ISSN 0819-2642
ISBN 0 7340 2604 8

THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 948

SEPTEMBER 2005

Small Concentration Asymptotics and Instrumental Variables Inference

by

D. S. Poskitt
&
Christopher L. Skeels

Department of Economics
The University of Melbourne
Melbourne Victoria 3010
Australia.

Small Concentration Asymptotics and Instrumental Variables Inference ⋆

D.S. Poskitt [a] C.L. Skeels [b,*]

[a] *Department of Econometrics and Business Statistics, Monash University, Vic 3800, Australia*

[b] *Department of Economics, The University of Melbourne, Vic 3010, Australia*

Abstract

Poskitt and Skeels (2005) 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. We present approximations to the sampling distributions of various functions of the IV estimator based upon small-concentration asymptotics, and investigate hypothesis testing procedures and confidence region construction using these approximations. We explore the relationship between our work and the K statistic of Kleibergen (2002) and demonstrate that our results can be used to explain the sampling behaviour of the K statistic in simultaneous equations models where identification is weak.

Key words: simultaneous equations model, IV estimator, weak identification, weak instruments, small-concentration asymptotics
JEL codes: C10, C12, C13, C30.

1 Introduction

In a recent contribution to the literature on instrumental variables (IV) estimation Poskitt and Skeels (2005) present a new approximation to the exact sampling distribution of the IV estimator of the coefficients on the endogenous

⋆ Preliminary and not to be quoted without permission. We would like to thank Murray Smith and Giovanni Forchini for helpful comments.
* Corresponding author.
Email address: `Chris.Skeels@unimelb.edu.au` (C.L. Skeels).

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 the distributions of 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 aim of this paper is 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; see the surveys of Stock, Wright, and Yogo (2002) and Hahn and Hausman (2003). 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 (2005b,a) and Stock and Yogo (2005). 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 (2005) 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 small-concentration asymptotic approximations. Our results provide insight into the sampling be-

[1] One interesting feature of the approximations of Poskitt and Skeels (2005) 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 Zivot, Startz, and Nelson (2005), 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.

haviour of some existing techniques that have been developed to deal with this special case.

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 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. In Section 5 we discuss the practical implemetation of our results and, in particular, relate them to the K statistic of Kleibergen (2002). We also provide some examples which illustrate that our analysis goes some way towards explaining the sampling characteristics of K, and other statistics, in the presence of weak instruments. Section 6 presents a brief conclusion. Finally, Appendix A establishes a non-central density referred to in Theorem 3 and all other proofs are presented in Appendix B.

2 The Model, Notation and Assumptions

Consider the classical structural equation model

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{u} \tag{1}$$

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 (1) — and $K = k + \nu$, then we are interested in making inferences about $\boldsymbol{\beta}$ using the IV estimator

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

where $\mathbf{P} = \mathbf{P}_{[\mathrm{X}\ \mathrm{Z}]} - \mathbf{P}_X = \mathbf{R}_{\mathrm{X}} - \mathbf{R}_{[\mathrm{X}\ \mathrm{Z}]}$. 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

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

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

$$\boldsymbol{\Sigma} = \begin{bmatrix} \omega^2 & \boldsymbol{\omega}' \\ \boldsymbol{\omega} & \boldsymbol{\Omega} \end{bmatrix}, \tag{4}$$

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.* $0 < \lambda_{\min}(\boldsymbol{\Sigma}) \leq \lambda_{\max}(\boldsymbol{\Sigma}) < \infty$. The components of the reduced form coefficient matrix $\boldsymbol{\Pi}$ — namely $\boldsymbol{\pi}_1$, $\boldsymbol{\Pi}_1$, $\boldsymbol{\pi}_2$ and $\boldsymbol{\Pi}_2$ — are of dimension $k \times 1$, $k \times n$, $\nu \times 1$ and $\nu \times n$, respectively.

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

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

which shall be denoted $\mathbf{S} \sim \mathcal{W}_{n+1}(\nu, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1/2})$.[2] We will also assume that the usual compatibility conditions hold; namely

$$\boldsymbol{\pi}_1 - \boldsymbol{\Pi}_1\boldsymbol{\beta} = \gamma\,, \qquad \boldsymbol{\pi}_2 = \boldsymbol{\Pi}_2\boldsymbol{\beta}\,, \qquad \sigma_{\mathrm{u}}^2 = \sigma_{\mathrm{u},\boldsymbol{\beta}}^2 \equiv [1, -\boldsymbol{\beta}']\boldsymbol{\Sigma}[1, -\boldsymbol{\beta}']'\,. \tag{5}$$

It follows that

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

where the partition of $\boldsymbol{\Lambda}$ occurs after the first row and column, as in (4).

[2] Clearly $\mathbf{S}$ will be non-central Wishart if $\mathrm{vec}[\mathbf{v}\ \mathbf{V}] \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_N)$, for then $\mathrm{vec}[\mathbf{y}\ \mathbf{Y}] \sim N(\mathrm{vec}([\mathbf{X}\ \mathbf{Z}]\boldsymbol{\Pi}), \boldsymbol{\Sigma} \otimes \mathbf{I}_N)$ 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.

Exploiting properties of the Wishart distribution, in conjunction with the compatibility conditions (5), Poskitt and Skeels (2005) show that if $\nu^{-1}\|\boldsymbol{\Delta}\|$ approaches zero 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

$$\boldsymbol{\mu}_\beta = \boldsymbol{\beta} + (\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})^{-1}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta}) \tag{6}$$

and dispersion parameter

$$\mathbf{D}_\beta = \frac{\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta}}{\sigma^2_{\mathrm{u},\boldsymbol{\beta}} - (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'(\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})^{-1}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})}\,, \tag{7}$$

Here, and in what follows, we use $\|\mathbf{A}\| = \sqrt{\mathrm{tr}\{\mathbf{A}'\mathbf{A}\}}$ to denote the Euclidean norm of a matrix $\mathbf{A}$. We will also 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 $\nu^{-1}\|\boldsymbol{\Delta}\| \to 0$, as compared to the more conventional asymptotic normal approximation and Edgeworth type expansions which require that the concentration parameter $\boldsymbol{\Delta}$ 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 (2005) for further discussion of this point.

3 Small Concentration Asymptotic Results

Let us begin by stating a basic result from Poskitt and Skeels (2005) which forms the foundation of subsequent developments.

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

$$\widehat{\mathbf{r}} = \{(\nu - n + 1)\mathbf{D}_\beta\}^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\,, \tag{8}$$

where $\widehat{\boldsymbol{\beta}}$, $\boldsymbol{\mu}_\beta$ *and* $\mathbf{D}_\beta$ *are as defined in equations* (2), (6), *and* (7), *respectively. Then as* $\nu^{-1}\|\boldsymbol{\Delta}\| \to 0$ *the vector* $\widehat{\mathbf{r}}$ *converges in probability to a random variable*

$\mathbf{r}$, *where the density function of* $\mathbf{r}$ *is given by*

$$f(\mathbf{r}) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{[(\nu-n+1)\pi]^{n/2}\Gamma\left(\frac{\nu-n+1}{2}\right)} \left[1 + \frac{\mathbf{r}'\mathbf{r}}{(\nu-n+1)}\right]^{-(\nu+1)/2} . \tag{9}$$

Lemma 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 1 in a different form.

Corollary 2 *Under the same conditions as in Lemma 1, the quadratic form* $(\nu-n+1)\mathcal{Q}(\boldsymbol{\beta})/n$ *converges in distribution to Snedecor's* F *distribution with degrees of freedom* n *and* $\nu-n+1$ *as* $\nu^{-1}\|\boldsymbol{\Delta}\| \to 0$, *where*

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

that is,

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

In the Appendix we establish both Corollary 2 and the following extension to it.

Theorem 3 *Under the same conditions as in Lemma 1, the quadratic form*

$$\mathcal{Q}(\boldsymbol{\beta}_0) = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \underset{a}{\sim} \frac{\Psi\{n, \nu-n+1, \kappa_0\}}{(\nu-n+1)q_0}$$

as $\nu^{-1}\|\boldsymbol{\Delta}\| \to 0$, *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

$$q_0 = \frac{\left[\sigma^2_{\mathrm{u},\boldsymbol{\beta}} - (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'(\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})^{-1}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta}))\right]_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}}{\sigma^2_{\mathrm{u},\boldsymbol{\beta}} - (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'(\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})^{-1}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})} .$$

The import of Theorem 3 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 either Theorem 3 or Corollary 2 it follows that an asymptotic size α critical region for testing $\mathcal{H}_0$ against $\mathcal{H}_1$ is given by

$$CR = \left\{ \widehat{\boldsymbol{\beta}} : \mathcal{Q}(\boldsymbol{\beta}_0) \geq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right\}$$

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. We will therefore consider the statistical properties of inferential procedures based upon the quadratic form $\mathcal{Q}(\boldsymbol{\beta})$.

4.1.1 Behaviour Under the Null Hypothesis

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

$$\begin{aligned}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\boldsymbol{\Omega}_\beta + \nu^{-1}\boldsymbol{\Delta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - 2(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta}) \\ + (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'(\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})^{-1}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})\end{aligned} \tag{10}$$

to

$$d_\beta = \sigma^2_{\mathrm{u},\beta} - (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'(\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})^{-1}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})\,. \tag{11}$$

As $\boldsymbol{\Delta}$ approaches zero it is obvious that $\boldsymbol{\Omega}+\nu^{-1}\boldsymbol{\Delta}$ approaches $\boldsymbol{\Omega}$ and a little algebra shows that the inverse $(\boldsymbol{\Omega}+\nu^{-1}\boldsymbol{\Delta})^{-1} = \boldsymbol{\Omega}^{-1} - \nu^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1} + o(\|\boldsymbol{\Delta}\|/\nu)$.

Expanding and rearranging terms in (10) and (11), whilst making use of the fact that

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

we find that the numerator in (10) equals

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

and that the denominator d_β equals

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

It follows that

$$\mathcal{Q}(\boldsymbol{\beta}) = \frac{L_\beta - \nu^{-1}Q_\beta}{\sigma_{\mathrm{u}}^2 + \nu^{-1}Q_\beta} + o(\|\boldsymbol{\Delta}\|/\nu)$$

where $\sigma_{\mathrm{u}}^2 = \omega^2 - \boldsymbol{\omega}'\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$, $Q_\beta = (\boldsymbol{\beta} - \boldsymbol{\Omega}^{-1}\boldsymbol{\omega})'\boldsymbol{\Delta}(\boldsymbol{\beta} - \boldsymbol{\Omega}^{-1}\boldsymbol{\omega})$ and

$$L_\beta = (\widehat{\boldsymbol{\beta}} - \boldsymbol{\Omega}^{-1}\boldsymbol{\omega})'\boldsymbol{\Omega}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\Omega}^{-1}\boldsymbol{\omega}) + \nu^{-1}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{\Delta}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\,.$$

Now, restricting attention to the null behaviour of $\mathcal{Q}(\boldsymbol{\beta})$, rearranging the inequality $\mathcal{Q}(\boldsymbol{\beta}_0) \geq n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}/(\nu - n + 1)$ we can see that the set $\{\boldsymbol{\beta} : \mathcal{Q}(\boldsymbol{\beta}_0) \geq n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}/(\nu - n + 1)\}$ is equivalent to $\{\boldsymbol{\beta} : Q_{\beta_0} \leq q_{(1-\alpha)}\}$ where

$$q_{(1-\alpha)} = \left[\frac{(\nu - n + 1)L_{\beta_0} - n\sigma_{\mathrm{u}}^2\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\} + (\nu - n + 1)}\right]\nu\,.$$

Thus, as $\|\boldsymbol{\Delta}\| \to 0$, $\mathcal{Q}(\boldsymbol{\beta}_0)$ will fall in the critical region for all $\boldsymbol{\beta} \in \{\boldsymbol{\beta} : Q_{\beta_0} \leq q_{(1-\alpha)}\}$, no matter how small α may be. Hence we find that a test based on CR will ultimately lead to the rejection of any hypothesized value $\boldsymbol{\beta}_0$ that lies in the interior of $\{\boldsymbol{\beta} : Q_{\beta_0} \leq q_{(1-\alpha)}\}$, the elliptical region in $\mathbb{R}^n$ centred at $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$, with principle axes of length $2(q_{(1-\alpha)}/\lambda_{\max}(\boldsymbol{\Delta}))^{1/2}, \ldots, 2(q_{(1-\alpha)}/\lambda_{\min}(\boldsymbol{\Delta}))^{1/2}$.

Such behaviour is not unreasonable. Unless $\boldsymbol{\beta} = \boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$, so that OLS is unbiased for $\boldsymbol{\beta}$, we know that values of $\boldsymbol{\beta}$ that lie in a neighbourhood of $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$ 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 $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$ should, perhaps, be considered acceptable. If $\boldsymbol{\beta} = \boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$, $Q_{\beta_0} = 0$ and CR reduces to a somewhat more familiar F-type rejection region:

$$\frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'(\boldsymbol{\Omega} + \nu^{-1}\boldsymbol{\Delta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)/n}{\sigma_{\mathrm{u}}^2/(\nu - n + 1)} \geq \mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}\,.$$

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 $\{\boldsymbol{\beta} : Q_\beta \leq q_{(1-\alpha)}\}$ will be rejected by tests based on CR. Although $\boldsymbol{\beta}$ is not itself random $\{\boldsymbol{\beta} : Q_\beta \leq q_{(1-\alpha)}\}$ is, of course, a realization of a random set, the randomness being a function of the distribution of the statistic $\widehat{\boldsymbol{\beta}}$ from which it is derived. Likewise, the probability that CR leads to a rejection will also be governed by the distributional properties of $\widehat{\boldsymbol{\beta}}$. In particular, we are interested in the impact that the distribution of $\widehat{\boldsymbol{\beta}}$ has

on the expected value of the indicator function ϕ_{CR}.[3]

Lemma 4 *Let*

$$\phi_{CR} = \begin{cases} 1, & \text{if } \hat{\boldsymbol{\beta}} \in CR, \\ 0, & \text{otherwise,} \end{cases}$$

so that $\pi_{CR}(\boldsymbol{\beta}) = E[\phi_{CR}]$ *denotes the power function of the test CR. Then*

$$\lim_{\nu^{-1}\|\boldsymbol{\Delta}\| \to 0} \pi_{CR}(\boldsymbol{\beta}) = 1 - \Psi\{\varrho_0; n, \nu - n + 1, \kappa_0\}$$

where q_0 *and* κ_0 *are as defined in Theorem 3, and* $\varrho_0 = nq_0\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}$.

It follows from Lemma 4 that CR defines a test procedure with a power function that is increasing in κ_0, but is inversely related to the magnitude of q_0. This suggests that $\pi_{CR}(\boldsymbol{\beta})$ could be subject to countervailing forces as the magnitudes of $\kappa_0 \geq 0$ and $q_0 \geq 0$ change as functions of $\boldsymbol{\beta}$.

To gain additional insight into the power of CR let us consider the behaviour of the non-centrality parameter κ_0. Substituting the expression

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

into κ_0, and using the second-order expansion

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

we find that

$$\kappa_0 = (\nu - n + 1)\frac{(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}{\nu^2 d_\beta} + o(\nu^{-2}\|\boldsymbol{\Delta}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\|^2)$$

where d_β is defined in equation (11). The denominator d_β converges to a constant as $\|\boldsymbol{\Delta}\| \to 0$ and the size of κ_0 is clearly controlled by the magnitude of $\nu^{-2}\|\boldsymbol{\Delta}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\|^2$. Since $\boldsymbol{\Delta}$ is presumed to be small, it follows that κ_0 can still be local to zero even if $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|$ is itself quite large.

Similarly, the difference $d_\beta - d_{\beta_0}$ equals

$$\begin{aligned}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'\left(\nu^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1} - \nu^{-2}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\right)(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta}) \\ - (\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta}_0)'\left(\nu^{-1}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1} - \nu^{-2}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\right)(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta}_0)\end{aligned}$$

plus terms $o(\nu^{-2}\|\boldsymbol{\Delta}\|^2)$. In general it is not possible to sign this difference, implying that $q_0 = d_{\boldsymbol{\beta}_0}/d_{\boldsymbol{\beta}}$ can be greater than, equal to, or less than one.

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

Hence changes in q_0 can either amplify or attenuate any changes in $\pi_{CR}(\boldsymbol{\beta})$ induced by changes in κ_0.

The upshot of these observations is that although CR may be, in the terminology of Wald (1941), a locally stringent test in some directions in the parameter space, it is possible for CR to exhibit low power even when the true value $\boldsymbol{\beta}$ deviates from $\boldsymbol{\beta}_0$ by a considerable 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 the critical region CR.[4] Following the previous development and applying this idea leads us to the confidence region

$$CI = \left\{ \boldsymbol{\beta} : \mathcal{Q}(\boldsymbol{\beta}) \leq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right\} .$$

The set CI determines a $(1-\alpha)100\%$ confidence region for $\boldsymbol{\beta}$, but unlike confidence regions constructed in many standard situations, the region given by 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 $\boldsymbol{\mu}_\beta$ and the dispersion parameter $\mathbf{D}_\beta$. Thus, although $\mathbf{D}_\beta^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$ has a spherically symmetric distribution, the confidence sets derived from $\mathcal{Q}(\boldsymbol{\beta})$ are not conventional Wald-type regions. Indeed, it follows from our previous arguments that CI will lie in the complement of $\{\boldsymbol{\beta} : Q_\beta \leq q_{(1-\alpha)}\}$ and hence the region CI need not be either convex or connected. Nor need CI be bounded.

To verify the latter point suppose that

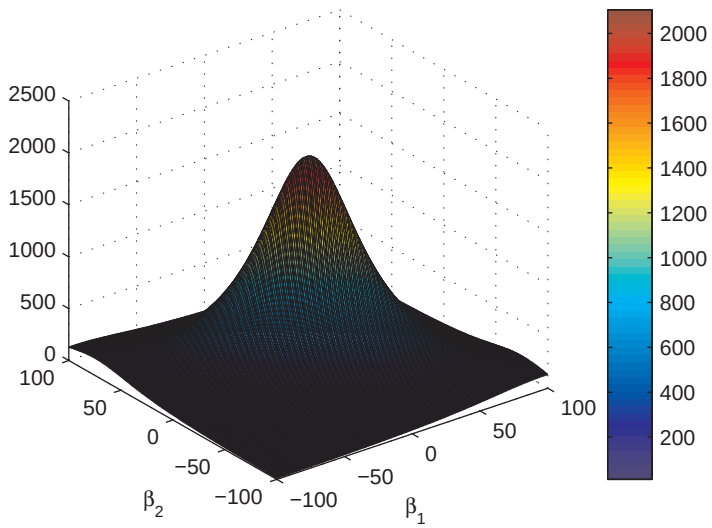
$$L_\beta \leq (\sigma_{\mathrm{u}}^2 + \nu^{-1} Q_\beta) \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} .$$

so that $\boldsymbol{\beta} \in CI$. Tedious rearrangement of the inequality $Q_\beta \geq q_{(1-\alpha)}$ yields

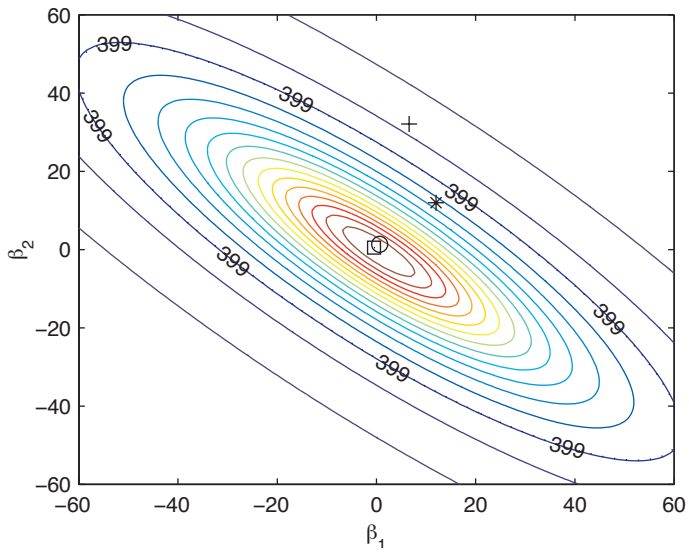
$$(1-\epsilon)\boldsymbol{\beta}'\boldsymbol{\Delta}\boldsymbol{\beta} + 2\boldsymbol{\beta}'\boldsymbol{\Delta}(\epsilon\widehat{\boldsymbol{\beta}} - \boldsymbol{\Omega}^{-1}\boldsymbol{\omega}) \geq \text{terms not involving } \boldsymbol{\beta},$$

where $0 < \epsilon = (\nu - n + 1)/[(\nu - n + 1) + n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}] < 1$. But if $\boldsymbol{\beta}$ satisfies this latter condition then so too does $\psi\boldsymbol{\beta}$ for any $\psi > 1$. We have thus exhibited a subset of CI in which $\|\boldsymbol{\beta}\|$, and therefore the diameter

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



(a) Three-dimensional plot of $\mathcal{Q}(\boldsymbol{\beta})$



(b) Contour plot of $\mathcal{Q}(\boldsymbol{\beta})$

Fig. 1. Graphs of $\mathcal{Q}(\boldsymbol{\beta})$: weakly identified case. The asterisk denotes $\boldsymbol{\beta}_0$, the plus sign $\widehat{\boldsymbol{\beta}}$, the circle $(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{y}$ and the square $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$. The black dotted contour represents the level curve at $2\mathcal{F}_{(0.95)}\{2,1\} = 399.0$.

of the subset, is unbounded. Hence inferential procedures based upon *CI* will not suffer from problems described by Dufour (1997). Theorem 3.3 of Dufour (1997) is satisfied and confidence regions for $\boldsymbol{\beta}$ constructed using $\mathcal{Q}(\boldsymbol{\beta})$ would be valid.

The phenomena described in the preceding paragraphs are clearly illustrated in Figure 1, which depicts the surface generated by the $\mathcal{Q}(\boldsymbol{\beta})$ in a neighbourhood of the ordinary least squares value $(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{y}$. The figure is based on a

hypothetical model in which $n = \nu = 2$, $\boldsymbol{\beta} = 12(1,1)'$,

$$\boldsymbol{\Sigma} = 12 \begin{bmatrix} 1 & -0.5 & 0.5 \\ -0.5 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Delta} = \begin{bmatrix} 0.0678 & 0.0508 \\ 0.0508 & 0.0520 \end{bmatrix}.$$

This gives $\|\boldsymbol{\Delta}\| = 0.1114$. The 95% confidence region *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 *CR* at the 5% level of significance. Note that the least squares value and $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$ lie roughly at the centre of this region. Because the model is weakly identified the conditional regression of $\mathbf{y}$ on $\mathbf{Y}$ implicit in the reduced form, namely $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$, is close to the ordinary least squares estimate. The latter estimate of $\boldsymbol{\beta}$ is known to be inconsistent however. And so it is natural to reject those values of $\boldsymbol{\beta}$ that lie in a neighbourhood of these two, almost coincident, points.

Note, in passing, that $q_{(1-\alpha)}$ decreases monotonically with α, so that as the level of significance falls the volume of $\{\boldsymbol{\beta} : Q_\beta \geq q_{(1-\alpha)}\}$ 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 $\{\boldsymbol{\beta} : Q_\beta \leq q_{(1-\alpha)}\}$,

$$\frac{(\pi q_{(1-\alpha)})^{n/2}}{\Gamma\left(\frac{n+1}{2}\right)(\det \boldsymbol{\Delta})^{1/2}},$$

will become unbounded as $\|\boldsymbol{\Delta}\| \to 0$ 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 $\mathcal{Q}(\boldsymbol{\beta})$ are presented for the same hypothetical model as before, except that a redundant 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.

5 Practical Implementation

In virtual all conceivable circumstances the construction of *CR* and *CI* will not be feasible because $\boldsymbol{\Omega}$ and $\boldsymbol{\Delta}$ will be unknown. Consequently we need to adapt our previous arguments to allow for this fact. The following result provides a parallel to Lemma 1 and Corollary 2 that provides a step in this direction.

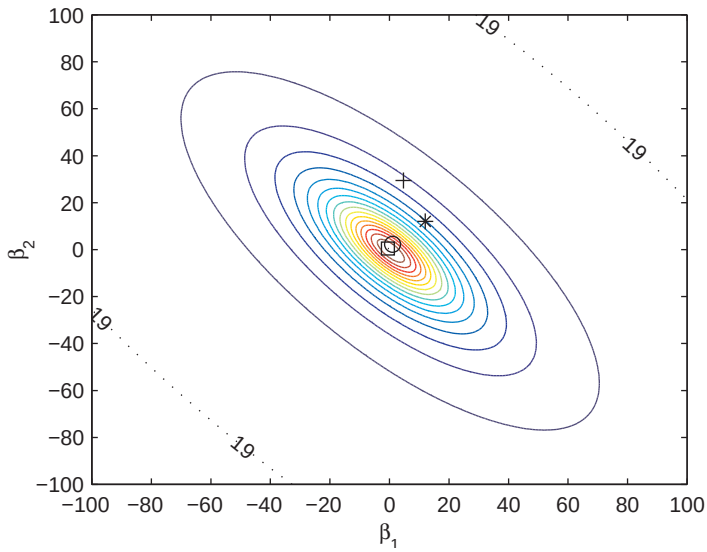


Fig. 2. Contour plot of $\mathcal{Q}(\boldsymbol{\beta})$: partially unidentified case. The asterisk denotes $\boldsymbol{\beta}_0$, the plus sign $\widehat{\boldsymbol{\beta}}$, the circle $(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{y}$ and the square $\boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$. The black dotted contour represents the level curve at $\mathcal{F}_{(0.95)}\{2,2\} = 19.0$.

Theorem 5 *Let* $\widetilde{y}_{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{Y}\boldsymbol{\beta}$ *and* $\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}} = \mathbf{X}\boldsymbol{\Pi}_1 + \mathbf{Z}\boldsymbol{\Pi}_2 + \widetilde{\mathbf{V}}$, *where*

$$\widetilde{\mathbf{V}} = \mathbf{V} - \mathbf{u}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'/\sigma^2_{\mathrm{u},\boldsymbol{\beta}}\,,$$

and set

$$\widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}} = \boldsymbol{\Omega} - \frac{(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'}{\sigma^2_{\mathrm{u},\boldsymbol{\beta}}}\,.$$

Then the matrix $[\widetilde{\mathbf{y}}_{\boldsymbol{\beta}}\ \widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}]'\mathbf{P}[\widetilde{\mathbf{y}}_{\boldsymbol{\beta}}\ \widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}]$ *has a non-central Wishart distribution with* ν *degrees of freedom, covariance matrix*

$$\widetilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma^2_{\mathrm{u},\boldsymbol{\beta}} & \mathbf{0}' \\ \mathbf{0} & \widetilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}} \end{bmatrix}$$

and non-centrality parameter

$$\widetilde{\boldsymbol{\Lambda}} = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Delta} \end{bmatrix}\,.$$

Furthermore, as $\nu^{-1}\|\boldsymbol{\Delta}\| \to 0$,

$$\widetilde{\boldsymbol{\beta}} = \left(\widetilde{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}\right)^{-1}\widetilde{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\widetilde{y}_{\boldsymbol{\beta}} \underset{a}{\sim} \mathbf{t}_n(\nu - n + 1, \mathbf{0}, \widetilde{\mathbf{D}}_{\beta})$$

and

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

where $\widetilde{\mathbf{D}}_\beta = (\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}} + \nu^{-1}\boldsymbol{\Delta})/\sigma^2_{\mathrm{u},\boldsymbol{\beta}}$.

Note that a consequence of Theorem 5 is that the mean vector required to construct our previous critical regions and confidence intervals, namely $\boldsymbol{\mu}_\beta$, has been mapped into zero. The dispersion parameter, however, $\widetilde{\mathbf{D}}_\beta$ in this transformed space, is still unknown, as is the transformation itself.

As $\nu^{-1}\|\boldsymbol{\Delta}\| \to 0$ we know from Theorem 1 of Poskitt and Skeels (2005) that the marginal distribution of $\widetilde{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}$ can be approximated by $W_n(\nu, \tilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}} + \nu^{-1}\boldsymbol{\Delta})$. Thus, for given $\boldsymbol{\beta}$, a natural estimator of $(\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}} + \nu^{-1}\boldsymbol{\Delta})$ would be $\nu^{-1}\widetilde{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}$ if $\boldsymbol{\Sigma}$ were known. Of course, as $\boldsymbol{\Sigma}$ is unknown we still cannot construct $\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}$. However,

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{(N-K)}[\mathbf{y},\mathbf{Y}]'(\mathbf{R}_X - \mathbf{R}_X\mathbf{Z}(\mathbf{Z}'\mathbf{R}_X\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}_X)[\mathbf{y},\mathbf{Y}] = \begin{bmatrix} \hat{\omega}^2 & \hat{\boldsymbol{\omega}}' \\ \hat{\boldsymbol{\omega}} & \widehat{\boldsymbol{\Omega}} \end{bmatrix},$$

provides a consistent estimator of $\boldsymbol{\Sigma}$, whatever the values of $\boldsymbol{\Pi}_2$ and $\boldsymbol{\Delta}$, and $\widehat{\boldsymbol{\Sigma}}$ can clearly be used to provide "plug in" values for the nuisance parameters. Replacing the unknown elements of

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} 1 & \mathbf{0}' \\ -\boldsymbol{\beta} & \mathbf{I}_n \end{bmatrix} \tag{12}$$

by

$$\widehat{\boldsymbol{\Phi}} = \begin{bmatrix} 1 & -\boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} \widehat{\boldsymbol{\Sigma}} \begin{bmatrix} 1 & \mathbf{0}' \\ -\boldsymbol{\beta} & \mathbf{I}_n \end{bmatrix},$$

gives us an empirical version of $\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}$, namely

$$\widehat{\mathbf{Y}}_{\boldsymbol{\beta}} = \mathbf{Y} - \tilde{y}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\omega}} - \widehat{\boldsymbol{\Omega}}\boldsymbol{\beta})/\hat{\sigma}^2_{\mathrm{u},\boldsymbol{\beta}}\,, \tag{13}$$

where $\hat{\sigma}^2_{\mathrm{u},\boldsymbol{\beta}} = [1, -\boldsymbol{\beta}']\widehat{\boldsymbol{\Sigma}}[1, -\boldsymbol{\beta}']'$. Substituting (13) for $\widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}$, and replacing $\tilde{\boldsymbol{\Omega}}_{\boldsymbol{\beta}} + \nu^{-1}\boldsymbol{\Delta}$ by $\nu^{-1}\widehat{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\widehat{\mathbf{Y}}_{\boldsymbol{\beta}}$ in the quadratic form $\tilde{\boldsymbol{\beta}}'\widetilde{\mathbf{D}}_\beta\tilde{\boldsymbol{\beta}}$ yields

$$\frac{1}{\nu}\frac{\tilde{y}'_{\boldsymbol{\beta}}\mathbf{P}\widehat{\mathbf{Y}}_{\boldsymbol{\beta}}\left(\widehat{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\widehat{\mathbf{Y}}_{\boldsymbol{\beta}}\right)^{-1}\widehat{\mathbf{Y}}'_{\boldsymbol{\beta}}\mathbf{P}\tilde{\mathbf{y}}_{\boldsymbol{\beta}}}{\hat{\sigma}^2_{\mathrm{u},\boldsymbol{\beta}}} = \frac{1}{\nu}K(\boldsymbol{\beta})\,,$$

where $K(\boldsymbol{\beta})$ denotes the K statistic of Kleibergen (2002).

Kleibergen (2002) shows that $K(\boldsymbol{\beta})$ 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 a consequence of the small concentration asymptotics that we have adopted here is that we have an alternative finite sample approximation to the sampling distribution of $K(\boldsymbol{\beta})$ based on the F distribution which, as $\nu \to \infty$, yields the same limiting chi-squared distribution as that originally obtained by Kleibergen (2002). The benefit of using our small concentration asymptotic approach is that it provides insights into the power properties of $K(\boldsymbol{\beta})$, when identification is weak, that are not available with existing results. This latter feature is well illustrated by the following example taken from Kleibergen (2002).

Consider a simple model in which

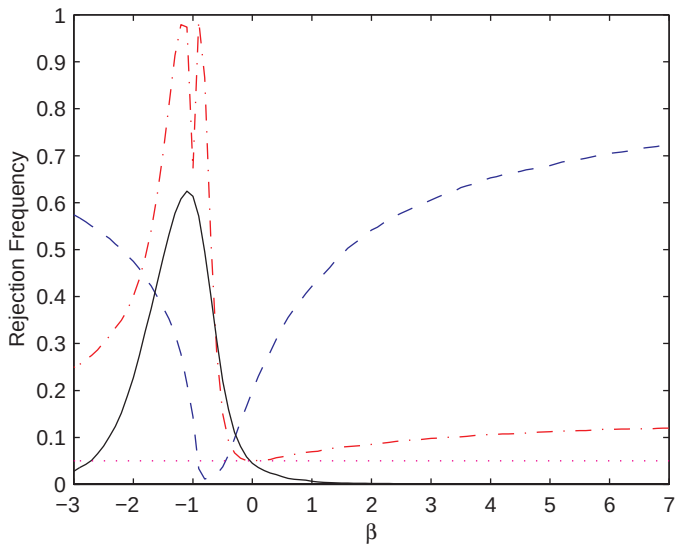
$$\begin{aligned} \mathbf{y} &= \mathbf{Y}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{Y} &= \mathbf{Z}\boldsymbol{\Pi}_2 + \mathbf{V} \end{aligned} \tag{14}$$

with $n = 1$, $N = 100$, $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_\nu \otimes \mathbf{I}_N)$, where $\nu = 5$, $\boldsymbol{\Pi}_2 = (0.1, 0, 0, 0, 0)'$, and $[\mathbf{u}\ \mathbf{V}] \sim N(\mathbf{0}, \boldsymbol{\Phi} \otimes \mathbf{I}_N)$, where

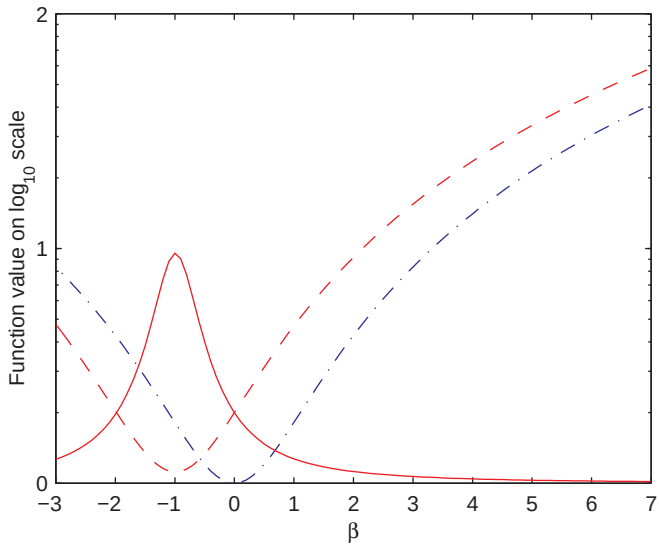
$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}$$

is the matrix defined in (12). The value of $\boldsymbol{\Pi}_2$ implies that four superfluous instruments are being used and the remaining instrument is weak. The instruments are held fixed across the replications and the implied value of $\boldsymbol{\Delta}$ is 0.9063. The correlation of 0.99 in the distribution of $[\mathbf{u}\ \mathbf{V}]$ implies that y and Y are strongly endogenous.

Figures 3a and 4a plot simulated power curves for $\mathcal{Q}(\boldsymbol{\beta})$, $K(\boldsymbol{\beta})$, and the *LIML* statistic of Bekker (1994) (denoted *BLIML* hereafter) which also has a $\chi^2(n)$ asymptotic distribution. These figures were generated on the basis of $10,000$ replications of (14). For $\mathcal{Q}(\boldsymbol{\beta})$ critical regions were constructed using the small concentration F approximation, whereas for $K(\boldsymbol{\beta})$ and *BLIML* they were determined on the basis of their asymptotic $\chi^2(n)$ approximations. In Figure 3a the null hypothesis is $\mathcal{H}_0 : \boldsymbol{\beta} = 0$ and in Figure 4a it is $\mathcal{H}_0 : \boldsymbol{\beta} = 5$. In both



(a) Simulated Power Curves: $\mathcal{Q}(\boldsymbol{\beta})$ (solid), $K(\boldsymbol{\beta})$ (dash-dot), *BLIML* (dash), 5% line (dotted)



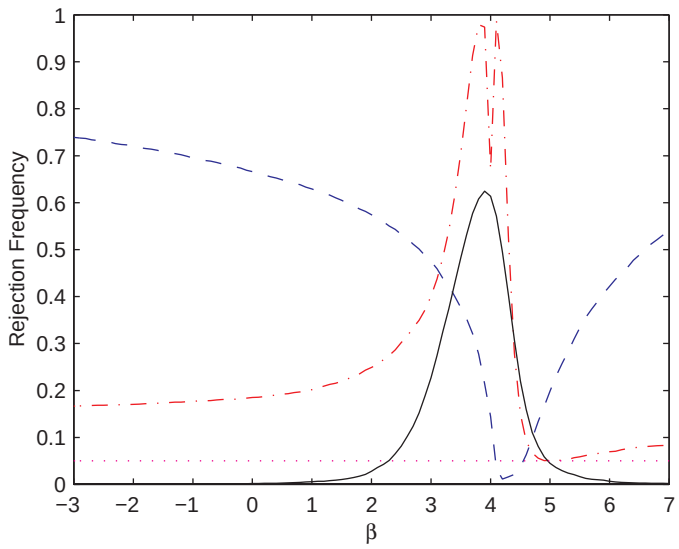
(b) Values $1 + \kappa_0$ (dash-dot), $1 + q_0$ (dashed) and $1 + 1/q_0$ (solid)

Fig. 3. Testing $\mathcal{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ against $\mathcal{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$

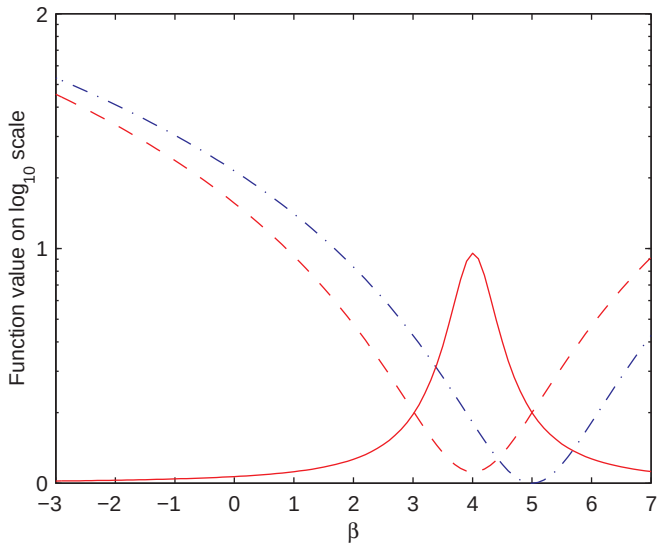
cases we are testing $\mathcal{H}_0$ against the two-sided alternative at the 5% significance level.

The key features of these figures are:

(1) Although $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ exhibit the correct nominal size, they both have very unusual looking power curves, with greatest power close to the null hypotheses but poor power elsewhere. In particular, the power functions are neither monotonically increasing functions of $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ nor are they symmetric about the null value.



(a) Simulated Power Curves: $\mathcal{Q}(\boldsymbol{\beta})$ (solid), $K(\boldsymbol{\beta})$ (dash-dot), *BLIML* (dash), 5% line (dotted)



(b) Values $1 + \kappa_0$ (dash-dot), $1 + q_0$ (dashed) and $1 + 1/q_0$ (solid)

Fig. 4. Testing $\mathcal{H}_0 : \boldsymbol{\beta} = 5$ against $\mathcal{H}_1 : \boldsymbol{\beta} \neq 5$

(2) Although $K(\boldsymbol{\beta})$ dominates $\mathcal{Q}(\boldsymbol{\beta})$ in terms of power, their power functions have profiles that are not too dissimilar. In particular, they both have greatest power at approximately $\boldsymbol{\beta}_0 - 1$.

(3) *BLIML* has very different properties to the other two. It has a much more conventional looking power curve but, as previously observed by Kleibergen (2002), at the expense of a serious size distortion.[5] Its power

[5] It is worth noting that the chi-squared approximation for $K(\boldsymbol{\beta})$ is a large N approximation whereas that for *BLIML* is a large ν approximation. One might imagine that for this particular experimental design, taken from Kleibergen (2002),

curve is asymmetric and appears to have a minimum at the point where the other tests have their maximum power.

(4) Both $\mathcal{Q}(\boldsymbol{\beta})$ and *BLIML* are biased tests, that is, there exist values of $\boldsymbol{\beta}$ such that the probability of a Type II error exceeds $1-\alpha$.

An important difference between $\mathcal{Q}(\boldsymbol{\beta})$, on the one hand, and $K(\boldsymbol{\beta})$ and *BLIML*, on the other, is that the latter pair are asymptotically pivotal statistics whereas the former is not because it depends on other, usually unknown, nuisance parameters. Nevertheless, the behaviour of $\mathcal{Q}(\boldsymbol{\beta})$ is of interest because (i) it displays similar patterns to those of $K(\boldsymbol{\beta})$ at values of $\boldsymbol{\beta}$ close to $\boldsymbol{\beta}_0$ and (ii) because the expression for the small concentration power of $\mathcal{Q}(\boldsymbol{\beta})$, which is presented in Lemma 4, obviously explains the behaviour of $\mathcal{Q}(\boldsymbol{\beta})$ observed here. In particular, we recall the dependence of the power function on the values of κ_0 and q_0. In the light of this we present in Figures 3b and 4b graphs showing how κ_0 and q_0 vary with $\boldsymbol{\beta}$.

From our previous arguments we know that the non-centrality parameter κ_0 is strictly positive for all $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, and from Lemma 4 it follows that whenever $q_0 \leq 1$ it will reinforce any reduction in the probability of a Type II error brought about by increases in κ_0. When $q_0 > 1$, however, increases in q_0 will lead to a reduction in power, other things being equal. We can therefore anticipate that $\pi_{CR}(\boldsymbol{\beta})$ will vary inversely with q_0 and in Figures 3b and 4b we have therefore also shown the variation in $1/q_0$ as a function of $\boldsymbol{\beta}$.

From these graphs it is clear that it is the variations in q_0 that are critical in determining the power characteristics of $\mathcal{Q}(\boldsymbol{\beta})$ rather than the variations in κ_0. Indeed, there is almost a one–to–one correspondence between the graphs of $1/q_0$ and the observed power function of $\mathcal{Q}(\boldsymbol{\beta})$. From Lemma 4 we can see that the impact of q_0 on ϱ_0 has greater influence on the value of $\Psi\{\varrho_0; n, \nu-n+1, \kappa_0\}$ than do changes in κ_0, hence we will couch subsequent discussion in terms of q_0 alone. Perhaps not too surprisingly, since we may view Kleibergen's statistic as a feasible version of our own in the transformed space, this correspondence also carries over to $K(\boldsymbol{\beta})$, albeit somewhat less strikingly. The rejection frequencies of *BLIML* behave very differently from those of $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$, but again they appear to be influenced directly by the value of q_0 more strongly than they do by the value of κ_0.

We observed above that the power functions of all three tests considered appear to have turning points at approximately $\boldsymbol{\beta}_0 - 1$, which corresponds to the minimum value of q_0. The point $\boldsymbol{\beta}_0 - 1$ is readily shown to be an artefact of the parameter choices of (14), although it demonstrates an interesting property of $\mathcal{Q}(\boldsymbol{\beta})$, and seemingly also of $K(\boldsymbol{\beta})$ and *BLIML*. It is a useful exercise

$K(\boldsymbol{\beta})$ is deeper into its approximating sequence with $N = 100$ than is *BLIML* with $\nu = 5$, and so the chi-squared approximation might work better for the former than it does the latter.

to consider why q_0 is minimized at about this value because it highlights an important implication of the experimental design.

To begin, as defined in Theorem 3, for given values of $\boldsymbol{\Sigma}$ and $\nu^{-1}\boldsymbol{\Delta}$ the ratio q_0 is a function of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. However, in this experimental design

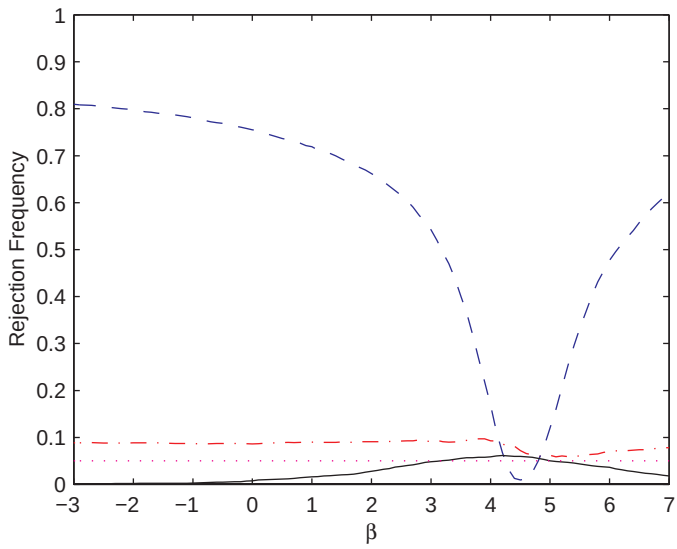
$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \boldsymbol{\beta}' \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} \boldsymbol{\Phi} \begin{bmatrix} 1 & \mathbf{0}' \\ \boldsymbol{\beta} & \mathbf{I}_n \end{bmatrix}, \quad \text{where} \quad \boldsymbol{\Phi} = \begin{bmatrix} \sigma^2_{\mathrm{u},\boldsymbol{\beta}} & \boldsymbol{\sigma}_{\mathrm{uV}} \\ \boldsymbol{\sigma}_{\mathrm{Vu}} & \boldsymbol{\Omega} \end{bmatrix} \quad \text{(say)},$$

is held fixed. Therefore $\boldsymbol{\Sigma}$ is itself implicitly a function of $\boldsymbol{\beta}$ for given $\boldsymbol{\Phi}$, so that q_0 is actually a function of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$, given $\nu^{-1}\boldsymbol{\Delta}$ and $\boldsymbol{\Phi}$. It is useful to re-express q_0 as a function of $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_0$. For given $\nu^{-1}\boldsymbol{\Delta}$ and $\boldsymbol{\Phi}$, it is a relatively simple matter to then show that q_0 is minimized at the point where $\boldsymbol{\beta} - \boldsymbol{\beta}_0 = -\boldsymbol{\Omega}^{-1}\boldsymbol{\sigma}_{\mathrm{Vu}}$. In the current example, $\boldsymbol{\Omega} = 1$ and $\boldsymbol{\sigma}_{\mathrm{Vu}} = 0.99$ and so, rather than $\boldsymbol{\beta}_0 - 1$, we see that the turning points are occurring at $\boldsymbol{\beta}_0 - 0.99$.

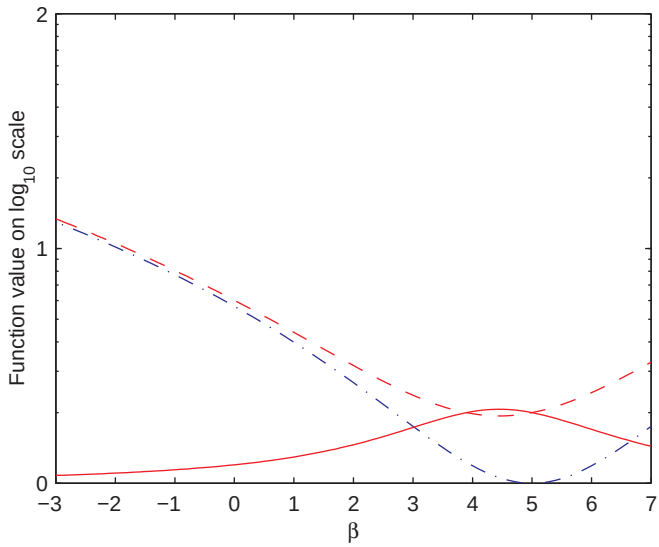
If instead of choosing $\boldsymbol{\sigma}_{\mathrm{Vu}} = 0.99$ we had chosen $\boldsymbol{\sigma}_{\mathrm{Vu}} = 0.55$, for example, then we see from Figure 5 that the turning points occur much closer to $\boldsymbol{\beta}_0 = 5$ than they did in Figure 4. As with Figure 4, $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ attain their maximum power at values of $\boldsymbol{\beta}$ in the closed interval defined by those values of $\boldsymbol{\beta}$ for which $q_0 \leq 1$; namely, the interval corresponding to the intersections of the curves $\ln(1 + q_0)$ and $\ln(1 + 1/q_0)$ in Figures 4b and 5b, respectively.[6] For $\boldsymbol{\sigma}_{\mathrm{Vu}} = 0.55$ this interval is much shorter than it was previously and the overall effect of this is that both $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ have no useful power. The *BLIML* statistic, on the other hand, tends to have useful power for those values of $\boldsymbol{\beta}$ for which $q_0 > 1$, and so this narrowing of the interval corresponds to an increase in power for *BLIML*. But *BLIML* still suffers from a significant size distortion, in contrast to both $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ which display good size properties.

Recall that a basic assumption underlying our theoretical development is that $[\mathbf{y}\ \mathbf{Y}]'\mathbf{P}[\mathbf{y}\ \mathbf{Y}] \sim \mathcal{W}_{n+1}(\nu, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1/2})$ where the parameters of the non–central Wishart distribution are determined from the regression coefficients and residual variance-covariance matrix of the reduced form in (3). Under this scenario the properties of the reduced form determine the underlying statistical features that characterize the observed behaviour of the data generating mechanism. The statistical properties of the structural equation, which represents an externally imposed theoretical economic construct, are derived from the reduced form via the compatability conditions in (5). This suggests that the experimental design of Kleibergen (2002), in which $\boldsymbol{\Phi}$ and $\nu^{-1}\boldsymbol{\Delta}$ are fixed and $\boldsymbol{\beta}$, and implicitly $\boldsymbol{\Sigma}$, are allowed to vary, is not in direct accord with the conceptual framework underlying such an analysis. Let us consider therefore

[6] Observe that these intersections occur when $q_0 = 1$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ will always be one of these points.



(a) Simulated Power Curves: $\mathcal{Q}(\boldsymbol{\beta})$ (solid), $K(\boldsymbol{\beta})$ (dash-dot), *BLIML* (dash), 5% line (dotted)



(b) Values $1 + \kappa_0$ (dash-dot), $1 + q_0$ (dashed) and $1 + 1/q_0$ (solid)

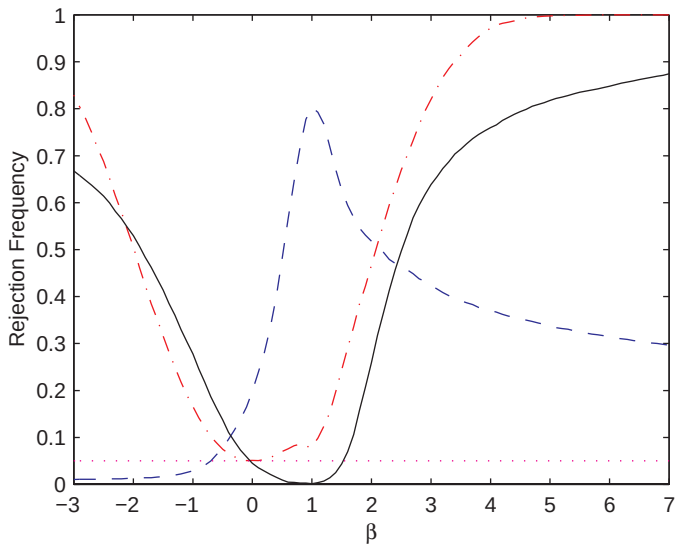
Fig. 5. Testing $\mathcal{H}_0 : \boldsymbol{\beta} = 5$ against $\mathcal{H}_1 : \boldsymbol{\beta} \neq 5$

a slightly different model in which

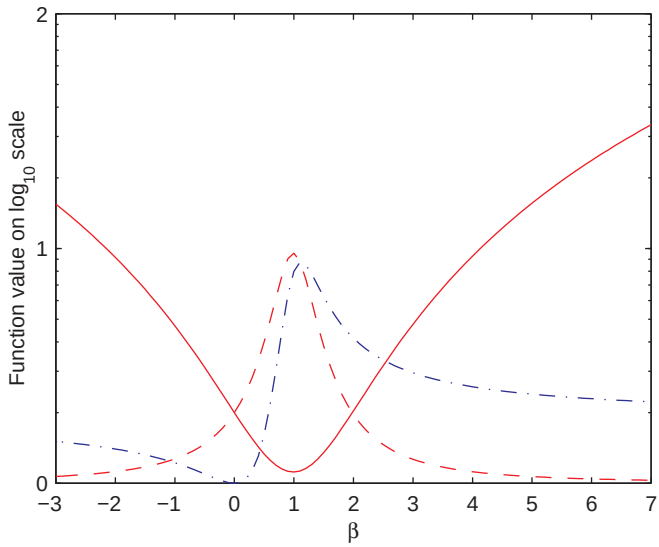
$$\begin{aligned} \mathbf{y} &= \mathbf{Z}\boldsymbol{\Pi}_2\boldsymbol{\beta} + \mathbf{v} \\ \mathbf{Y} &= \mathbf{Z}\boldsymbol{\Pi}_2 + \mathbf{V} \end{aligned}$$

where $[\mathbf{v}\ \mathbf{V}] \sim N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_N)$ and $\boldsymbol{\Sigma}$ is held fixed at

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix},$$



(a) Simulated Power Curves: $\mathcal{Q}(\boldsymbol{\beta})$ (solid), $K(\boldsymbol{\beta})$ (dash-dot), *BLIML* (dash), 5% line (dotted)



(b) Values $1 + \kappa_0$ (dash-dot), $1 + q_0$ (dashed) and $1 + 1/q_0$ (solid)

Fig. 6. Testing $\mathcal{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ against $\mathcal{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$

with n, ν, N, $\mathbf{Z}$ and $\mathbf{\Pi}_2$ specified as previously.

Figure 6a provides a counterpart to Figure 3a, and presents the power curves obtained using this alternative experimental design when testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta} = 0$ at the 5% significance level. We now find that various properties of $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$, and *BLIML* are reversed. In particular, $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ still exhibit the correct nominal size, with power functions that are neither monotonically increasing functions of $\boldsymbol{\beta} - \boldsymbol{\beta}_0$ nor symmetric about the null value, but their power functions have profiles that are more conventional in appearance. The *BLIML* statistic once again has very different properties.

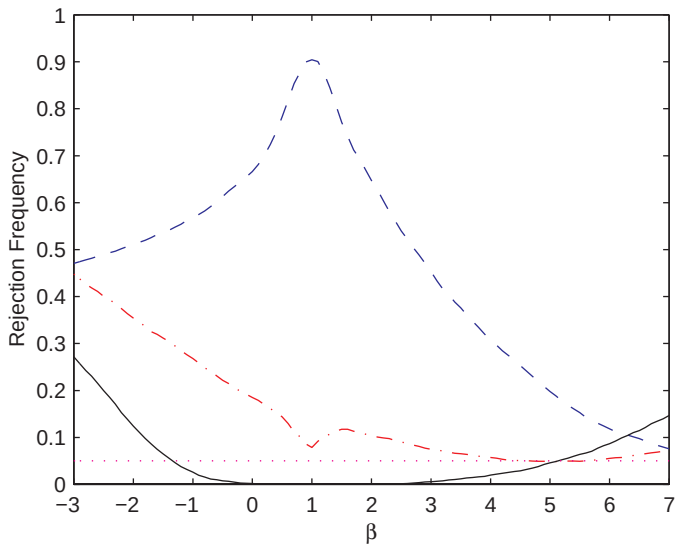
It still has serious size distortion, and now this is accompanied by a very unusual looking power curve with greatest power close to the null hypotheses but poor power elsewhere. Its power curve is asymmetric and appears to have a maximum at the point where $\mathcal{Q}(\boldsymbol{\beta})$ has its minimum power. Both $\mathcal{Q}(\boldsymbol{\beta})$ and *BLIML* are biased tests.

Once again it is clear that variations in q_0 are more critical in determining the power characteristics of the tests than are variations in κ_0. Under the current scenario we find that q_0 is maximized at the point where $\boldsymbol{\beta} = \boldsymbol{\Omega}^{-1}\boldsymbol{\omega}$ and in the current example, $\boldsymbol{\Omega} = 1$ and $\boldsymbol{\omega} = 0.99$, and so the turning points are occurring at $\boldsymbol{\beta} = 0.99$. As previously, $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ exhibit useful power for values of $\boldsymbol{\beta}$ for which $q_0 \leq 1$, whereas *BLIML* only has any useful power for those values of $\boldsymbol{\beta}$ for which $q_0 > 1$. In Figure 6b the values of $\boldsymbol{\beta}$ for which $q_0 > 1$ now correspond to the open interval delineated by the intersections of the curves $\ln(1+q_0)$ and $\ln(1+1/q_0)$, rather than its complement, as in Figure 3b. Hence the reversal of the observed power characteristics.

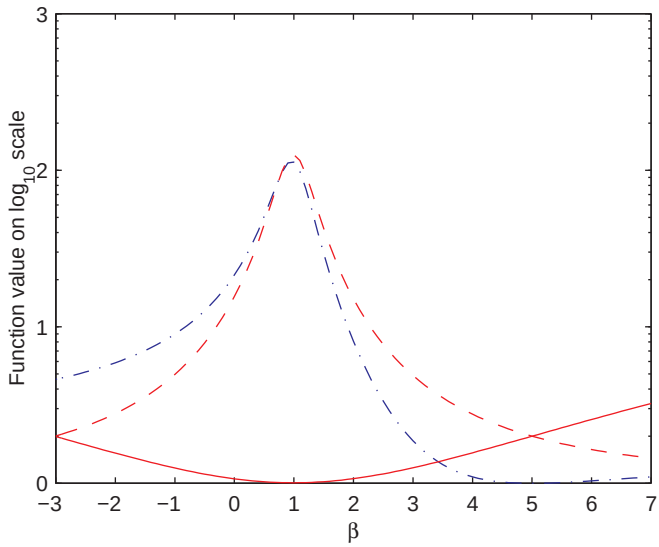
Finally, Figure 7 repeats the experiment underlying Figure 6, except now we are testing the null hypothesis $\mathcal{H}_0 : \boldsymbol{\beta} = 5$ at the 5% significance level. The first observation is that now the power curves are not simply translated along the horizontal axis as they were previously, contrast this with Figures 3 and 4, their shape is contingent on the value of $\boldsymbol{\beta}_0$. Second, although $\mathcal{Q}(\boldsymbol{\beta})$ and $K(\boldsymbol{\beta})$ exhibit U-shaped power curves they do not have much power and, indeed, $\mathcal{Q}(\boldsymbol{\beta})$ is biased over a reasonably long interval. Third, the *BLIML* statistic is still size distorted, with an unusual looking power curve with greatest power at $\boldsymbol{\beta} = 0.99$, the point where the maximum of q_0 occurs. Once more it is readily apparent that the observed power characteristics of the tests can be explained by reference to our analytical results 'and, of course, simulation evidence never can be viewed as a substitute for an analytical theory.' (Dufour and Taamouti, 2005, p. 1352)

6 Conclusion

This paper has been concerned with the characterization of confidence intervals and hypothesis tests for the coefficients on the endogenous regressors ($\boldsymbol{\beta}$) in a single equation in a simultaneous equations model. Our approach has exploited a small-concentration asymptotic approximation to the sampling distribution of the 2SLS estimator. This has enabled us to provide both the null distribution and the power function for $\mathcal{Q}(\boldsymbol{\beta})$, a function of the 2SLS estimator. On the basis of these results we are able to investigate the impact of various parameters on the behaviour of $\mathcal{Q}(\boldsymbol{\beta})$. We construct confidence regions by inverting $\mathcal{Q}(\boldsymbol{\beta})$ and demonstrate that these regions possess certain desirable properties.



(a) Simulated Power Curves: $\mathcal{Q}(\boldsymbol{\beta})$ (solid), $K(\boldsymbol{\beta})$ (dash-dot), *BLIML* (dash), 5% line (dotted)



(b) Values $1 + \kappa_0$ (dash-dot), $1 + q_0$ (dashed) and $1 + 1/q_0$ (solid)

Fig. 7. Testing $\mathcal{H}_0 : \boldsymbol{\beta} = 5$ against $\mathcal{H}_1 : \boldsymbol{\beta} \neq 5$

Unfortunately $\mathcal{Q}(\boldsymbol{\beta})$ is not a feasible statistic, being a function of unknown parameters whose estimation is problematic. Nevertheless, it is shown that replacement of these parameters by natural plug-in statistics yields, after suitable transformation, a feasible statistic whose empirical behaviour is similar to that of $\mathcal{Q}(\boldsymbol{\beta})$ in important ways. It transpires that this statistic is equivalent to the $K(\boldsymbol{\beta})$ statistic of Kleibergen (2002), and so our developments provide analytical insight into the sampling behaviour of $K(\boldsymbol{\beta})$ which heretofore has been missing from the literature. It is also the case that these observations go some way towards explaining the behaviour of the *LIML*-based statistic of Bekker (1994), although developing a stronger theoretical basis for these

links remains the subject of ongoing research. Finally, our results indicate that any attempt to rank these statistics is likely to be fraught with difficulties, particularly as such rankings will themselves be contingent on the conceptual and analytical framework adopted.

References

Anderson, T. W., Sawa, T., 1973. Distributions of estimates of coefficients of a single equation in a simultaneous system and their asymptotic expansions. Econometrica 41 (4), 683–714.

Anderson, T. W., Sawa, T., 1979. Evaluation of the distribution function of the two-stage least squares estimate. Econometrica 47 (1), 163–182.

Bekker, P., Kleibergen, F., 2003. Finite-sample instrumental variables inference using an asymptotically pivotal statistic. Econometric Theory 19 (5), 744–753.

Bekker, P. A., 1994. Alternative approximations to the distributions of instrumental variable estimators. Econometrica 62 (3), 657–681.

Chao, J. C., Swanson, N. R., 2005a. Asymptotic normality of single-equation estimators for the case with a large number of weak instruments. In: Corbae, P. D., Durlauf, S. N., Hansen, B. E. (Eds.), Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, New York, Ch. 4, pp. 82–124.

Chao, J. C., Swanson, N. R., 2005b. Consistent estimation with a large number of weak instruments. Econometrica 73 (5), 1673–1692.

Choi, I., Phillips, P. C. B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified models. Journal of Econometrics 51 (1–2), 113–150.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65 (6), 1365–1387.

Dufour, J.-M., Taamouti, M., 2005. Prohjection-based statistical inference in linear structral models with possibly weak instruments. Econometrica 73 (4), 1351–1365.

Gradshteyn, I. S., Ryzhik, I. M., 1980. Table of Integrals, Series, and Products, corrected and enlarged Edition. Academic Press, Inc., San Diego.

Hahn, J., Hausman, J., 2003. Weak instruments: Diagnosis and cures in empirical econometrics. The American Economic Review 93 (2), 118–125.

Herz, C. S., 1955. Bessel functions of matrix argument. Annals of Mathematics 61 (3), 474–523.

Kleibergen, F., 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70 (5), 1781–1803.

Nelson, C. R., Startz, R., 1990. Some further results on the exact small sample properties of the instrumental variable estimator. Econometrica 58 (4), 967–976.

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., May 1980. The exact distribution of instrumental variable estimators in an equation containing n+1 endogenous variables. Econometrica 48 (4), 861–878.

Phillips, P. C. B., 1989. Partially identified econometric models. Econometric Theory 5 (2), 181–240.

Poskitt, D. S., Skeels, C. L., 2005. Approximating the distribution of the instrumental variables estimator when the concentration parameter is small. Journal of Econometrics , forthcoming.

Rothenberg, T. J., 1984. Approximating the distributions of econometric estimators and test statistics. In: Griliches, Z., Intriligator, M. D. (Eds.), Handbook of Econometrics. Vol. 2. North Holland, Amsterdam, Ch. 15, pp. 881–935.

Sargan, J. D., Mikhail, W. M., 1971. A general approximation to the distribution of instrumental variables estimates. Econometrica 39 (1), 131–169.

Slater, L. J., 1966. Generalized Hypergeometric Functions. Cambridge University Press, Cambridge.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.

Stock, J. H., Wright, J., Yogo, M., October 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business & Economic Statistics 20 (4), 518–529.

Stock, J. H., Yogo, M., 2005. Asymptotic distributions of instrumental variables statistics with many weak instruments. In: Stock, J. H., Andrews, D. W. K. (Eds.), Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, UK, Ch. 6, pp. 109–120.

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

Wang, J., Zivot, E., 1998. Inference on structural parameters in instrumental variables regression with weak instruments. Econometrica 66 (6), 1389–1404.

Zivot, E., Startz, R., Nelson, C. R., 1998. Valid confidence intervals and inference in the presence of weak instruments. International Economic Review 39 (4), 1119–1144.

Zivot, E., Startz, R., Nelson, C. R., 2005. Improved inference in weakly identified instrumental variables regression. In: Corbae, P. D., Durlauf, S. N., Hansen, B. E. (Eds.), Econometric Theory and Practice: Frontiers of Analysis and Applied Research. Cambridge University Press, New York, Ch. 5, pp. 125–163.

A On Quadratic Forms in Non-Central Multivariate t Random Variables

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

$$\psi(r : p, \nu, \kappa) = \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}\right)} r^{(p-2)/2}[\nu + \kappa + r]^{-(\nu+p)/2} \times {}_2F_1\left(\frac{\nu+p}{4}, \frac{\nu+p+2}{4}; \frac{p}{2}; \frac{4\kappa r}{[\nu+\kappa+r]^2}\right) \quad \text{(A.1)}$$

wherein ${}_2F_1\left(k, l; m; x\right)$ *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,* $\kappa = \boldsymbol{\mu}'\boldsymbol{\mu}$. *The corresponding distribution function, denoted* $\Psi\{\varrho; p, \nu, \kappa\} = P(r \leq \varrho)$, *is*

$$\Psi\{\varrho; p, \nu, \kappa\} = \sum_{j=0}^{\infty} w_j \frac{\varrho^{j+p/2}}{j + p/2} {}_2F_1\left(2j + \frac{\nu+p}{2}, j + \frac{p}{2}; j + \frac{p+2}{2}; -\frac{\varrho}{\nu+\kappa}\right) \quad \text{(A.2)}$$

where

$$w_j = \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}\right)} \frac{\left(\frac{\nu+p}{4}\right)_j \left(\frac{\nu+p+2}{4}\right)_j}{j!\left(\frac{p}{2}\right)_j} \frac{(4\kappa)^j}{(\nu+\kappa)^{2j+(\nu+p)/2}}.$$

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

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

Using the result

$$s^{-\alpha}\Gamma(\alpha) = \int_0^{\infty} e^{-sx} x^{\alpha-1} dx \quad \text{(A.3)}$$

we obtain

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

where $c_1 = \left[(\nu\pi)^{p/2}\Gamma\left(\frac{\nu}{2}\right)\right]^{-1}$. Next transform from $\mathbf{t}$ to $\mathbf{v}r^{1/2}$ where $r = \mathbf{t}'\mathbf{t} > 0$,

$\mathbf{v} = \mathbf{t}(\mathbf{t}'\mathbf{t})^{-1/2}$, so that $\mathbf{v}'\mathbf{v} = 1$, and $d\mathbf{t} = 2^{-1}r^{(p-2)/2}d\mathbf{v}dr$. Hence

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

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, writing $\kappa = \boldsymbol{\mu}'\boldsymbol{\mu}$, we obtain

$$f(r) = \frac{r^{(p-2)/2}}{\nu^{p/2}\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}\right)}\int_0^\infty \exp\{-[1+\nu^{-1}(r+\kappa)]x\} \\ x^{(\nu+p-2)/2}\,{}_0F_1\left(\frac{p}{2};\nu^{-2}r\kappa x^2\right)dx\,. \quad \text{(A.4)}$$

Expanding the hypergeometric function in (A.4) and using (A.3) to integrate term by term we now find that

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

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

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

Using Hölder's inequality — and noting that $\nu > 0$, $\kappa \geq 0$ and $r \geq 0$ — we

see that

$$0 \leq \frac{4\kappa r}{[\nu+\kappa+r]^2} < \frac{4\kappa r}{[\kappa+r]^2} \leq 1\,,$$

and so the hypergeometric function is convergent (Slater, 1966, Section 1.1.1). Substituting this final expression for $\mathcal{I}$ back into (A.4) yields the density as given in (A.1).

Expanding the hypergeometric function in (A.1) and integrating term by term yields

$$\Psi\{\varrho; p, \nu, \kappa\} = \sum_{j=0}^{\infty} w_j \int_0^{\varrho} \frac{r^{j+(p-2)/2}}{[1+r/(\nu+\kappa)]^{2j+(\nu+p)/2}}\, dr\,, \tag{A.5}$$

where

$$w_j = \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}\right)} \frac{\left(\frac{\nu+p}{4}\right)_j \left(\frac{\nu+p+2}{4}\right)_j}{j!\left(\frac{p}{2}\right)_j} \frac{(4\kappa)^j}{(\nu+\kappa)^{2j+(\nu+p)/2}}\,.$$

Resolving the integral in (A.5) using Gradshteyn and Ryzhik (1980, Equation 3.194.1) gives (A.2). □

We shall use $\Psi\{p, \nu, \kappa\}$ to denote a random variable with probability density and distribution functions as given in (A.1) and (A.2), respectively. Note that if $\kappa = 0$ then (A.1) collapses to

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

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

B Proofs

In this Appendix we gather proofs of results developed in Sections 3–5.

Proof of Corollary 2 Let $\widehat{\mathbf{q}} = \mathbf{D}_\beta^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)$. Substituting into (9) using (8), noting that the Jacobian of the mapping from $\widehat{\boldsymbol{\beta}}$ to $\widehat{\mathbf{q}}$ is $|\mathbf{D}_\beta|^{-1/2}$, gives

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

for the asymptotic distribution of $\widehat{\mathbf{q}}$. Transforming from rectangular to polar co-ordinates in (B.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 First observe that $\mathbf{D}_{\beta_0} = \mathbf{D}_\beta/q_0$. It follows that

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

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 Lemma 4 Using notation from the proof of Theorem 3

$$\begin{aligned} P\left((\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)'\mathbf{D}_\beta\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\mu}_\beta)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} \geq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\}}{(\nu-n+1)} \right) \\ = P\left(\frac{\widehat{\mathbf{t}}_0'\widehat{\mathbf{t}}_0}{(\nu-n+1)q_0} \geq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\}}{(\nu-n+1)} \right) \\ = 1 - P\left(\widehat{\mathbf{t}}_0'\widehat{\mathbf{t}}_0 < nq_0\mathcal{F}_{(1-\alpha)}\{n, \nu-n+1\} \right) . \end{aligned}$$

The statement of the lemma follows directly. □

Proof of Theorem 5 Let

$$\mathbf{T} = \begin{bmatrix} 1 & \mathbf{0}' \\ -\boldsymbol{\beta} & \mathbf{I}_n \end{bmatrix} \begin{bmatrix} 1 & -(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'/\sigma^2_{\mathrm{u},\boldsymbol{\beta}} \\ \mathbf{0} & \mathbf{I}_n \end{bmatrix} = \begin{bmatrix} 1 & -(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'/\sigma^2_{\mathrm{u},\boldsymbol{\beta}} \\ -\boldsymbol{\beta} & \mathbf{I}_n + \boldsymbol{\beta}(\boldsymbol{\omega} - \boldsymbol{\Omega}\boldsymbol{\beta})'/\sigma^2_{\mathrm{u},\boldsymbol{\beta}} \end{bmatrix}$$

By assumption $[\mathbf{y}\ \mathbf{Y}]'\mathbf{P}[\mathbf{y}\ \mathbf{Y}] \sim \mathcal{W}_{n+1}(\nu, \boldsymbol{\Sigma}, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1/2})$ and, by construction, $[\widetilde{\mathbf{y}}_{\boldsymbol{\beta}}\ \widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}] = [\mathbf{y}\ \mathbf{Y}]\mathbf{T}$. Some tedious algebra reveals that $\mathbf{T}'\boldsymbol{\Sigma}\mathbf{T} = \widetilde{\boldsymbol{\Sigma}}$ and $\mathbf{T}'\boldsymbol{\Lambda}\mathbf{T} = \widetilde{\boldsymbol{\Lambda}}$. Hence $[\widetilde{\mathbf{y}}_{\boldsymbol{\beta}}\ \widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}]'\mathbf{P}[\widetilde{\mathbf{y}}_{\boldsymbol{\beta}}\ \widetilde{\mathbf{Y}}_{\boldsymbol{\beta}}] \sim \mathcal{W}_{n+1}(\nu, \widetilde{\boldsymbol{\Sigma}}, \widetilde{\boldsymbol{\Sigma}}^{-1/2}\widetilde{\boldsymbol{\Lambda}}\widetilde{\boldsymbol{\Sigma}}^{-1/2})$, which establishes the first part of the theorem. Noting from (1) that the implied structural equation for the transformed variables is $\widetilde{y}_{\boldsymbol{\beta}} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{u}$, the remaining results of the theorem follow immediately on application of Lemma 1 and Corollary 2, respectively. □