Factor and factor loading augmented estimators for panel regression

Jad BEYHUM (ORSTAT, KU Leuven) & Eric GAUTIER (Toulouse School of Economics)

EC^2 Conference, December 11-12, 2020, CREST-ENSAE & ESSEC

Objectives

- Develop and study the properties of a new methodology to estimate the regression parameter in panel data models with interactive fixed effects.
- The estimation procedure is composed of two steps:

(1) Estimate the factors and factor loadings;
(2) Regress the outcome on the regressors and the estimates of the factors and the factor loadings.

The model

We consider a panel data model with N individuals and T dates following the relationship

$$Y_{it} = \sum_{k=1}^{K} \beta_k X_{kit} + \sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_j + E_{it},$$

$$\forall k = 1, \dots, K,\ X_{kit} = \sum_{j=1}^{r_N} \lambda_{ij} f_{tj} \delta_{kj} + E_{kit}.$$

- $Y_{it}, X_{1it}, \dots X_{Kit}, E_{it}, E_{1it}, \dots, E_{Kit}$ are scalar random variables, $\beta \in \mathbb{R}^p$.
- $\{\lambda_i\}_{i=1}^N$ (resp. $\{f_t\}_{t=1}^T$) are random vectors of factors loadings (resp. factors) with support in $\mathbb{R}^{r_N}$.
- $\delta, \delta_1, \dots, \delta_K$ are vectors in $\mathbb{R}^{r_N}$ allowing the factors and factor loadings to appear in both equations.
- $E, E_1, \dots, E_K$ are error terms with small operator norm.

Related literature

- There are three main estimation methods of this type of models.

→ The common correlated effects (CCE) estimator of Pesaran (2006);
→ The non-ordinary least squares estimator of Bai (2009);
→ The Factor-augmented regression (FAR) as in Grenaway-McGrevy et al. (2012). The latter is similar to our approach but only the factors are estimated and used in the second step estimator.

- These approaches make the following assumptions:

→ The number of factors r_N is fixed;
→ Strong factor assumption;
→ The CCE and FAR approaches assume that the regressors also have a factor structure.

The estimator

$$\widehat{\beta} \in \underset{b \in \mathbb{R}^K}{\operatorname{argmin}} \min_{\substack{\phi_1, \dots, \phi_T \in \mathbb{R}^{\widehat{r}_u},\\ l_1, \dots, l_N \in \mathbb{R}^{\widehat{r}_v}}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - X_{it}^\top b - \widehat{\lambda}_i^\top \phi_t - l_i^\top \widehat{f}_t\right)^2,$$

where $\widehat{\lambda}_i$ and $\widehat{f}_t$ are estimates of the factor loadings and the factors, respectively.

Contributions

- Give sufficient conditions on generic estimators $\widehat{\lambda}_i$ and $\widehat{f}_t$ for $\widehat{\beta}$ to be asymptotically normal.
- Propose a method to estimate the factor and the factor loadings that relies on PCA.
- Under this first step estimator, we show that $\widehat{\beta}$ can be asymptotically normal even when

(1) There are weak factors;
(2) The number of factors goes to infinity with the sample size.

- This technique based on PCA allows to relax assumptions on N and T needed for asymptotic normality of the factor augmented regression estimator.
- Our simulations show that our estimator exhibits better small sample properties than alternatives.

First-step estimator based on PCA

Let $Y_u = (Y, X_1, \dots, X_K)$ and $Y_v = (Y^\top, X_1^\top, \dots, X_K^\top)$.

(1) Estimate the number of factors with the eigenvalue ratio estimator

$$\widehat{r}_z \in \arg \max_{j \in \{1, \dots, \lfloor \sqrt{N \wedge T} \rfloor\}} \frac{\sigma_j(Y_z)}{\sigma_{j+1}(Y_z)},$$

where $\sigma_j(Y_z)$ is the j^{th} singular value of Y_z.

(2) The estimates of the factor loadings (resp. factors) are the $\widehat{r}_u$ (resp. $\widehat{r}_v$) first left singular vectors of Y_u (resp. Y_v).