Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: https://www.tandfonline.com/loi/lecr20

LIML in the static linear panel data model

Tom Wansbeek & Dennis Prak

To cite this article: Tom Wansbeek & Dennis Prak (2017) LIML in the static linear panel data model, Econometric Reviews, 36:1-3, 385-395, DOI: 10.1080/07474938.2015.1114566

To link to this article: https://doi.org/10.1080/07474938.2015.1114566

ECONOMETRIC REVIEWS
2017, VOL. 36, NOS. 1–3, 385–395
http://dx.doi.org/10.1080/07474938.2015.1114566

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

LIML in the static linear panel data model

Tom Wansbeek and Dennis Prak

Faculty of Economics and Business, University of Groningen Groningen, The Netherlands

ABSTRACT
We consider the static linear panel data model with a single regressor. For this model, we derive the LIML estimator. We study the asymptotic behavior of this estimator under many-instruments asymptotics, by showing its consistency, deriving its asymptotic variance, and by presenting an estimator of the asymptotic variance that is consistent under many-instruments asymptotics. We briefly indicate the extension to the static panel data model with multiple regressors.

KEYWORDS
Bekker standard errors; LIML, panel data; many-instruments asymptotics; weak instruments

JEL CLASSIFICATION
C23; C26

1. Introduction

Regression with endogeneity is at the core of econometrics. Using instrumental variables is the standard approach. The expression "instrumental variables" is due to Reiersøl (1941). Theil (1953) developed "repeated least squares," pioneering more instruments than regressors. This later on became known as two-stage least squares (2SLS), which is now a standard tool.

Just before the onset of 2SLS, Anderson and Rubin (1949, 1950) introduced the limited-information maximum likelihood (LIML) estimator as a way to deal with endogeneity. Its small-sample qualities vis-à-vis 2SLS got appreciated in the fifties in the development of k-class estimators like LIML (Nagar, 1959), but applied researchers showed little eagerness and LIML vanished from sight.

A revival of the LIML estimator had to wait till the development of "many-instruments asymptotics." The argument, adapted from Wansbeek and Meijer (2000), is as follows. The LIML estimator is the maximum likelihood (ML) estimator of β in the model $y = X\beta + u$ and $X = Z\Pi + V$, with Z of order $N \times h$ exogenous instruments. The errors are independent and identically distributed (i.i.d.) normal. Let, writing $P_C = C(C'C)^{-1}C'$ and $M_C = I - P_C$ for any C of full column rank,

$$S_P \equiv (y, X)'P_Z(y, X),$$
$$S_M \equiv (y, X)'M_Z(y, X).$$

The LIML estimator $\hat{\beta}$ follows from maximizing the likelihood. After concentrating out the nuisance parameters, Π, and the joint covariance matrix of u and v, the first-order condition is

$$\left(S_P - \hat{\lambda} S_M\right)\begin{pmatrix} 1 \\ -\hat{\beta} \end{pmatrix} = 0, \tag{1}$$

with $\hat{\lambda}$ the smallest value for which $S_P - \hat{\lambda} S_M$ is singular. Let $\Sigma_P \equiv \mathrm{E}(S_P)$ and $\Sigma_M \equiv \mathrm{E}(S_M)$, and let $\lambda \equiv h/(N-h)$. If a and b are random $N \times 1$ with $\mathrm{E}(ab') = \theta I_N$ for some θ, then $\mathrm{E}(a'P_Z b) = h\theta$ and $\mathrm{E}(a'M_Z b) = (N-h)\theta$. Hence $\mathrm{E}\left(a'(P_Z - \lambda M_Z)b\right) = 0$ and

$$(\Sigma_P - \lambda\Sigma_M)\begin{pmatrix} 1 \\ -\beta \end{pmatrix} = \mathrm{E}\left((y, X)'(P_Z - \lambda M_Z)(y, X)\right)\begin{pmatrix} 1 \\ -\beta \end{pmatrix} = \mathrm{E}\left((y, X)'(P_Z - \lambda M_Z)u\right) = 0. \tag{2}$$

CONTACT Tom Wansbeek t.j.wansbeek@rug.nl Faculty of Economics and Business, P. O. Box 800, University of Groningen, AV Groningen 9700, The Netherlands.

So, the LIML estimator satisfies a relation, (1), that is the sample analog of a major model implication, (2). As Bekker (1994), Section 3, extensively argues, this observation has an implication when studying the asymptotic distribution of the LIML estimator. In general, an asymptotic distribution is based on parameter sequences. Their choice should be motivated by the quality of the approximation that the asymptotic distribution provides to the exact distribution of the estimators. According to Bekker (1994), "the sequence should be directed so that it generates acceptable approximations of known distributional properties of related statistics. When an alternative sequence provides better approximations to these known properties, it can also be expected to provide better approximations to the distributions of the estimators p.661." As statistics related to the estimators of β, Bekker (1994) considers the matrices P_Z and M_Z; known distributional properties of them are given by their expectations. This suggests to study an asymptotic sequence where $\hat{\lambda}$ does not vanish, i.e., where the number of instruments grows along with the number of observations. This motivates the use of LIML with many-instruments asymptotics.

The situation is different for the 2SLS estimator, $\tilde{\beta} = (X'P_ZX)^{-1}X'P_Zy$. A little bit of algebra shows that it satisfies

$$\left(S_P - \tilde{\lambda}\, e_1e_1'\right)\begin{pmatrix} 1 \\ -\tilde{\beta} \end{pmatrix} = 0, \tag{3}$$

with e_1 the first unit vector and $\tilde{\lambda} \equiv (y - X\tilde{\beta})'P_Z(y - X\tilde{\beta})$. The result of taking the expectation in (3) is at variance with (2), and suggests a worse performance of 2SLS relative to LIML. Evidently, the difference between 2SLS and LIML is small when $\lambda \approx 0$, so when N is large relative to the number of instruments, or when $\Sigma_M \approx c \cdot e_1e_1'$, so when the instruments are not weak and explain the regressors well. But when there are many instruments or when the instruments are weak, LIML should perform better.

Bekker (1994) was the first to fully describe such alternative asymptotics, now usually called many-instruments asymptotics. He also gave a consistent estimator of the alternative asymptotic variance, offering an inference procedure with often much better coverage rates than 2SLS. In his words, this was "a remarkable result with practical implications." The result is often known as "Bekker standard errors."

To the best of our knowledge, LIML estimators have not yet been developed for panel data models, at least not in the sense of estimators obtained from maximizing the likelihood. LIML-like estimators have been developed by, e.g., Alvarez and Arellano (2003), Akashi and Kunitomo (2012), and Moral-Benito (2013) for the dynamic panel data model, but these are least-variance ratio estimators and not "true" LIML estimators obtained from maximizing a likelihood function. So, there appears to be a gap in the literature. The objective of our study is to fill this gap by deriving the LIML estimator for the static linear panel data model and investigating its properties, in a framework of many-instruments asymptotics since the *raison d'être* of LIML lies there.

The article is organized as follows. We formulate our model in Section 2. We indicate what we mean by "limited information." The loglikelihood is formulated and maximized over the parameter space. This yields the panel LIML estimator. In Section 3 we define many-instruments asymptotics, and show that our estimator is consistent under such asymptotics. We derive the asymptotic variance, and present an estimator of this variance that is consistent under many-instruments asymptotics, leading to the so-called "Bekker standard errors." Section 4 deals with the static panel data model with multiple regressors, which appears to be relatively straightforward. In Section 5 we present some Monte Carlo results on the median bias and the coverage rate of the panel LIML estimator relative to the 2SLS estimator. Section 6 concludes.

2. LIML in the panel data model

In this section we consider the case of a single regressor since this captures the essential elements, and the notation can be kept simple. The extension to the the static panel data model with multiple regressors is dealt with in Section 4. We first formulate the model and derive the loglikelihood. We next maximize the loglikelihood to arrive at the panel LIML estimator.

2.1. The model and the loglikelihood

As in the case of a single cross-section, the model is given by two equations, one containing the structural relation under study and the second one relating the endogenous regressor to exogenous instruments:

$$y_n = \beta\, x_n + u_n, \tag{4}$$
$$x_n = \Pi' z_n + v_n, \tag{5}$$

for $n = 1, \ldots, N$, where x_n, y_n, u_n, and v_n are T-vectors and z_n is an h-vector; Π is of order $h \times T$. In matrix form, with X, Y, U, and V $(N \times T)$ and Z $(N \times h)$, we get for (4) $Y = \beta X + U$ and for (5) $X = Z\Pi + V$. The error terms satisfy

$$e_n \equiv \begin{pmatrix} u_n \\ v_n \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Omega), \qquad \Omega = \begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix}.$$

The estimation method is limited-information in the sense that we do not exploit any structure that may be present in Π (e.g., due to wave-specific instruments) or in Ω (e.g., due to random effects). When a specification with fixed-effects is deemed appropriate, we assume that they have been eliminated by an appropriate data transformation like taking first differences or the within-transformation, where one wave of the panel has to be discarded to avoid singularity of Ω.

We now turn to inference in this model. Let $S_e \equiv \sum_n e_n e_n'/N$. The logarithm of the joint density of the e_n is, apart from constants,

$$\log f = \log|\Omega| + \operatorname{tr}(\Omega^{-1} S_e).$$

Substitution of $y_n - \beta x_n$ for u_n and $x_n - \Pi' z_n$ for v_n turns this into the logarithm of the likelihood, $\mathcal{L}$, as the Jacobian of the transformation from u_n and v_n to y_n and x_n is 1. We maximize the likelihood in a few steps, using the same symbol $\mathcal{L}$ throughout and neglecting constants.

2.2. Maximizing the loglikelihood

First, we concentrate out Ω. As $\partial \log\mathcal{L}/\partial\Omega = \Omega^{-1} - \Omega^{-1} S_e \Omega^{-1}$, the optimal value for Ω is $\hat{\Omega} = S_e$, and the concentrated loglikelihood can be simplified to

$$\mathcal{L} = |S_e| = |(U, V)'(U, V)| = |U'U|\;|V'M_U V|, \tag{6}$$

using the formula for the determinant of a partitioned matrix. We next concentrate out Π, using $M_{(U,Z)} = M_U - M_U Z(Z'M_U Z)^{-1} Z' M_U$ and $\widetilde{\Pi} \equiv (Z'M_U Z)^{-1} Z' M_U X$ so $\widetilde{\Pi} - \Pi = (Z'M_U Z)^{-1} Z' M_U V$, to obtain

$$\begin{aligned} V'M_U V &= V'M_{(U,Z)} V + V'M_U Z(Z'M_U Z)^{-1} Z'M_U V \\ &= X'M_{(U,Z)} X + (\widetilde{\Pi} - \Pi)' Z'M_U Z(\widetilde{\Pi} - \Pi). \end{aligned}$$

On taking $\Pi = \widetilde{\Pi}$ in (6), we get in the optimum $\mathcal{L} = |U'U|\;|X'M_{(U,Z)} X|$, depending only on β. Applying the formula for the determinant of a partitioned matrix in two ways to $\begin{pmatrix} X'M_Z X & X'M_Z U \\ U'M_Z X & U'M_Z U \end{pmatrix}$ gives

$$|X'M_Z X|\;|U'M_{(X,Z)} U| = |U'M_Z U|\;|X'M_{(U,Z)} X|,$$

where $X'M_Z X$ does not depend on β; neither does $U'M_{(X,Z)} U$ as it is equal to $Y'M_{(X,Z)} Y$. So

$$\mathcal{L} = |U'U|\;|X'M_{(U,Z)} X| \propto \frac{|U'U|}{|U'M_Z U|} \qquad \text{or} \qquad \log\mathcal{L} = \log|U'U| - \log|U'M_Z U|. \tag{7}$$

Using the general result, for Ψ a positive-definite matrix and θ a scalar parameter,

$$\frac{\partial \log|\Psi|}{\partial\theta} = \operatorname{tr}\left(\Psi^{-1} \frac{\partial\Psi}{\partial\theta}\right),$$

we can differentiate $\log\mathcal{L}$ with respect to β and set the result equal to zero. This gives

$$\hat{\beta} = \frac{\text{tr}[(\widehat{U}'\widehat{U})^{-1}Y'X - (\widehat{U}'M_Z\widehat{U})^{-1}Y'M_ZX]}{\text{tr}[(\widehat{U}'\widehat{U})^{-1}X'X - (\widehat{U}'M_Z\widehat{U})^{-1}X'M_ZX]}, \tag{8}$$

with $\widehat{U} \equiv Y - X\hat{\beta}$; hence this equation is nonlinear in $\hat{\beta}$. Solving it is numerically easy by substitution of the 2SLS estimator of β in $\widehat{U}$,

$$\hat{\beta}_{2\text{SLS}} = \frac{\text{tr}[Y'P_ZX]}{\text{tr}[X'P_ZX]},$$

yielding a new value of $\hat{\beta}$, and iteration until convergence. In the simulations below, iteration took but a few steps, nearly always less than ten, and in most cases less than five. Notice that the panel LIML estimator is not the solution to an eigenequation, unlike in the case of a single cross-section, $T = 1$.

3. Asymptotic properties

In this section we study the asymptotic properties of the panel LIML estimator. The interesting and relevant case is the one of many-instruments asymptotics. We first indicate what we mean by that, and present the simple calculus implied. We next prove the consistency of the panel LIML estimator under many-instruments asymptotics. We conclude this section by deriving the asymptotic variance and present a consistent estimator of this variance. Derivations for the cross-sectional multiple-regressors case have been given by Bekker (1994) and Newey (2004).

3.1. Many-instruments asymptotics

We will consider the properties of the panel LIML estimator under asymptotics characterized by

$$N \to \infty, \qquad \frac{h}{N} \to \alpha, \qquad \frac{1}{N}\Pi'Z'Z\Pi \to Q \geq 0.$$

This means that the explanatory power of the instruments remains constant as $N \to \infty$ although the number of instruments grows along with N. We indicate many-instruments asymptotics by an asterisk to distinguish it from the usual asymptotics with $N \to \infty$ only. For example, with

$$\text{E}\left(X'M_ZU\right) = \text{E}\left(V'M_ZU\right) = \text{E}\sum_n v_n(M_Z)_{nn}u_n' = \sum_n (M_Z)_{nn}\Omega_{vu} = \text{tr}(M_Z)\Omega_{vu} = (N-h)\Omega_{vu},$$

we have $\text{plim}_{N\to\infty} X'M_ZU/N = \Omega_{vu}$ with the usual large-N asymptotics. But with many-instruments asymptotics, we have $\text{tr}(M_Z)/N = 1 - h/N \to 1 - \alpha$, so

$$\text{plim}^* \frac{1}{N}X'M_ZU = (1-\alpha)\,\Omega_{vu}.$$

Along these lines,

$$\frac{1}{N}\begin{pmatrix} U' \\ X' \end{pmatrix}(U, X) \xrightarrow{p^*} \begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix}, \tag{9}$$

$$\frac{1}{N}\begin{pmatrix} U' \\ X' \end{pmatrix}M_Z(U, X) \xrightarrow{p^*} (1-\alpha)\begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix}. \tag{10}$$

This directly yields

$$A \equiv (U'U)^{-1}U'X - (U'M_ZU)^{-1}U'M_ZX \xrightarrow{p^*} 0, \tag{11}$$

$$B \equiv (U'U)^{-1}X'X - (U'M_ZU)^{-1}X'M_ZX \xrightarrow{p^*} \Omega_{uu}^{-1}Q. \tag{12}$$

We are now in a position to study the asymptotic properties of the panel LIML estimator.

3.2. Consistency

Consistency can be shown by adapting the approach of Newey (2004) for the single-equation case to the case of panel data. The essential step is to show that $\mathcal{L} = |U'U|/|U'M_Z U|$ converges to a function of β with a unique minimum in the true value β_0, say. With $U = Y - \beta X = Y - \beta_0 X - (\beta - \beta_0)X$ we get

$$\Psi \equiv \text{plim}^* \frac{1}{N} U'U = \Omega_{uu} - (\beta - \beta_0)(\Omega_{uv} + \Omega_{vu}) + (\beta - \beta_0)^2 (Q + \Omega_{vv})$$

$$\Psi_\perp \equiv \text{plim}^* \frac{1}{N} U'M_Z U = (1 - \alpha)\left[\Omega_{uu} - (\beta - \beta_0)(\Omega_{uv} + \Omega_{vu}) + (\beta - \beta_0)^2 \Omega_{vv}\right].$$

So $\Psi_\perp = (1 - \alpha)\left[\Psi - (\beta - \beta_0)^2 Q\right]$, hence

$$\text{plim}^* \mathcal{L} = \text{plim}^* \frac{|U'U|}{|U'M_Z U|} = \frac{|\Psi|}{|\Psi_\perp|} = \frac{(1-\alpha)^T}{|I_T - \Psi^{-\frac{1}{2}} Q \Psi^{-\frac{1}{2}} (\beta - \beta_0)^2|}.$$

This function has a unique minimum in $\beta = \beta_0$.

It is of some interest to compare this with the behavior of the 2SLS estimator,

$$\tilde{\beta} = \frac{\text{tr}(X'P_Z Y)}{\text{tr}(X'P_Z X)} = \beta + \frac{\text{tr}(X'P_Z U)}{\text{tr}(X'P_Z X)}.$$

From (9) and (10), we obtain

$$\text{plim}^* \frac{1}{N} X'P_Z U = \alpha\ \Omega_{vu},$$

$$\text{plim}^* \frac{1}{N} X'P_Z X = Q + \alpha\ \Omega_{vv}.$$

So with many-instruments asymptotics the 2SLS estimator is inconsistent,

$$\text{plim}^* \tilde{\beta} = \beta + \frac{\alpha\ \text{tr}(\Omega_{vu})}{\text{tr}(Q) + \alpha\ \text{tr}(\Omega_{vv})} \neq \beta. \tag{13}$$

With large-N asymptotics, $\alpha = 0$, and the 2SLS estimator is consistent, the textbook case.

3.3. Asymptotic variance

Consider the (infeasible) estimator

$$\tilde{\beta} = \frac{\text{tr}\left[(U'U)^{-1} Y'X - (U'M_Z U)^{-1} Y'M_Z X\right]}{\text{tr}\left[(U'U)^{-1} X'X - (U'M_Z U)^{-1} X'M_Z X\right]} = \frac{\text{tr}\,[A]}{\text{tr}\,[B]}.$$

It has the same asymptotic variance as $\hat{\beta}$, so

$$V(\hat{\beta}) = \text{plim}^* N(\tilde{\beta} - \beta)^2 = \frac{\text{plim}^* N\,[\text{tr}(A)]^2}{\left[\text{plim}^* \text{tr}(B)\right]^2}. \tag{14}$$

As to A,

$$\begin{aligned} A &= (U'U)^{-1} U'X - (U'M_Z U)^{-1} U'M_Z X \\ &= (U'U)^{-1} U'Z\Pi + (U'U)^{-1} U'V - (U'M_Z U)^{-1} U'M_Z V \\ &= (U'U)^{-1} U'Z\Pi + (U'U)^{-1} U'\tilde{V} - (U'M_Z U)^{-1} U'M_Z \tilde{V} \end{aligned}$$

for any $\tilde{V}$ and Γ with $\tilde{V} = V + U\Gamma$. We choose $\Gamma = -\Omega_{uu}^{-1}\Omega_{uv}$, cf. Nagar (1959), which makes U and $\tilde{V}$ independent. There holds

$$\frac{1}{N} \text{E}(\tilde{V}'\tilde{V}) = \Omega_{vv \cdot u} \equiv \Omega_{vv} - \Omega_{vu}\Omega_{uu}^{-1}\Omega_{uv}.$$

Next, let

$$q \equiv N \begin{pmatrix} \mathrm{vec}(\widehat{U}'\widehat{U})^{-1} \\ \mathrm{vec}(\widehat{U}'\widehat{U})^{-1} \\ \mathrm{vec}(\widehat{U}'M_Z\widehat{U})^{-1} \end{pmatrix} \xrightarrow{p^*} \begin{pmatrix} 1 \\ 1 \\ \frac{1}{1-\alpha} \end{pmatrix} \otimes \mathrm{vec}(\Omega_{uu}^{-1})$$

$$u \equiv \mathrm{vec}(U')$$

$$d \equiv \begin{pmatrix} \Pi'Z' \otimes I_T \\ \tilde{V}' \otimes I_T \\ -\tilde{V}'M_Z \otimes I_T \end{pmatrix} u.$$

Then $\mathrm{E}(uu') = I_N \otimes \Omega_{uu}$ and

$$\lim{}^* \frac{1}{N}\mathrm{E}(dd') = \begin{pmatrix} Q & 0 & 0 \\ 0 & \Omega_{vv\cdot u} & -(1-\alpha)\,\Omega_{vv\cdot u} \\ 0 & -(1-\alpha)\,\Omega_{vv\cdot u} & (1-\alpha)\,\Omega_{vv\cdot u} \end{pmatrix} \otimes \Omega_{uu}.$$

Using

$$(1, \tfrac{1}{1-\alpha}) \begin{pmatrix} 1 & -(1-\alpha) \\ -(1-\alpha) & 1-\alpha \end{pmatrix} \begin{pmatrix} 1 \\ \frac{1}{1-\alpha} \end{pmatrix} = \tfrac{\alpha}{1-\alpha},$$

we obtain, since $\mathrm{tr}(\hat{A}) = \frac{1}{N}q'd$,

$$\begin{aligned} \mathrm{plim}^*\, N\left[\mathrm{tr}(\hat{A})\right]^2 &= (\mathrm{plim}^*\, q)' \left\{\lim{}^* \frac{1}{N}\mathrm{E}(dd')\right\} (\mathrm{plim}^*\, q) \\ &= \mathrm{tr}\,[\Omega_{uu}^{-1}(Q + \tfrac{\alpha}{1-\alpha}\Omega_{vv\cdot u})]. \end{aligned} \tag{15}$$

So the asymptotic variance of $\hat{\beta}$ is given by (14), with (12) and (15) inserted, thus generalizing the result for the cross-sectional case as given by Newey (2004).

3.4. Bekker standard errors

In order to make this formula operational and obtain Bekker standard errors, it remains to find an estimator $\hat{V}(\hat{\beta})$ of $V(\hat{\beta})$ that is consistent under many-instruments asymptotics. With $\hat{\alpha} = h/N$ and

$$\begin{aligned} H &\equiv (1-\hat{\alpha})\,P_Z - \hat{\alpha}\,M_Z \\ W &\equiv (1-\hat{\alpha})^2\,P_Z + \hat{\alpha}^2\,M_Z - \hat{\alpha}(1-\hat{\alpha})\,P_{\widehat{U}}, \end{aligned}$$

one such expression is

$$\hat{V}(\hat{\beta}) = \frac{\mathrm{tr}[(\widehat{U}'\widehat{U})^{-1}X'WX]}{[\mathrm{tr}((\widehat{U}'\widehat{U})^{-1}X'HX)]^2}. \tag{16}$$

By way of comparison, the usual estimator for the variance of the 2SLS estimator $\tilde{\beta}$ can be written as

$$\hat{V}(\tilde{\beta}) = \frac{1}{\mathrm{tr}[(\widehat{U}'\widehat{U})^{-1}X'P_ZX]}. \tag{17}$$

The consistency of (16) under many-instruments asymptotics follows from

$$\begin{aligned} \frac{1}{N}X'P_ZX &\xrightarrow{p^*} Q + \alpha\,\Omega_{vv}, \\ \frac{1}{N}X'M_ZX &\xrightarrow{p^*} (1-\alpha)\,\Omega_{vv}, \\ \frac{1}{N}X'P_{\widehat{U}}X &\xrightarrow{p^*} \Omega_{vu}\Omega_{uu}^{-1}\Omega_{uv}. \end{aligned}$$

On substitution, this readily gives

$$
\begin{aligned}
\frac{1}{N} X' H X &\xrightarrow{p^*} (1-\alpha)\,(Q+\alpha\,\Omega_{vv}) - \alpha(1-\alpha)\,\Omega_{vv} \\
&= (1-\alpha)\,Q, \\
\frac{1}{N} X' W X &\xrightarrow{p^*} (1-\alpha)^2\,(Q+\alpha\,\Omega_{vv}) + \alpha^2(1-\alpha)\,\Omega_{vv} - \alpha(1-\alpha)\,\Omega_{vu}\Omega_{uu}^{-1}\Omega_{uv} \\
&= (1-\alpha)^2\left\{Q+\alpha\,\Omega_{vv}+\tfrac{\alpha^2}{1-\alpha}\,\Omega_{vv}-\tfrac{\alpha}{1-\alpha}\,\Omega_{vu}\Omega_{uu}^{-1}\Omega_{uv}\right\} \\
&= (1-\alpha)^2\left\{Q+\tfrac{\alpha}{1-\alpha}\,\Omega_{vv\cdot u}\right\},
\end{aligned}
$$

from which the many-instruments consistency of $\hat{V}(\hat{\beta})$ in (16) follows directly.

4. Multiple regression

Up till now, we considered the case of a single regressor. We presently turn to the case where there are K regressors, all related to the same set of instruments. Model (4)–(5) then generalizes to

$$
\begin{aligned}
y_n &= x_{1n}\beta_1 + \ldots + x_{Kn}\beta_K + u_n, \\
x_{1n} &= \Pi_1' z_n + v_{1n}, \\
&\;\;\vdots \\
x_{Kn} &= \Pi_K' z_n + v_{Kn}.
\end{aligned}
$$

With $\beta \equiv (\beta_1,\ldots,\beta_K)', X \equiv (X_1,\ldots,X_K), E \equiv (E_1,\ldots,E_K), \Pi \equiv (\Pi_1,\ldots,\Pi_K), V \equiv (V_1,\ldots,V_K)$, we get for all n

$$
\begin{aligned}
Y &= X(\beta \otimes I_T) + U, \\
X &= Z\Pi + V.
\end{aligned}
$$

The notion of limited information is stretched as, for each n, *all* elements of X and *all* elements of Z are related, over regressors and over time.

The derivation of the LIML estimator above for the single-regressor case is hardly affected when there are multiple regressors. This is evident from the algebra from Section 2.2; we only need to read $U = Y - X(\beta \otimes I_T)$ rather than $U = Y - \beta\, X$. So the panel LIML estimator is the solution of

$$\hat{\beta} = \widehat{G}^{-1}\hat{g},$$

where

$$
\begin{aligned}
(\widehat{G})_{k\ell} &\equiv \operatorname{tr}[(\widehat{U}'\widehat{U})^{-1}X_k'X_\ell - (\widehat{U}'M_Z\widehat{U})^{-1}X_k'M_ZX_\ell], \\
(\hat{g})_k &\equiv \operatorname{tr}[(\widehat{U}'\widehat{U})^{-1}Y'X_k - (\widehat{U}'M_Z\widehat{U})^{-1}Y'M_ZX_k],
\end{aligned}
$$

for $k, \ell = 1,\ldots,K$. The generalization carries through in a straightforward way all the way to Bekker standard errors; (16) generalizes to

$$\hat{V}(\hat{\beta}) = \widetilde{A}^{-1}\widetilde{B}\widetilde{A}^{-1},$$

with

$$
\begin{aligned}
(\widetilde{A})_{k\ell} &= \operatorname{tr}[(\widehat{U}'\widehat{U})^{-1}X_k'HX_\ell], \\
(\widetilde{B})_{k\ell} &= \operatorname{tr}[(\widehat{U}'\widehat{U})^{-1}X_k'WX_\ell],
\end{aligned}
$$

with H and W as in (16).

This extension covers the case where there are multiple endogenous variables. When there are, in addition, multiple exogenous variables, the derivation is highly similar. Again, the algebra in Section 2.2 remains the core, and the only adaptation again is a further redefinition of U to include the exogenous regressors and their coefficients.

5. A Monte Carlo study

In this section, we examine by means of a Monte Carlo experiment the quality of asymptotic inference based on LIML with Bekker standard errors. We first describe the design used and then report the results.

5.1. Design

The design that we employ extends, to the case of $T = 2$, the design introduced by Bekker and Wansbeek (2014) for the case of a single cross-section:

$$\begin{pmatrix} y_{n1} \\ y_{n2} \end{pmatrix} = \begin{pmatrix} x_{n1} \\ x_{n2} \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_{n1} \\ \varepsilon_{n2} \end{pmatrix},$$
$$\begin{pmatrix} x_{n1} \\ x_{n2} \end{pmatrix} = \Pi' z_n + \omega \begin{pmatrix} \varepsilon_{n1} \\ \varepsilon_{n2} \end{pmatrix} + \begin{pmatrix} \tilde{v}_{n1} \\ \tilde{v}_{n2} \end{pmatrix}.$$

Since the number of instruments is h, the parameter matrix Π is of order $h \times 2$. The number of observations is $N = 500$, all variables have mean zero, and we further make the following choices. We let $\beta = 1$ and

$$\Pi = \begin{pmatrix} \pi & 0 & \cdots & 0 \\ \pi & 0 & \cdots & 0 \end{pmatrix}'.$$

The random vectors $(\varepsilon_{n1}, \varepsilon_{n2})'$ and $(\tilde{v}_{n1}, \tilde{v}_{n2})'$ are i.i.d. $N(0, I_2)$. The elements of z_n are i.i.d. $N(0, 1)$.

The implications of this design are as follows. The covariance matrix of the error terms is

$$\Omega = \begin{pmatrix} 1 & \omega \\ \omega & \omega^2 + 1 \end{pmatrix} \otimes I_2,$$

and $Q = \pi^2 \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. The OLS estimator of β converges to $\omega/(\pi^2 + \omega^2 + 1)$, and from (13) we see that the bias of 2SLS under many-instruments asymptotics is $\alpha\omega/[\pi^2 + \alpha(\omega^2 + 1)]$.

In our simulations, we are interested in the effect of the following issues:

- The number of instruments. We let $h = 10$ and $h = 30$. With $N = 500$, this means $\alpha = 0.02$ and $\alpha = 0.06$, respectively.
- The degree of endogeneity. This is driven by ω, where $\omega = 0$ corresponds with absence of endogeneity. We let $\omega = 0.5$ and $\omega = 2$.
- The strength of the instruments. This is measured by the F value of the regression of x on z. We let $F = 3$, $F = 5$, and $F = 10$, obtained by appropriate choices of π; in the population, there holds for the regression of x on z

$$R^2 = \frac{\pi^2}{\pi^2 + \omega^2 + 1}.$$

The way π drives F then follows directly from $F = (N - h)R^2/h(1 - R^2)$.

We consider two estimators, LIML and 2SLS. For each parameter configuration, we report for LIML the median bias (in absolute value) and two figures for the coverage rate, one based on many-instruments asymptotics and one on large-N asymptotics. For 2SLS, we report the absolute median bias and just a single figure for the coverage rate, based on the textbook expression for its asymptotic variance. We use the median rather than the mean since the mean of the LIML estimator does not exist in the cross-sectional case, and we conjecture that this also holds in the panel case. A comparison of the two figures for the coverage rate with LIML allows us to disentangle the first-order effect from the second-order effect: do the results for LIML differ from those for 2SLS due to a difference in centering or to the use of a different expression for the standard errors? In any case, the coverage rate is computed as the fraction of cases (in 50,000 replications) where the computed confidence interval contained the true value.

5.2. Results

The upper half of Table 1 presents the results for the median bias, in absolute value. As expected, it shows better performance of LIML relative to 2SLS when there are many instruments (right half) or when the instruments are weak (low F).

The lower half of the table presents the results for the 95% coverage rates. The performance of LIML is almost perfect, across all parameter combinations. By contrast, the performance of 2SLS ranges from reasonably good (90%) to extremely poor (4%); when there are many instruments that hardly explain the regressor, 2SLS cannot be trusted for inferential purposes.

The figures in italic are the coverage rates for LIML based on the textbook expression for the standard error. The results are between the LIML figures with many-instruments asymptotics and the figures for 2SLS. Overall, they are quite acceptable, suggesting that the better performance of LIML over 2SLS is primarily due to better centering, and less to better variance estimation.

The results were derived with data generated from the normal distribution, on which LIML is based, unlike 2SLS. Hence the results might be biased in a favor of LIML. To investigate this conjecture, we repeated the above analysis with data generated from the t-distribution, in order to get more mass in

Table 1. Median bias and coverage rate of LIML (in bold) and 2SLS, with data generated from the normal distribution; in italics, the coverage rate of LIML based on large-N asymptotics.

	$h = 10, \alpha = 0.02$				$h = 30, \alpha = 0.06$			
	$\omega = 0.5$		$\omega = 2$		$\omega = 0.5$		$\omega = 2$	
				abs. median bias ×1000				
$F = 3$	**0**	94	**1**	96	**1**	95	**1**	94
$F = 5$	**0**	63	**1**	63	**1**	63	**0**	63
$F = 10$	**0**	34	**0**	34	**0**	34	**0**	34
				95% coverage rate				
$F = 3$	**96**	81	**95**	38	**95**	59	**95**	4
	88		*90*		*87*		*91*	
$F = 5$	**95**	86	**95**	56	**95**	70	**95**	17
	91		*92*		*91*		*92*	
$F = 10$	**95**	90	**95**	74	**95**	81	**95**	43
	93		*93*		*93*		*94*	

Table 2. Median bias and coverage rate of LIML (in bold) and 2SLS, with data generated from the t distribution; in italics, the coverage rate of LIML based on large-N asymptotics.

	$h = 10, \alpha = 0.02$				$h = 30, \alpha = 0.06$			
	$\omega = 0.5$		$\omega = 2$		$\omega = 0.5$		$\omega = 2$	
				abs. median bias ×1000				
$F = 3$	**16**	184	**6**	188	**1**	185	**1**	184
$F = 5$	**6**	140	**2**	138	**0**	137	**0**	135
$F = 10$	**1**	84	**0**	84	**1**	84	**0**	83
				95% coverage rate				
$F = 3$	**95**	65	**92**	9	**95**	30	**95**	0
	72		*81*		*71*		*81*	
$F = 5$	**95**	73	**94**	21	**95**	42	**95**	1
	81		*87*		*80*		*87*	
$F = 10$	**96**	81	**95**	44	**95**	61	**95**	8
	89		*91*		*88*		*91*	

the tails of the distribution. We did so for the extreme case of three degrees of freedom (increasing this number to infinity leads to the normal distribution) and report the results in Table 2. It appears that there is a slight increase in the median bias of LIML but even in the most extreme case is is negligible. However, the already poor perfomance of 2SLS becomes twice as bad. The coverage rate of LIML remains almost perfect, while the coverage rate of 2SLS deteriorates compared to the case of normality, where it already was poor. We conclude that, in this very limited sensitivity analysis, LIML is robust to a deviation from normality.

6. Concluding remarks

We have derived the LIML estimator for the static linear normal panel data model, thus filling an apparent gap in the literature. We presented all derivations for the panel LIML estimator and its variance under many-instruments asymptotics. In simulations, LIML appeared to have an excellent coverage rate, also in the cases where 2SLS is (highly) off the mark.

There are various topics for further research. One is to develop an LIML version of the Hausman-Taylor estimator (Hausman and Taylor, 1981). This is not only a widely used, instruments-based panel data estimator, but also one where there are variants with many instruments, becoming available when the time dimension of the data is exploited (Amemiya and MaCurdy, 1986, and Breusch et al., 1989).

Another topic concerns heteroskedasticity. In the case of a single cross-section, heteroskedasticity has recently been the focus of LIML research, e.g., Hausman et al. (2012). Heteroskedasticity in a panel data setting is a topic deserving attention.

A final challenge is to see how far the material of this article, in particular on many-instruments asymptotics, remains relevant when the model is dynamic. We already referred to Alvarez and Arellano (2003), Akashi and Kunitomo (2012), and Moral-Benito (2013). Another starting point could be to adapt the maximum likelihood analysis of the linear panel data model by Hsiao et al. (2002) for many-instruments asymptotics.

Acknowledgments

We would like to thank Paul Bekker, Geert Dhaene, Erik Meijer, Jelle van Essen, and two referees for their helpful comments and useful suggestions, and participants of the 2013 Xiamen International Symposium on Analysis of Panel Data in honor of Cheng Hsiao, and the 2013 Workshop on Panel Data of the Amsterdam School of Economics for inspiring discussions.

References

Akashi, K., Kunitomo, N. (2012). Some properties of the LIML estimator in a dynamic panel structural equation. *Journal of Econometrics* 166:167–183.

Alvarez, J., Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71:1121–1159.

Amemiya, T., MaCurdy, T. E. (1986). Instrumental-variable estimation of an error components model. *Econometrica* 54:869–881.

Anderson, T. W., Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics* 20(1):46–63.

Anderson, T. W., Rubin, H. (1950). The asymptotic properties of estimates of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 21:570–582.

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

Bekker, P. A., Wansbeek, T. J. (2014). *LIML in IV form*. Working paper, University of Groningen.

Breusch, T. S., Mizon, G. E., Schmidt, P. (1989). Efficient estimation using panel data. *Econometrica* 57:695–700.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J., Swanson, N. R. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3:211–255.

Hausman, J. A., Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica* 49:1377–1398.

Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* 109:107–150.

Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business & Economic Statistics* 31:451–472.
Nagar, A. L. (1959). The bias and moment matrix of the general k−class estimators of the parameters in simultaneous equations. *Econometrica* 27:575–595.
Newey, W. K. (2004). *Many instrument asymptotics*. Working paper, MIT.
Reiersøl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica* 9:1–24.
Theil, H. (1953). Repeated least squares applied to complete equation systems. The Hague: Central Planning Bureau.
Wansbeek, T. J., Meijer, E. (2000). *Measurement Error and Latent Variables in Econometrics*. North-Holland: Amsterdam.