LIML in the static linear panel data model

Tom Wansbeek Dennis Prak*

This version May 30, 2014

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 proving 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 regression.

1 Introduction

Regression with endogeneity is at the core of econometrics. 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 has come to dominate the practice of econometric estimation.

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, but applied researchers showed little eagerness and LIML vanished from sight.

A revival of LIML had to wait till the development of an appropriate framework relevant for studying the asymptotic distribution of the LIML estimator, with the following considerations. 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. The parameter sequence should be such that it generates acceptable approximations of known distributional properties of related statistics. This suggests studying asymptotic behavior under alternative asymptotics that stays close to the model. When there are many instruments h relative to the number of observations N, or when the instruments are weak, the usual asymptotics with h fixed and $N \to \infty$ may lead to poor approximations, and an alternative is called for.

Anderson (1976) was the first to describe such alternative asymptotics, now usually called "many-instruments asymptotics". This result had a theoretical relevance only, until Bekker (1994) introduced a consistent estimator of the alternative asymptotic variance. He thus made the alternative asymptotics useful for empirical work, by 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" (not his words).

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

*Faculty of Economics and Business, University of Groningen. We would like to thank Paul Bekker, Geert Dhaene and Erik Meijer for their helpful comments, 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.

been developed by e.g. Alvarez and Arellano (2003) and Akashi and Kunitomo (2012), 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. We do so in a framework of many-instruments asymptotics since the *raison d'être* of LIML lies there.

The paper 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 coverage rate of the panel LIML estimator relative to the 2SLS estimator. Section 6 concludes. In an appendix we argue the relevance of many-instruments asymptotics.

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

Like 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:

$$y_n = \beta x_n + u_n \tag{1}$$
$$x_n = \Pi' z_n + v_n, \tag{2}$$

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 $h \times T$. The error terms satisfy

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

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 on Ω (e.g. due to a 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,

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

Substitution of $y_n - \beta x_n$ for u_n and $x_n - \Pi' z_n$ for v_n turns this into the loglikelihood, $\log\mathcal{L}$, as the Jacobian of the transformation from u_n and v_n to y_n and x_n is 1. We maximize the loglikelihood in a few steps, using a single symbol ($\log\mathcal{L}$) throughout and neglecting constants. The derivation is an adaptation to the panel data case of the relatively succinct derivation given by Wansbeek and Meijer (2000) for the case of a single cross-section.

By way of notational convention, we write, for generic C of full column rank, $M_A = I - C(C'C)^{-1}C'$ and $P_A = C(C'C)^{-1}C'$.

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 $\log\mathcal{L} = |S_e|$. This simplification is only possible when Ω is not restricted.

The next step is to concentrate out Π. To this end we first write the model in matrix form. With X, Y, U and V $(N \times T)$ and Z $(N \times h)$, the model for all n is

$$\begin{aligned} Y &= \beta X + U \\ X &= Z\Pi + V. \end{aligned}$$

Let $R \equiv (U, Z)$ and $V_* \equiv P_R X - Z\Pi$. Then

$$\begin{aligned} V &= X - Z\Pi \\ &= M_R X + P_R X - Z\Pi \\ &= M_R X + V_*, \end{aligned}$$

and

$$\begin{aligned} M_U V_* &= M_U \left((U, Z)(R'R)^{-1}R'X - Z\Pi\right) \\ &= M_U Z\left((0, I_h)(R'R)^{-1}R'X - \Pi\right) \\ &\equiv M_U Z(\hat{\Pi} - \Pi), \end{aligned}$$

with $\hat{\Pi}$ implicitly defined. This implies the useful auxiliary result

$$V_*' M_U V_* = (\hat{\Pi} - \Pi)' Z' M_U Z(\hat{\Pi} - \Pi) \geq 0, \tag{3}$$

with equality if $\hat{\Pi} = \Pi$. Substituting $M_R X + V_*$ for V in the loglikelihood gives

$$\begin{aligned} \log\mathcal{L} &= |(U, V)'(U, V)| \\ &= |(U, M_R X + V_*)'(U, M_R X + V_*)| \\ &= \begin{vmatrix} U'U & U'V_* \\ V_*'U & X'M_R X + V_*'V_* \end{vmatrix} \\ &= |U'U|\, |X'M_R X + V_*' M_U V_*| \\ &= |U'U|\, |X'M_R X|, \end{aligned}$$

where the last step holds in the optimum on taking $\Pi = \hat{\Pi}$, cf. (3). This is only possible when Π is not restricted. After concentrating out Ω first and then Π, the loglikelihood depends on β only. Let

$$W \equiv \begin{vmatrix} X'M_Z X & X'M_Z U \\ U'M_Z X & U'M_Z U \end{vmatrix}.$$

Then

$$\begin{aligned} |W| &= |X'M_Z X|\, |U'M_Z U - U'M_Z X(X'M_Z X)^{-1}X'M_Z U| \\ &= |X'M_Z X|\, |U'M_{(X,Z)} U|. \end{aligned}$$

Interchanging X and U does not affect the determinant, so

$$|X'M_ZX|\,|U'M_{(X,Z)}U| = |U'M_ZU|\,|X'M_{(U,Z)}X|.$$

We notice that $X'M_ZX$ does not depend on β. Neither does $U'M_{(X,Z)}U$ as it is equal to $Y'M_{(X,Z)}Y$. Since $R = (U, Z)$, $X'M_{(U,Z)}X = X'M_RX$. So the loglikelihood $\log\mathcal{L} = |U'U|\,|X'M_RX|$ can be rewritten as

$$\log\mathcal{L} = \frac{|U'U|}{|U'M_ZU|}.$$

Using the general result, for $\Psi > 0$ and θ a scalar parameter,

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

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

$$\hat{\beta} = \frac{\mathrm{tr}[(\hat{U}'\hat{U})^{-1}Y'X - (\hat{U}'M_Z\hat{U})^{-1}Y'M_ZX]}{\mathrm{tr}[(\hat{U}'\hat{U})^{-1}X'X - (\hat{U}'M_Z\hat{U})^{-1}X'M_ZX]}, \tag{4}$$

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

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

yielding a new value of $\hat{\beta}$, and iteration until convergence. In the simulations below, iteration to the correct value always took but a few steps. 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 regression 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

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

This means that the instruments behave well in the limit even though their number increases with N. We indicate many-instruments asymptotics by an asterisk to distinguish it from the usual asymptotics with $N \to \infty$ only. By way of example, under the latter there holds

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

so $\mathrm{plim}_{N\to\infty}\, X'M_ZU/N = \Omega_{vu}$. By contrast, under many-instruments asymptotics we have $\mathrm{tr}(M_Z)/N = 1 - h/N \to 1 - \alpha$, so

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

Along these lines,

$$\frac{1}{N}\begin{pmatrix} U' \\ X' \end{pmatrix}(U,X) \quad \xrightarrow{p^*} \quad \begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & Q \end{pmatrix} \tag{5}$$

$$\frac{1}{N}\begin{pmatrix} U' \\ X' \end{pmatrix}M_Z(U,X) \quad \xrightarrow{p^*} \quad (1-\alpha)\begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix}. \tag{6}$$

This directly yields

$$A \quad \equiv \quad (U'U)^{-1}U'X - (U'M_ZU)^{-1}U'M_ZX \xrightarrow{p^*} 0 \tag{7}$$

$$B \quad \equiv \quad (U'U)^{-1}X'X - (U'M_ZU)^{-1}X'M_ZX \xrightarrow{p^*} \Omega_{uu}^{-1}Q. \tag{8}$$

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

3.2 Consistency

Rewriting (4), the panel LIML estimator solves the equation

$$\hat{\beta} - \frac{\mathrm{tr}\left[(\hat{U}'\hat{U})^{-1}Y'X - (\hat{U}'M_Z\hat{U})^{-1}Y'M_ZX\right]}{\mathrm{tr}\left[(\hat{U}'\hat{U})^{-1}X'X - (\hat{U}'M_Z\hat{U})^{-1}X'M_ZX\right]} = 0,$$

where $\hat{U} = \hat{U}(\hat{\beta}) \equiv Y - \hat{\beta}\, X$. This equation can be written as $f(\hat{\beta}) = 0$, with $f(\cdot)$ implicitly defined; $f(\cdot)$ is continuous in $\hat{\beta}$, since it is a composite function of continuous transformations of $\hat{\beta}$. Since $f(\hat{\beta}) = 0$, trivially, $\mathrm{plim}^* f(\hat{\beta}) = 0$. Now consider $f(\beta)$, and note that $\hat{U}(\beta) = U$, so

$$f(\beta) = \beta - \frac{\mathrm{tr}\left[(U'U)^{-1}Y'X - (U'M_ZU)^{-1}Y'M_ZX\right]}{\mathrm{tr}\left[(U'U)^{-1}X'X - (U'M_ZU)^{-1}X'M_ZX\right]} = \frac{\mathrm{tr}(A)}{\mathrm{tr}(B)} \xrightarrow{p^*} 0$$

because of (7) and (8). Since $\mathrm{plim}^*\, f(\beta) = \mathrm{plim}^*\, f(\hat{\beta}) = 0$, applying the continuous mapping theorem twice yields $\mathrm{plim}^*\, \hat{\beta} = \mathrm{plim}^*\, f^{-1}(0) = \beta$, establishing the consistency of the panel LIML estimator under many-instruments asymptotics.

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

$$\tilde{\beta} = \frac{\mathrm{tr}(X'P_ZY)}{\mathrm{tr}(X'P_ZX)} = \beta + \frac{\mathrm{tr}(X'P_ZU)}{\mathrm{tr}(X'P_ZX)}.$$

From (5) and (6) we obtain

$$\mathrm{plim}^*\, \frac{1}{N}X'P_ZU \quad = \quad \alpha\, \Omega_{vu}$$

$$\mathrm{plim}^*\, \frac{1}{N}X'P_ZX \quad = \quad Q + \alpha\, \Omega_{vv}.$$

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

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

Under the usual asymptotics, $\alpha = 0$, and the 2SLS estimator is consistent as yet.

3.3 Asymptotic variance

Consider the (infeasible) estimator

$$\tilde{\beta} = \frac{\mathrm{tr}\left\{(U'U)^{-1}Y'X - (U'M_ZU)^{-1}Y'M_ZX\right\}}{\mathrm{tr}\left\{(U'U)^{-1}X'X - (U'M_ZU)^{-1}X'M_ZX\right\}} = \frac{\mathrm{tr}\,\{A\}}{\mathrm{tr}\,\{B\}}.$$

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

$$V(\hat{\beta}) = \mathrm{plim}^*\ N(\tilde{\beta} - \beta)^2 = \frac{\mathrm{plim}^*\ N\,\{\mathrm{tr}(A)\}^2}{\{\mathrm{plim}^*\ \mathrm{tr}(B)\}^2}. \tag{9}$$

As to A,

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

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

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

Next, let

$$\begin{aligned} q &\equiv N\begin{pmatrix} \mathrm{vec}(\hat{U}'\hat{U})^{-1} \\ \mathrm{vec}(\hat{U}'\hat{U})^{-1} \\ \mathrm{vec}(\hat{U}'M_Z\hat{U})^{-1} \end{pmatrix} \xrightarrow{p^*} \begin{pmatrix} 1 \\ 1 \\ \frac{1}{1-\alpha} \end{pmatrix} \otimes \mathrm{vec}(\Omega_{uu}^{-1}) \\ u &\equiv \mathrm{vec}(U') \\ d &\equiv \begin{pmatrix} \Pi'Z' \otimes I_T \\ \tilde{V}' \otimes I_T \\ -\tilde{V}'M_Z \otimes I_T \end{pmatrix} u. \end{aligned}$$

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

$$\begin{aligned} \mathrm{plim}^*\ N\left[\mathrm{tr}(\hat{A})\right]^2 &= (\mathrm{plim}^*\ q)'\left\{\lim{}^* \frac{1}{N}\,\mathrm{E}(dd')\right\}(\mathrm{plim}^*\ q) \\ &= \mathrm{tr}\ [\Omega_{uu}^{-1}(Q + \tfrac{\alpha}{1-\alpha}\Omega_{vv\cdot u})]. \end{aligned} \tag{10}$$

So the asymptotic variance of $\hat{\beta}$ is given by (9), with (8) and (10) inserted, generalizing 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

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

one such expression is

$$\hat{V}(\hat{\beta}) = \frac{\mathrm{tr}[(\hat{U}'\hat{U})^{-1}X'WX]}{[\mathrm{tr}((\hat{U}'\hat{U})^{-1}X'HX)]^2}. \tag{11}$$

The consistency of this expression under many-instruments asymptotics follows from

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

On substitution this readily gives

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

from which the many-instruments consistency of $\hat{V}(\hat{\beta})$ in (11) 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 (1)–(2) then generalizes to

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

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

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

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.

It appears that the derivation of the panel LIML estimator for the case of the static panel data model with multiple regressors is not essentially different from the case with a single regressor. Also now we obtain as the loglikelihood $\log\mathcal{L} = |U'U|/|U'M_Z U|$, but with an adapted U of course. The panel LIML estimator is the solution of

$$\hat{\beta} = \hat{H}^{-1}\hat{h},$$

where

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

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

5 A Monte Carlo study

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

5.1 Design

In the simulations, we consider model (1)–(2), with $T = 2$, and let u_{nt} and v_{nt} be AR(1) with autocorrelation parameters ρ and η, respectively, and standard normal innovations. We take all elements of Π $(h \times 2)$ identical, equal to π, and let the instruments z_{nh} all be i.i.d. standard normal. As to parameter choices, we let $N = 100$ and $\beta = 1$. We let α and π run from 0.05 to 0.5. Hence the number of instruments $h = \alpha N$ runs from 5 to 50.

For each parameter combination 10,000 simulations are performed, where in each repetition the coverage rate of a 95% confidence interval of LIML with Bekker standard errors is compared with the coverage rate of a 95% confidence interval of 2SLS with classical standard errors. The Bekker standard errors are based on (11). This expression does not guarantee nonnegativity. In the very rare cases where the expression produced a negative number, we took the absolute value, a continuous transformation which does not affect consistency.

5.2 Results

Figures 1, 2, and 3 depict the difference between the 95% coverage rates of 2SLS with classical standard errors and LIML with Bekker standard errors, for given values of η and ρ.

The figures are rather similar for different values of N (not reported here) but depend heavily on the relation between U and V. In figure 1, for example, improvements vary from 1.8% for $\alpha = 0.05$ and $\pi = 0.5$ to 84% for $\alpha = 0.5$ and $\pi = 0.05$. In figure 3, both the smallest and the largest improvement are achieved at $\pi = 0.5$, the smallest at $\alpha = 0.05$ (23%) and the largest at $\alpha = 0.5$ (94%). When η increases, then the coverage rates decrease, but less for LIML than for 2SLS. Especially with many and/or weak instruments, 2SLS coverage rates closely tend to 0, whereas LIML coverage rates (based on Bekker standard errors) remain substantially larger, even in the configuration of figure 3. In this extreme setting, the largest 2SLS coverage rate is 64% ($\alpha = 0.05, \pi = 0.5$), whereas all LIML coverage rates for $\pi > 0.3$ are larger than this and some even still closely resemble 95%. In the more ideal setting of figure 1, all LIML coverage rates are larger than 85%, whereas for more than half of the (α, π)-loci the 2SLS coverage rates are below 75%.

Furthermore, these figures show that the locus of largest improvement depends on the relation between U and V. If this relation is weak, LIML attains the largest improvement for large α and small π. That is, for many and weak instruments. However, when η increases, the point with the largest difference in coverage rate moves gradually towards the configuration with strong instruments. An explanation is that, when η increases, the LIML coverage rates decrease. Both LIML and 2SLS are most precise when π is large. Hence, in the situation where η is small, both estimators perform well and LIML attains the largest improvement in the well-studied situation with many and weak instruments but, when η increases, LIML also loses efficiency (but at a slower rate than 2SLS), mostly in the regions where α and π are low since there the relative efficiency gain was largest. Therefore, the point where LIML outperforms 2SLS most shifts gradually.

Note, however, that in all cases studied LIML coverage rates are larger than 2SLS coverage rates, such that inference using LIML is uniformly preferred over inference with 2SLS. Coverage rates of the latter are generally substantially lower than their theoretical values, which may lead to misleading conclusions in hypothesis testing. Of all cases depicted in figures 1, 2, and 3, the average increase in coverage rate is 48% for a 95% confidence interval.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

6 Concluding remarks

We have derived the LIML estimator for the simplest static linear 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.

There are various topics for further research. One such topic would be to develop a 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, Mizon, and Schmidt, 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, Newey, Woutersen, Chao, and Swanson (2012).

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

Appendix: The case for many-instruments asymptotics

In this appendix, based on Wansbeek and Meijer (2000), section 6.6, we give the core of the argument to motivate many-instruments asymptotics in the cross-sectional case. In that case, the LIML estimator is the ML estimator of β in

$$\begin{aligned} y &= X\beta + u \\ X &= Z\Pi + V, \end{aligned}$$

with Z of order $N \times h$ exogenous instruments. The errors are i.i.d. normal. Let

$$\begin{aligned} S_P &\equiv (y, X)' P_Z (y, X) \\ S_M &\equiv (y, X)' M_Z (y, X). \end{aligned}$$

The LIML estimator $\hat{\beta}$ follows from the first-order condition

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

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 $\ell \equiv h/(N-h)$. After some algebraic manipulations, it appears that a direct implication of the model is

$$(\Sigma_P - \lambda \Sigma_M) \begin{pmatrix} 1 \\ -\beta \end{pmatrix} = 0;$$

So the LIML estimator satisfies a relation that is the sample analog of a model implication; it stays "close to the data".

This observation has an implication when studying the asymptotic distribution, and in particular the asymptotic variance, 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. This suggests to study an asymptotic sequence where ℓ does not vanish, i.e. where the number of instruments grows along with the number of observations, thus staying close to the data. This motivates "many-instruments asymptotics."

Another implication is that the LIML estimator stays closer to the data than the IV or 2SLS estimator,

$$\tilde{\beta} = (X' P_Z X)^{-1} X' P_Z y.$$

This estimator has model counterpart $\beta = \Sigma_{P \cdot 22}^{-1} \sigma_{P \cdot 21}$; with e_1 the first unit vector, this corresponds with

$$(\Sigma_P - \mu\, e_1 e_1') \begin{pmatrix} 1 \\ -\beta \end{pmatrix} = 0.$$

This is at variance with the model, and suggests a better performance of LIML over IV. Evidently, the difference between IV and LIML is small when $\ell \approx 0$, so when N is large relative to the number of instruments, or when $\Sigma_M \approx c \cdot e_1 e_1'$, so when the instruments are not weak and explain the regressors well. But when there are many instruments or when the instruments are weak, the LIML is to be preferred, with standard errors based on many-instruments asymptotics.

References

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

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

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

Anderson, T.W. (1976), "Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics", *Journal of the Royal Statistical Society B*, **38**, 1–36.

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

Anderson, T.W. and H. Rubin (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**, 657–681.

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

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

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

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

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.

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*, Central Planning Bureau, The Hague.

Wansbeek, T.J. and E. Meijer (2000), *Measurement error and latent variables in econometrics*, North-Holland.

List of Figures

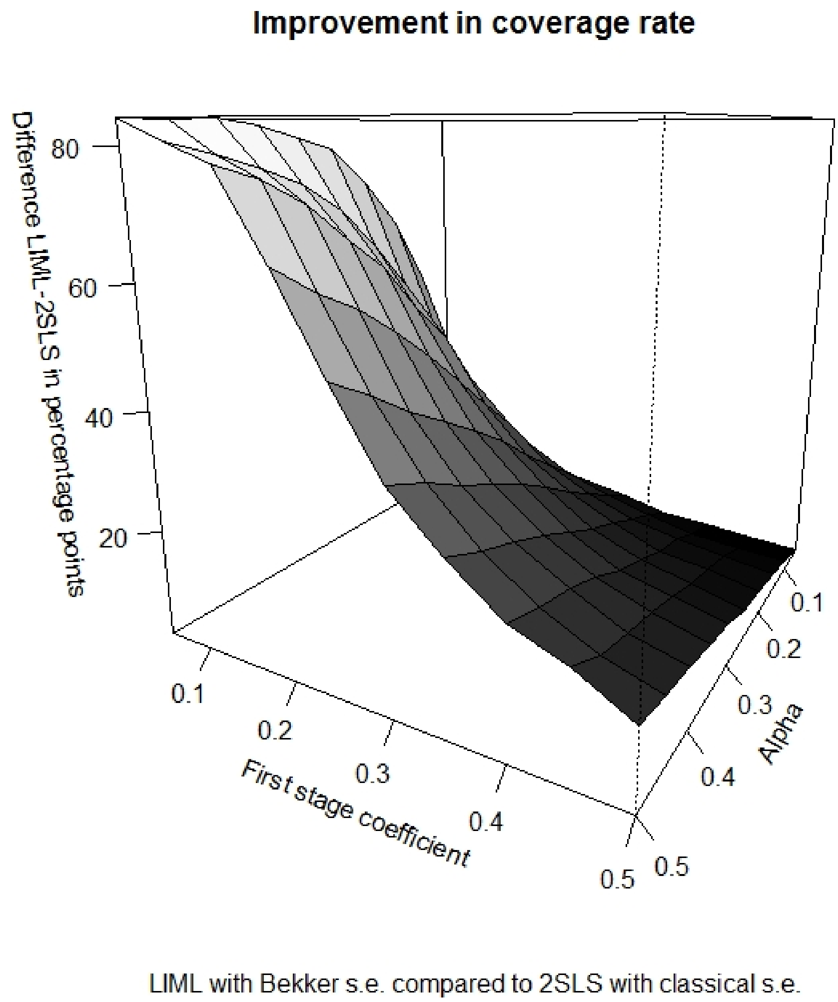

Figure 1: $N = 100$, $\rho = 0.4$, $\eta = 0.6$

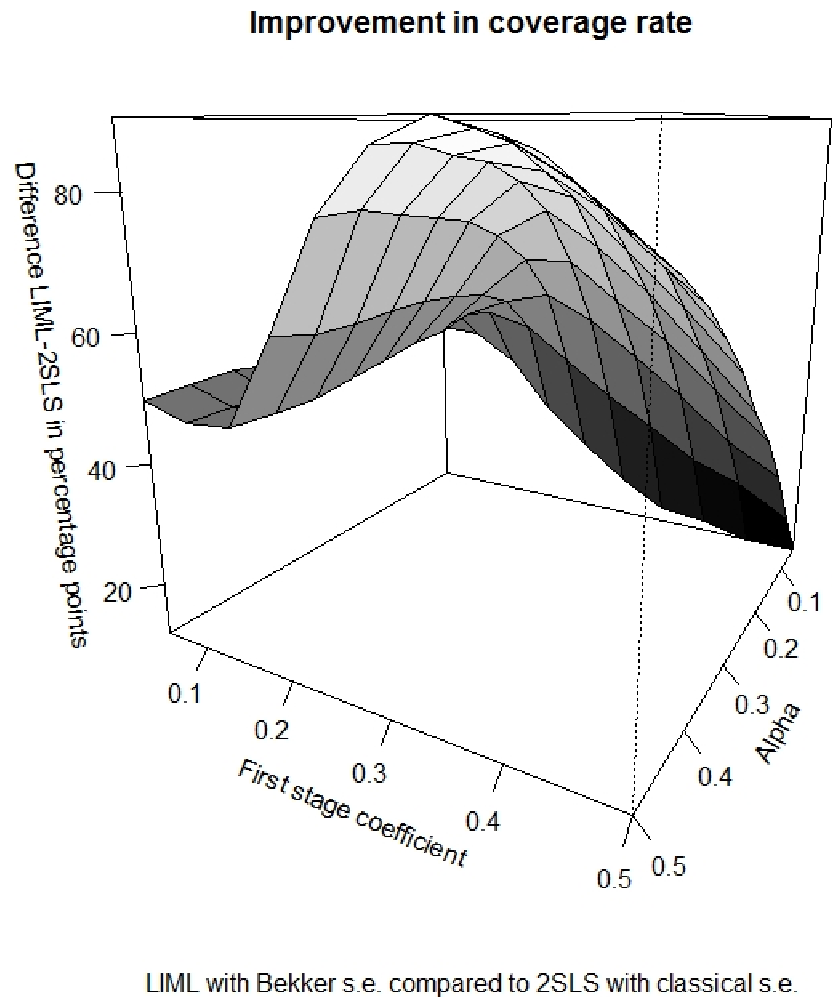

Figure 2: $N = 100$, $\rho = 0.8$, $\eta = 1.2$

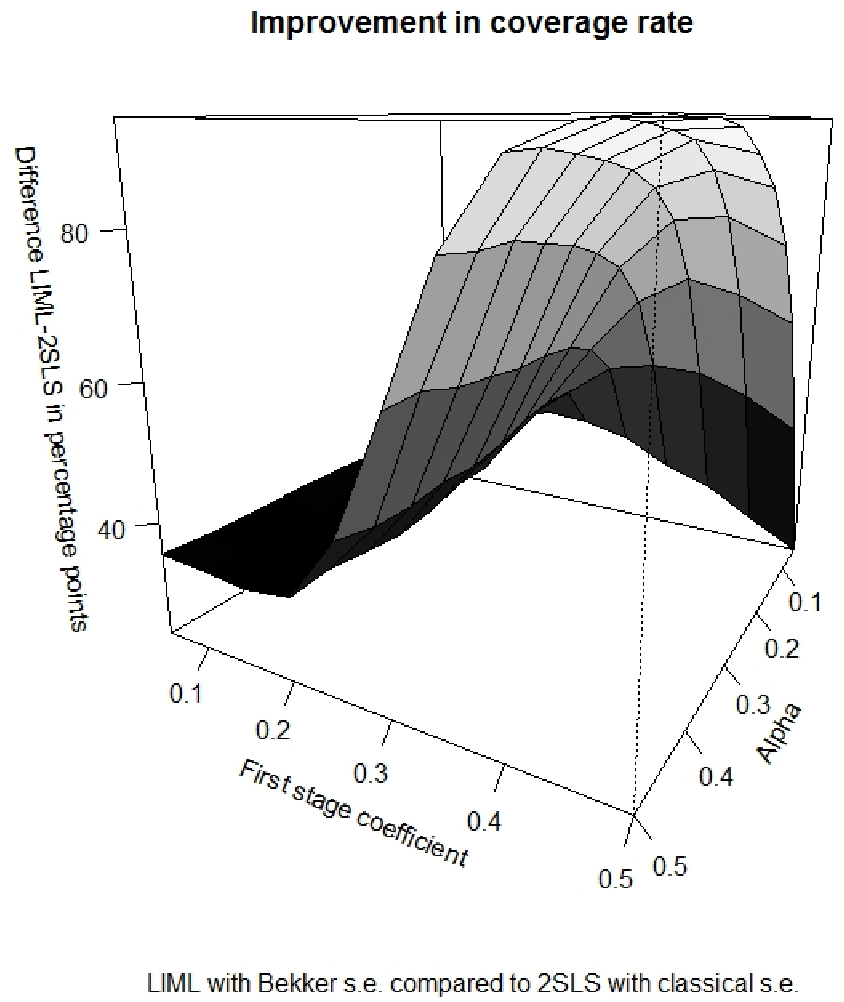

Figure 3: $N = 100$, $\rho = 0.3$, $\eta = 3.0$