ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

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

Tests for Over-identifying Restrictions in Partially Identified Linear Structural Equations

Giovanni Forchini

November 2006

Working Paper 20/06

Tests for Over-identifying Restrictions in Partially Identified Linear Structural Equations

by

G. Forchini[1][2]

Monash University

November 20, 2006

Abstract

Cragg and Donald (1996) have pointed out that the asymptotic size of tests for overidentifying restrictions can be much smaller than the asymptotic nominal size when the structural equation is partially identified. This may lead to misleading inference if the critical values are obtained from a chi-square distribution. To overcome this problem we derive the exact asymptotic distribution of the Byron test statistic. This allows the calculation of asymptotic critical values and p-values corrected for possible failure of identification.

JEL Classification: C12, C30

Key Words: Invariant Tests, Over-identifying restrictions, Partially identified structural equation

[1] Address for correspondence: Giovanni Forchini, Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria 3800, Australia.
E-mail: Giovanni.Forchini@BusEco.monash.edu.au

[2] I thank Les Godfrey, Grant Hillier, Peter Phillips and Katharina Hauck for helpful comments and discussions.

1. Introduction

Tests for over-identifying restrictions are certainly one the most important tools available to practitioners for detecting misspecification of linear structural equations. They have been studied, among many others, by Sargan (1958), Basmann (1960a), Basmann (1960b), Byron (1974) and Hansen (1982).

Over-identification is distinct from the concept of identification which has recently attracted a lot of attention in econometrics. Over-identification refers to the compatibility between the structural equation and the reduced form, so that tests for over-identification serve as checks for the coherency of structural equation and reduced form. On the other hand, identification pertains to the fact that some of the structural parameters may not be uniquely defined given a correctly specified model. The implication of having a structural equation that is misspecified or unidentified are very different. In the first case, the researcher can try to improve the model, hoping to find one that is not clearly misspecified. In the second case, the model use is well specified, but it is not informative about the parameters of interest.

Tests for over-identifying restrictions are often investigated taking identification of the structural parameters for granted. However, Phillips (1989), Choi and Phillips (1992) and Staiger and Stock (1997) have convincingly argued that identification of the structural parameters may fail in very common situations. Recently, considerable attention has been given to the fact that the parameters of a linear structural equation may be *unidentified* (e.g Sims (1980), Sargan (1983), Phillips (1983) and Hillier (1985)), *partially identified* (e.g. Phillips (1989) and Choi and Phillips (1992)) or *weakly identified* (e.g. Staiger and Stock (1997)). Concerns have been raised about the severity of the consequences of various forms of lack of identification of econometric models. However, the robustness of tests for over-identification to identification failures has not been fully studied.

Cragg and Donald (1996) have shown that commonly used tests for over-identification may lead to misleading inference when identification fails. Precisely, they show that lack of identification tends to concentrate the probability mass of a test statistic for over-identification around zero. Therefore, such tests tend to suggest a correct specification more often than one would expect under classical conditions.

In this paper, we investigate the properties of tests for over-identification and focus on their robustness to partial identification in the spirit of Phillips (1989) and

Choi and Phillips (1992). By refining results of Cragg and Donald (1996), we derive an asymptotic representation of the Byron test statistic that holds when identification or the over-identifying conditions fail. This allows us to find a closed form expression for the asymptotic distribution of Byron test that can be used to calculate asymptotic critical values under partial identification.

We propose a procedure to consistently test for over-identifying restrictions that has the correct asymptotic size in models that are only partially unidentified. Our method is based on the test statistic suggested by Byron (1974) (or the asymptotically equivalent statistic recommended by Basmann (1960a) and Basmann (1960b)) for which we modify the critical values to take into account the estimated rank of the matrix of correlations between the endogenous variables included as regressors in the structural equation and the instruments. Our procedure differs from the one suggested by Cragg and Donald (1996) in that (i) we modify the critical values of existing tests rather than the test statistics themseves; and (ii) we do not need to establish which structural parameters are identified and which ones are not. A further contribution of our work is the realization that the problem of deriving the asymptotic distributions of tests for over-identifying restriction using a GMM approach can be considerably simplified using simple invariance arguments.

The paper is structured as follows. Section 2 presents a commonly used Linear structural equation model, briefly discusses identification and over-identification. Section 3 formulates the testing problem in a GMM framework, and lists the assumptions used. Section 4 investigates the asymptotic properties of Byron test for over-identifying restrictions. Some numerical results are presented and discussed in Section 5. Section 6 concludes. Proofs are in the appendix.

2. The model

We consider a linear structural equation of the form

(1) $$y_1 = Y_2\beta + Z_1\gamma + u$$

where y_1 and Y_2 are, respectively, a $(T \times 1)$ vector and a $(T \times n)$ matrix of endogenous variables, Z_1 is a $(T \times k_1)$ matrix of exogenous variables, and u is a $(T \times 1)$ vector of random variables. The structural parameters β and γ are of dimension $(n \times 1)$ and $(k_1 \times 1)$, respectively. The reduced form associated with (1) is

(2) $$\left[y_1, Y_2\right] = Z_1\Phi + Z_2\Pi + \left[v_1, V_2\right]$$

where Z_2 is a $(T \times k_2)$ matrix of exogenous variables excluded from the structural equation with $k_2 \geq n$, and the random matrix $\left[v_1, V_2\right]$ is partitioned conformably to $\left[y_1, Y_2\right]$. The reduced form parameters Φ and Π are of dimension $(k_1 \times n+1)$ and $(k_2 \times n+1)$ respectively. We also assume that the rows of $\left[v_1, V_2\right]$ conditional on $\left[Z_1, Z_2\right]$ have covariance matrix Ω of dimension $(n+1 \times n+1)$.

Practitioners tend to interpret the *i-th* component of β as the unit change in the endogenous variable on the left-hand-side of (1) *caused* by a unit change in the *i-th* endogenous variable on the right-hand-side of (1). This, often unspoken, causality relation leads to the specification of the structural equation in (1), and prevents practitioners from specifying the structural equation with no explicit normalization as advocated by Hillier (1990), despite its advantages (see also Hillier (2006)).

By specifying the reduced form (2) we are implicitly assuming that the conditional distribution of $\left[y_1, Y_2\right]$ given $\left[Z_1, Z_2\right]$ can provide information about Φ, Π and Ω, and functions thereof only. The structural parameters are regarded as functionals on the space of distributions of $\left[y_1, Y_2\right]$ given $\left[Z_1, Z_2\right]$, and can be written in terms of the reduced form parameters. To see this we partition $\Pi = \left[\pi_1, \Pi_2\right]$ and $\Phi = \left[\phi_1, \Phi_2\right]$ conformably to $\left[y_1, Y_2\right]$ and insert the reduced form (2) into the structural equation (1) to obtain

(3) $$Z_1\phi_1 + Z_2\pi_1 + v_1 = \left(Z_1\phi_2 + Z_2\Pi_2 + V_2\right)\beta + Z_1\gamma + u .$$

For the structural equation to be *compatible* with the reduced form we must have

(4) $$\pi_1 = \Pi_2\beta$$

(5) $$\phi_1 = \Phi_2\beta + \gamma$$

and

(6) $$v_1 = V_2\beta + u .$$

Equation (4), (5) and (6) define β, γ and u, and are known as the *overidentifying* restrictions (e.g. Byron (1974) and Hausman (1983)), or the *identification* condition (e.g. Phillips (1983)).

The following result is well known.

Proposition 1. *Necessary and sufficient conditions for the structural parameters* β *to be identified is that (i) equation (4) holds and (ii)* Π_2 *has rank* n*. Necessary and sufficient condition for the structural parameter* γ *to be identified without further restrictions on the reduced form parameter* Φ_2 *is that* β *is identified.*

Notice that even if β is unidentified, the parameter γ could be identified provided further restrictions on Φ_2 are imposed. For example, if Π_2 has rank zero, then γ is identified if $\Phi_2 = 0$ (e.g. Phillips (1989) and Choi and Phillips (1992)).

Proposition 1 acknowledges that identification of the structural parameters relies on the simultaneous conditions that equation (4) holds and Π_2 has rank n. The first condition is the null hypothesis for tests of over-identifying restrictions, the second one is the focus of tests of identification. Notice that although both are needed to achieve identification, only one of the conditions is usually tested with the other being regarded as satisfied.

3. Byron test statistic and assumptions

Equation (4) can be written in the equivalent form

$$(Q\pi_1)' M_{(Q\Pi_2)} (Q\pi_1) = 0 \tag{7}$$

where Q is an arbitrary non-singular ($k_2 \times k_2$) fixed matrix, and for any ($k_2 \times n$) matrix A of rank r, $M_A = I_{k_2} - AA^\dagger$ and $A^\dagger$ denotes the Moore-Penrose inverse of A. Thus, a test for the validity of the over-identifying restrictions (4) is just a test for the null hypothesis that (7) holds against the alternative that it does not. A GMM test for the validity of (7) (or equivalently (4)) can be based on

$$(\hat{Q}\hat{\pi}_1)' M_{(\hat{Q}\hat{\Pi}_2)} (\hat{Q}\hat{\pi}_1) \tag{8}$$

where $\left[\hat{\pi}_1, \hat{\Pi}_2\right]$ is the OLS estimator of $\Pi = \left[\pi_1, \Pi_2\right]$ in the reduced form given in (11) and $\hat{Q}$ can be chosen as $\hat{Q} = \left(T^{-1} Z_2' M_{Z_1} Z_2\right)^{1/2}$. This justifies the use of a statistic having the asymptotic form

$$B = \frac{T (\hat{Q}\hat{\pi}_1)' M_{(\hat{Q}\hat{\Pi}_2)} (\hat{Q}\hat{\pi}_1)}{\left(1 + \beta^{*\prime} \beta^*\right) \omega_{11.2}} \tag{9}$$

where β^* denotes the *canonical* coefficients of the endogenous variables

(10) $$\beta^* = \left(\Omega_{22}^{1/2}\beta - \Omega_{22}^{-1/2}\omega_{21}\right)/\omega_{11.2}^{1/2}$$

in the structural equation (e.g. Phillips (1983)), $\omega_{11.2} = \omega_{11} - \omega_{21}'\Omega_{22}\omega_{21}$ and

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{pmatrix}$$

is partitioned conformably to $[y_1, Y_2]$. Under standard assumptions $\hat{B} \to^p \chi^2(k_2 - n)$ if β is identified. Notice that the denominator of (9) is the variance of the error in the structural equation.

The tests for over-identifying restrictions of Sargan (1958), Basmann (1960a), Basmann (1960b) and Hansen (1982) replace the denominator of (9) with $\hat{u}_{TSLS}'\hat{u}_{TSLS}/T$ where $\hat{u}_{TSLS}$ is the vector of TSLS residuals in (1). The test of Byron (1974) – denoted by $\hat{B}$ – uses β and Ω estimated, respectively, with TSLS and $\hat{\Omega}$. Provided the reduced form is correctly specified, a consistent estimator of Ω can be obtained from equation (2), and this will not be affected by failure of identification of β.

In order to derive the asymptotic distribution of $\hat{B}$ we make essentially the same assumptions as Cragg and Donald (1996).

Assumption 1. *The following conditions hold:*

(a) $\hat{Q} = T^{-1}Z_2'M_{Z_1}Z_2 \to^p Q$ *where* Q *is a fixed, finite, positive definite* ($k_2 \times k_2$) *matrix;*

(b) $\hat{\Omega} = [y_1, Y_2]'M_{[Z_1,Z_2]}[y_1, Y_2]/T \to^p \Omega$*;*

(c) The OLS estimator of $\Pi = [\pi_1, \Pi_2]$,

(11) $$\left[\hat{\pi}_1, \hat{\Pi}_2\right] = \left(Z_2'M_{Z_1}Z_2\right)^{-1} Z_2'M_{Z_1}[y_1, Y_2]$$

satisfies

(12) $$T^{1/2}\left(\left[\hat{\pi}_1, \hat{\Pi}_2\right] - [\pi_1, \Pi_2]\right) \to^d N\left(0, Q^{-1} \otimes \Omega\right).$$

Assumption 2. *The rank of* Π_2 *is* n_1 *where* $0 \le n_1 \le n$ *and is unknown.*

Assumption 1 is standard. Assumption 2 allows the structural equation to be partially identified in the sense of Phillips (1989) and Choi and Phillips (1992). Notice that our set-up can be further simplified without loss of generality: the problem of testing the null hypothesis that (7) holds against the alternative that it does not, and the statistic (9) have an invariance property described by the following lemma that has not been noticed before. This allows us to simplify the set-up considerably without compromising the generality of our results.

Lemma 1. *Both the testing problem and* B *are invariant to the transformations*

(13) $$\left[\hat{\pi}_1, \hat{\Pi}_2\right] \rightarrow \left[\hat{\pi}_1, \hat{\Pi}_2\right] L$$

where L *is the* $(n+1 \times n+1)$ *matrix*

(14) $$L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{pmatrix},$$

with L_{22} *being a non-singular* $(n \times n)$ *matrix and* $l_{11} > 0$. *Therefore, there is no loss of generality in imposing the following restrictions:*

(a) $\Omega = I_{n+1}$;

(b) $\Pi_2 = \left[\Pi_{21}, 0\right]$ *where* Π_{21} *is a matrix of dimension* $(k_2 \times n_1)$ *with rank* $0 \le n_1 \le n$, *and* 0 *denotes a* $(k_2 \times n_2)$, $n_2 = n - n_1$, *matrix of zeros;*

(c) β^* *can be partitioned conformably to* Π_2 *as* $\beta^* = \left[\beta_1^{*}{}', \beta_2^{*}{}'\right]'$ *and* β_1^* *is identified while* β_2^* *is unidentified.*

The particular block-triangular form of the matrix *L* reflects the fact that post multiplication by *L* must leave unchanged both the over-identifying condition (4) and the rank of Π_2. If we would insist that *L* is non-singular only, one or both of these conditions would be violated.

We now need to specify in what way the compatibility condition (4) may be violated. In this case π_1 is not a linear combination of the columns of Π_2, that is, we set

(15) $$\pi_1 = \Pi_{21}\beta_1^* + T^{-1/2}\Pi_{21}^{\perp}\beta^{\perp}$$

where $\left(\Pi_{21}^{\perp}\right)' Q \Pi_{21}^{\perp} = I_{k_2 - n_1}$ and $\left(\Pi_{21}\right)' Q \Pi_{21}^{\perp} = 0$.

4. Asymptotic properties of Byron test statistic $\hat{B}$

In this section we study the asymptotic properties of $\hat{B}$.

Theorem 1. *Suppose that Assumptions 1 and 2, and equation* (15) *hold. Then*

(16) $$\hat{B} \to^{d} b = \frac{\tau}{1 + r_2' r_2}$$

where

(17) $$r_2 \mid \delta \sim N\left((\delta'\delta)^{-1} \delta' \beta^{\perp} / \left(1 + \beta_1^{*}{}' \beta_1^{*}\right)^{1/2}, (\delta'\delta)^{-1} \right)$$

(18) $$\tau \mid \delta \sim \chi^2\left(k_2 - n, \frac{\beta^{\perp}{}' M_{\delta} \beta^{\perp}}{1 + \beta_1^{*}{}' \beta_1^{*}} \right)$$

and

(19) $$\delta \sim N\left(0, I_{k_2 - n_1} \otimes I_{n_2}\right).$$

Moreover, r_2 and τ are independent conditional on δ.

Theorem 1 gives an explicit asymptotic representation for the distribution of $\hat{B}$. Several known results can be obtained as special cases. If the model is identified $\hat{B} \to^{d} \chi^2(k_2 - n)$ (e.g. Byron (1974)). For local failure of the compatibility condition (4) but with rank of Π_2 equal to n, $\hat{B} \to^{d} \chi^2\left(k_2 - n, \beta^{\perp}{}'\beta^{\perp} / \left(1 + \beta_1^{*}{}'\beta_1^{*}\right)\right)$, indicating that the test is asymptotically unbiased and consistent.

If $\beta^{\perp} = 0$ and the rank of Π_2 equals $n_1 < n$ then $\tau \sim \chi^2(k_2 - n)$ and $r_2 \mid \delta \sim N\left(0, (\delta'\delta)^{-1}\right)$ are independent. Notice that in this case $\Pr\left\{\hat{B} \geq c\right\} \leq \int_0^c d\chi^2(k_2 - n)$ (c.f. Theorem 4 of Cragg and Donald (1996)). The following corollary shows that in such a case $\hat{B}$ has a non-standard non-central chi-squared distribution.

Corollary 1.1 *Suppose that Assumptions 1 and 2, and equation* (15) *hold. If $\beta^{\perp} = 0$ and the rank of Π_2 equals $n_1 < n = n_1 + n_2$ then the asymptotic distribution of $\hat{B}$ is*

$$(20)\qquad pdf\left(b\right)=\frac{e^{-b/2}b^{(k_2-n)/2-1}}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)}\frac{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i}{2}+1\right)}{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i+1}{2}\right)}\Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{b}{2}\right)$$

where Ψ *denotes a Tricomi confluent hypergeometric function (e.g. Lebedev (1972)). Moreover,*

$$CDF\left(b\right)=\frac{2\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i}{2}\right)}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i+1}{2}\right)}\times$$

$$(21)\qquad \left\{\frac{\sqrt{\pi}}{\Gamma\left(\frac{n_2+1}{2}\right)}\frac{b^{\frac{k_2-n}{2}}}{k_2-n}\,{}_2F_2\left(\frac{1-n_2}{2},\frac{k_2-n}{2};\frac{k_2-n}{2}+1,\frac{1}{2};-\frac{b}{2}\right)\right.$$

$$\left.-\frac{\sqrt{2\pi}}{\Gamma\left(\frac{n_2}{2}\right)}\frac{b^{\frac{k_2-n+1}{2}}}{k_2-n+1}\,{}_2F_2\left(\frac{2-n_2}{2},\frac{k_2-n+1}{2};\frac{k_2-n+1}{2}+1,\frac{3}{2};-\frac{b}{2}\right)\right\}$$

where ${}_2F_2$ *is Gauss hypergeometric function (e.g. Lebedev (1972)).*

Thus, we can use Byron's test for over-identifying restrictions even if the rank condition fails. In fact, if we know n_1, equation (21) allows us to find the correct asymptotic p-values for Byron test. If we do not know n_1 we can apply a two-step procedure. In the first step the rank of Π_2 is estimated as $\hat{n}_1$. This can be done with several consistent methods (e.g. Cragg and Donald (1996) and Robin and Smith (2000)) that use only the reduced form of Y_2, and, thus, do not involve the over-identifying restrictions themselves. In the second step, Byron or Basmann test statistics can be calculated and their p-values can obtained from (21) with n_1 replaced by $\hat{n}_1$.

Our procedure has two advantages over the one proposed by Cragg and Donald (1996) (Section 2.3). First, it is simple and relies on test statistics that are computed by standard packages, whereas Cragg and Donald (1996) suggests modifying the test statistics. Second, our procedure can be applied without having to select the identified parameters.

When the compatibility condition fails and the rank of Π_2 equals $n_1 < n$, the test is consistent, but for local departures as in equation (15) it is difficult to disentangle the two effects as the following result shows.

Corollary 1.2 *Suppose that Assumptions 1 and 2, and equation* (15) *hold. If the rank of* Π_2 *equals* $n_1 < n = n_1 + n_2$ *then the asymptotic distribution of* $\hat{B}$ *is*

$$pdf(b) = \frac{e^{-b/2} b^{(k_2-n)/2-1}}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)} \tag{22}$$
$$\times \frac{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i}{2}+1\right)}{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i+1}{2}\right)}\exp\left\{-\frac{1}{2}\lambda\right\}$$
$$\times \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{\left(\frac{k_2-n_1+1}{2}\right)_j\left(\frac{n_2}{2}\right)_j}{i!\,j!\left(\frac{k_2-n_1}{2}\right)_{j+i}}\left[\frac{1}{2}\lambda\right]^{i+j}\left(\frac{b}{2}\right)^i\Psi\left(\frac{n_2}{2}+j;\frac{1}{2}+i;\frac{b}{2}\right)$$

where Ψ *denotes a Tricomi confluent hypergeometric function (e.g. Lebedev (1972)), and* $\lambda = \beta^{\perp}{}'\beta^{\perp}/\left(1+\beta_1^*{}'\beta_1^*\right)$.

Using results of Cragg and Donald (1996) one can easily show that Byron’s test statistic is asymptotically equivalent to Basmann’s test statistic in all situations considered in this paper. Therefore, Theorem 1 and its corollaries also characterize the asymptotic distribution of the latter.

5. Numerical results

We now illustrate some of the properties of the asymptotic distribution of Byron’s test statistic using graphs. First we study the effect of lack of identification on the density of $\hat{B}$. Figure 1 shows the typical density of $\hat{B}$ for fixed $k_2 = 10$ and $n = 3$, and $n_2 = 0$ (solid line), $n_2 = 1$ (dotted line), $n_2 = 2$ (dashed line) and $n_2 = 3$ (dotted-dashed line). Clearly, the density of $\hat{B}$ tends to shift to towards the origin as the number of unidentified components of β increases. Therefore, if we choose the critical value for the test from the tables of a $\chi^2(k_2 - n)$, the test may be seriously undersized as Table 1 shows. This is especially true for small k_2.

[Figure 1 approximately here]

[Table 1 approximately here]

Next we consider the combined effect of lack of identification and violation of the compatibility conditions. Figure 2 shows a typical graph for the asymptotic density of $\hat{B}$ for $k_2 = 10$, $n = 3$ when $n_2 = 0$, i.e. the structural parameters are identified, (dashed line), $n_2 = 2$ and no violation of the over-identifying restrictions (dashed line), and $n_2 = 2$ and violation of the over-identifying restrictions with $\lambda = 12$ (solid line). The dot-dashed line is the density of a non-central chi-squared distribution with $k_2 - n = 7$ degrees of freedom and non-centrality parameter equal to $\lambda = 12$. The effect of violation of the over-identifying restrictions on the density of $\hat{B}$ when $n_2 = 2$ are not as marked as in the case where the structural parameters are completely identified.

Figure 3 shows the potential loss of asymptotic power due to the use of the incorrect critical values from a chi-square distribution when the rank condition is violated. The striking, but not unexpected, feature is that the test is asymptotically biased if the structural equation is partially unidentified and the critical values are not adjusted.

[Figure 2 approximately here]

[Figure 3 approximately here]

Next, we assess the goodness of the approximation offered by the asymptotic theory. We compare the asymptotic and the small sample size of Byron test for different values of T and n_2. Table 2 shows some representative results for $n = 4$ and $k_2 = 8$. In Table 2, first two columns, the size is based on the critical values obtained from (20). However, for the results in the first column n_2 is calculated by estimating $n_1 = n - n_2$ with the procedure suggested by Robin and Smith (2000) while, in the second column, n_2 is taken as known. The third column contain the size of the test when the critical values are obtained from a chi-square distribution $\chi^2(k_2 - n)$. The random variates are generated as independent $(T-k_2)\hat{\Omega} \sim W_{n+1}(T-k_2, I_n)$ and $\hat{\Pi} \sim \begin{pmatrix} \beta & I_{n_1} \\ 0 & 0 \end{pmatrix} + T^{-1/2} N\left(0, I_{k_2} \otimes I_{n+1}\right)$, and β is a vector of ones, and $\hat{Q}$ is taken to be an identity matrix. The size of the rank test used in estimating the rank of Π_2 is α=.01ln(100)/ln(T). The number of replications employed in the Monte Carlo test is 30,000.

It is evident from Table 2 that there are only small differences between the first two columns, so that estimating n_2 does not have a significant effect on the size of Byron test when the critical values are obtained from Corollary 1.1. The size of the classical Byron test is strongly affected by failure of the rank condition. For small sample size ($T < 100$, say) all three versions of Byron test seem to be oversized independently of the method used to calculate the critical values

[Table 2 approximately here]

6. Conclusions

Classical tests for over-identification may be seriously misleading in partially identified linear structural equations, however, by modifying the critical value of Byron or Basmann tests to take into account such a possible failure of identification, we can construct a consistent testing procedure having asymptotically the correct size. In contrast to the method of Cragg and Donald (1996) our procedure can be implemented without the need to modify the test statistic and to select the identified parameters.

Appendix: Proofs

Proof of Lemma 1

Invariance of the testing problem. The transformation (13) induces the transformations

$$\Pi_2 \to \Pi_2 L_{22} = \bar{\Pi}_2$$
$$\pi_1 \to l_{11}\pi_1 - \Pi_2 l_{21} = \bar{\pi}_1$$
$$\Omega \to L'\Omega L = \bar{\Omega}$$

in the parameter space. Note that if (7) holds then

$$\left(Q\bar{\pi}_1\right)' M_{(Q\bar{\Pi}_2)} \left(Q\bar{\pi}_1\right) = \left(Q\pi_1\right)' M_{(Q\Pi_2)} \left(Q\pi_1\right) = 0,$$

otherwise $\left(Q\bar{\pi}_1\right)' M_{(Q\bar{\Pi}_2)} \left(Q\bar{\pi}_1\right) > 0$.

Invariance of the test statistic. The statistics $\left[\hat{\pi}_1, \hat{\Pi}_2\right]$ transform as

$$\hat{\Pi}_2 \to \hat{\Pi}_2 L_{22}$$
$$\hat{\pi}_2 \to l_{11}\hat{\pi} - \hat{\Pi}_2 l_{21}.$$

Replacing these in (9), the numerator changes according to

$$\left(\hat{Q}\hat{\pi}_1\right)' M_{(\hat{Q}\hat{\Pi}_2)} \left(\hat{Q}\hat{\pi}_1\right) \to l_{11}^2 \left(\hat{Q}\hat{\pi}_1\right)' M_{(\hat{Q}\hat{\Pi}_2)} \left(\hat{Q}\hat{\pi}_1\right).$$

It can be easily checked that $\omega_{11.2} \to l_{11}^2 \omega_{11.2} = \bar{\omega}_{11.2}$, which shows that the statistic is also invariant to the transformations (13).

It follows that there is no loss of generality to assume that $\Omega = I_{n+1}$, because we can transform the model using

$$L = \begin{pmatrix} \omega_{11.2}^{-1/2} & 0 \\ -\omega_{11.2}^{-1/2}\Omega_{22}^{-1}\omega_{21} & \Omega_{22}^{-1/2} \end{pmatrix}$$

such that $L'\Omega L = I_{n+1}$. Note that in this case

$$\bar{\Pi}_2 = \Pi_2 \Omega_{22}^{-1/2}$$
$$\bar{\pi}_1 = \omega_{11.2}^{-1/2}\left(\pi_1 - \Pi_2 \Omega_{22}^{-1}\omega_{21}\right),$$

and that if (4) holds then $\bar{\pi}_1 = \bar{\Pi}_2 \beta^*$, otherwise $\bar{\pi}_1 \neq \bar{\Pi}_2 \beta^*$. Thus, we can assume that the structural equation is in *canonical form* (e.g. Phillips (1983)). The invariance property above also applies to the model when the structural equation is reduced to its canonical form. In this case, we can choose another matrix L of the form

$$L = \begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix}$$

where H is an $(n \times n)$ orthogonal matrix such that $\bar{\Pi}_2 H = \left[\Pi_{21}, 0\right]$ and the rank of Π_{21} is the same as the rank of Π_2. That is, if identification fails we can separate identified and unidentified components of β as suggested by Phillips (1989) and Choi and Phillips (1992):

(23) $$H'\beta^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix}$$

and β_1^* is identified while β_2^* is unidentified.

Proof of Theorem 1

We first prove the following lemma.

Lemma 2. *Suppose that assumptions 1 and 2 and equation (15) hold then:*

(i) $M_{\hat{Q}\hat{\Pi}} = \Pi_{21}^{\perp} M_{\delta} \Pi_{21}^{\perp}{}' + o_p(1)$,

(ii) $\hat{\Pi}_{22}{}'\hat{Q} M_{\hat{Q}\hat{\Pi}_{21}} \hat{Q}\hat{\Pi}_{22} = T^{-1}\delta'\delta + o_p(T^{-1})$,

(iii) $\Pi_{21}^{\perp}{}'\hat{Q}\left(\hat{\pi} - \hat{\Pi}_{21}\beta_1^*\right) = T^{-1/2}\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1/2}\right)$,

(iv) $\hat{\Pi}_{22}{}'\hat{Q} M_{\hat{Q}\hat{\Pi}_{21}} \hat{Q}\hat{\pi}_1 = T^{-1}\delta'\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1}\right)$, *and*

(v) $\hat{\pi}_1{}'\hat{Q} M_{\hat{Q}\hat{\Pi}_{21}} \hat{Q}\hat{\pi}_1 = T^{-1}\left(z + \beta^{\perp} - W\beta_1^*\right)'\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1}\right)$

where $\delta \sim N\left(0, I_{k_2-n_1} \otimes I_{n_2}\right)$, $z \sim N\left(0, I_{k_2-n_1}\right)$ *and* $W \sim N\left(0, I_{k_2-n_1} \otimes I_{n_1}\right)$ *are independent.*

Proof of Lemma 2

Let $\hat{\Pi}_2$ be partitioned conformably to $\Pi_2 = \left[\Pi_{21}, 0\right]$ as $\hat{\Pi}_2 = \left[\hat{\Pi}_{21}, \hat{\Pi}_{22}\right]$. Assumption 1(c) implies that

(24) $$\left(\hat{\pi}_2, \hat{\Pi}_{21}, \hat{\Pi}_{22}\right) = \left(\pi_2, \Pi_{21}, 0\right) + T^{-1/2} Q^{-1}\left(x, X_1, X_2\right) + o_p\left(T^{-1/2}\right)$$

where $\left(x, X_1, X_2\right) \sim N\left(0, I_{k_2} \otimes I_{n+1}\right)$. Note that the mapping $Q\Pi_{21} \to M_{Q\Pi_{21}}$ is continuous (e.g. Forchini (2005)) so that $M_{\hat{Q}\hat{\Pi}_{21}} = M_{Q\Pi_{21}} + o_p(1)$ by the continuous mapping theorem. Moreover, $\hat{Q} = Q + o_p(1)$.

Write $M_{Q\Pi_{21}} = \Pi_{21}^{\perp}\Pi_{21}^{\perp}{}'$. Then, $w = \Pi_{21}^{\perp}{}'Q^{1/2}x \sim N\left(0, I_{k_2-n_1}\right)$, $\delta = \Pi_{21}^{\perp}{}'Q^{1/2}X_2 \sim N\left(0, I_{k_2-n_1} \otimes I_{n_2}\right)$ and $W = \Pi_{21}^{\perp}{}'Q^{1/2}X_1 \sim N\left(0, I_{k_2-n_1} \otimes I_{n_1}\right)$, are independent. Then, (*i*) can be proved as follows

$$M_{\hat{Q}\hat{\Pi}} = M_{\hat{Q}\hat{\Pi}_{21}} - M_{\hat{Q}\hat{\Pi}_{21}}\left(\hat{Q}\hat{\Pi}_{22}\right)\left[\left(\hat{Q}\hat{\Pi}_{22}\right)' M_{\hat{Q}\hat{\Pi}_{21}}\left(\hat{Q}\hat{\Pi}_{22}\right)\right]^{-1}\left(\hat{Q}\hat{\Pi}_{22}\right)' M_{\hat{Q}\hat{\Pi}_{21}}$$
$$= \Pi_{21}^{\perp}\Pi_{21}^{\perp}{}' + \Pi_{21}^{\perp}\delta\left[\delta'\delta\right]^{-1}\delta'\Pi_{21}^{\perp}{}' + o_p(1)$$
$$= \Pi_{21}^{\perp} M_{\delta}\Pi_{21}^{\perp}{}' + o_p(1).$$

Equation (*ii*) follows from (*i*)

$$\hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\Pi}_{22} = \left(T^{-1/2}Q^{-1/2}X_2\right)' QM_{Q\Pi_{21}}Q\left(T^{-1/2}Q^{-1/2}X_2\right) + o_p\left(T^{-1}\right)$$
$$= T^{-1}X_2'Q^{1/2}M_{Q\Pi_{21}}Q^{1/2}X_2 + o_p\left(T^{-1}\right)$$
$$= T^{-1}X_2'Q^{1/2}\Pi_{21}^{\perp}\Pi_{21}^{\perp}{}'Q^{1/2}X_2 + o_p\left(T^{-1}\right)$$
$$= T^{-1}\delta'\delta + o_p\left(T^{-1}\right).$$

To prove (*iii*) note that

$$\Pi_{21}^{\perp}{}'\hat{Q}\left(\hat{\pi} - \hat{\Pi}_{21}\beta_1^*\right) = T^{-1/2}\Pi_{21}^{\perp}{}'\left(Q^{1/2}x - Q^{1/2}X_1\beta_1^* + Q\Pi_{21}^{\perp}\beta^{\perp}\right) + o_p\left(T^{-1/2}\right)$$
$$= T^{-1/2}\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1/2}\right).$$

Finally,

$$\hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\pi}_1 = \hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*\right)$$
$$= T^{-1}X_2'Q^{1/2}\Pi_{21}^{\perp}\Pi_{21}^{\perp}{}'\left(Q^{1/2}x - Q^{1/2}X_1\beta_1^* + Q\Pi_{21}^{\perp}\beta^{\perp}\right) + o_p\left(T^{-1}\right)$$
$$= T^{-1}\delta'\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1}\right)$$

and

$$\hat{\pi}_1'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\pi}_1 = \left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*\right)'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*\right)$$
$$= T^{-1}\left(Q^{1/2}x - Q^{1/2}X_1\beta_1^* + Q\Pi_{21}^{\perp}\beta^{\perp}\right)'\Pi_{21}^{\perp}\Pi_{21}^{\perp}{}'\left(Q^{1/2}x - Q^{1/2}X_1\beta_1^* + Q\Pi_{21}^{\perp}\beta^{\perp}\right) + o_p\left(T^{-1}\right)$$
$$= T^{-1}\left(z + \beta^{\perp} - W\beta_1^*\right)'\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1}\right)$$

and Lemma 2 is proved.

We can now prove Theorem 1. Write

$$\hat{r}_2 = \left(\hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\Pi}_{22}\right)^{-1}\hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\pi}$$
$$\hat{\tau} = \frac{\left(\hat{Q}\hat{\pi}\right)' M_{\hat{Q}\hat{\Pi}}\left(\hat{Q}\hat{\pi}\right)}{1 + \hat{r}_1'\hat{r}_1}$$
$$\hat{B}_{TSLS} = \frac{T\hat{\tau}}{1 + \hat{r}_2'\hat{r}_2/\left(1 + \hat{r}_1'\hat{r}_1\right)},$$

where $\hat{r}_1$ and $\hat{r}_2$ are the TSLS estimators of β_1^* and β_2^* respectively. Note that $\hat{r}_1 = \beta_1^* + o_p(1)$ from the results of Choi and Phillips (1992). So using the notation of Lemma 2 we have that

$$
\begin{aligned}
\hat{r}_2 &= \left(\hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\Pi}_{22}\right)^{-1}\hat{\Pi}_{22}'\hat{Q}M_{\hat{Q}\hat{\Pi}_{21}}\hat{Q}\hat{\pi}_1 \\
&= \left(T^{-1}\delta'\delta + o_p\left(T^{-1}\right)\right)^{-1}\left(T^{-1}\delta'\left(z+\beta^{\perp}-W\beta_1^*\right)+o_p\left(T^{-1}\right)\right) \\
&= \left(\delta'\delta\right)^{-1}\delta'\left(z+\beta^{\perp}-W\beta_1^*\right)+o_p(1) \\
&= \left(\delta'\delta\right)^{-1}\delta'\left(z+\beta^{\perp},W\right)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix}+o_p(1) \\
&= \tilde{r}_2 + o_p(1),
\end{aligned}
$$

say, and

$$
\begin{aligned}
\hat{\tau} &= T^{-1}\begin{pmatrix}1 & -\beta_1^*\end{pmatrix}\left(z+\beta^{\perp},W\right)M_{\delta}\left(z+\beta^{\perp},W\right)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix}/\left(1+\beta_1^{*}{}'\beta_1^*\right)+o_p\left(T^{-1}\right) \\
&= T^{-1}\tau + o_p\left(T^{-1}\right).
\end{aligned} \tag{25}
$$

Note that

$$
\begin{aligned}
\left(z+\beta^{\perp},W\right)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix} &\sim N\left(\beta^{\perp},\left(1+\beta_1^{*}{}'\beta_1^*\right)I_{k_2-n_1}\right) \\
&\sim \left(1+\beta_1^{*}{}'\beta_1^*\right)^{1/2}N\left(\beta^{\perp}/\left(1+\beta_1^{*}{}'\beta_1^*\right)^{1/2},I_{k_2-n_1}\right)
\end{aligned}
$$

and this is independent of δ. So, conditioning on δ we have

$$
\begin{aligned}
& r_2 = \left(1+\beta_1^{*}{}'\beta_1^*\right)^{-1/2}\tilde{r}_2 \mid \delta \sim N\left(\left(\delta'\delta\right)^{-1}\delta'\beta^{\perp}/\left(1+\beta_1^{*}{}'\beta_1^*\right)^{1/2},\left(\delta'\delta\right)^{-1}\right) \\
& \tau \mid \delta \sim \chi^2\left(k_2-n,\beta^{\perp}{}'M_{\delta}\beta^{\perp}/\left(1+\beta_1^{*}{}'\beta_1^*\right)\right).
\end{aligned}
$$

Conditional independence of these two statistics follows from the fact that $M_{\delta}\delta = 0$.

Proofs of Corollaries 1.1 and 1.2

The joint density of (τ, r_2) is

$$pdf\left(\tau ,r_2\right)=\frac{\exp\left\{-\frac{m'm}{2}\right\}}{2^{\left(k_2-n\right)/2}\Gamma\left(\frac{k_2-n}{2}\right)\left(2\pi\right)^{n_2\left(k_2-n_1+1\right)/2}}\exp\left\{-\frac{\tau}{2}\right\}\tau^{\left(k_2-n\right)/2-1}$$

(26)
$$\int_{\mathbb{R}^{(k_2-n_1)n_2}} etr\left\{-\frac{1}{2}\left(I_{n_2}+r_2r_2'\right)\delta'\delta\right\}\left|\delta'\delta\right|^{1/2}$$

$$_0F_1\left(\frac{k_2-n}{2};\frac{\tau}{4}m'M_\delta m\right)\exp\left\{m'\delta r_2\right\}d\delta$$

where $m=\beta^{\perp}/\left(1+\beta_1^{*}{}'\beta_1^{*}\right)^{1/2}$. To evaluate the integral we transform δ as $\delta=VR^{1/2}$ where $R=\delta'\delta>0$ and $V=\delta\left(\delta'\delta\right)^{-1/2}$ satisfies $V'V=I_{n_2}$. The Jacobian of the transformation is $2^{-n_2}\left|R\right|^{\left(k_2-n_1\right)/2-\left(n_2+1\right)/2}$. We also note that the integral over $V'V=I_{n_2}$ is invariant to the transformation $R^{1/2}r_2m'\to R^{1/2}r_2m'H$ where H is an $(n_2\times n_2)$ orthogonal matrix. Then using Theorem 7.4.1 of Muirhead (1982) we have

$$\int_{H'H=HH'=I_{n_2}} etr\left\{R^{1/2}r_2m'H\right\}\left(\underline{dH}\right)={}_0F_1\left(\frac{n_2}{2};\frac{1}{4}r_2m'VV'mr_2'R\right),$$

where $\left(\underline{dH}\right)$ represent the standardized Haar measure on the group of $(n_2\times n_2)$ orthogonal matrices. Thus, the integral over $R>0$ can be evaluated using Theorem 7.3.4 of Muirhead (1982) to yield after some simplifications

$$pdf\left(\tau ,r_2\right)=\frac{\exp\left\{-\frac{m'm}{2}\right\}\Gamma_{n_2}\left(\frac{k_2-n_1+1}{2}\right)}{2^{n_2+\left(k_2-n\right)/2}\Gamma\left(\frac{k_2-n}{2}\right)\pi^{n_2\left(k_2-n_1+1\right)/2}}\frac{\exp\left\{-\frac{\tau}{2}\right\}\tau^{\left(k_2-n\right)/2-1}}{\left(1+r_2'r_2\right)^{\left(k_2-n_1+1\right)/2}}$$

$$\int_{V'V=I_{n_2}} {}_0F_1\left(\frac{k_2-n}{2};\frac{\tau}{4}m'\left(I_{k_2-n_1}-VV'\right)m\right)$$

$$_1F_1\left(\frac{k_2-n_1+1}{2};\frac{n_2}{2};\frac{1}{2}m'VV'm\frac{r_2'r_2}{1+r_2'r_2}\right)\left(dV\right).$$

The function $\Gamma_{n_2}\left(a\right)$ is defined in Theorem 2.1.12 of Muirhead (1982). We now let $b=\left(1+r_2'r_2\right)\tau$ and transform r_2 to polar coordinates $r_2=vq^{1/2}$ where $v'v=1$ and $q=r_2'r_2>0$ so that

$$pdf\left(b\right)=\frac{\exp\left\{-\frac{m'm}{2}\right\}\Gamma_{n_2}\left(\frac{k_2-n_1+1}{2}\right)}{2^{n_2+(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)\pi^{n_2(k_2-n_1)/2}\Gamma\left(\frac{n_2}{2}\right)}\exp\left\{-\frac{b}{2}\right\}b^{(k_2-n)/2-1}$$

(27)
$$\int_{q>0}\exp\left\{-\frac{qb}{2}\right\}q^{n_2/2-1}\left(1+q\right)^{-(n_2+1)/2}$$
$$\int_{V'V=I_{n_2}}{}_0F_1\left(\frac{k_2-n}{2};\frac{b(1+q)}{4}m'\left(I_{k_2-n_1}-VV'\right)m\right)$$
$${}_1F_1\left(\frac{k_2-n_1+1}{2};\frac{n_2}{2};\frac{1}{2}m'VV'm\frac{q}{1+q}\right)(dV)\,dq$$

where we used $\int_{v'v}(dv)=\frac{2\pi^{n_2/2}}{\Gamma(n_2/2)}$. Corollary 1.1 follows easily by setting $m=0$ in (27) and noting that

(28)
$$\int_{V'V=I_{n_2}}(dV)=\frac{2^{n_2}\pi^{(k_2-n_1)n_2/2}}{\Gamma_{n_2}\left(\frac{k_2-n_1}{2}\right)}$$

and

(29)
$$\int_{q>0}\exp\left\{-\frac{qb}{2}\right\}q^{n_2/2-1}\left(1+q\right)^{-(n_2+1)/2}dq=\Gamma\left(\frac{n_2}{2}\right)\Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{b}{2}\right).$$

To prove Corollary 1.2 we expand the hypergeometric functions in infinite series and integrate term by term. The integral over $q>0$ is similar to (29) and produces Tricomi confluent hypergeometric function

$$\Gamma\left(\frac{n_2}{2}\right)\left(\frac{n_2}{2}\right)_j\Psi\left(\frac{n_2}{2}+j;\frac{1}{2}+i;\frac{b}{2}\right).$$

The integral over $V'V=I_{n_2}$ is

(30)
$$\int_{V'V=I_{n_2}}\left[m'\left(I_{k_2-n_1}-VV'\right)m\right]^i\left(m'VV'm\right)^j(dV)=\sum_{s=0}^{i}\binom{i}{s}\left(m'm\right)^{i-s}(-1)^s$$
$$\int_{V'V=I_{n_2}}\left(m'VV'm\right)^{j+s}(dV).$$

We now interpret V as the matrix formed by the first n_2 columns on an $(k_2-n_1\times k_2-n_1)$ orthogonal matrix and write the integrand as a top-order zonal polynomial, so that by reformulating the integral over a standardized measure we have

$$\int_{V'V=I_{n_2}} \left(m'VV'm\right)^{j+s} \left(dV\right) = \frac{2^{n_2}\pi^{(k_2-n_1)n_2/2}}{\Gamma_{n_2}\left(\frac{k_2-n_1}{2}\right)} \int_{H'H=HH'=I_{k_2-n1}} C_{[j+s]}\left(mm'H\begin{pmatrix} I_{n_2} & 0 \\ 0 & 0 \end{pmatrix}H'\right)\left(\underline{dH}\right)$$

$$= \frac{2^{n_2}\pi^{(k_2-n_1)n_2/2}}{\Gamma_{n_2}\left(\frac{k_2-n_1}{2}\right)} \frac{C_{[j+s]}\left(mm'\right)C_{[j+s]}\left(\begin{pmatrix} I_{n_2} & 0 \\ 0 & 0 \end{pmatrix}\right)}{C_{[j+s]}\left(I_{k_2-n_1}\right)}$$

$$= \frac{2^{n_2}\pi^{(k_2-n_1)n_2/2}}{\Gamma_{n_2}\left(\frac{k_2-n_1}{2}\right)} \frac{\left(m'm\right)^{j+s}\left(n_2/2\right)_{j+s}}{\left(\left(k_2-n_1\right)/2\right)_{j+s}}.$$

Corollary 2.1 follows from noting that

$$\sum_{s=0}^{i}\binom{i}{s}(-1)^s \frac{\left(n_2/2\right)_{j+s}}{\left(\left(k_2-n_1\right)/2\right)_{j+s}} = \frac{\left(n_2/2\right)_j\left(\left(k_2-n\right)/2\right)_i}{\left(\left(k_2-n_1\right)/2\right)_{j+i}}.$$

To prove the second part of Corollary 1.1 we need the following lemma.

Lemma 3. *For* $\delta > 0$, $b > 0$ *and* $\gamma \neq 0$ *we have*

$$\int_{0<t<b} \exp\left\{-t/2\right\} t^{\delta-1}\,{}_1F_1\left(\alpha;\gamma;t/2\right)dt = \left(b^a/a\right){}_2F_2\left(\gamma-\alpha,\delta;\delta+1,\gamma;-b/2\right).$$

Proof of Lemma 3

Using Kummer transformation we can write the integrand above as

$$t^{\delta-1}\,{}_1F_1\left(\gamma-\alpha;\gamma;-t/2\right).$$

We let $t = bx$, so that the desired integral becomes

$$\int_{0<x<b} t^{\delta-1}\,{}_1F_1\left(\gamma-\alpha;\gamma;-t/2\right)dt\,.$$

We can now expand the hypergeometric function as a power series and integrate term by term to obtain

$$\int_{0<x<b} t^{\delta-1}\,{}_1F_1\left(\gamma-\alpha;\gamma;-t/2\right)dt = \sum_{i=0}^{\infty}\frac{\left(\gamma-\alpha\right)_i}{i!\left(\gamma\right)_i}\left(-\frac{1}{2}\right)^i \int_{0<x<b} t^{\delta+i-1}dt$$

$$= \sum_{i=0}^{\infty}\frac{\left(\gamma-\alpha\right)_i}{i!\left(\gamma\right)_i}\left(-\frac{1}{2}\right)^i\left(\frac{b^{\delta+i}}{\delta+i}\right)$$

$$= \frac{b^a}{a}\sum_{i=0}^{\infty}\frac{\left(\gamma-\alpha\right)_i\left(\delta\right)_i}{i!\left(\gamma\right)_i\left(\delta+1\right)_i}\left(-\frac{b}{2}\right)^i.$$

The results stated in Lemma 3 follows.

We can now focus on equation (21). Note that

(31)

$$CDF(b)=\frac{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i}{2}\right)}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i+1}{2}\right)}\int_{0<t<b} e^{-t/2}t^{(k_2-n)/2-1}\Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{t}{2}\right)dt\ .$$

Using equation (9.10.3) of Slater (1960) we can write

$$\Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{t}{2}\right)=\frac{\sqrt{\pi}}{\Gamma\left(\frac{n_2+1}{2}\right)}\,{}_1F_1\left(\frac{n_2}{2};\frac{1}{2};\frac{b}{2}\right)-\frac{\sqrt{2\pi b}}{\Gamma\left(\frac{n_2}{2}\right)}\,{}_1F_1\left(\frac{n_2+1}{2};\frac{3}{2};\frac{b}{2}\right).$$

So, using Lemma 3 we have

$$\begin{aligned}\int_{0<t<b} e^{-t/2}t^{\frac{k_2-n}{2}-1}\Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{t}{2}\right)dt &= \frac{\sqrt{\pi}}{\Gamma\left(\frac{n_2+1}{2}\right)}\int_{0<t<b} e^{-t/2}t^{(k_2-n)/2-1}\,{}_1F_1\left(\frac{n_2}{2};\frac{1}{2};\frac{b}{2}\right)dt\\ &\quad -\frac{\sqrt{2\pi}}{\Gamma\left(\frac{n_2}{2}\right)}\int_{0<t<b} e^{-t/2}t^{(k_2-n+1)/2-1}\,{}_1F_1\left(\frac{n_2+1}{2};\frac{3}{2};\frac{b}{2}\right)dt\\ &= \frac{\sqrt{\pi}}{\Gamma\left(\frac{n_2+1}{2}\right)}\frac{2b^{\frac{k_2-n}{2}}}{k_2-n}\,{}_2F_2\left(\frac{1-n_2}{2},\frac{k_2-n}{2};\frac{k_2-n}{2}+1,\frac{1}{2};-\frac{b}{2}\right)\\ &\quad -\frac{\sqrt{2\pi}}{\Gamma\left(\frac{n_2}{2}\right)}\frac{2b^{\frac{k_2-n+1}{2}}}{k_2-n+1}\,{}_2F_2\left(\frac{2-n_2}{2},\frac{k_2-n+1}{2};\frac{k_2-n+1}{2}+1,\frac{3}{2};-\frac{b}{2}\right).\end{aligned}$$

Inserting this in (31) we obtain the desired result.

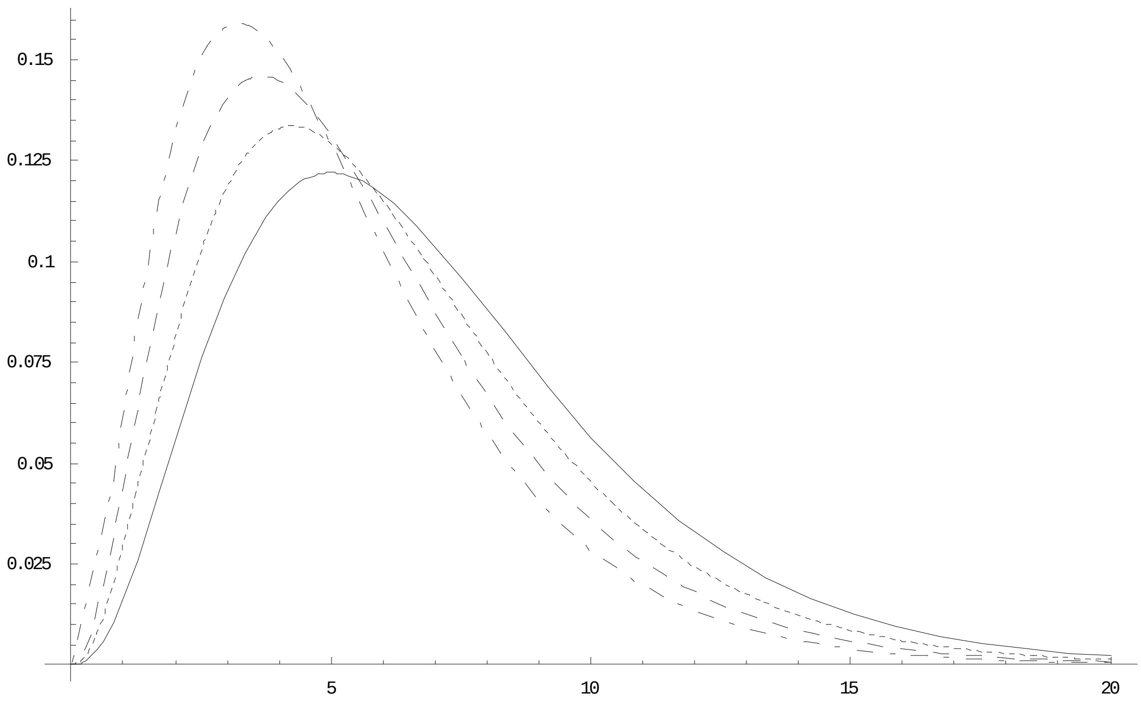

Figure 1: Asymptotic density of $\hat{B}$ for $k_2 = 10$, $n = 3$ and $n_2 = 0$ (solid line), $n_2 = 1$ (dotted line), $n_2 = 2$ (dashed line) and $n_2 = 3$ (dotted-dashed line).

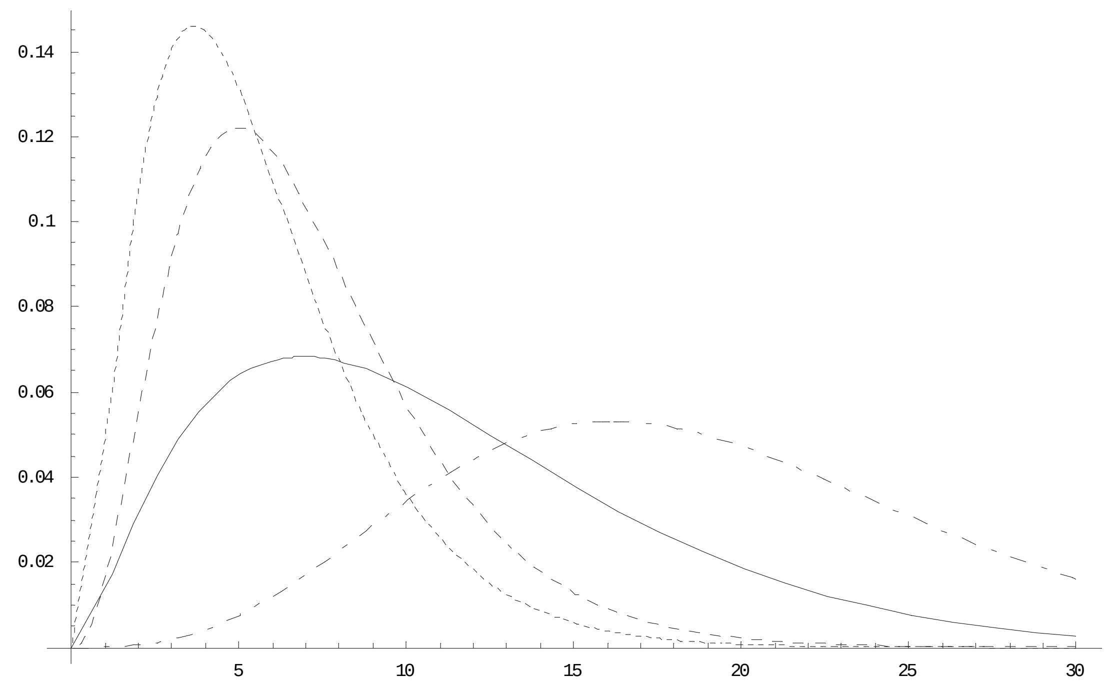

Figure 2: Comparison of the density of a chi-squared distribution with $k_2 - n$ degrees of freedom (dashed line), the density of a non-central chi-squared distribution with $k_2 - n$ degrees of freedom (dotted-dashed line) and $\lambda = 12$, the asymptotic density of $\hat{\mathrm{B}}$ given in Corollary 1.1 (dotted line), and the asymptotic density of $\hat{\mathrm{B}}$ given in Corollary 1.2 (solid line) for $k_2 = 10$, $n = 3$, $n_2 = 2$ and $\lambda = 12$.

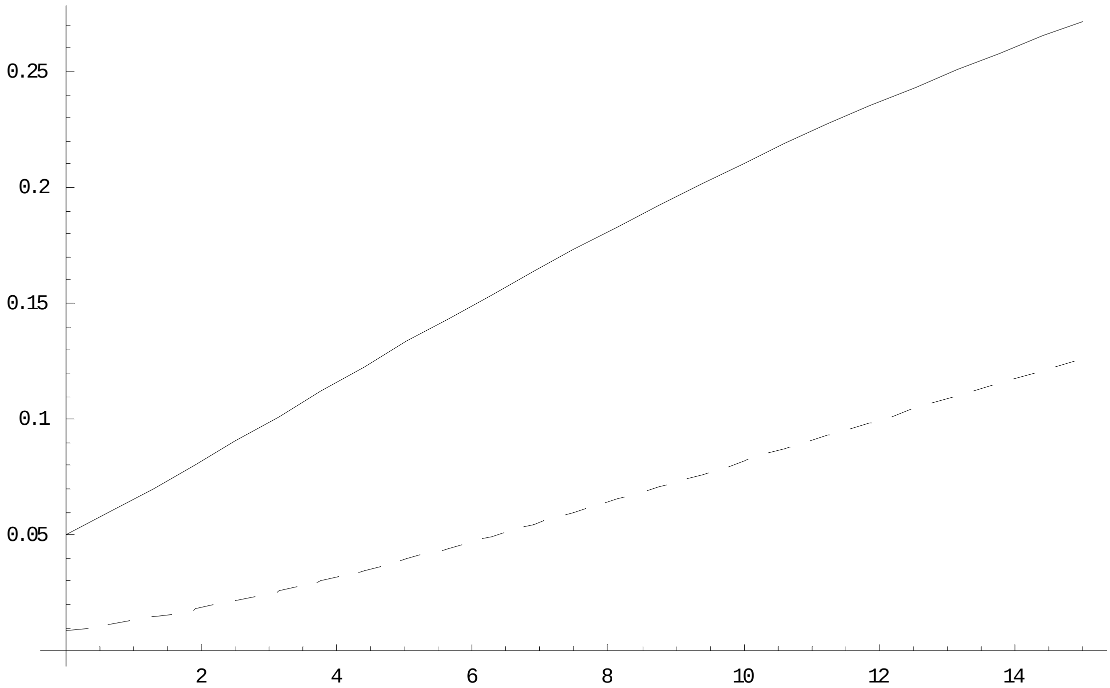

Figure 3: Asymptotic power for the case where the critical values are from a chi-squared distribution with $k_2 - n$ degrees of freedom (dashed line) and from the distribution in Corollary 1.1 (solid line) for $k_2 = 10$, $n = 4$, $n_2 = 4$ and $\lambda \in [0,15]$.

			k_2=5	k_2=10	k_2=20	k_2=40	k_2=80
n=1	n_2=	*0*	5.00	5.00	5.00	5.00	5.00
		1	2.91	3.33	3.70	4.02	4.27
n=2	n_2=	*0*	5.00	5.00	5.00	5.00	5.00
		1	2.77	3.26	3.68	4.01	4.27
		2	1.62	2.16	2.71	3.21	3.64
n=3	n_2=	*0*	5.00	5.00	5.00	5.00	5.00
		1	2.59	3.19	3.65	4.00	4.26
		2	1.47	2.08	2.67	3.19	3.63
		3	0.88	1.38	1.96	2.54	3.08
n=4	n_2=	*0*	5.00	5.00	5.00	5.00	5.00
		1	2.37	3.11	3.62	3.99	4.26
		2	1.30	1.99	2.63	3.17	3.62
		3	0.77	1.29	1.91	2.52	3.07
		4	0.49	0.86	1.40	2.00	2.60

Table 1: Asymptotic size of Byron's test (in %) using a nominal 5% level from $\chi^2(k_2 - n)$

n_2	n_2 estimated	n_2 known	$\chi^2(k_2 - n)$
		T=25	
0	29.61	23.19	23.19
1	23.59	21.12	16.42
2	19.52	19.51	11.70
3	18.80	18.80	8.94
4	18.16	18.16	6.56
		T=50	
0	15.31	12.85	12.85
1	11.21	11.21	7.89
2	10.30	10.31	5.12
3	10.27	10.27	3.35
4	9.86	9.86	2.51
		T=100	
0	9.38	9.08	9.08
1	7.61	7.64	4.89
2	7.34	7.36	3.05
3	7.23	7.24	1.94
4	7.09	7.10	1.34
		T=400	
0	6.08	6.08	6.08
1	5.51	5.44	3.32
2	5.43	5.44	2.04
3	5.55	5.56	1.24
4	5.40	5.42	0.76
		T=1600	
0	5.23	5.23	5.23
1	5.25	5.26	3.07
2	5.13	5.13	1.84
3	4.87	4.89	1.20
4	5.06	5.08	0.71
		T=6400	
0	5.15	5.15	5.15
1	4.91	4.91	2.83
2	5.02	5.04	1.82
3	5.07	5.09	1.17
4	5.07	5.08	0.77
		Asymptotics	
0	5.00	5.00	5.00
1	5.00	5.00	2.91
2	5.00	5.00	1.76
3	5.00	5.00	1.10
4	5.00	5.00	0.71

Table 2: Size of Byron's test (in %) using a nominal 5% value, $k_2 = 8$.

References

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

Basmann, R. L. (1960b), "On the Asymptotic Distribution of Generalised Linear Estimators", *Econometrica* 28, 97-107.

Byron, R. P. (1974), "Testing Structural Specification Using the Unrestricted Reduced Form", *Econometrica* 42, 869-883.

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

Cragg, J. G. and S. G. Donald (1996), "Testing Overidentifying Restrictions in Unidentified Models", *Unpublished UBC discussion paper* 96/20.

Forchini, G. (2005), "A Note on the Continuity of Projection Matrices with Application to the Asymptotic Distribution of Quadratic Forms", *Applicationes Mathematicae* 32, 51-55.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica* 40, 1029-1054.

Hausman, J. A. (1983), "Specification and Estimation of Simultaneous Equation Models", *Handbook of Econometrics, Volume*, M. D. Intriligator, Amsterdam, North-Holland Publishing Company**,** 391-448.

Hillier, G. H. (1985), "On the Joint and Marginal Densities of Instrumental Variable Estimators in a General Structural Equation", *Econometric Theory* 1, 53-72.

Hillier, G. H. (1990), "On the Normalization of Structural Equations: Properties of Direction Estimators", *Econometrica* 58, 1181-1194.

Hillier, G. H. (2006), "Yet More on The Exact Properties of IV Estimators", *EconometricTheory, Forthcoming*.

Lebedev, N. N. (1972), *Special Functions and Their Applications* New York, Dover Publications Inc.

Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory* New York, John Wiley and Sons, Inc.

Phillips, P. C. B. (1983), "Exact Small Sample Theory in the Simultaneous Equation Model", *Handbook of Econometrics,* Z. Griliches, Amsterdam, North Holland**,** 449-516.

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

Robin, J.-M. and R. J. Smith (2000), "Tests of Rank", *Econometric Theory* 16, 151-176.

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

Sargan, J. D. (1983), "Identification and Lack of Identification", *Econometrica* 51, 1605-1633.

Sims, C. A. (1980), "Macroeconomics and Reality", *Econometrica* 48, 1-48.

Slater, L. J. (1960), *Confluent Hypergeometric Functions* London, Cambridge University Press.

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