Estimation and Inference in High-Dimensional Panel Data Models with Interactive Fixed Effects

Maximilian Rücker
Ulm University

Joint work with Oliver Linton (Cambridge University), Michael Vogt (Ulm University) and Christopher Walsh (Newcastle University)

IPDC 2024 Conference, Orléans
July 04, 2024

Model setting

- Consider panel data

$$\{(Y_{it}, X_{it}) : 1 \leq t \leq T, 1 \leq i \leq n\}$$

Model setting

- Consider panel data

$$\{(Y_{it}, X_{it}) : 1 \leq t \leq T, 1 \leq i \leq n\}$$

- Data generated according to linear model

$$Y_{it} = X_{it}\beta + e_{it}$$

  - Covariates X_{it} in $\mathbb{R}^p$
  - β in $\mathbb{R}^p$ is high-dimensional sparse vector

Model setting

- Consider panel data

$$\{(Y_{it}, X_{it}) : 1 \leq t \leq T, 1 \leq i \leq n\}$$

- Data generated according to linear model

$$Y_{it} = X_{it}\beta + e_{it}$$

  - Covariates X_{it} in $\mathbb{R}^p$
  - β in $\mathbb{R}^p$ is high-dimensional sparse vector
- Error structure

$$e_{it} = \gamma_i^\top F_t + \varepsilon_{it}$$

Model setting

- Consider panel data

$$\{(Y_{it}, X_{it}) : 1 \leq t \leq T, 1 \leq i \leq n\}$$

- Data generated according to linear model

$$Y_{it} = X_{it}\beta + e_{it}$$

  - Covariates X_{it} in $\mathbb{R}^p$
  - β in $\mathbb{R}^p$ is high-dimensional sparse vector
- Error structure

$$e_{it} = \gamma_i^\top F_t + \varepsilon_{it}$$

- Covariate structure

$$X_{it} = \Gamma_i F_t + Z_{it}$$

Model specification

- Interactive fixed effects model with common correlation.

Model specification

- Interactive fixed effects model with common correlation.
- Model contains the standard fixed effects model as a special case.

Model specification

- Interactive fixed effects model with common correlation.
- Model contains the standard fixed effects model as a special case.
- Standard case (Low Dimensional): Studied by Pesaran in [2006]
    - p fixed and $n, T \to \infty$
    - p low-dimensional, i.e. $p < T < nT$

Model specification

- Interactive fixed effects model with common correlation.
- Model contains the standard fixed effects model as a special case.
- Standard case (Low Dimensional): Studied by Pesaran in [2006]
  - p fixed and $n, T \to \infty$
  - p low-dimensional, i.e. $p < T < nT$
- Our model setting (High-dimensional)
  - $p \to \infty$, i.e. possibly $p >> nT$
  - $n \to \infty$ and $T \to \infty$ (Large-T-case)
  - $n \to \infty$ and T fixed (Small-T-case)
  - β is a sparse parameter vector

Aim of this work

- Consistent estimation of unknown sparse parameter vector β.

Aim of this work

- Consistent estimation of unknown sparse parameter vector β.
- Develop inference procedure to obtain e.g. confidence bands or statistical tests for β.

Aim of this work

- Consistent estimation of unknown sparse parameter vector β.
- Develop inference procedure to obtain e.g. confidence bands or statistical tests for β.
- Implement R-package for simple computation of our estimator.

Estimation strategy

- Step 1: Stack the equations across t to obtain:

$$\begin{aligned} Y_i &= \boldsymbol{X}_i \beta + \boldsymbol{F} \gamma_i + \varepsilon_i \\ \boldsymbol{X}_i &= \boldsymbol{F} \boldsymbol{\Gamma}_i^\top + \boldsymbol{Z}_i, \end{aligned}$$

for $1 \leq i \leq n$, where

$$\underset{(T \times 1)}{Y_i} = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iT} \end{pmatrix}, \quad \underset{(T \times p)}{\boldsymbol{X}_i} = \begin{pmatrix} X_{i1}^\top \\ \vdots \\ X_{iT}^\top \end{pmatrix}, \quad \underset{(T \times K)}{\boldsymbol{F}} = \begin{pmatrix} F_1^\top \\ \vdots \\ F_T^\top \end{pmatrix}, \ldots$$

Estimation strategy

- Step 2: Project away the unknown factors $\boldsymbol{F}$ with some suitable projection matrix $\widehat{\boldsymbol{\Pi}}$:

$$\widehat{\boldsymbol{\Pi}} Y_i = \widehat{\boldsymbol{\Pi}} \boldsymbol{X}_i \beta + \underbrace{\widehat{\boldsymbol{\Pi}} \boldsymbol{F} \gamma_i}_{\approx 0} + \widehat{\boldsymbol{\Pi}} \varepsilon_i$$

Estimation strategy

- Step 2: Project away the unknown factors $\boldsymbol{F}$ with some suitable projection matrix $\widehat{\boldsymbol{\Pi}}$:

$$\widehat{\boldsymbol{\Pi}} Y_i = \widehat{\boldsymbol{\Pi}} \boldsymbol{X}_i \beta + \underbrace{\widehat{\boldsymbol{\Pi}} \boldsymbol{F} \gamma_i}_{\approx 0} + \widehat{\boldsymbol{\Pi}} \varepsilon_i$$

  - Problem: How to choose unknown projection matrix $\widehat{\boldsymbol{\Pi}}$?
    - ! Pesaran's CCE approach collapses in high-dimensions

Estimation strategy

- Step 3: Apply the lasso estimator to the projected regression problem:

$$\widehat{\beta}_\lambda \in \underset{b\in\mathbb{R}^p}{\operatorname{argmin}}\left\{ \frac{1}{nT} \sum_{i=1}^{n} \left\| \widehat{\mathbf{\Pi}} Y_i - \widehat{\mathbf{\Pi}} \boldsymbol{X}_i b \right\|_2^2 + \lambda \|b\|_1 \right\}$$

Estimation strategy

- Step 3: Apply the lasso estimator to the projected regression problem:

$$\widehat{\beta}_\lambda \in \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{nT} \sum_{i=1}^{n} \left\| \widehat{\mathbf{\Pi}} Y_i - \widehat{\mathbf{\Pi}} \boldsymbol{X}_i b \right\|_2^2 + \lambda \|b\|_1 \right\}$$

- Problem: Analysis of lasso estimator $\widehat{\beta}_\lambda$
  - ! Least squares estimator not available in high-dimensions anymore.
  - ! Investigate restricted eigenvalue condition.

Estimation strategy

- Step 3: Apply the lasso estimator to the projected regression problem:

$$\widehat{\beta}_\lambda \in \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{nT} \sum_{i=1}^{n} \left\| \widehat{\mathbf{\Pi}} Y_i - \widehat{\mathbf{\Pi}} \boldsymbol{X}_i b \right\|_2^2 + \lambda \|b\|_1 \right\}$$

- Problem: Analysis of lasso estimator $\widehat{\beta}_\lambda$
  - ! Least squares estimator not available in high-dimensions anymore.
  - ! Investigate restricted eigenvalue condition.
  - ! Analysis of the lasso's effective noise and derive a suitable asymptotics for the penalty parameter λ.

Construction of $\widehat{\mathbf{\Pi}}$

(1) Pesaran's CCE approach: Uses cross-sectional mean matrix $\overline{\boldsymbol{X}} \in \mathbb{R}^{T \times p}$ to construct $\widehat{\mathbf{\Pi}}$. Idea: $\overline{\boldsymbol{X}} \approx \boldsymbol{F}\mathbf{\Gamma}^\top$, i.e. $\overline{\boldsymbol{X}}$ and $\boldsymbol{F}$ have a similar image.

Construction of $\widehat{\boldsymbol{\Pi}}$

(1) Pesaran's CCE approach: Uses cross-sectional mean matrix $\overline{\boldsymbol{X}} \in \mathbb{R}^{T \times p}$ to construct $\widehat{\boldsymbol{\Pi}}$. Idea: $\overline{\boldsymbol{X}} \approx \boldsymbol{F}\boldsymbol{\Gamma}^\top$, i.e. $\overline{\boldsymbol{X}}$ and $\boldsymbol{F}$ have a similar image.

- Set
$$\widehat{\boldsymbol{\Pi}} = \boldsymbol{I} - \overline{\boldsymbol{X}}(\overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}})^{-1}\overline{\boldsymbol{X}}^\top$$

Construction of $\widehat{\boldsymbol{\Pi}}$

(1) Pesaran's CCE approach: Uses cross-sectional mean matrix $\overline{\boldsymbol{X}} \in \mathbb{R}^{T \times p}$ to construct $\widehat{\boldsymbol{\Pi}}$. Idea: $\overline{\boldsymbol{X}} \approx \boldsymbol{F}\boldsymbol{\Gamma}^\top$, i.e. $\overline{\boldsymbol{X}}$ and $\boldsymbol{F}$ have a similar image.

- Set

$$\widehat{\boldsymbol{\Pi}} = \boldsymbol{I} - \overline{\boldsymbol{X}}(\overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}})^{-1} \overline{\boldsymbol{X}}^\top$$

(2) Collapses in high dimensions: $\overline{\boldsymbol{X}}$ spans whole $\mathbb{R}^T$.

Construction of $\widehat{\boldsymbol{\Pi}}$

(1) Pesaran's CCE approach: Uses cross-sectional mean matrix $\overline{\boldsymbol{X}} \in \mathbb{R}^{T \times p}$ to construct $\widehat{\boldsymbol{\Pi}}$. Idea: $\overline{\boldsymbol{X}} \approx \boldsymbol{F}\boldsymbol{\Gamma}^\top$, i.e. $\overline{\boldsymbol{X}}$ and $\boldsymbol{F}$ have a similar image.

- Set

$$\widehat{\boldsymbol{\Pi}} = \boldsymbol{I} - \overline{\boldsymbol{X}}(\overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}})^{-1} \overline{\boldsymbol{X}}^\top$$

(2) Collapses in high dimensions: $\overline{\boldsymbol{X}}$ spans whole $\mathbb{R}^T$.

(3) Data driven estimator for number of factors K with techniques from factor analysis.

Construction of $\widehat{\boldsymbol{\Pi}}$

(1) Pesaran's CCE approach: Uses cross-sectional mean matrix $\overline{\boldsymbol{X}} \in \mathbb{R}^{T \times p}$ to construct $\widehat{\boldsymbol{\Pi}}$. Idea: $\overline{\boldsymbol{X}} \approx \boldsymbol{F}\boldsymbol{\Gamma}^\top$, i.e. $\overline{\boldsymbol{X}}$ and $\boldsymbol{F}$ have a similar image.

- Set

$$\widehat{\boldsymbol{\Pi}} = \boldsymbol{I} - \overline{\boldsymbol{X}}(\overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}})^{-1}\overline{\boldsymbol{X}}^\top$$

(2) Collapses in high dimensions: $\overline{\boldsymbol{X}}$ spans whole $\mathbb{R}^T$.

(3) Data driven estimator for number of factors K with techniques from factor analysis.

(4) Define $\overline{\boldsymbol{W}} = \overline{\boldsymbol{X}}\widehat{\boldsymbol{U}}$ for some suitable matrix $\widehat{\boldsymbol{U}}$ obtained by spectral analysis of $\frac{1}{T}\overline{\boldsymbol{X}}^\top \overline{\boldsymbol{X}}$ and set

$$\widehat{\boldsymbol{\Pi}} = \boldsymbol{I} - \overline{\boldsymbol{W}}(\overline{\boldsymbol{W}}^\top \overline{\boldsymbol{W}})^{-1}\overline{\boldsymbol{W}}^\top.$$

Estimation result for large-T-case

Theorem (Convergence of the HD-CCE estimator).

Consider the large-T-case. Assume that certain regularity conditions are satisfied. Let the penalty parameter λ be equal to $\lambda = h_n \log(npT)/\min\{n, \sqrt{nT}\}$. Then

$$\|\widehat{\beta}_\lambda - \beta\|_1 = O_p\Big(s\,\frac{h_n \log(npT)}{\min\{n, \sqrt{nT}\}}\Big).$$

Estimation result for large-T-case

Theorem (Convergence of the HD-CCE estimator).
Consider the large-T-case. Assume that certain regularity conditions are satisfied. Let the penalty parameter λ be equal to $\lambda = h_n \log(npT)/\min\{n, \sqrt{nT}\}$. Then

$$\|\widehat{\beta}_\lambda - \beta\|_1 = O_p\Big(s\,\frac{h_n \log(npT)}{\min\{n, \sqrt{nT}\}}\Big).$$

- Example: $n = T$, then the rate becomes

$$\|\widehat{\beta}_\lambda - \beta\|_1 = O_p\Big(s\,\frac{h_n \log(npT)}{\sqrt{nT}}\Big)$$

which is up to log-factors the standard lasso rate.

  - For fixed s this is almost the rate $\frac{1}{\sqrt{nT}}$.

Estimation result for large-T-case

Theorem (Convergence of the HD-CCE estimator).
Consider the large-T-case. Assume that certain regularity conditions are satisfied. Let the penalty parameter λ be equal to $\lambda = h_n \log(npT)/\min\{n, \sqrt{nT}\}$. Then

$$\|\widehat{\beta}_\lambda - \beta\|_1 = O_p\Big(s\,\frac{h_n \log(npT)}{\min\{n, \sqrt{nT}\}}\Big).$$

- Example: $n = T$, then the rate becomes

$$\|\widehat{\beta}_\lambda - \beta\|_1 = O_p\Big(s\,\frac{h_n \log(npT)}{\sqrt{nT}}\Big)$$

  which is up to log-factors the standard lasso rate.
  - For fixed s this is almost the rate $\frac{1}{\sqrt{nT}}$.
- Small-T-case: Obtain a similar estimation result

Inference

- Motivation:
  - Testing $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$
  - Confidence intervals for β_j

Inference

- Motivation:
  - Testing $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$
  - Confidence intervals for β_j
- Problems:
  - ! Lasso estimate has high bias
  - ! Complicated limiting distribution (Fu, W. & Knight K. (2000))
- Taking estimator directly for inference is not suitable.

Inference

- Motivation:
  - Testing $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$
  - Confidence intervals for β_j
- Problems:
  - ! Lasso estimate has high bias
  - ! Complicated limiting distribution (Fu, W. & Knight K. (2000))
- Taking estimator directly for inference is not suitable.
- Solution: Follow debiasing procedure of Zhang and Zhang, (2014) and Van de Geer et. al. (2014)

Inference strategy

- Step 1: Extend the model setting with nodewise regression structure

  $$X_{it,j} = X_{it,-j}\theta + \nu_i^\top F_t + u_{it}$$

  for all i and t or stacked across t

  $$X_{i(j)} = \boldsymbol{X}_{i(-j)}\theta + \boldsymbol{F}\nu_i + u_i$$

  for all i.
  - θ is sparse
  - ν_i is a vector of factor loadings
  - u_{it} is a mean-zero error term

Inference strategy

- Step 2: Project away the unknown factors $\boldsymbol{F}$ with some suitable projection matrix $\widetilde{\boldsymbol{\Pi}}$:

$$\widetilde{\boldsymbol{\Pi}} Y_i = \widetilde{\boldsymbol{\Pi}} \boldsymbol{X}_i \beta + \widetilde{\boldsymbol{\Pi}} \boldsymbol{F} \gamma_i + \widetilde{\boldsymbol{\Pi}} \varepsilon_i$$

  - Projection $\widetilde{\boldsymbol{\Pi}}$ built in the same fashion as for estimation but only based on $p-1$ regressors.

Inference strategy

- Step 2: Project away the unknown factors $\boldsymbol{F}$ with some suitable projection matrix $\widetilde{\boldsymbol{\Pi}}$:

$$\widetilde{\boldsymbol{\Pi}} Y_i = \widetilde{\boldsymbol{\Pi}} \boldsymbol{X}_i \beta + \widetilde{\boldsymbol{\Pi}} \boldsymbol{F} \gamma_i + \widetilde{\boldsymbol{\Pi}} \varepsilon_i$$

  - Projection $\widetilde{\boldsymbol{\Pi}}$ built in the same fashion as for estimation but only based on $p-1$ regressors.

- Step 3: Apply the lasso estimator to the projected regression problem:

$$\widetilde{\beta}_\lambda \in \underset{b \in \mathbb{R}^p}{\operatorname{argmin}} \left\{ \frac{1}{nT} \sum_{i=1}^{n} \left\| \widetilde{\boldsymbol{\Pi}} Y_i - \widetilde{\boldsymbol{\Pi}} \boldsymbol{X}_i b \right\|_2^2 + \lambda \|b\|_1 \right\}$$

Inference strategy

- Step 4: Apply the lasso estimator to the nodewise regression problem:

$$\widetilde{\theta}_\kappa \in \underset{b \in \mathbb{R}^{p-1}}{\operatorname{argmin}} \left\{ \frac{1}{nT} \sum_{i=1}^{n} \left\| \widetilde{\boldsymbol{\Pi}} X_{i(j)} - \widetilde{\boldsymbol{\Pi}} \boldsymbol{X}_{i(-j)} b \right\|_2^2 + \kappa \|b\|_1 \right\}$$

Inference strategy

- Step 4: Apply the lasso estimator to the nodewise regression problem:

$$\widetilde{\theta}_\kappa \in \operatorname*{argmin}_{b \in \mathbb{R}^{p-1}} \left\{ \frac{1}{nT} \sum_{i=1}^{n} \left\| \widetilde{\mathbf{\Pi}} X_{i(j)} - \widetilde{\mathbf{\Pi}} \boldsymbol{X}_{i(-j)} b \right\|_2^2 + \kappa \|b\|_1 \right\}$$

- Step 5: Debiasing the lasso

$$\widetilde{b}_j = \widetilde{\beta}_{\lambda,j} + \frac{\widetilde{R}^\top (\widetilde{Y} - \widetilde{\boldsymbol{X}} \widetilde{\beta}_\lambda)}{\widetilde{R}^\top \widetilde{X}_{(j)}}$$

Inference strategy

- Step 4: Apply the lasso estimator to the nodewise regression problem:

$$\widetilde{\theta}_\kappa \in \underset{b\in\mathbb{R}^{p-1}}{\operatorname{argmin}}\left\{ \frac{1}{nT}\sum_{i=1}^{n} \left\|\widetilde{\boldsymbol{\Pi}}X_{i(j)} - \widetilde{\boldsymbol{\Pi}}\boldsymbol{X}_{i(-j)}b\right\|_2^2 + \kappa\|b\|_1 \right\}$$

- Step 5: Debiasing the lasso

$$\widetilde{b}_j = \widetilde{\beta}_{\lambda,j} + \frac{\widetilde{R}^\top(\widetilde{Y} - \widetilde{\boldsymbol{X}}\widetilde{\beta}_\lambda)}{\widetilde{R}^\top\widetilde{X}_{(j)}}$$

  - $\widetilde{R}_i = \widetilde{\boldsymbol{\Pi}}X_{i(j)} - \widetilde{\boldsymbol{\Pi}}\boldsymbol{X}_{i(-j)}\widetilde{\theta}_\kappa$ are the nodewise lasso residuals.
  - $\widetilde{R} = (\widetilde{R}_1^\top, \ldots, \widetilde{R}_n^\top)^\top$.
  - $\tilde{Y} = \begin{pmatrix} \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_n \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\Pi}}Y_1 \\ \vdots \\ \widetilde{\boldsymbol{\Pi}}Y_n \end{pmatrix}, \tilde{\boldsymbol{X}} = \begin{pmatrix} \widetilde{\boldsymbol{\Pi}}\boldsymbol{X}_1 \\ \vdots \\ \widetilde{\boldsymbol{\Pi}}\boldsymbol{X}_n \end{pmatrix}, \tilde{X}_{(j)} = \begin{pmatrix} \widetilde{\boldsymbol{\Pi}}X_{1(j)} \\ \vdots \\ \widetilde{\boldsymbol{\Pi}}X_{n(j)} \end{pmatrix}$

Inference – Distributional result

Theorem (Asymptotic distribution of debiased lasso). Under certain regularity conditions it holds that

$$\frac{\widetilde{R}^\top \widetilde{X}_{(j)}}{\|\widetilde{R}\|_2}(\widetilde{b}_j - \beta_j) \xrightarrow{d} N(0, \sigma_\varepsilon^2).$$

- Estimate σ_ε^2, by

$$\widetilde{\sigma}_\varepsilon^2 = \left(\frac{T}{T-K}\right)\frac{\|\widetilde{\varepsilon}\|_2^2}{nT}$$

where $\widetilde{\varepsilon}$ are the Lasso residuals.

Summary

- Estimation and inference in high-dimensional interactive fixed effects model.

Summary

- Estimation and inference in high-dimensional interactive fixed effects model.
- R-package containing the high-dimensional estimator.

! Work in progress: Empirical example, package for debiased lasso estimator.

Literature

- Bai (2009). Panel data models with interactive fixed effects. *Econometrica*.
- Belloni, Chernozhukov, Hansen & Kozbur (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*.
- Belloni, Chen, Padilla & Wang (2019). High dimensional latent panel quantile regression with an application to asset pricing. Preprint.
- Fan, Liao & Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B*.
- Lu & Su (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics.*
- Pesaran (2006). Estimation and inference in large heterogeneous panels with a multifactor error. *Econometrica*.
- Vogt, Walsh & Linton(2022). CCE Estimation of High-Dimensional Panel Data Models with Interactive Fixed Effects. *arXiv:2206.12152.*
- Fu, W. & Knight K. (2000). Asymptotics for lasso-type estimators. *The Annals of Statistics*