ISSN 1440-771X

MONASH University

Australia

Department of Econometrics
and Business Statistics

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

The Asymptotic Distribution of the LIML Estimator in a Partially Identified Structural Equation

G Forchini

January 2006

Working Paper 1/06

The Asymptotic Distribution of the LIML Estimator in a Partially Identified Structural Equation

G. Forchini[1]
Monash University

January 27, 2006

Abstract

We derive general formulae for the asymptotic distribution of the LIML estimator for the coefficients of both endogenous and exogenous variables in a partially identified linear structural equation. We extend previous results of Phillips (1989) and Choi and Phillips (1992) where the focus was on IV estimators. We show that partial failure of identification affects the LIML in that its moments do not exist even asymptotically.

JEL Classification: C13, C30
Key Words: LIML estimator, Partial Identification, Linear structural equation, Asymptotic distribution

[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

1. Introduction

The problem of identification and possible failures of identification has been the focus of much econometric research in the last few years. The fact that some of the parameters of a model may be only weakly identified (e.g. Staiger and Stock (1997)) or partially unidentified (e.g. Phillips (1989) and Choi and Phillips (1992)) has been shown to be a common problem faced in applied econometrics, and has raised a lot of inferential issues for theorists.

The two-stage-least-squares (TSLS) estimator of the coefficients of the endogenous and the exogenous variables has been fully investigated both in the asymptotic and in the finite sample literature. The exact distribution theory can be found in the work of Phillips (1983), Phillips (1984a), Phillips (1989), Hillier (1985), Choi and Phillips (1992) and Skeels (1995) for the general case of $n+1$ endogenous variables. The TSLS estimator of identified structural parameters is consistent and asymptotically normal (Fujikoshi, Morimune, Kunitomo and Taniguchi (1982)), but consistency fails when the structural equation is totally unidentified (Phillips (1983) and Phillips (1989)), partially identified (Phillips (1989) and Choi and Phillips (1992)), and weakly identified (Staiger and Stock (1997)). The TSLS estimator is inconsistent when the number of instruments increases with the sample size (Bekker (1994)), even if the parameters are identified.

Limited information maximum likelihood (LIML) has also been extensively studied. The exact distribution of the LIML estimator for the coefficients of the n endogenous variables included as regressors in a linear structural equation is given by Phillips (1984b), Phillips (1985) and Hillier and Skeels (1993). This is expressed in terms of infinite series of invariant matrix polynomials, but simplifies to a multivariate Cauchy distribution if the coefficients of the endogenous variables are totally unidentified. Similar simplifications can potentially be obtained for partially identified models, but this task is extremely complicated. Higher order expansions for the identified case are given by Fujikoshi, Morimune, Kunitomo and Taniguchi (1982).

One important characteristic of the exact distribution of the LIML estimator is that it has no integer moments in finite samples (e.g. Phillips (1984b), Phillips (1985), Hillier and Skeels (1993) and references therein). This property carries on asymptotically when the structural equation is totally unidentified because, in this case, the exact distribution is invariant to the sample size. However, if the model is identified, then the LIML estimator has an asymptotic normal distribution that it approaches more quickly than the TSLS estimator (see Anderson, Kunitomo and Sawa (1982) and Fujikoshi, Morimune, Kunitomo and Taniguchi (1982) for further discussion on higher order expansions for the LIML estimator).

The LIML estimator is asymptotically equivalent to the TSLS estimator if the structural equation is identified, but this is not the case if the instruments are weak (Staiger and Stock (1997)). Then, the LIML estimator is inconsistent and has a non-standard asymptotic distribution. Chao and Swanson (2005) and Han and Phillips (2005) have shown that the LIML estimator is consistent when the number of, respectively, weak instruments and weak moment conditions increases with the sample size.

The existing literature is uninformative about the properties of LIML in partially identified linear structural equations, even though it is very difficult to distinguish between weak and irrelevant instruments in practice. The aim of our paper is to fill this gap and derive the asymptotic distribution of the LIML estimator for the coefficients of both endogenous and exogenous variables in a partially identified linear structural equation. We study the similarities and the differences between the distributions of the LIML and the TSLS estimators when identification partially fails.

For the sake of simplicity, and with no loss of generality, we consider a linear structural equation where the canonical transformations described in Phillips (1983) and the rotations of coordinates in the space of both the endogenous and the exogenous variables of Phillips (1989) and Choi and Phillips (1992) have been carried out. This allows us to partition the vector of coefficients (for both endogenous and exogenous variables) in two sub-vectors containing respectively the identified and the unidentified parameters. We obtain an approximation for the LIML estimator for both sub-vectors of parameters, and use it to study the asymptotic properties of LIML. We find that: *(i)* the LIML estimators for the identified coefficients of both endogenous and exogenous variables are consistent and have covariance matrix mixed normal limiting distributions; *(ii)* they have asymptotic normal distributions conditional on the LIML estimators of the coefficients of the unidentified endogenous variables; *(iii)* the estimators for the unidentified coefficients converge in law to non-degenerate distributions proportional to the multivariate Cauchy, and have no finite integer moments.

The last result implies that without, canonical transformations and rotations of coordinates, the LIML estimator does not have integer moments even asymptotically when the parameters are only partially identified. This unexpected result suggests that identification failures affect the LIML estimator more than the TSLS estimator. It also indicates that using asymptotic mean-square-error type measures to choose the instruments may be inadequate when one or more of the instruments could be irrelevant or could be close to being irrelevant.

The structure of the paper is as follows. Section 2 specifies the model and some preliminary results. Section 3 gives the asymptotic distribution for the LIML estimator of the identified coefficients of both endogenous and exogenous variables. Section 4 discusses the "no moment" problem in weakly identified models and Section 5 concludes.

2. The model and preliminary results

We consider a structural equation

$$(1) \qquad \underset{(T\times 1)}{y} = \underset{(T\times n_1)}{Y_1} \underset{(n_1\times 1)}{\beta_1^*} + \underset{(T\times n_2)}{Y_2} \underset{(n_2\times 1)}{\beta_2^*} + \underset{(k_2\times 1)}{Z_1} \begin{pmatrix} \underset{(k_{11}\times 1)}{\gamma_1^*} \\ \underset{(k_{12}\times 1)}{\gamma_2^*} \end{pmatrix} + u,$$

with corresponding reduced form

$$(2) \qquad \left(y, Y_1, Y_2\right) = \underset{(T\times k_1)}{Z_1} \begin{pmatrix} \underset{(k_{11}\times 1)}{\phi_1} & \underset{(k_{11}\times n_1)}{\Phi_{11}} & \underset{(k_{11}\times n_2)}{\Phi_{22}} \\ \underset{(k_{12}\times 1)}{\phi_2} & \underset{(k_{12}\times n_1)}{\Phi_{21}} & \underset{(k_{12}\times n_2)}{\Phi_{22}} \end{pmatrix} + \underset{(T\times k_2)}{Z_2} \left(\underset{(k_2\times 1)}{\pi}, \underset{(k_2\times n_1)}{\Pi_1}, \underset{(k_2\times n_2)}{\Pi_2} \right) + \left(v, V_1, V_2\right),$$

where $k_1 = k_{11} + k_{12}$ and $n_1 + n_2 = n \le k_2$. The matrices Z_1 and Z_2 contain observations on the exogenous variables, and y, Y_1 and Y_2 denote matrices of endogenous variables. The dimensions of vectors and matrices are reported in square brackets the first time they are used, unless they are obvious from the context.

We assume:

Assumption 1. *(Identification) The following restrictions are satisfied:*
(a) the model specified by equations (1) *and* (2) *is in canonical form (e.g. Phillips (1983)) and partially identified (e.g. Phillips (1989) and Choi and Phillips (1992)) in the sense that* $\Phi_{21} = 0$ *and* $\Pi_1 = 0$*;*

(b) the compatibility conditions

(1) $\pi = \Pi_2 \beta_2^*$ and $u = v - V_1 \beta_1^* - V_2 \beta_2^*$ and

(2) $\gamma_1^* = \phi_1 - \Phi_{11} \beta_1^* - \Phi_{12} \beta_2^*$, $\gamma_2^* = \phi_2 - \Phi_{22} \beta_2^*$;

hold;

(c) the rank conditions

(1) $rank(\Pi_2) = n_2$ and

(2) $rank(\Phi_{22}) = k_{12} = rank \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ 0 & \Phi_{22} \end{pmatrix}$

hold.

Assumption 2 *(Moment conditions)*

(a) $T^{-1}(v, V_1, V_2)'(v, V_1, V_2) \to^P \Omega = I_{n+1}$;

(b) $T^{-1}(Z_1, Z_2)'(Z_1, Z_2) \to^P Q = I_{k_1 + k_2}$;

(c) $T^{-1}(Z_1, Z_2)'(v, V_1, V_2) \to^P 0$;

(d) $T^{-1/2}(Z_1, Z_2)'(v, V_1, V_2) \to^D N(0, Q \otimes \Omega)$.

Assumption 1(a) implies that the rotations of coordinates in the space of endogenous and exogenous variables have already been carried out. We follow Phillips (1989) and Choi and Phillips (1992) and assume that the model is in canonical form. By so doing we simplify the analysis without compromising the generality of our results. The asymptotic distribution for the LIML estimator of the parameters of a structural equation that is not in canonical form can be easily obtained from our results by linear transformations (see Phillips (1983) for details). Assumptions 1(b) and 1(c) make the reduced form (2) compatible with the structural equation (1). These restrictions are known as *over-identifying* restrictions (e.g. Byron (1974) and Hausman (1983)), or *identification* conditions (e.g. Phillips (1983)). Assumptions 1(b) and 1(c) imply that the parameters β_2^* and γ_2^* are identified and can be written uniquely in terms of the reduced form parameters, whereas the parameters β_1^* and γ_1^* are unidentified.

Assumption 2 is a set of standard moment conditions expressed in the matrix notation of Muirhead (1982), and holds in a large variety of situations. In Assumptions 2(a) and 2(b) we have set $\Omega = I_{n+1}$ and $Q = I_{k_1+k_2}$ since, following Phillips (1989) and Choi and Phillips (1992), we have already assumed that the structural parameters are in canonical form.

Let

$$\left(\tilde{\pi}, \tilde{\Pi}_1, \tilde{\Pi}_2\right) = \left(Z_2' M_{Z_1} Z_2\right)^{-1/2} Z_2' M_{Z_1}\left(y, Y_1, Y_2\right) \tag{3}$$

$$\left(\hat{\phi}, \hat{\Phi}_1, \hat{\Phi}_2\right) = \left(Z_1' Z_1\right)^{-1} Z_1'\left(y, Y_1, Y_2\right) \tag{4}$$

and

$$S = \left(y, Y_1, Y_2\right)' M_{(Z_1, Z_2)}\left(y, Y_1, Y_2\right), \tag{5}$$

where $P_Z = Z\left(Z'Z\right)^{-1}Z'$ and $M_Z = I_T - P_Z$ for any $T \times p$ full column rank matrix Z. Then, the LIML estimator of $\left(\beta_1^*{}', \beta_2^*{}'\right)'$, $\left(\hat{\beta}_1', \hat{\beta}_2'\right)'$, minimizes the ratio

$$\nu\left(\hat{\beta}_1, \hat{\beta}_2\right) / \delta\left(\hat{\beta}_1, \hat{\beta}_2\right), \tag{6}$$

where

$$\nu\left(\hat{\beta}_1, \hat{\beta}_2\right) = \begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}' \left(\tilde{\pi}, \tilde{\Pi}_1, \tilde{\Pi}_2\right)' \left(\tilde{\pi}, \tilde{\Pi}_1, \tilde{\Pi}_2\right) \begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}$$

and

$$\delta\left(\hat{\beta}_1, \hat{\beta}_2\right) = \begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}' S \begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}.$$

We can show that:

Theorem 1. *If Assumptions 1 and 2 hold, the estimator of the unidentified parameters,* $\hat{\beta}_1$, *is* $\hat{\beta}_1 = -\Delta_1^{-1}\Delta_2$ *where* $\Delta = \left(\Delta_1, \Delta_2'\right)'$ *minimizes*

$$\frac{\Delta'\left(\tilde{\pi}, \tilde{\Pi}_1\right)' M_{\tilde{\Pi}_2}\left(\tilde{\pi}, \tilde{\Pi}_1\right)\Delta}{\Delta'\left(S_{11.2} + \left(\tilde{A} - S_{22}^{-1}S_{21}\right)' S_{22}\left(\tilde{A} - S_{22}^{-1}S_{21}\right)\right)\Delta}, \tag{7}$$

where S *has been partitioned as*

$$S=\begin{pmatrix} \underset{(n_1+1\times n_1+1)}{S_{11}} & S_{21}' \\ \underset{(n_2\times n_1+1)}{S_{21}} & \underset{(n_2\times n_2)}{S_{22}} \end{pmatrix}$$

and $S_{11.2}=S_{11}-S_{21}'S_{22}^{-1}S_{21}$. *The estimator of the identified parameters,* $\hat{\beta}_2$, *is*

(8)
$$\hat{\beta}_2=\tilde{A}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}+o_P(1),$$

where

(9)
$$\tilde{A}=\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\left(\tilde{\pi},\tilde{\Pi}_1\right).$$

The intuition behind Theorem 1 is as follows. Since $T^{-1}S\to^P \Omega=I_{n+1}$, we have

$$T^{-1}\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)\to^P\begin{pmatrix}1\\-\hat{\beta}_1\\-\hat{\beta}_2\end{pmatrix}'\Omega\begin{pmatrix}1\\-\hat{\beta}_1\\-\hat{\beta}_2\end{pmatrix}.$$

We can write the numerator of (6) as

(10)
$$\begin{aligned}\nu\left(\hat{\beta}_1,\hat{\beta}_2\right)=&\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}'\left(\tilde{\pi},\tilde{\Pi}_1\right)M_{\tilde{\Pi}_2}\left(\tilde{\pi},\tilde{\Pi}_1\right)\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}\\&+\left(\hat{\beta}_2-\tilde{A}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}\right)'\tilde{\Pi}_2'\tilde{\Pi}_2\left(\hat{\beta}_2-\tilde{A}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}\right).\end{aligned}$$

where $\tilde{A}$ has been defined in (9). The first term of (10) is

$$\left(\tilde{\pi},\tilde{\Pi}_1\right)'M_{\tilde{\Pi}_2}\left(\tilde{\pi},\tilde{\Pi}_1\right)=\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1\right)'M_{\tilde{\Pi}_2}\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1\right)=o_P(1),$$

and can be neglected so that the LIML estimator of the identified parameter β_2^* is approximately given by equation (8). If we insert equation (8) into (6) and maximize the resulting quantity with respect to $\hat{\beta}_1$ we see that the LIML estimator of β_1^* solves the minimization problem in equation (7).

Theorem 1 suggests an interpretation of the effect of partial lack of identification on the LIML estimator of β_2^*. We rewrite equation (8) as

$$\hat{\beta}_2=\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\tilde{\pi}-\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\tilde{\Pi}_1\hat{\beta}_1+o_P(1).$$

The first component $\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\tilde{\pi}$ is the TSLS estimator of β_2^* in the model where all endogenous variables with unidentified coefficients have been dropped,

(11) $$y = Y_2\beta_2^* + Z_1\gamma^* + u\ ,$$

with corresponding reduced form

(12) $$\left(y, Y_2\right) = Z_1\left(\phi_1, \Phi_2\right) + Z_2\left(\pi, \Pi_2\right) + \left(v, V_2\right).$$

The term $\left(\tilde{\Pi}_2{}'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2{}'\tilde{\pi}$ is consistent to β_2^* and has asymptotically a normal distribution. It is also an efficient estimator of β_2^* in the identified model above. The second component of $\hat{\beta}_2$, $\left(\tilde{\Pi}_2{}'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2{}'\tilde{\Pi}_1\hat{\beta}_1$, does not contain any useful information about β_2^* and captures the effect of the lack of identification of β_1^* on the LIML estimator of the identified parameters β_2^*. It is noise added to an asymptotically consistent and efficient estimator of β_2^*.

The LIML estimator of $\gamma^* = \left(\gamma_1^*{}', \gamma_2^*{}'\right)'$ is

$$\hat{\gamma} = \left(\hat{\phi}, \hat{\Phi}_1, \hat{\Phi}_2\right)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix}.$$

3. Distributional results

We shall study the asymptotic distribution of the LIML estimator for both the identified and the unidentified parameters under Assumptions 1 and 2.

Theorem 2. *(Coefficients of the endogenous variables) Suppose that Assumptions 1 and 2 hold. Then,*

(1) the estimator of the unidentified parameter, $\hat{\beta}_1$, *has the following asymptotic distribution*

$$\left(1 + \beta_2^*{}'\beta_2^*\right)^{-1/2}\hat{\beta}_1 \to^D C_{n_1}\ ,$$

where C_{n_1} *denotes a random vector having a multivariate Cauchy distribution in* $\mathbb{R}^{n_1}$*;*

(2) the estimator of the identified parameters satisfies:

$$\hat{\beta}_2 \to^P \beta_2^*\ ,$$

$$\frac{T^{1/2}\left(\hat{\beta}_2-\beta_2^*\right)}{\sqrt{1+\beta_2^*{}'\beta_2^*+\hat{\beta}_1{}'\hat{\beta}_1}}\to^D N\left(0,\left(\Pi_2{}'Q\Pi_2\right)^{-1}\right),$$

and

$$T^{1/2}\left(\hat{\beta}_2-\beta_2^*\right)\to^D \frac{\Gamma\left(\frac{n_1+1}{2}\right)}{\pi}\int N\left(0,\left(1+\beta_2^*{}'\beta_2^*\right)\left(1+C_{n_1}{}'C_{n_1}\right)\left(\Pi_2{}'Q\Pi_2\right)^{-1}\right)dC_{n_1}.$$

Theorem 3.1 contains results analogous to those of Phillips (1989) and Choi and Phillips (1992) for the TSLS estimator: the LIML estimator for the identified parameters, $\hat{\beta}_2$, is consistent to β_2^*, but the one for the unidentified parameters, $\hat{\beta}_1$, converges in distribution to a non-degenerate random vector.

The asymptotic distribution of $\hat{\beta}_1$ is proportional to a multivariate Cauchy distribution (with a coefficient of proportionality depending on β_2^*), so that it has no finite integer moments. Thus, the LIML seems to capture the uncertainty about the unidentified parameters more than the TSLS estimator because the latter is proportional to a multivariate t-distribution with k_2-n_1 degrees of freedom. If the model is totally unidentified (i.e. $\Pi_2=0$) then we obtain the asymptotic version of the standard result of Phillips (1984b) and Hillier and Skeels (1993): $\hat{\beta}\to^D C_n$, where C_n denotes the multivariate Cauchy distribution in $\mathbb{R}^n$.

The asymptotic distribution of $\hat{\beta}_2$ is covariance matrix mixture normal, and thus it is non-standard, indicating that lack of identification of β_1^* affects the LIML estimator of β_2^*. However, $\hat{\beta}_2$ has a normal asymptotic distribution if we condition on $\hat{\beta}_1$. If the model is identified (i.e. $n_1=0$), we obtain the well-known result that $T^{1/2}\left(\hat{\beta}-\beta^*\right)\to^D N\left(0,\left(1+\beta^*{}'\beta^*\right)\left(\Pi'Q\Pi\right)^{-1}\right)$.

In general one needs to rotate coordinates in the space of the endogenous variables to obtain the specification of the structural and reduced forms in equations (1) and (2). This means that the effect of partial identification will manifest itself in the original coordinates in the fact that the LIML, $\hat{\beta}$, has a non-standard, non-degenerate distribution with no finite integer moments.

We now turn to the coefficients of the exogenous variables.

Theorem 3. *(Coefficients of the exogenous variables) Suppose that assumptions 1 and 2 hold. Then,*

(1) the estimator of the unidentified parameter, $\hat{\gamma}_1$, has the following asymptotic distribution

$$\hat{\gamma}_1 \to^D \phi_1 - \Phi_{12}\beta_2^* - \left(1+\beta_2^{*\prime}\beta_2^*\right)^{1/2}\Phi_{11}C_{n_1},$$

where C_{n_1} denotes a random vector having a multivariate Cauchy distribution in $\mathbb{R}^{n_1}$;

(2) the estimator of the identified parameters, $\hat{\gamma}_2$, satisfies:

$$\hat{\gamma}_2 \to^P \gamma_2^\star,$$

$$\frac{T^{1/2}\left(\hat{\gamma}_2-\gamma_2^\star\right)}{\sqrt{1+\hat{\beta}_1'\hat{\beta}_1+\beta_2^{*\prime}\beta_2^*}} \to^D N\left(0, I_{k_{12}}+\Phi_{22}\left(\Pi_2'\Pi_2\right)^{-1}\Phi_{22}'\right),$$

and

$$T^{1/2}\left(\hat{\gamma}_2-\gamma_2^\star\right) \to^D \int N\left(0,\left(1+\beta_2^{\star\prime}\beta_2^\star\right)\left(1+C_{n_1}'C_{n_1}\right)\left(I_{k_{12}}+\Phi_{22}\left(\Pi_2'\Pi_2\right)^{-1}\Phi_{22}'\right)\right)dC_{n_1}.$$

The LIML estimator $\hat{\gamma}_2$ for γ_2^* is consistent but has a non-standard asymptotic distribution. The distribution of $\hat{\gamma}_2$ is a covariance matrix mixture normal, and has essentially the same structure as the distribution of $\hat{\beta}_2$. The LIML estimator of the unidentified coefficients γ_1^* converges to a non-degenerate distribution centred on the point $\phi_1 - \Phi_{12}\beta_2^*$ and has no finite integer moments.

The effect of partial identification on the LIML estimator of the coefficients of the exogenous variables in the original coordinates before structural and reduced form are transformed into (1) and (2) results in $\hat{\gamma}$ having a non-degenerate non-standard asymptotic distribution with no finite integer moments.

4. Weak instruments and the "moment problem"

We have shown that in partially identified models the LIML estimator of the unidentified parameters does not have moments even asymptotically. We now briefly discuss the weak instruments case.

When instruments are weak and $\Pi = T^{-1/2}C$ (e.g. Staiger and Stock (1997)), the LIML estimator is $\hat{\beta} = -\Delta_1^{-1}\Delta_2$ where $\Delta = (\Delta_1, \Delta_2')'$ is the eigenvector associated with the smallest eigenvalue of

$$\left(T^{-1}S\right)^{-1/2}\left(\tilde{\pi},\tilde{\Pi}\right)'\left(\tilde{\pi},\tilde{\Pi}\right)\left(T^{-1}S\right)^{-1/2} \to^{P} W_{n+1}\left(k_2 - n, I_{n+1}, \left(\beta^*, I_n\right)C'C\left(\beta^*, I_n\right)\right).$$

It follows from the continuous mapping theorem that the LIML estimator does not have finite moments even asymptotically since the leading term in the expansion of its asymptotic density is proportional to a multivariate Cauchy distribution in $\mathbb{R}^n$. Thus, the presence of weak instruments changes the asymptotic distribution of the LIML estimator by making it more similar to the small sample distribution of the LIMLK estimator under normality as studied by Phillips (1984b) and Phillips (1985). A modification of the analysis of the previous sections shows that the asymptotic moments of the LIML estimator do not exist for weakly identified parameters even if strongly identified parameters are present.

The lack of asymptotic moments for the LIML estimator under partial and weak identification is a remarkable property which has not been emphasised before in the literature: the precision of the LIML estimator measured by its asymptotic Fisher information matrix is zero, and the sample is not informative about the interest parameters. This implies that the mean squared error may not be a suitable tool for choosing instruments and comparing estimators even when the sample size is infinitely large.

5. Conclusions

This paper studies the asymptotic distribution theory for the LIML estimator in partially identified linear structural equations models. General formulae are given for the asymptotic distribution of the LIML estimator for the coefficient vectors of both the endogenous and exogenous variables. For the sake of simplicity, we assume that the structural parameters are in canonical form and that the rotations of coordinates in the space of endogenous and exogenous variables to separate identified and unidentified parameters have been carried out. Since these are affine transformations' the results for the unstandardized case follow easily.

The LIML estimators for the identified parameters are consistent but have non-standard asymptotic distributions expressed as covariance matrix mixed normals. These results are simpler than those for the TSLS estimator obtained by Phillips (1983) and Choi and Phillips (1992).

The LIML estimators for the unidentified parameters are obviously inconsistent, but have non-degenerate asymptotic distributions. We find that these are affine transformations of a random vector having a multivariate Cauchy distribution, and, consequently, they do not have any finite integer moments even asymptotically. This implies that the LIML estimator of the coefficients of both endogenous and exogenous variables does not have moments even asymptotically in partially identified linear structural equations which are not in canonical form or for which the identified and the unidentified parameters are not separated by rotations of coordinates.

Appendix: Proofs

The asymptotic properties of the statistics identified in Section 2 are described in the following lemma.

Lemma 1. If Assumptions 1 and 2 hold then

(i) $T^{-1}S \to^{P} \Omega = I_{n_1+n_2+1}$;

(ii) $\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right) \to^{P} \left(\Pi_2\beta_2^*,0,\Pi_2\right)$;

(iii) $T^{1/2}\left[\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)-\left(\Pi_2\beta_2^*,0,\Pi_2\right)\right] \to^{D} N\left(0,I_{k_2}\otimes I_{n_1+n_2+1}\right)$;

(iv) $\tilde{A}=\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\left(\tilde{\pi},\tilde{\Pi}_1\right) \to^{P} \left(\beta_2^*,0\right)$;

(v) $T^{1/2}\left(A-\left(\beta_2^*,0\right)\right) \to^{D} N\left(0,\left(\Pi_2'\Pi_2\right)^{-1}\otimes\Sigma\right)$ where

$$\Sigma=\begin{bmatrix}1+\beta_2^{*}{}'\beta_2^* & 0\\ 0 & I_{n_1}\end{bmatrix}; \tag{13}$$

(vi) $T\left(\tilde{\pi},\tilde{\Pi}_1\right)'M_{\tilde{\Pi}_2}\left(\tilde{\pi},\tilde{\Pi}_1\right) \to^{D} W_{n_1+1}\left(k_2-n_2,\Sigma\right)$ and it is asymptotically independent of $T^{1/2}\left(A-\left(\beta_2^*,0\right)\right)$;

(vii) $\hat{\Phi}=\left(\hat{\phi},\hat{\Phi}_1,\hat{\Phi}_2\right) \to^{P} \begin{pmatrix}\phi_1 & \Phi_{11} & \Phi_{12}\\ \phi_2 & 0 & \Phi_{22}\end{pmatrix}$;

(viii) $T^{1/2}\left(\left(\hat{\phi},\hat{\Phi}_1,\hat{\Phi}_2\right)-\begin{pmatrix}\phi_1 & \Phi_{11} & \Phi_{12}\\ \phi_2 & 0 & \Phi_{22}\end{pmatrix}\right)\to^D N\left(0,I_{k_1}^{-1}\otimes I_{n+1}\right)$, and it is independent of $T^{1/2}\left[\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)-\left(\Pi_2\beta_2^*,0,\Pi_2\right)\right]$.

Proof. Parts *(i)-(iv)* and *(vii)-(viii)* are standard results. We now prove parts *(v)* and (vi). We have

$$\begin{aligned}T^{1/2}\left(A-\left(\beta_2^*,0\right)\right)&=T^{1/2}\left(\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\left(\tilde{\pi},\tilde{\Pi}_1\right)\left(\tilde{\pi},\tilde{\Pi}_1\right)-\left(\beta_2^*,0\right)\right)\\&=\left[\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\right]T^{1/2}\left(\left(\tilde{\pi},\tilde{\Pi}_1\right)-\tilde{\Pi}_2\left(\beta_2^*,0\right)\right)\\&=\left[\left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\right]T^{1/2}\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1\right).\end{aligned}$$

Moreover, let $\tilde{C}$ be a matrix such that $\tilde{C}\tilde{C}'=M_{\tilde{\Pi}_2}$ and $\tilde{C}'\tilde{C}=I_{k-n_2}$ and note that $\tilde{\Pi}_2\to\tilde{C}$ is continuous. Also, by construction $\tilde{C}'\tilde{\Pi}_2=0$ so that $\tilde{C}'\left(\tilde{\pi},\tilde{\Pi}_1\right)=\tilde{C}'\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1\right)$. Part *(iii)* entails that

$$T^{1/2}\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1,\tilde{\Pi}_2-\Pi_2\right)\to^D N\left(0,I_k\otimes\begin{bmatrix}1+\beta_2^*{}'\beta_2^* & 0 & -\beta_2^*{}'\\ 0 & I_{n_1} & 0\\ -\beta_2^* & 0 & I_{n_2}\end{bmatrix}\right),$$

so we obtain

$$T^{1/2}\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1\right)\to^D N\left(0,I_k\otimes\Sigma\right).$$

The continuous mapping theorem implies that

$$\begin{aligned}&\begin{pmatrix}\tilde{C}'\\ \left(\tilde{\Pi}_2'\tilde{\Pi}_2\right)^{-1}\tilde{\Pi}_2'\end{pmatrix}T^{1/2}\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_1\right)\to^D\begin{pmatrix}C'\\ \left(\Pi_2'\Pi_2\right)^{-1}\Pi_2'\end{pmatrix}N\left(0,I_k\otimes\Sigma\right)\\&=N\left(0,\begin{pmatrix}I_{k-n_2} & 0\\ 0 & \left(\Pi_2'\Pi_2\right)^{-1}\end{pmatrix}\otimes\Sigma\right),\end{aligned}$$

so the random variates $T^{1/2}\left(A-\left(\beta_2^*,0\right)\right)$ and $T^{1/2}\tilde{C}'\left(\tilde{\pi}-\tilde{\Pi}_2\beta_2^*,\tilde{\Pi}_2\right)$ are asymptotically independent.

Proof of Theorem 1

The first order condition for a minimum of

(14) $$\nu\left(\hat{\beta}_1,\hat{\beta}_2\right)/\left[T^{-1}\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)\right]$$

is

(15) $$D_{\hat{\beta}_2}\frac{\nu\left(\hat{\beta}_1,\hat{\beta}_2\right)}{\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)}=\frac{1}{\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)}D_{\hat{\beta}_2}\nu\left(\hat{\beta}_1,\hat{\beta}_2\right)-\frac{1}{\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)}\frac{\nu\left(\hat{\beta}_1,\hat{\beta}_2\right)}{\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)}D_{\hat{\beta}_2}\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)=0,$$

where $\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)>0$ and

$$D_{\hat{\beta}_2}\nu\left(\hat{\beta}_1,\hat{\beta}_2\right)=2\tilde{\Pi}_2{}'\tilde{\Pi}_2\left(\hat{\beta}_2-\tilde{A}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}\right)$$

$$D_{\hat{\beta}_2}\delta\left(\hat{\beta}_1,\hat{\beta}_2\right)=2S_{22}\left(\hat{\beta}_2-S_{22}S_{21}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}\right).$$

Replacing these in (15) we obtain

$$\hat{\beta}_2=\tilde{A}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}-\left\{\min_{\beta_1,\beta_2}\frac{\nu\left(\beta_1,\beta_2\right)}{\delta\left(\beta_1,\beta_2\right)}\right\}\left(\tilde{\Pi}_2{}'\tilde{\Pi}_2\right)^{-1}S_{22}\left(\hat{\beta}_2-S_{22}S_{21}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}\right).$$

Since

(16) $$\min_{\beta_1,\beta_2}\frac{\nu\left(\beta_1,\beta_2\right)}{\delta\left(\beta_1,\beta_2\right)}=\min_{\Delta'\Delta=1}\frac{\Delta'\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)'\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)\Delta}{\Delta'S\Delta},$$

the right hand side of (16) is the smallest solution to the determinental equation

(17) $$\left|\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)'\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)-\lambda S\right|=0.$$

The smallest solution of (17) is a continuous function of $\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)'\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)$ and S, and equals zero if $\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)'\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)$ is a singular matrix. The first part of the theorem follows from the continuous mapping theorem and the fact that $\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)'\left(\tilde{\pi},\tilde{\Pi}_1,\tilde{\Pi}_2\right)$ converges in probability to a singular matrix. The second part of the theorem follows by inserting $\hat{\beta}_2=\tilde{A}\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}$ in (14) and minimizing it with respect to $\hat{\beta}_1$.

Proof of Theorem 2

(1) Theorem 1 shows that the LIML estimator for β_1 minimizes (7). Let

$$\tilde{Q} = \left(S_{11.2} + \left(\tilde{A} - S_{22}^{-1} S_{21} \right)' S_{22} \left(\tilde{A} - S_{22}^{-1} S_{21} \right) \right)^{1/2} \to^{D} \Sigma^{1/2}$$

where Σ is defined in (13), and let

$$\Delta = \tilde{Q} \tilde{\Delta}.$$

The problem of minimizing (13) can be written in terms of $\tilde{\Delta}$ as

$$\frac{\tilde{\Delta}' \left(\tilde{Q}^{-1} \right)' \left(\tilde{\pi}, \tilde{\Pi}_1 \right)' M_{\tilde{\Pi}_2} \left(\tilde{\pi}, \tilde{\Pi}_1 \right) \tilde{Q}^{-1} \tilde{\Delta}}{\tilde{\Delta}' \tilde{\Delta}}.$$

The continuous mapping theorem and Lemma 1 imply that

$$\left(\tilde{Q}^{-1} \right)' \left(\tilde{\pi}, \tilde{\Pi}_1 \right)' M_{\tilde{\Pi}_2} \left(\tilde{\pi}, \tilde{\Pi}_1 \right) \tilde{Q}^{-1} \to^{D} \Sigma^{-1/2} W_{n_1+1} \left(k - n_2, \Sigma \right) \Sigma^{-1/2} = W_{n_1+1} \left(k - n_2, I_{n_1+1} \right).$$

Let $\tilde{\Delta}_m$ be the eigenvector associated to the smallest eigenvalue of

$$\left(\tilde{Q}^{-1} \right)' \left(\tilde{\pi}, \tilde{\Pi}_1 \right)' M_{\tilde{\Pi}_2} \left(\tilde{\pi}, \tilde{\Pi}_1 \right) \tilde{Q}^{-1}.$$

It is well known that $\tilde{\Delta}_m$ tends in distribution to a random vector v uniformly distributed on the unit sphere in $\mathbb{R}^{n_1+1}$. The continuous mapping theorem implies that $\Delta \to^{D} \Sigma^{1/2} v$. It follows that $\hat{\beta}_1 = -\Delta_{m1}^{-1} \Delta_{m2}$ where $\Delta_m = \left(\Delta_{m1}, \Delta_{m2}' \right)'$ is the eigenvector associated with the smallest eigenvalue of (13) and tends in distribution to a $\left(1 + \beta_2^{*\prime} \beta_2^{*} \right)^{1/2} C_{n_1}$ where C_{n_1} has a Cauchy distribution in R^{n_1}.

(2.i) Since $\tilde{A} \to^{P} \left(\beta_2^{*}, 0 \right)$ (Lemma 1 (iii)) and $\hat{\beta}_1 \to^{D} \left(1 + \beta_2^{*\prime} \beta_2^{*} \right)^{1/2} C_{n_1}$ (Theorem 2), it follows from the continuous mapping theorem that

$$\hat{\beta}_2 = \tilde{A} \begin{pmatrix} 1 \\ -\hat{\beta}_1 \end{pmatrix} \to^{P} \left(\beta_2^{*}, 0 \right) \begin{pmatrix} 1 \\ -\left(1 + \beta_2^{*}{}' \beta_2^{*} \right)^{1/2} C_{n_1} \end{pmatrix} = \beta_2^{*}.$$

(2.ii) and (2.iii) Note that

$$T^{1/2} \left(A - \left(\beta_2^{*}, 0 \right) \right) \begin{pmatrix} 1 \\ -\hat{\beta}_1 \end{pmatrix} = T^{1/2} \left(\hat{\beta}_2 - \beta_2^{*} \right)$$

and

$$T^{1/2} \left(A - \left(\beta_2^{*}, 0 \right) \right) \to^{D} N \left(0, \left(\Pi_2' Q \Pi_2 \right)^{-1} \otimes \Sigma \right).$$

Moreover, $T^{1/2}\left(A-\left(\beta_2^*,0\right)\right)$ is asymptotically independent of $T\left(\tilde{\pi},\tilde{\Pi}_1\right)' M_{\tilde{\Pi}_2}\left(\tilde{\pi},\tilde{\Pi}_1\right)$ (i.e. of $\hat{\beta}_1$). Then

$$T^{1/2}\left(\hat{\beta}_2-\beta_2^*\right)|\,\hat{\beta}_1 \to^D N\left(0,\left(1+\beta_2^{*}{}'\beta_2^*+\hat{\beta}_1{}'\hat{\beta}_1\right)\left(\Pi_2{}'Q\Pi_2\right)^{-1}\right).$$

Proof of Theorem 3

We known from Lemma 1*(vii)* that

$$\hat{\Phi} \to^P \begin{pmatrix} \phi_1 & \Phi_{11} & \Phi_{12} \\ \phi_2 & 0 & \Phi_{22} \end{pmatrix}$$

so

$$\hat{\gamma}=\hat{\Phi}\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix} \to^P \begin{pmatrix} \phi_1 & \Phi_{11} & \Phi_{12} \\ \phi_2 & 0 & \Phi_{22} \end{pmatrix}\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\beta_2^* \end{pmatrix} = \begin{pmatrix} \phi_1-\Phi_{12}\beta_2^*-\Phi_{11}\hat{\beta}_1 \\ \phi_2-\Phi_{22}\beta_2^* \end{pmatrix} = \begin{pmatrix} \phi_1-\Phi_{12}\beta_2^*-\Phi_{11}\hat{\beta}_1 \\ \gamma_2^* \end{pmatrix},$$

So part (1) and consistency in part (2) are proved. To prove the rest of the theorem consider

$$\begin{aligned} T^{1/2}\left(\hat{\gamma}_2-\gamma_2^\star\right) &= T^{1/2}E_2\left(\hat{\Phi}-\begin{pmatrix} \phi_1 & \Phi_{11} & \Phi_{12} \\ \phi_2 & 0 & \Phi_{22} \end{pmatrix}\right)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix} - T^{1/2}\Phi_{22}\left(\hat{\beta}_2-\beta_2^*\right) \\ &= T^{1/2}\left(\hat{\Phi}_2-\begin{pmatrix} \phi_2 & 0 & \Phi_{22} \end{pmatrix}\right)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \\ -\hat{\beta}_2 \end{pmatrix} - T^{1/2}\Phi_{22}\left(\tilde{A}-\left(\beta_2^*,0\right)\right)\begin{pmatrix} 1 \\ -\hat{\beta}_1 \end{pmatrix} \end{aligned}$$

where $E_2=\left(0,I_{k_{12}}\right)$ be an $\left(k_{12}\times k_1\right)$ matrix and $\hat{\Phi}_2=E_2\hat{\Phi}$. It follows from Lemma 1*(viii)* that

$$T^{1/2}\left(\hat{\Phi}_2-\begin{pmatrix} \phi_2 & 0 & \Phi_{22} \end{pmatrix}\right) \to^D N\left(0,I_{k_{12}}\otimes I_{n+1}\right),$$

and this is independent of

$$T^{1/2}\left(A-\left(\beta_2^*,0\right)\right) \to^D N\left(0,\left(\Pi_2{}'\Pi_2\right)^{-1}\otimes\Sigma\right).$$

Therefore, the continuous mapping theorem implies that

$$T^{1/2}\left(\hat{\gamma}_2-\gamma_2^{\star}\right)\to^{D} N\left(0,I_{k_{12}}\otimes I_{n+1}\right)\begin{pmatrix}1\\-\hat{\beta}_1\\-\beta_2^{*}\end{pmatrix}-\Phi_{22}N\left(0,\left(\Pi_2{}'\Pi_2\right)^{-1}\otimes\Sigma\right)\begin{pmatrix}1\\-\hat{\beta}_1\end{pmatrix}$$

and the theorem follows easily.

References

Anderson, T. W., N. Kunitomo and T. Sawa (1982). "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator." *Econometrica* 50: 1009-1028.

Bekker, P. A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators." *Econometrica* 62: 657-681.

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

Chao, J. C. and N. R. Swanson (2005). "Consistent Estimation with a Large Number of Weak Instruments." *Econometrica* 73(5): 1673-1692.

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.

Fujikoshi, Y., K. Morimune, N. Kunitomo and M. Taniguchi (1982). "Asymptotic Expansions of the Distributions of the Estimates of Coefficients in a Simultaneous Equation System." *Journal of Econometrics* 18.

Han, C. and P. C. B. Phillips (2005). "GMM with Many Moment Conditions." *Econometrica* Forthcoming.

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. and C. L. Skeels (1993). "Some Further Exact Results for Structural Equation Estimators". *Models, Methods and Applications of Econometrics. Essays in Honor of A.R. Bergstroms*. P. C. B. Phillips. New York, Basil Blackwell**:** 117-139.

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. (1984a). "The Exact Distribution of Exogenous Variable Coefficient Estimators." *Journal of Econometrics* 26: 387-398.

Phillips, P. C. B. (1984b). "The Exact Distribution of LIML: I." *International Economic Review* 25: 249-261.

Phillips, P. C. B. (1985). "The Exact Distribution of the LIML: II." *International Economic Review* 26: 21-36.

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

Skeels, C. L. (1995). "Some Exact Results for Estimators of the Coefficients on the Exogenous Variables in a Single Equation." *Econometric Theory* 11: 484-497.

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