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
Panel Data Analysis

1. **Data Structure:** Dependent Variable $y_{it}$ and Independent Variable $x_{it} = (X_{1,it}, X_{2,it}, \ldots, X_{p,it})$ with $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.

2. **Aim:** Accurately model and estimate the relation between $y_{it}$ and $x_{it}$ 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).
Common factor models are widely used to capture cross-sectional dependence in panel data sets:

\[ y_{it} = x_{it}^\top \beta + e_{it}, \quad e_{it} = \lambda_i^\top F_t + \varepsilon_{it} \]  

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

- \( \beta \) is a \( p \)-dimensional unknown parameter;
- \( \{F_t\} \) are unknown \( r \)-dimensional common factors;
- \( \{\lambda_i\} \) are corresponding factor loadings.

Advantages of factor models:

- heterogenous effects of common shocks;
- Appropriate flexibility.
Bai (2009) proposes an iterative numerical method to approximate the minimizer of the least squares objective function:

\[
SSR = \sum_{i=1}^{N} \sum_{t=1}^{T} \left( y_{it} - x_{it}^\top \beta - \lambda_i^\top F_t \right)^2 \tag{2}
\]

▶ Estimate \( \beta \) by least squares method;
▶ Estimate \( \lambda_i \) and \( F_t \) by PCA method;
▶ Repeat until convergence.

▶ Extensions:
▶ Ando and Bai (2014).

▶ Challenges:
▶ Poor performance with endogenous factors (see Jiang et al., 2017).
Pesaran (2006) proposes valid proxies for $F_t$ in the following model:

$$
\begin{pmatrix}
    y_{it} \\
    x_{it}
\end{pmatrix}
= 
\begin{pmatrix}
    \lambda_i^\top + \beta_i^\top \gamma_i^\top \\
    \gamma_i^\top
\end{pmatrix}
F_t + 
\begin{pmatrix}
    \varepsilon_{it} + \beta_i^\top \eta_{it} \\
    \eta_{it}
\end{pmatrix},
$$

(3)

where $\{\gamma_i\}$ are unknown factor loadings.


Challenges:

- Rank condition $r \leq p + 1$,
- No estimators for $F_t, \lambda_i$. 

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):
    \[ y_{it} = x_{it}^\top \beta_t + f_t + \alpha_i + \epsilon_{it}; \]
    \hspace{1cm} (4)

  where \( \beta_t = \beta(\tau_t) \) 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:
    \[ \beta_i = \beta + u_i. \] (5)
  - Su et al. (2016)'s unknown group pattern:
    \[ \beta_i = \sum_{k=1}^{K} \beta^{(k)} \mathbb{1}\{i \in G_k\}, \] (6)
    where \( K \) is known and fixed but \( G_k \) is unknown.
  - Gao et al. (2019)'s complete heterogeneity:
    \[ y_{it} = x_{it}^\top \beta_i + f_{it} + \alpha_i + \varepsilon_{it}, \] (7)
    where \( f_{it} = f_i(\tau_t) \).
We consider the following model:

\[ y_{it} = x_{it}^\top \beta_{it} + \lambda_i^\top F_t + \varepsilon_{it}, \]  

(8)

where

- \( x_{it} \) and \( y_{it} \) are observable;
- \( \beta_{it} = \beta_i(\tau_t) \) is an unknown deterministic function;
- \( x_{it} \) can be correlated with \( \{\lambda_i, F_t\} \).
Outline of Contribution


2. Unified Estimation Approach: observed, unobserved or partially observed factors.

3. Asymptotic Theory: reconcile computational elements (iteration steps) with statistical properties.

Proposed Estimation Approach

Recall the heterogeneous model:

\[ y_{it} = x_{it}^\top \beta_i(\tau_t) + \lambda_i^\top F_t + \varepsilon_{it}. \]

The idea of iteration:

- With given \( F_t \), we can estimate \( \beta_i(\tau) \) and \( \lambda_i \) by a profile method.
- With \( \beta_i(\tau) \) and \( \lambda_i, F_t \) can be estimated by OLS method.
Estimation Procedure

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

(2) With \( \hat{F}_t^{(n)} \) and by regarding \( \lambda_i \) as known, \( \beta_i(\tau) \) can be estimated by local linear method. For \( \tau \in (0, 1) \)

\[
\min_{a_i(\tau), b_i(\tau)} \sum_{t=1}^T \left( y_{it} - \lambda_i^\top \hat{F}_t^{(n)} - x_{it}^\top \left( a_i(\tau) + \left( \frac{t - \tau T}{Th} \right) b_i(\tau) \right) \right)^2 K \left( \frac{t - \tau T}{Th} \right),
\]

we have

\[
\hat{\beta}_i^{(n+1)}(\tau, \lambda_i) = [I_p, 0_p] \left[ M_i(\tau)^\top W(\tau) M_i(\tau) \right]^{-1} M_i(\tau)^\top W(\tau) \left[ y_i - \hat{F}^{(n)} \lambda_i \right].
\]

(3) With \( \hat{\beta}_i(\tau, \lambda_i) \), we can estimate \( \lambda_i \) by the least squares method:

\[
\min_{\lambda_i} \sum_{t=1}^T \left( y_{it} - x_{it}^\top \hat{\beta}_i^{(n+1)}(\tau, \lambda_i) - \lambda_i^\top \hat{F}_t^{(n)} \right)^2.
\]
Estimation Procedure

We have

\[
\hat{\lambda}_i^{(n+1)} = \left[ \hat{F}^{(n)\top} (I - S_i)\top (I - S_i)\hat{F}^{(n)} \right]^{-1} \hat{F}^{(n)\top} (I - S_i)\top (I - S_i)y_i, \tag{12}
\]

where

\[
S_i = \left( s_i(1/T)^\top x_i1, \ldots, s_i(T/T)^\top x_iT \right)^\top,
\]

with

\[
s_i(\tau) = [I_p, 0_p][M_i(\tau)^\top W(\tau)M_i(\tau)]^{-1}M_i(\tau)^\top W(\tau).
\]

After plugging \( \hat{\lambda}_i \) back into \( \hat{\beta}_i(\tau, \lambda_i) \), we have

\[
\hat{\beta}_i^{(n+1)}(\tau) = [I_p, 0_p] \left[ M_i(\tau)^\top W(\tau)M_i(\tau) \right]^{-1} M_i(\tau)^\top W(\tau) \left[ y_i - \hat{F}^{(n)\top} \hat{\lambda}_i^{(n+1)} \right], \tag{13}
\]

for \( i = 1, \ldots, N \).
Estimation Procedure

(4) With $\hat{\beta}_i^{(n+1)}(\tau)$ and $\hat{\lambda}_i^{(n+1)}$, we can estimate $F_t$ by OLS method:

$$\hat{F}_t^{(n+1)} = \left( \hat{\Lambda}^{(n+1)\top} \hat{\Lambda}^{(n+1)} \right)^{-1} \hat{\Lambda}^{(n+1)\top} R_{1,t}^{(n+1)}$$

where $R_{1,t}^{(n+1)} = \left( y_{1t} - x_{1t}^\top \hat{\beta}_1^{(n+1)}(\tau_t), \ldots, y_{Nt} - x_{Nt}^\top \hat{\beta}_N^{(n+1)}(\tau_t) \right)^\top$.

(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 \( R_F^{(n)} = \hat{F}^{(n)} - F^0 \). For the initial estimator \( \hat{F}^{(0)} \), suppose that

\[
T^{-1/2} \| R_F^{(0)} \| = O_P (\delta_{F,0}) \quad \text{and} \quad (Th)^{-1/2} \| W(\tau)^\top R_F^{(0)} \| = O_P (\delta_{F,0}),
\]

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. $N^{-1/2} \| \hat{\Lambda}^{(n)} - \Lambda \| = O_p(\max \{ \delta_{F,0}, \delta_{NT} \})$;
2. $T^{-1/2} \| \hat{F}^{(n)} - F \| = O_p(\max \{ \delta_{F,0}, \delta_{NT} \})$,

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

Assume that

\[ x_{it} = g_i(\tau_t) + v_{it}. \]  \hfill (14)

Notations:

- \[ \Sigma_{v,i} = E\left[ v_i v_i^\top \right], \quad \Sigma_F = E\left[ F_1 F_1^\top \right], \quad \Sigma_{v,F,i} = E\left[ v_i F_t^\top \right], \quad \Sigma_{v,\lambda,i} = E\left[ v_i \lambda_i^0 \right], \]
- \[ \Sigma_{X,i}(\tau) = g_i(\tau) g_i^\top(\tau) + \Sigma_{v,i}, \quad \Omega_{F,i} = \Sigma_F - \Sigma_{v,F,i} \int_0^1 \Sigma_{X,i}^{-1}(\tau) d\tau \Sigma_{v,F,i}, \]
- \[ \sigma_{ij,ts} = E[\varepsilon_{it}\varepsilon_{js}], \quad z_{it} = F_t^0 - \Sigma_{v,F,i}^\top \Sigma_{X,i}^{-1}(\tau_t)x_{it}, \quad \Sigma_\lambda = \lim_{N \to \infty} N^{-1} \sum_{i=1}^N \lambda_i^0 \lambda_i^{0\top}, \]
- \[ \Delta_{F,i} = \Sigma_{v,F,i} \Omega_{F,i}^{-1} \Sigma_{v,F,i}^\top, \quad \lambda_i^+(\tau) = \Sigma_{X,i}^{-1}(\tau) \left( \Sigma_{v,\lambda,i}(\tau) + g_i(\tau) \lambda_i^{0\top} \right), \]
- \[ \Omega_1(t,s) = N^{-1} \sum_{i=1}^N E\left[ \lambda_i^0 \lambda_i^{0\top} x_{it} \Sigma_{X,i}^{-1}(\tau_t)x_{is} \right], \quad \Omega_2(t,s) = N^{-1} \sum_{i=1}^N E\left[ \lambda_i^0 \lambda_i^{0\top} z_{it} \Omega_{F,i}^{-1} z_{is} \right], \]
- \[ \Omega_3(t,s) = \Sigma_\lambda^{-1}(h^{-1} K_{s,0}(\tau_t) \Omega_1(t,s) + \Omega_2(t,s)), \]
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

\[
\sqrt{N} \left( \hat{F}_t(n) - F_0 - b_{F,t}^+(n) \right) \xrightarrow{D} \mathcal{N}(\sqrt{c_1} d_{F,t}, \Sigma_{F,t}),
\]

where \( \Sigma_{F,t} = \Sigma_{\lambda}^{-1} \Sigma_{F,t} \Sigma_{\lambda}^{-1} \),

\[
b_{F,t}^+(n) = T^{-n} \sum_{s_1,s_2,...,s_n=1}^T \Omega_3(t,s_1) \prod_{j=1}^{n-1} \Omega_3(s_j,s_{j+1}) \mathbf{R}_{F,s_n}^{(0)},
\]

\[
d_{F,t} = \lim_{N,T \to \infty} 1/(N \sqrt{T}) \Sigma_{\lambda}^{-1} \sum_{i=1}^N \sum_{s=1}^T \Omega_{F,i}^{-1} \Sigma_{v,F,i}^{\top} \Sigma_{X,i}^{-1} (\tau_s) \mathbf{g}_i(\tau_s) \sigma_{ii,ts}.
\]
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

\[
\sqrt{T} \left( \hat{\lambda}_i^{(n)} - \lambda^0_i - b_{\lambda,i}^{+(n)} \right) \xrightarrow{D} \mathcal{N} \left( \sqrt{c_2} d_{\lambda,i}, \Sigma_{\lambda,i} \right),
\]

where \( \Sigma_{\lambda,i} = \Omega_{F,i}^{-1} \Sigma_{\lambda,i}^0 \Omega_{F,i}^{-1} \),

\[
b_{\lambda,i}^{+(n)} = T^{-1} \Omega_{F,i}^{-1} \Sigma_{\lambda,i}^\top \sum_{t=1}^T \lambda_i^{+(\tau_t)} b_{F,t}^{+(n-1)},
\]

\[
d_{\lambda,i}^* = 1/\sqrt{N} \Omega_{F,i}^{-1} \Sigma_{\lambda}^{-1} \mu_{\lambda} \sum_{j=1}^N \sigma_{ij,11}.
\]
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 \((i, \tau)\), we have

\[
\sqrt{T} h \left( \hat{\beta}_i^{(n)}(\tau) - \beta_i(\tau) - a_i(\tau) h^2 - b_{\beta,i}^{\dagger(n)}(\tau) \right) \xrightarrow{D} \mathcal{N}(0_p, \Sigma_{\beta,i}(\tau)),
\]

where

\[
a_i(\tau) = \frac{\mu_2}{2} \beta_i''(\tau)(1 + o(1)), \quad \Sigma_{\beta,i}(\tau) = \Sigma_{X,i}(\tau) \Sigma_{\beta,i}(\tau) \Sigma_{X,i}(\tau),
\]

\[
\mu_2 = \int u^2 K(u) du, \quad \text{and}
\]

\[
b_{\beta,i}^{\dagger(n)}(\tau) = -T^{-1} \Sigma_{X,i}(\tau) \sum_{t=1}^{T} \left( h^{-1} K_{t,0}(\tau) \Sigma_{X,i}(\tau) + \Delta_{F,i} \right) \lambda_i^{\dagger}(\tau_t) b_{F,t}^{\dagger(n-1)}.
\]

See Assumptions
Asymptotic Properties

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

1. \( \sqrt{N} \left( \hat{F}_t^{(n)} - F_t^0 - b_{F,t}^{(n)} \right) \xrightarrow{D} \mathcal{N}(0, \Sigma_{F,t}) 
\)

2. \( \sqrt{T} \left( \hat{\lambda}_i^{(n)} - \lambda_i^0 - b_{\lambda,i}^{(n)} \right) \xrightarrow{D} \mathcal{N}(0, \Sigma_{\lambda,i}^*) 
\)

3. \( \sqrt{Th} \left( \hat{\beta}_i^{(n)}(\tau) - \beta_i(\tau) - a_i(\tau)h^2 - b_{\beta,i}^{(n)}(\tau) \right) \xrightarrow{D} \mathcal{N}(0, \Sigma_{\beta,i}^*(\tau)) 
\)

where \( \Sigma_{\lambda,i}^* = \Omega_{F,i}^{-1} \sigma^2_\varepsilon \) and \( \Sigma_{\beta,i}^*(\tau) = v_0 \Omega_{X,i}^{-1}(\tau) \sigma^2_\varepsilon \).
Asymptotic Properties

Define

\[ \kappa = \lim_{N,T \to \infty} (NT)^{-1} \sum_{s=1}^{T} \sum_{i=1}^{N} g_{i}^\top(\tau_{t}) \Sigma_{X,i}^{-1}(\tau_{t}) \left( \Sigma_{v,F,i} \Omega_{F,i}^{-1} \Sigma_{v,F,i}^{-1} \Sigma_{X,i}^{-1}(\tau_{s}) g_{i}(\tau_{s}) + g_{i}(\tau_{t}) \right) \in [0, 1). \]

**Theorem 2.3 (CLT, n → ∞)** 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, \( N/T \rightarrow c_{1} < \infty \),

\[ \sqrt{N} \left( \hat{F}_{t}^{(n)} - F_{t}^{0} \right) \xrightarrow{D} \mathcal{N}(\sqrt{c_{1}}d_{F,t}, \Sigma_{F,t}). \]

2. If, in addition, \( T/N \rightarrow c_{2} < \infty \),

\[ \sqrt{T} \left( \hat{\lambda}_{i}^{(n)} - \lambda_{i}^{0} \right) \xrightarrow{D} \mathcal{N}(\sqrt{c_{2}}d_{\lambda,i}, \Sigma_{\lambda,i}). \]

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

\[ \sqrt{Th} \left( \hat{\beta}_{i}^{(n)}(\tau) - \beta_{i}(\tau) - a_{i}(\tau)h^{2} \right) \xrightarrow{D} \mathcal{N}(0_{p}, \Sigma_{\beta,i}(\tau)). \]
Asymptotic Properties

Consider the following mean-group estimator (MGE)

\[ \hat{\beta}_w^n(\tau) = \sum_{i=1}^{N} w_{N,i} \hat{\beta}_i^{(n)}(\tau), \]

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

\[ \sqrt{\gamma_{N,w} Th} \left( \hat{\beta}_w^n(\tau) - \beta_w(\tau) - a_w(\tau) h^2 \right) \xrightarrow{D} \mathcal{N}(0_p, \Sigma_{\beta,w}), \quad (15) \]

where

- \( \gamma_{N,w} = \left( \sum_{i=1}^{N} w_{N,i}^2 \right)^{-1}, \)
- \( a_w(\tau) = \frac{\mu^2}{2} \sum_{i=1}^{N} w_{N,i} \beta''_i(\tau)(1 + o_P(1)), \quad \beta_w(\tau) = \sum_{i=1}^{N} w_{N,i} \beta_i(\tau). \)

See Assumptions
Discussions on initial estimators
Exogenous factor models

Step 1. First, by the local linear method:

\[
\hat{\beta}_i^{(0)}(\tau) = [I_p, 0_p] \left( M_i^\top(\tau)W(\tau)M_i(\tau) \right)^{-1} M_i^\top(\tau)W(\tau)y_i. \tag{16}
\]

Step 2. Second, by PCA:

\[
\frac{1}{NT} \sum_{i=1}^{N} R_{2,i}R_{2,i}^\top \hat{F}^{(0)} = \hat{F}^{(0)}V_{NT,1}, \tag{17}
\]

where \( R_{2,i} = (R_{i1}(\hat{\beta}_i^{(0)}(\tau_1)), \cdots, R_{iT}(\hat{\beta}_i^{(0)}(\tau_T)))^\top \) with \( R_{it}(\beta) = y_{it} - x_{it}^\top \beta \), and \( V_{NT,1} \) is an \( r \times r \) diagonal matrix with diagonal elements being the first \( r \) largest eigenvalues of

the matrix \( (NT)^{-1} \sum_{i=1}^{N} R_{2,i}R_{2,i}^\top. \)
Corollary 3.1 (CLT, exogenous factor case) Let Assumptions 1.(i-v), 2-3, 5 hold. Suppose

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

We have Theorem 2.3.(1-3) holds.
Discussions on initial estimators

Endogenous factor models

Assume that

\[ x_{it} = g_i(\tau_t) + v_{it}, \quad v_{it} = \gamma_i^0 \! \top \! F_t^0 + \eta_{it}. \] (18)

Step 1. First, by the local linear method:

\[ \hat{g}_i^{(w)}(\tau) = [1, 0] \left( M_T^\top(\tau) W(\tau) M_T(\tau) \right)^{-1} M_T^\top(\tau) W(\tau) \tilde{x}_i^{(w)} \] (19)

where \( \hat{g}_i^{(w)}(\tau) \) is the \( w \)-th element of \( \hat{g}_i(\tau) \), \( \tilde{x}_i^{(w)} = (x_{i1}^{(w)}, \cdots, x_{iT}^{(w)})^\top \) and \( x_{it}^{(w)} \) is the \( w \)-th element of \( x_{it} \), for \( w = 1, 2, \ldots, p \).

Step 2. Second, by PCA:

\[ \left( \frac{1}{NTp} \sum_{w=1}^p \tilde{R}_g^{(w)} \tilde{R}_g^{(w)^\top} \right) \hat{F}(0) = \hat{F}(0) V_{NT,2} \] (20)

where \( \tilde{R}_g^{(w)} = (\tilde{R}_g^{(w),1}, \ldots, \tilde{R}_g^{(w),N}) \), \( \tilde{R}_g^{(w)} = (R_g^{(w),1}, \ldots, R_g^{(w),iT})^\top \) with \( R_g^{(w),it} \) being the \( w \)-th element of \( R_g, \) \( x_{it} = x_{it} - \hat{g}_i(\tau_t) \), and \( V_{NT,2} \) is an \( r \times r \) diagonal matrix with diagonal elements being the first \( r \) largest eigenvalues of the matrix \( (NTp)^{-1} \sum_{w=1}^p \tilde{R}_g^{(w)} \tilde{R}_g^{(w)^\top} \).
Discussions on initial estimators

Endogenous factor models

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

\[
\max \left\{ \sqrt{\frac{N}{T h}}, \sqrt{\frac{T}{N}} \right\} \kappa^{n-2} \to 0.
\]

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

- \((\beta_{1i}(u), \beta_{2i}(u)) = (\sin(\pi u) + \cos(0.25\pi i), \cos(\pi u) + 0.5 \sin(0.25\pi i))\);
- \(X_{it,1} = g_{i1}(\tau_t) + \gamma_{i1,1} G_{t,1} + \gamma_{i2,1} G_{t,2} + \eta_{it,1}\);
- \(X_{it,2} = g_{i2}(\tau_t) + \gamma_{i1,2} G_{t,1} + \gamma_{i2,2} G_{t,2} + \eta_{it,2}\);
- \((g_{i1}(u), g_{i2}(u)) = (3 \cos(\pi(u + 0.25i)), 5 \sin(\pi(u + 0.25i)))\);
- \(F_{t,1} = \rho_{F1} F_{t-1,1} + v_{F1,t} \text{ with } \rho_{F1} = 0.6;\)
- \(F_{t,2} = \rho_{F2} F_{t-1,2} + v_{F2,t} \text{ with } \rho_{F2} = 0.4;\)
- \((G_{t,1}, G_{t,2}) \sim i.i.d.N(0, 1); \) the loadings and error terms: \(\sigma_{ij,1} = 0.8|i-j|;\)
Simulation studies

An example with exogenous factors

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

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

For $\hat{F}_t^{(n)}$ and $\hat{\lambda}_i^{(n)}$,

- $r = 2$ as given.
Simulation studies

An example with exogenous factors

- Replication times: \( R = 1000 \) times;
- For each replication,

\[
\text{MSE}(\hat{\beta}_{l,w}^{(n)}) = \frac{1}{T} \sum_{t=1}^{T} \left( \hat{\beta}_{l,w}^{(n)}(\tau_t) - \beta_{l,w}(\tau_t) \right)^2,
\]

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 \( \{\hat{\lambda}_{i}^{(n)}\} \) and \( \{\lambda_i\} \), \( \{\hat{F}_i^{(n)}\} \) and \( F_t \) 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

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

<table>
<thead>
<tr>
<th>SCC</th>
<th>$\hat{\lambda}_i^{(n)}$</th>
<th>$\hat{F}_i^{(n)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>$N/T$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3619</td>
<td>0.4877</td>
<td>0.5527</td>
</tr>
<tr>
<td>(0.2266)</td>
<td>(0.2346)</td>
<td>(0.2349)</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4461</td>
<td>0.6297</td>
<td>0.7433</td>
</tr>
<tr>
<td>(0.2570)</td>
<td>(0.2388)</td>
<td>(0.1931)</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5667</td>
<td>0.8081</td>
<td>0.8985</td>
</tr>
<tr>
<td>(0.2688)</td>
<td>(0.1668)</td>
<td>(0.0597)</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6934</td>
<td>0.9178</td>
<td>0.9514</td>
</tr>
<tr>
<td>(0.2491)</td>
<td>(0.0565)</td>
<td>(0.0213)</td>
</tr>
</tbody>
</table>
Simulation studies

An example with endogenous factors

Example 2 Consider the following data generating process:

\[
X_{it,1} = g_{i,1}(\tau_t) + \gamma_{i,1,1}F_{t,1} + \gamma_{i,2,1}F_{t,2} + \eta_{it,1}
\]

\[
X_{it,2} = g_{i,2}(\tau_t) + \gamma_{i,1,2}F_{t,1} + \gamma_{i,2,2}F_{t,2} + \eta_{it,2}
\]  

(21)

where \((g_{i,1}(u), g_{i,2}(u)) = (3 \cos(\pi u), 5u)\). \((\gamma_{i,1,1}, \gamma_{i,1,2}), (F_{t,1}, F_{t,2})\) 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

<table>
<thead>
<tr>
<th>MSE</th>
<th>$\hat{\beta}_{w,1}^{(n)}$</th>
<th>$\hat{\beta}_{w,2}^{(n)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N/T$</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>0.2790</td>
<td>0.0883</td>
</tr>
<tr>
<td></td>
<td>(0.5040)</td>
<td>(0.0414)</td>
</tr>
<tr>
<td>20</td>
<td>0.1514</td>
<td>0.0607</td>
</tr>
<tr>
<td></td>
<td>(0.1648)</td>
<td>(0.0257)</td>
</tr>
<tr>
<td>40</td>
<td>0.1119</td>
<td>0.0537</td>
</tr>
<tr>
<td></td>
<td>(0.0783)</td>
<td>(0.0148)</td>
</tr>
<tr>
<td>80</td>
<td>0.0906</td>
<td>0.0437</td>
</tr>
<tr>
<td></td>
<td>(0.0304)</td>
<td>(0.0100)</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>SCC</th>
<th>$\lambda_i^{(n)}$</th>
<th>$\hat{F}_i^{(n)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>$N/T$</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>0.4638</td>
<td>0.5178</td>
</tr>
<tr>
<td></td>
<td>(0.2444)</td>
<td>(0.2326)</td>
</tr>
<tr>
<td>20</td>
<td>0.5328</td>
<td>0.6467</td>
</tr>
<tr>
<td></td>
<td>(0.2512)</td>
<td>(0.2188)</td>
</tr>
<tr>
<td>40</td>
<td>0.6824</td>
<td>0.8007</td>
</tr>
<tr>
<td></td>
<td>(0.2029)</td>
<td>(0.1391)</td>
</tr>
<tr>
<td>80</td>
<td>0.7202</td>
<td>0.8952</td>
</tr>
<tr>
<td></td>
<td>(0.2119)</td>
<td>(0.0901)</td>
</tr>
</tbody>
</table>
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:

\[
HE_{it} = \beta_{1, it} GDP_{it} + \beta_{2, it} DR_{it}^{young} + \beta_{3, it} DR_{it}^{old} + \beta_{4, it} GHE_{it} + \sum_{m=1}^{r} \lambda_{mi} f_{mt} + \varepsilon_{it},
\]  

(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.
An empirical application in health economics

The number of factors

The criterion proposed by Bai and Ng (2002):

\[
IC(r) = \log \left( \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 \right) + r \left( \frac{N + T}{NT} \right) \log \left( \min\{N, T\} \right)
\]

(23)

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

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

<table>
<thead>
<tr>
<th>( r )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( IC(r) )</td>
<td>-6.6058</td>
<td>-6.5600</td>
<td>-6.5538</td>
<td>-6.4607</td>
<td>-6.4057</td>
<td>-6.3390</td>
<td>-6.2940</td>
<td>-6.2798</td>
</tr>
</tbody>
</table>
An empirical application in health economics

Figure 3: The estimated elasticities and confidence intervals
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.
Figure 4: The estimated elasticities and confidence intervals (European OECD countries)
An empirical application in health economics

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

- $\mathbf{W}(\tau) = \text{diag}(K(\frac{1-\tau T}{Th}), \ldots, K(\frac{T-\tau T}{Th}))$

- $\tilde{\mathbf{W}}(\tau) = \mathbf{W}(\tau) \otimes \mathbf{I}_N$

- $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_T^\top)^\top$

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

(24)
Define

\[ \mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})^\top, \quad \mathbf{x}_t = (x_{1t}, x_{2t}, \ldots, x_{Nt}) \]

\[ \mathbf{V} = (v_1, v_2, \ldots, v_N)^\top, \quad \mathbf{\tilde{F}}_t = (\tilde{F}_{1t}, \tilde{F}_{jt}, \ldots, \tilde{F}_{rt})^\top, \]

\[ \mathbf{\tilde{F}} = (\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_T)^\top, \quad \mathbf{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})^\top. \]
Let $W_0(\tau) = \text{diag}(K(\frac{1-\tau T}{Th}), \ldots, K(\frac{T-\tau T}{Th}))$, $W(\tau) = W_0(\tau) \otimes I_N$, $\tilde{y}_t = M_V y_t$, $\tilde{x}_t = x_t M_V$ and

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

Notation

Define

\[ y_i = (y_{i1}, \ldots, y_{iT})^\top, \quad W(\tau) = \left( K \left( \frac{1 - \tau T}{Th} \right), \ldots, K \left( \frac{T - \tau T}{Th} \right) \right) \]

and

\[ M_i = \begin{pmatrix} x_{i1}^\top & \frac{1 - \tau T}{Th} x_{i1}^\top \\ \vdots & \vdots \\ x_{iT}^\top & \frac{T - \tau T}{Th} x_{iT}^\top \end{pmatrix}. \]
Notations:

- $\Omega_3(t, s) = \Sigma_\lambda^{-1}(h^{-1}K_{s,0}(\tau_t)\Omega_1(t, s) + \Omega_2(t, s))$, 

- $\lambda_i^\dagger(\tau_t) = \Sigma_{X,i}^{-1}(\tau_t) \left( \Sigma_{X,\lambda,i}(\tau_t) + E[x_{it}]\lambda_i^\top \right)$, 

- $\Delta_{F,i} = \Sigma_{v,F}^{-1} \Sigma_{v,F}^\top, \Sigma_{X,\lambda,i}(\tau_t) = E\left[ x_{it}\lambda_i^\top \right]$
Appendix

Assumptions

Assumption 1.

(i) \( \alpha \)-mixing conditions on panel data are assumed as follows: \( \{v_t, \varepsilon_t, F_t^0\} \) are strictly stationary and \( \alpha \)-mixing across \( t \); Let \( \alpha_{ij}(|t-s|) \) represent the \( \alpha \)-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 \( E \left[ \|\omega_{it}\|^{4+\delta} \right] < \infty \) with \( \omega_{it} \in \{\lambda_i^0, F_t^0, \varepsilon_{it}, v_{it}\} \). Let \( \alpha(|t-s|) \) represent the \( \alpha \)-mixing coefficient between \( \{v_{it}, F_t^0\} \) and \( \{v_{is}, F_s^0\} \). 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 \( \{F_s^0, \lambda_j^0, \varepsilon_{js}\} \), for any \( i, j, t, s \).

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

(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 \( R_F^{(n)} = \hat{F}^{(n)} - F^0 \). For the initial estimator \( \hat{F}^{(0)} \), suppose that

\[
T^{-1/2} \|R_F^{(0)}\| = O_P(\delta_{F,0}) \quad \text{and} \quad (Th)^{-1/2} \|W(\tau)^	op R_F^{(0)}\| = O_P(\delta_{F,0}),
\]

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 \( N, T \to \infty \).
Appendix
Assumptions

Notation:

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

\[ v_0 = \int K(u)^2 du, \quad \Sigma^0_{\beta,i}(\tau) = v_0 \left( \sigma_{v,i}^2 + \sigma_{\varepsilon,0}^2 g_i(\tau) g_i^\top(\tau) \right), \]

\[ \xi_{1,it} = \lambda_i^{0\top} \mathbf{F}_i^0, \quad \xi_{2,it} = \mathbf{v}_{it} \lambda_i^{0\top}, \quad \sigma_{F,\varepsilon,0}^2 = \sigma_\varepsilon^2 \mathbf{F} + 2 \sum_{t=2}^{\infty} E[\varepsilon_{i1}\varepsilon_{it}] E\left[\mathbf{F}_i^0 \mathbf{F}_i^{0\top}\right], \]

\[ \Sigma_{\lambda,i}^0 = \sigma_{F,\varepsilon,0}^2 - \int_0^1 \sum_{\mathbf{v}_i} \mathbf{F}_i \mathbf{X}_i^{-1}(v) \left( \sigma_{v,i}^2 + \sigma_{\varepsilon,0}^2 g_i(v) g_i^\top(v) \right) \mathbf{X}_i^{-1}(v) \sum_{\mathbf{v}_i} d\mathbf{v} \]

Assumption 2.

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

\[ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} |\text{Cov}(\varepsilon_{i1}, \varepsilon_{it2}, \varepsilon_{jt3}, \varepsilon_{jt4})| \leq CNT^2 \]

\[ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} |\text{Cov}(\xi_{1,it1}, \xi_{1,it2}, \xi_{1,jt3}, \xi_{1,jt4})| \leq CNT^2 \]

\[ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} \|\text{Cov}(\xi_{2,it1}, \xi_{2,it2}, \xi_{2,jt3}, \xi_{2,jt4})\| \leq CNT^2 \]
Assumption 2.

(ii) Assume that $\Sigma_{v,i}, \Sigma_F, \Sigma^{0}_{\beta,i}(\tau)$ and $\Sigma^{0}_{\lambda,i}$ are positive definite and $\sigma^2_{\epsilon}$ is a positive scalar.

(iii) Suppose that $\left\|N^{-1} \sum_{i=1}^{N} \lambda^{0}_i \lambda^{0T}_i - \Sigma_{\lambda}\right\| = O_P(N^{-1/2})$ and

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

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

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

Let $E \left[ \lambda_i^0 \lambda_i^{0\top} | v_{i1}, \ldots, v_{iT}, F_{1}^{0\top}, \ldots, F_{T}^{0\top} \right] = \Sigma_{\lambda}$ almost surely, where $
abla_{\lambda} = \lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} \lambda_i^0 \lambda_i^{0\top}$ is positive definite.
Appendix

Assumptions for the heterogeneous model

Assumption 4.

(i) Assume that $E\left[v_{it}\lambda_i^{0\top}\right] = E\left[v_tF_t^{0\top}\right] = 0_{p \times r}$ and $E[\lambda_i] = 0_r$.

(ii) Define that

\[ \tilde{\sigma}_{v,\epsilon}^2(i, j, \tau) = \Sigma_{X,i}^{-1}(\tau)\sigma_{v,\epsilon}^2(i, j)\Sigma_{X,j}^{-1}(\tau), \]

\[ \tilde{\sigma}_{\epsilon}^2(i, j, \tau) = \sigma_{\epsilon}^2(i, j)\Sigma_{X,i}^{-1}(\tau)g_i(\tau)g_j^\top(\tau)\Sigma_{X,j}^{-1}(\tau), \]

\[ \Sigma_{\beta,\omega}(\tau) = \lim_{N \to \infty} \gamma_{N,\omega} v_0 \sum_{i=1}^N \sum_{j=1}^N w_{N,i}w_{N,j} \left( \tilde{\sigma}_{\epsilon}^2(i, j, \tau) + \tilde{\sigma}_{v,\epsilon}^2(i, j, \tau) \right). \]

We assume $\Omega_{F,i}$ and $\Sigma_{\beta,\omega}(\tau)$ are positive-definite matrices, where $\Omega_{F,i}$ is defined in Theorem 1.

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

Assumption 5.

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

$$N^{-1}\hat{\Lambda}^{(0)\top} \hat{\Lambda}^{(0)} = \text{diagonal} \quad \text{and} \quad T^{-1}\hat{F}^{(0)\top} \hat{F}^{(0)} = I_r.$$ 

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

(iii) Suppose $F^0_t$ is conditionally uncorrelated with $\Lambda^0, v_1, \ldots, v_T$:

$$E \left[ F^0_t | \Lambda^0, v_1, \ldots, v_T \right] = 0_r.$$ 

In addition, we assume $\{F^0_t | \Lambda^0, v_1, \ldots, v_T\}$ satisfies the $\alpha$-mixing condition in Assumption 1.

(iv) Suppose the following moment conditions can hold:

$$\sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} \left\| E \left[ F^0_{t_1} F^0_{t_2}^{\top} F^0_{t_3} F^0_{t_4}^{\top} \right] \right\| \leq CT^2,$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3 \neq t_1}^{T} \sum_{t_4 \neq t_2}^{T} \left| E \left[ \varepsilon_{i t_1} \varepsilon_{j t_2} \varepsilon_{i t_3} \varepsilon_{j t_4} \right] \right| \leq CNT^2.$$
Appendix

Assumptions

Assumption 6.

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

\[
N^{-1} \sum_{i=1}^{N} \hat{\gamma}^{(w,0)^\top}_i \hat{\gamma}^{(w,0)}_i = \text{diagonal} \quad \text{and} \quad T^{-1} \hat{F}^{(0)^\top} \hat{F}^{(0)} = I_r,
\]

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

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

(iii) The unknown deterministic function \( g_i(\tau) \) 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:

\[
\sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3=1}^{T} \sum_{t_4=1}^{T} \left\| E \left[ F^0_{t_1} F^0_{t_2}^\top F^0_{t_3} F^0_{t_4}^\top \right] \right\| \leq C T^2,
\]

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t_1=1}^{T} \sum_{t_2=1}^{T} \sum_{t_3 \neq t_1}^{T} \sum_{t_4 \neq t_2}^{T} \left\| E \left[ \eta^\top_{it_1} \eta^\top_{jt_2} \eta^\top_{it_3} \eta^\top_{jt_4} \right] \right\| \leq C N T^2.
\]
Appendix

Estimated loadings and factors

Figure 7: The estimated loadings and factors
Appendix

Bootstrapping

The details for our bootstrapping method are as follows:

Step 1. Calculate the residuals \( \{\varepsilon_{it}\} \) for the estimation method discussed in Section 2.

Step 2. Resample the residuals and obtain \( \{\varepsilon^*_it\} \), 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 \( \{\varepsilon^*_it\} \).

Step 3. The bootstrapping estimator \( \bar{\beta}^*_t \) can be obtained using the data set \( \{Y^*_it\} \).

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

Discussions on initial estimator: exogenous factors

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

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

$$\hat{\beta}_{i}^{(0)}(\tau) = [I_p, 0_p] \left( M_i^\top(\tau) W(\tau) M_i(\tau) \right)^{-1} M_i^\top(\tau) W(\tau) y_i,$$

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

(2) Then estimate $F$ using the PCA method as follows:

$$\frac{1}{NT} \sum_{i=1}^{N} R_{3,i} R_{3,i}^\top \hat{F}^{(0)} = \hat{F}^{(0)} V_{NT,F},$$  \hspace{1cm} (25)

where $R_{3,i} = \left( R_{i1}(\hat{\beta}_{1}^{(0)}(\tau_1)), \ldots, R_{iT}(\hat{\beta}_{i}^{(0)}(\tau_T)) \right)^\top$ and $R_{it}(\beta) = y_{it} - x_{it}^\top \beta(\tau_t)$. 
Corollary 3.2 Under some regularity conditions and $\hat{F}^{(0)}$ satisfies (25),

$$
\frac{1}{\sqrt{T}} \left\| \hat{F}^{(0)} - FH_1 \right\| = O_p \left( \max\{ (Th)^{-1/2}, N^{-1/2}, h^2 \} \right), \tag{26}
$$

where $H_1 = (NT)^{-1} \sum_{i=1}^{N} \lambda_i \lambda_i^\top F^\top \hat{F}^{(0)} V_{NT,1}^{-1}$. 
Appendix

Discussions on initial estimator: endogenous factors

Consider the following model:

\[ y_{it} = x_{it}^\top \beta_{it} + \lambda_i^\top F_t + \varepsilon_{it} \]

\[ x_{it} = g_i(\tau_t) + \gamma_i^\top F_t + \eta_{it} \]

PCA method to estimate \( \hat{F}^{(0)} \),

1. We first estimate the \( g_i(\tau) \) using local linear method:

\[ \hat{g}^{(w)}_i(\tau) = [1, 0] \left( M_T(\tau)W(\tau)M_T(\tau) \right)^{-1} M_T(\tau) W(\tau) \tilde{x}_i^{(w)} \]  

(27)

where \( \hat{g}^{(w)}_i(\tau) \) is the \( w \)-th element of \( \hat{g}_i(\tau) \), \( \tilde{x}_i^{(w)} = (x_{i1}^{(w)}, \ldots, x_{iT}^{(w)})^\top \) and \( x_{it}^{(w)} \) is the \( w \)-th element of \( x_{it} \).

2. Then \( F_t \) can be estimated by the PCA method:

\[ \left( \frac{1}{NTp} \sum_{w=1}^p \tilde{R}_g^{(w)} \tilde{R}_g^{(w)^\top} \right) \hat{F}^{(0)} = \hat{F}^{(0)} V_{NTg} \]  

(28)

where \( \tilde{R}_g^{(w)} = (\tilde{R}_{g,1}^{(w)}, \ldots, \tilde{R}_{g,N}^{(w)}) \), \( \tilde{R}_g^{(w)} = (R_{g,1}^{(w)}, \ldots, R_{g,iT}^{(w)})^\top \) and \( R_{g,it}^{(w)} \) is the \( w \)-th element of \( R_{g,it} = x_{it} - \hat{g}_i(\tau_t) \).
Appendix

Discussions on initial estimator: endogenous factors

**Corollary 3.3** Under some regularity conditions and \( \hat{F}^{(0)} \) satisfies (28),

\[
\frac{1}{\sqrt{T}} \| \hat{F}^{(0)} - FH_1 \| = \text{Op} \left( \max \{ (Th)^{-1/2}, N^{-1/2}, h^2 \} \right),
\]

(29)

where \( H_2 = \frac{1}{NTp} \sum_{w=1}^{p} \sum_{i=1}^{N} \gamma_i^{(w)} \gamma_i^{(w)\top} F^\top \hat{F}^{(0)} V_{NT,2}^{-1} \).
Reference


