The Asymptotic Distribution of Tests for Over-identification in Partially Identified Linear Structural Equations

by

G. Forchini[1][2]

Monash University

November 12, 2007

Abstract

We study the asymptotic distribution of the statistics suggested by Sargan/Byron/Wegge and Basmann for testing for over-identification in partially identified linear structural equations and derive closed form expressions for their asymptotic densities. These allow us to understand the asymptotic behaviour of these statistics in cases where identification of the structural parameters fails. We also consider variants of these statistics using estimators of the structural variance based on the LIML estimator and show that they have unexpected properties.

[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, Chris Skeels and Katharina Hauck for helpful comments and discussions. This research was partially supported by Australian Research Council grant DP0771445.

1. Introduction

Considerable attention has been recently 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 lack of identification for the parameters of econometric models on the properties of test statistics and estimators.

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), Wegge (1978), Hwang (1980) and Hansen (1982) under the assumption that all the structural parameters are identified. The robustness of these tests to identification failures has been investigated by Cragg and Donald (1996)[3]. They argue that commonly used tests for over-identification may lead to misleading inference when identification of the structural parameters fails. Precisely, they show that lack of identification tends to concentrate the probability mass of a test statistic for over-identification around zero implying that such tests tend to suggest a correct specification more often than one would expect for the nominal size of the test.

This paper further investigates the asymptotic distribution of tests of over-identification in partially identified linear structural equations and contributes to the literature in several ways. First, by refining results of Cragg and Donald (1996), we derive an asymptotic representation of the statistics suggested by Sargan (1958), Basmann (1960a), Basmann (1960b), Byron (1974) and Wegge (1978), that holds when the over-identifying conditions fail in a partially identified structural equation (e.g. Phillips (1989) and Choi and Phillips (1992)). This is used to find exact expressions for the asymptotic distributions of these statistics that can be used, for example, to calculate asymptotic critical values under partial identification. Our results do not apply in the presence of weak instruments however in such a case they yield bounds for the p-values and the critical values of tests of over-identification.

Second, we find that under the null hypothesis of a correctly specified model, the asymptotic density depends only on the rank of the matrix of correlations between the instruments and the endogenous variables included as regressors. If this is unknown, it can be consistently estimated. We propose a procedure to consistently test for over-identifying

[3] Notice that Cragg and Donald (1996) use a different terminology. Our Sargan/Byron/Wegge test is their Basmann test and our Basmann test is their Byron test. A short document available at http://www-personal.buseco.monash.edu.au/~forchini/ reviews the relationship between the test statistics suggested by Sargan (1958), Basmann (1960a), Basmann (1960b), Byron (1974) and Wegge (1978).

restrictions that has the correct asymptotic size in models for which the structural parameters are only partially unidentified. 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 themselves; and (ii) we do not need to establish which structural parameters are identified and which ones are not.

Third, a variant of the asymptotic distribution of the Sargan, Basmann, Byron and Wegge test statistics where the structural variance is estimated using the limited information maximum likelihood (LIML) estimator is considered. This modification has non-trivial consequences on the asymptotic distributions of these statistics, and suggests that one should be very careful in modifying standard test statistics when identification may fails.

Fourth, our work shows that the problem of deriving the asymptotic distributions of tests for over-identifying restriction using a minimal distance approach can be considerably simplified using simple invariance arguments.

The paper is structured as follows. Section 2 sets up the model, briefly discusses identification and over-identification, and motivates tests for over-identification in the presence of partial identification. Section 3 formulates the testing problem and lists the assumptions used. Section 4 investigates the asymptotic properties of the Sargan, Basmann and Byron test statistics. Some numerical results are presented and discussed in Section 5. Section 6 concludes. Proofs are in the appendix.

2. The model, over-identification and identification

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 maps on the space of distributions of $\left[y_1, Y_2\right]$ given $\left[Z_1, Z_2\right]$ to an Euclidean space, 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)). Notice that (4) does not imply nor is implied by the fact that Π_2 has rank n, that is (4) states that β is defined in terms of the reduced form parameters, but does not imply that β is uniquely defined by them.

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)). We have implicitly assumed that the model is informative about Φ, Π and Ω only, and consider the case where exclusion restrictions only are allowed to achieve identification of the structural parameters. Although this is a very important case, identification of the structural parameters can be achieved in other ways (e.g. Hsiao (1983)).

Having set up the problem we now discuss two possible reasons why one should be interested in testing for over-identification when identification of the structural parameters fails.

2.1. Tests for over-identification as misspecification tests

The over-identifying restrictions (4) are compatibility conditions between the structural and the reduced form equation, therefore a test for over-identification is a test for model misspecification. Although proposition 1 shows that identification of the structural parameters relies on the simultaneous conditions that equation (4) holds and Π_2 has rank n, we will now argue that it is unreasonable to assume that Π_2 has rank n when testing for over-identification.

First, over-identifying restrictions imply the compatibility between structural and reduced form equations. As such, they define the structural parameters in terms of the reduced form parameters, and set the context in which identification may or may not occur. For instance, equation (4) defines β as the vector of coefficients for which π_1 is a linear combination of the columns of Π_2. The definition of the interest parameter logically precedes any statement about its uniqueness (i.e. identification). This implies that the validity of the over-identifying restrictions must be established before identification of a structural equation is investigated.

Second, in practical applications, the correct model (1) is unknown to the researcher, and the instruments need to be chosen. Different sets of instruments imply different matrices $[\pi_1, \Pi_2]$, and, thus, they implicitly define different parameter vectors β satisfying the over-identifying restrictions in (4). Even if we are willing to assume that the structural parameter β is identified in the true but unknown structural model, there is no reason to assume that this is the case for every selection of instruments that we may consider (i.e. for every matrix $[\pi_1, \Pi_2]$).

Third, it is very important to distinguish between violations of identification of the structural parameters and over-identification because they have very different consequences for the applied researcher. If over-identification fails, the model is misspecified and the researcher can try to improve upon it. However, if identification fails, the model is well specified but uninformative about the parameters of interest so that nothing can be done to improve inference. Existing tests for over-identification are based on the assumption that under the null hypothesis both identification and over-identification hold. As such, these tests are based on critical values from a chi-squared distribution with $k_2 - n$ degrees of freedom and may suggest that a model containing irrelevant instruments is well specified even if it is not. If this is the case, the researcher may conclude that the model is partially identified when in fact it is misspecified (cf. Murray (2006)).

2.2. Tests for over-identification as tests for linear restrictions

Tests of over-identification are closely related to tests of linear restrictions on the structural parameters (e.g. Dhrymes (1969) and Kadane (1974)). Assuming, without loss of generality, that one wishes to test $H_0 : \beta = 0$ in (1), the Anderson-Rubin test (Anderson and Rubin (1949) and Anderson and Rubin (1950)) is a special case of tests for over-identifying restrictions (set $M_{(\hat{Q}\hat{\Pi}_2)} = I_{k_2}$ in equation (10) below to obtain the Anderson-Rubin test statistic).

Equation (1) can be regarded as the structural equation

$$y_1 = Y_2\beta + Y_3\delta + Z_1\gamma + u \tag{7}$$

with corresponding reduced form

$$\left[y_1, Y_2, Y_3\right] = Z_1\left[\phi_1, \Phi_2, \Phi_3\right] + Z_2\left[\pi_1, \Pi_2, \Pi_3\right] + \left[v_1, V_2, V_3\right], \tag{8}$$

under the hypothesis $H_0 : \delta = 0$ (the case $H_0 : \delta = \delta_0 \neq 0$ can be reduced to the previous case by suitable transformations of the structural and the reduced forms). In this case one must assume that the structural coefficients in δ are (weakly) identified but may wish to allow some of the coefficients in β to be unidentified in the spirit of Phillips (1989) and Choi and Phillips (1992). Thus, over-identification tests allow a researcher to perform tests on (possibly weakly) identified parameters without imposing restrictions on the identification of the remaining ones.

Clearly, in order, to interpret a rejection of the null hypothesis in a test for over-identifying restriction in (1) as rejection of $\delta = 0$ we need to be sure that (7) is well specified using a test for over-identification. One may use Bonferroni inequality to control the overall size of the test.

3. Test statistics and assumptions

Equation (4) can be written in the equivalent form

$$\left(P\pi_1\right)' M_{(P\Pi_2)} \left(P\pi_1\right) = 0\,, \tag{9}$$

where P 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 (9) holds against the alternative that it does not. A minimum distance test for the validity of (9) (or equivalently (4)) can be based on

$$(10) \qquad \left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{(\hat{Q}^{1/2}\hat{\Pi}_2)} \left(\hat{Q}^{1/2}\hat{\pi}_1\right)$$

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 (14) and $\hat{Q}$ can be chosen as $\hat{Q} = T^{-1}Z_2{}'M_{Z_1}Z_2$. This justifies the use of a statistic having the asymptotic form

$$(11) \qquad B = \frac{T\left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{(\hat{Q}^{1/2}\hat{\Pi}_2)} \left(\hat{Q}^{1/2}\hat{\pi}_1\right)}{\sigma_u^2}$$

where $\sigma_u^2 = (1, -\beta')\Omega(1, -\beta')' = \left(1 + \beta^*{}'\beta^*\right)\omega_{11.2}$ is the variance of the error in the structural equation, β^* denotes the *canonical* coefficients of the endogenous variables

$$(12) \qquad \beta^* = \left(\Omega_{22}^{1/2}\beta - \Omega_{22}^{-1/2}\omega_{21}\right) / \omega_{11.2}^{1/2}$$

in the structural equation (e.g. Phillips (1983)), $\omega_{11.2} = \omega_{11} - \omega_{21}{}'\Omega_{22}^{-1}\omega_{21}$ and

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

is partitioned conformably to $\left[y_1, Y_2\right]$. Under standard assumptions $B \to^p \chi^2\left(k_2 - n\right)$ when the over-identifying restrictions hold, irrespective of identification (e.g. Lemma 2 of Cragg and Donald (1996)).

The test for over-identifying restrictions of Sargan (1958), Byron (1974) and Wegge (1978) replaces the denominator of (11) with $\hat{\sigma}_u^2 = \hat{u}_{TSLS}{}'\hat{u}_{TSLS} / T$ where $\hat{u}_{TSLS}$ is the vector of TSLS residuals in (1). Basmann (1960a) and Basmann (1960b) uses a slightly different estimator or the structural variance of the form

$$(13) \qquad \begin{aligned} \hat{\sigma}_u^2 &= \hat{u}_{TSLS}{}'M_Z\hat{u}_{TSLS} / T \\ &= \hat{u}_{TSLS}{}'\hat{u}_{TSLS} / T - \left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{(\hat{Q}^{1/2}\hat{\Pi}_2)} \left(\hat{Q}^{1/2}\hat{\pi}_1\right) \\ &= \left(1, -\hat{\beta}_{TSLS}{}'\right)\hat{\Omega}\left(1, -\hat{\beta}_{TSLS}{}'\right)' \end{aligned}$$

where $\hat{\Omega} = \left[y_1, Y_2\right]' M_{[Z_1, Z_2]} \left[y_1, Y_2\right] / T$. If the over-identifying hold then $\left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{(\hat{Q}^{1/2}\hat{\Pi}_2)} \left(\hat{Q}^{1/2}\hat{\pi}_1\right)$ converges in probability to zero, and Basmann and Sargan/Byron/Wegge test statistics are asymptotically equivalent (e.g. Hwang (1980)). Notice that both estimators of the structural variance are consistent if and only if $\hat{\beta}_{TSLS}$ is consistent.

Clearly, one could use a different estimator of β in estimating σ_u^2, and an obvious choice is the LIML estimator $\hat{\beta}_{LIML}$. This version of the Sargan/Byron/Wegge statistic uses $\hat{\sigma}_u^2 = \hat{u}_{LIML}'\hat{u}_{LIML} / T$ where $\hat{u}_{LIML}$ is the vector of LIML residuals in (1). Similarly, the LIML version of Basmann test statistic $\hat{\sigma}_u^2 = \left(1, -\hat{\beta}_{LIML}'\right)\hat{\Omega}\left(1, -\hat{\beta}_{LIML}'\right)'$ is close to but it is not the same thing as the likelihood ratio statistic for testing for over-identifying restriction suggested by Anderson and Rubin (1949) and Anderson and Rubin (1950). We will see later on that this modification has non-trivial consequences on the asymptotic properties of the Sargan/Byron/Wegge and Basmann test statistics.

In order to derive the asymptotic distribution of test statistics based on (11) 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 $\Pi = [\pi_1, \Pi_2]$,

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

satisfies

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

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 because the problem of testing the null hypothesis that (9) holds against the alternative that it does not and the statistic (11) 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*

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

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

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

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) $\Pi_2 = [\Pi_{21}, 0]$ *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.*

Lemma 1 shows that there is no loss of generality in assuming that the structural form is in canonical form, and that the rotations of coordinates in the space of endogenous variables to separate identified and unidentified parameters have already been carried out as in Phillips (1989) and Choi and Phillips (1992).

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 make one further assumption.

Assumption 3. $\pi_1 = \Pi_{21}\beta_1^* + T^{-1/2}Q^{1/2}\Pi_{21}^{\perp}\beta^{\perp}$ *where* $\left(\Pi_{21}^{\perp}\right)'Q\Pi_{21}^{\perp} = I_{k_2 - n_1}$ *and* $\left(\Pi_{21}\right)'Q\Pi_{21}^{\perp} = 0$.

Notice that Assumption 3 has already been formulated in terms of the canonical model, and that the value of β_1^* is arbitrary so that there is no loss of generality in assuming $\left(\Pi_{21}\right)'Q\Pi_{21}^{\perp} = 0$. Assumption 3 allows for the case where the matrix $\Pi_{21}^{\perp}$ is fixed and the omitted structural parameters $T^{-1/2}\beta^{\perp}$ are in an $O\left(T^{-1/2}\right)$ neighbourhood of zero, or $\beta^{\perp}$ is fixed and the reduced form coefficients associated to it are of the form $T^{-1/2}\Pi_{21}^{\perp}$ (weak identification).

4. Asymptotic properties of Sargan/Byron/Wegge and Basmann and test statistics

In this section, we study the asymptotic properties of the Sargan/Byron/Wegge and Basmann test statistics. The first result establishes the asymptotic equivalence of the Sargan/Byron/Wegge and Basman estimators of σ_u^2.

Lemma 2. (i) $T^{-1}\hat{u}_{LIML}'\hat{u}_{LIML} \geq \left(1,-\hat{\beta}_{LIML}'\right)\hat{\Omega}\left(1,-\hat{\beta}_{LIML}'\right)'$.

(ii) $T^{-1}\hat{u}_{TSLS}'\hat{u}_{TSLS} \geq \left(1,-\hat{\beta}_{TSLS}'\right)\hat{\Omega}\left(1,-\hat{\beta}_{TSLS}'\right)'$.

If *Assumptions 1, 2 and 3 hold, then,*

(iii) $T^{-1}\hat{u}_{LIML}'\hat{u}_{LIML} = \left(1,-\hat{\beta}_{LIML}'\right)\hat{\Omega}\left(1,-\hat{\beta}_{LIML}'\right)' + o_p(1)$ *and*

(iv) $T^{-1}\hat{u}_{TSLS}'\hat{u}_{TSLS} = \left(1,-\hat{\beta}_{TSLS}'\right)\hat{\Omega}\left(1,-\hat{\beta}_{TSLS}'\right)' + o_p(1)$

Lemma 2 (i) and (ii) imply that in finite samples the Sargan/Byron/Wegge statistic is always smaller than Basmann statistic (c.f. Theorem 4 of Cragg and Donald (1996)) and the fact that $B \rightarrow^p \chi^2\left(k_2 - n\right)$ imply that the Sargan/Byron/Wegge and Basmann versions of the test statistics are asymptotically equivalent in all situations considered in this paper. Therefore, we will denote both statistics based on the TSLS estimator of σ_u^2 by $\hat{B}_{TSLS}$ and those based on the LIML estimator of σ_u^2 by $\hat{B}_{LIML}$.

Theorem 1. *Let* $\hat{B}_{TSLS} = T\left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{\left(\hat{Q}^{1/2}\hat{\Pi}_2\right)}\left(\hat{Q}^{1/2}\hat{\pi}_1\right)/\hat{\sigma}_u^2$ *where either* $\hat{\sigma}_u^2 = \hat{u}_{TSLS}'\hat{u}_{TSLS}/T$ *or* $\hat{\sigma}_u^2 = \hat{u}_{TSLS}'M_Z\hat{u}_{TSLS}/T$ *or* $\hat{\sigma}_u^2 = \left(1,-\hat{\beta}_{TSLS}'\right)\hat{\Omega}\left(1,-\hat{\beta}_{TSLS}'\right)'$ *and suppose that Assumptions 1, 2 and 3 hold. Then*

(18) $$\hat{B}_{TSLS} \rightarrow^d b = \frac{\tau}{1 + r_2'r_2}$$

where

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

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

and

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

Moreover, r_2 *and* τ *are independent conditional on* δ.

Theorem 1 gives an explicit asymptotic representation for the distribution of $\hat{B}_{TSLS}$. Several known results can be obtained as special cases. If the model is identified $\hat{B}_{TSLS} \to^{d} \chi^2(k_2 - n)$ (e.g. Sargan (1958), Basmann (1960b), Basmann (1960b), Byron (1974) and Cragg and Donald (1996)). For local failure of the compatibility condition (4) but with rank of Π_2 equal to n, $\hat{B}_{TSLS} \to^{d} \chi^2\left(k_2 - n, \beta^{\perp}{}'\beta^{\perp} / \left(1+\beta_1^{*}{}'\beta_1^{*}\right)\right)$, indicating that the test is asymptotically unbiased and consistent (e.g. Cragg and Donald (1996)).

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 it follows easily that $\Pr\left\{\hat{B}_{TSLS} \geq c\right\} \leq \int_0^c d\chi^2(k_2 - n)$ (c.f. Theorem 4 of Cragg and Donald (1996)). The following corollary shows that in such a case $\hat{B}_{TSLS}$ has an asymptotic non-standard non-central chi-squared distribution.

Corollary 1.1 *Suppose that Assumptions 1, 2 and 3 hold. If $\beta^{\perp} = 0$ and the rank of Π_2 equals $n_1 < n = n_1 + n_2$ then the asymptotic density function of $\hat{B}_{TSLS}$ is*

$$pdf_{\hat{B}_{TSLS}}(b) = \frac{e^{-b/2} b^{(k_2-n)/2-1}}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)} \times \frac{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i}{2}+1\right)}{\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i+1}{2}\right)} \Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{b}{2}\right), \tag{22}$$

where Ψ denotes a Tricomi confluent hypergeometric function (e.g. Lebedev (1972)). Moreover, its asymptotic distribution is

$$\begin{aligned} CDF_{\hat{B}_{TSLS}}(b) = {} & \frac{2\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i}{2}\right)}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)\prod_{i=1}^{n_2}\Gamma\left(\frac{k_2-n_1-i+1}{2}\right)} \times \\ & \left\{ \frac{\sqrt{\pi}}{\Gamma\left(\frac{n_2+1}{2}\right)} \frac{b^{\frac{k_2-n}{2}}}{k_2-n} {}_2F_2\left(\frac{1-n_2}{2}, \frac{k_2-n}{2}; \frac{k_2-n}{2}+1, \frac{1}{2}; -\frac{b}{2}\right) \right. \\ & \left. - \frac{\sqrt{2\pi}}{\Gamma\left(\frac{n_2}{2}\right)} \frac{b^{\frac{k_2-n+1}{2}}}{k_2-n+1} {}_2F_2\left(\frac{2-n_2}{2}, \frac{k_2-n+1}{2}; \frac{k_2-n+1}{2}+1, \frac{3}{2}; -\frac{b}{2}\right) \right\} \end{aligned} \tag{23}$$

where ${}_2F_2(a_1, a_2; b_1, b_2; x)$ is defined by (e.g. Lebedev (1972)),

$$ {}_2F_2\left(a_1,a_2;b_1,b_2;x\right)=\sum_{j=0}^{\infty}\frac{\left(a_1\right)_j\left(a_2\right)_j}{j!\left(b_1\right)_j\left(b_2\right)_j}x^j\ , $$

and $\left(a\right)_j=\Gamma\left(a+j\right)/\Gamma\left(a\right)$.

If we know n_1, equation (23) allows us to find the correct asymptotic p-values for the Sargan/Byron/Wegge and Basmann tests. 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, Sargan/Byron/Wegge or Basmann test statistics can be calculated and their p-values can obtained from (23) with n_1 replaced by $\hat{n}_1$.

The procedure above 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 Assumption 3 it is difficult to disentangle the two effects as the following result shows.

Corollary 1.2 *Suppose that Assumptions 1, 2 and 3 hold. If the rank of* Π_2 *equals* $n_1<n=n_1+n_2$ *then the asymptotic density function of* $\hat{B}_{TSLS}$ *is*

(24)
$$ \begin{aligned} pdf_{\hat{B}_{TSLS}}\left(b\right)=&\frac{e^{-b/2}b^{\left(k_2-n\right)/2-1}}{2^{\left(k_2-n\right)/2}\Gamma\left(\dfrac{k_2-n}{2}\right)}\\ &\times\frac{\prod_{i=1}^{n_2}\Gamma\left(\dfrac{k_2-n_1-i}{2}+1\right)}{\prod_{i=1}^{n_2}\Gamma\left(\dfrac{k_2-n_1-i+1}{2}\right)}\exp\left\{-\frac{1}{2}\lambda\right\}\\ &\times\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\frac{\left(\dfrac{k_2-n_1+1}{2}\right)_j\left(\dfrac{n_2}{2}\right)_j}{i!\,j!\left(\dfrac{k_2-n_1}{2}\right)_{j+i}}\left[\frac{1}{2}\lambda\right]^{i+j}\left(\frac{b}{2}\right)^i\Psi\left(\frac{n_2}{2}+j;\frac{1}{2}+i;\frac{b}{2}\right) \end{aligned} $$

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

Corollary 1.1 together with Proposition 9 of Cragg and Donald (1996) provides bounds for the asymptotic p-values of the tests of Sargan/Byron/Wegge and Basmann under weak instruments (in the sense that $\Pi_2 = O\left(T^{-1/2}\right)$, since these will always be between $1 - CDF_{\hat{B}_{TSLS}}(b)$ calculated with $n_1 = 0$ and $\Pr\left\{\chi^2(k_2 - n) > b\right\}$. Precisely,

Corollary 1.3 *Suppose that Assumptions 1, and 3 hold with* $\beta^{\perp} = 0$ *. Let* $\overline{CDF}_{\hat{B}_{TSLS}}(b)$ *denote* $CDF_{\hat{B}_{TSLS}}(b)$ *for* $n_1 = 0$ *. Then, under weak identification (i.e.* $\Pi_2 = T^{-1/2}C$ *for a matrix* C *of rank* n *)*

$$1 - \overline{CDF}_{\hat{B}_{TSLS}}(b) \le \Pr\left\{\hat{B}_{TSLS} > b\right\} \le \Pr\left\{\chi^2(k_2 - n) > b\right\}.$$

Alternatively, if cv_L *is the critical values calculated from* $\overline{CDF}_{\hat{B}_{TSLS}}(b)$ *and* cv_U *be the critical values calculated from a* $\chi^2(k_2 - n)$ *distribution, the critical values under weak instruments are always in the interval* $[cv_L, cv_U]$.

Therefore one could have a test where the null hypothesis that the over-identifying restrictions hold is (i) accepted if $\hat{B}_{TSLS} < cv_L$, (ii) rejected if $\hat{B}_{TSLS} > cv_U$ and no decision is taken if $cv_L \le \hat{B}_{TSLS} \le cv_U$.

Theorem 1 and its corollary fully characterize the asymptotic density of the classical Sargan/Byron/Wegge and Basmann test statistics under partial identification of the structural parameters. We now focus on the version of these two statistics where the structural variance is estimated with LIML.

Theorem 2. *Let* $\hat{B}_{LIML} = T\left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{\left(\hat{Q}^{1/2}\hat{\Pi}_2\right)}\left(\hat{Q}^{1/2}\hat{\pi}_1\right) / \hat{\sigma}_u^2$ *where either* $\hat{\sigma}_u^2 = \hat{u}_{LIML}'\hat{u}_{LIML} / T$ *or* $\hat{\sigma}_u^2 = \hat{u}_{LIML}' M_Z \hat{u}_{LIML} / T$ *or* $\hat{\sigma}_u^2 = \left(1 + \hat{\beta}_{LIML}^{*}{}'\hat{\beta}_{LIML}^{*}\right)\hat{\omega}_{11.2}$ *and suppose that Assumptions 1, 2 and 3 hold. Then*

$$\hat{B}_{LIML} \Rightarrow f_1\left(1 - f_1 \frac{r_2' W_{22}^{-1} r_2}{1 + r_2' r_2}\right). \tag{25}$$

$r_2 = -\Delta_1^{-1}\Delta_2$ *and* $\Delta = (\Delta_1, \Delta_2')'$ *is the eigenvector associated to the smallest eigenvalue* f_1 *of the matrix*

$$W = \begin{pmatrix} w_{11} & w_{21}' \\ w_{21} & W_{22} \end{pmatrix} \sim W_{n_2+1}\left(k_2 - n_1, I_{n_2+1}, \left(\beta^{\perp}, 0\right)'\left(\beta^{\perp}, 0\right) / \left(1 + \beta_1^{*}{}'\beta_1^{*}\right)\right)$$

where w_{11} *is a scalar.*

Theorem 2 is more complex than the corresponding result for the standard test for over-identifying restrictions in which the TSLS estimator is used instead of the LIML estimator. The statistic $\hat{B}_{LIML}$ depends, in a elaborate way, on r_2, the limit of the normalized LIML estimator of β_2^*, and two quantities affected by the identification of the structural parameters: f_1, the smallest eigenvalue of W, and W_{22}.

If the rank of Π_2 equals n then $\hat{B}_{LIML} - \hat{B}_{TSLS} = o_p(1)$. Thus, $\hat{B}_{LIML}$ and Sargan/Byron/Wegge and Basmann test statistics are equivalent in the sense that they have the same asymptotic chi-square distribution (central if the over-identifying restrictions hold and non-central if they do not hold). If the over-identifying restrictions hold but the rank condition fails, the behaviour of $\hat{B}_{LIML}$ is very different from that of Sargan/Byron/Wegge and Basmann test statistics. In this case, f_1 is close to zero, and $r_2'W_{22}^{-1}r_2/(1+r_2'r_2) < r_2'W_{22}^{-1}r_2/(r_2'r_2) < 1/f_1$ tends to be large ($W_{22} \geq f_1 I_{n_2}$) so that when the sample size is large, $\hat{B}_{LIML}$ will be close to zero. Precisely,

Corollary 2.1. *If Assumptions 1, 2 and 3 hold with $n_1 < n$, then, for any $0 < c < \infty$,*

$$\lim_{T\to\infty} \Pr\left\{\hat{B}_{LIML} \leq c\right\} > \Pr\left\{\chi^2(k_2 - n) \leq c\right\}.$$

It follows from Corollary 2.1 that a test based on $\hat{B}_{LIML}$ cannot be consistent against failure of the over-identifying restrictions when Π_2 is rank deficient. One may infer from this result that the estimation of the structural variance is crucial for the performance of tests of over-identification.

Notice that $\tilde{r}_2 = W_{22}^{-1}w_{21}$ and that $\hat{B}_{TSLS} / \hat{B}_{LIML} \Rightarrow (1+r_2'r_2)/(1+\tilde{r}_2'\tilde{r}_2)$ (e.g. (40) in the proof of Theorem 2). Thus,

Corollary 2.2. *If Assumptions 1, 2 and 3 hold with $n_1 < n$, then*

$$\lim_{T\to\infty} \Pr\left\{\hat{B}_{TSLS} / \hat{B}_{LIML} > 1\right\} = 1.$$

If identification of the structural parameters fails the probability mass of $\hat{B}_{LIML}$ tends to concentrate around the origin more than that of $\hat{B}_{TSLS}$. This suggests that $\hat{B}_{LIML}$ can be used differently from the standard test statistic. First, it can be used to construct an asymptotically unbiased test for failure of the rank condition on Π_2 (taking no notice of the validity of the compatibility condition). Since this test is inconsistent, its practical usefulness is doubtful although its asymptotic power can be close to one when k_2 is large as suggested by the numerical results reported below. Second, there may be some value in using both $\hat{B}_{TSLS}$ and $\hat{B}_{LIML}$ because if both identification and over-identification fail one would get a large $\hat{B}_{TSLS}$ and a very small $\hat{B}_{LIML}$ but if identification holds the two statistics would be approximately equal.

5. Numerical results

We now illustrate some of the properties of the asymptotic distribution of the Sargan/Byron/Wegge and Basmann test statistics using graphs and tables[4]. First we study the effect of lack of identification on the density of $\hat{B}_{TSLS}$. Figure 1 shows the typical density of $\hat{B}_{TSLS}$ 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 $\hat{B}_{TSLS}$ 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 of the structural parameters and violation of the compatibility conditions. Figure 2 shows typical densities of $\hat{B}_{TSLS}$ for $k_2 = 10$, $n = 3$ and $\lambda = 12$ for different values of for different values of n_2. Comparing Figures 1 and 2 one notices that the effect of identification on the asymptotic density of $\hat{B}_{TSLS}$ tends to be larger under the alternative ($\lambda > 0$) than under the null hypothesis ($\lambda = 0$).

[Figure 2 approximately here]

---

[4] Mathematica 6 files that carry out the computations in this section as well as others are available from the author's web-page (http://www-personal.buseco.monash.edu.au/~forchini). These run on the Mathematica Player freely downloadable from http://www.wolfram.com, and allow the user to plot the asymptotic density functions for the statistics studied and to re-run the simulations reported below for different values of the relevant parameters.

We now investigate how good the asymptotic distribution of $\hat{B}_{TSLS}$ is to the finite sample distribution for the Sargan/Byron/Wegge and Basmann test statistics. Table 2 shows some representative results for $n = 4$ and $k_2 = 8$. In Table 2, In the second, third, fifth and sixth columns, the size is based on the critical values obtained from (22). However, for the results in the second and fifth columns n_2 is calculated by estimating $n_1 = n - n_2$ with the procedure suggested by Robin and Smith (2000) while, in the third and sixth columns, n_2 is taken as known. The fourth and seventh columns contain the size of the test when the critical values are obtained from a chi-square distribution with $k_2 - n$ degrees of freedom. The random variates are generated as independent $(T - k_2)\hat{\Omega} \sim W_{n+1}(T - k_2, I_n)$ and $\hat{\Pi} \sim \Pi_2(\beta, I_4) + T^{-1/2} N(0, I_{k_2} \otimes I_{n+1})$, β is a vector of ones and $\Pi_2 = \begin{pmatrix} D_{n_1} \\ 0 \end{pmatrix}$, D_{n_1} is a diagonal matrix for which the first n_1 diagonal elements are the first n_1 elements of the list $(1.6, 1.2, .8, .4)$ and the remaining are zero, 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 $\alpha = .01\ln(100)/\ln(\mathrm{T})$. The number of replications employed in the Monte Carlo test is 50,000.

It is evident from Table 2 that there are only small differences between the second and the third columns and between the fifth and the sixth columns, so that estimating n_2 does not have a significant effect on the size of the Basmann test when the critical values are obtained from Corollary 1.1. The size of the classical Basmann test is strongly affected by failure of the rank condition. For small sample size ($T \leq 100$, say) all three versions of Basmann test seem to be oversized independently of the method used to calculate the critical values. Similar results apply to the Sargan/Byron/Wegge test. Notice that for small sample sizes the Sargan/Byron/Wegge test has better size properties that Basmann test although they both tend to be oversized (e.g. Hahn and Hausman (2002)).

[Table 2 approximately here]

Figure 3 shows the P-P plots for the asymptotic simulated distributions of $\hat{B}_{LIML}$ (dotted line) and $\hat{B}_{TSLS}$ (dashed line) in comparison to the chi-squared distribution with $k_2 - n$ degrees of freedom. Here we take $n = 4$ and $k_2 = 5, 10, 20$. As n_2 increases the distributions of both $\hat{B}_{LIML}$ and $\hat{B}_{TSLS}$ become different from that of a $\chi^2(k_2 - n)$, but as expected, the probability mass of $\hat{B}_{LIML}$ is shifted more towards zero than that of $\hat{B}_{TSLS}$. An increase of k_2 changes the distribution of $\hat{B}_{TSLS}$ slightly, but affects the distribution of $\hat{B}_{LIML}$ substantially.

[Figure 3 approximately here]
[Figure 4 approximately here]
[Figure 5 approximately here]

To investigate the effect of failures of the over-identifying restrictions we set $\beta^{\perp}$ equal to a $k_2 - n_1$ vector with components equal to $5/\sqrt{k_2 - n_1}$ in Figure 4 and $10/\sqrt{k_2 - n_1}$ in Figure 5. Both Figures 4 and 5 show the P-P plots for $n = 4$ and different values of n_2 and k_2. $\hat{B}_{TSLS}$ behaves as expected since it is affected by failures of both the identification condition and over-identification. As $\sqrt{\beta^{\perp}{}'\beta^{\perp}}$ increases the effects due to the failure of the over-identifying restrictions tend to dominate. $\hat{B}_{LIML}$ has its probability mass concentrated around zero as expected from Corollary 2.1. Moreover, it seems the distribution of $\hat{B}_{LIML}$ becomes more concentrated around zero as both k_2 and $\sqrt{\beta^{\perp}{}'\beta^{\perp}}$ increase. These results are representative of the distribution of $\hat{B}_{LIML}$ and $\hat{B}_{TSLS}$ for other values of n and $\beta^{\perp}$.

6. Conclusions

The asymptotic distributions of the statistics for tests of over-identification suggested by Sargan (1958), Basmann (1960a) Basmann (1960b) and Byron (1974) have been derived in closed form for linear partially identified structural equations models. Variants of these statistics for which the structural variance is based on the LIML estimator have been considered and have been shown to have quirky properties in partially identified structural equations.

Our results can be summarised as follows:

1) The asymptotic distribution of the Sargan/Byron/Wegge and Basmann test statistics in partially identified linear structural equations is a modification of the chi-square distribution with degrees of freedom equal to the degree of over-identification. Under the null hypothesis of a correctly specified structural equation the only parameter affecting the asymptotic distribution of the Sargan/Byron/Wegge and Basmann test statistics is the rank of the matrix of correlations between the instruments and the endogenous variables included as regressors in the structural equation (and this can be consistently estimated in the cases considered).

2) the Sargan/Byron/Wegge and Basmann tests for over-identification may be seriously misleading in partially identified linear structural equations, however, by modifying the critical values of these tests to take into account such a possible failure of identification of the structural parameters, 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 statistics and to select the identified parameters.

3) One has to be very careful about the estimation of the structural variance because the asymptotic properties of the Sargan/Byron/Wegge and Basmann tests change dramatically if we consider different estimators thereof based on the LIML estimator. The simultaneous use of different versions of these statistics may provide a diagnostic tool for detecting failure of over-identification and identification of structural parameters.

4) The asymptotic results derived above allow us to construct tests for over-identification for weakly identified parameters by providing bounds on the asymptotic p-values and critical values.

Appendix: Proofs

Proof of Lemma 1

Invariance of the testing problem. The transformation (16) induces the transformations

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

in the parameter space. Note that if (9) holds then

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

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

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

$$\left[\hat{\Pi}_2, \hat{\pi}_1\right] \to \left[\hat{\Pi}_2 L_{22}, l_{11}\hat{\pi}_1 - \hat{\Pi}_2 l_{21}\right].$$

Replacing this in (11), the numerator changes according to

$$\left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{(\hat{Q}^{1/2}\hat{\Pi}_2)} \left(\hat{Q}^{1/2}\hat{\pi}_1\right) \to l_{11}^2 \left(\hat{Q}^{1/2}\hat{\pi}_1\right)' M_{(\hat{Q}^{1/2}\hat{\Pi}_2)} \left(\hat{Q}^{1/2}\hat{\pi}_1\right).$$

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 (16).

It follows that there is no loss of generality to assume that $\Omega = I_{n+1}$, because one 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

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

and that if (4) holds then $\overline{\pi}_1 = \overline{\Pi}_2 \beta^*$, otherwise $\overline{\pi}_1 \neq \overline{\Pi}_2 \beta^*$. Thus, one 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, one 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 $\overline{\Pi}_2 H = \left[\Pi_{21}, 0\right]$ and the rank of Π_{21} is the same as the rank of Π_2. That is, if identification fails, identified and unidentified components of β can be separated as suggested by Phillips (1989) and Choi and Phillips (1992):

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

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

Proof of Lemma 2

In view of Lemma 1 there is no loss of generality in assuming that the structural model is in canonical form and that the rotations of coordinates in the space of the endogenous variables to separate identified and unidentified parameters have already been carried out.

(ii) and (iv) are essentially Lemma 3 of Cragg and Donald (1996). (i) and (iii) can be proved similarly. Since $\hat{\gamma}_{LIML} = \left(\hat{\phi}_1, \hat{\Phi}_2\right)\begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix} = \hat{\phi}_1 - \hat{\Phi}_2 \hat{\beta}_{LIML}$ we can write

$$\begin{aligned} \hat{u}_{LIML} &= y_1 - Y_2 \hat{\beta}_{LIML} - Z_1 \hat{\gamma} \\ &= y_1 - Y_2 \hat{\beta}_{LIML} - Z_1\left(\hat{\phi}_1 - \hat{\Phi}_2 \hat{\beta}_{LIMIL}\right) \\ &= \left(y_1 - Z_1\hat{\phi}_1 - Z_2\hat{\pi}_1, Y_2 - Z_1\hat{\Phi}_2 - Z_2\hat{\Pi}_2\right)\begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix} + Z_2\left(\hat{\pi}_1 - \hat{\Pi}_2\hat{\beta}_{LIML}\right) \\ &= \left(\hat{v}_1, \hat{V}_2\right)\begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix} + Z_2\left(\hat{\pi}_1 - \hat{\Pi}_2\hat{\beta}_{LIML}\right). \end{aligned}$$

Thus

$$T^{-1}\hat{u}_{LIML}'\hat{u}_{LIML} = \begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix}' \left[T^{-1}\left(\hat{v}_1, \hat{V}_2\right)'\left(\hat{v}_1, \hat{V}_2\right)\right] \begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix}$$
$$+\left(\hat{\pi}_1 - \hat{\Pi}_2\hat{\beta}_{LIML}\right)'\left[T^{-1}Z_2'Z_2\right]\left(\hat{\pi}_1 - \hat{\Pi}_2\hat{\beta}_{LIML}\right)$$
$$= \begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix}' \hat{\Omega} \begin{pmatrix} 1 \\ -\hat{\beta}_{LIML} \end{pmatrix} + \left(\hat{\pi}_1 - \hat{\Pi}_2\hat{\beta}_{LIML}\right)'\left[T^{-1}Z_2'Z_2\right]\left(\hat{\pi}_1 - \hat{\Pi}_2\hat{\beta}_{LIML}\right)$$

and (i) follows easily. (iii) follows from the fact that $\hat{\Pi}_2 \to^P (\Pi_{21}, 0)$, $\hat{\pi}_1 \to^P \pi_1 = \Pi_{21}\beta_1^*$, $\hat{\beta}_{LIML} \Rightarrow \begin{pmatrix} \beta_1^* \\ r \end{pmatrix}$ where r is a random variable having a Cauchy distribution (c.f. Forchini (2006)).

Preliminary results needed for in the Proof of Theorem 1 and its corollaries.

The following two lemmas will be used to prove the Theorem 1 and its corollaries. In view of Lemma 1 there is no loss of generality in assuming that the structural model is in canonical form and that the rotations of coordinates in the space of the endogenous variables to separate identified and unidentified parameters have already been carried out.

Lemma 3. *Suppose that Assumptions 1 and 2 and 3 hold. Let* $\hat{\Pi}_2$ *be partitioned conformably to* $\Pi_2 = [\Pi_{21}, 0]$ *as* $\hat{\Pi}_2 = \left[\hat{\Pi}_{21}, \hat{\Pi}_{22}\right]$*, where* Π_{21} *has full column rank, then:*

(i) $M_{\hat{Q}^{1/2}\hat{\Pi}} = \Pi_{21}^{\perp} M_{\delta} \Pi_{21}^{\perp}{}' + o_p(1)$,

(ii) $\hat{\Pi}_{22}'\hat{Q}^{1/2} M_{\hat{Q}^{1/2}\hat{\Pi}_{21}} \hat{Q}^{1/2}\hat{\Pi}_{22} = T^{-1}\delta'\delta + o_p\left(T^{-1}\right)$,

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

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

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

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 3. Assumption 1(c) implies that

(27) $$\left(\hat{\pi}_1, \hat{\Pi}_{21}, \hat{\Pi}_{22}\right) = \left(\pi_1, \Pi_{21}, 0\right) + T^{-1/2}\left(x, X_1, X_2\right) + o_p\left(T^{-1/2}\right)$$

where $\left(x, X_1, X_2\right) \sim N\left(0, Q^{-1} \otimes I_{n+1}\right)$. 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 $M_{Q^{1/2}\Pi_{21}} = \Pi_{21}^{\perp}\Pi_{21}^{\perp}{}'$. Then, $z = \Pi_{21}^{\perp}{}'Q^{1/2}x \sim N\left(0, I_{k_2-n_1}\right)$, $\delta = \Pi_{21}^{\perp}{}'Q^{1/2}X_2 \sim N\left(0, I_{k_2-n_1} \otimes I_{n_2}\right)$ 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

$$\begin{aligned} M_{\hat{Q}^{1/2}\hat{\Pi}} &= M_{\hat{Q}^{1/2}\hat{\Pi}_{21}} - M_{\hat{Q}^{1/2}\hat{\Pi}_{21}}\left(\hat{Q}^{1/2}\hat{\Pi}_{22}\right)\left[\left(\hat{Q}^{1/2}\hat{\Pi}_{22}\right)' M_{\hat{Q}^{1/2}\hat{\Pi}_{21}}\left(\hat{Q}^{1/2}\hat{\Pi}_{22}\right)\right]^{-1}\left(\hat{Q}^{1/2}\hat{\Pi}_{22}\right)' M_{\hat{Q}^{1/2}\hat{\Pi}_{21}} \\ &= \Pi_{21}^{\perp}\Pi_{21}^{\perp}{}' - \Pi_{21}^{\perp}\delta\left[\delta'\delta\right]^{-1}\delta'\Pi_{21}^{\perp}{}' + o_p(1) \\ &= \Pi_{21}^{\perp}M_{\delta}\Pi_{21}^{\perp}{}' + o_p(1). \end{aligned}$$

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

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

To prove (*iii*) note that

$$\begin{aligned} \Pi_{21}^{\perp}{}'\hat{Q}^{1/2}\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*\right) &= \Pi_{21}^{\perp}{}'\hat{Q}^{1/2}\left(\pi_1 + T^{-1/2}x - \left(\Pi_{21} + T^{-1/2}X_1\right)\beta_1^*\right) \\ &\Pi_{21}^{\perp}{}'\hat{Q}^{1/2}\left(T^{-1/2}x - T^{-1/2}X_1\beta_1^* + \pi_1 - \Pi_{21}\beta_1^*\right) \\ &= T^{-1/2}\left(\Pi_{21}^{\perp}{}'Q^{1/2}x - \Pi_{21}^{\perp}{}'Q^{1/2}X_1\beta_1^* + \Pi_{21}^{\perp}{}'Q\Pi_{21}^{\perp}\beta^{\perp}\right) + o_p\left(T^{-1/2}\right) \\ &= T^{-1/2}\left(z + \beta^{\perp} - W\beta_1^*\right) + o_p\left(T^{-1/2}\right). \end{aligned}$$

Finally,

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

and

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

and Lemma 3 is proved.

Lemma 4. *For* $\delta > 0$, $b > 0$ *and* $\gamma \neq 0$

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

Proof of Lemma 4. Using Kummer transformation we can write the integrand above as

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

The desired integral becomes

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

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

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

The results stated in Lemma 4 follows.

Proof of Theorem 1

Let

$$\begin{aligned}\hat{r}_2 &= \left(\hat{\Pi}_{22}'\hat{Q}^{1/2}M_{\hat{Q}^{1/2}\hat{\Pi}_{21}}\hat{Q}^{1/2}\hat{\Pi}_{22}\right)^{-1}\hat{\Pi}_{22}'\hat{Q}^{1/2}M_{\hat{Q}^{1/2}\hat{\Pi}_{21}}\hat{Q}^{1/2}\hat{\pi}_1 \\ \hat{\tau} &= \frac{\left(\hat{Q}^{1/2}\hat{\pi}_1\right)'M_{\hat{Q}^{1/2}\hat{\Pi}}\left(\hat{Q}^{1/2}\hat{\pi}_1\right)}{1+\hat{r}_1'\hat{r}_1},\end{aligned}$$

then, in view of Lemma 2 and the fact that that $\hat{r}_1=\beta_1^*+o_p(1)$ from the results of Choi and Phillips (1992), Sargan/Byron/Wegge and Basmann test statistics can be written as

$$\hat{B}_{TSLS}=\frac{T\hat{\tau}}{\left(1,-\hat{r}_1,-\hat{r}_2\right)\hat{\Omega}\begin{pmatrix}1\\-\hat{r}_1\\-\hat{r}_2\end{pmatrix}/\left(1+\hat{r}_1'\hat{r}_1\right)+o_p(1)},$$

where $\hat{r}_1$ and $\hat{r}_2$ are the TSLS estimators of β_1^* and β_2^* respectively. Clearly for Basmann tests statistics the $o_p(1)$ term is identically zero. So using Lemma 3 we have that

$$\begin{aligned}\hat{r}_2 &= \left(\hat{\Pi}_{22}'\hat{Q}^{1/2}M_{\hat{Q}^{1/2}\hat{\Pi}_{21}}\hat{Q}^{1/2}\hat{\Pi}_{22}\right)^{-1}\hat{\Pi}_{22}'\hat{Q}^{1/2}M_{\hat{Q}^{1/2}\hat{\Pi}_{21}}\hat{Q}^{1/2}\hat{\pi}_1 \\ &= \left(T^{-1}\delta'\delta+o_p\left(T^{-1}\right)\right)^{-1}\left(T^{-1}\delta'\left(z+\beta^{\perp}-W\beta_1^*\right)+o_p\left(T^{-1}\right)\right) \\ &= \left(\delta'\delta\right)^{-1}\delta'\left(z+\beta^{\perp}-W\beta_1^*\right)+o_p(1) \\ &= \left(\delta'\delta\right)^{-1}\delta'\left(z+\beta^{\perp},W\right)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix}+o_p(1) \\ &= \tilde{r}_2+o_p(1),\end{aligned}$$

where $\tilde{r}_2 = (\delta'\delta)^{-1}\delta'(z+\beta^{\perp},W)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix}$, and

$$\hat{\tau} = \frac{\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*\right)'\hat{Q}^{1/2}M_{\hat{Q}^{1/2}\hat{\Pi}}\hat{Q}^{1/2}\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*\right)}{1+\hat{r}_1'\hat{r}_1}$$

(28) $$= T^{-1}\left(1 \quad -\beta_1^*\right)\left(z+\beta^{\perp},W\right)M_{\delta}\left(z+\beta^{\perp},W\right)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix}/\left(1+\beta_1^*{}'\beta_1^*\right)+o_p\left(T^{-1}\right)$$

$$= T^{-1}\tau + o_p\left(T^{-1}\right),$$

where $\tau = \left(1 \quad -\beta_1^*\right)\left(z+\beta^{\perp},W\right)M_{\delta}\left(z+\beta^{\perp},W\right)\begin{pmatrix}1\\-\beta_1^*\end{pmatrix}/\left(1+\beta_1^*{}'\beta_1^*\right)$. Note that

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

$$\sim \left(1+\beta_1^*{}'\beta_1^*\right)^{1/2} N\left(\beta^{\perp}/\left(1+\beta_1^*{}'\beta_1^*\right)^{1/2}, I_{k_2-n_1}\right)$$

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

$$r_2 = \left(1+\beta_1^*{}'\beta_1^*\right)^{-1/2}\tilde{r}_2 \mid \delta \sim N\left(\left(\delta'\delta\right)^{-1}\delta'\beta^{\perp}/\left(1+\beta_1^*{}'\beta_1^*\right)^{1/2},\left(\delta'\delta\right)^{-1}\right)$$

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

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

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

$$\int_{\mathbb{R}^{(k_2-n_1)n_2}} etr\left\{-\frac{1}{2}\left(I_{n_2}+r_2r_2'\right)\delta'\delta\right\}\left|\delta'\delta\right|^{1/2} {}_0F_1\left(\frac{k_2-n}{2};\frac{\tau}{4}m'M_{\delta}m\right)\exp\left\{m'\delta r_2\right\}d\delta$$

where $m = \beta^{\perp}/\left(1+\beta_1^*{}'\beta_1^*\right)^{1/2}$. Notice that we are interested in deriving the distribution of $b = \tau/\left(1+r_2'r_2\right)$ and this is not affected by transforming $r_2 \to Hr_2$ where H is an orthogonal matrix. So we can replace (29) with

(30)

$$\frac{\exp\left\{-\frac{m'm}{2}\right\}}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)(2\pi)^{n_2(k_2-n_1+1)/2}}\exp\left\{-\frac{\tau}{2}\right\}\tau^{(k_2-n)/2-1}$$
$$\int_{O(n_2)}\int_{\mathbb{R}^{(k_2-n_1)n_2}} etr\left\{-\frac{1}{2}\left(I_{n_2}+Hr_2r_2'H'\right)\delta'\delta\right\}\left|\delta'\delta\right|^{1/2}$$
$${}_0F_1\left(\frac{k_2-n}{2};\frac{\tau}{4}m'M_\delta m\right)\exp\left\{m'\delta Hr_2\right\}d\delta\underline{(dH)}$$

where $\underline{(dH)}$ represent the standardized Haar measure on the group of ($n_2 \times n_2$) orthogonal matrices $O(n_2)$. Transform $\delta \to \delta H\left(I_{n_2}+r_2r_2'\right)^{-1/2}H'$ (the Jacobian is $\left(1+r_2'r_2\right)^{-(k_2-n_1)/2}$) so that (30) becomes

$$\frac{\exp\left\{-\frac{m'm}{2}\right\}}{2^{(k_2-n)/2}\Gamma\left(\frac{k_2-n}{2}\right)(2\pi)^{n_2(k_2-n_1+1)/2}}\exp\left\{-\frac{\tau}{2}\right\}\tau^{(k_2-n)/2-1}\left(1+r_2'r_2\right)^{-(k_2-n_1+1)/2}$$
$$\int_{\mathbb{R}^{(k_2-n_1)n_2}} etr\left\{-\frac{1}{2}\delta'\delta\right\}\left|\delta'\delta\right|^{1/2}{}_0F_1\left(\frac{k_2-n}{2};\frac{\tau}{4}m'M_\delta m\right)\int_{O(n_2)}\exp\left\{m'\delta H\left(I_{n_2}+r_2r_2'\right)^{-1/2}r_2\right\}\underline{(dH)}d\delta$$

Averaging over $O(n_2)$ using Theorem 7.4.1 of Muirhead (1982) we have

$$\int_{O(n_2)}\exp\left\{\left(I_{n_2}+r_2r_2'\right)^{-1/2}r_2m'\delta H\right\}\underline{(dH)}={}_0F_1\left(\frac{n_2}{2};\frac{1}{4}mr_2'\left(I_{n_2}+r_2r_2'\right)^{-1}r_2m'\delta\delta'\right).$$

To evaluate the remaining 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 $2^{-n_2}\left|R\right|^{(k_2-n_1)/2-(n_2+1)/2}$. The integral over $R>0$ can be evaluated using Theorem 7.3.4 of Muirhead (1982) to yield, after noticing that $r_2'\left(I_{n_2}+r_2r_2'\right)^{-1}r_2=r_2'r_2/\left(1+r_2'r_2\right)$,

$$\int_{R>0} etr\left\{-\frac{1}{2}R\right\}\left|R\right|^{\frac{k_2-n_1+1}{2}-\frac{n_2+1}{2}}{}_0F_1\left(\frac{n_2}{2};\frac{1}{4}V'mr_2'\left(I_{n_2}+r_2r_2'\right)^{-1}r_2m'VR\right)dR$$
$$=2^{\frac{(k_2-n_1+1)n_2}{2}}\Gamma_{n_2}\left(\frac{k_2-n_1+1}{2}\right){}_1F_1\left(\frac{k_2-n_1+1}{2};\frac{n_2}{2};\frac{1}{2}\frac{mr_2'r_2m'}{\left(1+r_2'r_2\right)}VV'\right),$$

where, the function $\Gamma_{n_2}(a)$ is defined in Theorem 2.1.12 of Muirhead (1982). Thus we are left with

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

$$\int_{V'V=I_{n_2}} {}_0F_1\left(\frac{k_2-n}{2};\frac{\tau}{4}m'\left(I_{k_2-n_1}-VV'\right)m\right){}_1F_1\left(\frac{k_2-n_1+1}{2};\frac{n_2}{2};\frac{1}{2}\frac{r_2'r_2}{\left(1+r_2'r_2\right)}m'VV'm\right)(dV).$$

Now, we let $\tau=\left(1+r_2'r_2\right)b$, transform r_2 to polar coordinates $r_2=vq^{1/2}$ where $v'v=1$ and $q=r_2'r_2>0$ and integrate out both v and q to obtain

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

(31)
$$\int_{q>0}\exp\left\{-\frac{qb}{2}\right\}q^{n_2/2-1}\left(1+q\right)^{-(n_2+1)/2}$$

$$\int_{V'V=I_{n_2}} {}_0F_1\left(\frac{k_2-n}{2};\frac{b(1+q)}{4}m'\left(I_{k_2-n_1}-VV'\right)m\right)$$

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

where we used $\int_{v'v=1}(dv)=\frac{2\pi^{n_2/2}}{\Gamma\left(n_2/2\right)}$. Corollary 1.1 follows easily by setting $m=0$ in (31) and noticing that

(32)
$$\int_{V'V=I_{n_2}}(dV)=\frac{2^{n_2}\pi^{(k_2-n_1)n_2/2}}{\Gamma_{n_2}\left(\frac{k_2-n_1}{2}\right)}$$

and

(33)
$$\int_{q>0}\exp\left\{-\frac{qb}{2}\right\}q^{n_2/2-1}\left(1+q\right)^{-(n_2+1)/2}dq=\Gamma\left(\frac{n_2}{2}\right)\Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{b}{2}\right)$$

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 (33) 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

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

We now interpret V as the matrix formed by the first n_2 columns of 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

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

Corollary 1.2 follows from noticing that

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

To prove the second part of Corollary 1.1, notice that

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

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

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

So, using Lemma 4 we obtain

$$\int_{0<t<b} e^{-t/2} t^{\frac{k_2-n}{2}-1} \Psi\left(\frac{n_2}{2};\frac{1}{2};\frac{t}{2}\right) dt = \frac{\sqrt{\pi}}{\Gamma\left(\frac{n_2+1}{2}\right)} \int_{0<t<b} e^{-t/2} t^{(k_2-n)/2-1} {}_1F_1\left(\frac{n_2}{2};\frac{1}{2};\frac{t}{2}\right) dt$$

$$- \frac{\sqrt{2\pi}}{\Gamma\left(\frac{n_2}{2}\right)} \int_{0<t<b} e^{-t/2} t^{(k_2-n+1)/2-1} {}_1F_1\left(\frac{n_2+1}{2};\frac{3}{2};\frac{t}{2}\right) dt$$

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

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

Inserting this in (37) we obtain the desired result.

Proof of Corollary 1.3

This result follows from proposition 9 of Cragg and Donald (1996) and Corollary 1.1.

Proof of Theorem 2

In view of Lemma 1 there is no loss of generality in assuming that the structural model is in canonical form and that the rotations of coordinates in the space of the endogenous variables to separate identified and unidentified parameters have already been carried out. The LIML estimator for the unidentified parameters is $\hat{\beta}_2 = \left(1+\beta_1^*{}'\beta_1^*\right)^{1/2} r_2$ and $r_2 = -\Delta_1^{-1}\Delta_2$ and $\Delta = (\Delta_1, \Delta_2{}')'$ is the eigenvector associated to the smallest eigenvalue of

$$\begin{aligned}\hat{W} &= T\Sigma^{-1/2}\left(\hat{\pi}_1, \hat{\Pi}_{22}\right)'\hat{Q}^{1/2} M_{\hat{Q}^{1/2}\hat{\Pi}_{21}} \hat{Q}^{1/2}\left(\hat{\pi}_1, \hat{\Pi}_{22}\right)\Sigma^{-1/2} \\ &= T\Sigma^{-1/2}\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*, \hat{\Pi}_{22}\right)'\hat{Q}^{1/2} M_{\hat{Q}^{1/2}\hat{\Pi}_{21}} \hat{Q}^{1/2}\left(\hat{\pi}_1 - \hat{\Pi}_{21}\beta_1^*, \hat{\Pi}_{22}\right)\Sigma^{-1/2}\end{aligned}$$

where $\Sigma = Diag\left(1+\beta_1^*{}'\beta_1^*, I_{n_2}\right)$ when the sample size is large (see Theorem 1 of Forchini (2006)). Using (1), (4) and (5) in Lemma 2 we have

(38) $$\hat{W} = W + o_p(1),$$

where $W = X'X$ and

$$X = \left[\left(z+\beta^{\perp} - W\beta_1^*\right), \delta\right]\Sigma^{-1/2} \sim N\left(\left[\beta^{\perp} / \left(1+\beta_1^*{}'\beta_1^*\right)^{1/2}, 0\right], I_{k_2-n_1} \otimes I_{n_2}\right).$$

Thus,

(39) $$W \sim W_{n_2+1}\left(k_2-n_1, I_{n_2+1}, \left(\beta^{\perp},0\right)'\left(\beta^{\perp},0\right) / \left(1+\beta_1^*{}'\beta_1^*\right)\right).$$

Partition W as

$$W = \begin{pmatrix} w_{11} & w_{21}' \\ w_{21} & W_{22} \end{pmatrix}$$

where w_{11} is a scalar, and let f_1 be the smallest eigenvalue of W . Then

$$\begin{pmatrix} w_{11} - f_1 & w_{21}' \\ w_{21} & W_{22} - f_1 I_{n_2} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_1 \begin{pmatrix} w_{11} - f_1 & w_{21}' \\ w_{21} & W_{22} - f_1 I_{n_2} \end{pmatrix} \begin{pmatrix} 1 \\ -r_2 \end{pmatrix} = 0 ,$$

which gives

(40)
$$\begin{aligned} w_{11} - f_1 - w_{21}' r_2 &= 0 \\ w_{21} - \left(W_{22} - f_1 I_{n_2} \right) r_2 &= 0. \end{aligned}$$

Multiplying $w_{21}' W_{22}^{-1}$ times the second equation we have

$$w_{21}' W_{22}^{-1} w_{21} - w_{21}' W_{22}^{-1} \left(W_{22} - f_1 I_{n_2} \right) r_2 = 0 .$$

Now we use this to find $\hat{\tau} = T^{-1}\tau + o_p\left(T^{-1}\right)$:

$$\begin{aligned} T^{-1}\tau &= w_{11} - w_{21}' W_{22} w_{21} \\ &= f_1 + f_1 w_{21}' W_{22}^{-1} r_2 \\ &= f_1 \left(1 + r_2' \left(W_{22} - f_1 I_{n_2} \right) W_{22}^{-1} r_2 \right) \\ &= f_1 \left(1 + r_2' r_2 \right) \left(1 - f_1 \frac{r_2' W_{22}^{-1} r_2}{1 + r_2' r_2} \right). \end{aligned}$$

The statement of the Theorem follows from the last display and the fact that

$$\hat{B}_{LIML} = \frac{T\hat{\tau}}{1 + \hat{\beta}_2' \hat{\beta}_2 / \left(1 + \hat{\beta}_1' \hat{\beta}_1 \right)} ,$$

$\hat{\beta}_1 = \beta_1 + o_p\left(1\right)$ (Forchini (2006), Theorem 2(2)).

<u>Proof of Corollary 2.1</u>

Since

$$f_1 = \min \frac{\Delta' W \Delta}{\Delta' \Delta} \le \min_i w_{ii}$$

where w_{ii} denotes the element in position *(i,i)* of W and

$$\begin{aligned} w_{11} &\sim \chi^2\left(k_2 - n_1, \beta^{\perp}' \beta^{\perp} / \left(1 + \beta_1^{*}' \beta_1^{*} \right) \right) \\ w_{ii} &\sim \chi^2\left(k_2 - n_1 - i + 1 \right), \quad i = 2, \ldots, n_2 + 1 \end{aligned}$$

are independent (e.g. Theorem 10.3.8 of Muirhead (1982)), we have for any finite $c > 0$ that

$$\begin{aligned}\Pr\{f_1 \le c\} &\ge \Pr\{\min_i w_{ii} \le c\}\\ &= 1-\Pr\{\min_i w_{ii} \ge c\}\\ &= 1-\Pr\left\{\chi^2\left(k_2-n_1,\frac{\beta^{\perp}{}'\beta^{\perp}}{1+\beta_1^*{}'\beta_1^*}\right)\ge c\right\}\times\prod_{i=2}^{n_2+1}\Pr\left\{\chi^2\left(k_2-n_1-i+1\right)\ge c\right\}\\ &\ge 1-\prod_{i=2}^{n_2+1}\Pr\left\{\chi^2\left(k_2-n_1-i+1\right)\ge c\right\}.\end{aligned}$$

Note that choosing $i = n_2 + 1$ we have $w_{n_2+1,n_2+1} \sim \chi^2\left(k_2 - n\right)$, and that for all other i we have $\Pr\left\{\chi^2\left(k_2-n_1-i+1\right)\ge c\right\}\le 1$, so

$$\begin{aligned}\prod_{i=2}^{n_2+1}\Pr\left\{\chi^2\left(k_2-n_1-i+1\right)\ge c\right\} &= \Pr\left\{\chi^2\left(k_2-n\right)\ge c\right\}\prod_{i=2}^{n_2}\Pr\left\{\chi^2\left(k_2-n_1-i+1\right)\ge c\right\}\\ &\le \Pr\left\{\chi^2\left(k_2-n\right)\ge c\right\}.\end{aligned}$$

The corollary follows by noting that $\lim_{T\to\infty}\Pr\left\{\hat{B}_{LIML}\le c\right\}\ge\Pr\{f_1\le c\}$.

<u>Proof of Corollary 2.2</u>

Notice that $W_{22} - f_1 I_{n_2}$ is positive definite with probability one, and from (40), $\tilde{r}_2 = \left(I_{n_2} - f_1 W_{22}^{-1}\right) r_2$. Then

$$\begin{aligned}0 &< 2 f_1\left(W_{22}^{-1}r_2\right)'\left(W_{22}-f_1 I_{n_2}\right)\left(W_{22}^{-1}r_2\right)\\ &\le 2 f_1 r_2{}'W_{22}^{-1}r - f_1^2 r_2{}'W_{22}^{-2}r_2\\ &= -r_2{}'\left(I_{n_2}-f_1W_{22}^{-1}\right)^2 r_2 + r_2{}'r_2\\ &= -\tilde{r}_2{}'\tilde{r}_2 + r_2{}'r_2.\end{aligned}$$

So

$$\begin{aligned}&r_2{}'r_2 > \tilde{r}_2{}'\tilde{r}_2\\ &1+r_2{}'r_2 > 1+\tilde{r}_2{}'\tilde{r}_2\\ &\left(1+r_2{}'r_2\right)/\left(1+\tilde{r}_2{}'\tilde{r}_2\right) > 1\end{aligned}$$

with probability one. The corollary follows from the fact that $\hat{B}_{TSLS}/\hat{B}_{LIML} \Rightarrow \left(1+r_2{}'r_2\right)/\left(1+\tilde{r}_2{}'\tilde{r}_2\right)$.

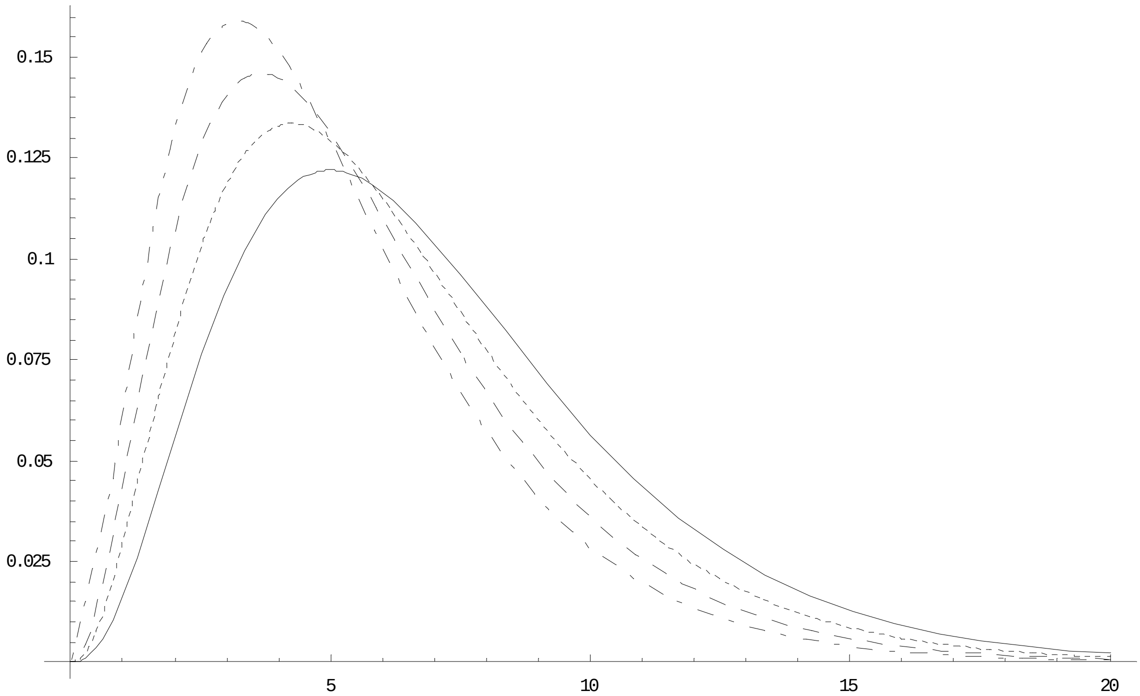

Figure 1: Asymptotic density of $\hat{B}_{TSLS}$ for $k_2 = 10$, $n = 3$ and $n_2 = 0$ (solid line), $n_2 = 1$ (dotted line), $n_2 = 2$ (dashed line) and $n_2 = 3$ (dotted-dashed line) when the over-identifying restrictions hold.

			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 $\hat{B}_{TSLS}$ (in %) using a nominal 5% level from $\chi^2\left(k_2 - n\right)$

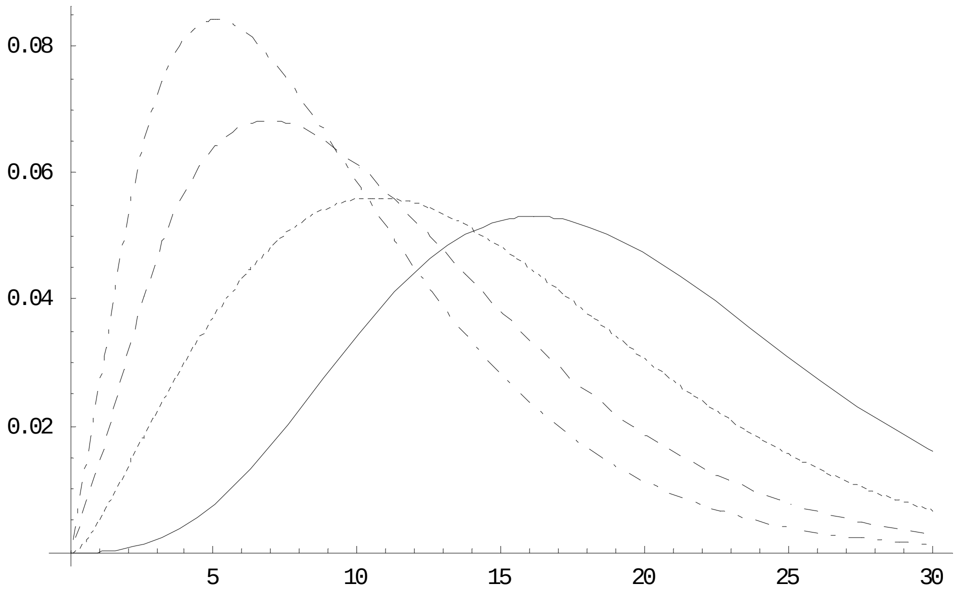

Figure 2: Comparison of the asymptotic densities of $\hat{\mathrm{B}}_{TSLS}$ for $k_2 = 10$, $n = 3$ for $n_2 = 0$ (solid line), $n_2 = 1$ (dotted line), $n_2 = 2$ (dashed line) and $n_2 = 3$ (dotted-dashed line) when the over-identifying restrictions do not hold with $\lambda = 12$.

		Basmann			Sargan/Byron/Wegge	
n_1	n_1 estimated	n_1 known	$\chi^2(k_2-n)$	n_2 estimated	n_2 known	$\chi^2(k_2-n)$
			T=25			
0	18.40	18.40	7.79	10.81	10.8	1.61
1	21.19	19.50	9.65	12.70	11.12	2.68
2	23.67	21.02	13.01	13.71	10.88	4.13
3	27.24	22.47	17.99	15.49	10.56	6.45
4	32.15	22.97	22.97	19.09	9.37	9.37
			T=50			
0	9.99	9.99	2.54	7.18	7.18	1.10
1	10.58	10.58	3.88	7.42	7.42	1.81
2	11.36	11.36	5.72	7.67	7.68	2.81
3	12.81	12.13	8.70	8.24	7.70	4.76
4	17.08	13.21	13.21	11.43	7.74	7.74
			T=100			
0	7.27	7.28	1.45	6.04	6.05	0.95
1	7.51	7.52	2.17	6.16	6.17	1.47
2	7.77	7.78	3.31	6.25	6.26	2.21
3	8.29	8.31	5.46	6.35	6.35	3.98
4	10.14	8.84	8.84	7.73	6.56	6.56
			T=400			
0	5.53	5.55	0.82	5.28	5.29	0.75
1	5.19	5.20	1.25	4.92	4.94	1.13
2	5.64	5.66	2.19	5.31	5.33	1.95
3	5.58	5.60	3.38	5.18	5.20	3.60
4	5.95	5.95	5.95	5.38	5.38	5.38
			T=1600			
0	5.10	5.12	0.69	5.00	50.2	0.66
1	5.00	5.02	1.09	4.94	4.96	1.17
2	5.27	5.28	1.88	5.18	5.19	1.85
3	5.05	5.06	2.98	4.96	4.98	2.91
4	5.23	5.23	5.23	5.13	5.13	5.13
			T=6400			
0	5.05	5.06	0.73	5.03	5.04	0.73
1	5.06	5.01	1.10	4.99	4.99	1.10
2	4.93	4.95	1.73	4.92	4.94	1.71
3	5.21	5.23	2.98	5.19	5.20	2.96
4	5.13	5.13	5.13	5.11	5.11	5.11
			T=infinity			
0	5.00	5.00	0.71	5.00	5.00	0.71
1	5.00	5.00	1.10	5.00	5.00	1.10
2	5.00	5.00	1.76	5.00	5.00	1.76
3	5.00	5.00	2.91	5.00	5.00	2.91
4	5.00	5.00	5.00	5.00	5.00	5.00

Table 2: Size of Basmann and Sargan/Byron/Wegge tests (in %) using a nominal 5% value, $n=4$ and $k_2=8$.

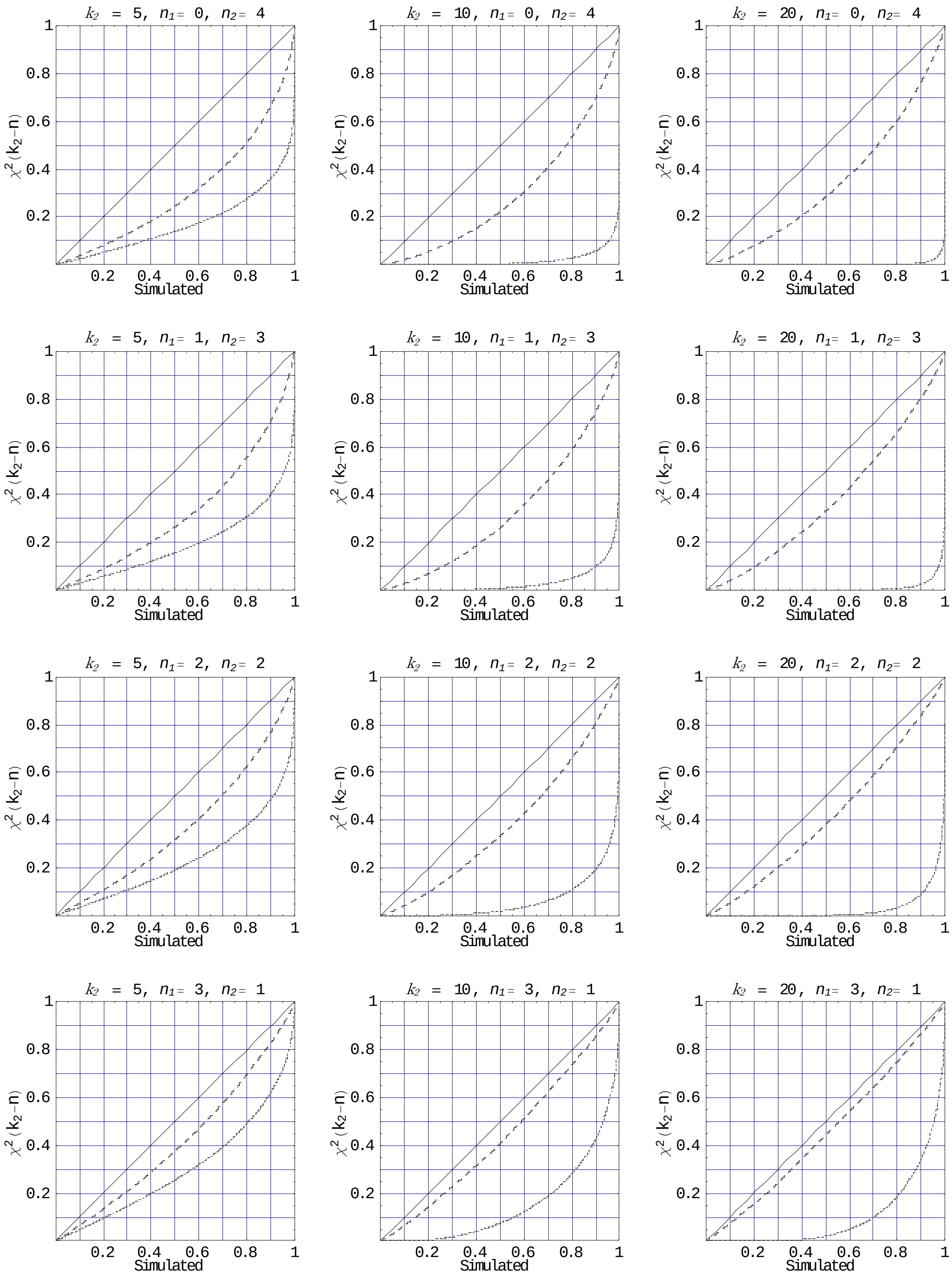

Figure 3: P-P plots for the asymptotic distribution of $\hat{B}_{LIML}$ (dotted line) and $\hat{B}_{TSLS}$ (dashed line) when the over-identifying restrictions hold.

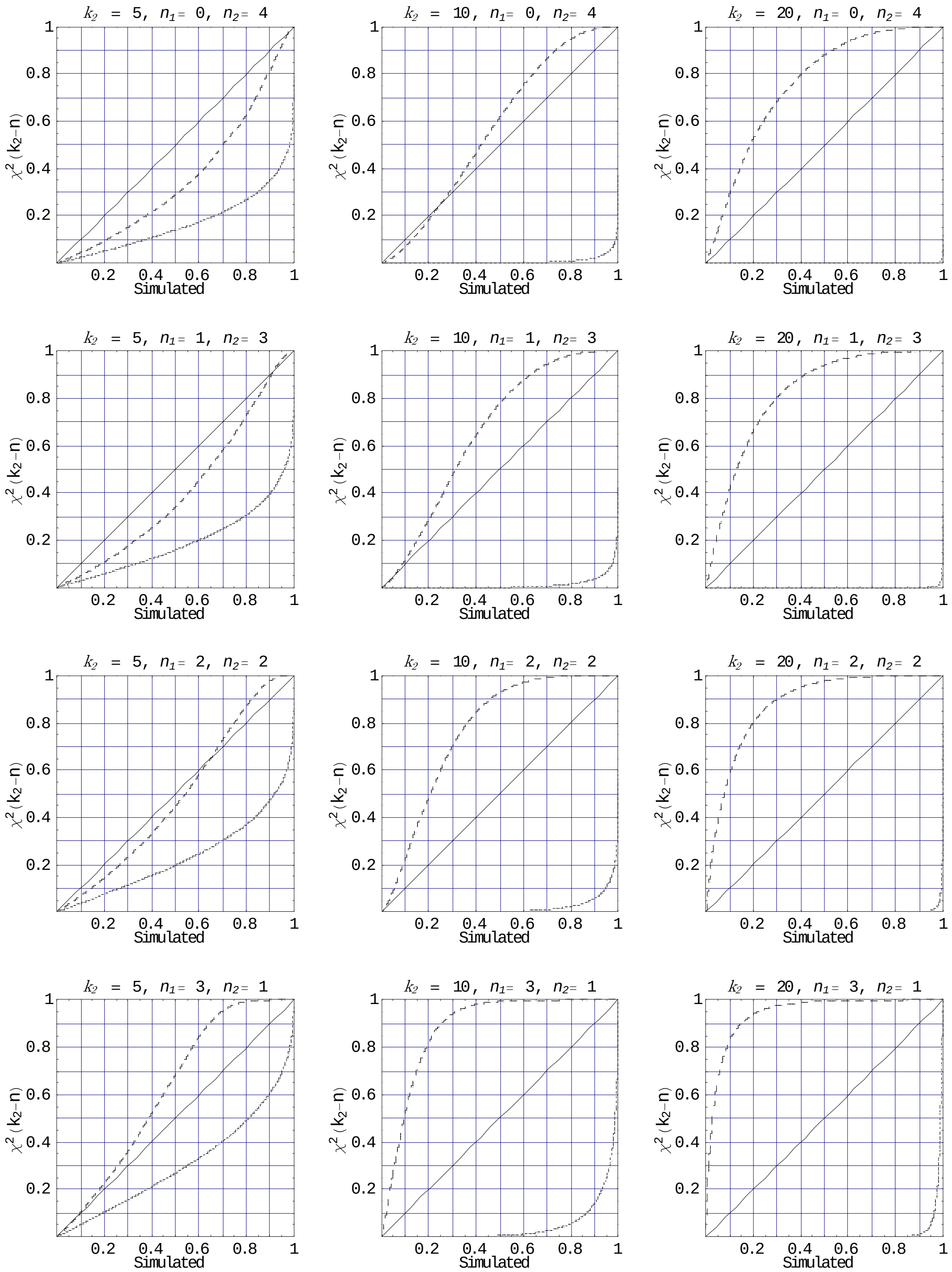

Figure 4: P-P plots for the asymptotic distribution of $\hat{B}_{LIML}$ (dotted line) and $\hat{B}_{TSLS}$ (dashed line) when the over-identifying restrictions fail with $\sqrt{\beta^{\perp}{}'\beta^{\perp}} = 5$.

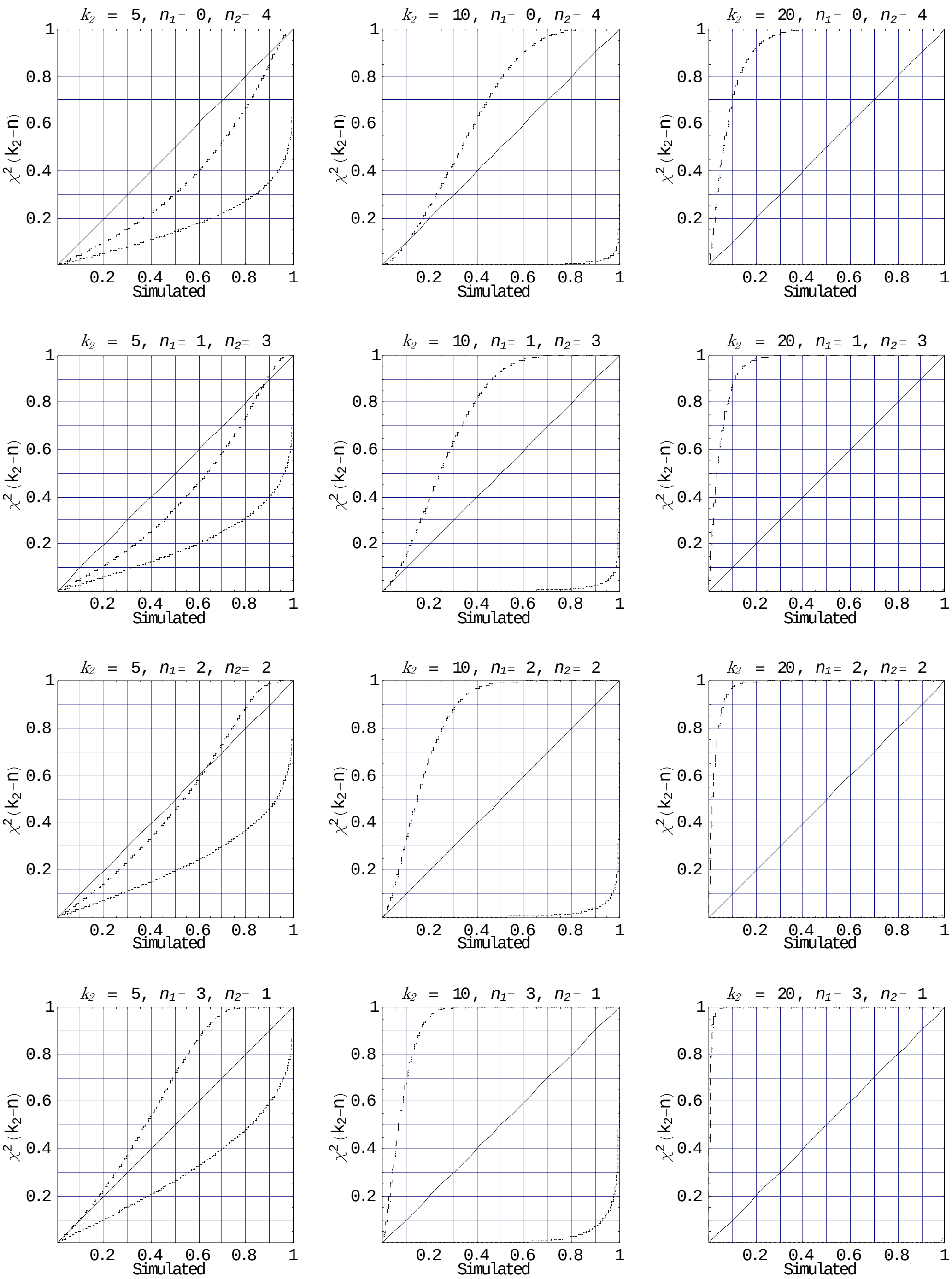

Figure 5: P-P plots for the asymptotic distribution of $\hat{B}_{LIML}$ (dotted line) and $\hat{B}_{TSLS}$ (dashed line) when the over-identifying restrictions fail with $\sqrt{\beta^{\perp}{}'\beta^{\perp}} = 10$.

References

Anderson, T. W. and H. Rubin (1949). "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." Annals of Mathematical Statistics 20: 46-62.

Anderson, T. W. and H. Rubin (1950). "The Asymptotic Properties of Estimators of the Parameters of a Single Equation in a Complete System of Stochastic Equations." Annals of Mathematical Statistics 21: 570-582.

Basmann, R. L. (1960a). "On Finite Sample Distributions of Generalized Classical Linear Identifiability 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.

Dhrymes, P. J. (1969). "Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS an 3SLS Estimated Parameters." The Review of Economic Studies 36 (2): 213-226.

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.

Forchini, G. (2006). "The Asymptotic Distribution of the LIML Estimator in a Partially identified Structural Equation." Monash University, Department of Econometrics and Business Statistics working paper, available at http://www-personal.buseco.monash.edu.au/~forchini/AsymLIML.pdf.

Hahn, J. and J. Hausman (2002). "A New Specification Test for the Validity of Instrumental Variables." Econometrica 70(1): 163-189.

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" in. *Handbook of Econometrics, Volume*. Z. Griliches and 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.

Hsiao, C. (1983). "Identification" in. *Handobook of Econometrics*. Z. Griliches and M. D. Intriligator. New York, North-Holland. Volume 1,**:** 223-283,.

Hwang, H.-S. (1980). "A Comparison of Tests of Overidentifying Restrictions." Econometrica 48 (7): 1821-1825.

Kadane, J. B. (1974). "Testing Subsets of Overidentifying Restrictions." Econometrica 42 (5): 853-867.

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.

Murray, M. P. (2006). "Avoiding Invalid Instruments and Coping with Weak Instruments." Journal of Economic Perspectives 20(4): 111-132.

Phillips, P. C. B. (1983). "Exact Small Sample Theory in the Simultaneous Equation Model" in. *Handbook of Econometrics*. M. D. Intriligator and 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.

Wegge, L. L. (1978). "constrained Indirect Least Squares Estimators." Econometrica 46: 435-499.