}\mathbf{X}_1)$ estimates $\sigma_{1\cdot2}^2$ rather than σ_{11}^2. Unfortunately, this does not provide a justification for Shea's statistic, the possibility of adjusting for the underestimation of σ_{11}^2 notwithstanding. As pointed out by Davidson and MacKinnon (1993), an *IV* test of $\bar{\mathcal{H}}_0$ against $\bar{\mathcal{H}}_1$ amounts to a joint test that the model in (4.1) is correctly specified and that $\mathbf{Z}$ contains a valid collection of instruments for that model. Not only does such a test not address the question of interest, namely, the relevance of the instruments in $\mathbf{Z}$ for the estimation of the coefficient on $\mathbf{X}_1$ in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$; but (4.1) and (4.2) are obtained by arbitrarily assuming that the additional regressors $\mathbf{X}_2$ that appear in $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ provide an adequate specification for $\mathbf{X}_1$ and there is no logical reason why this should be so. Thus we have no *a priori* justification for believing that either $\bar{\mathcal{H}}_0$ or $\bar{\mathcal{H}}_1$ will be true, indeed, it is much more likely that they are not, and the basic tenets of such a test are therefore undermined.

To relate the artificial regressions in (4.1) and (4.2) to A_p^2 note from Lemma A.1 that $\widetilde{\mathbf{X}}_1'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{X}}_1 = \mathbf{X}_1'(\mathbf{R}_{X_2} - \mathbf{P}_{\widetilde{Z}})\mathbf{X}_1 = \mathbf{X}_1'\mathbf{R}_{[X_2\ Z]}\mathbf{X}_1$, the residual sum of squares in the regression of $\mathbf{X}_1$ on both $\mathbf{X}_2$ and $\mathbf{Z}$ (what is in effect the reduced rank version of the Frisch-Waugh-Lovell theorem) and $\mathbf{X}_1'\mathbf{R}_{X_2}\mathbf{X}_1$ is the residual sum of squares from the regression of $\mathbf{X}_1$ on $\mathbf{X}_2$ alone. It follows that F corresponds to the F-statistic that would be obtained when testing the significance of the coefficient vector $\boldsymbol{\gamma}$ in the model (4.2). Bound et al. (1995, §2.2) argue that an examination of the value of the F-statistic on common excluded instruments in the first stage IV regression can provide valuable information on finite sample bias. Our results lend strong support to their conclusion, since an implication of the results established above is that

$$\mathsf{E}\left[\frac{(N-\rho-2)(1-A_p^2)}{A_p^2}\ |[\mathbf{X}_2\ \mathbf{Z}]\right] = \rho + \frac{\sigma_{11}^2\tau^2}{\sigma_{1\cdot 2}^2}$$

where τ^2 is Basmann's (Basmann, 1960) concentration parameter, which is inversely proportional to the bias of the IV estimator of β_1.

Hahn and Hausman (2002, pp. 166-7) argue against the use of goodness-of-fit measures on the grounds that the bias of IV estimators in these models depends upon more than one fundamental parameter, of which goodness-of-fit measures are concerned with only one. In order to say anything about the exact extent of bias it will be necessary to know something about all of the factors that determine it: These are the degree of over-identification, the number of endogenous regressors, the correlation(s) between the endogenous regressor(s) and the structural disturbance of the equation of interest, and the non-centrality parameter. Hence, one might reasonably conclude that a complete examination of bias will require the examination of measures of these correlations, as explored in tests of exogeneity, and measures of the non-centrality parameters, as explored by goodness-of fit measures. Procedures based on combinations of measures have been proposed by Hall and Peixe (2001) and the statistics advanced in this paper might also be included in such a mix.

It has also been argued that the absence of appropriate distribution theory mitigates against the use of goodness-of-fit measures such as those considered in this paper. Our results clearly address this shortcoming and indicate that A_p^2 can be employed to construct a useful measure of instrument relevance. It is important to emphasize, however, that the rationale for the measure A_p^2 and the associated statistic given in equation (3.4) is based upon the regression in (3.1) and the implied model (3.3), the equivalence to the F-statistic derived from the regressions in (4.1) and (4.2) is an algebraic equality that yields, at best, a convenient computational device.

Conceptually A_p^2, via F and p_{obs}, measures and calibrates the relevance of the instruments via the reduced form, it does not rely on an artificial specification.

5 Multivariate Measures

At this point in the paper we wish to consider extending the analysis in Section 3 to cover the case where interest focuses not on the estimation of the coefficient on one particular endogenous regressor but on the overall estimation of the coefficient vector associated with all the endogenous variables appearing in the model or equation of interest. We will continue to express this situation in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.1}$$

where $\mathbf{y}$, $\mathbf{X}$ and the instrument set $\mathbf{Z}$ are defined as above. Now, however, $\mathbf{X}$ will be divided into $\mathbf{Y}$, an $(N \times k_1)$ matrix of observations on all the endogenous variables that appear in (5.1), and $\mathbf{X}_2$, an $(N \times k_2)$ matrix of observations on the exogenous variables, with $k_1 + k_2 = k$, to give

$$\mathbf{y} = [\mathbf{Y}\ \mathbf{X}_2] \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \boldsymbol{\varepsilon}. \tag{5.2}$$

Technically we should introduce a new expression for $\mathbf{X}_2$ in the partition of $\mathbf{X}$ but we feel that the meaning should be clear from the context and in order to avoid a proliferation of notation we have used the same symbol for the additional regressors.

An amalgam of the concepts used in Section 3 and the results presented in this section is possible wherein $\mathbf{X} = [\mathbf{Y}\ \mathbf{X}_2]$ and $\mathbf{X}_2$ is allowed to include both exogenous and additional endogenous variables. Because we find it difficult to envisage a situation where one would only be interested in the estimation of a partial list of endogenous variables we do not discuss this latter case in detail. Suffice it to say that although the presence of endogenous variables in $\mathbf{X}_2$ introduces a more complicated theoretical derivation and leads to minor changes in interpretation, such a change has no impact on the practical implementation of the measures derived.

We are interested in measuring the relevance of the instruments in $\mathbf{Z}$ for the estimation of the coefficient vector $\boldsymbol{\beta}_1$ that appears on the endogenous regressors. Our motivation is based on the idea that when assessing instrument relevance, just as a measure of relevance for a single coefficient needs to 'partial out' the effects of other regressors, so to must allowance be made for the fact that individual relevance may be low whilst at the same time overall relevance is high. (The latter being analogous to the situation oft-quoted in text books where individual regressors have small 't-ratios' whilst the overall regression is highly significant.)

5.1 Instrument Relevance and Alienation

It seems reasonable to seek a measure for the multivariate situation that is similar to A_p^2 both in terms of its construction and interpretation, and an obvious candidate for consideration is

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

To interpret $\mathcal{A}_p^2$ observe that it takes the value one when $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are orthogonal and zero if there exists a matrix $\mathbf{D}$ of full column rank such that $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Z}}\mathbf{D}$. Thus $\mathcal{A}_p^2$ can be viewed as a measure of the perpendicularity between $\mathbf{Y}$ and $\mathbf{Z}$ having adjusted for the effects of $\mathbf{X}_2$. Indeed, $\mathcal{A}_p^2$ is a natural generalization of A_p^2 from the univariate case in that $1-\mathcal{A}_p^2$ represents the proportion of the generalized variance of $\widetilde{\mathbf{Y}}$ that can be attributed to $\widetilde{\mathbf{Z}}$ in the multivariate regression of $\widetilde{\mathbf{Y}}$ on $\widetilde{\mathbf{Z}}$. In point of fact, $\mathcal{A}_P^2$ 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.

In order to derive the distribution of $\mathcal{A}_p^2$ consider once again the sub-system reduced form $\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{E}$ where $\langle \mathrm{vec}(\mathbf{E})|\mathbf{Z}\rangle \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I})$. Pre-multiplying the reduced form by $\mathbf{R}_{X_2}$ we see that $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Z}}\mathbf{\Pi} + \mathbf{R}_{X_2}\mathbf{E}$, where $\langle \mathrm{vec}(\mathbf{R}_{X_2}\mathbf{E})|[\mathbf{X}_2\ \mathbf{Z}]\rangle \sim N(\mathbf{0}, \mathbf{\Sigma}\otimes\mathbf{R}_{X_2})$, and post-multiplying by the constant vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k_1})'$ we obtain the equation

$$\widetilde{\mathbf{Y}}\boldsymbol{\alpha} = \widetilde{\mathbf{Z}}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{5.3}$$

where now $\boldsymbol{\gamma} = \mathbf{\Pi}\boldsymbol{\alpha}$ and $\langle \boldsymbol{\eta}|[\mathbf{Z}\ \mathbf{X}_2]\rangle \sim N(\mathbf{0}, \sigma_\alpha^2\mathbf{R}_{X_2})$, $\sigma_\alpha^2 = \boldsymbol{\alpha}'\mathbf{\Sigma}\boldsymbol{\alpha}$. From the argument used in Section 3 we know that $\mathbf{R}_{\widetilde{Z}}$ is idempotent with rank $N-\rho$, $\mathbf{P}_{\widetilde{Z}}$ is idempotent with rank equal to ρ and $\mathbf{R}_{\widetilde{Z}}\mathbf{R}_{X_2}\mathbf{P}_{\widetilde{Z}} = \mathbf{0}$. It follows that for given $\mathbf{Z}$ and $\mathbf{X}_2$ the quadratic forms $\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}\boldsymbol{\alpha}$ will be independently distributed as $\sigma_\alpha^2\cdot\chi^2(\rho, \lambda)$, $\lambda = \boldsymbol{\gamma}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma}/\sigma_\alpha^2$, and $\sigma_\alpha^2\cdot\chi^2(N-\rho)$ random variables respectively, where we recall that $\rho = \rho\{\widetilde{\mathbf{Z}}_1\}$. 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:

$$\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}} \sim\ \mathcal{W}_{k_1}(\rho, \mathbf{\Sigma}, \mathbf{\Sigma}^-\mathbf{\Pi}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{\Pi})$$

and

$$\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}} \sim\ \mathcal{W}_{k_1}(N-\rho, \mathbf{\Sigma})\ .$$

In directions $\boldsymbol{\theta} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|$ such that $\boldsymbol{\theta}'\mathbf{\Pi}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\mathbf{\Pi}\boldsymbol{\theta} = 0$ the non-centrality parameter $\lambda = 0$ 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 $\mathcal{A}_p^2$ 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 $\lambda = \boldsymbol{\gamma}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma} = 0$ the statistic $\mathcal{A}_p^2$ will possess Wilks'-Λ distribution

$$\Lambda(k_1, N-\rho, \rho) \sim \begin{cases} \prod_{i=1}^{k_1} \mathcal{B}(\frac{N-\rho+1-i}{2}, \frac{\rho}{2}) & \rho \geq k_1, \\ \prod_{i=1}^{\rho} \mathcal{B}(\frac{N-\rho-k_1+i}{2}, \frac{k_1}{2}) & \text{otherwise,} \end{cases}$$

the product of independent Beta random variables, see (Wilks, 1962, §18.5.1). From equation (5.3) we see that $\boldsymbol{\alpha}'\boldsymbol{\Pi}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\Pi}\boldsymbol{\alpha} = 0$ is equivalent to the statement that $\widetilde{\mathbf{Y}}\boldsymbol{\alpha} \perp \widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ since $\mathsf{E}[\boldsymbol{\alpha}'\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma}] = \mathsf{E}[\boldsymbol{\eta}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma}] = \mathsf{E}[\boldsymbol{\alpha}'\mathbf{E}'\widetilde{\mathbf{Z}}\boldsymbol{\gamma}] = 0$. Wilks'-$\Lambda$ distribution can therefore be used to calibrate the significance of departures of the measure $\mathcal{A}_p^2$ from one in a manner similar to the way the $\mathcal{F}$ distribution is used to calibrate A_p^2. In the special case where $k_1 = 1$ it is of interest to note that

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

from which the probabilistic relationship of our multivariate measure $\mathcal{A}_p^2$ to the scalar version A_p^2 considered in Section 3 is readily apparent.

Box (1949) provides a series expansion for Wilks'-Λ distribution in terms of Chi-squared distributions. Banerjee (1958) uses Mellin transforms to construct an exact expression for the distribution of $\Lambda(k_1, N-\rho, \rho)$, involving sums, products and ratios of Gamma functions, that depends on whether k_1 and ρ are even or odd, and Schatzoff (1966) gives exact closed form representations applicable when k_1 or ρ is even and supplies tables of correction factors that can be used to convert Chi-squared percentile points to percentile points of $\Lambda(k_1, N-\rho, \rho)$ for k_1 or ρ even and $k_1\rho \leq 70$. In the current situation $k_1 \geq 2$ and $\rho \geq 1$. When $k_1 = 2$ we can use the exact result that

$$F = \left(\frac{N-\rho-1}{\rho}\right)\left(\frac{1-\mathcal{A}_p}{\mathcal{A}_p}\right) \sim \mathcal{F}\{2\rho, 2(N-\rho-1)\}$$

for any ρ to calculate p_{obs}. 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

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

where

$$m = N - \frac{k_1+\rho+1}{2},$$

will converge in distribution to $\chi^2(k_1\rho)$ as $N \to \infty$. A closer asymptotic approximation correct up to terms of order $O(N^{-3})$ can be constructed using a second order version of Box's expansion and Box (1949) also presents an $\mathcal{F}$ approximation for $-m\ln(\mathcal{A}_p^2)$ that has a remainder term $O(N^{-3})$. Box found that the latter gives close agreement with the exact distribution even when the sample size is small, $10 \leq N \leq 20$ say. An even more precise $\mathcal{F}$ approximation is given by Rao (1951). Rao's approximation implies that

$$F = \left(\frac{ms-2q}{k_1\rho}\right)\left(\frac{1-\mathcal{A}_p^{2/s}}{\mathcal{A}_p^{2/s}}\right),$$

where

$$s = \sqrt{\frac{(k_1\rho)^2-4}{k_1^2+\rho^2-5}} \quad \text{and} \quad q = \frac{k_1\rho-2}{4},$$

may be treated as an $\mathcal{F}\{k_1\rho, ms-2q\}$ 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 $O(N^{-4})$ but the structure of the approximation also has a certain appeal in the current circumstances since the statistic F can be employed to evaluate p_{obs} via $\mathcal{F}\{k_1\rho, ms-2q\}$ and thereby calibrate $\mathcal{A}_p^2$ in the same way it was used in the univariate case.

5.2 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

$$\mathcal{R}_p^2 = \frac{\det[\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}]}{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]}\ , \tag{5.4}$$

a partial version of Hotelling's 'coefficient of vector correlation'. Writing $\mathcal{R}_p^2$ 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 (5.3) shows that when the non-centrality parameter $\lambda = \boldsymbol{\alpha}'\boldsymbol{\Pi}'\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\boldsymbol{\Pi}\boldsymbol{\alpha}$ 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 $\mathcal{W}_{k_1}(\rho,\boldsymbol{\Sigma})$ and $\mathcal{W}_{k_1}(N-\rho,\boldsymbol{\Sigma})$ random variables, respectively, leads to the conclusion that the statistic $\mathcal{R}_p^2$ will possess Wilks'-$\Lambda(k_1,\rho,N-\rho)$ distribution.

Wilks'-Λ distribution can therefore be used to calibrate the measure $\mathcal{R}_p^2$ in much the same way it is used to calibrate $\mathcal{A}_p^2$, only now it will be large values of the statistic

$$F = \left(\frac{ms-2q}{k_1(N-\rho)}\right)\left(\frac{1-\mathcal{R}_p^{2/s}}{\mathcal{R}_p^{2/s}}\right),$$

where

$$m = \frac{N - k_1 + \rho - 1}{2}, \quad s = \sqrt{\frac{(k_1(N-\rho))^2 - 4}{k_1^2 + (N-\rho)^2 - 5}} \quad \text{and} \quad q = \frac{k_1(N-\rho) - 2}{4},$$

that will lend support to the hypothesis that $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are orthogonal, with small values of $p_{obs} = \mathbb{P}(\mathcal{F}\{k_1(N-\rho), ms - 2q\} \leq F)$ indicating that the regressors in $\widetilde{\mathbf{Z}}$ contain components that are sufficiently important to $\widetilde{\mathbf{Y}}$ to make $\mathcal{R}_p^2$ sufficiently large. It is important to observe, however, that since $\mathcal{R}_p^2 \neq 1 - \mathcal{A}_p^2$ probability calculations based on $\mathcal{A}_p^2$ and $\mathcal{R}_p^2$ will not be identical, as is the case with the univariate measures A_p^2 and $R_p^2 = 1 - A_p^2$. This raises the question of which measure should be used in practice, an issue to which we will return in Section 6.

5.3 A Multivariate Shea Statistic

In expression (2.1) we have seen that Shea's univariate statistic r^2 can be represented as the ratio of the variance of the *OLS* estimator to the variance of the *IV* estimator. Let us therefore define the multivariate version of Shea's measure, which we will denote by ∇^2, as the ratio of the generalised variances of the *OLS* and *IV* estimators. This gives us the analogous expression

$$\nabla^2 = \frac{\det[\widehat{\mathbf{Y}}'\mathbf{R}_{\widehat{X}_2}\widehat{\mathbf{Y}}]}{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]} \,. \tag{5.5}$$

In this guise the numerator of ∇^2 has a form that corresponds to that of the statistic $\mathcal{A}_p^2$ in that it is structured in terms of the operator $\mathbf{R}$. The numerator of ∇^2 is based on the prior projection of $\mathbf{Y}$ and $\mathbf{X}_2$ on to the space spanned by $\mathbf{Z}$, however, rather then the residual from the projection of $\mathbf{Y}$ and $\mathbf{Z}$ onto the space spanned by $\mathbf{X}_2$ as is the case with $\mathcal{A}_p^2$. Moreover, from the algebraic manipulations conducted in Section 2.4 we know that $\widehat{\mathbf{Y}}'\mathbf{R}_{\widehat{X}_2}\widehat{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ because under present assumptions $\mathbf{X}_2$ is exogenous and is assumed to be acting as its own instrument. Thus although at first it might appear that ∇^2 will behave in a manner similar to that of $\mathcal{A}_p^2$, the opposite is in fact true because

$$\nabla^2 = \mathcal{R}_p^2 \,.$$

6 Alienation, Canonical Correlation, Partial R^2 and Independence

In the light of the interpretation of $\mathcal{A}_p^2$ 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 $\det[\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}] = \det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}]$ into the product of $\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]$ and $\det[\mathbf{I} - (\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-}\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}]$ shows that

$$\mathcal{A}_p^2 = \prod_{i=1}^{f}(1 - c_p^2(i)) \tag{6.1}$$

where $c_p^2(1) \geq \ldots \geq c_p^2(f)$, $f = \min\{k_1, \rho\}$, lists in descending order the partial canonical correlations between $\mathbf{Y}$ and $\mathbf{Z}$ having adjusted for the effects of $\mathbf{X}_2$.

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). Hall et al. (1996) have advocated using the smallest canonical correlations between $\mathbf{Z}$ and $\mathbf{X}$ 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 estimate $\widehat{\mathbf{X}}$ is 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 a standard hypothesis testing procedure based on the likelihood principle.

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 coefficient matrix in the reduced form equation $\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{E}$ has rank k_1. A version of their procedure for testing the rank of $\boldsymbol{\Pi}$ 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 $\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ in the metric of $\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$. See hypothesis H_I^0 and Theorem 3 of Cragg and Donald (1993). Using the relationship $\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}} + \widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}}$ gives us the expression

$$\det[\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}} - \lambda\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}}] = \det[(1+\lambda)\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}] \cdot \det[(\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}})^{-}\widetilde{\mathbf{Y}}'\mathbf{P}_{\widetilde{Z}}\widetilde{\mathbf{Y}} - \frac{\lambda}{1+\lambda}\mathbf{I}]\,. \tag{6.2}$$

From (6.2) we can conclude that $\lambda/(1+\lambda) = c_p^2$ and hence that this version of Cragg and Donald (1993)'s statistic is equivalent to testing the significance of the smallest canonical correlation.

To relate such ideas 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 $\mathbf{R}_{X_2}\mathbf{Y}$ and $\mathbf{R}_{X_2}\mathbf{Z}$. Given $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ the squared partial correlation

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

estimates the corresponding population partial correlation coefficient and the region

$$\left\{R_p^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) > R_c^2\right\} \ , \tag{6.3}$$

where R_c^2 is an appropriate critical value, may be taken to indicate the presence of components in $\mathbf{Y}$ and $\mathbf{Z}$ that induce a significant partial correlation. The intersection

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

of all regions of the type given in (6.3) 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.

By analogy with the Union-Intersection principle 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

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

for if the smallest partial correlation between $\mathbf{Y}\boldsymbol{\alpha}$ and $\mathbf{Z}\boldsymbol{\gamma}$ lies in (6.3) then all partial correlations of such linear combinations must do so. Similarly, the region

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

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 $\mathbf{Y}\boldsymbol{\alpha}$ and $\mathbf{Z}\boldsymbol{\gamma}$ is non-zero. It is a standard exercise to show that $\max_{\boldsymbol{\alpha}\boldsymbol{\gamma}} R_p^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) = c_p^2(1)$ and $\min_{\boldsymbol{\alpha}\boldsymbol{\gamma}} R_p^2(\boldsymbol{\alpha},\boldsymbol{\gamma}) = c_p^2(f)$. 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 these extrema are equal to $c_p^2(2) \geq \ldots \geq c_p^2(f-1)$. This suggests that if we are interested in looking for evidence of linear combinations that yield evidence against the hypothesis that $\widetilde{\mathbf{Y}}\boldsymbol{\alpha} \perp \widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ then we should examine the size of the smallest partial canonical correlations, ultimately leading to a procedure akin to those considered by Cragg and Donald (1993) and Hall et al. (1996).

Let us now address the question of which multivariate measure, $\mathcal{A}_p^2$ or $\mathcal{R}_p^2$, should be employed in practice. Assume, for the sake of argument, that complete lack of correlation between $\widetilde{\mathbf{Y}}$ on $\widetilde{\mathbf{Z}}$ is characterised by all the partial canonical correlations being zero whereas exact correlation means that $c_p^2(1) = \ldots = c_p^2(f) = 1$. From the expression in (6.1), that is,

$$\mathcal{A}_p^2 = \prod_{i=1}^{f}(1 - c_p^2(i))$$

and the corresponding representation of $\mathcal{R}_p^2$, namely,

$$\mathcal{R}_p^2 = \prod_{i=1}^{f} c_p^2(i)$$

it follows that it is only necessary for the largest (smallest) partial canonical correlation to deviate substantially from zero (one) for $\mathcal{A}_p^2$ ($\mathcal{R}_p^2$) to deviate significantly from unity. Thus, whereas $\mathcal{A}_p^2$ will be sensitive to departures from complete lack of correlation $\mathcal{R}_p^2$ 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 λ being equal to zero, which we have already observed is equivalent to the hypothesis that $\widetilde{\mathbf{Y}}\boldsymbol{\alpha} \perp \widetilde{\mathbf{Z}}\boldsymbol{\gamma}$ and hence that $c_p^2(1) = \ldots = c_p^2(f) = 0$. Therefore $\mathcal{A}_p^2$ appears to be more in accord with the basic assumption underlying the application of the distribution than is $\mathcal{R}_p^2$. Furthermore, we can interpret $1 - \mathcal{A}_p^2$ as the proportion of the generalized variance of $\mathbf{Y}$ that can be attributed to $\mathbf{Z}$ in the reduced form regression $\mathbf{Y} = \mathbf{Z}\boldsymbol{\Pi} + \mathbf{E}$, but since $\mathcal{R}_p^2 \neq 1 - \mathcal{A}_p^2$ we cannot interpret $\mathcal{R}_p^2$ in this manner. Given these features and given that we are not seeking to detect exact or perfect correlation, the measure $\mathcal{A}_p^2$ appears to be far more suited to our purpose.

From the previous analysis it is apparent that although $\mathcal{A}_p^2$ has been derived from a rather different perspective it uses some of the same building blocks as the asymptotic test procedures considered by Cragg and Donald (1993) and Hall et al. (1996). Shea (1997) has criticized the use by Hall et al. (1996) of the canonical correlations between $\mathbf{Z}$ and $\mathbf{X}$ on the grounds that (i) they do not map directly into particular regressors and that (ii) they do not distinguish problems due to instrument relevance from those due to poor conditioning.

Shea's first criticism cannot be redirected at the current measure, for to suggest that $\mathcal{A}_p^2$ is inadequate because it cannot be used to target particular regressors when the whole *raison-d'être* for $\mathcal{A}_p^2$ is to provide a *joint* measure of the relevance of $\mathbf{Z}$ for *all* of the variables in $\mathbf{Y}$ is nonsensical. The measure $\mathcal{A}_p^2$ is monotonically decreasing in $c_p^2(i)$, $i = 1, \ldots, f$, and very small values of $\mathcal{A}_p^2$ will provide evidence of the significance of all the partial canonical correlations and will lead to small values of p_{obs}, both of which are indicative of instrument relevance. When only one or a small number of partial canonical correlations provide testimony to the existence of linear dependencies between components of $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{Y}}$ their significance might be obscured by the overall measure. In such circumstances the researcher may wish to assign an individual measure of relevance to particular regressors and recourse can be made to the univariate measure A_p^2.

Shea's second criticism might be levelled at $\mathcal{A}_p^2$, but $\mathcal{A}_p^2$, like partial R^2, adjusts for the effects of the exogenous regressors in $\mathbf{X}_2$ by examining the relationships between $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{Y}}$. Furthermore, the measure is defined relative to the overall dispersion of $\mathbf{Y}$ and therefore it will not confound the internal variance-covariance structure of $\mathbf{Y}$ with instrument relevance.

To clarify the later two points let $\mathbf{W}$ denote the set difference $\mathbf{Z} \setminus \mathbf{X}_2$ and consider the equation system

$$[\mathbf{Y}\ \mathbf{W}] = \mathbf{X}_2[\boldsymbol{\Delta}_1\ \boldsymbol{\Delta}_2] + [\mathbf{N}_1\ \mathbf{N}_2] \tag{6.4}$$

where the conditional distribution of $\mathbf{N} = [\mathbf{N}_1\ \mathbf{N}_2]$ given $\mathbf{X}_2$ is Gaussian with mean zero and variance-covariance $\boldsymbol{\Omega} \otimes \mathbf{I}$, $\langle \mathrm{vec}(\mathbf{N})|\mathbf{X}_2\rangle \sim N(\mathbf{0}, \boldsymbol{\Omega} \otimes \mathbf{I})$. We have already noted that if IV estimation is thought of as a two-stage procedure then the previous measures can be thought of as statistics that arise from an examination of the properties of a first stage regression. In the same vane, if we regard (6.4) as a specification for the joint distribution of $[\mathbf{Y} : \mathbf{W}]$ conditional on $\mathbf{X}_2$ we can contemplate testing that the instruments are orthogonal to the endogenous regressors by testing the hypothesis that $\mathbf{N}_1 \perp \mathbf{N}_2$, i.e. that $\boldsymbol{\Omega}_{12} = \mathbf{0}$.

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

$$-\frac{N}{2}\ln\det[\begin{bmatrix}\mathbf{Y}' \\ \mathbf{W}'\end{bmatrix}\mathbf{R}_{X_2}(\mathbf{Y}\ \mathbf{W})] - \frac{N}{2}(k_1 + n - k_2)(1 + \ln 2\pi)$$

in the unrestricted parameter space and

$$-\frac{N}{2}(\ln\det[\mathbf{Y}'\mathbf{R}_{X_2}\mathbf{Y}] + \ln\det[\mathbf{W}'\mathbf{R}_{X_2}\mathbf{W}]) - \frac{N}{2}(k_1 + n - k_2)(1 + \ln 2\pi)$$

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

$$-2\ln LR_N = N\{\ln\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}] + \ln\det[\widetilde{\mathbf{W}}'\widetilde{\mathbf{W}}] - \ln\det[\begin{bmatrix}\widetilde{\mathbf{Y}}' \\ \widetilde{\mathbf{W}}'\end{bmatrix}(\widetilde{\mathbf{Y}}\ \widetilde{\mathbf{W}})]\}$$

and

$$LR_N^{2/N} = \det[\begin{bmatrix}\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}} & \widetilde{\mathbf{Y}}'\widetilde{\mathbf{W}} \\ \widetilde{\mathbf{W}}'\widetilde{\mathbf{Y}} & \widetilde{\mathbf{W}}'\widetilde{\mathbf{W}}\end{bmatrix}]/\left\{\det[\widetilde{\mathbf{Y}}'\widetilde{\mathbf{Y}}]\cdot\det[\widetilde{\mathbf{W}}'\widetilde{\mathbf{W}}]\right\}\ . \tag{6.5}$$

It now follows from the equality $\mathbf{Z} = [\mathbf{W}\ (\mathbf{Z} \cap \mathbf{X}_2)]$ that $\widetilde{\mathbf{Z}} = \mathbf{R}_{X_2}\mathbf{Z} = [\widetilde{\mathbf{W}}\ \mathbf{0}]$ and hence that $\widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{Z}}\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Y}}'\mathbf{R}_{\widetilde{W}}\widetilde{\mathbf{Y}}$, from which we can readily deduce that $LR_N^{2/N} = \mathcal{A}_p^2$.

From equation (6.5) it is clear that $\mathcal{A}_p^2$ may be interpreted as arising out of a likelihood ratio test of multivariate orthogonality between the instruments and the endogenous regressors and that it depends on the relative magnitudes of the generalised variances of these two sets of variables. It seems reasonable to suppose therefore that $\mathcal{A}_p^2$ will provide a precise indication of any lack of correlation between $\mathbf{Z}$ and $\mathbf{Y}$, after having adjusted for the effects of $\mathbf{X}_2$, that will be independent of the internal variance-covariance structure of both the instruments and the endogenous regressors and thus that $\mathcal{A}_p^2$ will yield a reliable measure of instrument relevance alone.

7 Conclusion

This paper has been concerned with the question of determining the relevance of instruments used in the construction of instrumental variables estimators of the coefficients in single equations from a linear simultaneous equations model. The first contribution of this paper has been to introduce a new measure of instrument relevance. We have approached the problem from a different perspective than that adopted in the existing literature; in particular, we explore notions of alienation rather than correlation. Whereas most of the measures in the literature focus on the question of whether an instrument is 'good', the measure that we propose addresses the converse question of whether an instrument is 'bad'. In the case of a single endogenous regressor our measure (A_p^2) is easily demonstrated to be the natural complement of the partial R^2 (R_p^2) of Bound et al. (1995).

One criticism that has been levelled at measures such as A_p^2 and R_p^2 is the absence of an associated distribution theory when the model contains more than one endogenous regressor. The second contribution of this paper was to develop the exact finite sample distribution theory for both of these statistics. That said, we are somewhat uncomfortable about the use of traditional inferential techniques associated with hypothesis testing for instrument selection and favour an approach to the measurement of instrument relevance based on the use of p-values as a calibration devise.

In equations with more than one endogenous regressor the relevance of given instruments for the estimation of the coefficient on a single endogenous regressor is clearly a partial notion of relevance, as all instruments will impinge on the estimation of all coefficients in the model. The third contribution of the paper was to generalize both A_p^2 and R_p^2 to multivariate measures and to develop the exact finite sample distribution theory for each of these generalizations. The multivariate measures, $\mathcal{A}_p^2$ and $\mathcal{R}_p^2$ respectively, assess the relevance of instruments for the estimation of the coefficients on all of the endogenous variables in the structural equation of interest and, in this sense, provide measures of the overall relevance of the instruments. Unfortunately the breakdown of the complementarity between alienation and correlation in multivariate settings results in potentially different inferences when using these statistics. Only $\mathcal{A}_p^2$ is in accord with the *desiderata* discussed in Section 6.

References

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.

Basmann, R. L. (1960), "On Finite-Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55, 650–659.

Bound, J., Jaeger, D. A., and Baker, R. M. (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. and Turkington, D. A. (1984), *Instrumental Variables*, Econometric Society Monographs No. 8, Cambridge: Cambridge University Press.

Box, G. E. P. (1949), "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, 36, 317–346.

Choi, I. and Phillips, P. C. B. (1992), "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Models," *Journal of Econometrics*, 51, 113–150.

Cragg, J. G. and Donald, S. G. (1993), "Testing Identifiability and Specification in Instrument Variable Models," *Econometric Theory*, 9, 222–240.

Davidson, R. and MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Dufour, J.-M. (1997), "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

Forchini, G. and Hillier, G. H. (1999), "Conditional Inference for Possibly Unidentified Structural Equations," Working Paper No. 9906, Discussion Papers in Economics and Econometrics, University of Southampton.

Godfrey, L. G. (1999), "Instrument Relevance in Multivariate Linear Models," *Review of Economics and Statistics*, LXXXI, 550–552.

Hahn, J. and Hausman, J. (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.

Hall, A. R. and Peixe, F. P. M. (2001), "A Consistent Method For The Selection of Relevant Instruments In Linear Models," Discussion Paper, Department of Economics, North Carolina State University, Raleigh, NC.

Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (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.

Nelson, C. R. and Startz, R. (1990a), "Some Further Results On the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

—— (1990b), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio when the Instrument is a Poor One," *Journal of Business*, 63, S125–S140.

Phillips, P. C. B. (1989), "Partially Identified Edconometric Models," *Econometric Theory*, 5, 181–240.

Rao, C. R. (1951), "An Asymptotic Expansion of the Distribution of Wilk's Λ Criterion," *Bulletin of The International Statistical Institute*, 33, 117–180.

—— (1973), *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons, Inc., 2nd ed.

Rao, C. R. and Mitra, S. K. (1971), *Generalized Inverse of Matrices and its Applications*, New York: John Wiley & Sons, Inc.

Sargan, J. D. (1958), "The Estimation of Economic Relationships using Instrumental Variables," *Econometrica*, 26, 393–415.

Schatzoff, M. (1966), "Exact Distributions of Wilk's Likelihood Ratio Criterion," *Biometrika*, 53, 347–358.

Shea, J. (1997), "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *The Review of Economics and Statistics*, LXXIX, 348–352.

Staiger, D. and Stock, J. H. (1997), "Instrumental Variables Regression With Weak Instruments," *Econometrica*, 65, 557–586.

Startz, R., Nelson, C. R., and Zivot, E. (2000), "Improved Inference for the Instrumental Variable Estimator," Paper presented at the Eighth World Congress of the Econometric Society, Seattle.

Stock, J. H., Wright, J., and Yogo, M. (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method Of Moments," *Journal of Business & Economic Statistics*, 20, 518–529.

Wang, J. and Zivot, E. (1998), "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica*, 66, 1389–1404.

Wilks, S. S. (1962), *Mathematical Statistics*, New York: John Wiley.

Woglom, G. (2001), "More Results on the Exact Small Sample properties of the Instrumental Variable Estimator," *Econometrica*, 69, 1381–1389.

Zivot, E., Startz, R., and Nelson, C. R. (1998), "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1144.

Appendix

Lemma A.1. *If* $\mathbf{A} = [\mathbf{B}\ \mathbf{C}]$ *then*

$$\begin{aligned} \mathbf{R}_{\mathbf{A}} &= \mathbf{R}_B - \mathbf{R}_B\mathbf{C}(\mathbf{C}'\mathbf{R}_B\mathbf{C})^{-}\mathbf{C}'\mathbf{R}_B && \text{(A.1a)} \\ &= \mathbf{R}_C - \mathbf{R}_C\mathbf{B}(\mathbf{B}'\mathbf{R}_C\mathbf{B})^{-}\mathbf{B}'\mathbf{R}_C, && \text{(A.1b)} \end{aligned}$$

where for any $(N \times m)$ *matrix* $\mathbf{W}$, $\mathbf{R}_W = \mathbf{I}_N - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}'$.

Proof. Rao and Mitra (1971, Chapter 2, Complement 6) provide the following result: If

$$\mathbf{M} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}$$

then a generalized inverse of $\mathbf{M}$ is

$$\begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -(\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{I} \end{bmatrix} \mathbf{S}^{-} \begin{bmatrix} -\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-} & \mathbf{I} \end{bmatrix}$$

with $\mathbf{S} = \mathbf{X}_2'\mathbf{R}_{X_1}\mathbf{X}_2$, from which equation (A.1a) follows. Equation (A.1b) is derived in a similar manor on noting that another generalized inverse of $\mathbf{M}$ is

$$\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{X}_2'\mathbf{X}_2)^{-} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -(\mathbf{X}_2'\mathbf{X}_2)^{-}\mathbf{X}_2'\mathbf{X}_1 \end{bmatrix} \mathbf{T}^{-} \begin{bmatrix} \mathbf{I} & -\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-} \end{bmatrix}$$

where $\mathbf{T} = \mathbf{X}_1'\mathbf{R}_{X_2}\mathbf{X}_1$. □

Lemma A.2. *For any* $(N \times m)$ *matrix* $\mathbf{W}$,

$$\mathbf{W}(\mathbf{W}'\mathbf{W})^{-}\mathbf{W}'\mathbf{W} = \mathbf{W} \qquad \text{(A.2)}$$

Proof. See Rao and Mitra (1971, Lemma 2.2.6(b)). □