Nonparametric Estimation in Panel Data Models with Heterogeneity and Time–Varyingness

Jiti Gao[†], Fei Liu[†], Yanrong Yang[‡]

Monash University[†]

Australian National University[‡]

Dec 12, 2019

An Econometric Problem

Panel Data Analysis

1. Data Structure: Dependent Variable y_{it} and Independent Variable $\mathbf{x}_{it} = (X_{1,it}, X_{2,it}, \ldots, X_{p,it})$ with $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.
2. Aim: Accurately model and estimate the relation between y_{it} and $\mathbf{x}_{it}$ for all cross-sections $i = 1, 2, \ldots, N$ and time-periods $t = 1, 2, \ldots, T$.
3. Major Benefit: Homogeneity (Blessing of Dimensionality).
4. Challenge: Heterogeneity (Curse of Dimensionality).

Literature Review

Bai (2009, Econometrica)

Common factor models are widely used to capture cross-sectional dependence in panel data sets:

$$y_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta} + e_{it}, \quad e_{it} = \boldsymbol{\lambda}_i^\top \mathbf{F}_t + \varepsilon_{it} \tag{1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where

- $\boldsymbol{\beta}$ is a p-dimensional unknown parameter;
- $\{\mathbf{F}_t\}$ are unknown r-dimensional common factors;
- $\{\boldsymbol{\lambda}_i\}$ are corresponding factor loadings.

Advantages of factor models:

- heterogenous effects of common shocks;
- Appropriate flexibility.

Literature Review

Bai (2009, Econometrica)

- Bai (2009) proposes an iterative numerical method to approximate the minimizer of the least squares objective function:

$$\mathrm{SSR} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \mathbf{x}_{it}^{\top} \boldsymbol{\beta} - \boldsymbol{\lambda}_i^{\top} \mathbf{F}_t \right)^2 \tag{2}$$

  - Estimate $\boldsymbol{\beta}$ by least squares method;
  - Estimate $\boldsymbol{\lambda}_i$ and $\mathbf{F}_t$ by PCA method;
  - Repeat until convergence.
- Extensions:
  - Ando and Bai (2014).
- Challenges:
  - Poor performance with endogenous factors (see Jiang et al., 2017).

Literature Review

Pesaran (2006, Econometrica)

- Pesaran (2006) proposes valid proxies for $\mathbf{F}_t$ in the following model:

$$\begin{pmatrix} y_{it} \\ \mathbf{x}_{it} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\lambda}_i^\top + \boldsymbol{\beta}_i^\top \gamma_i^\top \\ \gamma_i^\top \end{pmatrix} \mathbf{F}_t + \begin{pmatrix} \varepsilon_{it} + \boldsymbol{\beta}_i^\top \boldsymbol{\eta}_{it} \\ \boldsymbol{\eta}_{it} \end{pmatrix}, \tag{3}$$

  where $\{\gamma_i\}$ are unknown factor loadings.
- Extensions: Chudik and Pesaran (2015).
- Challenges:
  - Rank condition $r \leq p + 1$,
  - No estimators for $\mathbf{F}_t$, $\boldsymbol{\lambda}_i$.

Literature Review

Time-varying panel data models

- Limitations of time-constant slope coefficients:
  - The risk of model misspecification;
  - The time-variation in parameters has been well recognized in many fields:
    - Silvapulle et al. (2017).
- Existing time-varying panel data models:
  - Li et al. (2011):

$$y_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta}_t + f_t + \alpha_i + \varepsilon_{it}; \tag{4}$$

where $\boldsymbol{\beta}_t = \boldsymbol{\beta}(\tau_t)$ and $f_t = f(\tau_t)$ with $\tau_t = \frac{t}{T}$.

Literature Review

Heterogeneous panel data models

- Existing heterogeneous panel data models:
  - Pesaran (2006)'s random coefficient assumption:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{u}_i. \tag{5}$$

  - Su et al. (2016)'s unknown group pattern:

$$\boldsymbol{\beta}_i = \sum_{k=1}^{K} \boldsymbol{\beta}^{(k)} \mathbf{1}\{i \in G_k\}, \tag{6}$$

    where K is known and fixed but G_k is unknown.
  - Gao et al. (2019)'s complete heterogeneity:

$$y_{it} = \mathbf{x}_{it}^{\top} \boldsymbol{\beta}_i + f_{it} + \alpha_i + \varepsilon_{it}, \tag{7}$$

    where $f_{it} = f_i(\tau_t)$.

Proposed Model

Our model

- We consider the following model:

$$y_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta}_{it} + \boldsymbol{\lambda}_i^\top \mathbf{F}_t + \varepsilon_{it}, \tag{8}$$

  where

  - $\mathbf{x}_{it}$ and y_{it} are observable;
  - $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_i(\tau_t)$ is an unknown deterministic function;
  - $\mathbf{x}_{it}$ can be correlated with $\{\boldsymbol{\lambda}_i, \mathbf{F}_t\}$.

Outline of Contribution

1. Generality of Model: Heterogeneous and Time-varying coefficients.
2. Unified Estimation Approach: observed, unobserved or partially observed factors.
3. Asymptotic Theory: reconcile computational elements (iteration steps) with statistical properties.
4. Empirical Application: relation between health care expenditure and income elasticity.

Proposed Estimation Approach

Recall the heterogeneous model:

$$y_{it} = \mathbf{x}_{it}^{\top} \boldsymbol{\beta}_i(\tau_t) + \boldsymbol{\lambda}_i^{\top} \mathbf{F}_t + \varepsilon_{it}.$$

The idea of iteration:

- With given $\mathbf{F}_t$, we can estimate $\boldsymbol{\beta}_i(\tau)$ and $\boldsymbol{\lambda}_i$ by a profile method.
- With $\boldsymbol{\beta}_i(\tau)$ and $\boldsymbol{\lambda}_i$, $\mathbf{F}_t$ can be estimated by OLS method.

Estimation Procedure

(1) Find an initial estimator $\widehat{\mathbf{F}}^{(0)} = (\widehat{\mathbf{F}}_1^{(0)}, \ldots, \widehat{\mathbf{F}}_T^{(0)})^\top$.

(2) With $\widehat{\mathbf{F}}_t^{(n)}$ and by regarding $\boldsymbol{\lambda}_i$ as known, $\boldsymbol{\beta}_i(\tau)$ can be estimated by local linear method. For $\tau \in (0,1)$

$$\min_{\mathbf{a}_i(\tau), \mathbf{b}_i(\tau)} \sum_{t=1}^{T} \left(y_{it} - \boldsymbol{\lambda}_i^\top \widehat{\mathbf{F}}_t^{(n)} - \mathbf{x}_{it}^\top \left(\mathbf{a}_i(\tau) + \left(\frac{t - \tau T}{Th} \right) \mathbf{b}_i(\tau) \right) \right)^2 K \left(\frac{t - \tau T}{Th} \right), \quad (9)$$

we have

$$\widehat{\boldsymbol{\beta}}_i^{(n+1)}(\tau, \boldsymbol{\lambda}_i) = [\mathbf{I}_p, \mathbf{0}_p] \left[\mathbf{M}_i(\tau)^\top \mathbf{W}(\tau) \mathbf{M}_i(\tau) \right]^{-1} \mathbf{M}_i(\tau)^\top \mathbf{W}(\tau) \left[\mathbf{y}_i - \widehat{\mathbf{F}}^{(n)} \boldsymbol{\lambda}_i \right]. \quad (10)$$

(3) With $\widehat{\boldsymbol{\beta}}_i(\tau, \boldsymbol{\lambda}_i)$, we can estimate $\boldsymbol{\lambda}_i$ by the least squares method:

$$\min_{\boldsymbol{\lambda}_i} \sum_{t=1}^{T} \left(y_{it} - \mathbf{x}_{it}^\top \widehat{\boldsymbol{\beta}}_i^{(n+1)}(\tau, \boldsymbol{\lambda}_i) - \boldsymbol{\lambda}_i^\top \widehat{\mathbf{F}}_t^{(n)} \right)^2. \quad (11)$$

See notation

Estimation Procedure

We have

$$\widehat{\boldsymbol{\lambda}}_i^{(n+1)} = \left[\widehat{\mathbf{F}}^{(n)\top}(\mathbf{I}-\mathbf{S}_i)^\top(\mathbf{I}-\mathbf{S}_i)\widehat{\mathbf{F}}^{(n)}\right]^{-1}\widehat{\mathbf{F}}^{(n)\top}(\mathbf{I}-\mathbf{S}_i)^\top(\mathbf{I}-\mathbf{S}_i)\mathbf{y}_i, \qquad (12)$$

where

$$\mathbf{S}_i = (\mathbf{s}_i(1/T)^\top\mathbf{x}_{i1},\ldots,\mathbf{s}_i(T/T)^\top\mathbf{x}_{iT})^\top,$$

with

$$\mathbf{s}_i(\tau) = [\mathbf{I}_p,\mathbf{0}_p][\mathbf{M}_i(\tau)^\top\mathbf{W}(\tau)\mathbf{M}_i(\tau)]^{-1}\mathbf{M}_i(\tau)^\top\mathbf{W}(\tau).$$

After plugging $\widehat{\boldsymbol{\lambda}}_i$ back into $\widehat{\boldsymbol{\beta}}_i(\tau,\boldsymbol{\lambda}_i)$, we have

$$\widehat{\boldsymbol{\beta}}_i^{(n+1)}(\tau) = [\mathbf{I}_p,\mathbf{0}_p]\left[\mathbf{M}_i(\tau)^\top\mathbf{W}(\tau)\mathbf{M}_i(\tau)\right]^{-1}\mathbf{M}_i(\tau)^\top\mathbf{W}(\tau)\left[\mathbf{y}_i-\widehat{\mathbf{F}}^{(n)}\widehat{\boldsymbol{\lambda}}_i^{(n+1)}\right] \qquad (13)$$

for $i=1,\ldots,N$.

Estimation Procedure

(4) With $\widehat{\boldsymbol{\beta}}_i^{(n+1)}(\tau)$ and $\widehat{\boldsymbol{\lambda}}_i^{(n+1)}$, we can estimate $\mathbf{F}_t$ by OLS method:

$$\widehat{\mathbf{F}}_t^{(n+1)} = \left(\widehat{\boldsymbol{\Lambda}}^{(n+1)\top}\widehat{\boldsymbol{\Lambda}}^{(n+1)}\right)^{-1}\widehat{\boldsymbol{\Lambda}}^{(n+1)\top}\mathbf{R}_{1,t}^{(n+1)}$$

where $\mathbf{R}_{1,t}^{(n+1)} = \left(y_{1t} - \mathbf{x}_{1t}^\top\widehat{\boldsymbol{\beta}}_1^{(n+1)}(\tau_t), \ldots, y_{Nt} - \mathbf{x}_{Nt}^\top\widehat{\boldsymbol{\beta}}_N^{(n+1)}(\tau_t)\right)^\top$.

(5) Repeat Steps 2-4 until convergence.

Asymptotic Properties

Assumption 1

(i-v) Regularity assumptions on weak serial and cross-sectional dependence and kernel estimation.

(vi) Let $\mathbf{R}_F^{(n)} = \widehat{\mathbf{F}}^{(n)} - \mathbf{F}^0$. For the initial estimator $\widehat{\mathbf{F}}^{(0)}$, suppose that

$$T^{-1/2}\|\mathbf{R}_F^{(0)}\| = O_P\left(\delta_{F,0}\right) \quad \text{and} \quad (Th)^{-1/2}\|\mathbf{W}(\tau)^\top \mathbf{R}_F^{(0)}\| = O_P\left(\delta_{F,0}\right),$$

where $\delta_{F,0}$ satisfies that $NTh^4\delta_{F,0}^2 \to 0$, $\delta_{F,0}^2/h \to 0$ and $\max\{N,T\}\delta_{F,0}^4/h \to 0$, as $N, T \to \infty$.

Assumption 2

(i-iv) Regularity assumptions on positive definiteness of asymptotic covariance matrices.

See Assumptions

Asymptotic Properties

Theorem 2.1 (Consistency) Under Assumption 1, as $N, T \to \infty$ simultaneously,

(1) $N^{-1/2} \left\| \widehat{\mathbf{\Lambda}}^{(n)} - \mathbf{\Lambda} \right\| = O_p \left(\max \left\{ \delta_{F,0}, \delta_{NT} \right\} \right)$;

(2) $T^{-1/2} \left\| \widehat{\mathbf{F}}^{(n)} - \mathbf{F} \right\| = O_p \left(\max \left\{ \delta_{F,0}, \delta_{NT} \right\} \right)$,

where $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}^{-1}$.

Asymptotic Properties

Assume that

$$\mathbf{x}_{it} = \mathbf{g}_i(\tau_t) + \mathbf{v}_{it}. \tag{14}$$

Notations:

$$\boldsymbol{\Sigma}_{v,i} = E\left[\mathbf{v}_{i1}\mathbf{v}_{i1}^\top\right], \quad \boldsymbol{\Sigma}_F = E\left[\mathbf{F}_1^0\mathbf{F}_1^{0\top}\right], \quad \boldsymbol{\Sigma}_{v,F,i} = E\left[\mathbf{v}_{it}\mathbf{F}_t^{0\top}\right], \quad \boldsymbol{\Sigma}_{v,\lambda,i} = E\left[\mathbf{v}_{it}\boldsymbol{\lambda}_i^{0\top}\right],$$

$$\boldsymbol{\Sigma}_{X,i}(\tau) = \mathbf{g}_i(\tau)\mathbf{g}_i^\top(\tau) + \boldsymbol{\Sigma}_{v,i}, \quad \boldsymbol{\Omega}_{F,i} = \boldsymbol{\Sigma}_F - \boldsymbol{\Sigma}_{v,F,i}^\top \int_0^1 \boldsymbol{\Sigma}_{X,i}^{-1}(\tau)d\tau\boldsymbol{\Sigma}_{v,F,i},$$

$$\sigma_{ij,ts} = E[\varepsilon_{it}\varepsilon_{js}], \quad \mathbf{z}_{it} = \mathbf{F}_t^0 - \boldsymbol{\Sigma}_{v,F,i}^\top\boldsymbol{\Sigma}_{X,i}^{-1}(\tau_t)\mathbf{x}_{it}, \quad \boldsymbol{\Sigma}_\lambda = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N \boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\top},$$

$$\boldsymbol{\Delta}_{F,i} = \boldsymbol{\Sigma}_{v,F,i}\boldsymbol{\Omega}_{F,i}^{-1}\boldsymbol{\Sigma}_{v,F,i}^\top, \quad \boldsymbol{\lambda}_i^\dagger(\tau) = \boldsymbol{\Sigma}_{X,i}^{-1}(\tau)\left(\boldsymbol{\Sigma}_{v,\lambda,i}(\tau) + \mathbf{g}_i(\tau)\boldsymbol{\lambda}_i^{0\top}\right),$$

$$\boldsymbol{\Omega}_1(t,s) = N^{-1}\sum_{i=1}^N E\left[\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\top}\mathbf{x}_{it}^\top\boldsymbol{\Sigma}_{X,i}^{-1}(\tau_t)\mathbf{x}_{is}\right], \quad \boldsymbol{\Omega}_2(t,s) = N^{-1}\sum_{i=1}^N E\left[\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\top}\mathbf{z}_{it}^\top\boldsymbol{\Omega}_{F,i}^{-1}\mathbf{z}_{is}\right],$$

$$\boldsymbol{\Omega}_3(t,s) = \boldsymbol{\Sigma}_\lambda^{-1}(h^{-1}K_{s,0}(\tau_t)\boldsymbol{\Omega}_1(t,s) + \boldsymbol{\Omega}_2(t,s)),$$

Asymptotic Properties

Theorem 2.2 (CLT, $n \geq 2$) Let Assumptions 1 and 2 hold. Then, as $N, T \to \infty$ simultaneously,

(1) if $N/T \to c_1 < \infty$, for any given t, we have

$$\sqrt{N}\left(\widehat{\mathbf{F}}_t^{(n)} - \mathbf{F}_t^0 - \mathbf{b}_{F,t}^{\dagger(n)}\right) \xrightarrow{D} \mathcal{N}(\sqrt{c_1}\mathbf{d}_{F,t}, \mathbf{\Sigma}_{F,t}),$$

where $\mathbf{\Sigma}_{F,t} = \mathbf{\Sigma}_\lambda^{-1}\mathbf{\Sigma}_{F,t}^0\mathbf{\Sigma}_\lambda^{-1}$,

$$\mathbf{b}_{F,t}^{\dagger(n)} = T^{-n} \sum_{s_1,s_2,\ldots,s_n=1}^{T} \mathbf{\Omega}_3(t, s_1) \prod_{j=1}^{n-1} \mathbf{\Omega}_3(s_j, s_{j+1}))\mathbf{R}_{F,s_n}^{(0)},$$

$$\mathbf{d}_{F,t} = \lim_{N,T\to\infty} 1/(N\sqrt{T})\mathbf{\Sigma}_\lambda^{-1} \sum_{i=1}^{N}\sum_{s=1}^{T} \mathbf{\Omega}_{F,i}^{-1}\mathbf{\Sigma}_{v,F,i}^{\top}\mathbf{\Sigma}_{X,i}^{-1}(\tau_s)\mathbf{g}_i(\tau_s)\sigma_{ii,ts}.$$

See Assumptions

Asymptotic Properties

Theorem 2.2 (CLT, $n \geq 2$) Let Assumptions 1 and 2 hold. Then, and as $N, T \to \infty$ simultaneously,

(2) if $T/N \to c_2 < \infty$, for any given i, we have

$$\sqrt{T}\left(\widehat{\boldsymbol{\lambda}}_i^{(n)} - \boldsymbol{\lambda}_i^0 - \mathbf{b}_{\lambda,i}^{\dagger(n)}\right) \xrightarrow{D} \mathcal{N}(\sqrt{c_2}\mathbf{d}_{\lambda,i}, \boldsymbol{\Sigma}_{\lambda,i}),$$

where $\boldsymbol{\Sigma}_{\lambda,i} = \boldsymbol{\Omega}_{F,i}^{-1}\boldsymbol{\Sigma}_{\lambda,i}^0\boldsymbol{\Omega}_{F,i}^{-1}$,

$$\mathbf{b}_{\lambda,i}^{\dagger(n)} = T^{-1}\boldsymbol{\Omega}_{F,i}^{-1}\boldsymbol{\Sigma}_{v,F,i}^{\top}\sum_{t=1}^{T}\boldsymbol{\lambda}_i^{\dagger}(\tau_t)\mathbf{b}_{F,t}^{\dagger(n-1)},$$

$$\mathbf{d}_{\lambda,i}^{*} = 1/\sqrt{N}\boldsymbol{\Omega}_{F,i}^{-1}\boldsymbol{\Sigma}_{\lambda}^{-1}\boldsymbol{\mu}_{\lambda}\sum_{j=1}^{N}\sigma_{ij,11}.$$

▸ See Assumptions

Asymptotic Properties

Theorem 2.2 (CLT, $n \geq 2$) Let Assumptions 1 and 2 hold. Then, as $N, T \to \infty$ simultaneously,

(3) for any given (i, τ), we have

$$\sqrt{Th}\left(\widehat{\boldsymbol{\beta}}_i^{(n)}(\tau) - \boldsymbol{\beta}_i(\tau) - \mathbf{a}_i(\tau)h^2 - \mathbf{b}_{\beta,i}^{\dagger(n)}(\tau)\right) \xrightarrow{D} \mathcal{N}(0_p, \boldsymbol{\Sigma}_{\beta,i}(\tau)),$$

where $\mathbf{a}_i(\tau) = \frac{\mu_2}{2}\boldsymbol{\beta}_i''(\tau)(1 + o(1))$, $\boldsymbol{\Sigma}_{\beta,i}(\tau) = \boldsymbol{\Sigma}_{X,i}^{-1}(\tau)\boldsymbol{\Sigma}_{\beta,i}^{0}(\tau)\boldsymbol{\Sigma}_{X,i}^{-1}(\tau)$, $\mu_2 = \int u^2 K(u)du$, and

$$\mathbf{b}_{\beta,i}^{\dagger(n)}(\tau) \quad = \quad -T^{-1}\boldsymbol{\Sigma}_{X,i}^{-1}(\tau)\sum_{t=1}^{T}\left(h^{-1}K_{t,0}(\tau)\boldsymbol{\Sigma}_{X,i}(\tau) + \boldsymbol{\Delta}_{F,i}\right)\boldsymbol{\lambda}_i^{\dagger}(\tau_t)\mathbf{b}_{F,t}^{\dagger(n-1)}.$$

See Assumptions

Asymptotic Properties

Corollary 2.1 (CLT, $n \geq 2$) Let Assumptions 1 and 2 hold. If ε_{it} is both serially and cross-sectionally uncorrelated, as $N, T \to \infty$ simultaneously,

(1) $\sqrt{N}\left(\widehat{\mathbf{F}}_t^{(n)} - \mathbf{F}_t^0 - \mathbf{b}_{F,t}^{\dagger(n)}\right) \xrightarrow{D} \mathcal{N}(0_r, \boldsymbol{\Sigma}_{F,t})$;

(2) $\sqrt{T}\left(\widehat{\boldsymbol{\lambda}}_i^{(n)} - \boldsymbol{\lambda}_i^0 - \mathbf{b}_{\lambda,i}^{\dagger(n)}\right) \xrightarrow{D} \mathcal{N}(0_r, \boldsymbol{\Sigma}_{\lambda,i}^*)$;

(3) $\sqrt{Th}\left(\widehat{\boldsymbol{\beta}}_i^{(n)}(\tau) - \boldsymbol{\beta}_i(\tau) - \mathbf{a}_i(\tau)h^2 - \mathbf{b}_{\beta,i}^{\dagger(n)}(\tau)\right) \xrightarrow{D} \mathcal{N}(0_p, \boldsymbol{\Sigma}_{\beta,i}^*(\tau))$;

where $\boldsymbol{\Sigma}_{\lambda,i}^* = \boldsymbol{\Omega}_{F,i}^{-1}\sigma_\varepsilon^2$ and $\boldsymbol{\Sigma}_{\beta,i}^*(\tau) = v_0\boldsymbol{\Sigma}_{X,i}^{-1}(\tau)\sigma_\varepsilon^2$.

Asymptotic Properties

Define

$$\kappa = \lim_{N,T\to\infty}(NT)^{-1}\sum_{s=1}^{T}\sum_{i=1}^{N}\mathbf{g}_i^\top(\tau_t)\mathbf{\Sigma}_{X,i}^{-1}(\tau_t)\left(\mathbf{\Sigma}_{v,F,i}\mathbf{\Omega}_{F,i}^{-1}\mathbf{\Sigma}_{v,F,i}^{-1}\mathbf{\Sigma}_{X,i}^{-1}(\tau_s)\mathbf{g}_i(\tau_s)+\mathbf{g}_i(\tau_t)\right)\in[0,1).$$

Theorem 2.3 (CLT, $n\to\infty$) Let Assumptions 1-3 hold. Suppose

$$\max\left\{\sqrt{N},\sqrt{T}\right\}\kappa^{n-2}\delta_{F,0}\to 0.$$

We have

(1) If, in addition, $N/T\to c_1<\infty$,

$$\sqrt{N}\left(\widehat{\mathbf{F}}_t^{(n)}-\mathbf{F}_t^0\right)\xrightarrow{D}\mathcal{N}(\sqrt{c_1}\mathbf{d}_{F,t},\mathbf{\Sigma}_{F,t}),$$

(2) If, in addition, $T/N\to c_2<\infty$,

$$\sqrt{T}\left(\widehat{\boldsymbol{\lambda}}_i^{(n)}-\boldsymbol{\lambda}_i^0\right)\xrightarrow{D}\mathcal{N}(\sqrt{c_2}\mathbf{d}_{\lambda,i},\mathbf{\Sigma}_{\lambda,i}),$$

(3) For any given $\tau\in(0,1)$,

$$\sqrt{Th}\left(\widehat{\boldsymbol{\beta}}_i^{(n)}(\tau)-\boldsymbol{\beta}_i(\tau)-\mathbf{a}_i(\tau)h^2\right)\xrightarrow{D}\mathcal{N}(0_p,\mathbf{\Sigma}_{\beta,i}(\tau)).$$

▸ See Assumptions

Asymptotic Properties

Consider the following mean-group estimator (MGE)

$$\widehat{\boldsymbol{\beta}}_w^{(n)}(\tau) = \sum_{i=1}^{N} w_{N,i} \widehat{\boldsymbol{\beta}}_i^{(n)}(\tau),$$

where $w_{N,i} \geq 0$ and $\sum_{i=1}^{N} w_{N,i} = 1$.

Theorem 2.4 (CLT, MGE) Let Assumptions 1-4 hold. Suppose

$$\sqrt{\gamma_{N,w} Th}\, \kappa^{n-2} \delta_{F,0} \to 0.$$

We have

$$\sqrt{\gamma_{N,w} Th} \left(\widehat{\boldsymbol{\beta}}_w^{(n)}(\tau) - \boldsymbol{\beta}_w(\tau) - \mathbf{a}_w(\tau) h^2 \right) \xrightarrow{D} \mathcal{N}(0_p, \boldsymbol{\Sigma}_{\beta,w}), \tag{15}$$

where

- $\gamma_{N,w} = \left(\sum_{i=1}^{N} w_{N,i}^2 \right)^{-1}$,
- $\mathbf{a}_w(\tau) = \frac{\mu_2}{2} \sum_{i=1}^{N} w_{N,i} \boldsymbol{\beta}_i''(\tau)(1 + o_P(1))$, $\boldsymbol{\beta}_w(\tau) = \sum_{i=1}^{N} w_{N,i} \boldsymbol{\beta}_i(\tau)$.

See Assumptions

Discussions on initial estimators

Exogenous factor models

Step 1. First, by the local linear method:

$$\widehat{\boldsymbol{\beta}}_i^{(0)}(\tau) = \left[\mathbf{I}_p, 0_p\right] \left(\mathbf{M}_i^\top(\tau)\mathbf{W}(\tau)\mathbf{M}_i(\tau)\right)^{-1} \mathbf{M}_i^\top(\tau)\mathbf{W}(\tau)\mathbf{y}_i. \tag{16}$$

Step 2. Second, by PCA:

$$\frac{1}{NT}\sum_{i=1}^{N}\mathbf{R}_{2,i}\mathbf{R}_{2,i}^\top\widehat{\mathbf{F}}^{(0)} = \widehat{\mathbf{F}}^{(0)}\mathbf{V}_{NT,1}, \tag{17}$$

where $\mathbf{R}_{2,i} = (R_{i1}(\widehat{\boldsymbol{\beta}}_i^{(0)}(\tau_1)), \cdots, R_{iT}(\widehat{\boldsymbol{\beta}}_i^{(0)}(\tau_T)))^\top$ with $R_{it}(\boldsymbol{\beta}) = y_{it} - \mathbf{x}_{it}^\top\boldsymbol{\beta}$, and $\mathbf{V}_{NT,1}$ is an $r \times r$ diagonal matrix with diagonal elements being the first r largest eigenvalues of the matrix $(NT)^{-1}\sum_{i=1}^{N}\mathbf{R}_{2,i}\mathbf{R}_{2,i}^\top$.

Discussions on initial estimators

Exogenous factor models

Corollary 3.1 (CLT, exogenous factor case) Let Assumptions 1.(i-v), 2-3, 5 hold.

Suppose

$$\max\left\{\sqrt{\frac{N}{Th}}, \sqrt{\frac{T}{N}}\right\}\kappa^{n-2} \to 0.$$

We have Theorem 2.3.(1-3) holds.

▸ See Assumptions

Discussions on initial estimators

Endogenous factor models

Assume that

$$\mathbf{x}_{it} = \mathbf{g}_i(\tau_t) + \mathbf{v}_{it}, \quad \mathbf{v}_{it} = \gamma_i^{0\top} \mathbf{F}_t^0 + \boldsymbol{\eta}_{it}. \tag{18}$$

Step 1. First, by the local linear method:

$$\widehat{g}_i^{(w)}(\tau) = [1, 0] \left(\mathbf{M}_T^\top(\tau) \mathbf{W}(\tau) \mathbf{M}_T(\tau) \right)^{-1} \mathbf{M}_T^\top(\tau) \mathbf{W}(\tau) \widetilde{\mathbf{x}}_i^{(w)} \tag{19}$$

where $\widehat{g}_i^{(w)}(\tau)$ is the w-th element of $\widehat{\mathbf{g}}_i(\tau)$, $\widetilde{\mathbf{x}}_i^{(w)} = \left(x_{i1}^{(w)}, \cdots, x_{iT}^{(w)} \right)^\top$ and $x_{it}^{(w)}$ is the w-th element of $\mathbf{x}_{it}$, for $w = 1, 2, \ldots, p$.

Step 2. Second, by PCA:

$$\left(\frac{1}{NTp} \sum_{w=1}^{p} \widetilde{\mathbf{R}}_g^{(w)} \widetilde{\mathbf{R}}_g^{(w)\top} \right) \widehat{\mathbf{F}}^{(0)} = \widehat{\mathbf{F}}^{(0)} \mathbf{V}_{NT,2} \tag{20}$$

where $\widetilde{\mathbf{R}}_g^{(w)} = \left(\widetilde{\mathbf{R}}_{g,1}^{(w)}, \ldots, \widetilde{\mathbf{R}}_{g,N}^{(w)} \right)$, $\widetilde{\mathbf{R}}_{g,i}^{(w)} = (R_{g,i1}^{(w)}, \ldots, R_{g,iT}^{(w)})^\top$ with $R_{g,it}^{(w)}$ being the w-th element of $\mathbf{R}_{g,it} = \mathbf{x}_{it} - \widehat{\mathbf{g}}_i(\tau_t)$, and $\mathbf{V}_{NT,2}$ is an $r \times r$ diagonal matrix with diagonal elements being the first r largest eigenvalues of the matrix $(NTp)^{-1} \sum_{w=1}^{p} \widetilde{\mathbf{R}}_g^{(w)} \widetilde{\mathbf{R}}_g^{(w)\top}$.

Discussions on initial estimators

Endogenous factor models

Corollary 3.2 (CLT, endogenous factor case) Let Assumptions 1.(i-v), 2-3, 6 hold.

Suppose

$$\max\left\{\sqrt{\frac{N}{Th}}, \sqrt{\frac{T}{N}}\right\}\kappa^{n-2} \to 0.$$

We have Theorem 2.3.(1-3) holds.

See Assumptions

Simulation studies

An example with exogenous factors

Example 1 Consider the following data generating process:

$$Y_{it} = X_{it,1}\beta_{1i}(\tau_t) + X_{it,2}\beta_{2i}(\tau_t) + \lambda_{i,1}F_{t,1} + \lambda_{i,2}F_{t,2} + \varepsilon_{it},$$

where

- $(\beta_{1i}(u), \beta_{2i}(u)) = (\sin(\pi u) + \cos(0.25\pi i), \cos(\pi u) + 0.5\sin(0.25\pi i))$;
- $X_{it,1} = g_{i1}(\tau_t) + \gamma_{i1,1}G_{t,1} + \gamma_{i2,1}G_{t,2} + \eta_{it,1}$;
- $X_{it,2} = g_{i2}(\tau_t) + \gamma_{i1,2}G_{t,1} + \gamma_{i2,2}G_{t,2} + \eta_{it,2}$;
- $(g_{i1}(u), g_{i2}(u)) = (3\cos(\pi(u + 0.25i)), 5\sin(\pi(u + 0.25i)))$;
- $F_{t,1} = \rho_{F_1}F_{t-1,1} + v_{F_1,t}$ with $\rho_{F_1} = 0.6$;
- $F_{t,2} = \rho_{F_2}F_{t-1,2} + v_{F_2,t}$ with $\rho_{F_2} = 0.4$;
- $(G_{t,1}, G_{t,2}) \sim i.i.d.N(0,1)$; the loadings and error terms: $\sigma_{ij,1} = 0.8^{|i-j|}$;

Simulation studies

An example with exogenous factors

For $\widehat{\boldsymbol{\beta}}_w^{(n)}(\tau)$

- $w_i = \frac{1}{N}$, for $i = 1, 2, \ldots, N$;
- h_{cv}: leave-one-out cross-validation method;
- Epanechnikov kernel is adopted.

For $\widehat{\mathbf{F}}_t^{(n)}$ and $\widehat{\boldsymbol{\lambda}}_i^{(n)}$,

- $r = 2$ as given.

- Replication times: $R = 1000$ times;
- For each replication,

$$\mathrm{MSE}(\widehat{\beta}_{l,w}^{(n)}) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{\beta}_{l,w}^{(n)}(\tau_t) - \beta_{l,w}(\tau_t)\right)^2,$$

for $l = 1, 2$, where $\beta_{l,w}(\tau_t) = N^{-1}\sum_{i=1}^{N}\beta_{l,i}(\tau_t)$ are true values.

- The second canonical correlation coefficients between $\{\widehat{\boldsymbol{\lambda}}_i^{(n)}\}$ and $\{\boldsymbol{\lambda}_i\}$, $\{\widehat{\mathbf{F}}_t^{(n)}\}$ and $\mathbf{F}_t$ are computed respectively for each replication.

Simulation studies

An example with exogenous factors

Table 1: Means and SDs of the mean squared errors for Example 4.1

MSE	$\widehat{\beta}_{w,1}^{(n)}$				$\widehat{\beta}_{w,2}^{(n)}$			
N/T	10	20	40	80	10	20	40	80
10	0.1771	0.0845	0.0454	0.0219	0.0531	0.0185	0.0077	0.0046
	(0.1755)	(0.0343)	(0.0203)	(0.0119)	(0.0775)	(0.0135)	(0.0034)	(0.0023)
20	0.1232	0.0650	0.0172	0.0123	0.0329	0.0133	0.0041	0.0026
	(0.0959)	(0.0174)	(0.0079)	(0.0051)	(0.0285)	(0.0075)	(0.0017)	(0.0010)
40	0.0954	0.0533	0.0154	0.0070	0.0225	0.0102	0.0036	0.0018
	(0.0209)	(0.0123)	(0.0053)	(0.0027)	(0.0147)	(0.0038)	(0.0009)	(0.0005)
80	0.0898	0.0455	0.0167	0.0046	0.0200	0.0083	0.0037	0.0015
	(0.0159)	(0.0084)	(0.0039)	(0.0017)	(0.0128)	(0.0020)	(0.0006)	(0.0004)

Simulation studies

An example with exogenous factors

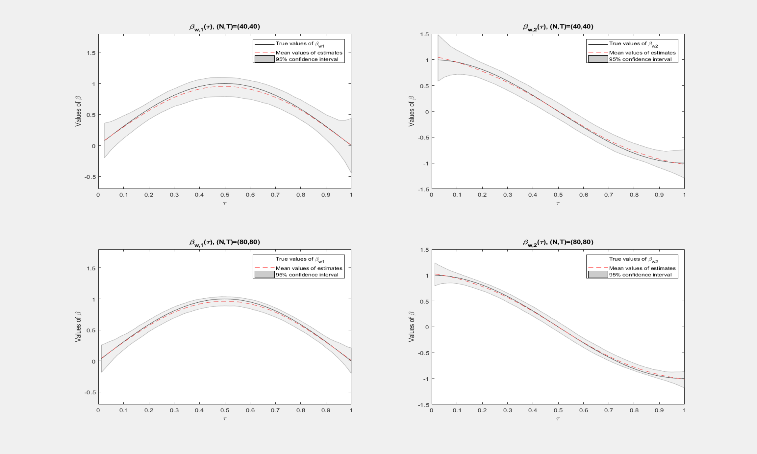

Figure 1: The simulated confidence intervals (Example 4.1)

Simulation studies

An example with exogenous factors

Table 2: Means and SDs of the second canonical coefficients for Example 4.1

SCC	$\widehat{\boldsymbol{\lambda}}_i^{(n)}$				$\widehat{\mathbf{F}}_t^{(n)}$			
N/T	10	20	40	80	10	20	40	80
10	0.3619	0.4877	0.5527	0.6042	0.4330	0.6693	0.8130	0.8736
	(0.2266)	(0.2346)	(0.2349)	(0.2342)	(0.2447)	(0.2696)	(0.2218)	(0.1961)
20	0.4461	0.6297	0.7433	0.8059	0.4455	0.7320	0.8914	0.9432
	(0.2570)	(0.2388)	(0.1931)	(0.1521)	(0.2470)	(0.2337)	(0.1687)	(0.1260)
40	0.5667	0.8081	0.8985	0.9213	0.5041	0.8374	0.9579	0.9818
	(0.2688)	(0.1668)	(0.0597)	(0.0440)	(0.2410)	(0.1641)	(0.0446)	(0.0308)
80	0.6934	0.9178	0.9514	0.9638	0.5573	0.9035	0.9718	0.9890
	(0.2491)	(0.0565)	(0.0213)	(0.0125)	(0.2315)	(0.0612)	(0.0152)	(0.0058)

Example 2 Consider the following data generating process:

$$\begin{aligned} X_{it,1} &= g_{i,1}(\tau_t) + \gamma_{i1,1}F_{t,1} + \gamma_{i2,1}F_{t,2} + \eta_{it,1} \\ X_{it,2} &= g_{i,2}(\tau_t) + \gamma_{i1,2}F_{t,1} + \gamma_{i2,2}F_{t,2} + \eta_{it,2} \end{aligned} \tag{21}$$

where $(g_{i1}(u), g_{i2}(u)) = (3\cos(\pi u), 5u)$. $(\gamma_{i1,1}, \gamma_{i1,2})$, $(F_{t,1}, F_{t,2})$ and $(\eta_{it,1}, \eta_{it,2})$ are following the same DGP in Example 1.

Simulation studies

An example with endogenous factors

Table 3: Means and SDs of the mean squared errors for Example 4.2

MSE	$\widehat{\beta}_{w,1}^{(n)}$				$\widehat{\beta}_{w,2}^{(n)}$			
N/T	10	20	40	80	10	20	40	80
10	0.2790	0.0883	0.0511	0.0181	0.0922	0.0213	0.0093	0.0051
	(0.5040)	(0.0414)	(0.0278)	(0.0152)	(0.1979)	(0.0238)	(0.0056)	(0.0038)
20	0.1514	0.0607	0.0192	0.0087	0.0599	0.0126	0.0047	0.0024
	(0.1648)	(0.0257)	(0.0103)	(0.0060)	(0.1353)	(0.0067)	(0.0021)	(0.0014)
40	0.1119	0.0537	0.0160	0.0045	0.0369	0.0107	0.0038	0.0015
	(0.0783)	(0.0148)	(0.0061)	(0.0030)	(0.1087)	(0.0040)	(0.0011)	(0.0006)
80	0.0906	0.0437	0.0128	0.0035	0.0250	0.0087	0.0032	0.0012
	(0.0304)	(0.0100)	(0.0038)	(0.0016)	(0.0135)	(0.0023)	(0.0007)	(0.0004)

Simulation studies

An example with endogenous factors

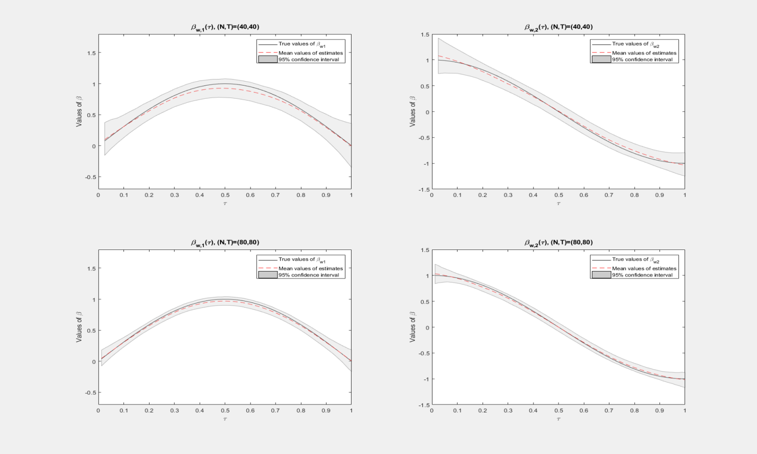

Figure 2: The simulated confidence intervals (Example 4.2)

Simulation studies

An example with endogenous factors

Table 4: Means and SDs of the second canonical coefficients for Example 4.2

SCC	$\widehat{\boldsymbol{\lambda}}_i^{(n)}$				$\widehat{\mathbf{F}}_t^{(n)}$			
N/T	10	20	40	80	10	20	40	80
10	0.4638	0.5178	0.5555	0.6054	0.3900	0.5814	0.7079	0.7652
	(0.2444)	(0.2326)	(0.2291)	(0.2335)	(0.2511)	(0.2673)	(0.2442)	(0.2446)
20	0.5328	0.6467	0.7218	0.7598	0.3888	0.6804	0.8091	0.8603
	(0.2512)	(0.2188)	(0.1895)	(0.1788)	(0.2284)	(0.2247)	(0.2003)	(0.1724)
40	0.6824	0.8007	0.8726	0.9032	0.4631	0.7906	0.9128	0.9510
	(0.2029)	(0.1391)	(0.0804)	(0.0658)	(0.2217)	(0.1357)	(0.0716)	(0.0527)
80	0.7202	0.8952	0.9426	0.9605	0.5079	0.8532	0.9475	0.9773
	(0.2119)	(0.0901)	(0.0404)	(0.0146)	(0.1958)	(0.0941)	(0.0410)	(0.0112)

An empirical application in health economics

Data description

The economic relationship between health care expenditure and income is reconsidered with the data set of OECD countries:

- The annual data is from 1971 to 2013 ($T = 43$) on 18 OECD countries ($N = 18$);
- Y_{it}: per capita health care expenditure (in US dollars, HE_{it});
- $X_{it,1}$: per capita GDP (in US dollars, GDP_{it});
- $X_{it,2}$: the proportion of population above 15 years over all population (DR_{it}^{young});
- $X_{it,3}$: the proportion of population above 65 years over all population (DR_{it}^{old});
- $X_{it,4}$: the proportion of government funding invested on health care industry in total health care expenditure (GHE_{it});
- all variables are expressed in natural logarithm.

An empirical application in health economics

Consider the following model:

$$HE_{it} = \beta_{1,it} GDP_{it} + \beta_{2,it} DR_{it}^{young} + \beta_{3,it} DR_{it}^{old} + \beta_{4,it} GHE_{it} + \sum_{m=1}^{r} \lambda_{mi} f_{mt} + \varepsilon_{it}, \qquad (22)$$

where

- $(\beta_{1,i}(\tau), \beta_{2,i}(\tau), \beta_{3,i}(\tau), \beta_{4,i}(\tau))$: unknown deterministic functions;
- $(f_{1t}, \ldots, f_{rt})$: common factors; $(\lambda_{1i}, \ldots, \lambda_{ri})$: loadings.

The criterion proposed by Bai and Ng (2002):

$$IC(r) = \log\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\varepsilon}_{it}^{2}\right) + r\left(\frac{N+T}{NT}\right)\log\left(\min\{N,T\}\right) \tag{23}$$

where $\widehat{\varepsilon}_{it}$ is the estimated residuals from model (22) with r factors.

Table 5: The values of $IC(r)$ in the determination of factor number

r	1	2	3	4	5	6	7	8
$IC(r)$	-6.6058	-6.5600	-6.5538	-6.4607	-6.4057	-6.3390	-6.2940	-6.2798

An empirical application in health economics

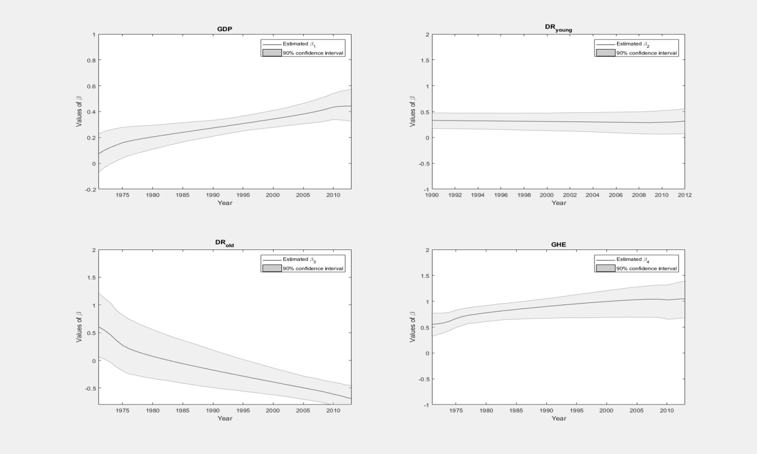

Figure 3: The estimated elasticities and confidence intervals

See bootstrap

An empirical application in health economics

Different groups:

- The European countries: Austria, Denmark, Finland, Germany, Iceland, Ireland, Netherlands, Norway, Portugal, Spain, Sweden and the UK;
- Non-European countries: Australia, Canada, Japan, Korea, New Zealand and the US.

An empirical application in health economics

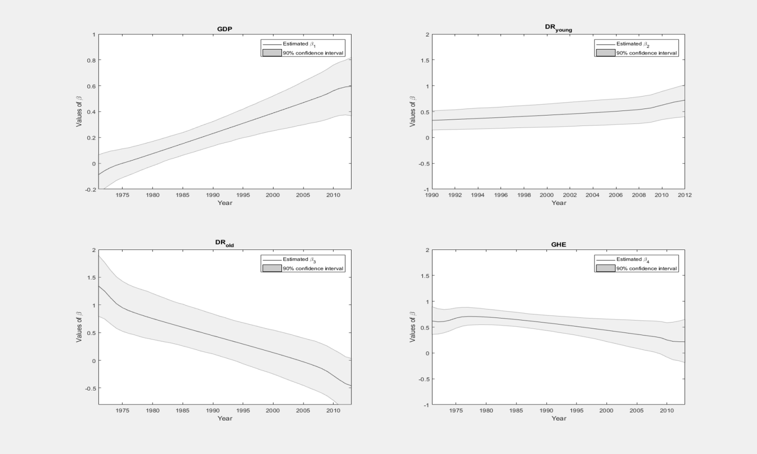

Figure 4: The estimated elasticities and confidence intervals (European OECD countries)

See bootstrap

An empirical application in health economics

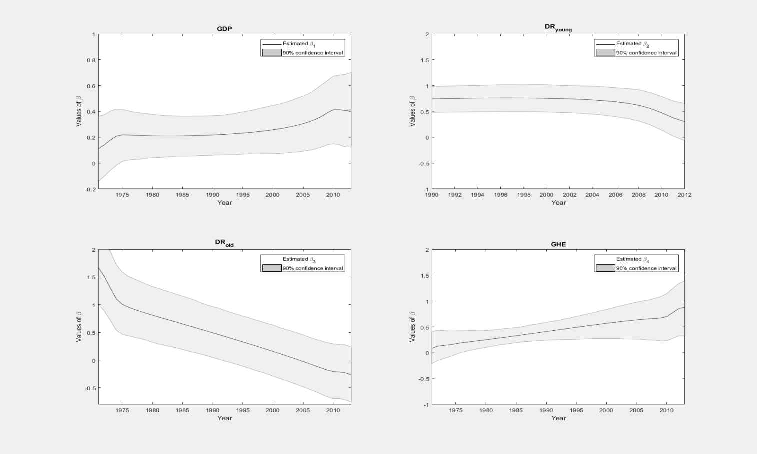

Figure 5: The estimated elasticities and confidence intervals (Non-European OECD countries)

See bootstrap

An empirical application in health economics

Estimated loadings and factors

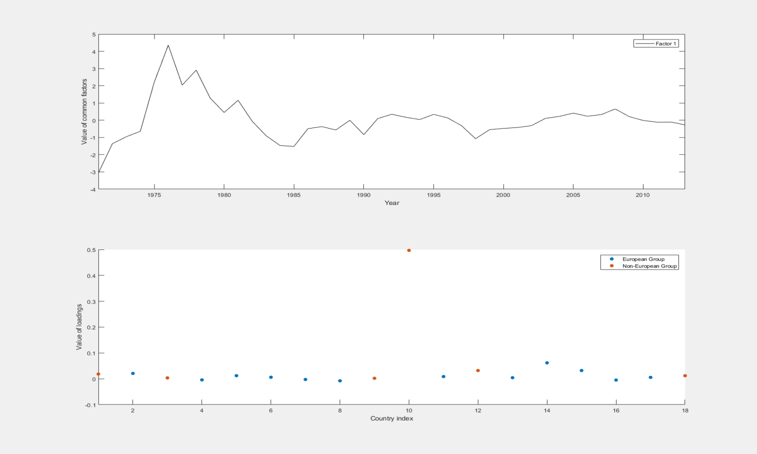

Figure 6: The estimated loadings and factors

Conclusions

Our contributions can be summarized as follows:

- Model:
  - Time-varying regression coefficients are introduced;
  - Heterogeneity is allowed.
- Method:
  - A recursive method is proposed to reduce the bias;
  - It can be generally used when the factors are exogenous or endogenous.
  - Asymptotic properties are established for the proposed estimators, including the factors and loadings.
- Empirical results: evidence of time-variation and heterogeneity in income elasticity of health care expenditure.

Thank You

Appendix

Notation

Define

- $\mathbf{W}(\tau) = \operatorname{diag}\left(K(\frac{1-\tau T}{Th}), \ldots, K(\frac{T-\tau T}{Th})\right)$
-
$$\overline{\mathbf{M}}(\tau) = \begin{pmatrix} \overline{\mathbf{x}}_1^\top & \frac{1-\tau T}{Th}\overline{\mathbf{x}}_1^\top \\ \vdots & \vdots \\ \overline{\mathbf{x}}_T^\top & \frac{T-\tau T}{Th}\overline{\mathbf{x}}_T^\top \end{pmatrix}. \tag{24}$$
- $\widetilde{\mathbf{W}}(\tau) = \mathbf{W}(\tau) \otimes \mathbf{I}_N$,
- $\overline{\mathbf{y}} = (\overline{\mathbf{y}}_1^\top, \cdots, \overline{\mathbf{y}}_T^\top)^\top$.

Return

Appendix

Notation

Define

$$
\begin{aligned}
\mathbf{y}_t &= (y_{1t}, y_{2t}, \ldots, y_{Nt})^\top, \quad \mathbf{x}_t = (\mathbf{x}_{1t}, \mathbf{x}_{2t}, \ldots, \mathbf{x}_{Nt}) \\
\mathbf{V} &= (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N)^\top, \quad \widetilde{\mathbf{F}}_t = \left(\widetilde{F}_{1t}, \widetilde{F}_{jt}, \ldots, \widetilde{F}_{rt}\right)^\top, \\
\widetilde{\mathbf{F}} &= \left(\widetilde{\mathbf{F}}_1, \widetilde{\mathbf{F}}_2, \ldots, \widetilde{\mathbf{F}}_T\right)^\top, \quad \boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2,t}, \ldots, \varepsilon_{Nt})^\top.
\end{aligned}
$$

Return

Appendix

Notation

Let $\mathbf{W}_0(\tau) = \text{diag}\left(K(\frac{1-\tau T}{Th}), \ldots, K(\frac{T-\tau T}{Th})\right)$, $\mathbf{W}(\tau) = \mathbf{W}_0(\tau) \otimes \mathbf{I}_N$, $\widetilde{\mathbf{y}}_t = \mathbf{M}_V \mathbf{y}_t$, $\widetilde{\mathbf{x}}_t = \mathbf{x}_t \mathbf{M}_V$ and

$$\mathbf{M}(\tau) = \begin{pmatrix} \widetilde{\mathbf{x}}_1^\top & \frac{1-\tau T}{Th} \widetilde{\mathbf{x}}_1^\top \\ \vdots & \vdots \\ \widetilde{\mathbf{x}}_T^\top & \frac{T-\tau T}{Th} \widetilde{\mathbf{x}}_T^\top \end{pmatrix}.$$

Return

Appendix

Notation

Define

$$\mathbf{y}_i = (y_{i1}, \cdots, y_{iT})^\top, \quad \mathbf{W}(\tau) = \left(K\left(\frac{1-\tau T}{Th}\right), \cdots, K\left(\frac{T-\tau T}{Th}\right) \right)$$

and

$$\mathbf{M}_i = \begin{pmatrix} \mathbf{x}_{i1}^\top & \frac{1-\tau T}{Th}\mathbf{x}_{i1}^\top \\ \vdots & \\ \mathbf{x}_{iT}^\top & \frac{T-\tau T}{Th}\mathbf{x}_{iT}^\top \end{pmatrix}.$$

Return

Notations:

- $\boldsymbol{\Omega}_3(t,s) = \boldsymbol{\Sigma}_\lambda^{-1}(h^{-1}K_{s,0}(\tau_t)\boldsymbol{\Omega}_1(t,s) + \boldsymbol{\Omega}_2(t,s)),$
- $\boldsymbol{\lambda}_i^\dagger(\tau_t) = \boldsymbol{\Sigma}_{X,i}^{-1}(\tau_t)\left(\boldsymbol{\Sigma}_{X,\lambda,i}(\tau_t) + E\left[\mathbf{x}_{it}\right]\boldsymbol{\lambda}_i^\top\right),$
- $\boldsymbol{\Delta}_{F,i} = \boldsymbol{\Sigma}_{v,F}\boldsymbol{\Omega}_{F,i}^{-1}\boldsymbol{\Sigma}_{v,F}^\top$, $\boldsymbol{\Sigma}_{X,\lambda,i}(\tau_t) = E\left[\mathbf{x}_{it}\boldsymbol{\lambda}_i^\top\right]$

Return

Appendix

Assumptions

Assumption 1.

(i) α–mixing conditions on panel data are assumed as follows: $\{\mathbf{v}_t, \boldsymbol{\varepsilon}_t, \mathbf{F}_t^0\}$ are strictly stationary and α–mixing across t; Let $\alpha_{ij}(|t-s|)$ represent the α-mixing coefficient between $\{\varepsilon_{it}\}$ and $\{\varepsilon_{js}\}$. Assume that

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left(\alpha_{ij}(t)\right)^{\delta/(4+\delta)} = O(N) \quad \text{and} \quad \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\alpha_{ij}(0)\right)^{\delta/(4+\delta)} = O(N),$$

where $\delta > 0$ is chosen such that $E\left[\|\boldsymbol{\omega}_{it}\|^{4+\delta}\right] < \infty$ with $\boldsymbol{\omega}_{it} \in \{\boldsymbol{\lambda}_i^0, \mathbf{F}_t^0, \varepsilon_{it}, \mathbf{v}_{it}\}$. Let $\alpha(|t-s|)$ represent the α-mixing coefficient between $\{\mathbf{v}_{it}, \mathbf{F}_t^0\}$ and $\{\mathbf{v}_{is}, \mathbf{F}_s^0\}$. Assume that

$$\alpha(t) = O(t^{-\theta}),$$

where $\theta > (4+\delta)/\delta$.

(ii) $\{\varepsilon_{it}\}$ are identically distributed across i with zero mean and independent of $\{\mathbf{F}_s^0, \boldsymbol{\lambda}_j^0, \mathbf{v}_{js}\}$, for any i, j, t, s.

(iii) The unknown deterministic functions $\{\boldsymbol{\beta}_i(\tau)\}$ have continuous derivatives of up to the second order on its support $\tau \in [0,1]$, and the functions $\{\mathbf{g}_i(\tau)\}$ are uniformly bounded: $\max_{1\le i\le N}\sup_{\tau\in[0,1]}||\mathbf{g}_i(\tau)|| < \infty$.

(iv) The kernel function $K(\cdot)$ is Lipschitz continuous with compact support on $[-1,1]$.

(v) As $N, T \to \infty$, the bandwidth satisfies that $h \to 0$, $\max\{N,T\}h^4 \to 0$ and $\min\{N,T\}h^2 \to \infty$.

(vi) Let $\mathbf{R}_F^{(n)} = \widehat{\mathbf{F}}^{(n)} - \mathbf{F}^0$. For the initial estimator $\widehat{\mathbf{F}}^{(0)}$, suppose that

$$T^{-1/2}\|\mathbf{R}_F^{(0)}\| = O_P\left(\delta_{F,0}\right) \quad \text{and} \quad (Th)^{-1/2}\|\mathbf{W}(\tau)^\top \mathbf{R}_F^{(0)}\| = O_P\left(\delta_{F,0}\right),$$

where $\delta_{F,0}$ satisfies that $NTh^4\delta_{F,0}^2 \to 0$, $\delta_{F,0}^2/h \to 0$ and $\max\{N,T\}\delta_{F,0}^4/h \to 0$, as $N, T \to \infty$.

◂ Return

Appendix

Assumptions

Notation:

$$\sigma^2_{v,\varepsilon,i} = \sigma^2_\varepsilon \mathbf{\Sigma}_{v,i} + 2\sum_{t=2}^{\infty} E\left[\varepsilon_{11}\varepsilon_{1t}\right] E\left[\mathbf{v}_{i1}\mathbf{v}_{it}^\top\right], \quad \sigma^2_{\varepsilon,0} = \sigma^2_\varepsilon + 2\sum_{t=2}^{\infty} E\left[\varepsilon_{11}\varepsilon_{1,t}\right], \quad \sigma^2_\varepsilon = E\left[\varepsilon^2_{11}\right],$$

$$v_0 = \int K(u)^2 du, \quad \mathbf{\Sigma}^0_{\beta,i}(\tau) = v_0\left(\boldsymbol{\sigma}^2_{v,\varepsilon,i} + \sigma^2_{\varepsilon,0}\mathbf{g}_i(\tau)\mathbf{g}_i^\top(\tau)\right),$$

$$\xi_{1,it} = \boldsymbol{\lambda}_i^{0\top}\mathbf{F}_t^0, \quad \boldsymbol{\xi}_{2,it} = \mathbf{v}_{it}\boldsymbol{\lambda}_i^{0\top}, \quad \boldsymbol{\sigma}^2_{F,\varepsilon,0} = \sigma^2_\varepsilon \mathbf{\Sigma}_F + 2\sum_{t=2}^{\infty} E\left[\varepsilon_{i1}\varepsilon_{it}\right] E\left[\mathbf{F}_1^0\mathbf{F}_t^{0\top}\right],$$

$$\mathbf{\Sigma}^0_{\lambda,i} = \boldsymbol{\sigma}^2_{F,\varepsilon,0} - \int_0^1 \mathbf{\Sigma}^\top_{v,F,i}\mathbf{\Sigma}^{-1}_{X,i}(v)\left(\boldsymbol{\sigma}^2_{v,\varepsilon,i} + \sigma^2_{\varepsilon,0}\mathbf{g}_i(v)\mathbf{g}_i^\top(v)\right)\mathbf{\Sigma}^{-1}_{X,i}(v)\mathbf{\Sigma}_{v,F,i}dv$$

Assumption 2.

(i) Assume the following moment conditions on $\{\varepsilon_{it}, \xi_{1,it}, \boldsymbol{\xi}_{2,it}\}$:

$$\begin{aligned}
\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T} |Cov(\varepsilon_{it_1}\varepsilon_{it_2}, \varepsilon_{jt_3}\varepsilon_{jt_4})| &\leq CNT^2 \\
\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T} |Cov(\xi_{1,it_1}\xi_{1,it_2}, \xi_{1,jt_3}\xi_{1,jt_4})| &\leq CNT^2 \\
\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T} \|Cov(\boldsymbol{\xi}_{2,it_1}\boldsymbol{\xi}^\top_{2,it_2}, \boldsymbol{\xi}_{2,jt_3}\boldsymbol{\xi}^\top_{2,jt_4})\| &\leq CNT^2
\end{aligned}$$

Assumption 2.

(ii) Assume that $\mathbf{\Sigma}_{v,i}$, $\mathbf{\Sigma}_F$, $\mathbf{\Sigma}^0_{\beta,i}(\tau)$ and $\mathbf{\Sigma}^0_{\lambda,i}$ are positive definite and σ^2_ε is a positive scalar.

(iii) Suppose that $\left\|N^{-1}\sum_{i=1}^N \boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\top} - \mathbf{\Sigma}_\lambda\right\| = O_P\left(N^{-1/2}\right)$ and

$$N^{-1/2}\sum_{i=1}^N \boldsymbol{\lambda}_i^0\varepsilon_{it} \xrightarrow{D} \mathcal{N}(0, \mathbf{\Sigma}^0_{F,t}),$$

for any fixed t, where both $\mathbf{\Sigma}_\lambda$, $\mathbf{\Sigma}^0_{F,t}$ are positive definite.

(iv) Let h satisfy $\limsup_{N,T\to\infty} NTh^5 < \infty$, $NT^{-(4+\delta^*)/4} \to 0$, $N^{\delta^\dagger}T^{-\theta}h^{-3-\theta}(\log T)^{1+2\theta} \to 0$, for $0 < \delta^* < \delta$ and $\delta^\dagger = (6+\delta)/(4+\delta) - 2(1+\theta)/(2+\delta)$, where θ and δ are defined in Assumption 1.

Return

Assumption 3.

Let $E\left[\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\top}|\mathbf{v}_{i1},\ldots,\mathbf{v}_{iT},\mathbf{F}_1^{0\top},\ldots,\mathbf{F}_T^{0\top}\right]=\boldsymbol{\Sigma}_\lambda$ almost surely, where $\boldsymbol{\Sigma}_\lambda=\lim_{N\to\infty}N^{-1}\sum_{i=1}^{N}\boldsymbol{\lambda}_i^0\boldsymbol{\lambda}_i^{0\top}$ is positive definite.

Return

Appendix

Assumptions for the heterogeneous model

Assumption 4.

(i) Assume that $E\left[\mathbf{v}_{it}\boldsymbol{\lambda}_i^{0\top}\right] = E\left[\mathbf{v}_{it}\mathbf{F}_t^{0\top}\right] = 0_{p\times r}$ and $E[\lambda_i] = 0_r$.

(ii) Define that

$$\begin{aligned}
\widetilde{\sigma}^2_{v,\varepsilon}(i,j,\tau) &= \boldsymbol{\Sigma}^{-1}_{X,i}(\tau)\boldsymbol{\sigma}^2_{v,\varepsilon}(i,j)\boldsymbol{\Sigma}^{-1}_{X,j}(\tau),\\
\widetilde{\sigma}^2_{\varepsilon}(i,j,\tau) &= \sigma^2_{\varepsilon}(i,j)\boldsymbol{\Sigma}^{-1}_{X,i}(\tau)\mathbf{g}_i(\tau)\mathbf{g}_j^{\top}(\tau)\boldsymbol{\Sigma}^{-1}_{X,j}(\tau),\\
\boldsymbol{\Sigma}_{\beta,w}(\tau) &= \lim_{N\to\infty}\gamma_{N,w}v_0\sum_{i=1}^{N}\sum_{j=1}^{N}w_{N,i}w_{N,j}\left(\widetilde{\sigma}^2_{\varepsilon}(i,j,\tau)+\widetilde{\sigma}^2_{v,\varepsilon}(i,j,\tau)\right).
\end{aligned}$$

We assume $\boldsymbol{\Omega}_{F,i}$ and $\boldsymbol{\Sigma}_{\beta,w}(\tau)$ are positive-definite matrices, where $\boldsymbol{\Omega}_{F,i}$ is defined in Theorem 1.

(iii) The bandwidth h satisfies that: $\lim_{N\to\infty}\gamma_{N,w}h^3 = 0$.

◂ Return

Appendix

Assumptions

Assumption 5.

(i) Assume the estimators $\widehat{\mathbf{F}}^{(0)}$ and $\widehat{\boldsymbol{\Lambda}}^{(0)}$ satisfy the following identification condition:

$$N^{-1}\widehat{\boldsymbol{\Lambda}}^{(0)\top}\widehat{\boldsymbol{\Lambda}}^{(0)} = \text{diagnal} \quad \text{and} \quad T^{-1}\widehat{\mathbf{F}}^{(0)\top}\widehat{\mathbf{F}}^{(0)} = \mathbf{I}_r.$$

(ii) Assume the true values $\mathbf{F}^0$ and $\boldsymbol{\Lambda}^0$ satisfy the identification conditions in Assumption 5.1.

(iii) Suppose $\mathbf{F}_t^0$ is conditionally uncorrelated with $\boldsymbol{\Lambda}^0$, $\mathbf{v}_1$, …, $\mathbf{v}_T$:

$$E\left[\mathbf{F}_t^0 | \boldsymbol{\Lambda}^0, \mathbf{v}_1, \ldots, \mathbf{v}_T\right] = 0_r.$$

In addition, we assume $\left\{\mathbf{F}_t^0 | \boldsymbol{\Lambda}^0, \mathbf{v}_1, \ldots, \mathbf{v}_T\right\}$ satisfies the α-mixing condition in Assumption 1.

(iv) Suppose the following moment conditions can hold:

$$\begin{aligned}
\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\left\|E\left[\mathbf{F}_{t_1}^0\mathbf{F}_{t_2}^{0\top}\mathbf{F}_{t_3}^0\mathbf{F}_{t_4}^{0\top}\right]\right\| &\leq CT^2, \\
\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3\neq t_1}^{T}\sum_{t_4\neq t_2}^{T}\left|E\left[\varepsilon_{it_1}\varepsilon_{jt_2}\varepsilon_{it_3}\varepsilon_{jt_4}\right]\right| &\leq CNT^2.
\end{aligned}$$

Return

Appendix

Assumptions

Assumption 6.

(i) Assume the estimators $\widehat{\mathbf{F}}^{(0)}$ and $\widehat{\gamma}_i^{(0)}$ satisfy the following identification condition:

$$N^{-1}\sum_{i=1}^{N}\widehat{\gamma}_i^{(w,0)\top}\widehat{\gamma}_i^{(w,0)} = \text{diagnal} \quad \text{and} \quad T^{-1}\widehat{\mathbf{F}}^{(0)\top}\widehat{\mathbf{F}}^{(0)} = \mathbf{I}_r,$$

for $w = 1, 2, \dots, p$, where $\widehat{\gamma}_i^{(w,0)}$ is the w-th column of $\widehat{\gamma}_i^{(0)\top}$.

(ii) Assume the true values $\mathbf{F}^0$ and $\boldsymbol{\lambda}^0$ satisfy the identification conditions in Assumption 5.1.

(iii) The unknown deterministic function $\mathbf{g}_i(\tau)$ has continuous derivatives of up to the second order on its support $\tau \in [0,1]$. Assume that the loadings $\{\gamma_i\}$ are deterministic and uniformly bounded.

(iv) Suppose we have the following moment conditions:

$$\begin{aligned}
\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}\left\|E\left[\mathbf{F}_{t_1}^0\mathbf{F}_{t_2}^{0\top}\mathbf{F}_{t_3}^0\mathbf{F}_{t_4}^{0\top}\right]\right\| &\leq CT^2, \\
\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3\neq t_1}^{T}\sum_{t_4\neq t_2}^{T}\left\|E\left[\boldsymbol{\eta}_{it_1}\boldsymbol{\eta}_{jt_2}^{\top}\boldsymbol{\eta}_{it_3}\boldsymbol{\eta}_{jt_4}^{\top}\right]\right\| &\leq CNT^2.
\end{aligned}$$

Return

Appendix

Estimated loadings and factors

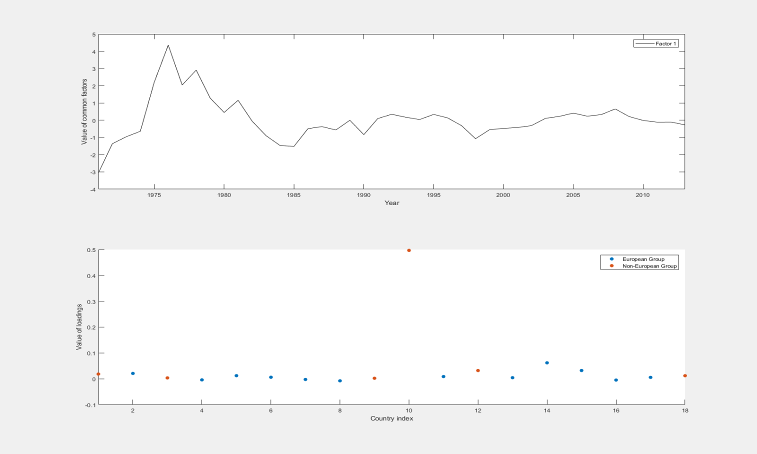

Figure 7: The estimated loadings and factors

Appendix

The details for our bootstrapping method are as follows:

Step 1. Calculate the residuals $\{\bar{\varepsilon}_{it}\}$ for the estimation method discussed in Section 2.

Step 2. Resample the residuals and obtain $\{\bar{\varepsilon}^*_{it}\}$, where $\varepsilon^*_{it} = \bar{\varepsilon}_k$ and k is randomly selected from $\{1, \ldots, T\}$. Then the bootstrapping sample $\{Y^*_{it}\}$ can be generated with $\{\bar{\varepsilon}^*_{it}\}$.

Step 3. The bootstrapping estimator $\overline{\boldsymbol{\beta}}^*_t$ can be obtained using the data set $\{Y^*_{it}\}$.

Step 4. Repeat Steps 2 and 3 1000 times to obtain the 90% confidence intervals.

Return to simulations Return to empirical

Appendix

Discussions on initial estimator: exogenous factors

PCA method to find $\widehat{\mathbf{F}}^{(0)}$:

(1) First, ignore the common factor part and estimate $\boldsymbol{\beta}_{it}$ using local linear method:

$$\widehat{\boldsymbol{\beta}}_i^{(0)}(\tau) = \left[\mathbf{I}_p, \mathbf{0}_p\right] \left(\mathbf{M}_i^\top(\tau)\mathbf{W}(\tau)\mathbf{M}_i(\tau)\right)^{-1} \mathbf{M}_i^\top(\tau)\mathbf{W}(\tau)\mathbf{y}_i,$$

for $i = 1, \ldots, N$.

(2) Then estimate $\mathbf{F}$ using the PCA method as follows:

$$\frac{1}{NT}\sum_{i=1}^{N} \mathbf{R}_{3,i}\mathbf{R}_{3,i}^\top \widehat{\mathbf{F}}^{(0)} = \widehat{\mathbf{F}}^{(0)}\mathbf{V}_{NT,F}, \tag{25}$$

where $\mathbf{R}_{3,i} = \left(R_{i1}(\widehat{\boldsymbol{\beta}}_1^{(0)}(\tau_1)), \ldots, R_{iT}(\widehat{\boldsymbol{\beta}}_i^{(0)}(\tau_T))\right)^\top$ and $R_{it}(\boldsymbol{\beta}) = y_{it} - \mathbf{x}_{it}^\top \boldsymbol{\beta}(\tau_t)$.

Appendix

Discussions on initial estimator: exogenous factors

Corollary 3.2 Under some regularity conditions and $\widehat{\mathbf{F}}^{(0)}$ satisfies (25),

$$\frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{F}}^{(0)} - \mathbf{F}\mathbf{H}_1 \right\| = O_p\left(\max\{(Th)^{-1/2}, N^{-1/2}, h^2\} \right), \tag{26}$$

where $\mathbf{H}_1 = (NT)^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^\top \mathbf{F}^\top \widehat{\mathbf{F}}^{(0)} \mathbf{V}_{NT,1}^{-1}$.

▸ See Assumptions

Appendix

Discussions on initial estimator: endogenous factors

Consider the following model:

$$y_{it} = \mathbf{x}_{it}^{\top}\boldsymbol{\beta}_{it} + \boldsymbol{\lambda}_i^{\top}\mathbf{F}_t + \varepsilon_{it}$$

$$\mathbf{x}_{it} = \mathbf{g}_i(\tau_t) + \gamma_i^{\top}\mathbf{F}_t + \boldsymbol{\eta}_{it}$$

PCA method to estimate $\widehat{\mathbf{F}}^{(0)}$,

(1) We first estimate the $\mathbf{g}_i(\tau)$ using local linear method:

$$\widehat{g}_i^{(w)}(\tau) = [1,0]\left(\mathbf{M}_T^{\top}(\tau)\mathbf{W}(\tau)\mathbf{M}_T(\tau)\right)^{-1}\mathbf{M}_T^{\top}(\tau)\mathbf{W}(\tau)\widetilde{\mathbf{x}}_i^{(w)} \tag{27}$$

where $\widehat{g}_i^{(w)}(\tau)$ is the w-th element of $\widehat{\mathbf{g}}_i(\tau)$, $\widetilde{\mathbf{x}}_i^{(w)} = \left(x_{i1}^{(w)}, \cdots, x_{iT}^{(w)}\right)^{\top}$ and $x_{it}^{(w)}$ is the w-th element of $\mathbf{x}_{it}$.

(2) Then $\mathbf{F}_t$ can be estimated by the PCA method:

$$\left(\frac{1}{NTp}\sum_{w=1}^{p}\widetilde{\mathbf{R}}_g^{(w)}\widetilde{\mathbf{R}}_g^{(w)\top}\right)\widehat{\mathbf{F}}^{(0)} = \widehat{\mathbf{F}}^{(0)}\mathbf{V}_{NT,g} \tag{28}$$

where $\widetilde{\mathbf{R}}_g^{(w)} = \left(\widetilde{\mathbf{R}}_{g,1}^{(w)}, \ldots, \widetilde{\mathbf{R}}_{g,N}^{(w)}\right)$, $\widetilde{\mathbf{R}}_{g,i}^{(w)} = (R_{g,i1}^{(w)}, \ldots, R_{g,iT}^{(w)})^{\top}$ and $R_{g,it}^{(w)}$ is the w-th element of $\mathbf{R}_{g,it} = \mathbf{x}_{it} - \widehat{\mathbf{g}}_i(\tau_t)$.

Appendix

Discussions on initial estimator: endogenous factors

Corollary 3.3 Under some regularity conditions and $\widehat{\mathbf{F}}^{(0)}$ satisfies (28),

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{F}}^{(0)}-\mathbf{F}\mathbf{H}_1\right\| = O_p\left(\max\{(Th)^{-1/2}, N^{-1/2}, h^2\}\right), \tag{29}$$

where $\mathbf{H}_2 = \frac{1}{NTp}\sum_{w=1}^{p}\sum_{i=1}^{N}\gamma_i^{(w)}\gamma_i^{(w)\top}\mathbf{F}^\top\widehat{\mathbf{F}}^{(0)}\mathbf{V}_{NT,2}^{-1}$.

▸ See Assumptions ▸ Return to Estimation

Reference

Ando, T. and Bai, J. (2014). Asset pricing with a general multifactor structure. *Journal of Financial Econometrics*, 13(3):556–604.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Chen, J., Gao, J., and Li, D. (2012). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2):393–420.

Gao, J., Xia, K., and Zhu, H. (2019). Heterogeneous panel data models with cross–sectional dependence. *Forthcoming in Journal of Econometrics*. Available at https://ideas.repec.org/p/msh/ebswps/2017-16.html.

Jiang, B., Yang, Y., Gao, J., and Hsiao, C. (2017). Recursive estimation in large panel data models: Theory and practice. Working paper at http://business.monash.edu/econometrics-and-business-statistics/research/publications/ebs/wp05-17.pdf.

Li, D., Chen, J., and Gao, J. (2011). Non-parametric time-varying coefficient panel data models with fixed effects. *The Econometrics Journal*, 14(3):387–408.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Silvapulle, P., Smyth, R., Zhang, X., and Fenech, J.-P. (2017). Nonparametric panel data model for crude oil and stock market prices in net oil importing countries. *Energy Economics*, 67:255–267.

Su, L., Shi, Z., and Phillips, P. C. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6):2215–2264.