Small Concentration Asymptotics and Instrumental Variables Inference

D.S. Poskitt\textsuperscript{a} C.L. Skeels\textsuperscript{b,*}

\textsuperscript{a}Department of Econometrics and Business Statistics, Monash University, Vic 3800, Australia
\textsuperscript{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, an approximation which is appropriate when the concentration parameter associated with the reduced form model is small. Here we extend those earlier results and investigate hypothesis testing procedures, under both null and alternative hypotheses, and confidence region construction. We also explore the relationship between our work and Kleibergen’s $K(\beta)$ statistic (Kleibergen, 2002) and demonstrate that our results can be used to explain the sampling behaviour of $K(\beta)$ in models where identification is weak.

Key words: simultaneous equations model, IV estimator, weak identification, weak instruments, power curves, quadratic forms, test statistics

JEL codes: C10, C12, C13, C30.

\* 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).
1 Introduction

In a recent contribution Poskitt and Skeels (2005) present a new approximation to the exact sampling distribution of the instrumental variables (IV) estimator of the coefficients on the endogenous regressors in a single equation from a linear system of simultaneous equations. More specifically, they examine the properties of the two-stage least squares estimator (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 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 that employ the small-concentration asymptotic approximations of Poskitt and Skeels (2005). Coincidentally our results allow us to explain the sampling behaviour of Kleibergen’s \( K(\beta) \) statistic (Kleibergen, 2002) when identification is weak.

Small-concentration asymptotics index the nesting sequence of problems by an ever diminishing value of the concentration parameter. The resulting approximations presented in Poskitt and Skeels (2005) have very simple functional forms and have been shown to be extremely accurate when, \textit{inter alia}, identification is weak.\(^1\) An interesting feature of the approximations 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.\(^2\) 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, 1

\(^{1}\) The consequences of weak identification for inference are by now well understood. Good surveys of this literature are available in Stock, Wright, and Yogo (2002), Hahn and Hausman (2003) and Andrews and Stock (2005), and we shall make no attempt to reproduce their content here.

\(^{2}\) The failure of standard asymptotics in weakly identified models 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 (2005a,b) 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.
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.

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 establishes the small-concentration asymptotic results upon which subsequent developments are based. Section 4 investigates 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 implementation of our results and, in particular, relate them to $K(\beta)$. We also provide examples which illustrate that our analysis explains the sampling characteristics of $K(\beta)$, and other statistics, in the presence of weak instruments. Section 6 presents a brief conclusion. Finally, Appendix A and Appendix B collect proofs of various results presented in the paper.

2 The Model, Notation and Assumptions

Consider the classical structural equation model

$$y = Y\beta + X\gamma + u$$ (1)

where the endogenous matrix variables $y$ and $Y$ are $N \times 1$ and $N \times n$, respectively, the matrix of exogenous variables $X$ is $N \times k$, and $u$ denotes a $N \times 1$ vector of uncorrelated stochastic disturbances with zero mean and variance $\sigma^2_u$. The vectors of structural coefficients $\beta$ and $\gamma$ are $n \times 1$ and $k \times 1$, respectively.

If we define $[X Z]$ to be the $N \times K$ instrument set, where $Z$ 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 $\beta$ using the IV estimator

$$\hat{\beta} = (Y'PY)^{-1}Y'Py,$$ (2)

where $P = P_{[X Z]} - P_X = R_X - R_{[X Z]}$. For any $N \times k$ matrix $X$ of full column rank $P_X$ denotes the idempotent, symmetric matrix $X(X'X)^{-1}X'$ and $R_X = I_N - P_X$. The matrix $P_X$ is of course the $N \times N$ (prediction) operator of rank $k$ that projects on to the space spanned by the columns of $X$ and $R_X$ 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 $[X Z]$ has full column rank, $\rho([X Z]) = K$ almost surely. In this case

$$P = R_X(Z/R_XZ)^{-1}Z/R_X$$
is an $N \times N$ matrix of rank $\nu \geq n$.

The corresponding reduced form model is

$$[y \ Y] = [X \ Z] \begin{bmatrix} \pi_1 \ \Pi_1 \\ \pi_2 \ \Pi_2 \end{bmatrix} + [v \ V]. \quad (3)$$

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

$$\Sigma = \begin{bmatrix} \omega^2 & \omega' \\ \omega & \Omega \end{bmatrix}, \quad (4)$$

$\omega^2$ scalar, where $[v \ V]$ is partitioned conformably with $[y \ Y]$. We assume that $0 < \Sigma < \infty$, meaning that the smallest and largest characteristic roots of $\Sigma$ are positive but bounded, viz. $0 < \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) < \infty$. The components of the reduced form coefficient matrix $\Pi$ — namely $\pi_1$, $\Pi_1$, $\pi_2$ and $\Pi_2$ — 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 $S = [y \ Y]'P[y \ Y]$ has a non-central Wishart distribution with $\nu$ degrees of freedom, covariance matrix $\Sigma$ and non-centrality parameter $\Sigma^{-1/2}\Lambda\Sigma^{-1/2}$, where $\Sigma^{1/2}$ is a symmetric matrix square root of $\Sigma$ and

$$\Lambda = [\pi_2 \ \Pi_2]'Z'R_XZ[\pi_2 \ \Pi_2],$$

which shall be denoted $S \sim \mathcal{W}_{n+1}(\nu, \Sigma, \Sigma^{-1/2}\Lambda\Sigma^{-1/2})$.\(^3\) We will also assume that the usual compatibility conditions hold; namely

$$\pi_1 - \Pi_1\beta = \gamma, \quad \pi_2 = \Pi_2\beta, \quad \sigma_u^2 = \sigma_{u,\beta}^2 \equiv [1, -\beta']\Sigma[1, -\beta']. \quad (5)$$

It follows that

$$\Lambda = [\beta, I_n]'\Pi_2'Z'R_XZ\Pi_2[\beta, I_n] = \begin{bmatrix} \delta^2 & \delta' \\ \delta & \Delta \end{bmatrix},$$

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

\(^3\) Clearly $S$ will be non-central Wishart if $\text{vec}[v \ V] \sim N(0, \Sigma \otimes I_N)$, for then $\text{vec}[y \ Y] \sim N(\text{vec}([X \ Z]\Pi), \Sigma \otimes I_N)$ and the result follows. It will also apply if $\text{vec}[v \ V]$ has a distribution from the elliptically symmetric family. We might also expect $S$ to be approximately non-central Wishart provided that the rows of $[y \ Y]$ satisfy an appropriate mixing condition, in which case the arguments underlying subsequent developments will still apply, with perhaps minor modifications.
3 Small Concentration Asymptotic Results

Exploiting properties of the Wishart distribution, in conjunction with the compatibility conditions (5), Poskitt and Skeels (2005) show that if \( \nu^{-1} \| \Delta \| \) approaches zero then \( \hat{\beta} \) converges in probability to a random vector possessing an \( n \)-variate \( t \)-distribution with \( \nu - n + 1 \) degrees of freedom, location parameter

\[
\mu_\beta = \beta + (\Omega + \nu^{-1} \Delta)^{-1}(\omega - \Omega \beta)
\]  

and dispersion parameter

\[
D_\beta = \frac{\Omega + \nu^{-1} \Delta}{\sigma^2_{\omega,\beta} - (\omega - \Omega \beta)'(\Omega + \nu^{-1} \Delta)^{-1}(\omega - \Omega \beta)};
\]

Here, and in what follows, we use \( \| A \| = \sqrt{\text{tr}\{A^\prime A\}} \) to denote the Euclidean norm of a matrix \( A \). We will also use the tilde symbol \( \sim \) underet with an ‘a’ to denote convergence in probability to a random variable with the stated distribution. Thus we shall write

\[
\hat{\beta} \sim_a t_n(\nu - n + 1, \mu_\beta, D_\beta).
\]

Note that the distribution \( t_n(\nu - n + 1, \mu_\beta, D_\beta) \) has mean vector \( \mu_\beta \) and, for \( \nu > n + 1 \), variance-covariance matrix \( [(\nu - n - 1)D_\beta]^{-1} \), where the notation is designed to highlight the dependence of both the mean vector and covariance matrix on \( \beta \). If one thinks of \( \nu \) as being fixed this result can be viewed as providing a small concentration asymptotic approximation since it is applicable as \( \nu^{-1} \| \Delta \| \to 0 \), as compared to the more conventional asymptotic normal approximation and Edgeworth type expansions which require that the concentration parameter \( \Delta \) be large, see Rothenberg (1984). If one allows for the possibility of \( \nu \) tending to infinity then this result can be used to demonstrate various aspects of many-instrument and many-weak-instrument asymptotics; see Poskitt and Skeels (2005) for further discussion of this point.

With this background we have the following result:

**Theorem 1** Using the definitions of (1) through (7), the quadratic form

\[
Q(\beta_0) = (\hat{\beta} - \mu_\beta)'D_\beta(\hat{\beta} - \mu_\beta)|_{\beta = \beta_0} \sim_a \frac{\Psi\{n, \nu - n + 1, \kappa_0\}}{(\nu - n + 1)q_0}
\]

as \( \nu^{-1} \| \Delta \| \to 0 \), where

\[
\kappa_0 = (\nu - n + 1)(\mu_\beta - \mu_{\beta_0})'D_\beta(\mu_\beta - \mu_{\beta_0})
\]
and
\[
q_0 = \frac{\left[ \sigma_{u,\beta}^2 - (\omega - \Omega\beta)'(\Omega + \nu^{-1}\Delta)^{-1}(\omega - \Omega\beta) \right]_{\beta = \beta_0}}{\sigma_{u,\beta}^2 - (\omega - \Omega\beta)'(\Omega + \nu^{-1}\Delta)^{-1}(\omega - \Omega\beta)},
\]
and \( \Psi\{n, \nu - n + 1, \kappa_0\} \) denotes a random variable with probability distribution as defined in Lemma A.1.

The import of Theorem 1 is that it gives us an approximation to the distribution of the quadratic form when calculated at an arbitrary point \( \beta_0 \) in the parameter space. In the special case where \( \beta_0 \) is equal to the erstwhile true parameter \( \beta, \kappa_0 = 0, q_0 = 1 \) and we have the following result:

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

\[
Q(\beta) = (\hat{\beta} - \mu_\beta)'D_\beta(\hat{\beta} - \mu_\beta);
\]

that is,

\[
Q(\beta) \sim_{\alpha} \frac{n\mathcal{F}\{n, \nu - n + 1\}}{n\nu - n + 1}.
\]

4 Inference

4.1 Hypothesis Testing

From Corollary 2 it follows that an asymptotic size \( \alpha \) critical region for testing the null hypothesis \( H_0 : \beta = \beta_0 \) against the alternative \( H_1 : \beta \neq \beta_0 \) is given by

\[
CR = \left\{ \hat{\beta} : Q(\beta_0) \geq \frac{n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right\}.
\]

where \( \mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\} \) denotes the \((1 - \alpha)100\%\) percentile point of an \( F \) distribution with \( n \) and \( \nu - n + 1 \) degrees of freedom.

4.1.1 Behaviour of CR Under the Null Hypothesis

As \( \Delta \) approaches zero it is obvious that \( \Omega + \nu^{-1}\Delta \) approaches \( \Omega \) and the inverse \( (\Omega + \nu^{-1}\Delta)^{-1} = \Omega^{-1} - \nu^{-1}\Omega^{-1}\Delta\Omega^{-1} + o(\|\Delta\|/\nu) \). Substituting the expressions for \( \mu_\beta \) and \( D_\beta \) into \( Q(\beta) \), whilst making use of the fact that

\[
\sigma_{u,\beta}^2 = (\omega - \Omega\beta)'\Omega^{-1}(\omega - \Omega\beta) + \omega'^{-1}\Omega^{-1}\omega,
\]
it is straightforward, if tedious, to establish that

\[ Q(\beta) = \frac{L_{\beta} - \nu^{-1}Q_{\beta}}{\sigma^2 + \nu^{-1}Q_{\beta}} + o(\|\Delta\|/\nu) \]

where \( \sigma^2 = \omega^2 - \omega'\Omega^{-1}\omega \), \( Q_{\beta} = (\beta - \Omega^{-1}\omega)'\Delta(\beta - \Omega^{-1}\omega) \) and

\[ L_{\beta} = (\hat{\beta} - \Omega^{-1}\omega)'\Omega(\hat{\beta} - \Omega^{-1}\omega) + \nu^{-1}(\hat{\beta} - \beta)'\Delta(\hat{\beta} - \beta). \]

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

\[ q_{(1-\alpha)} = \left[ \frac{(\nu - n + 1)L_{\beta_0} - n\sigma^2 F_{(1-\alpha)}\{n, \nu - n + 1\}}{n F_{(1-\alpha)}\{n, \nu - n + 1\} + (\nu - n + 1)} \right] \nu. \]

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

Such behaviour is not unreasonable. Unless \( \beta = \Omega^{-1}\omega \), so that OLS is unbiased for \( \beta \), we know that values of \( \beta \) that lie in a neighbourhood of \( \Omega^{-1}\omega \) are unlikely to be sensible candidates for the true value, and only values of \( \beta \) that lie in a region of \( \mathbb{R}^n \) that is outside a neighbourhood of \( \Omega^{-1}\omega \) should, perhaps, be considered acceptable. If \( \beta = \Omega^{-1}\omega \), \( Q_{\beta_0} = 0 \) and \( CR \) reduces to a somewhat more familiar F-type rejection region:

\[ \frac{(\hat{\beta} - \beta_0)'(\Omega + \nu^{-1}\Delta)(\hat{\beta} - \beta_0)/n}{\sigma^2/(\nu - n + 1)} \geq F_{(1-\alpha)}\{n, \nu - n + 1\}. \]

4.1.2 Power Properties of CR

We have seen that, for given values of the endogenous variables \([y, Y]\) and the exogenous variables \([X, Z]\), any hypothesized value \( \beta_0 \) that lies in the region in the parameter space given by \( \{\beta : Q_{\beta} \leq q_{(1-\alpha)}\} \) will be rejected by tests based on \( CR \). Although \( \hat{\beta} \) is not itself random \( \{\beta : Q_{\beta} \leq q_{(1-\alpha)}\} \) is, of course, a realization of a random set, the randomness being a function of the distribution of the statistic \( \hat{\beta} \) from which it is derived. Likewise, the probability that \( CR \) leads to a rejection will also be governed by the distributional properties of \( \hat{\beta} \). In particular, we are interested in the impact that the distribution of \( \hat{\beta} \) has
on the expected value of the indicator function \( \phi_{CR} \).\(^4\)

**Theorem 3** Let

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

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

\[
\lim_{\nu^{-1}\|\Delta\| \to 0} \pi_{CR}(\beta) = 1 - \Psi\{q_0; n, \nu - n + 1, \kappa_0\}
\]

where \( q_0 \) and \( \kappa_0 \) are as defined in Theorem 1, \( q_0 = nq_0F_{(1-\alpha)}\{n, \nu - n + 1\} \), and \( \Psi\{\cdot; \cdot, \cdot, \cdot\} \) denotes the distribution function given in (A.2).

It follows from Theorem 3 that CR defines a test procedure with a power function that is increasing in \( \kappa_0 \), but is inversely related to the magnitude of \( q_0 \). This suggests that \( \pi_{CR}(\beta) \) could be subject to countervailing forces as the magnitudes of \( \kappa_0 \geq 0 \) and \( q_0 \geq 0 \) change as functions of \( \beta \).

To gain additional insight into the power of CR let us consider the behaviour of the non-centrality parameter \( \kappa_0 \). Substituting the expression

\[
\mu_{\beta} - \mu_{\beta_0} = \left[I - (\Omega + \nu^{-1}\Delta)^{-1}\Omega\right](\beta - \beta_0)
\]

into \( \kappa_0 \), and using the second-order expansion

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

we find that

\[
\kappa_0 = (\nu - n + 1)\frac{(\beta - \beta_0)'\Delta\Omega^{-1}\Delta(\beta - \beta_0)}{\nu^2d_\beta} + o(\nu^{-2}\|\Delta(\beta - \beta_0)\|^2)
\]

where

\[
d_\beta = \omega^2 - \omega'\Omega^{-1}\omega + \nu^{-1}(\beta - \Omega^{-1}\omega)'\Delta(\beta - \Omega^{-1}\omega).
\]

The denominator \( d_\beta \) converges to \( \sigma^2 \) as \( \nu^{-1}\|\Delta\| \to 0 \) and the size of \( \kappa_0 \) is clearly controlled by the magnitude of \( \nu^{-2}\|\Delta(\beta - \beta_0)\|^2 \). Since \( \nu^{-1}\Delta \) is presumed to be local to zero, it follows that \( \kappa_0 \) can still be small even if \( \|\beta - \beta_0\| \) is itself quite large.

Next, consider the difference \( d_\beta - d_{\beta_0} \) which equals

\[
(\omega - \Omega\beta)'\left(\nu^{-1}\Omega^{-1}\Delta\Omega^{-1} - \nu^{-2}\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\right)(\omega - \Omega\beta) - (\omega - \Omega\beta_0)'\left(\nu^{-1}\Omega^{-1}\Delta\Omega^{-1} - \nu^{-2}\Omega^{-1}\Delta\Omega^{-1}\Delta\Omega^{-1}\right)(\omega - \Omega\beta_0)
\]

\(^4\) Strictly speaking, our notation should indicate that \( \phi_{CR} \) is a function of the data \([y, Y] \) and \([X, Z]\), but this dependence is omitted for simplicity.
plus terms $o(\nu^{-2}\|\Delta\|^2)$. In general it is not possible to sign this difference, implying that $q_0 = d_{\beta_0}/d_{\beta}$ can be greater than, equal to, or less than one. Hence changes in $q_0$ can either amplify or attenuate any changes in $\pi_{CR}(\beta)$ induced by changes in $\kappa_0$.

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 $\beta$ deviates from $\beta_0$ by a considerable margin.

### 4.2 Confidence Region Construction

Thinking of a confidence region as being equivalent to those values of $\beta$ that are consistent with the data indicates that confidence regions with the correct asymptotic coverage probability can be constructed by inverting the critical region $CR$. Following the previous development and applying this idea leads us to the confidence region

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

The set $CI$ determines a $(1 - \alpha)100\%$ confidence region for $\beta$, 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 non-linear manner in which the parameter $\beta$ enters into both the location parameter $\mu_\beta$ and the dispersion parameter $D_\beta$. Thus, although $D_\beta^{1/2}(\hat{\beta} - \mu_\beta)$ has a spherically symmetric distribution, the confidence sets derived from $Q(\beta)$ are not conventional Wald-type regions. Indeed, it follows from our previous arguments that $CI$ will lie in the complement of $\{\beta : Q_\beta \leq q_{(1-\alpha)}\}$ and hence the region $CI$ need not be either convex or connected. Nor need $CI$ be bounded.

To verify the latter point suppose that

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

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

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

where $0 < \epsilon = (\nu - n + 1)/[(\nu - n + 1) + n\mathcal{F}_{(1-\alpha)}\{n, \nu - n + 1\}] < 1$. But if $\beta$ satisfies this latter condition then so too does $\psi\beta$ for any $\psi > 1$. We

---

5 The viewpoint taken here was first espoused by Neyman (1937) and is now commonly adopted.
Figure 1. Contour plot of $Q(\beta)$: weakly identified case. The asterisk denotes $\beta_0$, the plus sign $\hat{\beta}$, the circle $(Y'Y)^{-1}Y'y$ and the square $\Omega^{-1}\omega$. The black dotted contour represents the level curve at $2F_{(0.95)}\{2, 1\} = 399.0$.

Have thus exhibited a subset of $CI$ in which $\|\beta\|$, and therefore the diameter 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 $\beta$ constructed using $Q(\beta)$ would be valid.

The phenomena described in the preceding paragraphs are clearly illustrated in Figure 1, which depicts the surface generated by $Q(\beta)$ in a neighbourhood of the ordinary least squares value $(Y'Y)^{-1}Y'y$ via a contour plot. The figure is based on a hypothetical model in which $n = \nu = 2$, $\beta = 12(1, 1)'$,

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

This gives $\|\Delta\| = 0.1114$. The 95% confidence region $CI$ consists of all those $\beta$’s that lie outside the area enclosed by the contour corresponding to the critical value $2F_{(0.95)}\{2, 1\} = 399$. Those $\beta$’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 $\Omega^{-1}\omega$ lie roughly at the centre of this region, close to where the maximum of $Q(\beta)$ occurs. Because the model is weakly identified the conditional regression of $y$ on $Y$ implicit in the reduced form, namely $\Omega^{-1}\omega$, is close to the ordinary least squares estimate. The latter estimate of $\beta$ is known to be inconsistent however. And so it is natural to reject those values of $\beta$ that lie in a neighbourhood of these two, almost coincident, points.
Figure 2. Contour plot of $Q(\beta)$: partially unidentified case. The asterisk denotes $\beta_0$, the plus sign $\bar{\beta}$, the circle $(\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\mathbf{y}$ and the square $\Omega^{-1}\omega$. The black dotted contour represents the level curve at $F_{(0.95)}\{2,2\} = 19.0$.

Note, in passing, that $q(1-\alpha)$ decreases monotonically with $\alpha$, so that as the level of significance falls the volume of $\{\beta : Q_\beta \geq q(1-\alpha)\}$ increases (ceteris paribus) and the set of values of $\beta$ 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 $\beta$ cannot be determined from the data. In this case the volume of $\{\beta : Q_\beta \leq q(1-\alpha)\}$,

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

will become unbounded as $\|\Delta\| \to 0$ and ultimately all values of $\beta$ will be deemed unacceptable, whatever the value of $\alpha$. This type of behaviour is seen in Figure 2, where the contours of $Q(\beta)$ are presented for the same hypothetical model as before except that $\nu = 3$, implying that a redundant instrument has been added, so that the model is partially unidentified. Virtually all $\beta$ 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: The Relationship of $Q(\beta)$ to $K(\beta)$

In virtually all conceivable circumstances the construction of $Q(\beta)$ will not be feasible because $\Omega$ and $\Delta$ will be unknown. Consequently we need to adapt our previous arguments to allow for this fact. The simplest way of proceeding is to replace the unknown parameters by appropriate ‘plug-in’ values. To motivate
the approach we will begin by exploring the relationship of \( Q(\beta) \) to \( K(\beta) \).

Kleibergen (2002) introduced the test statistic \( K(\beta) \) in the context of the linear simultaneous equations model and the associated test procedure has been subsequently explored in Bekker and Kleibergen (2003); see also Kleibergen (2005). It is an asymptotically pivotal statistic and hence yields an asymptotically similar test, even in the presence of weak identification. However, it must be said that, when identification is weak, Kleibergen’s experimental results suggested that \( K(\beta) \) has unusual, and unattractive, power properties; see for example Kleibergen (2002, Figures 4 – 6). There has been little attempt in the literature to understand what drives this behaviour, a notable exception being Andrews, Moreira, and Stock (2004); as we shall see, our earlier results provide such an explanation.

The statistic \( K(\beta) \) is most easily thought of as a modified Anderson and Rubin (1949) statistic applied in a transformed space. We shall characterize the transformation as

\[
\begin{pmatrix}
  \tilde{y}_{\beta_0} \\
  \tilde{Y}_{\beta_0}
\end{pmatrix} = [y \ Y] T_0 = \begin{pmatrix}
  y - Y_{\beta_0}, Y - \bar{y}_{\beta_0} (\omega - \Omega_{\beta_0}) / \sigma_{u,\beta_0}^2
\end{pmatrix},
\]

where

\[
T_0 = \begin{bmatrix}
  1 & - (\omega - \Omega_{\beta_0})' / \sigma_{u,\beta_0}^2 \\
  -\beta_0 & I_n + \beta_0 (\omega - \Omega_{\beta_0})' / \sigma_{u,\beta_0}^2
\end{bmatrix}.
\]

In this transformed space the implied reduced form is

\[
\begin{pmatrix}
  \tilde{y}_{\beta_0} \\
  \tilde{Y}_{\beta_0}
\end{pmatrix} = [X \ Z] \tilde{\Pi} + [v \ V] T_0, \quad (8)
\]

where \( \tilde{\Pi} = \Pi T_0 \) and \( [v_{\beta_0} \ \tilde{V}_{\beta_0}] = [v \ V] T_0 \), and the structural equation in (1) becomes

\[
\bar{y}_{\beta} = \tilde{Y}_{\beta} \beta^* + X \gamma + v_{\beta_0}, \quad (9)
\]

where \( v_{\beta_0} = v_{\beta_0} - \tilde{V}_{\beta_0} \beta^* \) and

\[
\beta^* = - \frac{(\beta - \beta_0)}{1 - (\beta - \beta_0)' (\omega - \Omega_{\beta_0}) / \sigma_{u,\beta_0}^2}.
\]

**Lemma 4** Let \( E_0 = [\beta - \beta_0, I_n - (\beta - \beta_0) (\omega - \Omega_{\beta_0})' / \sigma_{u,\beta_0}^2] \). Then

\[
[\tilde{y}_{\beta_0} \ \tilde{Y}_{\beta_0}]' P [\tilde{y}_{\beta_0} \ \tilde{Y}_{\beta_0}] \sim W_{n+1} (\nu, \bar{\Sigma}_{\beta_0}, \bar{\Sigma}_{\beta_0}^{-1/2} \bar{A}_{\beta_0} \bar{S}_{\beta_0}^{-1/2})
\]

where \( \bar{A}_{\beta_0} = E_0' \Delta E_0 \) and \( \bar{\Sigma}_{\beta_0} = E_0' \Sigma E_0 \). Moreover,

\[
\bar{\Sigma}_{\beta_0} = \begin{bmatrix}
  \sigma_{u,\beta_0}^2 & 0' \\
  0 & \tilde{\Omega}_{\beta_0}
\end{bmatrix} = \begin{bmatrix}
  \sigma_{u,\beta_0}^2 & 0' \\
  0 & \sigma_{u,\beta_0}^2 \Omega - (\omega - \Omega_{\beta_0}) (\omega - \Omega_{\beta_0})' / \sigma_{u,\beta_0}^2
\end{bmatrix}.
\]

11
We note from Lemma 4 that the fundamental purpose of the transformation is to obtain the block diagonal structure of \( \hat{\Sigma}_{\beta_0} \), so that the elements of \( \tilde{v}_{\beta_0} \) are uncorrelated with those of \( \tilde{V}_{\beta_0} \).

Let \( \tilde{Q}(\beta^*_0) \) denote the value of the quadratic form \( Q(\cdot) \) when constructed in the transformed space and evaluated at the arbitrary point \( \beta^*_0 \). From equations (8) and (9) and Lemma 4 it is clear that the model in this transformed space is a special case of that considered in Theorem 1, implying that Theorem 1 applies to \( \tilde{Q}(\beta^*_0) \) in its entirety. In particular, we note that the hypothesis \( H_0 : \beta = \beta_0 \) in (1) corresponds to \( H^*_0 : \beta^* = 0 \) in (9). This, coupled with the lack of correlation between \( \tilde{v}_{\beta_0} \) and \( \tilde{V}_{\beta_0} \), leads to various simplifications, and application of Corollary 2 gives us

\[
\tilde{Q}(0) = \tilde{\beta}' \tilde{D}_{\beta_0} \tilde{\beta} \frac{\gamma_0}{\alpha} \frac{n \mathcal{F}(n, \nu - n + 1)}{(\nu - n + 1)},
\]

where \( \tilde{D}_{\beta_0} = (\tilde{\Omega}_{\beta_0} + \nu^{-1} \Delta)/\sigma^2_{u, \beta_0} \) and \( \tilde{\beta} = (\tilde{Y}_{\beta_0}' \tilde{P} \tilde{Y}_{\beta_0})^{-1} \tilde{Y}_{\beta_0}' \tilde{P} \tilde{Y}_{\beta_0} \).

One consequence of working in the transformed space is that the mean vector required to construct our previous critical regions and confidence intervals, namely \( \mu_{\beta} \), has been mapped into zero. The dispersion parameter, however, \( \tilde{D}_{\beta} \) in this transformed space, is still unknown because \( \Omega \) and \( \Delta \) will be unknown, so the construction of \( \tilde{Q}(0) \) will not be feasible. Indeed, not only is \( \tilde{D}_{\beta} \) unknown, but so too is the transformation \( T_0 \) itself.

Recall that, by assumption, \( \nu^{-1} \| \Delta \| \to 0 \). It follows from Lemma 4 and Theorem 1 of Poskitt and Skeels (2005) that under \( H_0 \) the marginal distribution of \( \tilde{Y}_{\beta_0}' \tilde{P} \tilde{Y}_{\beta_0} \) is approximately \( W_n(\nu, \tilde{\Omega}_{\beta_0} + \nu^{-1} \Delta) \). Thus, a natural estimator of \( (\tilde{\Omega}_{\beta_0} + \nu^{-1} \Delta) \) would be the sufficient statistic \( \nu^{-1} \tilde{Y}_{\beta_0}' \tilde{P} \tilde{Y}_{\beta_0} \) if \( \Sigma \) were known. Of course, as \( \Sigma \) is unknown we still cannot construct \( [\tilde{v}_{\beta_0}, \tilde{Y}_{\beta_0}] \). However,

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

Note that, from Lemma 4, under the alternative hypothesis

\[
\nu^{-1} \tilde{Y}_{\beta_0}' \tilde{P} \tilde{Y}_{\beta_0} \sim W_n(\nu, \tilde{\Omega}_{\beta_0} + \nu^{-1} \Delta_{\beta_0}^*),
\]

where \( \Delta_{\beta_0}^* = E'_{02} \Delta E_{02} \) with \( E'_{02} = I_n - (\beta - \beta_0)(\omega - \Omega_{\beta_0})'/\sigma^2_{u, \beta_0} \), so that \( \nu^{-1} \tilde{Y}_{\beta_0}' \tilde{P} \tilde{Y}_{\beta_0} \) remains sufficient for \( \tilde{\Omega}_{\beta_0} + \nu^{-1} \Delta_{\beta_0}^* \). Application of Theorem 1 then yields an approximate power function for \( \tilde{Q}(0) \). We might expect this approximation to be reasonable in regions of the parameter space wherein the magnitude of \( \beta - \beta_0 \) doesn’t swamp the impact of \( \nu^{-1} \| \Delta \| \to 0 \).
provides a consistent estimator of $\Sigma$, whatever the values of $\Pi_2$ and $\Delta$, and $\hat{\Sigma}$ can clearly be used to provide “plug in” values for the nuisance parameters.

Replacing the elements of $\Sigma$ by those of $\hat{\Sigma}$ in the definition of $T_0$ gives us an empirical version of $[\hat{y}_{\beta_0}, \hat{Y}_{\beta_0}]$, namely

$$[\hat{y}_{\beta_0}, \hat{Y}_{\beta_0}] = [y, Y] \hat{T}_0 = [y - Y_{\beta_0}, Y - \bar{y}_{\beta_0}(\hat{\omega} - \hat{\Omega}_{\beta_0})/\hat{\sigma}^2_{u,\beta_0}], \quad (10)$$

where $\hat{\sigma}^2_{u,\beta_0} = [1, -\beta_0^T \hat{\Sigma}[1, -\beta_0]'$. Substituting (10) for $[\hat{y}_{\beta_0}, \hat{Y}_{\beta_0}]$, and replacing $\hat{\Omega}_{\beta_0} + \nu^{-1} \Delta$ by $\nu^{-1} \hat{y}_{\beta_0}' P \hat{Y}_{\beta_0}$ in the quadratic form $\beta' D \beta$ yields

$$\frac{1}{\nu} \frac{1}{\hat{\sigma}^2_{u,\beta_0}} P \hat{Y}_{\beta_0} (\hat{y}_{\beta_0}' P \hat{Y}_{\beta_0})^{-1} \hat{y}_{\beta_0}' P \hat{Y}_{\beta_0} = \frac{1}{\nu} K(\beta_0).$$

Thus we find that a practically implementable version of our statistic $Q(\beta)$ constructed in the transformed space is equivalent to $K(\beta)$.

Kleibergen (2002) shows that, under $H_0$, $K(\beta_0)$ 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 $K(\beta_0)$ and suggest using critical values from the $F\{n, N - K\}$ distribution for the lower and, when re-scaled by $N/(N - k)$, upper bound exact critical values of $K(\beta_0)/n$. 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 $K(\beta_0)$ in the totally unidentified case. They show that if $\nu \to \infty$ as $N \to \infty$, such that $\nu/N \to \tau > 0$, then $(1 - \nu/N) K(\beta_0)$ 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 $K(\beta_0)$ 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 $K(\beta_0)$, 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

$$y = Y\beta + u$$
$$Y = Z\Pi_2 + V$$

with $n = 1$, $N = 100$, $Z \sim N(0, I_\nu \otimes I_N)$, where $\nu = 5$, $\Pi_2 = (0.1, 0, 0, 0, 0)'$, 

13
and \([u \ V] \sim N(0, \Phi \otimes I_N)\), where

\[
\Phi = \begin{bmatrix}
1 & 0.99 \\
0.99 & 1
\end{bmatrix}
\]

is the matrix defined in (6). The value of \(\Pi_2\) 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 \(\Delta\) is 0.9063. The correlation of 0.99 in the distribution of \([u \ V]\) implies that \(y\) and \(Y\) are strongly endogenous. We shall refer to this specification as DGP 1.

Figures 3a and 4a plot simulated power curves for \(Q(\beta)\), \(K(\beta)\), 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 DGP 1. For \(Q(\beta)\) critical regions were constructed using the small concentration \(F\) approximation, whereas for \(K(\beta)\) and \(BLIML\) they were determined on the basis of their asymptotic \(\chi^2(n)\) approximations. In Figure 3a the null hypothesis is \(H_0: \beta = 0\) and in Figure 4a it is \(H_0: \beta = 5\). In both cases we are testing \(H_0\) against the two-sided alternative at the 5\% significance level.

The key features of these figures are:

1. Although \(Q(\beta)\) and \(K(\beta)\) 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 \(\beta - \beta_0\) nor are they symmetric about the null value.

2. Although \(K(\beta)\) dominates \(Q(\beta)\) in terms of power, their power functions have profiles that are not too dissimilar. In particular, they both have greatest power at approximately \(\beta_0 - 1\).

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.\(^7\) Its power curve is asymmetric and appears to have a minimum at the point where the other tests have their maximum power.

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

\(^7\) It is worth noting that the chi-squared approximation for \(K(\beta)\) is a large \(N\) approximation whereas that for \(BLIML\) is a large \(\nu\) approximation. One might imagine that for this particular experimental design \(K(\beta)\) 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.
An important difference between $Q(\beta)$, on the one hand, and $K(\beta)$ 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 $Q(\beta)$ is of interest because (i) it displays similar patterns to those of $K(\beta)$ at values of $\beta$ close to $\beta_0$ and (ii) because the expression for the small concentration power of $Q(\beta)$, which is presented in Theorem 3, obviously explains the behaviour of $Q(\beta)$ observed here. In particular, we recall the dependence of the power function on the values of $\kappa_0$ and $q_0$. In the light of this we present in Figures 3b and 4b graphs showing how $\kappa_0$ and $q_0$ vary with $\beta$. 

Figure 3. Testing $\mathcal{H}_0: \beta = 0$ against $\mathcal{H}_1: \beta \neq 0$
Figure 4. Testing $\mathcal{H}_0 : \beta = 5$ against $\mathcal{H}_1 : \beta \neq 5$

From our previous arguments we know that the non-centrality parameter $\kappa_0$ is strictly positive for all $\beta \neq \beta_0$, and from Theorem 3 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 $\kappa_0$. 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 $\pi_{CR}(\beta)$ 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 $\beta$.

From these graphs it is clear that it is the variations in $q_0$ that are critical in determining the power characteristics of $Q(\beta)$ rather than the varia-
tions in \( \kappa_0 \). Indeed, there is almost a one–to–one correspondence between the graphs of \( 1/q_0 \) and the observed power function of \( Q(\beta) \). From Theorem 3 we can see that the impact of \( q_0 \) on \( \varrho_0 \) has greater influence on the value of \( \Psi\{q_0; n, \nu - n + 1, \kappa_0\} \) than do changes in \( \kappa_0 \), 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 \( K(\beta) \), albeit somewhat less strikingly. The rejection frequencies of \( BLIML \) behave very differently from those of \( Q(\beta) \) and \( K(\beta) \), but again they appear to be influenced directly by the value of \( q_0 \) more strongly than they do by the value of \( \kappa_0 \).

We observed above that the power functions of all three tests considered appear to have turning points at approximately \( \beta_0 - 1 \), which corresponds to the minimum value of \( q_0 \). The point \( \beta_0 - 1 \) is readily shown to be an artefact of the parameter choices in DGP 1, although it demonstrates an interesting property of \( Q(\beta) \), and seemingly also of \( K(\beta) \) and \( BLIML \). It is a useful exercise 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 1, for given values of \( \Sigma \) and \( \nu^{-1}\Delta \) the ratio \( q_0 \) is a function of \( \beta \) and \( \beta_0 \). However, in this experimental design

\[
\Sigma = \begin{bmatrix} 1 & \beta' \\ 0 & I_n \end{bmatrix} \Phi \begin{bmatrix} 1 & 0' \\ \beta' & I_n \end{bmatrix}, \quad \text{where} \quad \Phi = \begin{bmatrix} \sigma^2_{u,\beta} & \sigma_{uV} \\ \sigma_{Vu} & \Omega \end{bmatrix} \quad \text{(say)},
\]

is held fixed. Therefore \( \Sigma \) is itself implicitly a function of \( \beta \) for given \( \Phi \), so that \( q_0 \) is actually a function of \( \beta \) and \( \beta_0 \), given \( \nu^{-1}\Delta \) and \( \Phi \). It is useful to re-express \( q_0 \) as a function of \( \beta - \beta_0 \) and \( \beta_0 \). For given \( \nu^{-1}\Delta \) and \( \Phi \), it is a relatively simple matter to then show that \( q_0 \) is minimized at the point where \( \beta - \beta_0 = -\Omega^{-1}\sigma_{Vu} \). In the current example, \( \Omega = 1 \) and \( \sigma_{Vu} = 0.99 \) and so, rather than \( \beta_0 - 1 \), we see that the turning points are occurring at \( \beta_0 - 0.99 \).

If instead of choosing \( \sigma_{Vu} = 0.99 \) we had chosen \( \sigma_{Vu} = 0.55 \), for example, then we see from Figure 5 that the turning points occur much closer to \( \beta_0 = 5 \) than they did in Figure 4. As with Figure 4, \( Q(\beta) \) and \( K(\beta) \) attain their maximum power at values of \( \beta \) in the closed interval defined by those values of \( \beta \) 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.\(^8\) For \( \sigma_{Vu} = 0.55 \) this interval is much shorter than it was previously and the overall effect of this is that both \( Q(\beta) \) and \( K(\beta) \) have no useful power. The \( BLIML \) statistic, on the other hand, tends to have useful power for those values of \( \beta \) for which \( q_0 > 1 \), and so this narrowing of the interval corresponds to an increase

\(^8\) Observe that these intersections occur when \( q_0 = 1 \) and \( \beta = \beta_0 \) will always be one of these points.
in power for BLIML. But BLIML still suffers from a significant size distortion, in contrast to both $Q(\beta)$ and $K(\beta)$ which display good size properties.

Recall that a basic assumption underlying our theoretical development is that $[y \ Y]'[\Pi y \ Y] \sim \mathcal{W}_{n+1}(\nu, \Sigma, \Sigma^{-1/2} \Lambda \Sigma^{-1/2})$, 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 repre-
resents an externally imposed theoretical economic construct, are derived from
the reduced form via the compatibility conditions in (5). This suggests that
the experimental design of Kleibergen (2002), in which \( \Phi \) and \( \nu^{-1}\Delta \) are fixed
and \( \beta \), and implicitly \( \Sigma \), are allowed to vary, is not in direct accord with the
conceptual framework underlying such an analysis. Let us consider therefore
a slightly different model, referred to hereafter as DGP 2, in which
\[ y = Z\Pi_2\beta + v \]
\[ Y = Z\Pi_2 + V \]
where \([v V] \sim N(0, \Sigma \otimes I_N)\) and \( \Sigma \) is held fixed at
\[ \Sigma = \begin{bmatrix} 1 & 0.99 \\ 0.99 & 1 \end{bmatrix}, \]
with \( n, \nu, N, Z \) and \( \Pi_2 \) 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 hy-
pothesis \( H_0 : \beta = 0 \) at the 5% significance level. We now find that various
properties of \( Q(\beta) \) and \( K(\beta) \), and \( BLIML \) are reversed. In particular, \( Q(\beta) \)
and \( K(\beta) \) still exhibit the correct nominal size, with power functions that are
neither monotonically increasing functions of \( \beta - \beta_0 \) 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 un-
usual 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 \( Q(\beta) \) has its minimum power. Both \( Q(\beta) \) 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 \( \kappa_0 \). Under the current
scenario we find that \( q_0 \) is maximized at the point where \( \beta = \Omega^{-1}\omega \) and in the
current example, \( \Omega = 1 \) and \( \omega = 0.99 \), and so the turning points are occurring
at \( \beta = 0.99 \). As previously, \( Q(\beta) \) and \( K(\beta) \) exhibit useful power for values
of \( \beta \) for which \( q_0 \leq 1 \), whereas \( BLIML \) only has any useful power for those
values of \( \beta \) for which \( q_0 > 1 \). In Figure 6b the values of \( \beta \) 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.

Figure 7 repeats the experiment underlying Figure 6, except now we are test-
ing the null hypothesis \( H_0 : \beta = 5 \) 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, their shape is contingent on the value of $\beta_0$, in contrast with Figures 3 and 4. Second, although $Q(\beta)$ and $K(\beta)$ exhibit U-shaped power curves they do not have much power and, indeed, $Q(\beta)$ 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 $\beta = 0.99$, the point where the maximum of $q_0$ occurs. Once more it is apparent that the observed power characteristics of the tests can be readily explained by reference to our analytical results.

Finally, in our simulation experiments we have also considered the behaviour
Figure 7. Testing $H_0: \beta = 5$ against $H_1: \beta \neq 5$

of a practically implementable version of $Q(\beta)$. The empirical version was constructed in a manner analogous to the way the correspondence between $\tilde{Q}(0)$ and $K(\beta_0)$ was derived, using the same rationale, but without moving into the transformed space: namely, replacing the unknown parameter $\Omega + \nu^{-1}\Delta$ by the sufficient statistic $\nu^{-1}Y'PY$ and substituting the elements of $\Sigma$ for those of $\Sigma$. This gives the statistic

$$\hat{Q}(\beta) = (\hat{\beta} - \hat{\mu}_\beta)'\hat{D}_\beta(\hat{\beta} - \hat{\mu}_\beta)$$
where $\hat{\mu}_{\beta} = \beta + \nu(Y'PY)^{-1}(\hat{\omega} - \hat{\Omega}\beta)$ and

$$\hat{D}_{\beta} = \nu \left( \hat{\sigma}_{u,\beta}^2 - \nu(\hat{\omega} - \hat{\Omega}\beta)'(Y'PY)^{-1}(\hat{\omega} - \hat{\Omega}\beta) \right).$$

The performance of $\hat{Q}(\beta)$ obtained using the small-concentration asymptotic critical values $nF_{(0.95)}\{n, \nu - n + 1\}/(\nu - n + 1)$ is illustrated in Figure 8. Panels 8a and 8b in this figure reproduce the first panels of Figure 3 and Figure 6, respectively, except that the observed power curve of BLIML is replaced by that of $\hat{Q}(\beta)$. Comparison with Figures 3a and 6a indicates that although $\hat{Q}(\beta)$ generally behaves like $Q(\beta)$, the empirical power of $\hat{Q}(\beta)$ looks like a mixture of that of $Q(\beta)$, $K(\beta)$ and BLIML, which is perhaps not too surprising given $\hat{Q}(\beta)$'s structure and pedigree.

6 Conclusion

This paper has been concerned with the characterization of confidence intervals and hypothesis tests for the coefficients on the endogenous regressors ($\beta$) 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 $Q(\beta)$, a quadratic 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 $Q(\beta)$. We construct confidence regions by inverting $Q(\beta)$ and demonstrate that these regions possess certain desirable properties.

In practical situations the construction of the quadratic $Q(\beta)$ is not feasible, 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 a statistic $\hat{Q}(\beta)$ whose empirical behaviour is similar to that of $Q(\beta)$ in important ways. It transpires that, after suitable transformation, $\hat{Q}(\beta)$ maps into the statistic $\tilde{Q}(\beta)$ which is equivalent to the $K(\beta)$ statistic of Kleibergen (2002), and so our developments provide analytical insight into the sampling behaviour of $K(\beta)$ 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 this link 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.
(a) Testing $\mathcal{H}_0 : \beta = 0$ against $\mathcal{H}_1 : \beta \neq 0$ in DGP 1

(b) Testing $\mathcal{H}_0 : \beta = 0$ against $\mathcal{H}_1 : \beta \neq 0$ in DGP 2

Figure 8. Simulated Power Curves: $\mathcal{Q}(\beta)$ (solid), $\mathcal{K}(\beta)$ (dash-dot) and $\hat{\mathcal{Q}}(\beta)$ (dash), with 5% line (dotted)

References


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

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

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

wherein ${}_2F_1(k, l; m; x)$ denotes the hypergeometric function $\sum_{j=0}^{\infty} \frac{(k)_j (l)_j}{(m)_j} x^j$, with $(k)_j = \Gamma(k+j)/\Gamma(k)$, Pochhammer’s forward factorial function, $\kappa = \mathbf{\mu}'\mathbf{\mu}$. The corresponding distribution function, denoted $\Psi\{\varrho; p, \nu, \kappa\} = P(r \leq \varrho)$, is

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

where

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

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

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

Using the result

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

we obtain

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

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

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

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

\[
x^{(\nu+p-2)/2} \exp\{2\nu^{-1}r^{1/2}x\mu'\mathbf{v}\}dx.
\]

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

\[
\int_{\mathbf{v}'\mathbf{v} = 1} \exp\{\nu'k\}d\mathbf{v} = \frac{2\pi^{p/2}}{\Gamma\left(\frac{p}{2}\right)} {}_0\text{F}_1\left(\frac{p}{2}; \frac{1}{4}k'^2\right)
\]

for any fixed \( p \)-vector \( \mathbf{k} \) and, writing \( \kappa = \mu'\mu \), we obtain

\[
f(r) = \frac{r^{(p-2)/2}}{\nu^{p/2} \Gamma\left(\frac{p}{2}\right) \Gamma\left(\frac{p}{2}\right)} \int_0^\infty \exp\{-[1 + \nu^{-1}(r + \kappa)]x\}
\]

\[
x^{(\nu+p-2)/2} {}_0\text{F}_1\left(\frac{p}{2}; \nu^{-2}r\kappa x^2\right)dx. \quad (A.4)
\]

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

\[
\mathcal{I} = \int_0^\infty \exp\{-[1 + \nu^{-1}(r + \kappa)]x\}x^{(\nu+p-2)/2} {}_0\text{F}_1\left(\frac{p}{2}; \nu^{-2}r\kappa x^2\right)dx
\]

\[
= \sum_{j=0}^{\infty} \frac{(\nu^{-2}r\kappa)^j}{j!} \int_0^\infty \exp\{-[1 + \nu^{-1}(r + \kappa)]x\}x^{2j+(\nu+p-2)/2}dx
\]

\[
= \sum_{j=0}^{\infty} \frac{(\nu^{-2}r\kappa)^j}{j!} \frac{(1 + \nu^{-1}(r + \kappa))^{-(2j+(\nu+p)/2)}}{\Gamma\left(2j + \frac{\nu + p}{2}\right)}
\]

\[
= \Gamma\left(\frac{\nu + p}{2}\right) \left[1 + \nu^{-1}(r + \kappa)\right]^{-(\nu+p)/2} \sum_{j=0}^{\infty} \frac{\left(\frac{\nu + p}{2}\right)_j}{j!} \left[\frac{4\kappa r}{[\nu + \kappa + r]^2}\right]^j.
\]

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

\[
\mathcal{I} = \Gamma\left(\frac{\nu + p}{2}\right) \left[1 + \nu^{-1}(r + \kappa)\right]^{-(\nu+p)/2} \sum_{j=0}^{\infty} \frac{\left(\frac{\nu + p}{4}\right)_j}{j!} \left[\frac{4\kappa r}{[\nu + \kappa + r]^2}\right]^j
\]

\[
= \Gamma\left(\frac{\nu + p}{2}\right) \left[1 + \nu^{-1}(r + \kappa)\right]^{-(\nu+p)/2} 2F_1\left(\frac{\nu + p}{4}, \frac{\nu + p + 2}{4}; \frac{p}{2}; \frac{4\kappa r}{[\nu + \kappa + r]^2}\right).
\]

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 \(I\) 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
\[
\Psi\{p; \nu, \kappa\} = \sum_{j=0}^{\infty} w_j \int_0^e \frac{r^{j+(p-2)/2}}{[1 + r/(\nu + \kappa)]^{2j+(\nu+p)/2}} \, dr,
\]
where
\[
w_j = \frac{\nu^{p/2}\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{p}{2}\right)} \frac{(\nu+p+2)_j}{j!} \frac{(4\kappa)_j}{(\nu+k+2j+(\nu+p)/2)}.
\]
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) \(F\{p, \nu\}\) distribution. We will therefore refer to \(\kappa\) 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 Theorem 1 First observe that \(D_{\beta_0} = D_{\beta}/q_0\). It follows that
\[
(\bar{\beta} - \mu_{\beta})' D_{\beta} (\bar{\beta} - \mu_{\beta}) |_{\beta = \beta_0} = \frac{\hat{t}_0 \hat{t}_0}{(\nu - n + 1)q_0},
\]

28
where
\[ \hat{t}_0 = ((\nu - n + 1)D_{\beta})^{1/2} (\beta - \mu_{\beta_0}) = \hat{r} + \mu, \]
with
\[ \hat{r} = ((\nu - n + 1)D_{\beta})^{1/2} (\beta - \mu_{\beta}), \]
and
\[ \mu = ((\nu - n + 1)D_{\beta})^{1/2} (\mu_{\beta} - \mu_{\beta_0}). \]

As \( \hat{\beta} \sim t_n((\nu - n + 1), \mu_{\beta}, D_{\beta}), \) it follows from Slutsky’s theorem and Lemma A.1 that
\[ \frac{\hat{t}_0' \hat{t}_0}{\mu_{\beta}'} \sim \psi\{n, \nu - n + 1, \mu' \mu\}, \]
which establishes the desired result. \( \square \)

**Proof of Corollary 2** If \( \beta_0 \) is equal to the true value \( \beta \) then, from Theorem 1, \( \kappa_0 = 0 \) and \( q_0 = 1. \) From (A.1) we see that \( \psi(r \colon n, \nu - n + 1, 0) \) collapses to
\[ \frac{\Gamma\left(\frac{\nu + 1}{2}\right)}{\Gamma\left(\frac{\nu - n + 1}{2}\right) \Gamma\left(\frac{\nu}{2}\right)} \frac{r^{(n - 2)/2}}{(\nu - n + 1)^{n/2}} \left[ 1 + \frac{r}{\nu - n + 1} \right]^{-(\nu + 1)/2}. \]

If we then make the transformation \( r = nF \) we see that \( F \sim F(n, \nu - n + 1). \)
That is, \( (\nu - n + 1)Q(\beta)/n \sim F(n, \nu - n + 1), \) which is the stated result.

**Proof of Theorem 3** Using notation from the proof of Theorem 1
\[ P\left( (\beta - \mu_{\beta})'D_{\beta} (\beta - \mu_{\beta}) \right|_{\beta = \beta_0} \geq \frac{nF_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right) \]
\[ = P\left( \frac{\hat{t}_0' \hat{t}_0}{(\nu - n + 1)q_0} \geq \frac{nF_{(1-\alpha)}\{n, \nu - n + 1\}}{(\nu - n + 1)} \right) \]
\[ = 1 - P\left( \frac{\hat{t}_0' \hat{t}_0}{(\nu - n + 1)q_0} < nq_0F_{(1-\alpha)}\{n, \nu - n + 1\} \right). \]

The statement of the theorem follows directly. \( \square \)

**Proof of Lemma 4** The first part of the lemma follows from basic properties of the Wishart distribution. To show that \( \Sigma_{\beta_0} = E_0' \Sigma E_0 \) is block diagonal with elements as described is a simple exercise in matrix algebra. \( \square \)