Likelihood approach to small T dynamic panel models with interactive effects

by Jushan Bai

The Cowles Foundation Conference, Econometrics
June 23, 2009

The Model

$$y_{it} = \alpha\, y_{it-1} + x_{it}'\beta + \delta_t + \lambda_i' f_t + \varepsilon_{it}$$

$$i = 1, 2, ..., N; t = 1, 2, ..., T$$

- λ_i and f_t are $r \times 1$ unobservable.
- factor error structure
- either λ_i or f_t or both are correlated with x_{it}.
- if $f_t \equiv 1$, the usual fixed effects model

Examples of multiple interactive effects

- (Multiple unobserved heterogeneities) In earnings studies, λ_i is a vector of unmeasured skills such as motivation, perseverance, and hard working; F_t is the price vector for the skills.
- In macroeconomics, F_t represents unobserved common shocks to all countries (states) and λ_i is the impact of the common shocks on country i.
- In finance, F_t : factor returns (systematic risks); λ_i : exposures to the systematic risks; ε_{it} : idiosyncratic returns.
- Multiple interactive effects can also capture cross-sectional correlations.

The Model

$$y_{it} = \alpha\, y_{it-1} + x_{it}'\beta + \delta_t + \lambda_i' f_t + \varepsilon_{it}$$

- objective: estimate α, β, and $\mathrm{var}(\varepsilon_{it}) = \sigma_t$.
- key feature: T is small. Treat f_t as parameters (conditional on f_t)
- Correlation between x_{it} and λ_i
- By definition, y_{it-1} is also correlated with λ_i

Related literature:

Kiefer(1980), Holtz-Eakin, Newey, and Rosen (1988), Ahn, Lee, and Schmidt (2001, 2006), Pesaran (2006), Bai (2005), ...

Earlier literature:
Goldberger (1971), Joreskog and Goldberger (1972), MaCurday (1982), ...

Omitting the literature on factor models for forecasting purposes, or unit root analysis because of either no regressors, or not modeling the correlation between regressors and the effects, or large T and not modeling the initial conditions.

Comments

• The within group estimator will not work.
The within transformation cannot eliminate the interactive effects.
• Even for additive effects, the within group estimator is not consistent under fixed T for dynamic panels.

Estimating interactive effects model by the least squares method

Write the model in vector form

$$y_i = x_i\beta + F\lambda_i + \varepsilon_i$$

where y_i is $T \times 1$, and x_i is $T \times p$, F is $T \times r$.
The least squares objective function

$$SSR = \sum_{i=1}^{N}(y_i - x_i\beta - F\lambda_i)'(y_i - x_i\beta - F\lambda_i)$$

We minimize SSR w.r.t. β, F, and Λ, where

$$F = \begin{bmatrix} F_1' \\ F_2' \\ \vdots \\ F_T' \end{bmatrix} \quad \Lambda = \begin{bmatrix} \lambda_1' \\ \lambda_2' \\ \vdots \\ \lambda_N' \end{bmatrix}$$

This paper: likelihood approach, small T dynamic model, heteroskedasticity
To control the correlation between x_{it} and λ_i

$$E(\lambda_i | x_{i1}, x_{i2}, ..., x_{iT}) = \lambda + \sum_{s=1}^{T} \phi_s \, x_{is} \tag{1}$$

$$E(\lambda_i | x_{i1}, x_{i2}, ..., x_{iT}) = \lambda + \phi \bar{x}_i \tag{2}$$

Write

$$\lambda_i = \lambda + \phi \bar{x}_i + \eta_i$$

and

$$y_{it} = (\delta_t + \lambda f_t) + x_{it}' \beta + f_t \phi \bar{x}_i + f_t' \eta_t + \varepsilon_{it}$$

In vector form

$$y_i = \delta + x_i\beta + F\phi\bar{x}_i + F\eta_i + \varepsilon_i$$

Assuming η_i and ε_i are independent normal, the likelihood for $F\eta_i + \varepsilon_i$ is easy. That is also the likelihood for y_i conditional on x_i. Let $\theta = (\delta, \beta, \phi, F, \Phi, D)$, the likelihood function is

$$\ell(\theta) = -\frac{N}{2}\log\det(\Omega) - \frac{1}{2}\sum_{i=1}^{N} u_i'\Omega^{-1}u_i$$

where

$$u_i = y_i - \delta - x_i\beta - F\phi\bar{x}_i$$

$\Omega = F\Phi F' + D$, $\Phi = \mathrm{var}(\eta_i)$, $D = \mathrm{var}(\varepsilon_i)$.

If one is willing to specify the DGP for x_{it}, then y and x can be modeled jointly and there is no need for the Chamberlain and Mundlak projections (details in the paper)

Dynamic panel

$$y_{it} = \delta_t + \alpha y_{i,t-1} + x_{it}'\beta + f_t'\lambda_i + \varepsilon_{it}$$

Under the assumption

$$E(\varepsilon_{it}|x_{i1},...,x_{iT},\lambda_i) = 0$$

x_{it} is strictly exogenous w.r.t. ε_{it},
but is allowed to be correlated with the effects λ_i.

The likelihood conditional on y_{i0} and x_i

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\alpha & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -\alpha & 1 \end{bmatrix} \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{bmatrix} y_{i0} + \delta + x_i \beta + F\lambda_i + \varepsilon_i$$

or

$$By_i = \alpha e_1 y_{i0} + \delta + x_i' \beta + F\lambda_i + \varepsilon_i$$

Projection

$$\lambda_i = \phi \bar{x}_i + \phi_0 y_{i0} + \eta_i$$

$$By_i = \alpha e_1 y_{i0} + \delta + x_i \beta + F\phi \bar{x}_i + F\phi_0 y_{i0} + F\eta_i + \varepsilon_i$$

The likelihood for $F\eta_i + \varepsilon_i$ is also the likelihood for By_i conditional on y_{i0} and x_i. Since $|B| = 1$, it is also the likelihood of y_i conditional on y_{i0} and x_i.

• Can test if the initial condition y_{i0} (endowment) is correlated with individual effect by testing if $\phi_0 = 0$.
• It is possible to derive the likelihood for the entire sequence $(y_{i0}, y_{i1}, ..., y_{iT})$. Need the reduced form for y_{i0} (in the paper).
• x_{it} predetermined, but weakly exogenous (in the paper)
• x_{it} predetermined, but not weakly exogenous, need joint modeling of y and x, through VAR (in the paper).

EM algorithm.
The complete date likelihood (joint density)

$$f(y_i, \eta_i|\theta) = f(y_i|\eta_i, \theta) f(\eta_i|\theta)$$

$$\log f(y_i, \eta_i|\theta) = \log f(y_i|\eta_i, \theta) + \log f(\eta_i|\theta)$$

Expectation:

$$Q(\theta|\theta^{(k)}) = E\Big[\log f(y_i, \eta_i|\theta) \Big| y_i, \theta^{(k)} \Big]$$

Maximization

$$\theta^{(k+1)} = \text{argmax}_\theta Q(\theta|\theta^{(k)})$$

In our case, the complete data likelihood is

$$-\frac{N}{2}\ln|D| - \frac{1}{2}\sum_{i=1}^{N}\left[u_i' D^{-1} u_i - 2u_i' D^{-1} F\eta_i + \mathrm{tr}(F' D^{-1} F\eta_i\eta_i')\right]$$

$$-\frac{N}{2}\ln|\Phi| - \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}(\Phi^{-1}\eta_i\eta_i')$$

where $u_i = y_i - \delta - x_i\beta - F\phi\bar{x}_i$.
The expected likelihood is

$$Q(\theta) = -\frac{N}{2}\ln|D| - \frac{1}{2}\sum_{i=1}^{N}\left[u_i' D^{-1} u_i - 2u_i' D^{-1} F\, E(\eta_i|u_i) + \right.$$

$$\left. \mathrm{tr}[F' D^{-1} F\, E(\eta_i\eta_i'|u_i)]\right]$$

$$-\frac{N}{2}\ln|\Phi| - \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}[\Phi^{-1} E(\eta_i\eta_i'|u_i)]$$

First order conditions

$$F = \sum_{i=1}^{N} v_i(\bar{x}_i'\phi' + \hat{\eta}_i')\Big[\sum_{i=1}^{N}\Big(\phi\bar{x}_i\bar{x}_i'\phi' + \hat{\eta}_i\bar{x}_i'\phi' + \phi\bar{x}_i\hat{\eta}_i' + \widehat{\eta_i\eta_i'}\Big)\Big]^{-1}$$

$$\delta = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i\beta - F\phi\bar{x}_i - F\hat{\eta}_i)$$

$$D = \mathrm{diag}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big(u_i u_i' - 2F\hat{\eta}_i u_i' + F\widehat{\eta_i\eta_i'}F'\Big)\Big]$$

$$\Phi = \frac{1}{N}\sum_{i=1}^{N}\widehat{\eta_i\eta_i'}$$

and

$$\theta_1 = \Big[\sum_{i=1}^{N} X_i' D^{-1} X_i\Big]^{-1}\sum_{i=1}^{N}\Big[X_i' D^{-1}\Big(y_i - \delta - F\hat{\eta}_i\Big)\Big]$$

where $v_i = y_i - \delta - x_i\beta$, $\theta_1 = (\beta, \phi)$.

ECM, Meng and Rubin, 1993

$$Q(\theta_1, \theta_2 | \theta_1^{(k)}, \theta_2^{(k)})$$

$$CM1: \quad \max_{\theta_2} \; Q(\theta_1^{(k)}, \theta_2 | \theta_1^{(k)}, \theta_2^{(k)})$$

$$CM2: \quad \max_{\theta_1} \; Q(\theta_1, \theta_2^{(k+1)} | \theta_1^{(k)}, \theta_2^{(k)})$$

CM1 gives $\theta_2^{(k+1)}$ and CM2 gives $\theta_1^{(k+1)}$. The obtained $(k+1)$ step estimator is different from the joint maximization of $Q(\theta|\theta^{(k)})$, but guarantees an increase in the likelihood function.

In our case, three groups, or two groups,

$$\theta_3 = (F, \Phi), \quad \theta_2 = (\delta, D), \quad \theta_1 = (\beta, \phi)$$

For two groups,

$$\theta_2 = (F, \delta, D, \Phi), \quad \theta_1 = (\beta, \phi)$$

and F and δ must be estimated simultaneously,

$$F^* = (F, \delta)$$

$$\eta_i^* = \begin{bmatrix} 1 \\ \eta_i \end{bmatrix}$$

so that

$$F^* \eta_i^* = \delta + F\delta_i$$

The conditional mean of η_i^* and conditional second moments for η_i^* is straightforward.

ECME, Liu and Rubin, 1994 (skipped this page)

Instead of maximizing the Q function, some of the CM steps can be taken by maximizing the actual likelihood function.
For example,

$$\theta_1^{(k+1)} = \Big[\sum_{i=1}^{N} X_i^{(k+1)\prime} (\Omega^{(k+1)})^{-1} X_i^{(k+1)} \Big]^{-1} \cdot$$

$$\sum_{i=1}^{N} \Big[X_i^{(k+1)\prime} (\Omega^{(k+1)})^{-1} \Big(y_i - \delta^{(k+1)} \Big) \Big].$$

Simulations

Data:

$$y_{it} = \delta_t + \beta_1 x_{it,1} + \beta_2 x_{it,2} + \lambda_i f_t + \sigma_t \varepsilon_{it}$$

$$x_{it,k} = 1 + \lambda_i' f_t + \xi_{it}$$

$$k = 1, 2$$

where $\lambda_i, f_t, \varepsilon_{it}, \xi_{it}$ are all iid N(0,1); $\beta_1 = 1, \beta_2 = 2$, $\sigma_t = \sqrt{t}$ so matrix D is

$$D = \mathrm{diag}(1, 2, ..., T)$$

Estimators: all with time effects and full projection
OLS
GLS
MLE

Means and standard deviations from 1000 repetitions

N	T	OLS		GLS		MLE	
		$\beta_1 = 1$	$\beta_2 = 2$	$\beta_1 = 1$	$\beta_2 = 2$	$\beta_1 = 1$	$\beta_2 = 2$
100	5	1.2360	2.2397	1.0166	2.0094	1.0365	2.0343
		0.0700	0.0697	0.2461	0.2537	0.1226	0.1225
500	5	1.4112	2.4097	1.0049	2.0045	1.0105	2.0081
		0.0280	0.0283	0.1014	0.0987	0.0710	0.0718
100	10	1.2644	2.2636	1.0265	2.0136	1.0238	2.0233
		0.0715	0.0678	0.3747	0.3772	0.1053	0.0965
500	10	1.2124	2.2119	1.0107	2.0094	1.0001	1.9992
		0.0299	0.0308	0.1436	0.1412	0.0268	0.0276

For each (N, T) combination, the first row is the sample mean and the second row is the sample standard deviation

OLS inconsistent, GLS consistent, less efficient than MLE.

Dynamic panels:

$$y_{it} = \alpha y_{it-1} + x_{it}'\beta + \lambda_i' f_t + \varepsilon_{it}$$

For each T (either 5 or 10), simulate 2T observations, and discard the first half. $\mathrm{var}(\varepsilon_{it}) = 1$ for the first half. The retained sample has variance

$$\mathrm{var}(\varepsilon_{it}) = t, \quad t = 1, 2, ..., T$$

$$\alpha = 0.5, \quad 1.0$$

all other variables are the same as the previous case

			OLS			GLS			MLE		
α	N	T	α	$\beta_1 = 1$	$\beta_2 = 2$	α	$\beta_1 = 1$	$\beta_2 = 2$	α	$\beta_1 = 1$	$\beta_2 = 2$
0.5	100	5	0.528	1.184	2.191	0.514	1.121	2.122	0.513	1.054	2.062
			0.019	0.071	0.078	0.064	0.228	0.222	0.037	0.085	0.100
	500	5	0.499	1.308	2.308	0.497	1.005	1.997	0.499	1.001	2.003
			0.008	0.030	0.030	0.025	0.101	0.106	0.012	0.032	0.035
	100	10	0.495	1.288	2.282	0.498	1.026	1.990	0.501	1.010	2.010
			0.016	0.073	0.067	0.039	0.378	0.384	0.023	0.081	0.080
	500	10	0.501	1.330	2.336	0.500	1.022	2.022	0.498	1.000	2.004
			0.006	0.027	0.027	0.018	0.150	0.142	0.009	0.028	0.028
1.0	100	5	0.989	1.174	2.180	1.003	1.077	2.081	1.003	1.036	2.040
			0.005	0.073	0.078	0.048	0.234	0.232	0.027	0.080	0.091
	500	5	1.002	1.308	2.308	1.001	1.006	1.997	1.000	1.001	2.003
			0.005	0.030	0.030	0.020	0.101	0.105	0.006	0.032	0.035
	100	10	0.994	1.282	2.275	0.998	1.022	1.993	1.000	1.014	2.015
			0.005	0.074	0.068	0.024	0.380	0.378	0.008	0.086	0.087
	500	10	0.995	1.326	2.332	1.002	1.022	2.021	1.000	1.000	2.004
			0.003	0.026	0.027	0.010	0.150	0.141	0.003	0.028	0.028

1. Interestingly, the OLS of α is not biased (due to large heteroskedasticity), but slope coefficients are biased.
2. GLS is consistent, but have large standard errors (this is a crude GLS, not the one proposed).
3. MLE is okay, still possible improvements (initial values, number of iterations, the number of CM steps, use of ECME, etc).
4. Not reported, the heteroskedasticities are estimated very well.

Summary:

- Small T panel data

factor error structure correlated with regressors

- Robust to heteroskedasticity and serial correlation
- Controlling the correlation by the Chamberlain and Mundlak method
- Treating F as parameters, and estimating a model of random effects
- Careful analysis of the initial conditions
- Predetermined regressors, weakly exogenous or non-weakly exogenous regressors
- Feasible computation

Thank you!