The Stata Journal (2017)
17, Number 3, pp. 517–545

Bias corrections for probit and logit models with two-way fixed effects

Mario Cruz-Gonzalez
Department of Economics
Boston University
Boston, MA
mgonza@bu.edu

Iván Fernández-Val
Department of Economics
Boston University
Boston, MA
ivanf@bu.edu

Martin Weidner
Department of Economics
University College London
London, UK
m.weidner@ucl.ac.uk

Abstract. In this article, we present the user-written commands `probitfe` and `logitfe`, which fit probit and logit panel-data models with individual and time unobserved effects. Fixed-effects panel-data methods that estimate the unobserved effects can be severely biased because of the incidental parameter problem (Neyman and Scott, 1948, *Econometrica* 16: 1–32). We tackle this problem using the analytical and jackknife bias corrections derived in Fernández-Val and Weidner (2016, *Journal of Econometrics* 192: 291–312) for panels where the two dimensions (N and T) are moderately large. We illustrate the commands with an empirical application to international trade and a Monte Carlo simulation calibrated to this application.

Keywords: st0485, probitfe, logitfe, probit, logit, panel, fixed effects, bias corrections, incidental parameter problem

1 Introduction

Panel data, which consist of multiple observations over time for a set of individuals, are commonly used in empirical analysis to control for unobserved individual and time heterogeneity. Researchers often do this by adding individual and time effects to the model and treating these unobserved effects as parameters to be estimated in the so-called fixed-effects (FEs) approach. However, FEs estimators of nonlinear models such as binary response models suffer from the incidental parameter problem (Neyman and Scott 1948). A special case is the logit model with individual effects, where one can use the conditional likelihood approach (Rasch 1960; Andersen, 1973; Chamberlain, 1984), implemented in `clogit` and `xtlogit` (see [R] **clogit** and [XT] **xtlogit**). This approach provides estimates of model coefficients, but it is not available for the probit model and also does not produce estimates of average partial effects (APE) or marginal effects, which are often the quantities of interest in binary response models. Moreover, `clogit` and `xtlogit` do not work well when the panel is long and when the model also includes time effects, because estimating the time effects introduces additional incidental parameter bias. Time effects are routinely used in empirical analysis to control for aggregate common shocks and to parsimoniously account for cross-sectional dependence.

We deal with the incidental parameter problem using the bias corrections recently developed by Fernández-Val and Weidner (2016) for nonlinear panel models with two-

 st0485

way FEs. These corrections apply to panel datasets or other pseudopanel data structures where the two dimensions (N and T) are moderately large; see Arellano and Hahn (2007) and Fernández-Val and Weidner (2017) for a survey on bias correction methods to deal with the incidental parameter problem. Examples of moderately long panel datasets include traditional microeconomic panel surveys with a long history of data such as the Panel Study of Income Dynamics and National Longitudinal Survey of Youth, international cross-country panels such as the PennWorld Table, U.S. state-level panels over time such as the Current Population Survey, and square pseudopanels of trade flows across countries such as the Feenstra's World Trade Flows and CEPII, where the indices correspond to the same countries indexed as importers and exporters. The `probitfe` and `logitfe` commands implement analytical and jackknife corrections for FEs estimators of logit and probit models with individual and time effects. They produce corrected estimates of the model coefficients and APE. To the best of our knowledge, these are the first commands in Stata to implement bias correction methods for nonlinear panel models.

The symbols $\to_P$ and $\to_d$ are used to denote convergence in probability and distribution, respectively.

The rest of this article is organized as follows: Section 2 describes the probit and logit panel models, the incidental parameter problem, and the bias corrections of Fernández-Val and Weidner (2016). Section 3 presents `probitfe`, `logitfe`, and their features. Section 4 provides an illustrative empirical application on international trade flows across countries, together with the results of a Monte Carlo simulation calibrated to the application. The expressions of the bias and variance, their estimators, and one-way models are given in the *Appendix*. We refer interested readers to Fernández-Val and Weidner (2016) for details on the assumptions, asymptotic theory, and proofs of all results presented in section 2.

2 Probit and logit models with two-way FEs

2.1 Models and estimators

We observe a binary response variable $Y_{it} \in \{0, 1\}$ together with a vector of covariates $\mathbf{X}_{it}$ for individual $i = 1, \ldots, N$ at time $t = 1, \ldots, T$. This definition of the indices i and t applies to standard panel datasets. More generally, i and t can specify any group structure in the data. For example, in the empirical application of section 4, i and t index the same countries as importers and exporters, respectively. The logit and probit models specify the probability of $Y_{it} = 1$ conditional on current and past values of the regressors $\mathbf{X}_i^t = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{it})$, unobserved individual specific effects $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$, and unobserved time specific effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_T)$, namely,

$$\mathbf{Pr}\left(Y_{it} = 1 \mid \mathbf{X}_i^t, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\beta}\right) = F(\mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i + \gamma_t)$$

where $F : \mathbb{R} \to [0, 1]$ is a cumulative distribution function (the standard normal distribution in the probit model and the standard logistic distribution in the logit model) and $\boldsymbol{\beta}$ is a vector of unknown model coefficients of the same dimension as $\mathbf{X}_{it}$. The vector $\mathbf{X}_{it}$ contains predetermined variables with respect to Y_{it}. In particular, $\mathbf{X}_{it}$ can include lags of Y_{it} to accommodate dynamic models. In some static models or in panels where t does not index time, $\mathbf{X}_{it}$ can be treated as strictly exogenous with respect to Y_{it} by replacing $\mathbf{X}_i^t$ by $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{iT})$ in the conditioning set. The model does not impose any restriction on the relation between the covariate vector and the unobserved effects. In empirical applications, the conditioning on the unobserved effects serves to control for endogeneity as the individual and time effects capture unobserved heterogeneity that can be related to the covariates.

We adopt an FEs approach and treat the individual and time effects as parameters to be estimated. We denote by $\boldsymbol{\beta}^0$, $\boldsymbol{\alpha}^0$, and $\boldsymbol{\gamma}^0$ the true values of the parameters, that is, the parameters that are assumed to generate the distribution of Y_{it} according to the model above. The (conditional) log-likelihood function of the observation (i, t) is

$$\begin{aligned}\ell_{it}(\boldsymbol{\beta}, \alpha_i, \gamma_t) := & Y_{it} \times \log\left\{F(\mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i + \gamma_t)\right\} \\ & + (1 - Y_{it}) \times \log\left\{1 - F\left(\mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i + \gamma_t\right)\right\}\end{aligned}$$

and the FEs estimators for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$ are obtained by maximizing the log-likelihood function of the sample,

$$\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}}\right) \in \operatorname{argmax}_{(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \in \mathbb{R}^{\dim \boldsymbol{\beta} + N + T}} \sum_{i,t} \ell_{it}(\boldsymbol{\beta}, \alpha_i, \gamma_t) \tag{1}$$

This is a smooth concave maximization program for the logit and probit models. However, there is a perfect collinearity problem because the log-likelihood function is invariant to the transformation $\alpha_i \mapsto \alpha_i + c$ and $\gamma_t \mapsto \gamma_t - c$ for any $c \in \mathbb{R}$. If $\mathbf{X}_{it}$ includes a constant term, we overcome this problem by dropping α_1 and γ_1, which normalizes $\alpha_1 = 0$ and $\gamma_1 = 0$. If $\mathbf{X}_{it}$ does not include a constant term, we need to drop only either α_1 or γ_1. As in linear panel models, the covariates $\mathbf{X}_{it}$, other than the constant term, need to vary both across i and over t to avoid further perfect collinearity problems, that is, to guarantee that the log-likelihood function is strictly concave.

The above FEs estimators can be implemented in Stata by using the existing `logit` and `probit` (see [R] **logit** and [R] **probit**) commands including individual and time binary indicators to account for α_i and γ_t. However, as we will explain in the next subsection, the FEs estimator $\widehat{\boldsymbol{\beta}}$ can be severely biased, and the existing routines do not incorporate any bias-correction method.

In many applications of the logit and probit models, the ultimate parameters of interest are the APE of the covariates, which take the form

$$\boldsymbol{\delta}^0 = \mathbb{E}\left\{\Delta(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, \boldsymbol{\gamma}^0)\right\}, \quad \Delta(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = (NT)^{-1} \sum_{i,t} \Delta(\mathbf{X}_{it}, \boldsymbol{\beta}, \alpha_i, \gamma_t) \tag{2}$$

where $\mathbb{E}$ denotes the expectation with respect to the joint distribution of the data and the unobserved effects. The expression of the partial effect function $\Delta(\mathbf{X}_{it}, \boldsymbol{\beta}, \alpha_i, \gamma_t)$ depends on the type of covariate. If $X_{it,k}$, the kth element of $\mathbf{X}_{it}$, is binary, then its partial effect on the conditional probability of Y_{it} is calculated using

$$\Delta(\mathbf{X}_{it}, \boldsymbol{\beta}, \alpha_i, \gamma_t) = F(\beta_k + \mathbf{X}_{it,-k}'\boldsymbol{\beta}_{-k} + \alpha_i + \gamma_t) - F(\mathbf{X}_{it,-k}'\boldsymbol{\beta}_{-k} + \alpha_i + \gamma_t)$$

where β_k is the kth element of $\boldsymbol{\beta}$ and $\mathbf{X}_{it,-k}$ and $\boldsymbol{\beta}_{-k}$ include all elements of $\mathbf{X}_{it}$ and $\boldsymbol{\beta}$ except for the kth element. This partial effect measures the impact of changing $X_{it,k}$ from 0 to 1 on the conditional probability of $Y_{it} = 1$ holding the rest of the covariates fixed at their observed values $\mathbf{X}_{it,-k}$. If $X_{it,k}$ is not binary, then the partial effect of $X_{it,k}$ on the conditional probability of Y_{it} is calculated using

$$\Delta(\mathbf{X}_{it}, \boldsymbol{\beta}, \alpha_i, \gamma_t) = \beta_k \partial F(\mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i + \gamma_t)$$

where ∂F is the derivative of F. This partial effect measures the impact of a marginal change in $X_{it,k}$ on the probability of $Y_{it} = 1$, conditional on the observed value of the covariates $\mathbf{X}_{it}$.

The FEs estimator of APE is obtained by plugging in estimators of the model parameters in the sample analog of (2); that is,

$$\widetilde{\boldsymbol{\delta}} = \Delta\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\gamma}}\right)$$

where $\widetilde{\boldsymbol{\beta}}$ is an estimator for $\boldsymbol{\beta}$, and

$$(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\gamma}}) \in \text{argmax}_{(\boldsymbol{\alpha},\boldsymbol{\gamma})\in\mathbb{R}^{N+T}} \sum_{i,t} \ell_{it}\left(\widetilde{\boldsymbol{\beta}}, \alpha_i, \gamma_t\right)$$

For example, if $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$, then $(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\gamma}}) = (\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}})$, where $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\gamma}})$ is the FEs estimator defined in (1). Again, there are Stata routines to calculate $\widetilde{\boldsymbol{\delta}}$, but they do not implement any bias correction.

2.2 Incidental parameter problem

The FEs estimators $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\delta}}$ suffer from the Neyman and Scott incidental parameter problem. In particular, these estimators are inconsistent under asymptotic sequences where T is fixed and $N \to \infty$ when the model has individual effects. They are also inconsistent when N is fixed and $T \to \infty$ when the model has time effects. The source of the problem is that there is only a fixed number of observations to estimate each unobserved effect, T observations for each individual effect or N observations for each time effect, rendering the corresponding estimators inconsistent. The nonlinearity of the model propagates the inconsistency in the estimation of the individual or time effects to all the model coefficients and APE.

A recent response to the incidental parameter problem is to consider an alternative asymptotic approximation where $N \to \infty$ and $T \to \infty$ (for example, Arellano and Hahn [2007]). The key insight of this so-called large-T panel-data literature is that, under this approximation, the incidental parameter problem becomes a bias problem that is easier to handle than the inconsistency problem under the traditional asymptotic approximation. In particular, Fernández-Val and Weidner (2016) show that as $N, T \to \infty$, with $N/T \to c > 0$, the limit distribution of $\widehat{\boldsymbol{\beta}}$ is described by

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 - \mathbf{B}^\beta/T - \mathbf{D}^\beta/N\right) \to_d \mathcal{N}\left(\mathbf{0}, \mathbf{V}^\beta\right)$$

where $\mathbf{V}^\beta$ is the asymptotic variance–covariance matrix, $\mathbf{B}^\beta$ is an asymptotic bias term coming from the estimation of the individual effects, and $\mathbf{D}^\beta$ is an asymptotic bias term coming from the estimation of the time effects.[1] The finite sample prediction of this result is that the FEs estimator can have significant bias relative to its dispersion even if N and T are of the same order. Moreover, confidence intervals constructed around the FEs estimator can severely undercover the true value of the parameter even in large samples. We show that this large-N large-T version of the incidental parameters problem provides a good approximation to the finite sample behavior of the FEs estimator through simulation examples in section 4.

For $\widetilde{\boldsymbol{\delta}}$, the situation is different because the order of the standard deviation (SD) of $\widetilde{\boldsymbol{\delta}}$, $1/\sqrt{N} + 1/\sqrt{T}$, is slower than the order of the SD of $\widehat{\boldsymbol{\beta}}$, $1/\sqrt{NT}$. In this case, Fernández-Val and Weidner (2016) show that as $N, T \to \infty$, with $N/T \to c > 0$, the limit distribution is

$$\sqrt{\min(N,T)} \left(\widetilde{\boldsymbol{\delta}} - \boldsymbol{\delta}^0 - \mathbf{B}^\delta/T - \mathbf{D}^\delta/N \right) \to_d \mathcal{N}\left(\mathbf{0}, \mathbf{V}^\delta\right)$$

where $\mathbf{V}^\delta$ is the asymptotic variance, $\mathbf{B}^\delta$ is the asymptotic bias coming from the estimation of the individual effects, and $\mathbf{D}^\delta$ is the asymptotic bias term coming from the estimation of the time effects.[2] Here the SD dominates both of the bias terms, implying that $\widetilde{\boldsymbol{\delta}}$ is asymptotically first-order unbiased. The biases can nevertheless be significant in small samples as we show in section 4 through simulation examples.

2.3 Analytical bias correction

The analytical bias correction consists of removing estimates of the leading terms of the bias from the FEs estimator of $\boldsymbol{\beta}$. Let $\widehat{\mathbf{B}}^\beta$ and $\widehat{\mathbf{D}}^\beta$ be consistent estimators of $\mathbf{B}^\beta$ and $\mathbf{D}^\beta$; that is, $\widehat{\mathbf{B}}^\beta \to_P \mathbf{B}^\beta$ and $\widehat{\mathbf{D}}^\beta \to_P \mathbf{D}^\beta$ as $N, T \to \infty$. The bias-corrected estimator can be formed as

$$\widetilde{\boldsymbol{\beta}}^A = \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{B}}^\beta/T - \widehat{\mathbf{D}}^\beta/N$$

As $N, T \to \infty$ with $N/T \to c > 0$, the limit distribution of $\widetilde{\boldsymbol{\beta}}^A$ is

$$\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}}^A - \boldsymbol{\beta}^0\right) \to_d \mathcal{N}\left(\mathbf{0}, \mathbf{V}^\beta\right)$$

The analytical correction therefore centers the asymptotic distribution at the true value of the parameter without increasing asymptotic variance. This result predicts that in large samples, the corrected estimator has small bias relative to dispersion, the correction does not increase dispersion, and the confidence intervals constructed around the corrected estimator have coverage probabilities close to the nominal levels. We show that these predictions provide a good approximation to the behavior of the corrections in section 4.

1. The expressions of $\mathbf{V}^\beta$, $\mathbf{B}^\beta$, and $\mathbf{D}^\beta$ for probit and logit models are given in the *Appendix*.
2. The expressions of $\mathbf{V}^\delta$, $\mathbf{B}^\delta$, and $\mathbf{D}^\delta$ for probit and logit models are given in the *Appendix*.

The bias-corrected APE can be constructed in the same fashion as

$$\widetilde{\boldsymbol{\delta}}^{A} = \widetilde{\boldsymbol{\delta}} - \widehat{\mathbf{B}}^{\delta}/T - \widehat{\mathbf{D}}^{\delta}/N$$

where $\widehat{\mathbf{B}}^{\delta}$ and $\widehat{\mathbf{D}}^{\delta}$ are consistent estimators of $\mathbf{B}^{\delta}$ and $\mathbf{D}^{\delta}$; that is, $\widehat{\mathbf{B}}^{\delta} \to_P \mathbf{B}^{\delta}$ and $\widehat{\mathbf{D}}^{\delta} \to_P \mathbf{D}^{\delta}$ as $N, T \to \infty$. The limit distribution of $\widetilde{\boldsymbol{\delta}}^{A}$ is

$$\sqrt{\min(N,T)}\left\{\widetilde{\boldsymbol{\delta}}^{A} - \boldsymbol{\delta}^0 + o_P(T^{-1} + N^{-1})\right\} \to_d \mathcal{N}(\mathbf{0}, \mathbf{V}^{\delta})$$

We give the details on how to compute $\widehat{\mathbf{B}}^{\beta}$, $\widehat{\mathbf{D}}^{\beta}$, $\widehat{\mathbf{B}}^{\delta}$, and $\widehat{\mathbf{D}}^{\delta}$ in the *Appendix*. The `probitfe` and `logitfe` commands compute these analytical bias corrections with the `analytical` option. If the regressors $\mathbf{X}_{it}$ are predetermined, for example, when lagged dependent variables are included, the calculation of $\widehat{\mathbf{B}}^{\beta}$ and $\widehat{\mathbf{B}}^{\delta}$, and thus of the bias corrections, requires the specification of a trimming parameter $L \in \{1, 2, 3, \ldots\}$ to estimate a spectral expectation. For the asymptotic theory, the requirement on L is that $L \to \infty$ such that $L/T \to 0$ because $T \to \infty$. In practice, we do not recommend using L larger than four, and we suggest computing the analytical bias corrections for different values of L as a robustness check. When the regressors $\mathbf{X}_{it}$ are strictly exogenous, L should be set to zero. The trimming parameter is set through the command options `lags(`*integer*`)`, as described below.

2.4 Jackknife bias correction

The `probitfe` and `logitfe` commands with the `jackknife` option allow for six different types of jackknife corrections, denoted as `ss1`, `ss2`, `js`, `sj`, `jj`, and `double`. We will briefly explain each correction and give some intuition about how they reduce bias. The jackknife corrections do not require explicit estimation of the bias but are computationally more intensive because they involve solving multiple FEs estimation programs. The methods are combinations of the leave-one-observation-out panel jackknife (PJ) of Hahn and Newey (2004) and the split-panel jackknife (SPJ) of Dhaene and Jochmans (2015) applied to the two dimensions of the panel.

Let $\mathscr{N} = \{1, \ldots, N\}$ and $\mathscr{T} = \{1, \ldots, T\}$. Define the FEs estimator of $\boldsymbol{\beta}$ in the subpanel with cross-sectional indices $\mathscr{A} \subseteq \mathscr{N}$ and time-series indices $\mathscr{B} \subseteq \mathscr{T}$ as

$$\widehat{\boldsymbol{\beta}}_{\mathscr{A},\mathscr{B}} \in \operatorname{argmax}_{\boldsymbol{\beta} \in \mathbb{R}^{\dim \boldsymbol{\beta}}} \max_{\alpha(\mathscr{A}) \in \mathbb{R}^{|\mathscr{A}|}} \max_{\gamma(\mathscr{B}) \in \mathbb{R}^{|\mathscr{B}|}} \sum_{i \in \mathscr{A}, t \in \mathscr{B}} \ell_{it}(\boldsymbol{\beta}, \alpha_i, \gamma_t)$$

where $\alpha(\mathscr{A}) = \{\alpha_i : i \in \mathscr{A}\}$ and $\gamma(\mathscr{B}) = \{\gamma_t : t \in \mathscr{B}\}$. Notice that the original FEs estimator $\widehat{\boldsymbol{\beta}}$ defined above is equal to $\widehat{\boldsymbol{\beta}}_{\mathscr{N},\mathscr{T}}$. Using this notation, we can now describe the six jackknife corrections:

- The correction `ss1` applies SPJ simultaneously to both dimensions of the panel. Let $\widetilde{\boldsymbol{\beta}}_{N/2,T/2}$ be the average of the four split jackknife estimators that leave out half the individuals and half the time periods; that is,

$$\widetilde{\boldsymbol{\beta}}_{N/2,T/2} = \frac{1}{4}\Big(\widehat{\boldsymbol{\beta}}_{\{i\,:\,i\leq\lceil N/2\rceil\},\{t\,:\,t\leq\lceil T/2\rceil\}} + \widehat{\boldsymbol{\beta}}_{\{i\,:\,i\geq\lfloor N/2+1\rfloor\},\{t\,:\,t\leq\lceil T/2\rceil\}} \\ + \widehat{\boldsymbol{\beta}}_{\{i\,:\,i\leq\lceil N/2\rceil\},\{t\,:\,t\geq\lfloor T/2+1\rfloor\}} + \widehat{\boldsymbol{\beta}}_{\{i\,:\,i\geq\lfloor N/2+\rfloor\},\{t\,:\,t\geq\lfloor T/2+1\rfloor\}}\Big)$$

  where $\lfloor . \rfloor$ and $\lceil . \rceil$ denote the floor and ceiling function, respectively. The `ss1` corrected estimator is

$$\widetilde{\boldsymbol{\beta}}^{\texttt{ss1}} = 2\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_{N/2,T/2}$$

- The correction `ss2` applies SPJ separately to both dimensions of the panel. Let $\widetilde{\boldsymbol{\beta}}_{N,T/2}$ be the average of the two split jackknife estimators that leave out the first and second halves of the time periods, and let $\widetilde{\boldsymbol{\beta}}_{N/2,T}$ be the average of the two split jackknife estimators that leave out half the individuals; that is,

$$\widetilde{\boldsymbol{\beta}}_{N,T/2} = \frac{1}{2}\left(\widehat{\boldsymbol{\beta}}_{\mathscr{N},\{t\,:\,t\leq\lceil T/2\rceil\}} + \widehat{\boldsymbol{\beta}}_{\mathscr{N},\{t\,:\,t\geq\lfloor T/2+1\rfloor\}}\right)$$
$$\widetilde{\boldsymbol{\beta}}_{N/2,T} = \frac{1}{2}\left(\widehat{\boldsymbol{\beta}}_{\{i\,:\,i\leq\lceil N/2\rceil\},\mathscr{T}} + \widehat{\boldsymbol{\beta}}_{\{i\,:\,i\geq\lfloor N/2+1\rfloor\},\mathscr{T}}\right)$$

  The `ss2` corrected estimator is

$$\widetilde{\boldsymbol{\beta}}^{\texttt{ss2}} = 3\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_{N,T/2} - \widetilde{\boldsymbol{\beta}}_{N/2,T}$$

- The correction `js` applies PJ to the individual dimension and SPJ to the time dimension. Let $\widetilde{\boldsymbol{\beta}}_{N,T/2}$ be defined as above, and let $\widetilde{\boldsymbol{\beta}}_{N-1,T}$ be the average of the N jackknife estimators that leave out one individual; that is,

$$\widetilde{\boldsymbol{\beta}}_{N-1,T} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\boldsymbol{\beta}}_{\mathscr{N}\setminus\{i\},\mathscr{T}}$$

  The `js` corrected estimator is

$$\widetilde{\boldsymbol{\beta}}^{\texttt{js}} = (N+1)\widehat{\boldsymbol{\beta}} - (N-1)\widetilde{\boldsymbol{\beta}}_{N-1,T} - \widetilde{\boldsymbol{\beta}}_{N,T/2}$$

- The correction `sj` applies SPJ to the individual dimension and PJ to the time dimension. Let $\widetilde{\boldsymbol{\beta}}_{N/2,T}$ be defined as above, and let $\widetilde{\boldsymbol{\beta}}_{N,T-1}$ be the average of the T jackknife estimators that leave out one time period; that is,

$$\widetilde{\boldsymbol{\beta}}_{N,T-1} = \frac{1}{T}\sum_{t=1}^{T}\widehat{\boldsymbol{\beta}}_{\mathscr{N},\mathscr{T}\setminus\{t\}}$$

  The `sj` corrected estimator is

$$\widetilde{\boldsymbol{\beta}}^{\texttt{sj}} = (T+1)\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_{N/2,T} - (T-1)\widetilde{\boldsymbol{\beta}}_{N,T-1}$$

- The correction `jj` applies PJ to both the individual and the time dimension. Let $\widetilde{\boldsymbol{\beta}}_{N-1,T}$ and $\widetilde{\boldsymbol{\beta}}_{N,T-1}$ be defined as above. The `jj` corrected estimator is

$$\widetilde{\boldsymbol{\beta}}^{\texttt{jj}} = (N+T-1)\widehat{\boldsymbol{\beta}} - (N-1)\widetilde{\boldsymbol{\beta}}_{N-1,T} - (T-1)\widetilde{\boldsymbol{\beta}}_{N,T-1}$$

- The correction `double` uses PJ for observations with the same cross-section and time-series indices. This type of correction makes sense only for panels where i and t index the same entities. For example, in country trade data, the cross-section dimension represents each country as an importer, and the "time-series dimension" represents each country as an exporter. Thus let $N = T$, and define $\widetilde{\boldsymbol{\beta}}_{N-1,N-1}$ as the average of the N jackknife estimators that leave one entity out; that is,

$$\widetilde{\boldsymbol{\beta}}_{N-1,N-1} = \frac{1}{N}\sum_{i=1}^{N} \widehat{\boldsymbol{\beta}}_{\mathscr{N}\setminus\{i\},\mathscr{N}\setminus\{i\}}$$

  The corrected estimator is

$$\widetilde{\boldsymbol{\beta}}^{\texttt{double}} = N\widehat{\boldsymbol{\beta}} - (N-1)\widetilde{\boldsymbol{\beta}}_{N-1,N-1}$$

To give some intuition on how these corrections reduce bias, we use a first-order approximation to the bias,

$$\text{bias}\left(\widehat{\boldsymbol{\beta}}_{\mathscr{A},\mathscr{B}}\right) \approx \mathbf{B}^{\boldsymbol{\beta}}/|\mathscr{A}| + \mathbf{D}^{\boldsymbol{\beta}}/|\mathscr{B}|$$

where $|\mathscr{A}|$ denotes the cardinality of the set $\mathscr{A}$. Consider, for example, the `ss1` option. Using the previous approximation, we see that

$$\text{bias}\left(\widetilde{\boldsymbol{\beta}}^{\texttt{ss1}}\right) \approx 2 \times \text{bias}\left(\widehat{\boldsymbol{\beta}}_{\mathscr{N},\mathscr{T}}\right) - \text{bias}\left(\widetilde{\boldsymbol{\beta}}_{N/2,T/2}\right) = 0$$

because the leading bias of $\widetilde{\boldsymbol{\beta}}_{N/2,T/2}$ is twice the leading bias in $\widehat{\boldsymbol{\beta}}_{\mathscr{N},\mathscr{T}}$ because the subpanels used to construct $\widetilde{\boldsymbol{\beta}}_{N/2,T/2}$ contain half the individuals and time periods. In other words, subtracting $(\widetilde{\boldsymbol{\beta}}_{N/2,T/2} - \widehat{\boldsymbol{\beta}}_{\mathscr{N},\mathscr{T}})$ from $\widehat{\boldsymbol{\beta}}$ removes a nonparametric estimator of the leading bias. Similarly, we can show that the leading bias of $\widehat{\boldsymbol{\beta}}$ is removed by the other corrections because they use appropriate choices of the size of the subpanels and corresponding coefficients in the linear combinations of the subpanel estimators.

There are panels for which there is no natural ordering of the observations along some of the dimensions, for example, the individuals in the Panel Study of Income Dynamics. In this case, there are multiple ways to select the subpanels to implement the `ss1` and `ss2` corrections. To avoid any arbitrariness in the choice of subpanels, `probitfe` and `logitfe` include the possibility of constructing $\widetilde{\boldsymbol{\beta}}_{N/2,T/2}$ and $\widetilde{\boldsymbol{\beta}}_{N/2,T}$ as the average of the estimators obtained from multiple orderings of the panels by randomly permuting the indices of the dimension that has no natural ordering of the observations. The `multiple()` option allows the user to specify the number of different permutations of the panel to use.

Fernández-Val and Weidner (2016) show that the correction `ss2` removes the bias without increasing dispersion in large samples. In particular, they show that the limit distribution of $\widetilde{\boldsymbol{\beta}}^{\texttt{ss2}}$ as $N, T \to \infty$ with $N/T \to c > 0$ is

$$\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}}^{\texttt{ss2}} - \boldsymbol{\beta}^0\right) \to_d \mathcal{N}(\mathbf{0}, \mathbf{V}^\beta)$$

the same as the limit distribution of the analytical correction. The assumptions required for this result include homogeneity conditions along the two dimensions of the panel to guarantee that the bias terms $\mathbf{B}^\beta$ and $\mathbf{D}^\beta$ are the same in all the subpanels. The analytical corrections described above do not require these type of conditions and are therefore more widely applicable.

Jackknife corrections for the APE are formed analogously. We compute estimates $\widetilde{\boldsymbol{\delta}}_{\mathscr{A},\mathscr{B}}$ from subpanels with cross-sectional indices $\mathscr{A} \subseteq \mathscr{N}$ and time-series indices $\mathscr{B} \subseteq \mathscr{T}$ and use the corrections described above replacing $\boldsymbol{\beta}$ by $\boldsymbol{\delta}$ everywhere.

2.5 One-way FEs

So far, we have focused on two-way FEs models with individual and time effects because they are the most commonly used in empirical applications. For completeness, the `probitfe` and `logitfe` commands also provide functionality for one-way FEs models that include only either individual effects or time effects (using the `ieffects()` and `teffects()` options, respectively), as well as the flexibility to choose whether the bias corrections should account only for either individual effects or time effects (using the `ibias()` and `tbias()` options, respectively). FEs estimators of these models also suffer from the incidental parameter problem. The commands implement the analytical and jackknife corrections of Hahn and Newey (2004) and Fernández-Val (2009) and the split-panel correction of Dhaene and Jochmans (2015). We do not describe these corrections in detail, because they are very similar to the ones described above for two-way models. For example, the analytical correction for $\boldsymbol{\beta}$ has the same form as $\widetilde{\boldsymbol{\beta}}^A$ after making one of the estimated bias terms equal to zero: $\widehat{\mathbf{D}}^\beta = \mathbf{0}$ for models without time effects or $\widehat{\mathbf{B}}^\beta = \mathbf{0}$ for models without individual effects. We give the expressions of $\mathbf{B}^\beta$ and $\mathbf{D}^\beta$ and describe the jackknife corrections for one-way FEs models in the *Appendix*.

2.6 Unbalanced panel data

In the description of the incidental parameter problem and bias corrections, we implicitly assumed that the panel was balanced; that is, we observe each individual, $i = 1, \ldots, N$, at each time period, $t = 1, \ldots, T$. Nevertheless, unbalanced panel datasets are common in empirical applications. Unbalancedness does not introduce special theoretical complications provided that the source of the missing observations is random. It does not introduce complications in the computation either, because `probitfe` and `logitfe` use Stata's time-series operators that account for missing observations, provided the data are declared to be time series.

Suppose, for example, that we have the following dataset:

```
. tsset
       panel variable:  id (weakly balanced)
        time variable:  time, 1 to 7, but with gaps
                delta:  1 unit
. list id time

      ┌───────────┐
      │ id   time │
      ├───────────┤
   1. │  1      1 │
   2. │  1      2 │
   3. │  1      4 │
   4. │  1      5 │
   5. │  1      7 │
      ├───────────┤
   6. │  2      2 │
   7. │  2      3 │
   8. │  2      5 │
   9. │  2      6 │
  10. │  2      7 │
      └───────────┘
```

This includes two individuals and seven time periods, but there are no observations for every time period for each individual. Time-series operators are important when the analytical correction is applied and the trimming parameter is higher than zero. If the trimming parameter is equal to 1, for example, `probitfe` and `logitfe` will correctly produce a missing value for $t = \{1, 4, 7\}$ for the first individual and produce a missing value for $t = \{2, 5\}$ for the second individual.

In the jackknife corrections, `probitfe` and `logitfe` identify the appropriate subset of observations for each individual because they use time as index instead of the observation number. If we apply, for example, the jackknife bias correction `ss1`, where the subpanels include half the time periods for each individual, the commands will correctly use $t = \{1, 2, 4\}$ for the first individual and $t = \{2, 3\}$ for the second individual.

3 The probitfe and logitfe commands

3.1 Syntax

Both `probitfe` and `logitfe` share the same syntax and options. Here we use the syntax for `probitfe`. One needs only to replace `probitfe` with `logitfe` if one wishes to fit a logit model.

Uncorrected estimator

`probitfe` *depvar indepvars* [*if*] [*in*], `noc`orrection [`ieffects(`*string*`)` `teffects(`*string*`)` `pop`ulation`(`*integer*`)`]

Analytical-corrected estimator

`probitfe` *depvar indepvars* [*if*] [*in*] [, `analytical lags(`*integer*`)`
`ieffects(`*string*`) teffects(`*string*`) ibias(`*string*`) tbias(`*string*`)`
`population(`*integer*`)`]

Jackknife-corrected estimator

`probitfe` *depvar indepvars* [*if*] [*in*], `jackknife` [`ss1` [`multiple(`*integer*`)`
`individuals time`] `ss2` [`multiple(`*integer*`) individuals time`] `js sj jj`
`double ieffects(`*string*`) teffects(`*string*`) ibias(`*string*`) tbias(`*string*`)`
`population(`*integer*`)`]

Both a panel variable and a time variable must be specified. *indepvars* may contain factor variables. *depvar* and *indepvars* may contain time-series operators.

3.2 Options for uncorrected estimator

`nocorrection` computes the probit FEs estimator without correcting for the bias because of the incidental parameter problem.

If the `nocorrection` option is specified without the type of included effects, the model will include both individual and time effects. `ieffects(no)` and `teffects(no)` cannot be combined.

`ieffects(`*string*`)` specifies whether the uncorrected estimator includes individual effects.

`ieffects(yes)`, the default, includes individual FEs.

`ieffects(no)` omits the individual FEs.

`teffects(`*string*`)` specifies whether the uncorrected estimator includes time effects.

`teffects(yes)`, the default, includes time FEs.

`teffects(no)` omits the time FEs.

`population(`*integer*`)` adjusts the estimation of the variance of the APE by a finite population correction (FPC). Let m be the number of original observations included in `probitfe`, and let $M \geq m$ be the number of observations for the entire population declared by the user. The computation of the variance of the APE is corrected by the factor $\text{FPC} = (M-m)/(M-1)$. The default is `population(1)`, corresponding to an infinite population. Notice that M makes reference to the total number of observations and not the total number of individuals. If, for example, the population has 100 individuals followed over 10 time periods, the user must specify `population(1000)` instead of `population(100)`.

3.3 Options for analytical-corrected estimator

`analytical`, the default, computes the probit FEs estimator using the analytical bias correction derived in Fernández-Val and Weidner (2016).

`lags(`*integer*`)` specifies the value of the trimming parameter to estimate spectral expectations. See the discussion in section 2.3 for details. The default is `lags(0)`; that is, the trimming parameter to estimate spectral expectations is set to zero. This option should be used when the model is static with strictly exogenous regressors.

The trimming parameter can be set to any value between 0 and $(T-1)$. A trimming parameter higher than 0 should be used when the model is dynamic or some of the regressors are weakly exogenous or predetermined. As mentioned in section 2.3, we do not recommend setting the value of the trimming parameter to a value higher than 4, unless T is very large.

If the `analytical` option is specified without the type of included effects, the model will include both individual and time effects. `ieffects(no)` and `teffects(no)` cannot be combined.

`ieffects(`*string*`)` specifies whether the model includes individual FEs.

`ieffects(yes)`, the default, includes individual FEs.

`ieffects(no)` omits the individual FEs.

`teffects(`*string*`)` specifies whether the model includes time FEs.

`teffects(yes)`, the default, includes time FEs.

`teffects(no)` omits the time FEs.

If the `analytical` option is specified without the type of correction, the model will include analytical bias correction for both individual and time effects. `ibias(no)` and `tbias(no)` cannot be combined.

`ibias(`*string*`)` specifies whether the analytical correction accounts for individual effects.

`ibias(yes)`, the default, corrects for the bias coming from the individual FEs.

`ibias(no)` omits the individual FEs analytical bias correction.

`tbias(`*string*`)` specifies whether the analytical correction accounts for time effects.

`tbias(yes)`, the default, corrects for the bias coming from the time FEs.

`tbias(no)` omits the time FEs analytical bias correction.

`population(`*integer*`)` adjusts the estimation of the variance of the APE by an FPC. Let m be the number of original observations included in `probitfe`, and let $M \geq m$ be the number of observations for the entire population declared by the user. The computation of the variance of the APE is corrected by the factor $\text{FPC} = (M - m)/(M-1)$. The default is `population(1)`, corresponding to an infinite population. Notice that M makes reference to the total number of observations and not the total number of individuals. If, for example, the population has 100 individuals followed over 10 time periods, the user must specify `population(1000)` instead of `population(100)`.

3.4 Options for jackknife-corrected estimator

`jackknife` computes the probit FEs estimator using the jackknife bias corrections described in Fernández-Val and Weidner (2016).

`ss1` [`multiple(`*integer*`)` `individuals` `time`] specifies SPJ in four nonoverlapping subpanels; in each subpanel half the individuals and half the time periods are left out. See previous section for the details.

`multiple(`*integer*`)` is an `ss1` suboption that allows for different multiple partitions, each one made on a randomization of the observations in the panel; the default is `multiple(0)`; that is, the partitions are made on the original order in the dataset. If `multiple(10)` is specified, for example, the `ss1` estimator is computed 10 times on 10 different randomizations of the observations in the panel; the resulting estimator is the mean of these 10 SPJ corrections. This option can be used if there is a dimension of the panel where there is no natural ordering of the observations.

If neither `individuals` nor `time` options are specified, the multiple partitions are made on both the cross-sectional and the time dimensions.

`individuals` specifies the multiple partitions to be made only on the cross-sectional dimension.

`time` specifies the multiple partitions to be made only on the time dimension.

`ss2` [`multiple(`*integer*`)` `individuals` `time`], the default, specifies SPJ in both dimensions. As in `ss1`, there are four subpanels: in two of them, half the individuals are left out, but all time periods are included; in the other two, half the time periods are left out, but all the individuals are included. See previous section for the details.

`multiple(`*integer*`)` is an `ss2` suboption that allows for different multiple partitions, each one made on a randomization of the observations in the panel; the default is `multiple(0)`; that is, the partitions are made on the original order in the dataset. If `multiple(10)` is specified, for example, then the `ss2` estimator is computed 10 times on 10 different randomizations of the observations in the panel; the resulting estimator is the mean of these 10 SPJ corrections. This option can be used if there is a dimension of the panel where there is no natural ordering of the observations.

If neither `individuals` nor `time` options are specified, the multiple partitions are made on both the cross-sectional and the time dimensions.

`individuals` specifies the multiple partitions to be made only on the cross-sectional dimension.

`time` specifies the multiple partitions to be made only on the time dimension.

`js` uses delete-one PJ in the cross-section and SPJ in the time series. See the previous section for details.

`sj` uses SPJ in the cross-section and delete-one PJ in the time series. See the previous section for details.

`jj` uses delete-one PJ in both the cross-section and the time series. See the previous section for details.

`double` uses delete-one jackknife for observations with the same cross-section and the time-series indices. See the previous section for details.

If the `jackknife` option is specified without the type of included effects, the model will include both individual and time effects. `ieffects(no)` and `teffects(no)` cannot be combined.

`ieffects(`*string*`)` specifies whether the model includes individual FEs.

`ieffects(yes)`, the default, includes individual FEs.

`ieffects(no)` omits the individual FEs.

`teffects(`*string*`)` specifies whether the model includes time FEs.

`teffects(yes)`, the default, includes time FEs.

`teffects(no)` omits the time FEs.

If the `jackknife` option is specified without the type of correction, the model will include jackknife correction for both individual and time effects. `ibias(no)` and `tbias(no)` cannot be combined.

`ibias(`*string*`)` specifies whether the jackknife correction accounts for the individual effects.

`ibias(yes)`, the default, corrects for the bias coming from the individual FEs.

`ibias(no)` omits the individual FEs jackknife correction. If this option and multiple partitions only in the time dimension are specified together (for the jackknife `ss1` or `ss2` corrections), the resulting estimator is equivalent to the one without multiple partitions.

`tbias(`*string*`)` specifies whether the jackknife correction accounts for the time effects.

`tbias(yes)`, the default, corrects for the bias coming from the time FEs.

`tbias(no)` omits the time FEs jackknife correction. If this option and multiple partitions only in the cross-section are specified together (for the jackknife `ss1` or `ss2` corrections), the resulting estimator is equivalent to the one without multiple partitions.

`population(`*integer*`)` adjusts the estimation of the variance of the APE by an FPC. Let m be the number of original observations included in `probitfe`, and let $M \geq m$

be the number of observations for the entire population declared by the user. The computation of the variance of the APE is corrected by the factor FPC = $(M - m)/(M-1)$. The default is `population(1)`, corresponding to an infinite population. Notice that M references the total number of observations and not the total number of individuals. If, for example, the population has 100 individuals followed over 10 time periods, the user must specify `population(1000)` instead of `population(100)`.

3.5 Stored results

`probitfe` and `logitfe` store the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(N_drop)`	number of observations dropped because of all positive or all zero outcomes
`e(N_group_drop)`	number of groups dropped because of all positive or all zero outcomes
`e(N_time_drop)`	number of time periods dropped because of all positive or all zero outcomes
`e(N_group)`	number of groups
`e(k)`	number of parameters excluding individual or time effects
`e(df_m)`	model degrees of freedom
`e(r2_p)`	pseudo-R-squared
`e(ll)`	log likelihood
`e(ll_0)`	log likelihood, constant-only model
`e(chi2)`	likelihood-ratio χ^2 model test
`e(p)`	significance of model test
`e(rankV)`	rank of `e(V)`
`e(rankV2)`	rank of `e(V2)`
`e(fpc)`	FPC factor
`e(T_min)`	smallest group size
`e(T_avg)`	average group size
`e(T_max)`	largest group size
Macros	
`e(cmd)`	`probitfe` or `logitfe`
`e(cmdline)`	command as typed
`e(depvar)`	name of dependent variable
`e(title)`	title in estimation output
`e(title1)`	type of included effects
`e(title2)`	type of correction
`e(title3)`	lags for trimming parameter or number of multiple partitions
`e(chi2type)`	`LR`; type of model χ^2 test
`e(properties)`	`b V`
`e(id)`	name of cross-section variable
`e(time)`	name of time variable
Matrices	
`e(b)`	coefficient vector
`e(b2)`	APE
`e(V)`	variance–covariance matrix of coefficient vector
`e(V2)`	variance–covariance matrix of APE
Functions	
`e(sample)`	marks estimation sample

4 Bilateral trade flows between countries

4.1 Empirical example

To illustrate the use of the bias corrections described in sections 2.4 and 2.5, we present an empirical application to bilateral trade flows between countries using data from Helpman, Melitz, and Rubinstein (2008). The dataset includes trade flows for 158 countries over the period from 1970 to 1997, as well as country-level data on geography, institutions, and culture (the variables used in the analysis are described below). We fit probit and logit models for the probability of positive trade between country pairs in 1986. The data structure is a pseudopanel where the two dimensions index countries, with `id` as importers and `jd` as exporters. There are $157 \times 156 = 24649$ possible country pairs.[3]

For each country pair, the outcome variable $\texttt{trade}_{ij}$ is an indicator equal to one if country i imports from country j and equal to zero otherwise. We use j instead of t to emphasize that the second dimension does not index time. The model specification is based on the gravity equation of Anderson and van Wincoop (2003) with various measures of trade barriers and enhancers as key determinants of international trade flows. We also include the presence of bilateral trade in 1985 to account for possible state dependence in trade decisions. Importer and exporter country FEs control for unobserved country heterogeneity such as size, natural resources, or trade openness. The probability that country i imports from country j, conditional on the observed variables, X_{ij}, the unobserved importer FE, α_i, and the unobserved exporter FE, γ_j, is modeled as

$$\mathbf{Pr}(\texttt{trade}_{ij} = 1 \mid X_{ij}, \alpha_i, \gamma_j) = F(X_{ij}'\boldsymbol{\beta} + \alpha_i + \gamma_j)$$

where $F(\cdot)$ is the standard normal cumulative distribution function for the probit model, or the logistic distribution for the logit model.

The set of explanatory variables, X_{ij}, includes the following:

1. $\texttt{ltrade}_{ij}$ is a binary variable equal to one if country i imported from country j in 1985 and equal to zero otherwise.
2. $\texttt{ldist}_{ij}$ specifies the logarithm of the distance (in kilometer) between country i and country j capitals.
3. $\texttt{border}_{ij}$ is a binary variable equal to one if country i and country j share a common physical boundary and equal to zero otherwise.
4. $\texttt{legal}_{ij}$ is a binary variable equal to one if country i and country j share the same legal origin (including civil law, common law, customary law, mixed or pluralistic law, and religious law) and equal to zero otherwise.
5. $\texttt{language}_{ij}$ is a binary variable equal to one if country i and country j share the same official language and equal to zero otherwise.

3. The original dataset included 158 countries, but we dropped Congo because it did not export to any country in 1986.

6. $\texttt{colony}_{ij}$ is a binary variable equal to one if country i ever colonized country j or vice versa and equal to zero otherwise.

7. $\texttt{currency}_{ij}$ is a binary variable equal to one if country i and country j use the same currency or if pair money was interchangeable within the country at a 1:1 exchange rate for an extended period of time and equal to zero otherwise.

8. $\texttt{fta}_{ij}$ is a binary variable equal to one if country i and country j belong to a common regional trade agreement and equal to zero otherwise.

9. $\texttt{islands}_{ij}$ is a binary variable equal to one if both country i and country j are islands and equal to zero otherwise.

10. $\texttt{religion}_{ij}$ specifies the sum of (percent Protestants in country $i \times$ percent Protestants in country j) + (percent Catholics in country $i \times$ percent Catholics in country j) + (percent Muslims in country $i \times$ percent Muslims in country j).

11. $\texttt{landlock}_{ij}$ is a binary variable equal to one if both country i and country j have no coastline or direct access to sea and equal to zero otherwise.

The specification of X_{ij} is the same as in table I of Helpman, Melitz, and Rubinstein (2008), except that we include $\texttt{ltrade}_{ij}$. Despite the inclusion of the lag dependent variable, X_{ij} can be treated as strictly exogenous because none of the two dimensions of the panel indexes time.

Tables 1 and 2 show the results of the logit model and probit model, respectively. In both tables, column (1) reports uncorrected FEs estimates, column (2) reports estimates of the analytical correction setting the trimming parameter equal to zero (AN-0), and columns (3) to (5) show estimates of the `ss2`, `jj`, and `double` jackknife corrections. The `double` correction makes sense because both dimensions of the panel index the same set of countries. Each table shows estimates of index coefficients and APE. The latter are reported in brackets. We also include standard errors (SEs) for the index coefficients in column (6) and SEs for the APE in columns (6) and (7). In the case of the APE, the SEs in column (7) are adjusted by the FPC parameter described in section 3.2, using a population equal to the sample size (24,492). There is only one set of SEs because the SEs for the uncorrected estimator are consistent for the corrected estimators (see Fernández-Val and Weidner [2016]).

Table 1. FEs logit model

	(1) FE	(2) AN-0	(3) JK-SS2	(4) JK-JJ	(5) Double	(6) Std. error	(7)
`ltrade`	2.838	2.741	2.786	2.743	2.745	(0.058)	
	[0.325]	[0.323]	[0.349]	[0.325]	[0.326]	(0.014)	(0.008)
`ldist`	−0.839	−0.819	−0.742	−0.812	−0.812	(0.044)	
	[−0.055]	[−0.055]	[−0.049]	[−0.055]	[−0.055]	(0.004)	(0.003)
`border`	−0.571	−0.557	−0.493	−0.564	−0.573	(0.195)	
	[−0.037]	[−0.037]	[−0.036]	[−0.037]	[−0.038]	(0.012)	(0.012)
`legal`	0.115	0.113	0.017	0.112	0.112	(0.062)	
	[0.008]	[0.008]	[0.003]	[0.008]	[0.008]	(0.004)	(0.004)
`language`	0.368	0.358	0.385	0.354	0.352	(0.080)	
	[0.025]	[0.025]	[0.026]	[0.024]	[0.024]	(0.005)	(0.005)
`colony`	0.492	0.435	−0.023	0.344	0.129	(0.633)	
	[0.034]	[0.030]	[0.002]	[0.021]	[0.004]	(0.045)	(0.045)
`currency`	0.984	0.961	2.464	1.009	1.079	(0.252)	
	[0.070]	[0.070]	[0.164]	[0.071]	[0.073]	(0.020)	(0.019)
`fta`	2.244	2.171	3.347	1.827	1.571	(0.657)	
	[0.178]	[0.177]	[0.285]	[0.142]	[0.118]	(0.062)	(0.061)
`islands`	0.406	0.395	0.393	0.396	0.396	(0.156)	
	[0.027]	[0.027]	[0.028]	[0.027]	[0.027]	(0.011)	(0.011)
`religion`	0.244	0.239	0.238	0.240	0.245	(0.123)	
	[0.016]	[0.016]	[0.017]	[0.016]	[0.017]	(0.008)	(0.008)
`landlock`	0.143	0.139	0.153	0.156	0.170	(0.221)	
	[0.010]	[0.010]	[0.014]	[0.010]	[0.011]	(0.015)	(0.015)
Obs.	24492	24492	24492	24492	24492		

Notes: APE in brackets. FE denotes uncorrected FEs estimator; AN-0 denotes analytical correction with 0 lags; JK-SS2 denotes SPJ in both dimensions; JK-JJ denotes delete-one jackknife in both dimensions; Double denotes delete-one jackknife for observations with the same index in the cross-section and the time series. For the APE, the SEs reported in column (7) are adjusted by the FPC parameter using a population equal to the number of observations (24,492).

Table 2. FEs probit model

	(1) FE	(2) AN-0	(3) JK-SS2	(4) JK-JJ	(5) Double	(6) Std. error	(7)
`ltrade`	1.631	1.586	1.625	1.587	1.588	(0.031)	
	[0.343]	[0.345]	[0.371]	[0.346]	[0.347]	(0.014)	(0.009)
`ldist`	−0.438	−0.426	−0.377	−0.423	−0.422	(0.023)	
	[−0.054]	[−0.054]	[−0.046]	[−0.054]	[−0.054]	(0.004)	(0.003)
`border`	−0.273	−0.265	−0.208	−0.268	−0.273	(0.107)	
	[−0.033]	[−0.033]	[−0.029]	[−0.033]	[−0.034]	(0.013)	(0.012)
`legal`	0.059	0.057	0.011	0.056	0.056	(0.033)	
	[0.007]	[0.007]	[0.003]	[0.007]	[0.007]	(0.004)	(0.004)
`language`	0.203	0.198	0.215	0.196	0.196	(0.042)	
	[0.025]	[0.025]	[0.027]	[0.025]	[0.025]	(0.005)	(0.005)
`colony`	0.287	0.253	0.005	0.207	0.099	(0.356)	
	[0.037]	[0.033]	[0.006]	[0.025]	[0.008]	(0.047)	(0.047)
`currency`	0.529	0.515	1.340	0.537	0.568	(0.139)	
	[0.069]	[0.070]	[0.166]	[0.070]	[0.072]	(0.020)	(0.019)
`fta`	1.235	1.192	1.807	1.067	0.991	(0.340)	
	[0.180]	[0.178]	[0.281]	[0.155]	[0.143]	(0.057)	(0.057)
`islands`	0.194	0.187	0.203	0.188	0.188	(0.084)	
	[0.024]	[0.024]	[0.026]	[0.024]	[0.024]	(0.011)	(0.011)
`religion`	0.134	0.132	0.133	0.133	0.135	(0.066)	
	[0.017]	[0.017]	[0.018]	[0.017]	[0.017]	(0.008)	(0.008)
`landlock`	0.041	0.041	0.033	0.044	0.049	(0.119)	
	[0.005]	[0.005]	[0.008]	[0.005]	[0.006]	(0.015)	(0.015)
Obs.	24492	24492	24492	24492	24492		

Notes: APE in brackets. FE denotes uncorrected FEs estimator; AN-0 denotes analytical correction with 0 lags; JK-SS2 denotes SPJ in both dimensions; JK-JJ denotes delete-one jackknife in both dimensions; Double denotes delete-one jackknife for observations with the same index in the cross-section and the time series. For the APE, the SEs reported in column (7) are adjusted by the FPC parameter using a population equal to the number of observations (24,492).

We focus on the results for the logit model. The conclusions from the probit model are analogous, especially in terms of APE, which, unlike index coefficients, are comparable across models. As shown in column (1), the probability that country i imports from country j is higher if country i already imported from country j in the previous year (`ltrade`), if the two countries are closer to each other (`ldist`), if they share the same language (`language`), if they share the same currency (`currency`), if they belong to the same regional free trade agreement (`fta`), if they are not islands (`islands`), or if they share the same religion (`religion`). As in Helpman, Melitz, and Rubinstein (2008), the probability that country i imports from country j decreases if both countries have a common land border (`border`), which they attribute to the effect of territorial border conflicts that suppress trade between neighbors. These effects go in the same direction regardless of the type of correction used. However, there are some differences in the magnitudes of the effects produced by the different estimators.

Comparing across columns, we see that AN-0, JK-JJ, and double produce very similar estimates of index coefficients and APE that are all within one SE of each other. The split-panel correction estimates of the index coefficients and APE of `ldist`, `legal`, `currency`, and `fta` in column (3) are two or more SEs away from the rest of the estimates in the same rows. We show in the next section that JK-SS2 is less accurate than AN-0 and double through a Monte Carlo simulation calibrated to this application. Relative to the uncorrected estimates in column (1), the corrected estimates of the index coefficient of `ltrade` are more than one SE lower. We attribute the similarity in the rest of the index coefficients and APE between uncorrected and bias-corrected estimates partly to the large sample size (except for JK-SS2). Thus we find more significant differences in the next section when we consider subpanels with less than 157 countries.

4.2 Calibrated Monte Carlo simulations

To evaluate the performance of the bias corrections, we conduct a Monte Carlo simulation that mimics the empirical example described above. We focus on the logit model, but we find similar results for the probit model that are not reported here. All the parameters are calibrated to the data used in the previous section, and their values are set to the uncorrected FEs estimates from column (1) in table 1. To speed up computation, we consider only two explanatory variables in X_{ij}: the presence of trade in the previous year (`ltrade`) and the log distance between country pairs (`ldist`).

For all possible country pairs, we first construct the index

$$\texttt{index}_{ij} = \widehat{\beta}_1 \texttt{ltrade}_{ij} + \widehat{\beta}_2 \texttt{ldist}_{ij} + \widehat{\alpha}_i + \widehat{\gamma}_j$$

where $\widehat{\beta}_1 = 2.838$, $\widehat{\beta}_2 = -0.839$, and $\widehat{\alpha}_i$ and $\widehat{\gamma}_j$ are the uncorrected estimates of the importer and exporter FEs (not reported in table 1). Next, we generate a new trade indicator for each country pair as

$$\texttt{trade}^*_{ij} = 1 \times \left\{ \texttt{index}_{ij} > \underbrace{\ln\left(\frac{1}{\text{runiform}(1,1)} - 1\right)}_{(*)} \right\}$$

where ln denotes the natural logarithm and $\text{runiform}(1,1)$ generates a random number from the uniform distribution in $(0,1)$, such that $(*)$ corresponds to a random draw from the standard logistic distribution.

We use the generated trade indicators to estimate the equation

$$\mathbf{Pr}(\texttt{trade}^*_{ij} = 1 \mid \texttt{ltrade}_{ij}, \texttt{ldist}_{ij}, \alpha_i, \gamma_j) = F(\beta_1^* \texttt{ltrade}_{ij} + \beta_2^* \texttt{ldist}_{ij} + \alpha_i^* + \gamma_j^*)$$

where $F(\cdot)$ is the logistic distribution and `ltrade` and `dist` are the variables from the original dataset. We repeat this procedure in 500 simulations for 5 different sample sizes: $N = 25$, $N = 50$, $N = 75$, $N = 100$, and $N = 157$ (full sample). For each sample size and simulation, we draw a random sample of N countries both as importers and exporters without replacement, so that the number of observations is $N \times (N - 1)$.

Table 3 reports the result for the uncorrected estimator (FE), analytical correction setting the trimming parameter equal to zero (AN-0), jackknife correction `ss2` (JK-SS2), and jackknife correction `double` (Double). We analyze the performance of these estimators in terms of bias and inference accuracy of their asymptotic distribution for both index coefficients and APE. In particular, we compute the biases (Bias), SDs (Std. dev.), and root mean squared errors (RMSE) of the estimators together with the ratio of average SEs to the simulation SDs (SE/SD) and the empirical coverages of confidence intervals with 95% nominal level ($p; 0.95$). The variance of the APE is adjusted by the `population(`*integer*`)` option, with the population being equal to the original sample size (24,492 observations). All the results are reported in percentage of the true parameter value.

Table 3. Calibrated Monte Carlo simulations for the logit model

		Index coefficients					APE				
		(1) Bias	(2) Std. dev.	(3) RMSE	(4) SE/SD	(5) $p;95$	(6) Bias	(7) Std. dev.	(8) RMSE	(9) SE/SD	(10) $p;0.95$
A. FE											
$N = 25$	`ltrade`	33.640	24.422	41.556	0.806	0.654	0.662	20.430	20.421	0.925	0.928
	`ldist`	27.470	46.597	54.051	0.857	0.898	−0.211	37.521	37.484	1.094	0.958
$N = 50$	`ltrade`	12.768	8.003	15.065	0.943	0.606	−0.727	10.776	10.790	0.938	0.924
	`ldist`	10.629	17.095	20.115	0.949	0.902	−1.847	17.273	17.354	1.148	0.960
$N = 75$	`ltrade`	7.681	5.132	9.235	0.914	0.614	−1.205	7.130	7.224	0.984	0.944
	`ldist`	6.421	11.101	12.815	0.924	0.896	−1.766	11.869	11.988	1.121	0.962
$N = 100$	`ltrade`	5.615	3.627	6.683	0.939	0.614	−1.184	5.058	5.189	1.042	0.946
	`ldist`	4.308	8.212	9.266	0.915	0.910	−2.003	8.771	8.988	1.132	0.962
$N = 157$	`ltrade`	3.534	2.174	4.148	0.965	0.592	−1.279	2.611	2.905	0.984	0.920
	`ldist`	2.694	4.555	5.288	1.025	0.920	−1.975	4.283	4.712	1.329	0.976
B. AN-0											
$N = 25$	`ltrade`	−1.726	14.631	14.716	1.334	0.986	−2.178	19.547	19.643	0.965	0.922
	`ldist`	−1.672	32.615	32.621	1.219	0.986	2.686	37.116	37.165	1.119	0.961
$N = 50$	`ltrade`	0.142	6.675	6.669	1.131	0.974	−2.352	10.841	11.082	0.931	0.905
	`ldist`	0.570	15.271	15.266	1.062	0.966	−1.276	17.117	17.147	1.158	0.959
$N = 75$	`ltrade`	−0.151	4.592	4.590	1.022	0.964	−2.358	7.174	7.544	0.977	0.929
	`ldist`	0.353	10.355	10.350	0.991	0.956	−1.392	11.783	11.853	1.129	0.961
$N = 100$	`ltrade`	−0.090	3.331	3.329	1.022	0.948	−2.081	5.065	5.471	1.040	0.930
	`ldist`	−0.029	7.804	7.796	0.962	0.934	−1.680	8.729	8.881	1.137	0.974
$N = 157$	`ltrade`	−0.020	2.069	2.067	1.014	0.972	−1.871	2.602	3.203	0.987	0.864
	`ldist`	0.059	4.416	4.412	1.058	0.960	−1.732	4.281	4.614	1.329	0.982

Continued on next page

		Index coefficients					APE				
		(1) Bias	(2) Std. dev.	(3) RMSE	(4) SE/SD	(5) $p;95$	(6) Bias	(7) Std. dev.	(8) RMSE	(9) SE/SD	(10) $p;0.95$
C. JK-SS2											
$N=25$	`ltrade`	−29.035	27.597	40.034	0.672	0.599	4.839	26.728	27.136	0.707	0.844
	`ldist`	−54.245	80.471	96.974	0.485	0.617	−20.435	66.444	69.452	0.618	0.806
$N=50$	`ltrade`	−3.259	7.836	8.479	0.963	0.932	−0.283	11.633	11.625	0.869	0.904
	`ldist`	−6.515	25.274	26.076	0.642	0.792	−5.351	23.101	23.690	0.858	0.890
$N=75$	`ltrade`	−1.158	5.079	5.204	0.924	0.920	−1.167	7.509	7.591	0.934	0.928
	`ldist`	−2.737	15.151	15.382	0.677	0.816	−3.377	15.085	15.444	0.882	0.892
$N=100$	`ltrade`	−0.457	3.575	3.601	0.952	0.922	−1.178	5.403	5.525	0.975	0.936
	`ldist`	−1.751	11.024	11.152	0.681	0.792	−2.928	11.014	11.386	0.901	0.908
$N=157$	`ltrade`	−0.119	2.158	2.159	0.972	0.952	−1.060	2.770	2.964	0.927	0.928
	`ldist`	−0.942	6.132	6.198	0.762	0.850	−2.210	5.821	6.221	0.978	0.916
D. JK-Double											
$N=25$	`ltrade`	−33.612	52.118	61.973	0.378	0.714	1.119	23.181	23.185	0.815	0.876
	`ldist`	−32.292	33.637	46.604	1.188	0.930	−8.335	37.439	38.319	1.096	0.926
$N=50$	`ltrade`	−3.152	6.346	7.080	1.189	0.958	−1.927	10.933	11.091	0.924	0.910
	`ldist`	−3.203	14.739	15.068	1.101	0.954	−2.834	17.118	17.334	1.158	0.960
$N=75$	`ltrade`	−1.286	4.488	4.664	1.045	0.962	−2.012	7.182	7.451	0.977	0.936
	`ldist`	−1.120	10.270	10.321	0.999	0.954	−2.052	11.862	12.027	1.122	0.960
$N=100$	`ltrade`	−0.643	3.288	3.347	1.035	0.946	−1.814	5.073	5.383	1.039	0.932
	`ldist`	−0.821	7.781	7.816	0.965	0.938	−2.108	8.744	8.986	1.135	0.970
$N=157$	`ltrade`	−0.180	2.062	2.068	1.017	0.968	−1.692	2.607	3.105	0.985	0.876
	`ldist`	−0.257	4.423	4.427	1.056	0.960	−1.951	4.305	4.723	1.322	0.976

Notes: FE denotes uncorrected FEs estimator; AN-0 denotes analytical correction with 0 lags; JK-SS2 denotes SPJ in both dimensions; JK-Double denotes delete-one jackknife for observations with the same index in the cross-section and the time-series.

For the uncorrected estimators in panel A, we observe in column (1) that there is significant bias in the index coefficients. This bias decreases with the sample size, but it is still larger than the SD for the coefficient of `ltrade` in the full sample. Moreover, column (5) shows that confidence intervals constructed around the uncorrected estimates suffer from severe undercoverage for all sample sizes. As in Fernández-Val and Weidner (2016), we find very little bias in the APE, despite the large bias in the index coefficients. In panel B, we see that the analytical correction substantially reduces the bias in the index coefficients, producing confidence intervals with coverage close to their nominal level for every sample size. This correction reduces SD, resulting in a reduction of more than 50% in RMSE for several sample sizes. The jackknife corrections also reduce bias and generally improve coverage but increase dispersion in small samples and require larger sample sizes than the analytical corrections to improve RMSE over the uncorrected estimator. The jackknife correction `double` performs similarly to the analytical correction, except for the smallest sample size. The jackknife correction `ss2` of the index coefficient of `ldist` has higher RMSE than the uncorrected estimator even for the full sample size. Overall, the SEs provide a good approximation to the SDs of all the estimators of both the index coefficients and APE.

To summarize, table 3 shows that the analytical correction substantially reduces the bias of the uncorrected estimator, producing more accurate point and interval estimators for all the sample sizes considered. The jackknife correction `double` performs similarly to the analytical correction, except for the smallest sample size $N = 25$. The split-panel correction `ss2` reduces bias but at the cost of increasing dispersion for most sample sizes. In this application, `ss2` is dominated by the other corrections uniformly across all the sample sizes in terms of RMSE. These results are consistent with the empirical evidence in table 1, where the uncorrected estimates of the index coefficient of `ltrade` were more than one SE below the corrected estimates, the estimates of the APE were very similar for the uncorrected and corrected estimators except for `ss2`, and the jackknife correction `ss2` produced estimates for `ldist` at odds with the other estimators.

5 Concluding remarks

The `probitfe` and `logitfe` commands implement the analytical and jackknife bias corrections of Fernández-Val and Weidner (2016) for logit and probit models with two-way FEs. The commands compute estimators of both index coefficients and APE, which are often the parameters of interest in these models. We also provide functionality for models with one-way FEs, offering an alternative to the `clogit` and `xtlogit` commands, which do not produce corrected estimates of APE. Logit and probit models are commonly used in empirical work, making the new commands a valuable addition to the applied econometrician's toolkit. Similar corrections can be implemented for other nonlinear panel models such as tobit models for censored outcome variables. We leave this extension to future research.

6 Acknowledgments

We thank an anonymous reviewer for helpful comments and suggestions. Mario Cruz-Gonzalez gratefully acknowledges support from the National Science Foundation grant No. SES-1060889. Iván Fernández-Val gratefully acknowledges support from the National Science Foundation grant No. SES-1060889. Martin Weidner gratefully acknowledges support from the Economic and Social Research Council through the Centre for Microdata Methods and Practice grant RES-589-28-0001 and also from the European Research Council grant ERC-2014-CoG-646917-ROMIA.

7 References

Andersen, E. B. 1973. *Conditional Inference and Models for Measuring*, vol. 5. Copenhagen: Mentalhygiejnisk forlag.

Anderson, J. E., and E. van Wincoop. 2003. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* 93: 170–192.

Arellano, M., and J. Hahn. 2007. Understanding bias in nonlinear panel models: Some recent developments. In *Advances in Economics and Econometrics: Theory and Applications*, ed. R. Blundell, W. K. Newey, and T. Persson, 381–409. Cambridge: Cambridge University Press.

Chamberlain, G. 1984. Panel data. In *Handbook of Econometrics, vol. 2*, ed. Z. Griliches and M. D. Intriligator, 1247–1318. Amsterdam: Elsevier.

Dhaene, G., and K. Jochmans. 2015. Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82: 991–1030.

Fernández-Val, I. 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150: 71–85.

Fernández-Val, I., and M. Weidner. 2016. Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics* 192: 291–312.

———. 2017. Fixed effects estimation of large T panel data models. Working paper, Boston University.

Hahn, J., and W. Newey. 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72: 1295–1319.

Helpman, E., M. Melitz, and Y. Rubinstein. 2008. Estimating Trade Flows: Trading Partners and Trading Volumes. *Quarterly Journal of Economics* 123: 441–487.

Neyman, J., and E. L. Scott. 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16: 1–32.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Pædagogiske Institut.

About the authors

Mario Cruz-Gonzalez is a PhD candidate in the Department of Economics at Boston University. His research interests include topics in labor economics, development economics, and econometrics.

Iván Fernández-Val is a professor of economics at Boston University. His research interests focus on topics in theoretical and applied econometrics, including panel-data analysis and distributional and quantile methods.

Martin Weidner is a lecturer in economics at University College London and a member of the Centre for Microdata Methods and Practice at the Institute for Fiscal Studies in London. He is working on theoretical and applied econometrics, with a special focus on panel data, factor models, demand estimation, and social networks.

Appendix

A.1. Expressions of the asymptotic bias and variance

Fernández-Val and Weidner (2016) show that the asymptotic bias and variance for $\boldsymbol{\beta}$ can be expressed as

$$\mathbf{B}^{\beta} = \mathbf{W}^{-1}\mathbf{B}, \quad \mathbf{D}^{\beta} = \mathbf{W}^{-1}\mathbf{D}, \quad \mathbf{V}^{\beta} = \mathbf{W}^{-1}$$

where

$$\mathbf{B} = \mathbb{E}\left(-\frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \left\{ H_{it}\partial^2 F_{it} \widetilde{\mathbf{X}}_{it} + 2\sum_{\tau=t+1}^{T} H_{it}(Y_{it} - F_{it})\omega_{i\tau}\widetilde{\mathbf{X}}_{i\tau} \right\}}{\sum_{t=1}^{T} \omega_{it}} \right)$$

$$\mathbf{D} = \mathbb{E}\left(-\frac{1}{2T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} H_{it}\partial^2 F_{it}\widetilde{\mathbf{X}}_{it}}{\sum_{i=1}^{N} \omega_{it}} \right)$$

$$\mathbf{W} = \mathbb{E}\left(\frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T} \omega_{it}\widetilde{\mathbf{X}}_{it}\widetilde{\mathbf{X}}'_{it} \right)$$

$\mathbb{E} := \text{plim}_{N,T\to\infty}$, $\omega_{it} = H_{it}\partial F_{it}$, $H_{it} = \partial F_{it}/\{F_{it}(1-F_{it})\}$,
$\partial^j G_{it} := \partial^j G(Z)|_{Z=\mathbf{X}'_{it}\boldsymbol{\beta}^0+\alpha_i^0+\gamma_t^0}$ for any function G and $j = 0, 1, 2$, and $\widetilde{\mathbf{X}}_{it}$ is the residual of the population projection of $\mathbf{X}_{it}$ on the space spanned by α_i and γ_t under a metric weighted by ω_{it}.

The expressions of the asymptotic bias terms for the APE are different depending on whether the APE are obtained from uncorrected or bias-corrected estimators of $\boldsymbol{\beta}$. The `probitfe` and `logitfe` commands implement the corrections on APE obtained from bias-corrected estimators of the parameters; that is, $\widetilde{\boldsymbol{\delta}}$ is obtained using $\widetilde{\boldsymbol{\beta}}$ equal to the bias-corrected estimator $\widetilde{\boldsymbol{\beta}}^A$ defined below. The expressions for the leading bias terms of $\widetilde{\boldsymbol{\delta}}$ then read

$$\mathbf{B}^{\delta} = \mathbb{E}\left(\frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \left\{ 2\sum_{\tau=t+1}^{T} H_{it}(Y_{it} - F_{it})\omega_{i\tau} \widetilde{\boldsymbol{\Psi}}_{i\tau} + \partial_{\alpha_i^2} \boldsymbol{\Delta}_{it} - \boldsymbol{\Psi}_{it} H_{it} \partial^2 F_{it} \right\}}{\sum_{t=1}^{T} \omega_{it}} \right)$$

$$\mathbf{D}^{\delta} = \mathbb{E}\left(\frac{1}{2T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \left\{ \partial_{\alpha_i^2} \boldsymbol{\Delta}_{it} - \boldsymbol{\Psi}_{it} H_{it} \partial^2 F_{it} \right\}}{\sum_{i=1}^{N} \omega_{it}} \right)$$

where $\boldsymbol{\Psi}_{it}$ and $\widetilde{\boldsymbol{\Psi}}_{it}$ are the fitted value and residual of the population regression of $-\partial_{\pi} \boldsymbol{\Delta}_{it}/\omega_{it}$ on the space spanned by α_i and γ_t under the metric given by ω_{it}. If all the components of $\mathbf{X}_{it}$ are strictly exogenous, the first term of $\mathbf{B}^{\delta}$ is zero. The asymptotic variance of the estimators of $\boldsymbol{\delta}$ is

$$\mathbf{V}^{\delta} = \mathbb{E}\left\{ \frac{r_{NT}^2}{N^2T^2} \sum_{i=1}^{N} \left(\sum_{t,\tau=1}^{T} \widetilde{\boldsymbol{\Delta}}_{it} \widetilde{\boldsymbol{\Delta}}_{i\tau}' + \sum_{j \neq i} \sum_{t=1}^{T} \widetilde{\boldsymbol{\Delta}}_{it} \widetilde{\boldsymbol{\Delta}}_{jt}' + \sum_{t=1}^{T} \boldsymbol{\Gamma}_{it} \boldsymbol{\Gamma}_{it}' \right) \right\}$$

where $r_{NT} = \sqrt{NT/(N+T-1)}$, $\widetilde{\boldsymbol{\Delta}}_{it} = \boldsymbol{\Delta}_{it} - \boldsymbol{\delta}^0$, $\boldsymbol{\Gamma}_{it} = (\mathbf{D}_{\beta}\boldsymbol{\Delta})' W_{\infty}^{-1} H_{it}(Y_{it} - F_{it}) \widetilde{\mathbf{X}}_{it} - \boldsymbol{\Psi}_{it} H_{it}(Y_{it} - F_{it})$ and

$$\mathbf{D}_{\beta}\boldsymbol{\Delta} = \mathbb{E}\left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \partial_{\alpha_i} \boldsymbol{\Delta}_{it} \widetilde{\mathbf{X}}_{it} \right)$$

A.2. Analytical correction

The analytical corrections are implemented using plugin estimators of the bias terms that replace expectations by sample averages and true parameter values by FEs estimators. Thus, for any function of the data, unobserved effects and parameters $g_{it}(\boldsymbol{\beta}, \alpha_i, \gamma_t)$, let $\widehat{g}_{it} = g_{it}(\widehat{\boldsymbol{\beta}}, \widehat{\alpha}_i, \widehat{\gamma}_t)$ denote the FEs estimator of $g_{it} = g_{it}(\boldsymbol{\beta}^0, \alpha_i^0, \gamma_t^0)$; for example, $\widehat{F}_{it} = F(\mathbf{X}_{it}'\widehat{\boldsymbol{\beta}} + \widehat{\alpha}_i + \widehat{\gamma}_t)$ denotes the FEs estimator of $F_{it} = F(\mathbf{X}_{it}'\boldsymbol{\beta}^0 + \alpha_i^0 + \gamma_t^0)$. The `probitfe` and `logitfe` commands with the `analytical` option compute the correction for $\boldsymbol{\beta}$

$$\widetilde{\boldsymbol{\beta}}^{A} = \widehat{\boldsymbol{\beta}} - \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}/T - \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{D}}/N$$

where

$$\widehat{\mathbf{B}} = -\frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \widehat{H}_{it} \partial^2 \widehat{F}_{it} \widehat{\widetilde{\mathbf{X}}}_{it} + 2\sum_{j=1}^{L} \{T/(T-j)\} \sum_{t=j+1}^{T} \widehat{H}_{i,t-j}(Y_{i,t-j} - \widehat{F}_{i,t-j}) \widehat{\omega}_{it} \widehat{\widetilde{\mathbf{X}}}_{it}}{\sum_{t=1}^{T} \widehat{\omega}_{it}}$$

$$\widehat{\mathbf{D}} = -\frac{1}{2T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \widehat{H}_{it} \partial^2 \widehat{F}_{it} \widehat{\widetilde{\mathbf{X}}}_{it}}{\sum_{i=1}^{N} \widehat{\omega}_{it}}$$

$$\widehat{\mathbf{W}} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\omega}_{it} \widehat{\widetilde{\mathbf{X}}}_{it} \widehat{\widetilde{\mathbf{X}}}_{it}'$$

$\widehat{\omega}_{it} = \widehat{H}_{it}\partial\widehat{F}_{it}$, $\widehat{\widetilde{\mathbf{X}}}_{it}$ is the residual of the least-squares projection of $\mathbf{X}_{it}$ on the space spanned by the incidental parameters under a metric weighted by $\widehat{\omega}_{it}$, and L is a trimming parameter for estimation of spectral expectations such that $L \to \infty$ and $L/T \to 0$. The factor $T/(T-j)$ is a degrees of freedom adjustment that rescales the time-series averages $T^{-1}\sum_{t=j+1}^{T}$ by the number of observations instead of by T.

Similarly, the analytical correction for $\boldsymbol{\delta}$ is computed as

$$\widetilde{\boldsymbol{\delta}}^{A} = \widetilde{\boldsymbol{\delta}} - \widehat{\mathbf{B}}^{\delta}/T - \widehat{\mathbf{D}}^{\delta}/N$$

where

$$\widetilde{\boldsymbol{\delta}} = \Delta\left(\widetilde{\boldsymbol{\beta}}^{A}, \widetilde{\boldsymbol{\alpha}}^{A}, \widetilde{\boldsymbol{\gamma}}^{A}\right)$$

$$\left(\widetilde{\boldsymbol{\alpha}}^{A}, \widetilde{\boldsymbol{\gamma}}^{A}\right) \in \operatorname{argmax}_{(\boldsymbol{\alpha},\boldsymbol{\gamma})\in\mathbb{R}^{N+T}} \sum_{i,t} \ell_{it}\left(\widetilde{\boldsymbol{\beta}}^{A}, \alpha_i, \gamma_t\right)$$

$$\widehat{\mathbf{B}}^{\delta} = \frac{1}{2N}\sum_{i=1}^{N} \frac{2\sum_{j=1}^{L}\{T/(T-j)\}\sum_{t=j+1}^{T}\widehat{H}_{i,t-j}(Y_{i,t-j}-\widehat{F}_{i,t-j})\widehat{\omega}_{it}\widehat{\widetilde{\boldsymbol{\Psi}}}_{it} + \sum_{t=1}^{T}\left(\partial_{\alpha_i^2}\widehat{\boldsymbol{\Delta}}_{it} - \widehat{\boldsymbol{\Psi}}_{it}\widehat{H}_{it}\partial^2\widehat{F}_{it}\right)}{\sum_{t=1}^{T}\widehat{\omega}_{it}}$$

$$\widehat{\mathbf{D}}^{\delta} = \frac{1}{2T}\sum_{t=1}^{T} \frac{\sum_{i=1}^{N}\left(\partial_{\alpha_i^2}\widehat{\boldsymbol{\Delta}}_{it} - \widehat{\boldsymbol{\Psi}}_{it}\widehat{H}_{it}\partial^2\widehat{F}_{it}\right)}{\sum_{i=1}^{N}\widehat{\omega}_{it}}$$

A.3. SEs

The SEs for all the estimators (uncorrected or corrected) of the kth component of $\boldsymbol{\beta}$ are computed as

$$\sqrt{\widehat{W}_{kk}^{-1}/(NT)}, \quad k = \{1, \ldots, \dim\boldsymbol{\beta}\}$$

where $\widehat{W}_{kk}^{-1}$ is the (k,k) element of the matrix $\widehat{\mathbf{W}}^{-1}$ defined above, which is based on the uncorrected FE estimator $\widehat{\boldsymbol{\beta}}$. The SEs for all the estimators of the APE are computed as

$$\frac{1}{NT}\left\{\sum_{i=1}^{N}\left(a_{NT}\sum_{t,\tau=1}^{T}\widehat{\widetilde{\boldsymbol{\Delta}}}_{it}\widehat{\widetilde{\boldsymbol{\Delta}}}_{i\tau}' + a_{NT}\sum_{t=1}^{T}\sum_{j\neq i}\widehat{\widetilde{\boldsymbol{\Delta}}}_{it}\widehat{\widetilde{\boldsymbol{\Delta}}}_{jt}' + \sum_{t=1}^{T}\widehat{\boldsymbol{\Gamma}}_{it}\widehat{\boldsymbol{\Gamma}}_{it}'\right)\right\}^{1/2}$$

where $\widehat{\widetilde{\boldsymbol{\Delta}}}_{it} = \widehat{\boldsymbol{\Delta}}_{it} - \widetilde{\boldsymbol{\delta}}$, $\widehat{\boldsymbol{\Gamma}}_{it} = (\mathbf{D}_{\beta}\widehat{\boldsymbol{\Delta}})'\widehat{\mathbf{W}}^{-1}\widehat{H}_{it}(Y_{it}-\widehat{F}_{it})\widehat{\widetilde{\mathbf{X}}}_{it} - \widehat{\boldsymbol{\Psi}}_{it}\widehat{H}_{it}(Y_{it}-\widehat{F}_{it})$ and

$$\mathbf{D}_{\beta}\widehat{\boldsymbol{\Delta}} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\partial_{\alpha_i}\widehat{\boldsymbol{\Delta}}_{it}\widehat{\widetilde{\mathbf{X}}}_{it}$$

The factor a_{NT} is an FPC term,

$$a_{NT} = (N_0T_0 - NT)/(N_0T_0 - 1)$$

where N_0 and T_0 are the population sizes of the two dimensions of the panel. For example, $a_{NT} = 1$ if at least one of the dimensions has infinite size in the population, and $a_{NT} = 0$ if we observe the entire population. The correction affects only the first two terms of the variance because they come from using a sample mean to estimate a population mean, whereas the third term is due to parameter estimation.

A.4. One-way FEs models

In models that include only individual effects, all the expressions of the asymptotic bias and variance are the same as for the two-way FEs models except for

$$\mathbf{D}^{\beta} = \mathbf{0}, \quad \mathbf{D}^{\delta} = \mathbf{0}, \quad \partial^j G_{it} := \partial^j G(Z)|_{Z=\mathbf{X}_{it}'\boldsymbol{\beta}^0+\alpha_i^0}$$

and $\widetilde{\mathbf{X}}_{it}$ is the residual of the population projection of $\mathbf{X}_{it}$ on the space spanned by α_i under a metric weighted by ω_{it}. Symmetrically, in models that include only time effects, all the expressions of the asymptotic bias and variance are the same as for the two-way FEs models except for

$$\mathbf{B}^{\beta} = \mathbf{0}, \quad \mathbf{B}^{\delta} = \mathbf{0}, \quad \partial^j G_{it} := \partial^j G(Z)|_{Z=\mathbf{X}_{it}'\boldsymbol{\beta}^0+\gamma_t^0}$$

and $\widetilde{\mathbf{X}}_{it}$ is the residual of the population projection of $\mathbf{X}_{it}$ on the space spanned by γ_t under a metric weighted by ω_{it}.

We do not provide explicit expressions for the analytical bias corrections and SEs because they are analogous to the expressions given in sections A.2 and A.3. For the jackknife, in models that include only individual effects, the following apply:

- The corrections `ss1`, `ss2`, and `sj` implement the SPJ of Dhaene and Jochmans (2015), which applies SPJ to the individual dimension; that is,

$$\widetilde{\boldsymbol{\beta}}^{\texttt{ss1}} = \widetilde{\boldsymbol{\beta}}^{\texttt{ss2}} = \widetilde{\boldsymbol{\beta}}^{\texttt{sj}} = 2\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_{N,T/2}$$

- The corrections `js`, `jj`, and `double` implement the jackknife correction of Hahn and Newey (2004), which applies PJ to the individual dimension; that is,

$$\widetilde{\boldsymbol{\beta}}^{\texttt{js}} = \widetilde{\boldsymbol{\beta}}^{\texttt{jj}} = \widetilde{\boldsymbol{\beta}}^{\texttt{double}} = N\widehat{\boldsymbol{\beta}} - (N-1)\widetilde{\boldsymbol{\beta}}_{N-1,T}$$

Similarly, in models that include only time effects, the following apply:

- The corrections `ss1`, `ss2`, and `js` implement the SPJ of Dhaene and Jochmans (2015), which applies SPJ to the time dimension; that is,

$$\widetilde{\boldsymbol{\beta}}^{\texttt{ss1}} = \widetilde{\boldsymbol{\beta}}^{\texttt{ss2}} = \widetilde{\boldsymbol{\beta}}^{\texttt{js}} = 2\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}}_{N/2,T}$$

- The corrections `sj`, `jj`, and `double` implement the jackknife correction of Hahn and Newey (2004), which applies PJ to the time dimension; that is,

$$\widetilde{\boldsymbol{\beta}}^{\texttt{sj}} = \widetilde{\boldsymbol{\beta}}^{\texttt{jj}} = \widetilde{\boldsymbol{\beta}}^{\texttt{double}} = T\widehat{\boldsymbol{\beta}} - (T-1)\widetilde{\boldsymbol{\beta}}_{N,T-1}$$