The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–30

ifete: Estimation and inference of treatment effects via interactive fixed effects

Guanpeng Yan
Shandong University of Finance and Economics
Jinan, China
guanpengyan@yeah.net

Qiang Chen
Shandong University
Jinan, China
qiang2chen2@126.com

Xingyu Li
Zhejiang University
Hangzhou, China
lixyecon@zju.edu.cn

Yan Shen
Peking University
Beijing, China
yshen@nsd.pku.edu.cn

Qiankun Zhou
Louisiana State University
LA, USA
qzhou@lsu.edu

Abstract. This paper introduces the Stata command **ifete**, designed for estimation and inference of treatment effects in panel data models with interactive fixed effects. As a major extension of two-way fixed effects, interactive fixed effects provide a flexible way to construct counterfactual predictions for causal inference. The **ifete** command implements the tall-wide algorithm for efficient estimation of treatment effects (Bai and Ng 2021, *Journal of the American Statistical Association*), while using residual-based wild bootstrap for valid inference (Li et al. 2024, *Journal of Econometrics*). This command presents point estimates of treatment effects as well as confidence intervals and p-values for inference, while covariates and nonstationary trends are allowed. We illustrate the use of the **ifete** command by revisiting examples of Hong Kong's economic integration with mainland China (Hsiao et al. 2012) and the California tobacco control program (Abadie et al. 2010).

Keywords: st0001, ifete, interactive fixed effects, treatment effects, confidence intervals

Journal of Econometrics 240 (2024) 105684

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Confidence intervals of treatment effects in panel data models with interactive fixed effects☆

Xingyu Li [a], Yan Shen [a,b,c], Qiankun Zhou [d,*]

[a] China Center for Economic Research, National School of Development, Peking University, China
[b] HSBC Business School, Peking University, China
[c] Institute of Digital Finance, Peking University, China
[d] Department of Economics, Louisiana State University, United States of America

ARTICLE INFO

JEL classification:
C01
C21
C31
I18

Keywords:
Bootstrap
Confidence interval
Treatment effects
Panel data analysis
Interactive fixed effects
Nonstationarity

ABSTRACT

We augment the factor-based estimation of treatment effects proposed by Bai and Ng (2021) with easy-to-implement and nonparametric confidence intervals of treatment effects on every treated unit at every post-treatment time. The construction of confidence intervals entails a residual-based bootstrap resampling procedure. This method does not rely on any parametric assumption on the distribution of idiosyncratic errors and it is robust to weak cross-sectional and time-series dependence among idiosyncratic errors. We prove the asymptotic validity of the proposed confidence intervals as the numbers of control units and pre-treatment times go to infinity. We also extend this method to cases where the common factors and covariates (if any) are unit root processes. Monte Carlo experiments show that the proposed confidence intervals are well-behaved in finite samples and outperform confidence intervals based on normal quantiles. Empirical applications with two classical datasets add informative confidence intervals to existing point estimates of treatment effects.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 116, NO. 536, 1746–1763: Theory and Methods Special Section on Synthetic Control Methods
https://doi.org/10.1080/01621459.2021.1967163

Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data

Jushan Bai[a] and Serena Ng[b]

[a]Department of Economics, Columbia University, New York, NY; [b]Department of Economics, Columbia University and NBER, New York, NY

ABSTRACT

This article proposes an imputation procedure that uses the factors estimated from a tall block along with the re-rotated loadings estimated from a wide block to impute missing values in a panel of data. Assuming that a strong factor structure holds for the full panel of data and its sub-blocks, it is shown that the common component can be consistently estimated at four different rates of convergence without requiring regularization or iteration. An asymptotic analysis of the estimation error is obtained. An application of our analysis is estimation of counterfactuals when potential outcomes have a factor structure. We study the estimation of average and individual treatment effects on the treated and establish a normal distribution theory that can be useful for hypothesis testing.

ARTICLE HISTORY
Received November 2019
Accepted August 2021

KEYWORDS
EM algorithm;
Missing-at-random;
Nuclear-norm regularization;
Synthetic controls

1. Introduction

We introduce a new Stata command *ifete*, which uses panel data models with interactive fixed effects (IFE) for estimation and inference of treatment effects.

IFE is a major extension of the conventional two-way fixed effects (TWFE). A panel data model with IFE has a linear factor structure:

$$y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{f}_t^T \boldsymbol{\lambda}_i + e_{it}, \tag{1}$$

where $\mathbf{f}_t = (f_{1t} \cdots f_{rt})^T$ is a vector of common factors (also known as command trends), and r is a small positive integer. The potentially heterogeneous effects of common trends on each unit is captured by the vector of factor loadings $\boldsymbol{\lambda}_i = (\lambda_{i1} \cdots \lambda_{ir})^T$.

Both $\boldsymbol{\lambda}_i$ and $\mathbf{f}_t$ are random variables, which may correlate with the covariates $\mathbf{x}_{it}$, hence the term $\mathbf{f}_t^T \boldsymbol{\lambda}_i$ is known as the interactive fixed effects.

TWFE is just a special case of IFE. For example, let $\mathbf{f}_t = (1 \ \eta_t)^T$ and $\boldsymbol{\lambda}_i = (u_i \ 1)^T$, then

$$\mathbf{f}_t^T \boldsymbol{\lambda}_i = (1 \ \eta_t) \begin{pmatrix} u_i \\ 1 \end{pmatrix} = u_i + \eta_t . \tag{2}$$

While TWFE only allows one-dimensional time and individual fixed effects, IFE constitutes a major extension by accommodating multidimensional time fixed effects with heterogenous factor loadings on individuals.

IFE provides a more flexible way of modelling unobserved fixed effects than TWFE.

On one hand, a popular application of TWFE is difference-in-differences (DID) models, which identifies treatment effects by parallel trends assumption (PTA).

However, PTA may not hold in practice, as the treatment and control groups may differ systematically in both observed and unobserved aspects.

On the other hand, if we upgrade TWFE to IFE, then PTA is no longer needed. Specifically, each individual is allows to have its own trend, as factor loadings differ across individuals.

Three major approaches to estimate the IFE model has appeared in the literature, which include the common correlated effects (CCE) approach (Pesaran, 2006), the least squares principal components (LSPC) approach (Bai, 2009) and the transformed approach (Hsiao et al., 2022).

The CCE estimator uses the cross-sectional averages of the regressors and the dependent variable to proxy for the unobserved common factors.

The LSPC estimator aims to directly estimate IFE. Specifically, given $\mathbf{f}_t$, we may estimate equation (1) by least squares (LS).

On the other hand, given $\boldsymbol{\beta}$, the following equation has a pure factor structure

$$y_{it} - \mathbf{x}_{it}^T \boldsymbol{\beta} = \mathbf{f}_t^T \boldsymbol{\lambda}_i + e_{it},$$

where the common factors $\mathbf{f}_t$ can be estimated by principal components (PC).

Starting from an initial OLS estimate of $\boldsymbol{\beta}$ ignoring the common factors, the LSPC estimator iterates between least squares and principal components until convergence.

While the CCE estimator proxies for IFE and the LSPC estimator estimates IFE, the transformed approach aims to eliminate IFE by some suitable transformation.

In a broad sense, the popular synthetic control method (Abadie, Diamond and Hainmueller, 2010) and regression control method (Hsiao, Ching and Wan, 2012) fall into this category, as they both use a linear factor model as the data generating process for the untreated potential outcomes.

In this paper, we follow the LSPC approach to directly estimate IFE for treatment evaluation.

Gobillon and Magnac (2016) first present the idea of using the LSPC estimator of Bai (2009) to predict counterfactual outcomes, but no asymptotic theory is developed other than evidence from Monte Carlo simulations.

Xu (2017) proposes "generalized synthetic control method" (GSC) based on a straightforward application of the LSPC estimator for counterfactual prediction. However, as pointed out by Bai and Ng (2021), "the theoretical analysis is incomplete" in Xu (2017).

Bai and Ng (2021) provide a sophisticated tall-wide (TW) algorithm for efficient imputation of counterfactual outcomes, and establish consistency and asymptotic normality of the resulting estimator of treatment effects. However, statistical inference for unit-level treatment effects still relies on strong assumptions such as identical normal distribution of the idiosyncratic error e_{it} across both *i* and *t*.

Li, Shen and Zhou (2024) provide a valid inference procedure via residual-based wild bootstrap, which is robust to weak cross-sectional and time-series dependence among idiosyncratic errors. They also extend the inferential procedure to allow for covariates and/or nonstationary trends.

2. Estimation

Consider a panel data setting with N cross-sectional units, indexed by $i = 1, \dots, N$, observed over T time periods, indexed by $t = 1, \dots, T$.

Assume that the policy intervention occurs to units $i = N_0 + 1, \dots, N$ at time $t = T_0 + 1, \dots, T$.

Let y_{it}^1 and y_{it}^0 be the potential outcomes of unit i in period t with and without intervention, respectively.

We are interested in estimating the treatment effect for unit i in period t, defined as the difference between the potential outcomes with and without treatment:

$$\Delta_{it} = y_{it}^1 - y_{it}^0.$$

Since the untreated potential outcome y_{it}^0 is the focus of our analysis, we abuse the notation a bit by using y_{it} as a shorthand for y_{it}^0, i.e., we denote $y_{it} = y_{it}^0$ in the rest of this paper. The observed outcome for unit i in period t is denoted as $\mathcal{Y}_{it}$, which takes the form

$$\mathcal{Y}_{it} = y_{it} + d_{it}\Delta_{it}\,, \tag{2}$$

where d_{it} denotes the treatment variable with $d_{it} = 1$ if unit i is treated in period t and $d_{it} = 0$ otherwise. Moreover, denote the treated set $\mathcal{I}_1 = \{(i,t)|d_{it} = 1\}$ and untreated set $\mathcal{I}_0 = \{(i,t)|d_{it} = 0\}$, with $\mathcal{I} = \mathcal{I}_0 \cup \mathcal{I}_1$ representing the set of all observations. Hence the observed outcomes satisfy $\mathcal{Y}_{i,t} = y_{it}$ for $(i,t) \in \mathcal{I}_0$ and $\mathcal{Y}_{i,t} = y_{it} + \Delta_{it}$ for $(i,t) \in \mathcal{I}_1$.

The estimation the treatment effects Δ_{it} in Equation (2) requires knowledge of both $\mathcal{Y}_{i,t}$ and y_{it}, but only one of $\mathcal{Y}_{it}$ and y_{it} can be observed simultaneously.

A prevalent approach in the causal inference literature is to construct counterfactual outcomes $\hat{y}_{it}$ as a proxy for the unobserved y_{it} for $(i,t) \in \mathcal{I}_1$ by assuming a data generating process of y_{it}.

Assume that untreated potential outcomes y_{it} are generated by a pure factor model:

$$y_{it} = c_{it} + e_{it} = \boldsymbol{f}_t^{\top}\boldsymbol{\lambda}_i + e_{it}\,, \tag{3}$$

where $\boldsymbol{f}_t$ is a $r \times 1$ vector of unobserved common factors at period t, $\boldsymbol{\lambda}_i$ is a $r \times 1$ vector of unobserved factor loadings of unit i, and $c_{it} = \boldsymbol{f}_t^{\top}\boldsymbol{\lambda}_i$ is the common component (IFE) of unit i in period t.

Following Bai and Ng (2021), $.e_{it}$ is the idiosyncratic error term allowing for weak cross-sectional and time-series dependence.

The number of unobserved common factors r is typically unknown to researchers, but we treat it as known because it can be consistently estimated by information criteria (Bai and Ng, 2002), cross-validation (Xu, 2017; Wei and Chen, 2020) or eigenvalue ratio test (Ahn and Horenstein, 2013), which are implemented in our command.

Our objective is to obtain the prediction of counterfactual outcomes $\hat{y}_{it}$ based on equation (3), and estimate the treatment effects as $\hat{\Delta}_{it} = y_{it} - \hat{y}_{it}$ for $(i,t) \in \mathcal{I}_1$.

Given i, stacking equation (3) over $t = 1, \cdots, T$ yields

$$\mathbf{y}_i = \mathbf{F}\boldsymbol{\lambda}_i + \mathbf{e}_i, \quad i = 1, \cdots, N, \tag{4}$$

where $\mathbf{y}_i = (y_{i1} \cdots y_{iT})^T$ and $\mathbf{e}_i = (e_{i1} \cdots e_{iT})^T$.

$\mathbf{F} = (\mathbf{f}_1 \cdots \mathbf{f}_T)^T$ is the $T \times r$ factor matrix with the r common factors as its columns.

Juxtaposing equation (4) horizontally over $i = 1, \cdots, N$ yields

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\Lambda}^T + \mathbf{e}, \tag{5}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1 \cdots \boldsymbol{\lambda}_N)^T$ is the $N \times r$ factor loading matrix, $\mathbf{e} = (\mathbf{e}_1 \cdots \mathbf{e}_N)$, and $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$ is the matrix of untreated potential outcomes:

$$\mathbf{Y} = \begin{bmatrix} y_{1,1} & \cdots & y_{N_0,1} & y_{N_0+1,1} & \cdots & y_{N,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{1,T_0} & \cdots & y_{N_0,T_0} & y_{N_0+1,T_0} & \cdots & y_{N,T_0} \\ y_{1,T_0+1} & \cdots & y_{N_0,T_0+1} & y_{N_0+1,T_0+1} & \cdots & y_{N,T_0+1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{1,T} & \cdots & y_{N_0,T} & y_{N_0+1,T} & \cdots & y_{N,T} \end{bmatrix}$$

Nothing on the right side of equation (5) is observable, thus $\mathbf{F}$ and $\mathbf{\Lambda}$ are not separately identifiable.

To achieve identification, normalizations in the form of $\mathbf{F}^T\mathbf{F}/T = \mathbf{I}_r$ and diagonal $\mathbf{\Lambda}^T\mathbf{\Lambda}$ are usually imposed, where $\mathbf{I}_r$ is the $r \times r$ identity matrix.

Moreover, since $\mathbf{F}\mathbf{\Lambda}^T = \mathbf{F}\mathbf{A}\mathbf{A}^{-1}\mathbf{\Lambda}^T$ for any arbitrary $r \times r$ invertible matrix, it is easy to see that $\mathbf{F}$ and $\mathbf{\Lambda}$ are only identified up to a rotation.

The southeast part of $\boldsymbol{Y}$ (the rows $(T_0+1):T$ and columns $(N_0+1):N$) corresponding to the treatment group after the treatment is unobservable, which is known as the missing block.

For the time being, suppose the entire matrix $\boldsymbol{Y}$ is observed. Then we could obtain an estimate of the factor matrix $\boldsymbol{F}$ as the corresponding first r principal components of $\boldsymbol{Y}$ by solving an eigenvalue problem.

In practice, this is often accomplished by singular value decomposition (SVD), which enjoys better numerical accuracy and computational efficiency than eigenvalue decomposition.

The normalized data $\boldsymbol{Y}/\sqrt{NT}$ has the following singular value decomposition:

$$\frac{\mathbf{Y}}{\sqrt{NT}} = \mathbf{P}\mathbf{D}\mathbf{Q}^T = \mathbf{P}_r\mathbf{D}_r\mathbf{Q}_r^T + \mathbf{P}_{N-r}\mathbf{D}_{N-r}\mathbf{Q}_{N-r}^T$$

where $\mathbf{D}_r$ is a $r \times r$ diagonal matrix with the r largest singular values of $\boldsymbol{Y}/\sqrt{NT}$, while the $T \times r$ matrix $\mathbf{P}_r$ contains the corresponding r left singular vectors, and the $N \times r$ matrix $\mathbf{Q}_r$ contains the corresponding r right singular vectors.

Similarly, $\mathbf{D}_{N-r}$ is a $(N-r)\times(N-r)$ diagonal matrix with the rest $(N-r)$ singular values, while $\mathbf{P}_{N-r}$ and $\mathbf{Q}_{N-r}$ contain the corresponding $(N-r)$ left and right singular vectors, respectively.

The principal component estimators for $\mathbf{F}$ and $\mathbf{\Lambda}$ are given by

$$\hat{\mathbf{F}} = (\hat{\mathbf{f}}_1 \cdots \hat{\mathbf{f}}_T)^T = \sqrt{T}\mathbf{P}_r, \quad \hat{\mathbf{\Lambda}} = (\hat{\boldsymbol{\lambda}}_1 \cdots \hat{\boldsymbol{\lambda}}_N)^T = \sqrt{N}\mathbf{Q}_r\mathbf{D}_r. \tag{6}$$

Under regularity conditions, as $N, T \to \infty$, $\hat{\mathbf{F}}$ and $\hat{\mathbf{\Lambda}}$ consistently estimate $\mathbf{F}$ and $\mathbf{\Lambda}$ up to a rotation, respectively; see Lemma 1 in Bai and Ng (2021).

In particular, for any $1 \le t \le T$ and any $1 \le i \le N$, we have

$$\hat{\mathbf{f}}_t - \mathbf{H}^T \mathbf{f}_t = o_p(1), \quad \hat{\boldsymbol{\lambda}}_i - \mathbf{H}^{-1} \boldsymbol{\lambda}_i = o_p(1), \tag{7}$$

where is a $r \times r$ unknown rotation matrix $\mathbf{H}$, and $o_p(1)$ signifies a random variable converging to 0 in probability.

Since the consistent estimation of factor and factor loading matrices relies on the assumption of $N, T \to \infty$ (i.e., large panels), it is labelled as asymptotic principal components (APC) by Bai and Ng (2021).

Given the APC estimator in equation (6), we may predict $\boldsymbol{Y}$ simply by $\hat{\mathbf{Y}} = \hat{\mathbf{F}} \hat{\boldsymbol{\Lambda}}^T$.

Now back to reality. The above APC algorithm is infeasible due to the missing block in the southeast part of $\boldsymbol{Y}$.

Bai and Ng (2021) overcome this missing data problem by making efficient use of the observable part of $\boldsymbol{Y}$.

Denote the tall block (left part) of $\boldsymbol{Y}$ as $\boldsymbol{Y}_{\text{tall}}$ corresponding the control group $(i = 1, \cdots, N_0)$:

$$\boldsymbol{Y}_{\text{tall}} = \begin{bmatrix} y_{1,1} & \cdots & y_{N_0,1} \\ \vdots & \ddots & \vdots \\ y_{1,T_0} & \cdots & y_{N_0,T_0} \\ y_{1,T_0+1} & \cdots & y_{N_0,T_0+1} \\ \vdots & \ddots & \vdots \\ y_{1,T} & \cdots & y_{N_0,T} \end{bmatrix}.$$

Similarly, denote the wide block (top part) of $\boldsymbol{Y}$ as $\boldsymbol{Y}_{\text{wide}}$ corresponding to the pretreatment periods $(t = 1, \cdots, T_0)$:

$$\boldsymbol{Y}_{\text{wide}} = \begin{bmatrix} y_{1,1} & \cdots & y_{N_0,1} & y_{N_0+1,1} & \cdots & y_{N,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ y_{1,T_0} & \cdots & y_{N_0,T_0} & y_{N_0+1,T_0} & \cdots & y_{N,T_0} \end{bmatrix}.$$

Both $\boldsymbol{Y}_{\text{tall}}$ and $\boldsymbol{Y}_{\text{wide}}$ are observable, and they overlap in the northwest part of $\boldsymbol{Y}$, which is known as the balance block.

Let $(\boldsymbol{F}_{\text{tall}}, \boldsymbol{\Lambda}_{\text{tall}})$ and $(\boldsymbol{F}_{\text{wide}}, \boldsymbol{\Lambda}_{\text{wide}})$ be the factor and factor loading matrices associated with $\boldsymbol{Y}_{\text{tall}}$ and $\boldsymbol{Y}_{\text{wide}}$, respectively.

Since $\boldsymbol{Y}_{\text{tall}}$ spans all periods $1, \cdots, T$, it follows that $\boldsymbol{F}_{\text{tall}} = \boldsymbol{F}$; but $\boldsymbol{\Lambda}_{tall} \neq \boldsymbol{\Lambda}$, since $\boldsymbol{\Lambda}_{\text{tall}}$ only has N_0 columns.

Similarly, since $\boldsymbol{Y}_{\text{wide}}$ includes all units $1,\cdots,N$, we have $\boldsymbol{\Lambda}_{\text{wide}} = \boldsymbol{\Lambda}$; but $\mathbf{F}_{wide} \neq \mathbf{F}$, since $\mathbf{F}_{wide}$ only has T_0 rows.

We may apply the APC algorithm using the tall block $\boldsymbol{Y}_{\text{tall}}$ to obtain the $T \times r$ matrix $\widehat{\boldsymbol{F}}_{\text{tall}}$ and the $N_0 \times r$ matrix $\widehat{\boldsymbol{\Lambda}}_{\text{tall}}$. Similarly, we can apply the APC algorithm using the wide block $\boldsymbol{Y}_{\text{wide}}$ to obtain the $T_0 \times r$ matrix $\widehat{\boldsymbol{F}}_{\text{wide}}$ and the $N \times r$ matrix $\widehat{\boldsymbol{\Lambda}}_{\text{wide}}$.

A naïve approach is to predict $\boldsymbol{Y}$ directly by $\widehat{\boldsymbol{F}}_{\text{tall}}\widehat{\boldsymbol{\Lambda}}_{\text{wide}}^{\mathsf{T}}$.

The problem is that they are estimated from different blocks of data, thus their rotation matrices do not match in general.

Fortunately, the data in the balance block contain information shared by both the tall and wide blocks and can be to re-rotate the data.

Specifically, for any unit i in the balance block, by equation (7) we have

$$\boldsymbol{\lambda}_i = \mathbf{H}_{tall}\hat{\boldsymbol{\lambda}}_{tall,i} + o_p(1) = \mathbf{H}_{wide}\hat{\boldsymbol{\lambda}}_{wide,i} + o_p(1),$$

which implies that for any unit i in the balance block,

$$\hat{\boldsymbol{\lambda}}_{tall,i} = \left(\mathbf{H}_{tall}^{-1}\mathbf{H}_{wide}\right)\hat{\boldsymbol{\lambda}}_{wide,i} + o_p(1). \tag{8}$$

Define a new $r \times r$ nonsingular rotation matrix

$$\mathbf{H}_{miss} = \mathbf{H}_{tall}^{-1}\mathbf{H}_{wide},$$

which can be used for the desirable re-rotation. While $\mathbf{H}_{miss}$ is still unknown, equation (8) suggests simple way to estimate it by linear regression.

Denote $\hat{\boldsymbol{\Lambda}}_{\text{wide},0}$ as the $N_0 \times r$ sub-matrix of $\hat{\boldsymbol{\Lambda}}_{\text{wide}}$ associated with the balance block.

Taking a transpose of equation (8) and stacking over $1,\cdots,N_0$ yields $\hat{\mathbf{\Lambda}}_{tall} = \hat{\mathbf{\Lambda}}_{wide,0}\mathbf{H}_{miss}^{T} + o_p(1)$. Thus, $\mathbf{H}_{miss}$ can be estimated by regressing $\widehat{\mathbf{\Lambda}}_{\text{tall}}$ on $\widehat{\mathbf{\Lambda}}_{\text{wide},0}$:

$$\widehat{\boldsymbol{H}}_{\text{miss}} = \widehat{\mathbf{\Lambda}}_{\text{tall}}^{\mathsf{T}}\widehat{\mathbf{\Lambda}}_{\text{wide},0}\left(\widehat{\mathbf{\Lambda}}_{\text{wide},0}^{\mathsf{T}}\widehat{\mathbf{\Lambda}}_{\text{wide},0}\right)^{-1}.$$

Consequently, we may predict $\boldsymbol{Y}$ by $\hat{\mathbf{Y}} = \widehat{\boldsymbol{F}}_{\text{tall}}\widehat{\boldsymbol{H}}_{\text{miss}}\widehat{\mathbf{\Lambda}}_{\text{wide}}^{\mathsf{T}}$, from which we may obtain the prediction $\hat{y}_{it} = \hat{c}_{it}$ as the (t, i) element of $\hat{\mathbf{Y}}$.

The residual is computed as the difference between observed outcome $y_{i,t}$ and the estimated common component $\hat{c}_{it}$, denoted as $\hat{e}_{it} = y_{it} - \hat{c}_{it}$.

In particular, for $(i,t) \in \mathcal{I}_1$, the estimated treatment effect is equal to the residual, i.e., $\widehat{\Delta}_{it} = \hat{e}_{it}$. This completes the point estimation of the treatment effect for $(i,t) \in \mathcal{I}_1$ by the tall-wide (TW) algorithm.

3. Inference

Under regularity conditions, for $(i,t) \in \mathcal{I}_1$, Bai and Ng (2021, Proposition 1) show that, $\delta_{NT}\left(\hat{c}_{it} - c_{it}\right)$ is asymptotically normal as $N_0, T_0 \to \infty$, where $\delta_{NT} = \min\left\{\sqrt{N_0}, \sqrt{T_0}\right\}$.

Thus, the rate of convergence is the smaller of $\sqrt{N_0}$ and $\sqrt{T_0}$.

The assumed regularity conditions (Assumptions A-D in Bai and Ng (2021)) include

(1) regularity conditions for factors, factor loadings and idiosyncratic errors such that the factors are strong and can be separated from the idiosyncratic errors for the whole sample and each sub-blocks (i.e., the balance, tall, wide and missing blocks);

(2) weak cross-sectional and time-series dependence for the errors;

(3) the number of factors r is not too large such that the order conditions $r < \left(TN_0/(T+N_0)\right)$ and $r < \left(T_0N/(T_0+N)\right)$ hold;

(4) both N_0 and T_0 grow faster than $\sqrt{N}$ and $\sqrt{T}$ such that the rate conditions $\left(\sqrt{N}/\min\{N_0,T_0\}\right) \to 0$ and $\left(\sqrt{T}/\min\{N_0,T_0\}\right) \to 0$ hold as $N_0, N, T_0, T \to \infty$; and (5) relevant central limit theorems.

However, constructing a confidence interval for Δ_{it} for $(i,t) \in \mathcal{I}_1$ is more complicated than this result of asymptotic normality. Recall $\mathcal{Y}_{it} = y_{it} + \Delta_{it}$ and $\widehat{\Delta}_{it} = \mathcal{Y}_{it} - \hat{c}_{it}$, we have

$$\widehat{\Delta}_{it} - \Delta_{it} = y_{it} - \hat{c}_{it} = e_{it} - (\hat{c}_{it} - c_{it}) .$$

Since the estimation error $(\hat{c}_{it} - c_{it})$ converges in probability to 0, the right hand side of the above equation it is dominated by the error e_{it}.

Thus, the distribution of $\widehat{\Delta}_{it} - \Delta_{it}$ depends on the distribution of e_{it}.

To make it worse, since we are interested in the cell-level treatment effect Δ_{it}, we cannot average e_{it} over $i = N_0 + 1, \cdots, N$ or $t = T_0 + 1, \cdots, T$.

To get around this issue, Bai and Ng (2021) makes a strong assumption such that e_{it} follows a normal distribution, and is identically distributed across both i and t.

This assumption is likely too restrictive in practice.

To relax these strong assumptions, Li, Shen and Zhou (2024) propose a robust inferential method using residual-based wild bootstrap.

Denote $\text{se}(\widehat{\Delta}_{it}-\Delta_{it})$ as the standard error of $\widehat{\Delta}_{it}-\Delta_{it}$. Let q_1 and q_2 be the $\alpha/2$ and $1-\alpha/2$ quantiles of the t statistic $(\widehat{\Delta}_{it}-\Delta_{it})/\text{se}(\widehat{\Delta}_{it}-\Delta_{it})$. It follows that

$$
\begin{aligned}
1-\alpha &= \mathrm{P}\left(q_1 \le \frac{\hat{\Delta}_{it}-\Delta_{it}}{se(\hat{\Delta}_{it}-\Delta_{it})} \le q_2 \right) \\
&= \mathrm{P}\left(\hat{\Delta}_{it} + q_1 se(\hat{\Delta}_{it}-\Delta_{it}) \le \Delta_{it} \le \hat{\Delta}_{it} + q_2 se(\hat{\Delta}_{it}-\Delta_{it}) \right).
\end{aligned}
$$

Hence, to construct a confidence interval for Δ_{it}, we need to compute $\text{se}(\widehat{\Delta}_{it}-\Delta_{it})$ and estimate q_1 and q_2, which we discuss in turn.

3.1 Estimation of standard errors

Since $\widehat{\Delta}_{it}-\Delta_{it} = e_{it} - (\hat{c}_{it}-c_{it})$, the variance of $(\widehat{\Delta}_{it}-\Delta_{it})$ has two components consisting of $Var(\hat{c}_{it}-c_{it})$ and $Var(e_{it})$, where the covariance is asymptotically

negligible and ignored. For $Var(e_{it})$, we may estimate it by $\hat{\sigma}_i^2 = \frac{1}{T_0}\sum_{s=1}^{T_0}\hat{e}_{is}^2$. For $Var(\hat{c}_{it} - c_{it})$, recall that $\hat{c}_{it}$ is the (t, i) element of $\widehat{\mathbf{Y}} = \widehat{\boldsymbol{F}}_{\text{tall}}\widehat{\boldsymbol{H}}_{\text{miss}}\widehat{\boldsymbol{\Lambda}}_{\text{wide}}^{\mathsf{T}}$, which has a close-form expression. Therefore, the variance of $(\hat{c}_{it} - c_{it})$ also has an analytic formula, albeit complicated. Using the heteroskedasticity and autocorrelation consistent (HAC) variance estimator by Newey and West (1987) to accommodate weak time-series dependence, the variance of $(\hat{c}_{it} - c_{it})$ can be estimated as by

$$\hat{v}_{it} = T_0\widehat{\boldsymbol{f}}_{\text{tall},t}^{\mathsf{T}}\left(\widehat{\boldsymbol{F}}_{\text{wide}}^{\mathsf{T}}\widehat{\boldsymbol{F}}_{\text{wide}}\right)^{-1}\widehat{\boldsymbol{\Phi}}_{0i}\left(\widehat{\boldsymbol{F}}_{\text{wide}}^{\mathsf{T}}\widehat{\boldsymbol{F}}_{\text{wide}}\right)^{-1}\widehat{\boldsymbol{f}}_{\text{tall},t} + N_0\widehat{\boldsymbol{\lambda}}_{\text{wide},i}^{\mathsf{T}}\left(\widehat{\boldsymbol{\Lambda}}_{\text{tall}}^{\mathsf{T}}\widehat{\boldsymbol{\Lambda}}_{\text{tall}}\right)^{-1}\widehat{\boldsymbol{\Gamma}}_{0t}\left(\widehat{\boldsymbol{\Lambda}}_{\text{tall}}^{\mathsf{T}}\widehat{\boldsymbol{\Lambda}}_{\text{tall}}\right)^{-1}\widehat{\boldsymbol{\lambda}}_{\text{wide},i},$$

where $\widehat{\boldsymbol{\Gamma}}_{0t} = \frac{1}{N_0}\sum_{j=1}^{N_0}\hat{e}_{jt}^2\,\widehat{\boldsymbol{\lambda}}_{\text{tall},j}\widehat{\boldsymbol{\lambda}}_{\text{tall},j}^{\mathsf{T}}$, $\boldsymbol{L}_{0,k,i} = \frac{1}{T_0}\sum_{s=k+1}^{T_0}\widehat{\boldsymbol{f}}_{\text{wide},s}\,\hat{e}_{is}\hat{e}_{i,s-k}\widehat{\boldsymbol{f}}_{\text{wide},s-k}^{\mathsf{T}}$, $\widehat{\boldsymbol{\Phi}}_{0i} = \boldsymbol{L}_{0,0,i} + \sum_{k=1}^{K}\left(1 - \frac{k}{K+1}\right)\left(\boldsymbol{L}_{0,k,i} + \boldsymbol{L}_{0,k,i}^{\mathsf{T}}\right)$ with the truncation parameter $K = \left\lfloor T_0^{1/5} \right\rfloor$, i.e., the largest integer smaller than or equal to $T_0^{1/5}$.

Li, Shen and Zhou (2024) considers the general case of K with $K \to \infty$ and $K/T_0^{1/4} \to 0$ as $T_0 \to \infty$. For practical implementation, we set $K = \left\lfloor T_0^{1/5} \right\rfloor$. Thus, the standard error of $(\widehat{\Delta}_{it} - \Delta_{it})$ is given by

$$\mathrm{se}\left(\widehat{\Delta}_{it} - \Delta_{it}\right) = \sqrt{\hat{v}_{it} + \hat{\sigma}_i^2}\,.$$

Moreover, the t statistic $(\widehat{\Delta}_{it} - \Delta_{it})/\mathrm{se}(\widehat{\Delta}_{it} - \Delta_{it})$ can be rewritten as

$$s_{it} = \frac{\hat{c}_{it} - y_{it}}{\sqrt{\hat{v}_{it} + \hat{\sigma}_i^2}}\,.$$

3.2 Estimation of empirical quantiles

We adopt residual-based bootstrap to construct the empirical distribution of s_{it}, which approximates the true distribution without imposing strong parametric assumptions. Specifically, we take the following steps to implement the residual-based bootstrap:

1. For treated observations $(i,t) \in \mathcal{I}_1$, use simple residual bootstrap by drawing the bootstrapped error e_{it}^* from the empirical distribution of the pretreatment residuals of the same unit $\{\hat{e}_{is}: s = 1, \dots, T_0\}$. Specifically, sample e_{it}^* independently from the discrete uniform distribution of $\{\hat{e}_{i1} - \overline{\hat{e}}_i, \dots, \hat{e}_{i,T_0} - \overline{\hat{e}}_i\}$, where $\overline{\hat{e}}_i = \frac{1}{T_0}\sum_{s=1}^{T_0} \hat{e}_{is}$ is the sample mean of the pretreatment residuals for unit i.

2. For untreated observations $(i,t) \in \mathcal{I}_0$, use wild bootstrap to capture the potential cross-sectional and serial correlations among the original error terms e_{it}. Specifically, we generate the bootstrapped error e_{it}^* by multiplying the residual $\hat{e}_{it}$ and a random multiplier u_{Tt}, i.e., $e_{it}^* = u_{Tt}\hat{e}_{it}$, where the subscript T of the random multiplier indicates its distribution depending on T. Following Li, Shen and Zhou (2024), the random multipliers are randomly drawn from a T-dimensional multivariate normal distribution with mean 0 and variance matrix $\Sigma_{u,T}$, where the (t,s) entry of Σ_{uT} is

given by $a\big((t-s)/\lceil T^{1/3}\rceil\big)$ for $t,s\in\{1,\dots,T\}$, $a(\cdot)$ is the Bartlett kernel with $a(z)=1-|z|$ for all $z\in[-1,1]$ and zero otherwise, and $\lceil T^{1/3}\rceil$ is the smallest integer greater than or equal to $T^{1/3}$).

3. For all observations $(i,t)\in\mathcal{I}$, construct the bootstrapped outcomes y_{it}^{*} by adding the bootstrapped errors to the estimated common components from the original sample, i.e., $y_{it}^{*}=\hat{c}_{it}+e_{it}^{*}$.

4. Apply the tall-wide algorithm described in Section 2 to the bootstrapped sample $\boldsymbol{Y}^{*}$ consisting of y_{it}^{*} for all $(i,t)\in\mathcal{I}$ to obtain the bootstrapped estimates $\hat{c}_{it}^{*}$, $\hat{v}_{it}^{*}$ and $\hat{\sigma}_{i}^{*2}$. The bootstrapped version of the t statistic s_{it} is computed as

$$s_{it}^{*}(b)=\frac{\hat{c}_{it}^{*}-y_{it}^{*}}{\sqrt{\hat{v}_{it}^{*}+\hat{\sigma}_{i}^{*2}}}$$

After executing the above bootstrap procedure for $b = 1, \dots, B$ times, we obtain B bootstrapped statistics denoted by $s_{it}^*(1), \dots, s_{it}^*(B)$ for every $(i,t) \in \mathcal{I}_1$. Let $q_{\alpha/2,it}$ and $q_{1-\alpha/2,it}$ be the $\alpha/2$ and $(1-\alpha/2)$ empirical quantiles of the set $\{s_{it}^*(1), \dots, s_{it}^*(B)\}$, respectively. The equal tailed $(1-\alpha)$ confidence interval of Δ_{it} is estimated by

$$\mathrm{EQ}_{1-\alpha,it} = \left[\widehat{\Delta}_{it} + q_{\frac{\alpha}{2},it}\sqrt{\hat{v}_{it} + \hat{\sigma}_i^2},\ \widehat{\Delta}_{it} + q_{1-\left(\frac{\alpha}{2}\right),it}\sqrt{\hat{v}_{it} + \hat{\sigma}_i^2}\right]$$

Moreover, let $p_{1-\alpha,it}$ be the $(1-\alpha)$ empirical quantile of the set $\{|s_{it}^*(1)|, \dots, |s_{it}^*(B)|\}$. The symmetric $(1-\alpha)$ confidence interval of Δ_{it} can be estimated by

$$\mathrm{SY}_{1-\alpha,it} = \left[\widehat{\Delta}_{it} - p_{1-\alpha,it}\sqrt{\hat{v}_{it} + \hat{\sigma}_i^2},\ \widehat{\Delta}_{it} + p_{1-\alpha,it}\sqrt{\hat{v}_{it} + \hat{\sigma}_i^2}\right]$$

Under regularity conditions, Li, Shen and Zhou (2024, Theorem 3.1) show that the constructed confidence intervals are asymptotically valid in the sense that the coverage

probabilities converge to the nominal level $(1-\alpha)$ as $(N_0, T_0) \to \infty$. Formally, the confidence intervals satisfy the properties $\lim_{N_0,T_0\to\infty} \mathrm{P}\left(\Delta_{it} \in \mathrm{EQ}_{1-\alpha,it}\right) = \lim_{N_0,T_0\to\infty} \mathrm{P}\left(\Delta_{it} \in \mathrm{SY}_{1-\alpha,it}\right) = 1-\alpha$. These regularity conditions include Assumptions A-D in Bai and Ng (2021), and additional assumptions introduced by Li, Shen and Zhou (2024) such as (1) the error process $\left\{e_{it}\right\}_{t=1}^{\infty}$ is strictly stationary and ergodic; (2) the cumulative distribution function of e_{it} is everywhere continuous; and (3) N_0, N, T_0, T grow at the same rate. Therefore, the residual-based bootstrap procedure proposed by Li, Shen and Zhou (2024) provides asymptotically valid inference for the TW algorithm of Bai and Ng (2021) under weak conditions that is robust to weak spatial and temporal dependence.

Furthermore, we may conduct hypothesis testing using the bootstrapped distribution of s_{it}. Specifically, for the null hypothesis $H_0: \Delta_{it} = 0$, assuming the distribution is symmetric, we may compute the two-sided *p*-value by

$$p_{it} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(|s_{it}^*(b)| \geq |s_{it}|\right),$$

where $\mathbf{1}(\cdot)$ is the indicator function, which equals 1 if the expression inside is true and 0 otherwise.

In the general case where the distribution is asymmetric, the corresponding two-sided the *p*-value can computed by

$$p_{it}^* = 2\min\left\{\frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(s_{it}^*(b) \geq s_{it}\right), \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\left(s_{it}^*(b) \leq s_{it}\right)\right\},$$

which is the default *p*-value reported by the *ifete* command, since it is robust to asymmetric distribution.

4. Extensions

In this section, we follow Bai and Ng (2021) and Li, Shen and Zhou (2024) to extend the TW algorithm and its inference to the cases with covariates and/or nonstationary trends.

4.1 The case with covariates

Exogenous covariates may be added to the pure factor model (3) to improve the efficiency of estimation. Following Bai (2009), we assume that the untreated potential outcome y_{it} is generated by

$$y_{it} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + c_{it} + e_{it} = \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta} + \boldsymbol{f}_{t}^{\top}\boldsymbol{\lambda}_{\boldsymbol{i}} + e_{it},$$

where $\boldsymbol{x}_{it}$ is a $p \times 1$ vector of time-varying exogenous covariates for unit i in period t, and $\boldsymbol{\beta}$ is the corresponding $p \times 1$ vector of coefficients. Covariates $\boldsymbol{x}_{it}$ are exogenous such that they uncorrelated with the error e_{it}.

The TW algorithm can be similarly applied to the case with covariates, with an additional preprocessing step to estimate $\boldsymbol{\beta}$ by the LSPC estimator discussed in Section 1.

Specifically, let $\widehat{\boldsymbol{\beta}}$ be the LSPC estimator using the tall block corresponding to the control group. Define $\widehat{\boldsymbol{R}}$ as the residualized outcome matrix consisting of $\hat{r}_{it} = y_{it} - \boldsymbol{x}_{it}^{\top}\widehat{\boldsymbol{\beta}}$. Then we can apply the TW algorithm without covariates in Section 2 using $\widehat{\boldsymbol{R}}$ in place of $\widehat{\boldsymbol{Y}}$.

Under additional regularity conditions about $\boldsymbol{x}_{it}$, Bai (2009) shows the LSPC estimator $\widehat{\boldsymbol{\beta}}$ converges to $\boldsymbol{\beta}$ at the rate of $1/\sqrt{N_0 T_0}$, which is faster than the convergent rates of estimated factors and factor loadings. Thus, the estimation error of $\boldsymbol{\beta}$ is negligible in large sample, and the asymptotic results in Bai and Ng (2021) and Li, Shen and Zhou (2024) remain valid.

4.2 The case with nonstationary trends

In Sections 2 and 3, the common factors and covariates (if any) are assumed to be stationary $I(0)$ processes. Li, Shen and Zhou (2024) extend the TW algorithm and its inference to handle nonstationary factors and/or covariates driven by $I(1)$ unit root processes. In particular, assume that the common factors $\boldsymbol{f}_t$ follow a vector-valued integrated process:

$$\mathbf{f}_t = \mathbf{f}_{t-1} + \boldsymbol{\eta}_t, \tag{9}$$

where $\boldsymbol{\eta}_t$ is a $r \times 1$ vector of zero-mean stationary process driving the stochastic trends. Recall that the APC for stationary data use normalization $\boldsymbol{F}^\top \boldsymbol{F}/T = \boldsymbol{I}_r$ for identification. However, when the common factors $\boldsymbol{f}_t$ are integrated, they grow at faster rates such that the appropriate normalization is $\boldsymbol{F}^\top \boldsymbol{F}/T^2 = \boldsymbol{I}_r$. Therefore, we should normalize the outcome matrix by $\boldsymbol{Y}/\sqrt{NT^2}$ instead of $\boldsymbol{Y}/\sqrt{NT}$. Other than that, the APC for nonstationary factors are the same as the APC for stationary data.

Next, we consider the case with nonstationary common factors and covariates. In additional to the common factors $\boldsymbol{f}_t$ following equation (9), the covariates $\boldsymbol{x}_{it}$ are also generated by a vector-valued integrated process:

$$\boldsymbol{x}_{it} = \boldsymbol{x}_{i,t-1} + \boldsymbol{\varepsilon}_{it}$$

where $\boldsymbol{\varepsilon}_{it}$ is a $p \times 1$ vector of zero-mean stationary process.

To estimate the common factors, factor loadings and coefficients on covariates, Li, Shen and Zhou (2024) modify the tall-white algorithm by replacing the LSPC estimator with the continuously-updated (CUP) estimator by Bai, Kao and Ng (2009). The CUP estimator may be viewed as a generalization of LSPC to panel cointegration with global stochastic trends, and readers are referred to Bai, Kao and Ng (2009) for details.

4.3 The number of factors

rmethod(*criterion***)** specifies the method used to determine the number of factors. *criterion* should be one of **loo**, **cv**, **cv(***K J***)**, **pc1**, **pc2**, **pc3**, **ic1**, **ic2**, **ic3**, **er**, or **gr**. The default is **rmethod(loo)**.

loo (default) specifies leave-one-out cross validation (Xu, 2017), which selects the number of factors by minimizing the mean squared prediction error.

cv or **cv(***K J***)** specifies the twice *K*-fold cross validation (Wei and Chen, 2020) for selecting the number of factors. The tall block is split into *K* folds in the time dimension and *J* folds in the cross-sectional dimension. For each candidate number of factors, the method computes the summed squared prediction error over all held-out blocks and selects the value that minimizes this criterion. If **cv** is specified without arguments, the default is *K* = *T* and *J* = *N0*. **cv(***K J***)** may also be specified, where 1 <= *K* <= *T* and 1 <= *J* <= *N0*.

pc1, **pc2**, **pc3**, **ic1**, **ic2** or **ic3** specifies Bai and Ng's information criteria (Bai and Ng, 2002), which minimize the sum of squared residuals in addition to a penalty term that increases with the number of factors.

er or **gr** specifies Ahn and Horenstein's eigenvalue-based criteria (Ahn and Horenstein, 2013), which determine the number of factors by examining the ratio of adjacent eigenvalues, with **er** relying on the eigenvalue ratio and **gr** on the growth ratio.

rmin(#) specifies the minimum number of factors considered by **rmethod(***criterion***)**, which defaults to 1. # must be a positive integer less than or equal to the maximum number of factors specified by **rmax(#)**.

rmax(#) specifies the maximum number of factors considered by **rmethod(***criterion***)**, which defaults to 10. # must be a positive integer that lies within the interval [1, min(*T0* - 1, *N0* - 1)], where *T0* denotes the number of pretreatment periods when all units are untreated, and *N0* denotes the number of control units. It is recommended to further increase **rmax(#)** if it turns out to be the same as the optimal number of factors chosen.

5. The ifete command

Title

ifete — estimation and inference of treatment effects in panel data with interactive fixed effects

Syntax

ifete *depvar* [*indepvars*], **treatvar**(*treatvarname*) [*options*]

options	Description
Model	
r(#)	number of factors
iterate(#)	maximum number of iterations
tolerance(*tol*)	tolerance criterion for convergence
trend({**0**|**1**)}	indicator of nonstationary trend
bootstrap(#)	number of bootstrap samples
seed(*int*)	seed used by the random number generator
Optimization	
rmethod(*criterion*)	method for estimating the number of factors
rmin(#)	minimum number of factors
rmax(#)	maximum number of factors
Reporting	
citype({**eq**|**sy**)}	type of reported confidence interval
frame(*framename*)	create a Stata frame storing generated variables.
nofigure	do not display figures
savegraph(*prefix*, [**asis replace**])	save all produced graphs to the current path.

6. Examples

To begin with, the Stata command `ifete` can be install from the SSC:

```
. ssc install ifete, all replace
```

The option "`all`" specifies downloading two example datasets (growth2.dta and smoking2.dta) attached to the `ifete` command.

The option "`replace`" instructs replacement of previous version of the ifete command if installed.

6.1 Example 1: Economic integration between Hong Kong and Chinese Mainland (Hsiao et al. 2012)

In this example, we revisit the case in Hsiao et al. (2012), which evaluates the effects of political and economic integration between Hong Kong and Chinese mainland on the economy of Hong Kong.

The panel dataset growth2.dta attached to the ifete command includes the following variables across Hong Kong and other 24 countries or regions from 1993Q1 to 2008Q1: the outcome variable gdp representing quarterly real GDP growth rates, the treatment variable pi indicating political integration, and the treatment variable ei indicating economic integration.

After loading the dataset growth2.dta, and declaring it a panel dataset by “xtset region time”, we use the third-party command “panelview” (installed from SSC) to examine and visualize the treatment status of political integration captured by the treatment variable ei:

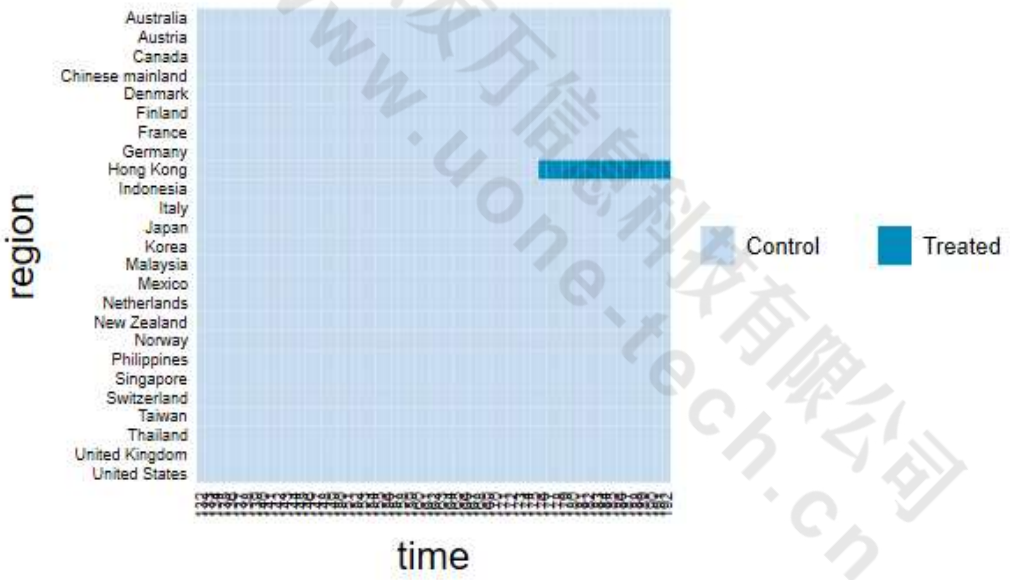
```
. sysuse growth2, clear
. xtset region time

Panel variable: region (strongly balanced)
 Time variable: time, 1993q1 to 2008q1
         Delta: 1 quarter
```

```
. panelview gdp ei, i(region) t(time) type(treat)
```

#	Variable	# Missing	% Missing
1	gdp	0	0.0
2	ei	0	0.0

Missing for how many variables?	Freq.	Percent	Cum.
0	1,525	100.00	100.00
Total	1,525	100.00	

Note: White cells represent missing values/observations in data.



```
. ifete gdp, treatvar(ei) frame(growth_ei)
```

The option "`treatvar(ei)`" designates `ei` as the treatment variable.

The option "`frame(growth_ei)`" instructs the creation of a Stata frame named `growth_ei`, which contains variables generated by ifete, including counterfactual outcomes, treatment effects, confidence intervals, and *p*-values.

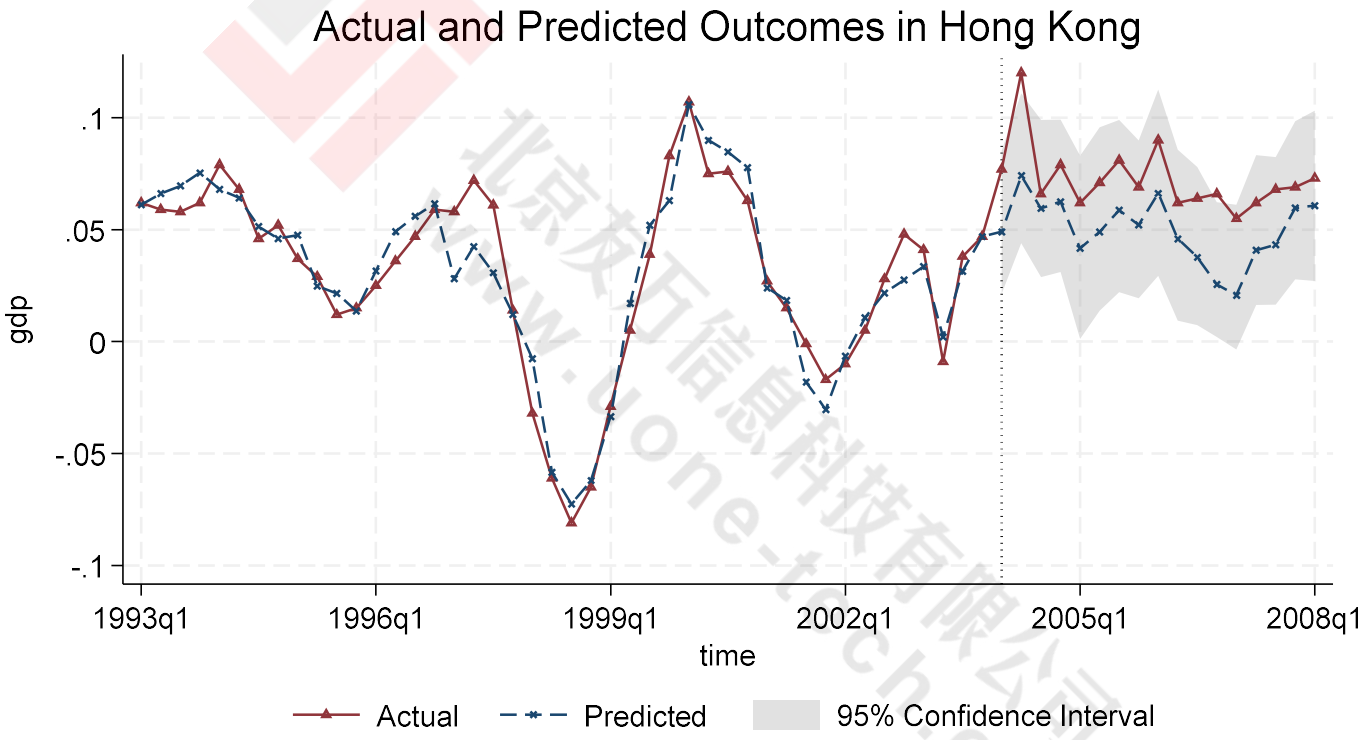
```
Estimation results based on the data from control units and the pretreatment data
>  of treated units:
------------------------------------------------------------------------------
 Number of Control Units   =         24    Size of Tall Block =      (61, 24)
 Number of Pretreatment Periods  =   44    Size of Wide Block =      (44, 25)
 Number of Covariates      =          0    Number of Factors  =             8
 Root Mean Squared Error   =      0.011    Pretreatment R²    =       0.91831
------------------------------------------------------------------------------

Note: The number of factors is estimated using the leave-one-out cross
      validation (loo) (Xu, 2017), with the number of factors selected from the
      range [1, 10].
```

Estimation and prediction results during the posttreatment periods in Hong Kong,
> with equal-tailed confidence intervals:

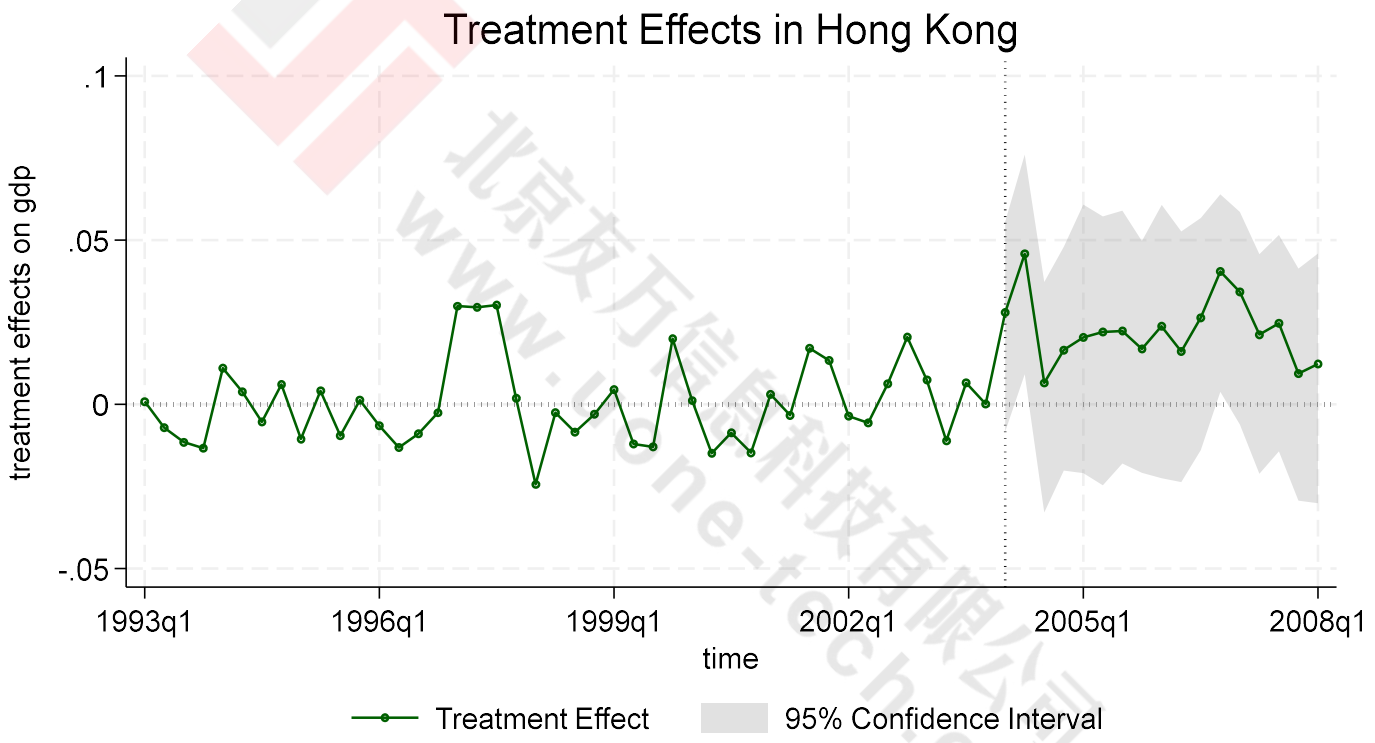
Time	Actual Outcome	Predicted Outcome	[95% Confidence Interval]	
2004q1	0.0770	0.0490	0.0213	0.0853
2004q2	0.1200	0.0741	0.0440	0.1109
2004q3	0.0660	0.0595	0.0288	0.0990
2004q4	0.0790	0.0625	0.0311	0.0992
2005q1	0.0620	0.0416	0.0012	0.0830
2005q2	0.0710	0.0490	0.0138	0.0956
2005q3	0.0810	0.0587	0.0221	0.0990
2005q4	0.0690	0.0521	0.0193	0.0899
2006q1	0.0900	0.0662	0.0293	0.1126
2006q2	0.0620	0.0459	0.0094	0.0857
2006q3	0.0640	0.0376	0.0073	0.0779
2006q4	0.0660	0.0255	0.0021	0.0622
2007q1	0.0550	0.0207	-0.0036	0.0611
2007q2	0.0620	0.0408	0.0163	0.0832
2007q3	0.0680	0.0433	0.0165	0.0823
2007q4	0.0690	0.0596	0.0277	0.0984
2008q1	0.0730	0.0607	0.0271	0.1031

Time	Treatment Effect	*p*-value	[95% Confidence Interval]	
2004q1	0.0280	0.116	-0.0083	0.0557
2004q2	0.0459**	0.016	0.0091	0.0760
2004q3	0.0065	0.540	-0.0330	0.0372
2004q4	0.0165	0.248	-0.0202	0.0479
2005q1	0.0204	0.212	-0.0210	0.0608
2005q2	0.0220	0.212	-0.0246	0.0572
2005q3	0.0223	0.228	-0.0180	0.0589
2005q4	0.0169	0.268	-0.0209	0.0497
2006q1	0.0238	0.192	-0.0226	0.0607
2006q2	0.0161	0.320	-0.0237	0.0526
2006q3	0.0264	0.192	-0.0139	0.0567
2006q4	0.0405**	0.032	0.0038	0.0639
2007q1	0.0343*	0.092	-0.0061	0.0586
2007q2	0.0212	0.260	-0.0212	0.0457
2007q3	0.0247	0.232	-0.0143	0.0515
2007q4	0.0094	0.480	-0.0294	0.0413
2008q1	0.0123	0.424	-0.0301	0.0459
Mean	0.0228			

Note: (1) The average treatment effect over the posttreatment period is 0.0228.
(2) ***, **, and * denote statistical significance of treatment effect at the 1%, 5%, and 10% level, respectively.



Note: The vertical dotted line indicates the start of the treatment period (2004q1).



Note: The vertical dotted line indicates the start of the treatment period (2004q1).

6.2 Example 2: California tobacco control program (Abadie et al. 2010)

To demonstrate the use of *ifete* of in a model with covariates and nonstationary trend, we revisit the case in Abadie et al. (2010), which evaluates the effects of California Tobacco Control Program (CTCP) on per capita cigarettes consumption in California.

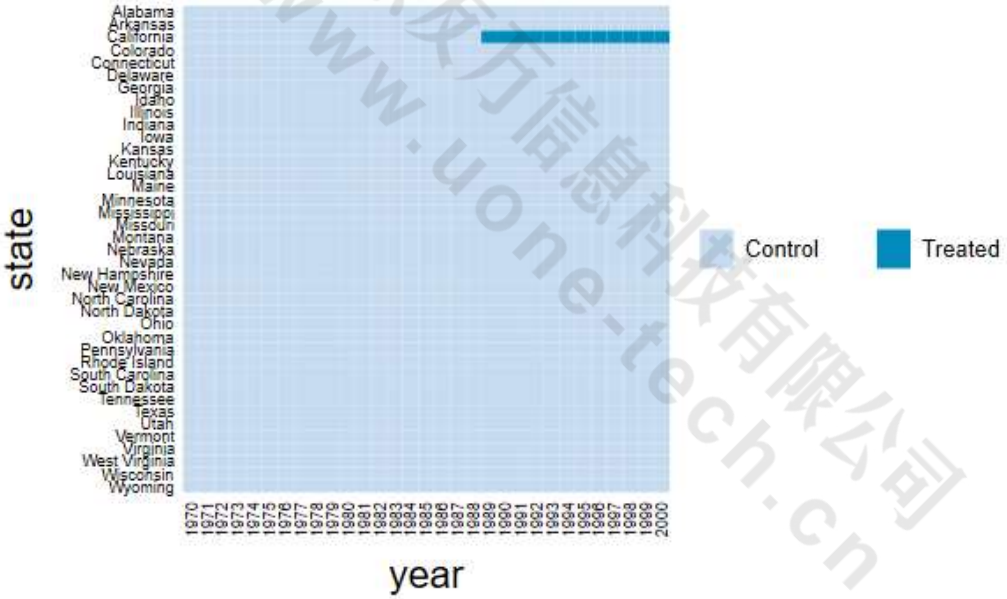
The dataset smoking2.dta, which is attached to the *ifete* command and constructed by Hsiao and Zhou (2019) for reevaluating the impact of CTCP, contains the following variables for 39 U.S. states from 1970 to 2000: the outcome variable `cigsale` (cigarette sales per capita in packs), the covariates `lnincome` (logged per capita state personal income), `eduattain` (percentage of the population aged 25 years and over with a college degree), `poverty` (poverty rates), and the treatment variable `ctcp` (indicator of CTCP).

```
. sysuse smoking2, clear

. xtset state year


Panel variable: state (strongly balanced)
 Time variable: year, 1970 to 2000
         Delta: 1 unit



. panelview cigsale ctcp, i(state) t(year) type(treat)
```

#	Variable	# Missing	% Missing
1	cigsale	0	0.0
2	ctcp	0	0.0

Missing for how many variables?	Freq.	Percent	Cum.
0	1,209	100.00	100.00
Total	1,209	100.00	

Note: White cells represent missing values/observations in data.



Prior to proceeding with the estimation, it is necessary to examine whether the tall and wide matrices of the outcome variable and covariates contain unit roots.

Following the approach of Li et al. (2024), we implement the `xtunitroot llc` command on the variables cigsale, lnincome, eduattain, and poverty to perform the Levin-Lin-Chu test (Levin et al. 2002).

The panel unit root test results reveal that both the tall block of cigsale and the wide blocks of cigsale and eduattain have insignificant *p*-values at the 5% level. Under the null hypothesis of panels containing unit roots, this suggests the presence of unit roots in the variables cigsale and eduattain.

We employ the `trend(1)` option to specify a model with a nonstationary trend and the `rmethod(er)` option to estimate the number of factors by the eigenvalue ratio criterion proposed by Ahn and Horensteain (2013).

For this example, using other criteria would overstate the number of factors, resulting in overfitting and unreliable results.

```
. ifete cigsale lnincome eduattain poverty,
treatvar(ctcp) trend(1) rmethod(er)
```

Estimation results based on the data from control units and the pretreatment data
> of treated units:

Number of Control Units	= 38	Size of Tall Block	= (31, 38)
Number of Pretreatment Periods	= 19	Size of Wide Block	= (19, 39)
Number of Covariates	= 3	Number of Factors	= 1
Root Mean Squared Error	= 9.594	Pretreatment R^2	= 0.91214

cigsale	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnincome	41.82516	4.74327	8.82	0.000	32.51833	51.13199
eduattain	-2.553667	.1742708	-14.65	0.000	-2.895606	-2.211728
poverty	-.290102	.1527914	-1.90	0.058	-.589896	.0096919
_cons	-332.3309	51.89091	-6.40	0.000	-434.1468	-230.5151

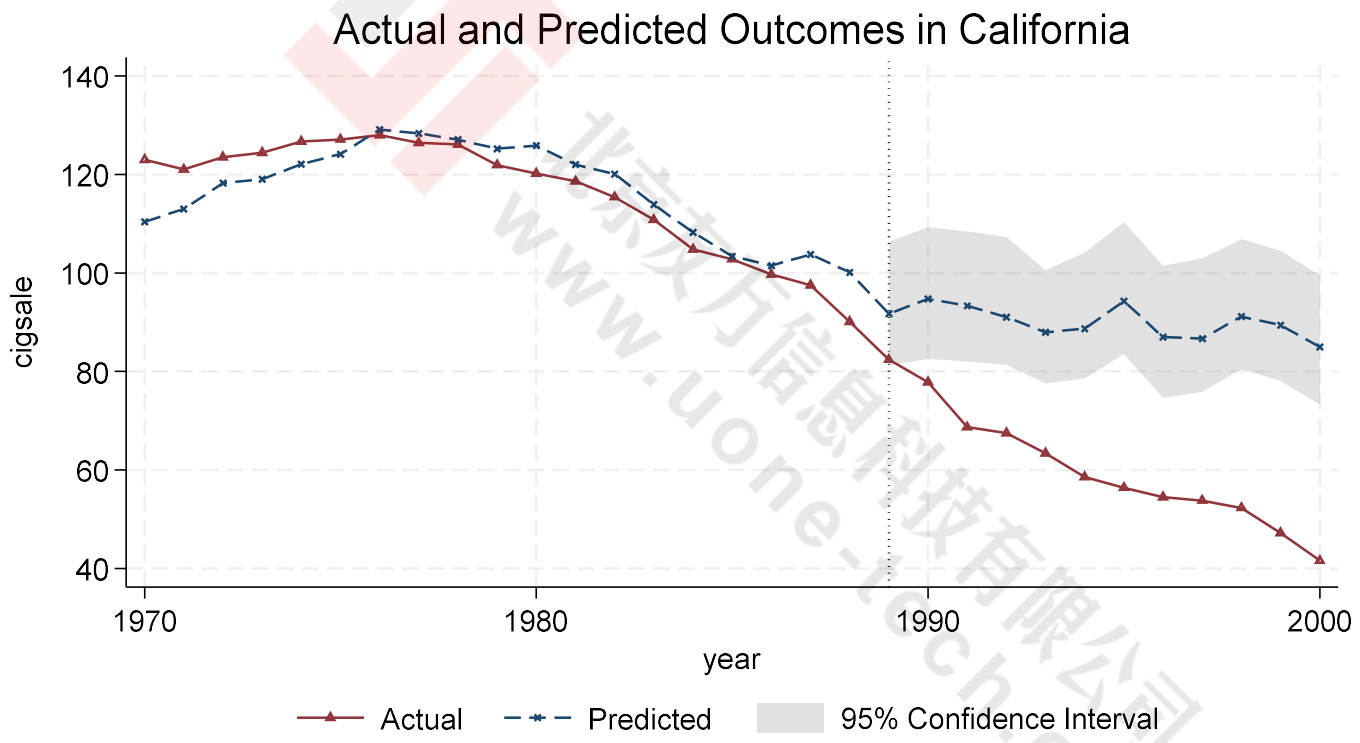
Note: The number of factors is estimated using the eigenvalue ratio criterion
      (Ahn and Horenstein, 2013), with the number of factors selected from the
      range [1, 10].

Estimation and prediction results during the posttreatment periods in California,
> with equal-tailed confidence intervals:

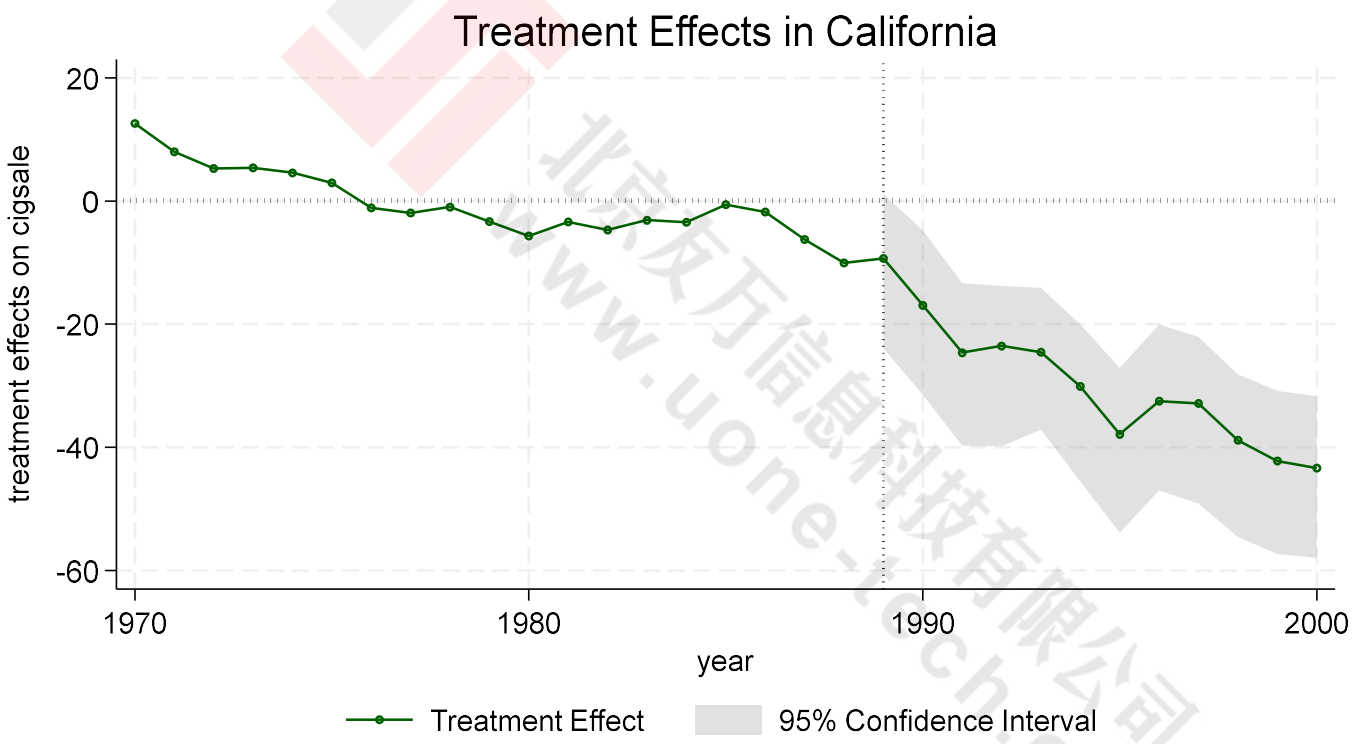
Time	Actual Outcome	Predicted Outcome	[95% Confidence Interval]	
1989	82.4000	91.7245	81.2485	106.3303
1990	77.8000	94.7342	82.5341	109.2785
1991	68.7000	93.3237	82.0242	108.4273
1992	67.5000	91.0326	81.2992	107.3241
1993	63.4000	87.9495	77.5207	100.5318
1994	58.6000	88.7011	78.6162	104.1397
1995	56.4000	94.2815	83.5088	110.2981
1996	54.5000	86.9832	74.5968	101.4922
1997	53.8000	86.6681	75.8361	102.9366
1998	52.3000	91.1378	80.4570	106.8708
1999	47.2000	89.4060	78.0402	104.5328
2000	41.6000	84.9636	73.2819	99.5071

Time	Treatment Effect	*p*-value	[95% Confidence Interval]	
1989	-9.3245*	0.084	-23.9303	1.1515
1990	-16.9342**	0.012	-31.4785	-4.7341
1991	-24.6237***	0.000	-39.7273	-13.3242
1992	-23.5326***	0.000	-39.8241	-13.7992
1993	-24.5495***	0.000	-37.1318	-14.1207
1994	-30.1011***	0.000	-45.5397	-20.0162
1995	-37.8815***	0.000	-53.8981	-27.1088
1996	-32.4832***	0.000	-46.9922	-20.0968
1997	-32.8681***	0.000	-49.1366	-22.0361
1998	-38.8378***	0.000	-54.5708	-28.1570
1999	-42.2060***	0.000	-57.3328	-30.8402
2000	-43.3636***	0.000	-57.9071	-31.6819
Mean	-29.7255			

Note: (1) The average treatment effect over the posttreatment period is -29.7255.
(2) ***, **, and * denote statistical significance of treatment effect at the 1%, 5%, and 10% level, respectively.



Note: The vertical dotted line indicates the start of the treatment period (1989).



Note: The vertical dotted line indicates the start of the treatment period (1989).

7. Conclusion

Initially proposed by Gobillon and Magnac (2016) and Xu (2017), the idea of using interactive fixed effects for counterfactual prediction was formalized by Bai and Ng (2021), while valid inference via residual-based wild bootstrap was provided by Li, Shen and Zhou (2024). This has become a powerful tool for causal inference using panel data with both large *N* and large *T*.

This paper presents the *ifete* command for easy implementation of the tall-wide algorithm of Bai and Ng (2021) for point estimation of treatment effects, as well as the confidence interval and *p*-values of Li, Shen and Zhou (2024) for inference.

This command accommodates models with or without covariates and/or nonstationary trends, making it suitable for a wide range of empirical settings.