ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

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

Small Concentration Asymptotics and Instrumental Variables Inference

D S Poskitt and C L Skeels

February 2005

Working Paper 4/05

Small Concentration Asymptotics and Instrumental Variables Inference

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

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

Abstract

Poskitt and Skeels (2003) provide a new approximation to the sampling distribution of the IV estimator in a simultaneous equations model, the approximation is appropriate when the concentration parameter associated with the reduced form model is small. A basic purpose of this paper is to provide the practitioner with easily implemented inferential tools based upon extensions to these small concentration asymptotic results. We present various approximations to the sampling distribution of functions of the IV estimator based upon small concentration asymptotics, and investigate hypothesis testing procedures and confidence region construction using these approximations. It is shown that the test statistics advanced are asymptotically pivotal and that the associated critical regions generate locally uniformly most powerful invariant tests. The confidence regions are also shown to be valid. The small-concentration asymptotic approximations lead to a non-standard application of standard distributions, facilitating numerical implementation using commonly available software.

Some key words: IV estimator, concentration parameter, small concentration asymptotics, hypothesis testing, confidence region construction, valid inference.

JEL Subject classifications: C12, C16, C30, C50.

1 Introduction

In a recent contribution to the literature on instrumental variables (IV) estimation Poskitt and Skeels (2003) present a new approximation to the exact sampling distribution of the 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 (2SLS) and show that when the 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), and they are applicable under circumstances that differ significantly from those for which the classical asymptotic normal approximation and Edgeworth type expansions of the distribution of the IV estimator, as described in Sargan and Mikhail (1971) and Anderson and Sawa (1973, 1979), are designed. The basic aims of this paper are to provide a guide as to how the Poskitt and Skeels (2003) approximation can be employed in practice for inferential purposes and to examine the properties of hypothesis testing procedures and confidence regions constructed using the approximation.

Asymptotic methods often yield simple approximations in situations where the evaluation of exact analytic solutions would be difficult or nigh impossible. Unfortunately such approximations can sometimes be poor. For example, large sample approximations to the sampling properties of 2SLS have been shown to perform poorly in the face of weak identification. This has motivated the development of alternative approaches, such as the many-instrument asymptotics considered in (Bekker, 1994), local-to-zero asymptotics as investigated in (Staiger and Stock, 1997), and the many-weak-instruments asymptotics considered by Chao and Swanson (2002, 2003) and Stock and Yogo (2003). No one of these alternative approaches is more correct than any other, they differ essentially in the structure of the hypothetical sequence in which they nest the problem of interest. The only criterion on which they might be compared is the usefulness of the statistical procedures that they ultimately yield.

The *small-concentration asymptotics* of Poskitt and Skeels (2003) indexes the nesting sequence of problems by an ever diminishing value of the concentration parameter. The resulting approximations have very simple functional forms and have been shown to be extremely accurate when, *inter alia*, identification is weak.[1] In this paper we explore inference based upon our small-concentration asymptotic approximations and show that it does not suffer from the problems associated with more conventional techniques that have motivated recent interest in this special case; see the surveys of Stock, Wright, and Yogo (2002) and Hahn and Hausman (2002).

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. Section 3 presents various approximations to the sampling distribution of functions of the IV estimator based on the

[1] One interesting feature of the approximations of Poskitt and Skeels (2003) is their ability to capture many of the stylized facts that have been obtained under the different asymptotic paradigms that have been used to analyze weak identification. They provide a framework that goes some way towards unifying the qualitatively similar but technically distinct results of Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz, and Nelson (1998) and Startz, Nelson, and Zivot (2000), on the one hand, and Phillips (1989), Nelson and Startz (1990) and Choi and Phillips (1992) on the other. Similarly, results constructed using the many-instrument asymptotics of Bekker (1994) can also be obtained as a special case.

application of small-concentration asymptotics. Section 4 develops appropriate inferential procedures using the approximations given in Section 3, both hypothesis testing and confidence region construction are addressed. It is shown that the test statistics advanced are asymptotically pivotal and that the associated critical regions generate locally uniformly most powerful invariant tests. The confidence regions are also shown to be valid. In Section 5 we compare our proposed statistic to two others well-known in the literature. Section 6 presents a brief conclusion. All proofs are assembled in the Appendix.

2 The Model, Notation and Assumptions

Consider the classical structural equation model

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{u} \tag{2.1}$$

where the endogenous matrix variables $\mathbf{y}$ and $\mathbf{Y}$ are $N \times 1$ and $N \times n$, respectively, the matrix of exogenous variables $\mathbf{X}$ is $N \times k$, and $\mathbf{u}$ denotes a $N \times 1$ vector of uncorrelated stochastic disturbances with zero mean and variance σ_{u}^2. 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 $N \times K$ instrument set, where $\mathbf{Z}$ denotes a $N \times \nu$ matrix of instruments — exogenous regressors not appearing in equation (2.1) — and $K = k + \nu$, then we are interested in making inferences about $\boldsymbol{\beta}$ using the IV estimator

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Y}'\mathbf{P}\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{P}\mathbf{y}, \tag{2.2}$$

where $\mathbf{P} = \mathbf{P}_{[\mathrm{X\ Z}]} - \mathbf{P}_X = \mathbf{R}_{\mathrm{X}} - \mathbf{R}_{[\mathrm{X\ Z}]}$. For any $N \times k$ matrix $\mathbf{X}$ of full column rank $\mathbf{P}_{\mathrm{X}}$ denotes the idempotent, symmetric matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{R}_{\mathrm{X}} = \mathbf{I}_N - \mathbf{P}_{\mathrm{X}}$. The matrix $\mathbf{P}_{\mathrm{X}}$ is of course the $N \times N$ (prediction) operator of rank k that projects on to the space spanned by the columns of $\mathbf{X}$ and $\mathbf{R}_{\mathrm{X}}$ is the associated (residual) operator of rank $N - k$ which projects on to the orthogonal complement of that space. 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. In this case

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

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

The corresponding reduced form model is

$$[\mathbf{y}\ \mathbf{Y}] = [\mathbf{X}\ \mathbf{Z}] \begin{bmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\Pi}_1 \\ \boldsymbol{\pi}_2 & \boldsymbol{\Pi}_2 \end{bmatrix} + [\mathbf{v}\ \mathbf{V}]\,. \tag{2.3}$$

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

$$\boldsymbol{\Sigma} = \begin{bmatrix} \omega^2 & \boldsymbol{\omega}' \\ \boldsymbol{\omega} & \boldsymbol{\Omega} \end{bmatrix}, \tag{2.4}$$

ω^2 scalar, where $[\mathbf{v}\ \mathbf{V}]$ is partitioned conformably with $[\mathbf{y}\ \mathbf{Y}]$. We assume that $0 < \boldsymbol{\Sigma} < \infty$, meaning that the smallest and largest characteristic roots of $\boldsymbol{\Sigma}$ are positive but bounded, *viz.* $0 < \lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\max}(\mathbf{\Sigma}) < \infty$. The components of the reduced form coefficient matrix $\mathbf{\Pi}$ — namely $\boldsymbol{\pi}_1$, $\mathbf{\Pi}_1$, $\boldsymbol{\pi}_2$ and $\mathbf{\Pi}_2$ — are of dimension $k \times 1$, $k \times n$, $\nu \times 1$ and $\nu \times n$, respectively.

We will assume that sufficient regularity can be imposed to ensure that $\mathbf{S} = [\mathbf{y}\ \mathbf{Y}]'\mathbf{P}[\mathbf{y}\ \mathbf{Y}]$ has a non-central Wishart distribution with ν degrees of freedom, covariance matrix $\mathbf{\Sigma}$ and non-centrality parameter $\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Lambda}\mathbf{\Sigma}^{-\frac{1}{2}}$, $\mathcal{W}_{n+1}(\nu, \mathbf{\Sigma}, \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Lambda}\mathbf{\Sigma}^{-\frac{1}{2}})$,[2] where $\mathbf{\Sigma}^{\frac{1}{2}}$ is a symmetric matrix square root of $\mathbf{\Sigma}$ and

$$\mathbf{\Lambda} = [\boldsymbol{\pi}_2\ \mathbf{\Pi}_2]'\mathbf{Z}'\mathbf{R}_X\mathbf{Z}[\boldsymbol{\pi}_2\ \mathbf{\Pi}_2]\,. \tag{2.5}$$

We will assume that the usual compatibility conditions hold; namely

$$\boldsymbol{\pi}_1 - \mathbf{\Pi}_1\boldsymbol{\beta} = \gamma\,, \qquad \boldsymbol{\pi}_2 = \mathbf{\Pi}_2\boldsymbol{\beta}\,, \qquad \sigma_{\mathrm{u}}^2 = [1, -\boldsymbol{\beta}']\mathbf{\Sigma}[1, -\boldsymbol{\beta}']'\,. \tag{2.6}$$

It follows that

$$\mathbf{\Lambda} = [\boldsymbol{\beta}, \mathbf{I}_n]'\mathbf{\Pi}_2'\mathbf{Z}'\mathbf{R}_{\mathrm{X}}\mathbf{Z}\mathbf{\Pi}_2[\boldsymbol{\beta}, \mathbf{I}_n] = \begin{bmatrix} \delta^2 & \boldsymbol{\delta}' \\ \boldsymbol{\delta} & \mathbf{\Delta} \end{bmatrix}\,, \tag{2.7}$$

where the partition of $\mathbf{\Lambda}$ occurs after the first row and column, as in (2.4).

Exploiting properties of the Wishart distribution, in conjunction with the compatibility conditions (2.6), Poskitt and Skeels (2003) show that if $\nu^{-1}\|\mathbf{\Gamma}\|$ approaches zero, where $\mathbf{\Gamma} = \mathbf{\Omega}^{-\frac{1}{2}}\mathbf{\Delta}\mathbf{\Omega}^{-\frac{1}{2}}$, then $\widehat{\boldsymbol{\beta}}$ converges in probability to a random vector possessing an n-variate t-distribution with $\nu - n + 1$ degrees of freedom, location parameter

$$\boldsymbol{\mu}_\beta = \boldsymbol{\beta} + \mathbf{\Omega}^{-\frac{1}{2}}(\mathbf{I}_n + \nu^{-1}\mathbf{\Gamma})^{-1}\boldsymbol{\rho}\sigma_{\mathrm{u}} \tag{2.8}$$

and dispersion parameter

$$\mathbf{D}_\beta = \frac{[\mathbf{\Omega}^{\frac{1}{2}}(\mathbf{I}_n + \nu^{-1}\mathbf{\Gamma})\mathbf{\Omega}^{\frac{1}{2}}]}{\sigma_{\mathrm{u}}^2(1 - \boldsymbol{\rho}'(\mathbf{I}_n + \nu^{-1}\mathbf{\Gamma})^{-1}\boldsymbol{\rho})}\,, \tag{2.9}$$

where $\boldsymbol{\rho} = \mathbf{\Omega}^{-\frac{1}{2}}(\boldsymbol{\omega} - \mathbf{\Omega}\boldsymbol{\beta})/\sigma_{\mathrm{u}}$. Here, and in what follows, we use $\|\mathbf{A}\| = \sqrt{\mathrm{tr}\{\mathbf{A}'\mathbf{A}\}}$ to denote the Euclidean norm of a matrix $\mathbf{A}$. Henceforth we will use the tilde symbol $\sim$ underset with an a to denote convergence in probability to a random variable with the stated distribution. Thus we shall write

$$\widehat{\boldsymbol{\beta}} \underset{a}{\sim} \mathbf{t}_n(\nu - n + 1, \boldsymbol{\mu}_\beta, \mathbf{D}_\beta)\,.$$

Note that the distribution $\mathbf{t}_n(\nu - n + 1, \boldsymbol{\mu}_\beta, \mathbf{D}_\beta)$ has mean vector $\boldsymbol{\mu}_\beta$ and, for $\nu > n + 1$, variance-covariance matrix $[(\nu - n - 1)\mathbf{D}_\beta]^{-1}$, where the notation is designed to highlight the dependence of both the mean vector and covariance matrix on $\boldsymbol{\beta}$. If one thinks of ν as being fixed this result can be viewed as providing a small concentration asymptotic approximation since it is applicable as $\nu^{-1}\|\mathbf{\Gamma}\| \to 0$, as compared to the more conventional asymptotic normal approximation and Edgeworth type expansions which require that the concentration parameter $\mathbf{\Gamma}$ be large, see Rothenberg (1984). If one allows for the possibility of ν tending to infinity then this result can be used to demonstrate various aspects of many-instrument asymptotics; see Poskitt and Skeels (2003) for further discussion of this point.

[2]Clearly $\mathbf{S}$ will be non-central Wishart if $\mathrm{vec}[\mathbf{v}\ \mathbf{V}] \sim N(\mathbf{0}, \mathbf{\Sigma}\otimes\mathbf{I}_N)$, for then $\mathrm{vec}[\mathbf{y}\ \mathbf{Y}] \sim N(\mathrm{vec}([\mathbf{X}\ \mathbf{Z}]\mathbf{\Pi}), \mathbf{\Sigma}\otimes\mathbf{I}_N)$ and the result follows. It will also apply if $\mathrm{vec}[\mathbf{v}\ \mathbf{V}]$ has a distribution from the elliptically symmetric family. We might also expect $\mathbf{S}$ to be approximately non-central Wishart provided that the rows of $[\mathbf{y}\ \mathbf{Y}]$ satisfy an appropriate mixing condition, in which case the arguments underlying subsequent developments will still apply, with perhaps minor modifications.

3 Small Concentration Asymptotics

Let us begin by stating a basic result from Poskitt and Skeels (2003) which forms the foundation of subsequent developments.

Lemma 3.1. *Suppose that* $\mathbf{S} = [\mathbf{y}\ \mathbf{Y}]'\mathbf{P}[\mathbf{y}\ \mathbf{Y}] \sim \mathcal{W}_{n+1}(\nu, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-\frac{1}{2}})$. *Let*

$$\widehat{\mathbf{r}} = \{(\nu - n + 1)\mathbf{D}_\beta\}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\,, \tag{3.1}$$

where $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_\beta$ *and* $\mathbf{D}_\beta$ *are as defined in equations* (2.2), (2.8), *and* (2.9), *respectively. Then as* $\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0$ *the vector* $\widehat{\mathbf{r}}$ *converges in probability to a random variable* $\mathbf{r}$, *where the density function of* $\mathbf{r}$ *is given by*

$$f(\mathbf{r}) = \frac{\Gamma\left[\frac{\nu+1}{2}\right]}{[(\nu-n+1)\pi]^{n/2}\Gamma\left[\frac{\nu-n+1}{2}\right]}\left[1 + \frac{\mathbf{r}'\mathbf{r}}{(\nu-n+1)}\right]^{-(\nu+1)/2}. \tag{3.2}$$

Lemma 3.1 implies that $\widehat{\boldsymbol{\beta}}$ has approximately an n-variate t distribution with $\nu - n + 1$ degrees of freedom, location parameter $\boldsymbol{\mu}_\beta$ and dispersion parameter $\mathbf{D}_\beta$. For the purposes of implementation in subsequent inferential applications it proves useful to re-couch Lemma 3.1 in a different form.

Corollary 3.1. *Under the same conditions as in Lemma 3.1, the quadratic form*

$$(\nu - n + 1)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)/n$$

converges in distribution to Snedecor's F *distribution with degrees of freedom* n *and* $\nu - n + 1$ *as* $\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0$; *that is,*

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta) \underset{a}{\sim} \frac{n\mathcal{F}\{n, \nu - n + 1\}}{(\nu - n + 1)}\,.$$

In the Appendix we establish the following extension to Corollary 3.1.

Theorem 3.1. *Under the same conditions as in Lemma 3.1, the quadratic form*

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\Big|_{\beta=\beta_0} \underset{a}{\sim} \frac{\Psi\{n, (\nu - n + 1), \kappa_0\}}{(\nu - n + 1)q_0}$$

as $\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0$, *where* Ψ *denotes the distribution defined in Lemma A.1,*

$$\kappa_0 = (\nu - n + 1)(\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_{\beta_0})'\mathbf{D}_\beta(\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_{\beta_0})$$

and

$$q_0 = \frac{\sigma_{\mathrm{u}}^2(1 - \boldsymbol{\rho}'(\mathbf{I}_n + \nu^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\rho})\big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}}{\sigma_{\mathrm{u}}^2(1 - \boldsymbol{\rho}'(\mathbf{I}_n + \nu^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\rho})}\,.$$

The import of Theorem 3.1 is that it gives us the distribution of the quadratic form when calculated at an arbitrary point $\boldsymbol{\beta}_0$ in the parameter space, rather than when evaluated at the erstwhile true parameter point $\boldsymbol{\beta}$.

4 Inference

4.1 Hypothesis Testing

Consider testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. From Lemma 3.1, see also Corollary 3.1, it follows that an asymptotic size α critical region for testing $\mathcal{H}_0$ against $\mathcal{H}_1$ is given by

$$CR = \left\{ \widehat{\boldsymbol{\beta}} : (\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)' \mathbf{D}_\beta (\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta) \Big|_{\beta=\beta_0} \geq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right\}$$

where $\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}$ denotes the $(1 - \alpha)100\%$ percentile point of Snedecor's F distribution with n and $\nu - n + 1$ degrees of freedom. But CR is not a feasible critical region as it stands because $\boldsymbol{\mu}_\beta$ and $\mathbf{D}_\beta$ depend on the unknown parameter values $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$.

Fortunately the nuisance parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ can be consistently estimated from the reduced form. First,

$$\widehat{\boldsymbol{\Sigma}} = N^{-1}([\mathbf{y}, \mathbf{Y}]'(\mathbf{R}_X - \mathbf{R}_X\mathbf{Z}(\mathbf{Z}'\mathbf{R}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_X)[\mathbf{y}, \mathbf{Y}]\,,$$

the residual mean square from the first stage reduced form regression of the endogenous variables $[\mathbf{y}, \mathbf{Y}]$ on the exogenous variables $[\mathbf{X}, \mathbf{Z}]$, yields a consistent estimate of $\boldsymbol{\Sigma}$. Second,

$$N^{-1}\widehat{\boldsymbol{\Delta}} = N^{-1}(\mathbf{Y}'\mathbf{R}_X\mathbf{Z}(\mathbf{Z}'\mathbf{R}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_X\mathbf{Y})$$

provides a consistent estimate of $\lim_{N\to\infty} N^{-1}\boldsymbol{\Delta}$. Both $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Delta}}$ yield consistent estimates whatever the values of $\boldsymbol{\Pi}_2$ and $\boldsymbol{\Gamma}$, and they can clearly be used as "plug in" values for the nuisance parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ to obtain

$$\widehat{\boldsymbol{\mu}}_\beta = \boldsymbol{\beta} + (\widehat{\boldsymbol{\Omega}} + \nu^{-1}\widehat{\boldsymbol{\Delta}})^{-1}(\widehat{\boldsymbol{\omega}} - \widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})$$

and

$$\widehat{\mathbf{D}}_\beta = \frac{[\widehat{\boldsymbol{\Omega}} + \nu^{-1}\widehat{\boldsymbol{\Delta}}]}{\widehat{\sigma}^2_{\mathrm{u},\beta} - (\widehat{\boldsymbol{\omega}} - \widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})'(\widehat{\boldsymbol{\Omega}} + \nu^{-1}\widehat{\boldsymbol{\Delta}})^{-1}(\widehat{\boldsymbol{\omega}} - \widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})}\,,$$

where $\widehat{\sigma}^2_{\mathrm{u},\beta} = [1, -\boldsymbol{\beta}']\widehat{\boldsymbol{\Sigma}}[1, -\boldsymbol{\beta}']'$. We will therefore consider the statistical properties of inferential procedures based upon the quadratic form

$$PS(\boldsymbol{\beta}) = (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)\,,$$

which we will henceforth refer to as the PS-statistic.

4.1.1 Behaviour Under the Null Hypothesis

Theorem 4.1. *For all $\nu^{-1}\|\boldsymbol{\Gamma}\|$ sufficiently small the statistic $PS(\boldsymbol{\beta})$ is asymptotically pivotal and the subset of the sample space given by*

$$\widehat{CR} = \left\{ \widehat{\boldsymbol{\beta}} : PS(\boldsymbol{\beta})|_{\beta=\beta_0} \geq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right\}$$

defines an asymptotically similar critical region of size $\alpha \in (0, 1)$ for testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ against the alternative $\mathcal{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$.

We have already observed that weak identification and weak instruments manifests itself in the magnitude of the concentration parameter being small and that this has a deleterious effect on many standard techniques of inference. Since the Poskitt and Skeels (2003) approximation is designed to work well as $\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0$ it is of interest to examine the behaviour of $\widehat{CR}$ in more detail.

Substituting the expressions for $\widehat{\boldsymbol{\mu}}_\beta$ and $\widehat{\mathbf{D}}_\beta$ into $(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\mu}}_\beta)$ it is straightforward to establish that the quadratic form equals the ratio of

$$(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\widehat{\boldsymbol{\Omega}}+\nu^{-1}\widehat{\boldsymbol{\Delta}})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - 2(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta}) + (\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})'(\widehat{\boldsymbol{\Omega}}+\nu^{-1}\widehat{\boldsymbol{\Delta}})^{-1}(\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta}) \quad (4.1)$$

to

$$\widehat{d}_\beta = \widehat{\sigma}^2_{\mathrm{u},\beta} - (\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})'(\widehat{\boldsymbol{\Omega}}+\nu^{-1}\widehat{\boldsymbol{\Delta}})^{-1}(\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})\,. \quad (4.2)$$

Suppose that $\|\boldsymbol{\Gamma}\| < \gamma$. Then $\|\boldsymbol{\Delta}\| \le \|\boldsymbol{\Omega}\|\gamma$. Since $\widehat{\boldsymbol{\Delta}}$ is consistent for $\boldsymbol{\Delta}$, if γ is small then we can anticipate that $\|\widehat{\boldsymbol{\Delta}}\|$ will also be small. Now, as $\widehat{\boldsymbol{\Delta}}$ approaches zero it is obvious that $\widehat{\boldsymbol{\Omega}}+\nu^{-1}\widehat{\boldsymbol{\Delta}}$ approaches $\widehat{\boldsymbol{\Omega}}$ and a little algebra shows that the inverse $(\widehat{\boldsymbol{\Omega}}+\nu^{-1}\widehat{\boldsymbol{\Delta}})^{-1} = \widehat{\boldsymbol{\Omega}}^{-1} - \nu^{-1}\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Omega}}^{-1} + o(\|\widehat{\boldsymbol{\Delta}}\|/\nu)$.

Expanding and rearranging terms in (4.1) and (4.2), whilst making use of the fact that

$$\widehat{\sigma}^2_{\mathrm{u},\beta} = (\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})'\widehat{\boldsymbol{\Omega}}^{-1}(\widehat{\boldsymbol{\omega}}-\widehat{\boldsymbol{\Omega}}\boldsymbol{\beta}) + \widehat{\omega}^2 - \widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}\,,$$

we find that the numerator in (4.1) equals

$$(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) + \nu^{-1}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\widehat{\boldsymbol{\Delta}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) - \nu^{-1}(\boldsymbol{\beta}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Delta}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) + o(\|\widehat{\boldsymbol{\Delta}}\|/\nu)$$

and that the denominator $\widehat{d}_\beta$ equals

$$\widehat{\omega}^2 - \widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}} + \nu^{-1}(\boldsymbol{\beta}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Delta}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) + o(\|\widehat{\boldsymbol{\Delta}}\|/\nu)\,.$$

It follows that

$$(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\mu}}_\beta) = \widehat{L}_\beta + \frac{\nu^{-1}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\widehat{\boldsymbol{\Delta}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\widehat{\sigma}^2_{\mathrm{u}} + \nu^{-1}\widehat{Q}_\beta} + o(\|\widehat{\boldsymbol{\Delta}}\|/\nu)$$

where $\widehat{\sigma}^2_{\mathrm{u}} = \widehat{\omega}^2 - \widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$,

$$\widehat{Q}_\beta = (\boldsymbol{\beta}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Delta}}(\boldsymbol{\beta}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}),$$

the generalised distance of $\boldsymbol{\beta}$ from $\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$, and the lower bound

$$\widehat{L}_\beta = \frac{(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) - \nu^{-1}\widehat{Q}_\beta}{\widehat{\sigma}^2_{\mathrm{u}} + \nu^{-1}\widehat{Q}_\beta}\,.$$

Now, by rearranging the inequality $\widehat{L}_\beta > n\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\}/(\nu-n+1)$ we can see that the set $\{\boldsymbol{\beta} : \widehat{L}_\beta > n\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\}/(\nu-n+1)\}$ is equivalent to $\{\boldsymbol{\beta} : \widehat{Q}_\beta < \hat{q}_{(1-\alpha)}\}$ where

$$\hat{q}_{(1-\alpha)} = \left[\frac{(\nu-n+1)(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) - n\widehat{\sigma}^2_{\mathrm{u}}\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\}}{n\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\} + (\nu-n+1)}\right]\nu\,.$$

Thus, as $\|\widehat{\mathbf{\Delta}}\| \to 0$, the quadratic form $(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)$ will fall in the critical region for all $\boldsymbol{\beta} \in \{\boldsymbol{\beta} : \widehat{Q}_\beta < \hat{q}_{(1-\alpha)}\}$, no matter how small α may be. Hence we find that a test based on $\widehat{CR}$ will ultimately lead to the rejection of any hypothesized value $\boldsymbol{\beta}_0$ that lies in the interior of $\{\boldsymbol{\beta} : \widehat{Q}_\beta < \hat{q}_{(1-\alpha)}\}$, the elliptical region in $\mathbb{R}^n$ centred at $\widehat{\mathbf{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$, with principle axes of length $2(\hat{q}_{(1-\alpha)}/\lambda_{\max}(\widehat{\mathbf{\Delta}}))^{\frac{1}{2}}, \ldots, 2(\hat{q}_{(1-\alpha)}/\lambda_{\min}(\widehat{\mathbf{\Delta}}))^{\frac{1}{2}}$.

Such behaviour is not unreasonable. Clearly $\widehat{\mathbf{\Delta}}$ being small presents *prima facie* evidence that $\|\mathbf{\Delta}\|$, and therefore $\|\mathbf{\Gamma}\|$, is small and that the model is at best only weakly identified. When the model is weakly identified the estimate of the conditional regression of $\mathbf{y}$ on $\mathbf{Y}$ implicit in the reduced form, namely $\widehat{\mathbf{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$, will be close to the ordinary least squares estimate $(\mathbf{Y}'\mathbf{R}_X\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{R}_X\mathbf{y}$. The latter estimate of $\boldsymbol{\beta}$ is known to be inconsistent however. Hence, values of $\boldsymbol{\beta}$ that lie in a neighbourhood of $\widehat{\mathbf{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$ are unlikely to be sensible candidates for the true value, and only values of $\boldsymbol{\beta}$ that lie in a region of $\mathbb{R}^n$ that is *outside* a neighbourhood of $\widehat{\mathbf{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$ should, perhaps, be considered acceptable. For $\widehat{CR}$ the set $\{\boldsymbol{\beta} : \widehat{Q}_\beta < \hat{q}_{(1-\alpha)}\}$ belongs to the aforementioned neighbourhood and the region in the parameter space that will be consistent with the data is a subset of $\{\boldsymbol{\beta} : \widehat{Q}_\beta > \hat{q}_{(1-\alpha)}\}$.

4.1.2 Power Properties

We have seen that, for given values of the endogenous variables $[\mathbf{y}, \mathbf{Y}]$ and the exogenous variables $[\mathbf{X}, \mathbf{Z}]$, any hypothesized value $\boldsymbol{\beta}_0$ that lies in the region in the parameter space given by $\{\boldsymbol{\beta} : \widehat{Q}_\beta \le \hat{q}_{(1-\alpha)}\}$ will be rejected by the test $\widehat{CR}$. Although $\boldsymbol{\beta}$ is not itself random $\{\boldsymbol{\beta} : \widehat{Q}_\beta \le \hat{q}_{(1-\alpha)}\}$ is, of course, a realization of a random set, the randomness being a function of the distribution of the statistics $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{\Delta}}$ and $\widehat{\mathbf{\Sigma}}$ from which it is derived. Likewise, the probability that either CR or $\widehat{CR}$ leads to a rejection will also be governed by the distributional properties of these statistics. In particular, we are interested in the impact that the distributions of $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{\Delta}}$ and $\widehat{\mathbf{\Sigma}}$ have on the expected value of their respective indicator functions ϕ_{CR} and $\phi_{\widehat{CR}}$.[3]

Lemma 4.1. *Let $\pi_{CR}(\boldsymbol{\beta}) = E[\phi_{CR}]$ denote the power function of the test CR. Then*

$$\lim_{\nu^{-1}\|\mathbf{\Gamma}\| \to 0} \pi_{CR}(\boldsymbol{\beta}) = \int_{cv_{(1-\alpha)}}^{\infty} \psi(\xi : n, (\nu - n + 1), \kappa_0) d\xi$$

where the lower limit of integration $cv_{(1-\alpha)} = q_0 n \mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}$, and q_0 and the non-centrality parameter κ_0 are as defined in Theorem 3.1.

It follows from Lemma 4.1, on comparison of $\Psi\{n, \nu - n + 1, 0\}$ with $\Psi\{n, \nu - n + 1, \kappa\}$ for $\kappa > 0$, that CR defines an asymptotically unbiased test. To establish other desirable optimality properties, however, it is necessary to confine attention to a class of invariant tests.

For the classical structural equation model it is the space spanned by the columns of $\mathbf{Y}$ that is important in determining the internal structure of the model and when making inferences, and not the co-ordinate system chosen to represent that space. It is natural

[3]Strictly speaking, our notation should indicate that ϕ_{CR} and $\phi_{\widehat{CR}}$ are functions of the data $[\mathbf{y}, \mathbf{Y}]$ and $[\mathbf{X}, \mathbf{Z}]$, but this dependence is omitted for simplicity.

therefore to consider the class of tests that are invariant under transformations of the endogenous regressors from $\mathbf{Y}$ to $\mathbf{YG}$ where $\mathbf{G}$ is a member of the general linear group, $\mathcal{G}\ell_n$, the group of $n \times n$ nonsingular matrices with group operation matrix multiplication. The mapping $\mathbf{Y} \mapsto \mathbf{YG}$ induces the transformations $\boldsymbol{\beta} \mapsto \mathbf{G}^{-1}\boldsymbol{\beta}$, $\boldsymbol{\Pi}_2 \mapsto \boldsymbol{\Pi}_2\mathbf{G}$, $\boldsymbol{\omega} \mapsto \mathbf{G}'\boldsymbol{\omega}$, and $\boldsymbol{\Omega} \mapsto \mathbf{G}'\boldsymbol{\Omega}\mathbf{G}$ on the parameter space and it is straightforward to verify that $\boldsymbol{\mu}_\beta \mapsto \mathbf{G}^{-1}\boldsymbol{\mu}_\beta$ and $\mathbf{D}_\beta \mapsto \mathbf{G}'\mathbf{D}_\beta\mathbf{G}$, but the concentration parameter is invariant, that is $\boldsymbol{\Gamma} \mapsto \boldsymbol{\Gamma}$. On the space of the statistics the triple $\{\widehat{\boldsymbol{\beta}} : \widehat{\boldsymbol{\Delta}} : \widehat{\boldsymbol{\Sigma}}\} \mapsto \{\mathbf{G}^{-1}\widehat{\boldsymbol{\beta}} : \mathbf{G}'\widehat{\boldsymbol{\Delta}}\mathbf{G} : \text{diag}\{1, \mathbf{G}'\}\widehat{\boldsymbol{\Sigma}}\text{diag}\{1, \mathbf{G}\}\}$.

Now let $\phi_T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta, \mathbf{D}_\beta)$ denote the test function of a test T that depends on $\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta$ and $\mathbf{D}_\beta$, and suppose that T is invariant with respect to all transformations $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta) \mapsto \mathbf{G}^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$ and $\mathbf{D}_\beta \mapsto \mathbf{G}'\mathbf{D}_\beta\mathbf{G}$, $\mathbf{G} \in \mathcal{G}\ell_n$. Choosing $\mathbf{G} = \mathbf{D}_\beta^{-\frac{1}{2}}$ we have

$$\phi_T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta, \mathbf{D}_\beta) = \phi_T(\mathbf{D}_\beta^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta), \mathbf{I}),$$

and from the singular value decomposition of $\mathbf{D}_\beta^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$ it follows that there exists an $n \times n$ orthogonal matrix $\mathbf{U}$ such that

$$\mathbf{U}'\mathbf{D}_\beta^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta) = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot \sqrt{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)},$$

and therefore

$$\phi_T(\mathbf{D}_\beta^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta), \mathbf{I}) = \phi_T((1, 0, \ldots, 0)'\{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\}^{\frac{1}{2}}, \mathbf{I}).$$

Hence T must be a function of the quadratic form $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$. Let us therefore call such a test a nonsingular invariant quadratic test.

Theorem 4.2. *For all $\boldsymbol{\Gamma}$ sufficiently small the test CR is locally at least as powerful as any other nonsingular invariant quadratic test T of $\mathcal{H}_0$ with no greater size. That is, for any nonsingular invariant quadratic test T of $\mathcal{H}_0$ versus $\mathcal{H}_1$ with level of significance α the inequality*

$$\lim_{\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0} \pi_{CR}(\boldsymbol{\beta}) \geq \lim_{\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0} \pi_T(\boldsymbol{\beta})$$

holds for all $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ where $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \to 0$.

Following the arguments employed in Lehmann (1986, Chapter 6) it can be shown that conditioning on sufficient statistics entails no loss of power. The fact that no loss is incurred arises because, by the law of the iterated expectation, if S is a sufficient statistic and T is any test, then the test function

$$\phi_T(S) = E[\phi_T|S]$$

has power

$$\pi_{\phi_T(S)} = E\left[E[\phi_T|S]\right] = E[\phi_T] = \pi_{\phi_T},$$

which is exactly the same as the power of T. Thus, T is equivalent to a test based on the sufficient statistic S. The following extension of Theorem 4.2 is almost an immediate consequence of this generic result.

Theorem 4.3. *Of all tests of $\mathcal{H}_0$ versus $\mathcal{H}_1$ that are invariant under transformations of the endogenous regressors from $\mathbf{Y}$ to $\mathbf{YG}$, where $\mathbf{G} \in \mathcal{G}\ell_n$, the test $\widehat{CR}$ is asymptotically the locally uniformly most powerful as $\nu^{-1}\|\mathbf{\Gamma}\| \to 0$.*

To gain additional insight into the power that is likely to be achieved in practice let us consider the behaviour of the non-centrality parameter

$$\widehat{\kappa}_0 = (\nu - n + 1)(\widehat{\boldsymbol{\mu}}_\beta - \widehat{\boldsymbol{\mu}}_{\beta_0})'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\mu}}_\beta - \widehat{\boldsymbol{\mu}}_{\beta_0})\,.$$

Substituting the expression

$$\widehat{\boldsymbol{\mu}}_\beta - \widehat{\boldsymbol{\mu}}_{\beta_0} = \left[\mathbf{I} - (\widehat{\mathbf{\Omega}} + \nu^{-1}\widehat{\mathbf{\Delta}})^{-1}\widehat{\mathbf{\Omega}}\right](\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

into $\widehat{\kappa}_0$ and using the expansion

$$(\widehat{\mathbf{\Omega}} + \nu^{-1}\widehat{\mathbf{\Delta}})^{-1} = \widehat{\mathbf{\Omega}}^{-1} - \nu^{-1}\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Omega}}^{-1} + \nu^{-2}\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Omega}}^{-1} + o(\nu^{-2}\|\widehat{\mathbf{\Delta}}\|^2)$$

we find that

$$\widehat{\kappa}_0 = (\nu - n + 1)\frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\widehat{\mathbf{\Delta}}\widehat{\mathbf{\Omega}}^{-1}\widehat{\mathbf{\Delta}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{\nu^2\widehat{d}_\beta} + o(\nu^{-2}\|\widehat{\mathbf{\Delta}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\|^2)$$

where $\widehat{d}_\beta$ is defined in equation (4.2). The denominator $\widehat{d}_\beta$ converges to a constant as $\|\widehat{\mathbf{\Delta}}\| \to 0$ and the size of $\widehat{\kappa}_0$ is clearly controlled by the magnitude of $\nu^{-2}\|\widehat{\mathbf{\Delta}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\|^2$. Since $\widehat{\mathbf{\Delta}}$ is presumed to be small, it follows that $\widehat{\kappa}_0$ can still be local to zero even if $\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|$ is itself quite large. Given that we have shown that asymptotically $\widehat{CR}$ yields, in the terminology of Wald (1941), the locally most stringent invariant test, this suggests that $\widehat{CR}$ may exhibit good power properties over a range of values of $\boldsymbol{\beta}$ that deviate from $\boldsymbol{\beta}_0$ by a not inconsiderable margin.

4.2 Confidence Region Construction

Thinking of a confidence region as being equivalent to those values of $\boldsymbol{\beta}$ that are consistent with the data indicates that confidence regions with the correct asymptotic coverage probability can be constructed by inverting an asymptotically pivotal statistic.[4] Following the previous development and applying this idea leads us to the feasible confidence region

$$\widehat{CI} = \left\{\boldsymbol{\beta}_0 : (\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)\Big|_{\beta=\beta_0} < \frac{p\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)}\right\}\,.$$

The set $\widehat{CI}$ determines a $(1-\alpha)100\%$ confidence region for $\boldsymbol{\beta}$, but unlike confidence regions constructed in many standard situations, the region given by $\widehat{CI}$ does not equate to the interior of an ellipsoid in $\mathbb{R}^n$ because of the nonlinear manner in which the parameter $\boldsymbol{\beta}$ enters into both the location parameter $\widehat{\boldsymbol{\mu}}_\beta$ and the dispersion parameter $\widehat{\mathbf{D}}_\beta$. Thus, although $\widehat{\mathbf{D}}_\beta^{1/2}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)$ has a spherically symmetric distribution, the confidence sets derived from $(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)$ are not conventional Wald-type regions. Indeed, it follows from

[4] The view point taken here was first espoused by Neyman (1937) and is now commonly adopted.

our previous arguments that $\widehat{CI}$ will lie in the complement of $\{\boldsymbol{\beta} : \widehat{Q}_\beta < \hat{q}_{(1-\alpha)}\}$ and hence the region $\widehat{CI}$ need not be convex, nor connected. Nor need $\widehat{CI}$ be bounded.

To verify the latter point suppose that

$$(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) + \nu^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\widehat{\boldsymbol{\Delta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ \leq (\widehat{\sigma}_{\mathrm{u}}^2 + \nu^{-1}\widehat{Q}_\beta)\frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)}\,.$$

Then $\boldsymbol{\beta} \in \widehat{CI}$ and

$$\widehat{Q}_\beta \geq \hat{q}_{(1-\alpha)} + \left(\frac{(\nu - n + 1)}{(\nu - n + 1) + n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\})}\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\widehat{\boldsymbol{\Delta}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\,.$$

This inequality states that the generalised distance of $\boldsymbol{\beta}$ from $\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$ exceeds $\hat{q}_{(1-\alpha)}$ by the proportion $(\nu - n + 1)/((\nu - n + 1) + n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\})$ of its corresponding distance from $\widehat{\boldsymbol{\beta}}$. But if $\boldsymbol{\beta}$ satisfies the latter condition then so too does $\psi\boldsymbol{\beta}$ for any $\psi > 1$.

We have thus exhibited a subset of $\widehat{CI}$ in which $\|\boldsymbol{\beta}\|$, and therefore the diameter of the subset, is unbounded. Hence inferential procedures based upon $\widehat{CI}$ will not suffer from the problems described by Dufour (1997). Theorem 3.3 of Dufour (1997) is satisfied and confidence regions for $\boldsymbol{\beta}$ constructed using the PS-statistic are valid.

The phenomena described in the preceding paragraphs are clearly illustrated in Figure 1, which depicts the surface generated by the PS-statistic in a neighbourhood of $\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$. The figure is based on a hypothetical model in which $n = \nu = 2$, $\boldsymbol{\beta} = 12(1,1)'$,

$$\boldsymbol{\Sigma} = 12\begin{pmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{pmatrix} \text{ and } \boldsymbol{\Delta} = \begin{pmatrix} 0.07 & 0.05 \\ 0.05 & 0.05 \end{pmatrix}.$$

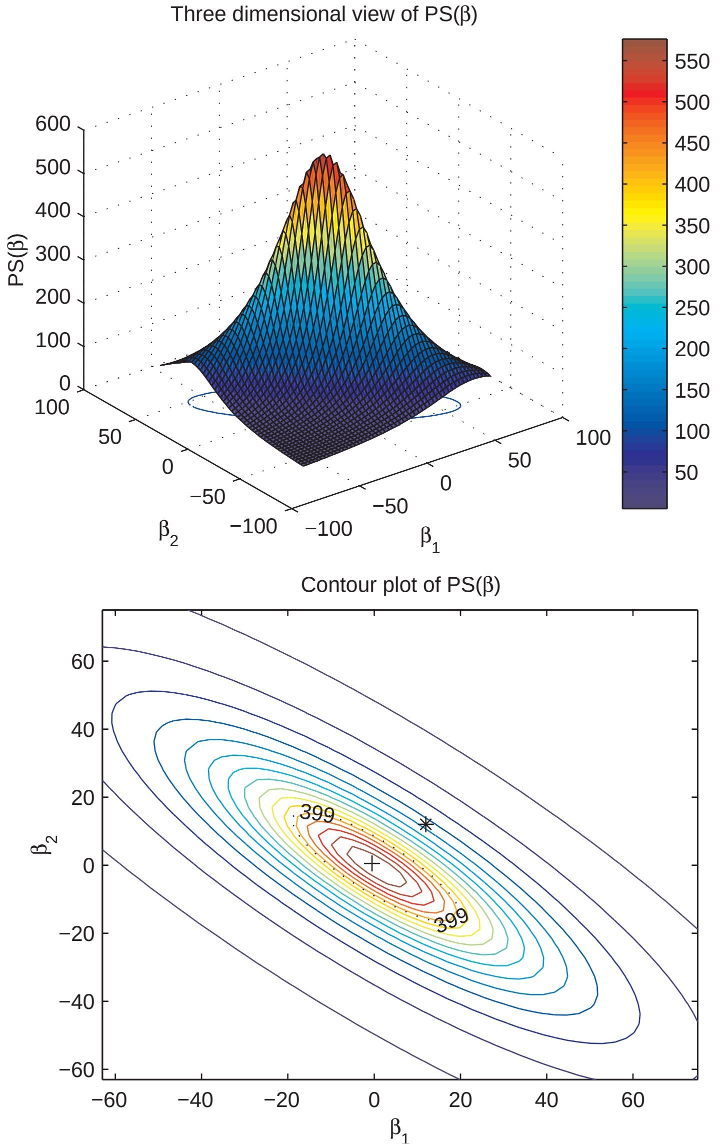
This gives $\|\boldsymbol{\Gamma}\| = 0.0092$ for the magnitude of the concentration parameter. The 95% confidence region $\widehat{CI}$ consists of all those $\boldsymbol{\beta}$'s that lie *outside* the area enclosed by the contour corresponding to the critical value $2\mathcal{F}_{(0.95)}\{2,1\} = 399.0$. Those $\boldsymbol{\beta}$'s that lie *inside* the region enclosed by the contour will be rejected by $\widehat{CR}$ at the 5% level of significance.

Note, in passing, that $\hat{q}_{(1-\alpha)}$ decreases monotonically with α, so that as the level of significance falls the volume of $\{\boldsymbol{\beta} : \widehat{Q}_\beta > \hat{q}_{(1-\alpha)}\}$ increases (*ceteris paribus*) and the set of values of $\boldsymbol{\beta}$ that are potentially consistent with the data increases. In the limit, of course, the model becomes totally unidentified as the concentration parameter approaches zero and $\boldsymbol{\beta}$ cannot be determined from the data. In this case the volume of $\{\boldsymbol{\beta} : \widehat{Q}_\beta < \hat{q}_{(1-\alpha)}\}$,

$$\frac{(\pi\hat{q}_{(1-\alpha)})^{n/2}}{\Gamma\left[\frac{n+1}{2}\right](\det\widehat{\boldsymbol{\Delta}})^{1/2}}\,,$$

will become unbounded as $\|\widehat{\boldsymbol{\Delta}}\| \to 0$ and ultimately all values of $\boldsymbol{\beta}$ will be deemed unacceptable, whatever the value of α. This type of behaviour is seen in Figure 2, where the contours of the PS-statistic are presented for the same hypothetical model as before, except that a redundent instrument has been added, implying that the model is partially unidentified. Virtually all $\boldsymbol{\beta}$ being considered are now rejected. Following Dufour (1997) we might interpret such an occurrence as providing evidence that the model is misspecified.

Figure 1: Graph of PS-statistic, weakly identified case. The asterisk denotes $\widehat{\boldsymbol{\beta}}$ and the plus sign $\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$. The black dotted contour represents the level curve at $2\mathcal{F}_{(0.95)}\{2,1\} = 399.0$.

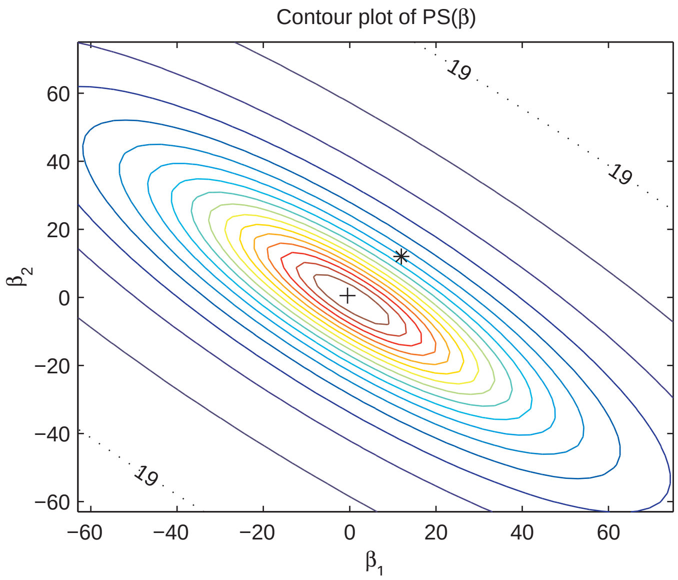


5 Other Procedures

Given that we have established certain optimality properties of our test statistic under the general linear group, $\mathcal{G}\ell_n$, it seems natural at this point to consider one or two other statistics that fall within the ambit of this analysis. An obvious candidate for consideration is the Anderson-Rubin statistic (Anderson and Rubin, 1949) which is, in the notation of our paper,

$$AR(\boldsymbol{\beta}) = \frac{(N-K)}{\nu}\,\frac{(\mathbf{y}-\mathbf{Y}\boldsymbol{\beta})'\mathbf{P}(\mathbf{y}-\mathbf{Y}\boldsymbol{\beta})}{(\mathbf{y}-\mathbf{Y}\boldsymbol{\beta})'\mathbf{R}_{[X,Z]}(\mathbf{y}-\mathbf{Y}\boldsymbol{\beta})}\,. \tag{5.1}$$

Figure 2: Graph of PS-statistic, partially unidentified case. The asterisk denotes $\widehat{\boldsymbol{\beta}}$ and the plus sign $\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}$. The black dotted contour represents the level curve at $\mathcal{F}_{(0.95)}\{2,2\} = 19.0$.



Under $\mathcal{H}_0$ the statistic $AR(\boldsymbol{\beta}_0)$ has an $\mathcal{F}\{\nu, N-K\}$ distribution, whatever the value of $\boldsymbol{\Gamma}$, and it provides an exact finite sample test of $\mathcal{H}_0$ verses $\mathcal{H}_1$ (under Gaussian assumptions) that is independent of nuisance parameters.

It is known, however, that the AR statistic has poor power when $\nu - n$ is large and Kleibergen (2002) suggests that the deficiency arises because the degrees of freedom of the numerator equals the number of instruments rather than the number of parameters under test. Kleibergen (2002) shows that if the projection matrix $\mathbf{P}$ in (5.1) is replaced by $\mathbf{P}_{\widetilde{Y}(\lambda)}$ where

$$\widetilde{\mathbf{Y}}(\boldsymbol{\lambda}) = \mathbf{P}(\mathbf{Y} - (\mathbf{y} - \mathbf{Y}\boldsymbol{\beta})\boldsymbol{\lambda})$$

and $\boldsymbol{\lambda} = (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})/\sigma_{\mathrm{u}}^2$, then the resulting statistic has an $\mathcal{F}\{n, N-K\}$ distribution when re-scaled by ν/n. The parameter $\boldsymbol{\lambda}$ is unknown, of course, but Kleibergen (2002) establishes that under $\mathcal{H}_1$ it can be replaced by the consistent estimate $\widehat{\boldsymbol{\lambda}} = (\widehat{\boldsymbol{\omega}} - \widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})/\widehat{\sigma}_{\mathrm{u},\beta}^2$ and the resulting statistic

$$K(\boldsymbol{\beta}) = (N-K)\frac{(\mathbf{y} - \mathbf{Y}\boldsymbol{\beta})'\mathbf{P}_{\widetilde{Y}(\widehat{\lambda})}(\mathbf{y} - \mathbf{Y}\boldsymbol{\beta})}{(\mathbf{y} - \mathbf{Y}\boldsymbol{\beta})'\mathbf{R}_{[X,Z]}(\mathbf{y} - \mathbf{Y}\boldsymbol{\beta})}$$

is asymptotically distributed as Chi-squared with n degrees of freedom, $\chi^2(n)$, as $N \to \infty$. Bekker and Kleibergen (2003) determine bounds for the exact distribution of $K(\boldsymbol{\beta})$ and suggest using critical values from the $\mathcal{F}\{n, N-K\}$ distribution for the lower and, when re-scaled by $N/(N-k)$, upper bound exact critical values of $K(\boldsymbol{\beta})/n$. We refer to Bekker and Kleibergen (2003) for more detailed particulars.

The arguments used in Bekker and Kleibergen (2003) depend upon an examination of the properties of $K(\boldsymbol{\beta})$ in the totally unidentified case. They show that if $\nu \to \infty$ as $N \to \infty$, such that $\nu/N \to \tau > 0$, then $(1 - \nu/N)K(\boldsymbol{\beta})$ has an asymptotic $\chi^2(n)$ distribution. Note that in this situation $(\nu - n + 1)PS(\boldsymbol{\beta})$ also has the same asymptotic $\chi^2(n)$ distribution. The ratio of the"denominator" degrees of freedom of the PS-statistic to that of the K-statistic, $(\nu - n+1)/(N-K)$, converges to $\tau/(1-\tau)$, however, and the difference $N^{-1}(\mathbf{y} - \mathbf{Y}\boldsymbol{\beta})'\mathbf{R}_{[X,Z]}(\mathbf{y} - \mathbf{Y}\boldsymbol{\beta}) - \widehat{d}_{\beta}$ equals

$$(\boldsymbol{\beta} - \widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}})'(\widehat{\boldsymbol{\Omega}} - \nu^{-1}\widehat{\boldsymbol{\Delta}})(\boldsymbol{\beta} - \widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\omega}}) \tag{5.2}$$

plus terms of order $o(\|\widehat{\boldsymbol{\Delta}}\|/\nu)$. The expression in (5.2) is $O(1)$ for all values of N and ν, no matter what the value of $\boldsymbol{\Gamma}$. Thus the PS-statistic and the K-statistic have a common asymptotic distribution with the minimal number of degrees of freedom, but $PS(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ are not equivalent, even asymptotically.

Now, it is a simple exercise to show that both $AR(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ are invariant under the mappings $\mathbf{Y} \mapsto \mathbf{YG}$ and $\boldsymbol{\beta} \mapsto \mathbf{G}^{-1}\boldsymbol{\beta}$ and the results in Section 4 indicate that the performance of $PS(\boldsymbol{\beta})$ will be at least as good as either of these two statistics in situations where $\boldsymbol{\Gamma}$ is small.

Finally, note that if $\widehat{\boldsymbol{\theta}} = \mathbf{A}\widehat{\boldsymbol{\beta}} + \mathbf{b}$, where $\mathbf{A}$ is a fixed $p \times n$ coefficient matrix of rank $p \leq n$ and $\mathbf{b}$ is a given p component constant vector, then $\widehat{\boldsymbol{\theta}}$ will also possess an asymptotic t distribution. This result follows from (3.1) on noting that

$$\widehat{\boldsymbol{\theta}} = \mathbf{A}\{(\nu - n + 1)\mathbf{D}_{\beta}\}^{-1/2}\widehat{\mathbf{r}} + (\mathbf{A}\boldsymbol{\mu}_{\beta} + \mathbf{b}),$$

where $\widehat{\mathbf{r}}$ converges in probability to $\mathbf{r}$, a standard n-variate t random variable with $\nu - n + 1$ degrees of freedom. It is well known, Cornish (1954), that the latter implies that $\mathbf{r}$ can be expressed as a vector of standard normal random variables divided by the square root of the ratio of an independent chi-squared random variable to its degrees of freedom. It follows via the same representation that, as $\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0$, the vector $\widehat{\boldsymbol{\theta}}$ converges in probability to a random variable that possesses a p-variate t distribution with $\nu - n + 1$ degrees of freedom, location parameter $\boldsymbol{\mu}_{\theta} = \mathbf{A}\boldsymbol{\mu}_{\beta} + \mathbf{b}$ and dispersion parameter $\mathbf{D}_{\theta} = (\mathbf{A}\mathbf{D}_{\beta}^{-1}\mathbf{A}')^{-1}$. Hence inferences about $\boldsymbol{\theta}$ can be made using the procedures described in Section 4 in conjunction with the more general statistic $PS(\boldsymbol{\theta}) = (\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\mu}}_{\theta})'\widehat{\mathbf{D}}_{\theta}(\widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\mu}}_{\theta})$. There is one important caveat to this observation. Suppose that one wishes to test hypotheses about $\boldsymbol{\theta}$ when some of the columns of $\mathbf{A}$ are zero, so that $\boldsymbol{\theta}$ is an affine combination of a subset of the coefficients in $\boldsymbol{\beta}$. Here both the location and dispersion parameters of the distribution of $\widehat{\boldsymbol{\theta}}$ still depend on all of the elements in $\boldsymbol{\beta}$ through $\boldsymbol{\rho}$, not just those restricted by the null hypothesis. Therefore, like the Anderson-Rubin and K statistics, *inter alia*, we must restrict attention to testing hypotheses which involve all of the elements in the $\boldsymbol{\beta}$.

6 Conclusion

The main contribution of this paper has been to provide the practitioner with optimal and valid methods of inductive inference that may be employed in conjunction with the IV estimator when the concentration parameter is small. These procedures are based on the application of small concentration asymptotic approximations that lead, interestingly enough, to a non-standard application of standard distributions, facilitating numerical implementation using commonly available software.

In practice, of course, the applied worker 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. Poskitt and Skeels (2004) present a method of ascertaining when the concentration parameter is small and hence when the use of small concentration asymptotic approximations are appropriate. They show that their measure provides a reliable guide to the magnitude of the concentration parameter that can be used to calibrate any subsequent statistical inference. Thus the relevance of the approximations is easy for practitioners to ascertain, making the inferential procedures considered in this paper potentially very useful for empirical work.

References

Anderson, T. W. and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Anderson, T. W. and Sawa, T. (1973), "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," *Econometrica*, 41, 683–714.

— (1979), "Evaluation of the Distribution Function of the Two-Stage Least Squares Estimate," *Econometrica*, 47, 163–182.

Bekker, P. and Kleibergen, F. (2003), "Finite-Sample Instrumental Variables Inference Using an Asymptotically Pivotal Statistic," *Econometric Theory*, 19, 744–753.

Bekker, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 58, 1443–1458.

Chao, J. C. and Swanson, N. R. (2002), "Consistent Estimation with a Large Number of Weak Instruments," Manuscript, University of Maryland.

— (2003), "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments," Manuscript, University of Maryland.

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

Cornish, E. A. (1954), "The Multivariate t-distribution Associated with a Set of Normal Random Deviates," *Australian Journal of Physics*, 7, 531–542.

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

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

Herz, C. S. (1955), "Bessel Functions of Matrix Argument," *Annals of Mathematics*, 61, 474–523.

Kleibergen, F. (2002), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70, 1781–1803.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, New York: Springer-Verlag, 2nd ed.

Nelson, C. R. and Startz, R. (1990), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business*, 63, S125–S140.

Neyman, J. (1937), "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Philosophical transactions of the Royal Society of London. Series A*, 236, 333–380.

Phillips, P. C. B. (1980), "The Exact Distribution of Instrumental Variable Estimators in an Equation Containing n+1 Endogenous Variables," *Econometrica*, 48, 861–878.

— (1989), "Partially Identified Edconometric Models," *Econometric Theory*, 5, 181–240.

Poskitt, D. S. and Skeels, C. L. (2003), "Approximating the Distribution of the Instrumental Variables Estimator when the Concentration Parameter is Small," Paper to be presented at the 14th EC^2 Conference, London.

— (2004), "Assessing the Magnitude of the Concentration Parameter in a Simultaneous Equations Model," .

Rothenberg, T. J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, eds. Griliches, Z. and Intriligator, M. D., Amsterdam: North Holland, vol. 2, chap. 15.

Sargan, J. D. and Mikhail, W. M. (1971), "A General Approximation to the Distribution of Instrumental Variables Estimates," *Econometrica*, 39, 131–169.

Slater, L. J. (1966), *Generalized Hypergeometric Functions*, Cambridge: Cambridge University Press.

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.

Stock, J. H. and Yogo, M. (2003), "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," Department of Economics, Harvard University.

Wald, A. (1941), "Asymptotically most powerful tests of statistical hypothesis," *Annals of Mathematical Statistics*, 12, 1–19.

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

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

Proof of Corollary 3.1 Let $\widehat{\mathbf{q}} = \mathbf{D}_{\beta}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_{\beta})$. Substituting into (3.2) using (3.1), noting that the Jacobian of the mapping from $\widehat{\boldsymbol{\beta}}$ to $\widehat{\mathbf{q}}$ is $|\mathbf{D}_{\beta}|^{-1/2}$, gives

$$\frac{\Gamma\left[\frac{\nu+1}{2}\right]}{[\pi]^{n/2}\Gamma\left[\frac{\nu-n+1}{2}\right]}\left[1+\widehat{\mathbf{q}}'\widehat{\mathbf{q}}\right]^{-(\nu+1)/2} \tag{A.1}$$

for the asymptotic distribution of $\widehat{\mathbf{q}}$. Transforming from rectangular to polar co-ordinates in (A.1), integrating with respect to the angular rotations, and applying Slutzky's theorem, we find that $\hat{q} = \widehat{\mathbf{q}}'\widehat{\mathbf{q}}$ converges in probability to q where the distribution of q is given by

$$\frac{\Gamma\left[\frac{\nu+1}{2}\right]}{\Gamma\left[\frac{n}{2}\right]\Gamma\left[\frac{\nu-n+1}{2}\right]} q^{n/2+1}(1+q)^{-(\nu+1)/2}\,.$$

The stated result now follows. □

Proof of Theorem 3.1 First observe that $\mathbf{D}_{\beta_0} = \mathbf{D}_{\beta}/q_0$, where

$$q_0 = \frac{\sigma_{\mathrm{u}}^2(1-\boldsymbol{\rho}'(\mathbf{I}_n+\nu^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\rho})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}}{\sigma_{\mathrm{u}}^2(1-\boldsymbol{\rho}'(\mathbf{I}_n+\nu^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\rho})}\,.$$

It follows that

$$(\widehat{\boldsymbol{\beta}}-\boldsymbol{\mu}_{\beta})'\mathbf{D}_{\beta}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\mu}_{\beta})|_{\beta=\beta_0} = \frac{\widehat{\mathbf{t}}_0'\widehat{\mathbf{t}}_0}{(\nu-n+1)q_0}$$

where

$$\widehat{\mathbf{t}}_0 = ((\nu-n+1)\mathbf{D}_{\beta})^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\mu}_{\beta_0}) = \widehat{\mathbf{r}}+\boldsymbol{\mu}\,,$$

with

$$\widehat{\mathbf{r}} = ((\nu-n+1)\mathbf{D}_{\beta})^{1/2}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\mu}_{\beta})$$

and

$$\boldsymbol{\mu} = ((\nu-n+1)\mathbf{D}_{\beta})^{1/2}(\boldsymbol{\mu}_{\beta}-\boldsymbol{\mu}_{\beta_0})\,.$$

As $\widehat{\boldsymbol{\beta}} \underset{a}{\sim} t_n((\nu-n+1),\boldsymbol{\mu}_{\beta},\mathbf{D}_{\beta})$, it follows from Slutsky's theorem and Lemma A.1 that

$$\widehat{\mathbf{t}}_0'\widehat{\mathbf{t}}_0 \underset{a}{\sim} \Psi\{n,\nu-n+1,\boldsymbol{\mu}'\boldsymbol{\mu}\}\,,$$

which establishes the desired result. □

Proof of Theorem 4.1 To begin, note via Lemma 3.1, or equivalently Corollary 3.1, that when $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ the distribution of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$ is independent of $\boldsymbol{\beta}$ as $\nu^{-1}\|\boldsymbol{\Gamma}\| \to 0$. But

$$(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\mu}}_\beta) = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta) + o_p(1) \tag{A.2}$$

and hence

$$PS(\boldsymbol{\beta}_0) \underset{a}{\sim} \frac{n\mathcal{F}\{n, \nu - n + 1\}}{(\nu - n + 1)}$$

as $N \to \infty$. Thus $PS(\boldsymbol{\beta})$ is a function of the data and $\boldsymbol{\beta}$ whose limiting null distribution is independent of $\boldsymbol{\beta}$ and hence it is asymptotically pivotal.

Now, by construction CR defines a critical region such that for given $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$ we have $\lim_{\nu^{-1}\|\boldsymbol{\Gamma}\|\to 0} P(CR) = \alpha$. Thus, for all $\epsilon > 0$ there exists a $\gamma > 0$ such that $\nu^{-1}\|\boldsymbol{\Gamma}\| < \gamma$ implies $|P(CR) - \alpha| < \frac{1}{2}\epsilon$. Suppose that $\nu^{-1}\|\boldsymbol{\Gamma}\| < \gamma$. Evaluating CR at $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Delta} = \widehat{\boldsymbol{\Delta}}$ gives $\widehat{CR}$. From (A.2) it follows that $\lim_{N\to\infty} P(\{CR \setminus \widehat{CR}\} \cup \{\widehat{CR} \setminus CR\}) = 0$, which implies that $P(\widehat{CR})$ lies in the interval $(\alpha - \epsilon, \alpha + \epsilon)$ as $N \to \infty$ and hence that $\lim_{N\to\infty} \lim_{\nu^{-1}\|\boldsymbol{\Gamma}\|\to 0} P(\widehat{CR}) = \alpha$. It follows that $\widehat{CR}$ defines an asymptotically similar critical region of size α, as stated. □

Proof of Lemma 4.1 The result follows directly from the definition of the critical region CR and Theorem 3.1. □

Proof of Theorem 4.2 A nonsingular invariant quadratic test of $\mathcal{H}_0$ is, by definition, a function of the quadratic form $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\Big|_{\beta=\beta_0}$. Now, it is easily verified that under the transformations induced by the mapping $\mathbf{Y} \mapsto \mathbf{YG}$ the quadratic form $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$ and the non-centrality parameter κ are invariant. Moreover, the hypotheses $\mathcal{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\mathcal{H}_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ imply that $\kappa = 0$ and $\kappa > 0$, respectively. From an application of Neyman-Pearson theory in conjunction with Theorem 3.1 we know: First, that the uniformly most powerful size α test of the hypotheses $\mathcal{H}_0 : \kappa = 0$ against $\mathcal{H}_1' : \kappa = \kappa' > 0$ is given by a critical region of the form $\{r : LR(r) > k_\alpha\}$ for suitably chosen k_α, where the likelihood ratio

$$LR(r) = q_0\psi(rq_0 : \nu - n + 1, n, \kappa')/\psi(r : \nu - n + 1, n, 0)$$

and $r = (\nu - n + 1)\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\Big|_{\beta=\beta_0}$; Second, that the locally uniformly most powerful test, ϕ_{LUMP}, is obtained by maximizing the gradient of $LR(r)$ in a neighbourhood of the null. It follows that the construction of the locally uniformly most powerful test for the current problem amounts to evaluating $\lim_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\to 0} \partial LR(r)/\partial\kappa^+$, where $\partial LR(r)/\partial\kappa^+$ denotes the right-hand derivative of $LR(r)$ with respect to κ at the origin. After the application of some basic analysis and algebraic manipulations we find that

$$\lim_{\|\boldsymbol{\beta}-\boldsymbol{\beta}_0\|\to 0} \frac{\partial LR(r)}{\partial\kappa^+} = \frac{\nu + 1}{2}\left[\frac{(\nu + 3)r}{n(\nu - n + 1 + r)^2} - \frac{1}{(\nu - n + 1 + r)}\right]. \tag{A.3}$$

The expression on the right hand side of (A.3) is a monotonically increasing function of r, implying that $\phi_{LUMP} = \{r : r > k_\alpha'\}$ for an appropriately chosen k_α'. This, however, is equivalent to the critical region specified by CR and hence the desired result is obtained. □

Proof of Theorem 4.3 Noting that $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\Delta}}$ are sufficient for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}$, and therefore $\widehat{\mu}_\beta$ and $\widehat{\mathbf{D}}_\beta$ are sufficient for μ_β and $\mathbf{D}_\beta$, it follows that any other test T of $\mathcal{H}_0$ against $\mathcal{H}_1$ is equivalent to a test based on $\widehat{\boldsymbol{\beta}} - \widehat{\mu}_\beta$ and $\widehat{\mathbf{D}}_\beta$ with test function $\phi_T(\widehat{\boldsymbol{\beta}} - \widehat{\mu}_\beta, \widehat{\mathbf{D}}_\beta)$, say. By assumption, however, T is invariant under the mapping $\mathbf{Y} \mapsto \mathbf{YG}$ and therefore T must be a function of $(\widehat{\boldsymbol{\beta}} - \widehat{\mu}_\beta)'\widehat{\mathbf{D}}_\beta(\widehat{\boldsymbol{\beta}} - \widehat{\mu}_\beta)\Big|_{\beta=\beta_0}$. From (A.2) we can conclude that T is asymptotically equivalent to a nonsingular invariant quadratic test and the result now follows from Theorem 4.2. □

Derivation of Non-Central Density

Lemma A.1. *Let $\mathbf{t} = \mathbf{r} + \boldsymbol{\mu}$, where $\mathbf{r} \sim t_p(a, \mathbf{0}, \mathbf{I})$; that is, $\mathbf{r}$ has a standard p-variate t distribution with a degrees of freedom, and $\boldsymbol{\mu}$ is an arbitrary vector. Then the probability distribution of $r = \mathbf{t}'\mathbf{t}$ will be denoted by $\Psi\{p, a, \kappa\}$ and the associated probability density function is given by*

$$\psi(r : p, a, \kappa) = \frac{a^{a/2}\Gamma\left(\frac{a+p}{2}\right)}{\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{p}{2}\right)} r^{(p-2)/2}[a+\kappa+r]^{-(a+p)/2} \times {}_2F_1\left(\frac{a+p}{4}, \frac{a+p+2}{4}; \frac{p}{2}; \frac{4\kappa r}{[a+\kappa+r]^2}\right) \quad \text{(A.4)}$$

wherein ${}_2F_1(k, l; m; x)$ denotes the hypergeometric function $\sum_{j=0}^{\infty} \frac{(k)_j (l)_j}{j!(m)_j} x^j$, with $(k)_j = \Gamma(k+j)/\Gamma(k)$, Pochhammer's forward factorial function, and $\kappa = \boldsymbol{\mu}'\boldsymbol{\mu}$.

Proof. Since the Jacobian of the transformation from $\mathbf{r}$ to $\mathbf{t} - \boldsymbol{\mu}$ is unity the probability density function of $\mathbf{t}$ is

$$f(\mathbf{t}) = \frac{\Gamma\left(\frac{a+p}{2}\right)}{(a\pi)^{p/2}\Gamma\left(\frac{a}{2}\right)} \left[1 + a^{-1}(\mathbf{t}-\boldsymbol{\mu})'(\mathbf{t}-\boldsymbol{\mu})\right]^{-(a+p)/2}.$$

Using the result

$$s^{-\alpha}\Gamma(\alpha) = \int_0^\infty e^{-sx}x^{\alpha-1}dx \quad \text{(A.5)}$$

we obtain

$$\begin{aligned} f(\mathbf{t}) &= c_1 \int_0^\infty \exp\{-[1 + a^{-1}(\mathbf{t}-\boldsymbol{\mu})'(\mathbf{t}-\boldsymbol{\mu})]x\} x^{(a+p-2)/2} dx \\ &= c_1 \int_0^\infty \exp\{-[1 + a^{-1}(\mathbf{tt} + \boldsymbol{\mu}'\boldsymbol{\mu})]x\} x^{(a+p-2)/2} \exp\{2a^{-1}x\boldsymbol{\mu}'\mathbf{t}\} dx, \end{aligned}$$

where $c_1 = \left[(a\pi)^{p/2}\Gamma\left(\frac{a}{2}\right)\right]^{-1}$. Next transform from $\mathbf{t}$ to $\mathbf{v}r^{1/2}$ where $\mathbf{v} = \mathbf{t}(\mathbf{t}'\mathbf{t})^{-1/2}$, $\mathbf{v}'\mathbf{v} = 1$, $r = \mathbf{t}'\mathbf{t} > 0$. The volume element becomes $d\mathbf{t} = 2^{-1}r^{(p-2)/2}d\mathbf{v}dr$ and hence

$$f(\mathbf{v}, r) = \frac{c_1}{2} r^{(p-2)/2} \int_0^\infty \exp\{-[1 + a^{-1}(r + \boldsymbol{\mu}'\boldsymbol{\mu})]x\} \, x^{(a+p-2)/2} \exp\{2a^{-1}r^{1/2}x\boldsymbol{\mu}'\mathbf{v}\} dx.$$

Averaging over the Stieffel manifold using Herz (1955, Lemma 3.7) we have

$$\int_{\mathbf{v}'\mathbf{v}=1} \exp\{\mathbf{v}'\mathbf{k}\} d\mathbf{v} = \frac{2\pi^{p/2}}{\Gamma\left(\frac{p}{2}\right)} {}_0F_1\left(\frac{p}{2}; \frac{1}{4}\mathbf{k}'\mathbf{k}\right)$$

for any fixed p-vector $\mathbf{k}$ and thus we obtain

$$f(r) = \frac{r^{(p-2)/2}}{a^{p/2}\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{p}{2}\right)} \int_0^\infty \exp\{-[1 + a^{-1}(r+\kappa)]x\} \\ x^{(a+p-2)/2} {}_0F_1\left(\frac{p}{2}; a^{-2}r\kappa x^2\right) dx\,. \quad \text{(A.6)}$$

where $\kappa = \boldsymbol{\mu}'\boldsymbol{\mu}$.

Expanding the hypergeometric function in (A.6) and using (A.5) to integrate term by term we now find that

$$\begin{aligned}
\mathcal{I} &= \int_0^\infty \exp\{-[1 + a^{-1}(r+\kappa)]x\} x^{(a+p-2)/2} {}_0F_1\left(\frac{p}{2}; a^{-2}r\kappa x^2\right) dx \\
&= \sum_{j=0}^\infty \frac{(a^{-2}r\kappa)^j}{j!\left(\frac{p}{2}\right)_j} \int_0^\infty \exp\{-[1 + a^{-1}(r+\kappa)]x\} x^{2j+(a+p-2)/2} dx \\
&= \sum_{j=0}^\infty \frac{(a^{-2}r\kappa)^j}{j!\left(\frac{p}{2}\right)_j} [1 + a^{-1}(r+\kappa)]^{-(2j+(a+p)/2)} \Gamma\left(2j + \frac{a+p}{2}\right) \\
&= \Gamma\left(\frac{a+p}{2}\right) [1 + a^{-1}(r+\kappa)]^{-(a+p)/2} \sum_{j=0}^\infty \frac{\left(\frac{a+p}{2}\right)_{2j}}{j!\left(\frac{p}{2}\right)_j} \left[\frac{\kappa r}{[a+\kappa+r]^2}\right]^j .
\end{aligned}$$

Finally, using the result $(c)_{2j} = (c/2)_j((c+1)/2)_j 2^{2j}$ (see Slater, 1966, I.25) we have

$$\begin{aligned}
\mathcal{I} &= \Gamma\left(\frac{a+p}{2}\right) [1 + a^{-1}(r+\kappa)]^{-(a+p)/2} \\
&\qquad \sum_{j=0}^\infty \frac{\left(\frac{a+p}{4}\right)_j \left(\frac{a+p+2}{4}\right)_j}{j!\left(\frac{p}{2}\right)_j} \left[\frac{4\kappa r}{[a+\kappa+r]^2}\right]^j \\
&= \Gamma\left(\frac{a+p}{2}\right) [1 + a^{-1}(r+\kappa)]^{-(a+p)/2} \\
&\qquad {}_2F_1\left(\frac{a+p}{4}, \frac{a+p+2}{4}; \frac{p}{2}; \frac{4\kappa r}{[a+\kappa+r]^2}\right) .
\end{aligned}$$

Substituting this result back into (A.6) yields the density as given in (A.4). □

Note that if $\kappa = 0$ then (A.4) collapses to

$$\frac{\Gamma\left(\frac{a+p}{2}\right)}{a^{p/2}\Gamma\left(\frac{a}{2}\right)\Gamma\left(\frac{p}{2}\right)} r^{(p-2)/2} \left[1 + \frac{r}{a}\right]^{-(a+p)/2} ,$$

which corresponds to the density function of the product of p times a random variable with the (central) $\mathcal{F}\{p, a\}$ distribution. We will therefore refer to κ as the non-centrality parameter. It should be emphasized, however, that the distribution Ψ does not equate to the standard non-central F distribution when $\kappa > 0$.