Inferences of panel data models with interactive effects using large covariance matrices

Jushan Bai and Yuan Liao

2017 China meeting

Consider

$$y_{it} = \beta' X_{it} + \lambda_i' f_t + u_{it}$$

- Goal: efficient estimation of β using GLS.

$$\min_{\beta} \min_{\Lambda, F} \sum_{t=1}^{T} (y_t - X_t\beta - \Lambda f_t) \Sigma_u^{-1} (y_t - X_t\beta - \Lambda f_t)$$

- Here $\Sigma_u = \mathrm{cov}(u_t)$.
- We assume:

  (i) Σ_u is sparse (locations of zeros are unknown)

  (ii) serial independence in $\{u_t : t \geq 1\}$

Findings

1. No cross-sectional correlation bias
2. Faster rate of convergence when $N = o(T)$.
3. Σ_u^{-1} is the first-order optimal weight.
4. Applied to studying the effect of law reform on divorce rates, concluded significance in long-run effects
5. Technically, it is challenging to show:

$$\text{the effect of } \Sigma_u^{-1} - \widehat{\Sigma}_u^{-1} \text{ is first order negligible}$$

$$\widehat{\beta}(W) = \arg\min_{\beta}\min_{\Lambda,F}\sum_{t=1}^{T}(y_t - X_t\beta - \Lambda f_t)W(y_t - X_t\beta - \Lambda f_t)$$

- Bai 09 showed:

$$\sqrt{NT}[\widehat{\beta}(I) - \beta - BIAS] \to^d N(0, V)$$

where $BIAS = O_P(\frac{1}{N})$ is due to cross-sectional correlations in u_{it}.

$$\|\widehat{\beta}(I) - \beta\| = O_P(\frac{1}{\sqrt{NT}} + \frac{1}{N})$$

- In contrast,

$$\sqrt{NT}[\widehat{\beta}(\widehat{\Sigma}_u^{-1}) - \beta] \to^d N(0, V_2)$$

- $V_2 \leq V$.

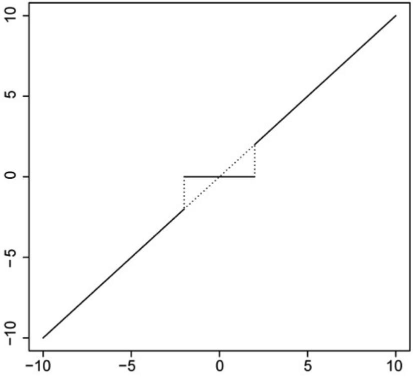
Estimating Σ_u

1. Obtain $\widehat{\beta}(I)$ and $\widehat{\Lambda}, \widehat{f}_t$
2. Compute $\widehat{u}_t = y_t - X_t\widehat{\beta}(I) - \widehat{\Lambda}\widehat{f}_t$ and $(\widehat{r}_{ij}) = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t\widehat{u}_t'$.
3. Apply (adaptive) thresholding

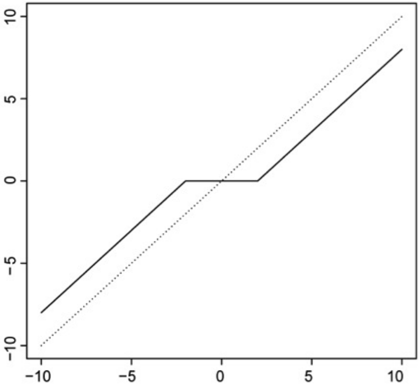
$$\widehat{\Sigma}_u = (s_{ij}(\hat{r}_{ij})).$$

$s_{ij}(\cdot)$: hard, soft, ,...

Positive definite in finite sample even when $N > T$



Hard-thresholding,



Soft-thresholding

- Hard-thresholding: $s(t) = t1\{|t| > \tau\}$

  Soft-thresholding: $s(t) = (t - \tau)1\{|t| > \tau\}$ when $t > 0$; symmetric when $t < 0$.

$$\tau = C\left(\sqrt{\frac{\log N}{T}} + \frac{1}{\sqrt{N}}\right)$$

- $C > 0$ is chosen to maintain positive definite for finite sample.

Both cross-sec. and serial correlations are present

$$\min_{\beta}\min_{F,\Lambda}[Y - X\beta - \mathrm{vec}(\Lambda F')]'\Sigma_U^{-1}[Y - X\beta - \mathrm{vec}(\Lambda F')]$$

where Y is $(NT) \times 1$ and X is $(NT) \times d$.

- Need to estimate $(NT) \times (NT)$ matrix Σ_U.
- Its (t, s) th block is $N \times N$ matrix $Eu_t u_s'$, which "decays" as $|t - s| \to \infty$.
- Construct $\widehat{\Sigma}_U$ as many "blocks" $\widehat{Eu_t u_s'}$

$$\widehat{Eu_t u_s'} = \begin{cases} \text{soft-thresholding}, & |t - s| < \ell_{TN} \\ 0, & |t - s| > \ell_{TN} \end{cases}$$

- Choice of ℓ_{TN} depends on (i) mixing condition, (ii) amount of accumulated estimation errors.

Technical challenges: $\widehat{\Sigma}_u^{-1} - \Sigma_u^{-1} \approx 0$

$$\hat{\beta} = f(D_T, \Sigma_u^{-1}),$$

- Substitute to obtain feasible estimator.

$$\hat{\beta}^* = f(D_T, \hat{\Sigma}_u^{-1})$$

- Similar to Bai 09, $\sqrt{NT}(\hat{\beta} - \beta) = \sqrt{NT} A_1 \Sigma_u^{-1} A_2 + o_p(1)$
- Show effect of estimating Σ is negligible:

$$\sqrt{NT}(\hat{\beta} - \hat{\beta}^*) = \sqrt{NT}\mathbf{A_1}(\Sigma_\mathbf{u}^{-1} - \widehat{\Sigma}_\mathbf{u}^{-1})\mathbf{A_2} + o_p(1) = o_p(1)$$

$$\|A_1\| = O_P(\sqrt{NT}) = \|A_2\|.$$

- In high dimensions, very challenging even if $\|\Sigma_u^{-1} - \widehat{\Sigma}_u^{-1}\|$ achieves the optimal rate

- Need **weighted convergence**

$$\sqrt{NT}A_1'(\widehat{\Sigma}_u^{-1} - \Sigma_u^{-1})A_2$$

- $\sqrt{NT}\|A_1\|\|A_2\|\|\widehat{\Sigma}_u^{-1} - \Sigma_u^{-1}\| \neq o_p(1)$
- Need to expand $A_1'(\widehat{\Sigma}_u^{-1} - \Sigma_u^{-1})A_2$
- Also need to assume

$$\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{(\Sigma_u)_{ij} \text{ is "large"}}(u_{is}u_{js} - Eu_{is}u_{js})w_{ji} = o_P(\sqrt{T}).$$

where $w_{ji} = O_P(1)$ is some random weight.

Empirical study: effect of law reforms

- Sharp increase of divorce rate in 1960-1970s in U.S.
- Also observe decrease trend eight years after law reforms
- Wolfers (2006) found that
  "the divorce rate rose sharply following the adoption of unilateral divorce laws, but this rise was reversed within about a decade."
- So law reform may have two-sided effects.

The model

$$y_{it} = \sum_{k=1}^{K} X_{it,k}\beta_k + g_{it} + f(\delta_i, t) + u_{it}, \tag{1}$$

$$X_{it,k} = \begin{cases} 1, & 2k-1 \leq t - T_i \leq 2k \\ 0 & \end{cases}$$

whether t is $2k$ years after the reform.

- Omitted variables may be present, leading to endogeneity.
- interactive effects control for unobserved heterogeneity
- Kim and Oka (14): $g_{it} = \lambda_i' f_t$.
- They concluded insignificance when $k > 4$.
- GLS: yields tighter confidence intervals.

Table: 95% confidence intervals; intervals with * are significant. No time trend

	WPC		PC		Relative
	estimate	con. interval	estimate	con. interval	efficiency
1-2 ys	0.014	[0.007, 0.021]*	0.018	[0.0091, 0.028]*	0.59
3-4 ys	0.034	[0.027, 0.041]*	0.042	[0.032, 0.053]*	0.59
5-6 ys	0.025	[0.017, 0.032]*	0.032	[0.022, 0.042]*	0.58
7-8 ys	0.015	[0.007, 0.023]*	0.030	[0.019, 0.04]*	0.56
9-10 ys	-0.006	[-0.014, 0.001]	0.008	[-0.002, 0.018]	0.56
11-12 ys	-0.008	[-0.015, -0.001]*	0.010	[-0.001, 0.02]	0.53
13-14 ys	-0.009	[-0.017, -0.001]*	0.005	[-0.005, 0.016]	0.53
15 ys+	0.009	[0.001, 0.017]*	0.031	[0.020, 0.042]*	0.55

Table: Linear time trend

	WPC		PC		Relative
	estimate	con. interval	estimate	con. interval	efficiency
1-2 ys	0.014	[0.006, 0.021]*	0.016	[0.006, 0.026]*	0.55
3-4 ys	0.032	[0.024, 0.039]*	0.037	[0.026, 0.047]*	0.54
5-6 ys	0.018	[0.010, 0.026]*	0.024	[0.012, 0.035]*	0.54
7-8 ys	0.006	[-0.002, 0.014]	0.017	[0.005, 0.028]*	0.52
9-10 ys	-0.017	[-0.025, -0.008]*	-0.007	[-0.019, 0.005]	0.52
11-12 ys	-0.019	[-0.028, -0.010]*	-0.006	[-0.018, 0.006]	0.51
13-14 ys	-0.021	[-0.030, -0.012]*	-0.012	[-0.025, 0.001]	0.50
15 ys+	-0.003	[-0.012, 0.006]	0.014	[0.000, 0.028]*	0.46

Simulations

$$y_{it} = \lambda_i' f_t + u_{it}$$

$$y_{it} = x_{it}'\beta + \lambda_i' f_t + u_{it}$$

- u_{it} follows a cross-sectional MA(3) model, so Σ_u is a banded matrix.

$$X_{it,1} = 2.5\lambda_{i1} f_{1,t} - 0.2\lambda_{i2} f_{2,t} - 1 + \eta_{it,1},$$

$$X_{it,2} = \lambda_{i1} f_{1,t} - 2\lambda_{i2} f_{2,t} + 1 + \eta_{it,2}$$

- true $\beta = (1, 3)'$.

Table: interactive effects

		$\beta_1 = 1$				$\beta_2 = 3$			
		Mean		Normalized SE		Mean		Norm.SE	
T	N	WPC	PC	WPC	PC	WPC	PC	WPC	PC
50	75	1.005	1.013	0.758	1.413	2.99	3.00	0.744	1.472
50	100	1.005	1.010	0.662	1.606	2.99	2.99	0.731	1.616
50	150	1.004	1.008	0.964	1.913	2.99	2.99	0.951	1.881
100	100	1.002	1.010	0.550	1.418	3.00	3.00	0.416	1.353
100	150	1.003	1.007	0.681	1.626	2.99	3.00	0.611	1.683
100	200	1.002	1.005	0.631	1.800	3.00	3.00	0.774	1.752
150	100	1.003	1.006	0.772	1.399	3.00	2.99	0.714	1.458
150	150	1.001	1.005	0.359	1.318	3.00	3.00	0.408	1.379
150	200	1.001	1.003	0.547	1.566	3.00	3.00	0.602	1.762

Summary

- Large covariances are important for inference
- Effect of estimating covariance on inference is negligible, but tech. non-trivial
- Current “absolute convergence” is restrictive (for inference)
- More efficient, smaller asym. variance