LIML IN THE LINEAR PANEL DATA MODEL

Tom Wansbeek & Dennis Prak
University of Groningen

July 2014

20th International Conference on Panel Data, Tokyo

Introduction

- Regression with endogeneity is at the core of econometrics. IV is the standard approach. The expression "instrumental variables" is due to Reiersøl (1941, 1945). Theil (1953) developed "repeated least squares", pioneering more instruments than regressors.
- LIML estimation was introduced by Anderson and Rubin (1949, 1950). Its qualities over IV got appreciated soon but applied researchers stayed away. Interest was revived by Angrist and Krueger (1991) by their work on weak and many instruments. Bekker (1994) inspired new theoretical work.
- LIML is not the only alternative to IV: split-sample IV (Angrist and Krueger, 1995), which is related to two-sample IV, jackknife IV (Angrist, Imbens, and Krueger, 1999), and symmetrically normalized IV (Arellano and Alonso-Borrego, 1999). LIML and alternatives are now a lively research area (Bekker and Wansbeek, 2014).

Panel data

- As to panel data, LIML-like estimators have been developed by e.g. Alvarez and Arellano (2003), Akashi and Kunitomo (2011), and Moral-Benito (2011), for the dynamic model. Their estimators are Least Variance Ratio (LVR) estimators, not “true” LIML obtained from maximizing the likelihood function; there seems to be gap in the literature.
- The objective of our study is to fill this gap by deriving the LIML estimator for the linear panel data model and investigating its properties.
- Since the LIML literature is often complicated, beginning with the derivation of the LIML estimator, we try to keep things as simple as possible.

LIML in the cross-sectional case

In general, the LIML estimator is the ML estimator of β in

$$y = X_1\beta_1 + X_2\beta_2 + u = X\beta + u,$$

X_1 $(N \times g_1)$ endogenous, X_2 $(N \times g_2)$ exogenous, and

$$X_1 = X_2\Pi_2 + X_3\Pi_3 + V = Z\Pi + V,$$

V $(N \times g_1)$ orthogonal to Z; $(u_n, v_n')'$ i.i.d. normal.

The LIML estimator

Throughout, P_A denotes projection on A and M_A denotes projection orthogonal to A, for any A. Let

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

Anderson and Rubin (1949) showed that the LIML estimator 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.

Why LIML?

LIML offers a better inferential context than IV. This follows from what the model implies (with $X_2 = 0$ for non-essential simplicity):

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

so, with g the number of regressors,

$$
\begin{aligned}
(y, X) &= (X\beta + u, X) \\
&= Z\Pi(\beta, I_g) + (u + V\beta, V) \\
&\equiv Z\Pi(\beta, I_g) + W.
\end{aligned}
$$

So $\mathrm{E}(Z'W) = 0$. Each row of W is i.i.d. $(0, \Psi)$, say.

Why LIML? (cont.)

Let h be the number of IVs (columns of Z). Then

$$\begin{aligned} \Sigma_P &\equiv \mathrm{E}(S_P) = (\beta, I_g)'\Pi'Z'Z\Pi(\beta, I_g) + h\Psi \\ \Sigma_M &\equiv \mathrm{E}(S_M) = (N-h)\Psi, \end{aligned}$$

so, eliminating Ψ,

$$\begin{aligned} \Sigma_P &= (\beta, I_g)'\Pi'Z'Z\Pi(\beta, I_g) + \lambda\Sigma_M, \quad \text{with} \\ \lambda &\equiv \frac{h}{N-h}. \end{aligned}$$

Rearranging and postmultiplication by $\begin{pmatrix} 1 \\ -\beta \end{pmatrix}$ gives

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

λ is the smallest value for which $\Sigma_P - \lambda\Sigma_M$ is singular yet ≥ 0.

Why LIML? (cont.)

Summarizing, the model implies, for smallest λ,

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

and the LIML estimator satisfies, for smallest $\hat{\lambda}$,

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

So the LIML estimator satisfies a relation that is the sample analog of a model implication; it stays "close to the data". This "sales pitch" is from Wansbeek and Meijer (2000), based on Bekker (1994) and van der Ploeg (1997).

Comparing with IV

Even more, the LIML estimator stays "closer to the data" than the IV estimator. Take again the model implication

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

The IV estmator

$$\hat{\beta}_{\text{IV}} = (X'P_Z X)^{-1}X'P_Z y$$

has model counterpart $\beta = \Sigma_{P\cdot 22}^{-1}\sigma_{P\cdot 21}$, which corresponds with

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

at variance with the model; better performance of LIML over IV.

When can you neglect this?

Evidently, the difference between IV and LIML is small when

$$\lambda = \frac{h}{N-h} \approx 0,$$

so when N is large, or when

$$\Sigma_M \approx c \cdot \; e_1 e_1',$$

that is, $\mathrm{E}(v_n v_n') \approx 0$, so the instruments are not weak and do a good job in explaining the regressors. If you cannot neglect this, there are implications for the investigation of the asymptotic behavior of the estimators.

Alternative asymptotics

- 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.
- Anderson (1976) first described such alternative asymptotics, now usually called "many-instruments asymptotics". Bekker (1994) introduced a consistent estimator of the alternative asymptotic variance, in his words, "a remarkable result with practical implications". The result is often known as "Bekker standard errors" (not his words).

Inconsistency of IV

Under such asymptotics, one implication of the model is

$$\Sigma_{P\cdot 22}^{-1}\sigma_{P\cdot 21} = \beta + \frac{\lambda}{\lambda+1}\Sigma_{P\cdot 22}^{-1}(\psi_{21} - \Psi_{22}\beta).$$

The LHS is the population counterpart of the IV estimator. The RHS equals β plus a term that vanishes under the traditional asymptotics but not under many-instruments asymptotics. Under the latter, which arguably gives a better approximation of the exact distribution, IV is inconsistent.

Now, the panel data model

We consider the case of a single regressor since this captures essential elements, and the notation can be kept simple. Then

$$\begin{aligned} y_n &= \beta \cdot x_n + u_n \\ x_n &= \Pi' z_n + v_n, \end{aligned}$$

where x_n, y_n, u_n, v_n are $T \times 1$, and z_n is $h \times 1$. For the error terms:

$$f_n \equiv \binom{u_n}{v_n} \sim N(0, \Omega), \quad \Omega = \begin{pmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{pmatrix},$$

i.i.d.The model is "limited-information" as we do not impose a structure on Π (and not on Ω), unlike the very simple specification $x_n = \pi \cdot z_n + e_n$, when $\Pi = \pi \cdot I_T$ or the case of instruments per wave, $x_t = Z_t \pi_t + e_t$, when $\Pi = \mathrm{diag}(\pi_1', \ldots, \pi_T')$.

Likelihood

Adaptation for the panel case (plus some simplification) from Wansbeek and Meijer (2000). Let $S_f \equiv \sum_n f_n f_n'/N$. The joint density of the f_n is

$$L \equiv \log|\Omega| + \mathrm{tr}(\Omega^{-1} S_f),$$

appart from constants. Substitution of $y_n - x_n\beta$ for u_n and $x_n - \Pi' z_n$ for v_n gives the loglikelihood as the Jacobian is unity.

First, concentrate out Ω. Because $\partial L/\partial\Omega = \Omega^{-1} - \Omega^{-1} S_f \Omega^{-1}$, the optimal value for Ω is S_f, and the concentrated likelihood can be simplified to $L = |S_f|$. This simplification is only possible when Ω is not restricted.

Likelihood (cont.)

With X, Y, U and V $(N \times T)$ and Z $(N \times h)$, the model for all n is

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

The key to a succinct derivation is through the definition $R \equiv (U, Z)$, interpretation elusive as yet. Then

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

leading to

$$V_*' M_U V_* = (\hat{\Pi} - \Pi)' Z' M_U Z(\hat{\Pi} - \Pi).$$

Likelihood (cont.)

We can now further elaborate the likelihood, with $V = M_R X + V_*$:

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

in the optimum. Now, L depends on β only. This simplification is only possible when Π is not restricted.

Likelihood (cont.)

For

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

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

Interchanging X and U does not affect the determinant, so

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

Since $X'M_Z X$ and $U'M_{(X,Z)}U$ $(= Y'M_{(X,Z)}Y)$ do not depend on β and $(U,Z) = R$, the likelihood $L = |U'U|\,|X'M_R X|$ becomes

$$L = \frac{|U'U|}{|U'M_Z U|}.$$

The determinants constitute "generalized variances".

The LIML estimator

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

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

the derivative w.r.t. β is obtained. Setting it equal to zero gives

$$\hat{\beta} = \frac{\mathrm{tr}\left\{ (\hat{U}'\hat{U})^{-1} Y'X - (\hat{U}'M_Z\hat{U})^{-1} Y'M_Z X \right\}}{\mathrm{tr}\left\{ (\hat{U}'\hat{U})^{-1} X'X - (\hat{U}'M_Z\hat{U})^{-1} X'M_Z X \right\}}.$$

This equation is non-linear in $\hat{\beta}$ as $\hat{U} = Y - \hat{\beta} \cdot X$. Solving it for β is numerically easy by substitution of an initial estimator of β, like the instrumental variables estimator, in $\hat{U}'\hat{U}$ and $\hat{U}'M_Z\hat{U}$, and iteration, "continuous updating", generalizing the eigenvalue.

Many-instruments asymptotics

We consider its properties under many-instruments asymptotics,

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

So 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. E.g., under the latter there holds

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

so $\mathrm{plim}_{N\to\infty} X' M_Z U / N = \Omega_{vu}$.Under many-instruments asymptotics, $(N-h)/N \to 1 - \alpha$, so

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

Auxiliary results

Along these lines,

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

This directly yields

$$
\begin{aligned}
A &\equiv (U'U)^{-1}U'X - (U'M_ZU)^{-1}U'M_ZX \xrightarrow{p^*} 0 \\
B &\equiv (U'U)^{-1}X'X - (U'M_ZU)^{-1}X'M_ZX \xrightarrow{p^*} \Omega_{uu}^{-1}Q.
\end{aligned}
$$

We write $\hat{A}$ and $\hat{B}$ for A and B when we have $\hat{U} \equiv Y - \hat{\beta} \cdot X$ instead of U; same plims.

Consistency

The panel LIML estimator solves the equation

$$\hat{\beta} - \beta = \frac{\operatorname{tr}\left\{\hat{A}\right\}}{\operatorname{tr}\left\{\hat{B}\right\}},$$

or briefly $f(\hat{\beta}) = 0$. Since $f(\hat{\beta}) = 0$, trivially, $\operatorname{plim}^* f(\hat{\beta}) = 0$. Now consider $f(\beta)$. It satisfies

$$f(\beta) = \frac{\operatorname{tr}\{A\}}{\operatorname{tr}\{B\}} \xrightarrow{p^*} 0.$$

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

Inconsistency of IV

Now consider the IV estimator $\hat{\beta}_{\text{IV}}$,

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

Since

$$\begin{aligned} \text{plim}^* \ \tfrac{1}{N} X'P_Z U &= \alpha \, \Omega_{vu} \\ \text{plim}^* \ \tfrac{1}{N} X'P_Z X &= Q + \alpha \, \Omega_{vv}, \end{aligned}$$

we have under many-instruments asymptotics

$$\text{plim}^* \ \hat{\beta}_{\text{IV}} = \beta + \frac{\alpha \ \text{tr}(\Omega_{vu})}{\text{tr}(Q) + \alpha \ \text{tr}(\Omega_{vv})} \neq \beta.$$

Under the usual asymptotics, $\alpha = 0$, and the bias disappears.

Asymptotic variance

Consider the (infeasible) estimator

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

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

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

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

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

Asymptotic variance (cont.)

In order to elaborate $\mathrm{tr}(A)$, let $u \equiv \mathrm{vec}(U)$ so $\mathrm{E}(uu') = \Omega_{uu} \otimes I_N$,

$$q \equiv \begin{pmatrix} \mathrm{vec}(U'U)^{-1} \\ \mathrm{vec}(U'U)^{-1} \\ \mathrm{vec}(U'M_Z U)^{-1} \end{pmatrix}, \quad \text{and} \quad d \equiv \begin{pmatrix} (I_T \otimes \Pi' Z')u \\ (I_T \otimes \tilde{V}')u \\ -(I_T \otimes \tilde{V}' M_Z)u \end{pmatrix}.$$

Then $\mathrm{tr}(A) = q'd$ and $\mathrm{plim}^* \; N \{\mathrm{tr}(A)\}^2 = \mathrm{plim}^* \; N \, q' \mathrm{E}(dd') q$;

$$\mathrm{E}(dd') = \Omega_{uu} \otimes \begin{pmatrix} \Pi' Z' Z \Pi & 0 & 0 \\ 0 & N \, \Omega_{vv \cdot u} & -(N-h)\, \Omega_{vv \cdot u} \\ 0 & -(N-h)\, \Omega_{vv \cdot u} & (N-h)\, \Omega_{vv \cdot u} \end{pmatrix}.$$

The independence of U and $\tilde{V}$ has simplified things essentially.

Asymptotic variance (cont.)

Collecting all the terms gives

$$\begin{aligned} q'\mathrm{E}(dd')q \quad = \quad & \mathrm{tr}\left\{(U'U)^{-1}\Omega_{uu}(U'U)^{-1}\Pi'Z'Z\Pi\right\} \\ & +\mathrm{tr}\left\{(U'U)^{-1}\Omega_{uu}(U'U)^{-1}\Omega_{vv\cdot u}\right\}\cdot N \\ & +\mathrm{tr}\left\{(U'M_ZU)^{-1}\Omega_{uu}(U'M_ZU)^{-1}\Omega_{vv\cdot u}\right\}\cdot(N-h) \\ & -2\mathrm{tr}\left\{(U'M_ZU)^{-1}\Omega_{uu}(U'U)^{-1}\Omega_{vv\cdot u}\right\}\cdot(N-h) \end{aligned}$$

and $\mathrm{plim}^* \; N\{\mathrm{tr}(A)\}^2$ follows readily.We already had obtained $\mathrm{plim}^* B = \Omega_{uu}^{-1}Q$, so the asymptotic variance of $\hat{\beta}$ is

$$V(\hat{\beta}) = \frac{\mathrm{tr}\left\{\Omega_{uu}^{-1}(Q + \frac{\alpha}{1-\alpha}\Omega_{vv\cdot u})\right\}}{\left\{\mathrm{tr}(\Omega_{uu}^{-1}Q)\right\}^2},$$

generalizing Newey (2004). In order to obtain Bekker standard errors, it remains to find a consistent estimator $\hat{V}(\hat{\beta})$ of $V(\hat{\beta})$.

Bekker standard errors

Bekker standard errors can have various forms. With $\hat{\alpha} = h/N$,

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

one form is

$$\hat{V}(\hat{\beta}) = \text{tr}\left\{(\hat{U}'\hat{U})^{-1} X'WX\right\} / \left\{\text{tr}((\hat{U}'\hat{U})^{-1} X'HX)\right\}^2.$$

By taking * limits, the consistency of $\hat{V}(\hat{\beta})$ follows immediately. An alternative is to replace the scalar factors $\hat{\alpha}$ and $(1-\hat{\alpha})$ by matrix factors $\hat{A}$ and $(I_T - \hat{A})$ outside the terms $X' \cdots X$, where

$$\begin{aligned} \hat{A} &= \hat{U}' P_Z \hat{U} (\hat{U}'\hat{U})^{-1} \xrightarrow{p^*} \alpha\ I_T \\ (I_T - \hat{A}) &= \hat{U}' M_Z \hat{U} (\hat{U}'\hat{U})^{-1} \xrightarrow{p^*} (1-\alpha)\ I_T. \end{aligned}$$

These constructions are a bit ad hoc; Bekker and Wansbeek (2014) may offer better alternative.

Simulation design

We extend to $T = 2$ the design of Bekker and Wansbeek (2014),

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

Π is $h \times 2$, $N = 500$, all variables have mean zero. Further, $\beta = 0$ and

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

$(\varepsilon_{n1}, \varepsilon_{n2})'$ and $(\tilde{v}_{n1}, \tilde{v}_{n2})'$ are i.i.d. $N(0, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right))$. The elements of z_n are i.i.d. $N(0, 1)$.

Elements of interest

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

- 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, driven by ω; $\omega = 0$ corresponds with absence of endogeneity. We let $\omega = 0.5$ and $\omega = 2$.
- The strength of the instruments, as measured by the F of the regression of x on z. We let $F = 3, F = 5, F = 10$, obtained by appropriate choices of π; 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 $F = (N - h)R^2/h(1 - R^2)$.
- The coherence in the panel. This is driven by ρ. We consider $\rho = 0, \rho = 0.3, \rho = 0.8$.

Simulation results, **LIML** and 2SLS, $N = 500, \rho = 0$

	10 instruments				30 instruments			
	$\omega = 0.5$		$\omega = 2$		$\omega = 0.5$		$\omega = 2$	
	abs. median bias$\times$1000							
$F = 3$	**104**	144	**40**	119	**55**	154	**28**	121
$F = 5$	**31**	70	**48**	111	**37**	113	**12**	61
$F = 10$	**35**	68	**47**	76	**16**	26	**3**	23
	95% coverage rate							
$F = 3$	**96**	82	**95**	52	**95**	60	**95**	9
$F = 5$	**95**	87	**95**	64	**95**	72	**95**	23
$F = 10$	**95**	90	**95**	78	**95**	82	**95**	48

Simulation results, **LIML** and 2SLS, $N = 500, \rho = 0.3$

	10 instruments				30 instruments			
	$\omega = 0.5$		$\omega = 2$		$\omega = 0.5$		$\omega = 2$	
	abs. median bias×1000							
$F = 3$	**170**	197	**34**	129	**79**	165	**45**	120
$F = 5$	**148**	74	**48**	4	**45**	77	**0**	63
$F = 10$	**31**	58	**25**	18	**13**	43	**6**	32
	95% coverage rate							
$F = 3$	**96**	84	**95**	58	**95**	66	**95**	15
$F = 5$	**96**	88	**95**	69	**95**	76	**95**	33
$F = 10$	**95**	91	**95**	81	**95**	85	**95**	57

Simulation results, **LIML** and 2SLS, $N = 500, \rho = 0.8$

	10 instruments				30 instruments			
	$\omega = 0.5$		$\omega = 2$		$\omega = 0.5$		$\omega = 2$	
	abs. median bias$\times$1000							
$F = 3$	**12**	143	**94**	36	**123**	231	**9**	72
$F = 5$	**32**	9	**44**	105	**7**	60	**54**	104
$F = 10$	**5**	2	**41**	14	**15**	39	**3**	4
	95% coverage rate							
$F = 3$	**96**	85	**94**	65	**95**	71	**95**	26
$F = 5$	**96**	88	**95**	74	**95**	79	**95**	44
$F = 10$	**95**	92	**95**	83	**95**	87	**95**	66

More regressors

With more regressors, all related to the same set of instruments:

$$\begin{aligned} y_n &= x_{1n}\beta_1 + \ldots + x_{kn}\beta_k + u_n \\ x_{1n} &= \Pi_1' z_n + v_{1n}, \\ &\vdots \\ x_{kn} &= \Pi_k' z_n + v_{kn}. \end{aligned}$$

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.

LIML estimation

The derivation of the LIML estimator for this extended case carries through as before, again leading to $L = |U'U|/|U'M_Z U|$. The LIML estimator is the solution of

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

where

$$\begin{aligned}
(\hat{H})_{ij} &\equiv \operatorname{tr}\left\{(\hat{U}'\hat{U})^{-1}X_i'X_j - (\hat{U}'M_Z\hat{U})^{-1}X_i'M_ZX_j\right\} \\
(\hat{h})_i &\equiv \operatorname{tr}\left\{(\hat{U}'\hat{U})^{-1}Y'X_j - (\hat{U}'M_Z\hat{U})^{-1}Y'M_ZX_j\right\},
\end{aligned}$$

for $i, j = 1, \ldots, k$. The generalization carries through in a straightforward way all the way to Bekker standard errors, as long as there are no exclusion restrictions on Π.

Concluding remarks

- We have derived the LIML estimator for the simplest panel data model, thus filling an apparent gap in the literature. Results are much like the $T = 1$ case, with a lot of "tr" added. We presented all derivations for (panel) LIML estimator and its variance under many-instruments asymptotics, and they are overall (fairly) simple.
- In simulations LIML has excellent coverage rate, also in the cases where 2SLS is (highly) off the mark.
- Further research questions include: Why is the gain in median bias of LIML over 2SLS in a panel less than in a cross-section? How can the link be made between T and h? What is the effect of heteroskedasticity? How does Hausman-Taylor (and its many-instruments extensions) fit into a LIML framework? In which cases should applied researchers really turn to LIML?