

Estimation of Nonlinear Panel Models with Multiple Unobserved Effects*

Mingli Chen[†]

Job Market Paper

November 16, 2014

Abstract

I propose a fixed effects expectation-maximization (EM) estimator that can be applied to a class of nonlinear panel data models with unobserved heterogeneity, which is modeled as individual effects and/or time effects. Of particular interest is the case of interactive effects, i.e. when the unobserved heterogeneity is modeled as a factor analytical structure. The estimator is obtained through a computationally simple, iterative two-step procedure, where the two steps have closed form solutions. I show that estimator is consistent in large panels and derive the asymptotic distribution for the case of the probit with interactive effects. I develop analytical bias corrections to deal with the incidental parameter problem. Monte Carlo experiments demonstrate that the proposed estimator has good finite-sample properties. I illustrate the use of the proposed model and estimator with an application to international trade networks.

Keywords: Nonlinear panel, latent variables, interactive effects, factor error structure, EM algorithm, incidental parameters, bias correction

JEL Classification: C13, C21, C22

*I would like to thank Victor Chernozhukov, Iván Fernández-Val, Pierre Perron and Marc Rysman for guidance, support and encouragement on this project. I would also like to thank Manuel Arellano for helpful suggestions and for pointing out this paper's estimator as a new EM-type estimator. I am grateful to Brad Larsen, Whitney Newey and Martin Weidner for very helpful discussions. This paper also benefited from conversations with Jushan Bai, Hiroaki Kaido, Ao Li, Yuan Liao, and Zhongjun Qu.

[†]Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215 (mlchen@bu.edu)

1 Introduction

Panel data allow the possibility of controlling for unobserved heterogeneity. Such heterogeneity can be an important phenomenon, and failure to control for it can result in misleading inference. For example, in demand estimation, unobserved individual heterogeneity is an important source of variation.

In this paper, I model unobserved heterogeneity as individual-specific effects to control for individual heterogeneity, and/or time specific effects to control for common shocks that occur to each individual. The way I control for those individual and time effects in nonlinear models is to treat each effect as a separate parameter to be estimated, and I propose a fixed effects expectation-maximization (EM) estimator that can be applied to a class of nonlinear panel data models with those individual and/or time effects. Of particular interest is the case of interactive effects, i.e., when the unobserved heterogeneity is modeled as a factor analytical structure. To the best of the author's knowledge, the current paper presents the first fixed effects EM-type estimator for nonlinear panel data models.

Interactive effects relax the invariant heterogeneity assumption and allow a more general model of time-varying heterogeneity. These interactive effects can be arbitrarily correlated with the observable covariates, which accommodates endogeneity and, at the same time, allows correlations between individual effects. As an example of why these interactive effects are important, Moon et al. (2014), in a demand estimation setting, demonstrate that interactive fixed effects can capture strong persistence in market shares across products and markets, and find evidence that the factors are indeed capturing much of the unobservable product and time effects leading to price endogeneity.

The nonlinear panel data models with unobserved fixed effects that I consider in this paper have the following latent representation:

$$Y_{it}^* = X_{it}'\beta + g(\alpha_i, \gamma_t) + \varepsilon_{it}, \quad (1)$$

$$Y_{it} = r(Y_{it}^*), \quad (2)$$

for $t = 1, \dots, T$ and $i = 1, \dots, N$. The econometrician observes Y_{it} , the dependent variable for individual i at time t (or t can be a group), and X_{it} , the time-variant $K \times 1$ regressor matrix. The econometrician does not observe Y_{it}^* (the latent dependent variable), α_i (the unobserved time-invariant individual effect), γ_t (the unobserved time effect), or ε_{it} (the unobserved error term). The vector β contains the main structural parameters of interest. The function $r(\cdot)$ is

a known transformation of the unobserved latent variable. The individual effects α_i and time effects γ_t are allowed to be correlated with the regressor matrix. I do not make parametric assumptions on the distribution of either individual effects or time effects, hence the model is semiparametric.¹ The method proposed here can be applied to many functional forms between α_i and γ_t . The leading case I consider is when $g(\alpha_i, \gamma_t) = \alpha_i' \gamma_t$ where both α_i and γ_t are $R \times 1$ vectors; note that this includes the special case settings with only individual effects or settings with additive individual and time effects.

Substantial theoretical and computational challenges are present in nonlinear panel models involving a large number of individual and time effects. In particular, in these models it is in general not possible to remove the unobserved effects by differencing as is commonly done in linear models. The incidental parameter problem, first pointed out by Neyman and Scott (1948), may also be present due to the fact that an estimator of β will be a function of the estimators of α_i and γ_t , which converges to their limits at slower convergence rates than that of β .

To deal with these problems, I propose a fixed effects expectation-maximization (EM) type estimator, which I denote IF-EM when applied to the interactive effects case. The estimator is obtained through an iterative two-step procedure, where the two steps have closed-form solutions. The first step (the “E”-step) involves obtaining the expectation of the mean utility function (the latent index) conditional on the observed dependent data.² The second step (the “M”-step) involves maximizing the resulting “linear” model. In practice, the estimator is simple and straightforward to compute. Monte Carlo simulations demonstrate it has good small-sample properties.

The incidental parameters problem might be present because estimates of fixed effects are partially consistent, and structural parameters of interest are functions of these estimates.³ For example, I discuss a panel probit model with interactive fixed effects (which I denote PPIF) and demonstrate that its estimator PPIF is biased. I develop analytical bias corrections to deal with the incidental parameter problem. The correction is based on adapting to my setting the general asymptotic expansion of fixed effects estimators with incidental

¹Relaxing parametric assumptions on the distribution of unobserved heterogeneity in nonlinear models is important, as often such restrictions cannot be justified by economic theory.

²As shown later, this is essentially an inverse distribution approach. For the exponential class of distributions, under Bregman loss, the conditional expectation is optimal in terms of MSE.

³The incidental parameters problem has different effects in different contexts and might not be present in some nonlinear models, e.g., Poisson models or slope coefficients in Tobit models. Additionally, marginal effects in probit models with individual fixed effects might not have bias or might have small bias, as shown in Fernandez-Val (2009).

parameters in multiple dimensions under asymptotic sequences where both dimensions of the panel grow with the sample size (as in Fernández-Val and Weidner (2014)). In addition to model parameters, I provide bias corrections for average partial effects, which are functions of the data, parameters, and individual and time effects in nonlinear models.

The proposed model and estimates can have wide applications in economics. For example, factor structures have been used in a probit setting to represent market structure (as in Elrod and Keane (1995)) or, in a linear setting, to explain labor and behavioral outcomes (Heckman et al. (2006)) or estimate the evolution of cognitive and noncognitive skills (Cunha and Heckman (2008); Cunha et al. (2010)). Another example where the fixed effects approaches are used is the international trade partner choice (as in Helpman et al. (2008)). The estimator is also particularly useful in empirical finance and in the setting with long time series, such as empirical work using PSID data. Furthermore, the estimation procedure can easily be extended to multinomial choice models.

This paper is related to multiple strands of the literature. First, it is related to the literature on linear panel data models with factor structures. Bai (2009a) estimates factors using the method of principal components. Moon et al. (2014) extend the standard BLP random coefficients discrete choice demand model and propose a two-step procedure to calculate the estimator. Other related papers include Holtz-Eakin et al. (1988); Ahn et al. (2001); Bai and Ng (2002); Bai (2003); Ahn et al. (2013); Andrews (2005); Pesaran (2006); Bai (2009b); Moon and Weidner (2010a), and Moon and Weidner (2010b). Some of these papers (e.g. Bai (2009b)) let $N \rightarrow \infty$ and $T \rightarrow \infty$ while others (e.g. Ahn et al. (2013)) have T fixed and $N \rightarrow \infty$.

This paper is also related to the literature on nonlinear panel data models and bias correction, such as Arellano and Hahn (2007); Hahn and Newey (2004); Hahn and Kuersteiner (2002); Fernández-Val (2009); Bester and Hansen (2009); Carro (2007); Fernández-Val and Vella (2011); Bonhomme (2012); Chamberlain (1980), and Dhaene and Jochmans (2010). Charbonneau (2012) extends the conditional fixed effects estimators to logit and Poisson models with exogenous regressors and additive individual and time effects. Fernández-Val and Weidner (2014) develop analytical and jackknife bias corrections for nonlinear panel data models with additive individual and time effects. Freyberger (2012) studies nonparametric panel data models with multidimensional, unobserved individual effects when T is fixed. Chen et al. (2013) develop analytical and jackknife estimators for a class of nonlinear panel data models with individual and time effects which enter the model interactively.

A final contribution of this paper is on the computation front, relating to the EM algo-

rithm and latent backfitting procedure. Related work includes Orchard et al. (1972); Dempster et al. (1977); Pan (2002); Meng and Rubin (1993); Laird (1985), and Pastorello et al. (2003).

The remainder of the paper is structured as follows. Section 2 introduces the model, the leading examples and their estimators. I also discuss the convergence of the estimation procedure. Section 3 presents consistency and asymptotic results for probit with interactive fixed effects. Section 4 presents some extensions and discussions. Section 5 contains Monte Carlo simulation results and Section 6 presents the empirical examples. Section 7 concludes. All proofs are contained in the Appendix.

2 Models and Estimators

In this section, I start with the panel probit with interactive individual and time effects case. I first specify the model and present the parameters and functional of interest and then show how the model can be estimated using the proposed EM procedure.

2.1 Panel probit with interactive fixed effects (PPIF)

2.1.1 Model

I consider the following interactive fixed effects probit model

$$\begin{aligned} Y_{it}^* &= X_{it}'\beta + \alpha_i'\gamma_t + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \tag{3}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. Here, Y_{it} is a scalar outcome variable of interest, X_{it} is a vector of explanatory variables, and β is a finite dimensional parameter vector. The variables α_i and γ_t are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks, respectively. The model is semiparametric in that I do not specify the distribution of these effects nor their relationship with the explanatory variables, but, given that I consider probit in this section, I do specify ε to be normally distributed with unit variance.

Denoting the cumulative distribution function of ε_{it} as $\Phi(\cdot)$, the standard normal distribution, the conditional distribution of Y_{it} can then be written using the single-index specification

$$P(Y_{it} = 1 | X_{it}, \beta, \alpha_i, \gamma_t) = \Phi(X_{it}'\beta + \alpha_i'\gamma_t).$$

For estimation, I adopt a fixed effects approach, treating the unobserved individual and time effects as parameters to be estimated. I collect all these effects in the vector $\phi_{NT} = (\alpha_1, \dots, \alpha_N, \gamma_1, \dots, \gamma_N)'$. The model parameter β usually includes regression coefficients of interest, while the unobserved effects ϕ_{NT} are treated as nuisance parameters. The true values of the parameters are denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, \dots, \alpha_N^0, \gamma_1^0, \dots, \gamma_T^0)'$. Other quantities of interest involve averages over the data and unobserved effects, such as average partial effects, which are often the ultimate quantities of interest in nonlinear models. These can be denoted

$$\delta_{NT}^0 = \mathbb{E}_\phi[\Delta_{NT}(\beta^0, \phi_{NT}^0)], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \beta, \alpha'_i \gamma_t), \quad (4)$$

where $\Delta(X_{it}, \beta, \alpha'_i \gamma_t)$ represents some partial effect of interest and \mathbb{E}_ϕ denotes the expectation with respect to the distribution of the data, conditional on ϕ_{NT}^0 and β^0 .

Some examples of partial effects are the following:

Example 2.1. (Average partial effects) If $X_{it,k}$, the k -th element of X_{it} , is binary, its partial effect for model specified by (3) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \beta, \alpha'_i \gamma_t) = \Phi(\beta_k + X'_{it,-k} \beta_{-k} + \alpha'_i \gamma_t) - \Phi(X'_{it,-k} \beta_{-k} + \alpha'_i \gamma_t), \quad (5)$$

where β_k is the k -th element of β , and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the k -th element. If $X_{it,k}$ is continuous, the partial effects of $X_{it,k}$ for model (3) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \alpha_i, \gamma_t) = \beta_k \phi_f(X'_{it} \beta + \alpha'_i \gamma_t), \quad (6)$$

here $\phi_f(\cdot)$ is the derivative of Φ .

A particular application of this model is the study of international trade partner choice. For example, Helpman et al. (2008) consider panel of unilateral trade flows between 158 countries for the year 1986. They use a probit model for the extensive margin of a gravity equation with exporter and importer country effects to allow for asymmetric trade.

Example 2.2. (International Trade)

$$Pr(\text{Trade}_{ij} = 1 | X_{ij}, \alpha_i \gamma_j) = \Phi(X'_{ij} \beta + \alpha'_i \gamma_j), \quad \forall i, j \in V, \quad i \neq j,$$

here V contains the identities of all the countries considered.

Here $Trade_{ij}$ is an indicator for positive trade from country j to country i , X_{ij} includes log of bilateral distance, and nine indicators for geography, institution and culture differences.⁴ In this setting, $N \approx T$. The estimated fixed effects can be used for forecasting network linkages or calculating average partial effects as well.

2.1.2 Estimator for panel probit with interactive fixed effects

In this section, I describe how the model with interactive fixed effects can be estimated using the proposed EM procedure. I discuss the case where the model has a known number of factors R .⁵ I will start with $R = 1$; the case for $R > 1$ will be discussed in Section 4. For full identification, I assume $\gamma_1 = 1$, though different normalization restrictions can be imposed and will require different maximization steps, but this does not affect the estimation of β as the factor structure enters into the model jointly as $\alpha_i \gamma_t$.

Definition 2.1. (PPIF) The EM procedure for estimating the panel probit model with interactive fixed effects is as follows:

- (1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X'_{it}\beta^{(k)} + \alpha_i^{(k)}\gamma_t^{(k)}$,
- (2) **E-step:** Calculate

$$\begin{aligned} \hat{Y}_{it}^{(k)} &: = E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)}) / \{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)}))\}, \end{aligned}$$

- (3) **M-step:** This contains three conditional maximization (CM) steps
 CM-step 1: Given α_i and γ_t , the parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it} X'_{it} \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it} \left(\hat{Y}_{it}^{(k)} - \alpha_i^{(k)} \gamma_t^{(k)} \right) \right\},$$

- CM-step 2: Given β and γ_t , the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \left\{ \sum_{t=1}^T (\hat{Y}_{it}^{(k)} - X'_{it}\beta^{(k+1)})\gamma_t^{(k)} \right\} / \sum_{t=1}^T \left\{ \gamma_t^{(k)} \right\}^2,$$

⁴See Helpman et al. (2008) for additional details.

⁵Choosing the number of factors is beyond the scope of this paper.

CM-step 3: Given β and α_i , the parameter γ_t can be updated by

$$\gamma_t^{(k+1)} = \left\{ \sum_{i=1}^N (\hat{Y}_{it}^{(k)} - X'_{it}\beta^{(k+1)})\alpha_i^{(k+1)} \right\} / \sum_{i=1}^N \left\{ \alpha_i^{(k+1)} \right\}^2,$$

(4) Iterate the above steps until convergence.

Convergence and consistency, along with the asymptotic distribution of β will be discussed in the next sections.

Note that the estimation procedure can be adapted to linear panel data models with interactive fixed effects, e.g. Bai (2009b). In a linear panel data model, Y^* is observed, and hence the E-step described here will not be needed. However, the conditional maximization procedure can still be applied to estimate a linear model.

The EM procedure proposed here is simple, easy to implement and has closed-form solutions in each step. The conditional maximization steps involves replacing the functional of the current estimates of the other parameters.⁶

Remark 2.1. Different normalizations for the individual and time effects can lead to different estimation procedures, even for linear models. For example, with the normalization $\gamma_1 = 1$, the linear panel data model with interactive fixed effects

$$Y_{it} = X'_{it}\beta + \alpha_i\gamma_t + \varepsilon_{it},$$

can be estimated as follows

CM-step 1: Given α_i and γ_t , the parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it}X'_{it} \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it} \left(Y_{it} - \alpha_i^{(k)}\gamma_t^{(k)} \right) \right\},$$

CM-step 2: Given β and γ_t , the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \left\{ \sum_{t=1}^T (Y_{it} - X'_{it}\beta^{(k+1)})\gamma_t^{(k)} \right\} / \sum_{t=1}^T \left\{ \gamma_t^{(k)} \right\}^2,$$

⁶This is an expectation and conditional maximization (ECM) procedure, see Meng and Rubin (1993) for more details about ECM.

CM-step 3: Given β and α_i , the parameter γ_t can be updated by

$$\gamma_t^{(k+1)} = \left\{ \sum_{i=1}^N (Y_{it} - X'_{it}\beta^{(k+1)})\alpha_i^{(k+1)} \right\} / \sum_{i=1}^N \left\{ \alpha_i^{(k+1)} \right\}^2,$$

Iterate until convergence.

Since individual effects and additive individual and time effects are special cases of interactive effects, I will present results for the individual effects case only.⁷ For the case with additive individual and time effects, see Appendix A.1.

2.2 Panel probit with only individual fixed effects

In this setting, I consider the following model

$$\begin{aligned} Y_{it}^* &= X'_{it}\beta + \alpha_i + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \tag{7}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. Here, Y_{it} is a scalar outcome variable of interest, X_{it} is a vector of explanatory variables, β is a finite-dimensional parameter vector, α_i are unobserved individual effects.

Similarly to Section (2.1), I model the conditional distribution of Y_{it} using the single-index specification

$$P(Y_{it} = 1 | X_{it}, \beta, \alpha_i) = \Phi(X_{it}\beta + \alpha_i),$$

and for estimation I adopt a fixed effects approach treating the unobserved individual effects as parameters to be estimated. I collect all these effects in the vector $\phi_{NT} = (\alpha_1, \dots, \alpha_N)'$. The true values of the parameters are denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, \dots, \alpha_N^0)'$. Other quantities of interest involve averages over the data and unobserved effects

$$\delta_{NT}^0 = \mathbb{E}[\Delta_{NT}(\beta^0, \phi_{NT}^0)], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \beta, \alpha_i), \tag{8}$$

and examples of partial effects (Δ) are the following:

⁷More precisely, when the unobserved individual and time effects are multidimensional, the additive individual and time effects case is a special case of the interactive effects case.

Example 2.3. (Average partial effects) If $X_{it,k}$, the k -th element of X_{it} , is binary, its partial effect for model (7) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \beta, \alpha_i) = \Phi(\beta_k + X'_{it,-k}\beta_{-k} + \alpha_i) - \Phi(X'_{it,-k}\beta_{-k} + \alpha_i), \quad (9)$$

where β_k is the k -th element of β , and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the k -th element. If $X_{it,k}$ is continuous, for model (7) the partial effects of $X_{it,k}$ on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \alpha_i) = \beta_k \phi_f(X'_{it}\beta + \alpha_i), \quad (10)$$

where $\phi_f(\cdot)$ is the derivative of Φ .

Definition 2.2. The fixed effects EM estimator for panel probit with individual fixed effects is defined by

- (1) Given initial $(\beta^{(k)}, \alpha_i^{(k)})$, denote $\mu_{it}^{(k)} = X'_{it}\beta^{(k)} + \alpha_i^{(k)}$,
- (2) **E-step:** Calculate

$$\begin{aligned} \hat{Y}_{it}^{(k)} : &= E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)}) / \{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)}))\}, \end{aligned}$$

- (3) **M-step:** This contains two conditional maximization steps

CM-step 1: Given α_i , the parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it} X'_{it} \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it} (\hat{Y}_{it}^{(k)} - \alpha_i^{(k)}) \right\},$$

CM-step 2: Given β , the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \frac{1}{T} \sum_{t=1}^T (\hat{Y}_{it}^{(k)} - X'_{it}\beta^{(k+1)}),$$

- (4) Iterate until converge.

This is essentially the case $\gamma_t = 1, \forall t = 1, \dots, T$. Note that the CM-step 2 here is just the average over time using $\hat{Y}_{it}^{(k)}$ as surrogate for Y_{it}^* . This estimation procedure does not involve computing the inverse of the Hessian, unlike the Netwon's method described in Greene (2004).

2.3 Nonlinear panel models with multiple unobserved effects

In this section, I describe how a general nonlinear panel data model with individual and time effects can be estimated using the proposed EM procedure.

Definition 2.3. The fixed effect EM estimator for a class of nonlinear panel data model with individual and time effects is defined by

- (1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$;
- (2) **E-step:** calculate $\hat{Y}_{it}^{(k)} := E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, g(\alpha_i^{(k)}, \gamma_t^{(k)})]$,
- (3) **M-step:**

$$(\beta^{(k+1)}, \alpha^{(k+1)}, \gamma^{(k+1)}) \in \arg \min_{\beta, \alpha, \gamma} S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)}) = (\hat{Y}_{it}^{(k)} - X'_{it}\beta - g(\alpha_i, \gamma_t))^2, \quad (11)$$

- (4) Iterate until convergence.

Convergence and consistency of $\hat{\beta}$, defined as the output from the iteration, will be discussed in the following sections. Note that this procedure is different from the traditional EM algorithm (discussed in Dempster et al. (1977)), which is used to maximize the expected log-likelihood function when there are latent variables, and its E-step is to augment the incomplete likelihood with conditional likelihood for $Y_{it}^* | Y_{it}$; while here, the E-step is to calculate a surrogate, \hat{Y}_{it} , for the unobserved Y_{it}^* when there are unobserved individual and time effects. This difference leads to a different strategy of proof. Specifically, I adopt the approach of using the conditional expectation of Y_{it}^* because under Bregman loss the conditional expectation is optimal in terms of mean squared error. Under certain conditions, e.g., the density of the error term is in the exponential class of distributions, as shown in Section 3, as well as for probit, those two have the same score functions. This is due to the quadratic loss function of the probit model.

Remark 2.2. Depending on the functional form of the individual and/or time effects, the M-step can be as follows:

CM-step 1: Given α_i and γ_t , the parameter β is updated via

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it} X'_{it} \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it} (\hat{Y}_{it}^{(k)} - g(\alpha_i^{(k)}, \gamma_t^{(k)})) \right\},$$

CM-step 2: Given β , the parameters α_i and γ_t are updated by maximizing

$$-\sum_{i=1}^N \sum_{t=1}^T (\hat{Y}_{it}^{(k)} - X'_{it}\beta - g(\alpha_i^{(k)}, \gamma_t^{(k)}))^2,$$

and this step can be implemented by using the method of least squares (or principal components).

2.3.1 Convergence

In this section, I show the resulting estimate from the estimation procedure converges to a point that maximizes the observed log-likelihood function. I focus on the interactive fixed effects case, which is more complex due to the high degree of nonlinearity of the unobserved effects term (all the other cases are concave in the fixed effects, though the convergence rates are different). Consistency results are discussed in Section 4. The IF-EM for probit suffers from asymptotic bias because the fixed effects converge slowly, which I address in Section 3.

For a binary model, denote the negative log-likelihood function

$$-\mathcal{L}_{NT} = -\sum_{i,t} \log F(q_{it}(X'_{it}\beta + \alpha'_i\gamma_t)),$$

where $q_{it} := 2Y_{it} - 1$ and F is the cdf of Y_{it} conditional on X_{it}, α_i and γ_t . For brevity, assume F is symmetric. Define the hazard function $h(\theta_1) := -\partial \log F(\theta_1) / \partial \theta_1$ for a particular argument θ_1 .

Recall the quadratic loss function $S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)}) = (\hat{Y}_{it}^{(k)} - X'_{it}\beta - g(\alpha_i, \gamma_t))^2$ of the M-step that the proposed fixed effects EM-type estimator depends on. The strategy of the proof is to show the negative log likelihood function of the model under consideration is majorized by this quadratic function (up to some constant), which is satisfied by the following propositions

Proposition 2.1. *Suppose X is a three-dimensional matrix with p sheets ($N \times T \times p$), β and $\tilde{\beta}$ are $p \times 1$ vectors, α and $\tilde{\alpha}$ are $N \times R$ matrices, and γ and $\tilde{\gamma}$ are $T \times R$ matrices. Define $\tilde{h}_{it} := h(q_{it}(X'_{it}\tilde{\beta} + \tilde{\alpha}'_i\tilde{\gamma}_t))$, then*

$$-\mathcal{L}_{NT}(\beta, \alpha, \gamma) \leq -\mathcal{L}_{NT}(\tilde{\beta}, \tilde{\alpha}, \tilde{\gamma}) - \frac{1}{2} \sum_{i,t} \tilde{h}_{it}^2 + \frac{1}{2} \sum_{i,t} (\tilde{z}_{it} - X'_{it}\beta - \alpha'_i\gamma_t)^2.$$

Proof: See Appendix A.2.

Proposition 2.2. (i) Up to a constant that depends on $(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$ but not on (β, α, γ) , the function $S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$ majorizes $-\mathcal{L}_{NT}(\beta, \alpha, \gamma)$ at $(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$.

(ii) Let $(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$, $k = 1, 2, \dots$, be a sequence obtained by the IF-EM procedure. Then $S(\beta^{(k)}, \alpha^{(k)}, \gamma^{(k)})$ decreases as k increases and converges to a local minimum of $-\mathcal{L}_{NT}(\beta, \alpha, \gamma)$ as k goes to infinity.

The proof of part (i) follows by applying the result from Proposition 2.1. The proof of part (ii) follows from the property of the quadratic majorization.

This proves the convergence of the general EM procedure. Note that although I show it for an interactive fixed effects model, the same proof procedure can be adapted to other single index models with individual and time fixed effects. I discuss consistency in Section 4. Since the asymptotic distribution differs for different models, in the next section I will show the asymptotic distribution for the probit model, in which case the incidental parameter problem occurs, and provide an analytical bias correction solution.

The EM procedure proposed here is simple, easy to implement, and has a closed form solution in each step. The method can be extended in a straightforward way to handle composite data which consists of both binary and continuous variables. While the binary variables are modeled with Bernoulli distributions, the continuous variables can be modeled with Gaussian distributions. Including some continuous variables corresponds to adding some Gaussian log-likelihood terms to the existing log-likelihood expression. Since the Gaussian log-likelihood is quadratic, the ultimate function would still be majorized by a quadratic function.⁸

3 Asymptotic theory for panel probit with interactive fixed effects

In this section, I discuss consistency and asymptotic bias of the proposed estimator. I do so in the context of PPIF but my method of proof can be extended to a wider class of models.

3.1 Consistency

I show PPIF is consistent but suffers from incidental parameters bias. I will also discuss bias corrections to the parameter and average partial effects in the next section.

I consider a panel probit model with scalar individual and time effects that enter the likelihood function interactively through $\pi_{it} = \alpha_i \gamma_t$. In this model, the dimension of the

⁸When there are no fixed effects, convergence is proved by the contraction mapping theorem argument. See Gourieroux et al. (1987)

incidental parameters is $\dim \phi_{NT} = N + T$. I prove the consistency of PPIF under assumptions on the indexes. Since the proposed fixed effects EM estimator has the same score as that of MLE, I derive its properties directly through the expansion of the score of its profile likelihood function.

In this section, the parametric part of the model takes the form

$$\log \Phi(q_{it}(X'_{it}\beta + \pi_{it})) = \ell_{it}(\beta, \pi_{it}).$$

Hence, the log-likelihood function is

$$\mathcal{L}_{NT}(\beta, \phi_{NT}) = \mathcal{L}_{NT}(\beta, \pi) = \frac{1}{NT} \sum_{i,t} \ell_{it}(\beta, \pi) = \frac{1}{NT} \sum_{i,t} \log \Phi(q_{it}(X'_{it}\beta + \pi_{it})).$$

I make the following assumptions:

Assumption 1. *Let $v > 0$ and $\mu > 4(8 + v)/v$. Let $\varepsilon > 0$ and let $\mathcal{B}_\varepsilon^0$ be a subset of $\mathbb{R}^{\dim \beta + 1}$ that contains an ε -neighborhood of (β^0, π_{it}^0) for all i, t, N, T .*

(i) *Asymptotics: Consider limits of sequences where $N/T \rightarrow \kappa^2$, $0 < \kappa < \infty$, as $N, T \rightarrow \infty$.*

(ii) *Sampling: Conditional on ϕ , $\{(Y_i^T, X_i^T) : 1 \leq i \leq N\}$ is independent across i , and for each i , $\{Y_{it}, X_{it} : 1 < t \leq T\}$ is α -mixing with mixing coefficients satisfying $\sup_i a_i(m) = O(m^{-\mu})$ as $m \rightarrow \infty$, where*

$$a_i(m) := \sup_t \sup_{A \in \mathcal{A}_t^i, B \in \mathcal{B}_{t+m}^i} |P(A \cap B) - P(A)P(B)|$$

and for $Z_{it} = (Y_{it}, X_{it})$, \mathcal{A}_t^i is the sigma field generated by $(Z_{it}, Z_{i,t-1}, \dots)$, and \mathcal{B}_t^i is the sigma field generated by $(Z_{it}, Z_{i,t+1}, \dots)$.

(iii) *Moments: The partial derivatives of $\ell_{it}(\beta, \pi)$ w.r.t. the elements of (β, π) up to fourth order are bounded in absolute value uniformly over $(\beta, \pi) \in \mathcal{B}_\varepsilon^0$ by a function $M(Z_{it}) > 0$ a.s., and $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+v}]$ is a.s. uniformly bounded over N, T . There exist constants b_{\min} and b_{\max} such that for all $(\beta, \pi) \in \mathcal{B}_\varepsilon^0$, $0 < b_{\min} \leq -\mathbb{E}_\phi[\partial_{\pi^2} \ell_{it}(\beta, \pi)] \leq b_{\max}$ a.s. uniformly over i, t, N, T .*

(iv) *Non-colinearity condition: $\exists c > 0$, such that w.p.a.1,*

$$\min_{\{\alpha \in \mathbb{R}, \|\alpha\|=1\}} \min_{\Lambda \in \mathbb{R}^{N \times 2}} \frac{1}{NT} \text{Tr}[(\alpha \cdot X)' M_\alpha (\alpha \cdot X)] > c$$

Assumption (i) defines the large- T asymptotic framework as in Hahn and Kuersteiner (2002); Fernández-Val and Weidner (2014); Chen et al. (2013). Assumption (ii) defines the data sampling conditions. Assumption (iii) defines the finite moment condition. Assumption (iv) states that no linear combination of the regressors converges to zero, even after projecting any two-dimensional factor loading α . Note that this rules out time-invariant and cross-sectional invariant regressors.

Define the fixed effects EM estimator for PPIF as $\hat{\beta}_{PPIF}$.

Lemma 3.1. *Under Assumption 1, $\hat{\beta}_{PPIF} = \beta^0 + o_P(1)$.*

The proof is found in Appendix B.1 and contains two steps. I first show the consistency of the index with the generalized residuals from the E-step. Then, in step two I show that the residuals satisfy the conditions imposed on the linear panel data models with interactive fixed effects as in Bai (2009b). The consistency of $\hat{\beta}_{PPIF}$ follows.

3.2 Asymptotic results

Define the nonlinear differencing operator

$$D_{\beta\pi^q}\ell_{it} := \partial_{\pi^{q+1}}\ell_{it}(X_{it} - \Xi_{it}), \quad \text{for } q = 0, 1, 2$$

where Ξ_{it} is a $\dim \beta$ -vector including the least squares projections of X_{it} on the space of incidental parameters spanned by $\alpha_i^0\gamma_t^0(\alpha_i + \gamma_t)$ weighted by $\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})$, i.e.,

$$\Xi_{it,k} = \alpha_i^0\gamma_t^0(\alpha_{i,k}^* + \gamma_{t,k}^*), \quad (12)$$

$$(\alpha_k^*, \gamma_k^*) \in \arg \min_{\alpha_{i,k}, \gamma_{t,k}} \sum_{i,t} \mathbb{E}_\phi[-\partial_{z^2}\ell_{it}(X_{it} - \alpha_i^0\gamma_t^0(\alpha_{i,k} + \gamma_{t,k}))^2].$$

Let $\bar{\mathcal{H}}$ be the $(N+T) \times (N+T)$ expected value of the Hessian matrix of the log-likelihood with respect to the nuisance parameters evaluated at the true parameters, i.e.,

$$\bar{\mathcal{H}} = \mathbb{E}_\phi[-\partial_{\phi\phi'}\mathcal{L}] = \begin{bmatrix} \bar{\mathcal{H}}_{(\alpha\alpha)} & \bar{\mathcal{H}}_{(\alpha\gamma)} \\ \bar{\mathcal{H}}'_{(\alpha\gamma)} & \bar{\mathcal{H}}_{(\gamma\gamma)} \end{bmatrix},$$

where $\bar{\mathcal{H}}_{(\alpha\alpha)} = \text{diag}(\sum_t(\gamma_t^0)^2\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/(NT)$, $\bar{\mathcal{H}}_{(\alpha\gamma)it} = (\alpha_i^0\gamma_t^0\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/(NT)$, and $\bar{\mathcal{H}}_{(\gamma\gamma)} = \text{diag}(\sum_i(\alpha_i^0)^2\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/(NT)$. Furthermore, let $\bar{\mathcal{H}}_{(\alpha\alpha)}^{-1}$, $\bar{\mathcal{H}}_{(\alpha\gamma)}^{-1}$, $\bar{\mathcal{H}}_{(\gamma\alpha)}^{-1}$, and $\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1}$

denote the $N \times N$, $N \times T$, $T \times N$ and $T \times T$ blocks of the inverse $\overline{\mathcal{H}}^{-1}$ of $\overline{\mathcal{H}}$. Then

$$\Xi_{it} = -\frac{1}{NT} \sum_{j=1}^N \sum_{\tau=1}^T (\overline{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \gamma_\tau^0 \gamma_t^0 + \overline{\mathcal{H}}_{(\alpha\gamma)it}^{-1} \alpha_j^0 \gamma_t^0 + \overline{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \alpha_i^0 \gamma_\tau^0 + \overline{\mathcal{H}}_{(\gamma\gamma)it}^{-1} \alpha_i^0 \alpha_j^0) \mathbb{E}_\phi(\partial_{\beta\pi} \ell_{jt}) \quad (13)$$

This nonlinear differencing operator generalizes to nonlinear models the partialing-out of individual and time effects in linear models. For example, if the model is linear, $\partial_{\pi^2} \ell_{it} = -1$, $\partial_{\beta\pi} \ell_{it} = -X_{it}$, and

$$\Xi_{it} = T^{-1} \sum_{t=1}^T \mathbb{E}_\phi[X_{it}] + N^{-1} \sum_{i=1}^N \mathbb{E}_\phi[X_{it}] - (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_\phi[X_{it}],$$

so that $D_{\beta\pi} \ell_{it} = -(X_{it} - \Xi_{it}) \partial_{\pi} \ell_{it}$, $D_{\beta\pi^2} \ell_{it} = -(X_{it} - \Xi_{it})$, and $D_{\beta\pi^2} \ell_{it} = 0$.

Let $\overline{\mathbb{E}} := \text{plim}_{N,T \rightarrow \infty}$. The following theorem establishes the asymptotic distribution of the fixed effects EM estimator for PPIF, $\hat{\beta}_{PPIF}$.

Theorem 3.1. (*Asymptotic distribution of $\hat{\beta}_{PPIF}$*). *Suppose that Assumption 1 holds, that the following limits exist*

$$\begin{aligned} \overline{B}_\infty &= -\overline{\mathbb{E}} \left[\frac{\frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \sum_{\tau=t}^T \gamma_t^0 \gamma_\tau^0 \mathbb{E}_\phi[\partial_{\pi} \ell_{it} D_{\beta\pi} \ell_{i\tau}] + \frac{1}{2} \sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it})}{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})} \right], \\ \overline{D}_\infty &= -\overline{\mathbb{E}} \left[\frac{\frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(\partial_{\pi} \ell_{it} D_{\beta\pi} \ell_{it} + \frac{1}{2} D_{\beta\pi^2} \ell_{it})}{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})} \right], \\ \overline{W}_\infty &= -\overline{\mathbb{E}} \left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_\phi(\partial_{\beta\beta'} \ell_{it} - \partial_{\pi^2} \ell_{it} \Xi_{it} \Xi'_{it}) \right], \end{aligned}$$

and that $\overline{W}_\infty > 0$. Then,

$$\sqrt{NT}(\hat{\beta}_{PPIF} - \beta^0) \xrightarrow{d} \overline{W}_\infty^{-1} N(\kappa \overline{B}_\infty + \kappa^{-1} \overline{D}_\infty, \overline{W}_\infty).$$

The detailed proof is in Appendix B.2.

Let $\tilde{X}_{it} = X_{it} - \Xi_{it}$ be the residual of the least squares projection of X_{it} on the space spanned by the incidental parameters weighted by $\mathbb{E}_\phi(\omega_{it})$, for $\omega_{it} = (\phi_f(X'_{it}\beta + \alpha_i^0 \gamma_t^0))^2 / [\Phi(X'_{it}\beta^0 + \alpha_i^0 \gamma_t^0)(1 - \Phi(X'_{it}\beta + \alpha_i^0 \gamma_t^0))]$.

Remark 3.1. For the probit model with X_{it} strictly exogenous, observe that

$$\begin{aligned}\bar{B}_\infty &= \mathbb{E}\left[\frac{1}{2N} \sum_{i=1}^N \frac{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi[\omega_{it} \tilde{X}_{it} \tilde{X}'_{it}]}{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi[\omega_{it}]} \right] \beta^0, \\ \bar{D}_\infty &= \mathbb{E}\left[\frac{1}{2T} \sum_{t=1}^T \frac{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi[\omega_{it} \tilde{X}_{it} \tilde{X}'_{it}]}{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi[\omega_{it}]} \right] \beta^0, \\ \bar{W}_\infty &= \mathbb{E}\left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_\phi[\omega_{it} \tilde{X}_{it} \tilde{X}'_{it}] \right].\end{aligned}$$

The asymptotic bias is therefore a positive-definite-matrix of the weighted average of the true parameters as in the case of the probit model with additive effects (see Fernández-Val and Weidner (2014)).

3.3 Asymptotic distribution of the average partial effects

In nonlinear models, the researcher is often interested in average partial effects in addition to the model structural parameters. These effects are averages of the data, parameters and unobserved effects as in equation (4). I impose the following sampling and moment conditions on the function Δ that defines the partial effects:

Assumption 2. (*Partial effects*). Let $v > 0$, $\epsilon > 0$, and \mathcal{B}_ϵ^0 all be as in Assumption 1

- (i) *Sampling:* for all $N, T, \{\alpha_i\}_N$ and $\{\gamma_t\}_T$ are deterministic;
- (ii) *Model:* for all i, t, N, T , the partial effects depend on α_i and γ_t through $\alpha_i \gamma_t$:

$$\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \alpha_i \gamma_t).$$

The realizations of the partial effects are denoted by $\Delta_{it} := \Delta_{it}(\beta^0, \alpha_i^0 \gamma_t^0)$.

(iii) *Moments:* The partial derivatives of $\Delta_{it}(\beta, \pi)$ with respect to the elements of (β, π) up to fourth order are bounded in absolute value uniformly over $(\beta, \pi) \in \mathcal{B}_\epsilon^0$ by a function $M(Z_{it}) > 0$ a.s., and $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+v}]$ is a.s. uniformly bounded over N, T .

(iv) *Non-degeneracy and moments:* $\min_{i,t} \text{Var}(\Delta_{it}) > 0$ and $\max_{i,t} \text{Var}(\Delta_{it}) < \infty$, uniformly over N, T .

Analogous to Ξ_{it} in equation (13), define

$$\Psi_{it} = -\frac{1}{NT} \sum_{j=1}^N \sum_{\tau=1}^T (\bar{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \gamma_\tau^0 \gamma_t^0 + \bar{\mathcal{H}}_{(\alpha\gamma)it}^{-1} \alpha_j^0 \gamma_t^0 + \bar{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \alpha_i^0 \gamma_\tau^0 + \bar{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \alpha_i^0 \alpha_j^0) \partial_\pi \Delta_{j\tau},$$

which is the population projection of $\partial_\pi \Delta_{it} / \mathbb{E}_\phi[\partial_{\pi^2} \ell_{it}]$ on the space spanned by the incidental parameters under the metric given by $\mathbb{E}_\phi[-\partial_{\pi^2} \ell_{it}]$. I use an analogous notation to the previous section for the derivatives with respect to β and higher order derivatives with respect to π .

Let δ_{NT}^0 be the APE as defined in equation (4), and $\hat{\delta}$ be its estimator $\Delta_{NT}(\hat{\beta}, \hat{\phi}_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \hat{\beta}, \hat{\alpha}_i \hat{\gamma}_t)$. The following theorem establishes the asymptotic distribution of $\hat{\delta}$.

Theorem 3.2. (*Asymptotic distribution of $\hat{\delta}$*). *Suppose that the assumptions of Theorem 3.1 and Assumption 2 hold, and that the following limits exist:*

$$\begin{aligned} \overline{(D_\beta \Delta)}_\infty &= \mathbb{E}\left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_\phi(\partial_\beta \Delta_{it} - \Xi_{it} \partial_\pi \Delta_{it})\right], \\ \overline{B}_\infty^\delta &= \overline{(D_\beta \Delta)}'_\infty \overline{W}_\infty^{-1} \overline{B}_\infty + \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N \frac{\sum_{t=1}^T \sum_{\tau=t}^T \gamma_t^0 \gamma_\tau^0 \mathbb{E}_\phi(\partial_\pi \ell_{it} \partial_{\pi^2} \ell_{i\tau} \Psi_{i\tau})}{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})}\right] \\ &\quad - \mathbb{E}\left[\frac{1}{2N} \sum_{i=1}^N \frac{\sum_{t=1}^T (\gamma_t^0)^2 [\mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3} \ell_{it}) \mathbb{E}_\phi(\Psi_{it})]}{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})}\right], \\ \overline{D}_\infty^\delta &= \overline{(D_\beta \Delta)}'_\infty \overline{W}_\infty^{-1} \overline{D}_\infty + \mathbb{E}\left[\frac{1}{T} \sum_{t=1}^T \frac{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(\partial_\pi \ell_{it} \partial_{\pi^2} \ell_{it} \Psi_{it})}{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})}\right] \\ &\quad - \mathbb{E}\left[\frac{1}{2T} \sum_{t=1}^T \frac{\sum_{i=1}^N (\alpha_i^0)^2 [\mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3} \ell_{it}) \mathbb{E}_\phi(\Psi_{it})]}{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})}\right], \\ \overline{V}_\infty^\delta &= \mathbb{E}\left\{\frac{1}{NT} \sum_{i=1}^N \left[\sum_{t,\tau=1}^T \mathbb{E}_\phi(\tilde{\Delta}_{it} \tilde{\Delta}'_{i\tau}) + \sum_{t=1}^T \mathbb{E}_\phi(\Gamma_{it} \Gamma'_{it}) \right]\right\}, \end{aligned}$$

for some $\overline{V}_\infty^\delta > 0$, where $\tilde{\Delta}_{it} = \Delta_{it} - \mathbb{E}(\Delta_{it})$ and $\Gamma_{it} = \overline{(D_\beta \Delta)}'_\infty \overline{W}_\infty^{-1} D_\beta \ell_{it} - \mathbb{E}_\phi(\Psi_{it}) \partial_\pi \ell_{it}$. Then,

$$\sqrt{NT}(\hat{\delta} - \delta_{NT}^0 - T^{-1} \overline{B}_\infty^\delta - N^{-1} \overline{D}_\infty^\delta) \xrightarrow{d} N(0, \overline{V}_\infty^\delta).$$

The bias and variance are of the same order asymptotically under the asymptotic sequence of Assumption 1(i).

Remark 3.2. (Average effects from bias-corrected estimators). As in the case of the probit with additive effects (Fernández-Val and Weidner (2014)), the first term in the expressions of the biases $\overline{B}_\infty^\delta$ and $\overline{D}_\infty^\delta$ comes from the bias of the estimator of β . It drops out when the APEs are constructed from asymptotically unbiased or bias-corrected estimators of the

parameter β , i.e.,

$$\tilde{\delta} = \Delta(\tilde{\beta}, \hat{\phi}(\tilde{\beta})),$$

where $\tilde{\beta}$ is such that $\sqrt{NT}(\tilde{\beta} - \beta^0) \xrightarrow{d} N(0, \bar{W}_\infty^{-1})$. The asymptotic variance of $\tilde{\delta}$ is the same as in Theorem 3.2.

In the following examples I assume that the APEs are constructed from asymptotically unbiased estimators of the model parameters.

Example 3.1. Consider the partial effects defined in (5) and (6) with

$$\Delta_{it}(\beta, \alpha_i \gamma_t) = \Phi(\beta_k + X'_{it,-k} \beta_{-k} + \alpha_i \gamma_t) - \Phi(X'_{it,-k} \beta_{-k} + \alpha_i \gamma_t)$$

and

$$\Delta_{it}(\beta, \alpha_i \gamma_t) = \beta_k \phi_f(X'_{it} \beta + \alpha_i \gamma_t).$$

Denote $H_{it} = \phi_f(X'_{it} \beta + \alpha_i^0 \gamma_t^0) / [\Phi(X'_{it} \beta^0 + \alpha_i^0 \gamma_t^0)(1 - \Phi(X'_{it} \beta + \alpha_i^0 \gamma_t^0))]$ and use notations previously introduced, the components of the asymptotic bias of $\tilde{\delta}$ are

$$\begin{aligned} \bar{B}_\infty^\delta &= \mathbb{E} \left[\frac{1}{2N} \sum_{i=1}^N \frac{\sum_{t=1}^T [2 \sum_{\tau=t+1}^T \mathbb{E}_\phi(H_{it}(Y_{it} - \Phi_{it}) \omega_{i\tau} \tilde{\Psi}_{i\tau}) - \mathbb{E}_\phi(\Psi_{it}) \mathbb{E}_\phi(H_{it} \partial^2 \Phi_{it}) + \mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it})]}{\sum_{t=1}^T \mathbb{E}_\phi(\omega_{it})} \right], \\ \bar{D}_\infty^\delta &= \mathbb{E} \left[\frac{1}{2T} \sum_{t=1}^T \frac{\sum_{i=1}^N [-\mathbb{E}_\phi(\Psi_{it})] \mathbb{E}_\phi(H_{it} \partial^2 \Phi_{it}) + \mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it})}{\sum_{i=1}^N \mathbb{E}_\phi(\omega_{it})} \right] \end{aligned}$$

where $\tilde{\Psi}_{it}$ is the residual of the population regression of $-\partial_\pi \Delta_{it} / \mathbb{E}_\phi[\omega_{it}]$ on the space spanned by the incidental parameters under the metric given by $\mathbb{E}_\phi[\omega_{it}]$. If all the components of X_{it} are strictly exogenous, the first term in the numerator of \bar{B}_∞^δ is zero.

3.4 Bias-corrected estimators

The results of the previous sections show that the asymptotic distributions of the interactive fixed effects estimators of the model parameters and APEs can have asymptotic bias under sequences where T grows at the same rate as N , as also discussed in Chen et al. (2013). This large- T version of the incidental parameters problem can invalidate any inference based on the asymptotic distribution. In this section I discuss how to construct analytical bias corrections for PPIF and give conditions for the asymptotic validity of the analytical bias corrections. The proof strategy here is similar to Fernández-Val and Weidner (2014) which is under the additive individual and time effects setting.

The analytical corrections are constructed using sample analogs of the expressions in Theorems 3.1 and 3.2, replacing the true values of β and ϕ by the estimated ones. To describe these corrections, I introduce some additional notation. For any function of the data, unobserved effects and parameters $\varphi_{itj}(\beta, \alpha_i \gamma_t, \alpha_i \gamma_{t-j})$ with $0 \leq j < t$, let $\hat{\varphi}_{itj} = \varphi_{it}(\hat{\beta}, \hat{\alpha}_i \hat{\gamma}_t, \hat{\alpha}_i \hat{\gamma}_{t-j})$ be its estimator, e.g., $\mathbb{E}_\phi[\widehat{\partial_{\pi^2} \ell_{it}}]$ denotes the estimator of $\mathbb{E}_\phi[\partial_{\pi^2} \ell_{it}]$. Let $\hat{\mathcal{H}}_{(\alpha\alpha)}^{-1}$, $\hat{\mathcal{H}}_{(\alpha\gamma)}^{-1}$, $\hat{\mathcal{H}}_{(\gamma\alpha)}^{-1}$ and $\hat{\mathcal{H}}_{(\gamma\gamma)}^{-1}$ denote the blocks of the matrix $\hat{\mathcal{H}}^{-1}$, where

$$\hat{\mathcal{H}} = \begin{pmatrix} \hat{\mathcal{H}}_{(\alpha\alpha)} & \hat{\mathcal{H}}_{(\alpha\gamma)} \\ \hat{\mathcal{H}}'_{(\alpha\gamma)} & \hat{\mathcal{H}}_{(\gamma\gamma)} \end{pmatrix},$$

$\hat{\mathcal{H}}_{(\alpha\alpha)} = \text{diag}(-\sum_t (\hat{\gamma}_t)^2 \mathbb{E}_\phi[\widehat{\partial_{\pi^2} \ell_{it}}]) / (NT)$, $\hat{\mathcal{H}}_{(\alpha\gamma)it} = -\hat{\alpha}_i \hat{\gamma}_t \mathbb{E}_\phi[\widehat{\partial_{\pi^2} \ell_{it}}] / (NT)$, and $\hat{\mathcal{H}}_{(\gamma\gamma)} = \text{diag}(-\sum_i (\hat{\alpha}_i)^2 \mathbb{E}_\phi[\widehat{\partial_{\pi^2} \ell_{it}}]) / (NT)$. Let

$$\hat{\Xi}_{it} = -\frac{1}{NT} \sum_{j=1}^N \sum_{\tau=1}^T (\hat{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \hat{\gamma}_\tau \hat{\gamma}_t + \hat{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1} \hat{\alpha}_j \hat{\gamma}_t + \hat{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \hat{\alpha}_i \hat{\gamma}_\tau + \hat{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \hat{\alpha}_i \hat{\alpha}_j) \mathbb{E}_\phi(\widehat{\partial_{\beta\pi} \ell_{j\tau}}),$$

the k th component of $\hat{\Xi}_{it}$ corresponds to a least square regression of X_{it} on the space spanned by the incidental parameters weighted by $-\mathbb{E}_\phi(\widehat{\partial_{\beta\pi} \ell_{it}})$. The analytical bias-corrected estimator of β^0 is

$$\tilde{\beta}^A = \hat{\beta} - \hat{B}/T - \hat{D}/N,$$

where

$$\hat{B} = -\frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=0}^L (T/(T-j)) \sum_{t=j+1}^T \hat{\gamma}_t \hat{\gamma}_\tau \mathbb{E}(\partial_{\pi} \widehat{\ell_{it}} D_{\beta\pi} \ell_{i\tau}) + \frac{1}{2} \sum_{t=1}^T (\hat{\gamma}_t)^2 \mathbb{E}(\widehat{D_{\beta\pi^2} \ell_{it}})}{\sum_{t=1}^T (\hat{\gamma}_t)^2 \mathbb{E}(\widehat{\partial_{\pi^2} \ell_{it}})},$$

$$\hat{D} = -\frac{1}{T} \sum_{t=1}^T \frac{\sum_{i=1}^N (\hat{\alpha}_i)^2 \mathbb{E}(\partial_{\pi} \widehat{\ell_{it}} D_{\beta\pi} \ell_{it}) + \frac{1}{2} \widehat{D_{\beta\pi^2} \ell_{it}}}{\sum_{i=1}^N (\hat{\alpha}_i)^2 \mathbb{E}(\widehat{\partial_{\pi^2} \ell_{it}})},$$

and L is a trimming parameter for estimation of spectral expectations such that $L \rightarrow \infty$ and $L/T \rightarrow 0$, see Hahn and Kuersteiner (2011).

Asymptotic $(1-p)$ - confidence intervals for the components of β^0 can be formed as

$$\tilde{\beta}_k^A \pm z_{1-p} \sqrt{\widehat{W}_{kk}^{-1} / (NT)}, \quad k = \{1, \dots, \dim \beta^0\}.$$

where z_{1-p} is the $(1-p)$ quantile of the standard normal distribution, and \widehat{W}_{kk}^{-1} is the

(k, k) -element of the matrix \widehat{W}^{-1} with

$$\widehat{W} = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \mathbb{E}_\phi(\widehat{\partial_{\beta\beta'} \ell_{it}}) - \mathbb{E}_\phi(\partial_{\pi^2} \widehat{\ell_{it}} \widehat{\Xi_{it}} \widehat{\Xi'_{it}}).$$

The analytical bias-corrected estimator of δ_{NT}^0 is

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

where I use $\widetilde{\delta}$, i.e., the APE constructed from a bias corrected estimator of β . Let

$$\widehat{\Psi}_{it} = -\frac{1}{NT} \sum_{j=1}^N \sum_{\tau=1}^T (\widehat{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \widehat{\gamma}_\tau \widehat{\gamma}_t + \widehat{\mathcal{H}}_{(\alpha\gamma)it}^{-1} \widehat{\alpha}_j \widehat{\gamma}_t + \widehat{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \widehat{\alpha}_i \widehat{\gamma}_\tau + \widehat{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \widehat{\alpha}_i \widehat{\alpha}_j) \widehat{\partial_\pi \Delta_{j\tau}},$$

then the estimated asymptotic biases are

$$\begin{aligned} \widehat{B}^\delta &= \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j=0}^L [T/(T-j)] \sum_{t=j+1}^T \widehat{\gamma}_t \widehat{\gamma}_\tau \mathbb{E}_\phi(\widehat{\partial_{\pi^2} \ell_{it}} \widehat{\Psi}_{it})}{\sum_{t=1}^T (\widehat{\gamma}_t)^2 \mathbb{E}_\phi(\widehat{\partial_{\pi^2} \ell_{it}})} \\ &\quad - \frac{1}{2N} \sum_{i=1}^N \frac{\sum_{t=1}^T (\widehat{\gamma}_t)^2 [\mathbb{E}_\phi(\widehat{\partial_{\pi^2} \Delta_{it}}) - \mathbb{E}_\phi(\widehat{\partial_{\pi^3} \ell_{it}}) \mathbb{E}_\phi(\widehat{\Psi}_{it})]}{\sum_{t=1}^T (\widehat{\gamma}_t)^2 \mathbb{E}_\phi(\widehat{\partial_{\pi^2} \ell_{it}})} \\ \widehat{D}^\delta &= \frac{1}{T} \sum_{t=1}^T \frac{\sum_{i=1}^N (\widehat{\alpha}_i)^2 [\mathbb{E}_\phi(\widehat{\partial_{\pi^2} \ell_{it}} \widehat{\Psi}_{it}) - \frac{1}{2} \mathbb{E}_\phi(\widehat{\partial_{\pi^2} \Delta_{it}}) + \frac{1}{2} \mathbb{E}_\phi(\widehat{\partial_{\pi^3} \ell_{it}}) \mathbb{E}_\phi(\widehat{\Psi}_{it})]}{\sum_{i=1}^N (\widehat{\alpha}_i)^2 \mathbb{E}_\phi(\widehat{\partial_{\pi^2} \ell_{it}})}. \end{aligned}$$

The estimator of the asymptotic variance depends on the assumptions about the distribution of the unobserved effects and explanatory variables. Assumption 2(i) requires imposing a homogeneity assumption on the distribution of the explanatory variables to estimate the first term of the asymptotic variance. For example, if $\{X_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$ is identically distributed over i , this term is given by

$$\widehat{V}^\delta = \frac{1}{NT} \sum_{i=1}^N \left[\sum_{t,\tau=1}^T \widehat{\Delta}_{it} \widehat{\Delta}'_{i\tau} + \sum_{t=1}^T \mathbb{E}_\phi(\widehat{\Gamma_{it}} \widehat{\Gamma}'_{it}) \right],$$

for $\widehat{\Delta}_{it} = \widehat{\Delta}_{it} - N^{-1} \sum_{i=1}^N \widehat{\Delta}_{it}$. Bias corrected estimators and confidence intervals can be constructed in the same fashion as for the model parameter.

The following theorems show that the analytical bias corrections eliminate the bias from the asymptotic distribution of the fixed effects estimators of the model parameters and

APEs without increasing the variance, and that the estimators of the asymptotic variances are consistent. Those are the main results of this section.

Theorem 3.3. (*Bias correction for $\hat{\beta}$*) Under the conditions of Theorem 3.1,

$$\widehat{W} \xrightarrow{p} \overline{W}_\infty,$$

and, if $L \rightarrow \infty$ and $L/T \rightarrow 0$,

$$\sqrt{NT}(\tilde{\beta}^A - \beta^0) \xrightarrow{d} N(0, \overline{W}_\infty^{-1}).$$

Theorem 3.4. (*Bias correction for $\hat{\delta}$*) Under the conditions of Theorems 3.1 and 3.2,

$$\widehat{V}^\delta \xrightarrow{p} \overline{V}_\infty^\delta,$$

and, if $L \rightarrow \infty$ and $L/T \rightarrow 0$,

$$\sqrt{NT}(\tilde{\delta}^A - \delta_{NT}^0) \xrightarrow{d} N(0, \overline{V}_\infty^\delta).$$

Remark 3.3. Split-panel jackknife as described in Chen et al. (2013); Fernández-Val and Weidner (2014) can also be applied.

4 Discussions and Extensions

4.1 Comparison with the existing estimators: No fixed effects or only individual effects

When there are no fixed effects, the model becomes

$$\begin{aligned} Y_{it}^* &= X_{it}'\beta + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \tag{14}$$

where all objects are as defined previously. The conditional distribution of Y_{it} is given by

$$P(Y_{it} = 1 | X_{it}, \beta) = \Phi(X_{it}\beta),$$

and for estimation the following EM procedure can be used:

Definition 4.1. (1) Given initial $\beta^{(k)}$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)}$;

(2) **E-step:** Calculate $\hat{Y}_{it}^{(k)} := E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}]$;

(3) **M-step:** The parameter β is updated via

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it} X_{it}' \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it} \hat{Y}_{it}^{(k)} \right\}.$$

(4) Iterate until convergence.

I start by comparing this estimation with existing methods.

Proposition 4.1. *For panel probit models, the proposed EM-type estimator is asymptotically equivalent to the MLE.*

Proof: See Appendix C.1.1. When applying the proposed fixed effects EM-type estimator to probit (or for the general exponential family), its E-step involves calculating the conditional expectation of the error, which is exactly the score of expected, complete data, log-likelihood function or the score of the observed log-likelihood (it also corresponds to the notion of generalized residuals proposed in Gourieroux et al. (1987) for cross-sectional data). Hence, the fixed effects EM-type estimator directly works with the observed score. For the case when there are no unobserved effects, the EM method is asymptotically equivalent to MLE and there is no asymptotic bias. For the cases when there are unobserved effects, and when there are incidental parameter problems, an iterated bias correction to the score can be easily implemented through the E-step.

Proposition 4.2. *For the panel probit model with individual effects, the difference between the proposed fixed effects EM-type estimator and Newton's method lies in whether inverting the Hessian of the observed data log-likelihood function.*

Proof: See Appendix C.1.2. I explicitly compare the two iterative steps of the fixed effects EM-type estimator and the Newton's method. Each iteration of the proposed fixed effects EM-type estimator is a least squares calculation (with the generalized residual); it does not use the inverse of the Hessian of the observed data log-likelihood function like Newton's method.⁹

⁹See Greene (2004) for more about estimation of nonlinear panel data models with individual fixed effects.

4.2 PPIF with multiple factors

In this setting, the model, written in matrix notation, is

$$Y = \mathbf{1}(X\beta + \alpha\gamma' + \varepsilon \geq 0),$$

where $Y = (Y_1, \dots, Y_N)'$ (with $Y_i = (Y_{i1}, \dots, Y_{iT})'$, a $T \times 1$ vector) is an $N \times T$ matrix and X (with $X_i = [X_{i1}, \dots, X_{iT}]'$ is a $T \times p$ matrix) is a three-dimensional matrix with p sheets ($N \times T \times p$), the ℓ -th sheet of which is associated with the ℓ -th element of β ($\ell = 1, \dots, p$). $\alpha = (\alpha_1, \dots, \alpha_N)'$ is an $N \times R$ matrix, while $\gamma = (\gamma_1, \dots, \gamma_T)'$ is a $T \times R$ matrix. The product $X\beta$ is an $N \times T$ matrix and $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$ is an $N \times T$ matrix.

Since $\alpha\gamma' = \alpha A^{-1} A \gamma'$ for any $R \times R$ invertible A , identification is not possible without restrictions.

Condition 1. (Normalization) (i) $\gamma'\gamma/T = I_R$; (ii) $\alpha'\alpha = \text{diagonal}$.

Under different normalization conditions, the estimation procedure (the conditional maximization steps) for the factor is different.

Definition 4.2. The EM procedure for estimating a panel probit model with multi-dimensional interactive fixed effects under Condition 1 is defined by the following:

- (1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X_{it}'\beta^{(k)} + (\alpha_i^{(k)})'\gamma_t^{(k)}$,
- (2) **E-step:** Calculate

$$\begin{aligned} \hat{Y}_{it}^{(k)} &: = E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)}) / \{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)}))\}, \end{aligned}$$

- (3) **M-step:** This contains three conditional maximization (CM) steps

CM-step 1: Given α_i and γ_t , the parameter β is updated via

$$\beta^{(k+1)} = \left(\sum_{i=1}^N X_i' X_i \right)^{-1} \left\{ \sum_{i=1}^N X_i' (\hat{Y}_i^{(k)} - \alpha_i^{(k)} \gamma^{(k)}) \right\},$$

CM-step 2: Given β and α_i , the parameter γ is updated via

$$\gamma^{(k+1)} = \text{eig} \left[\frac{1}{NT} \sum_{i=1}^N (\hat{Y}_i^{(k)} - X_i \beta^{(k+1)}) (\hat{Y}_i^{(k)} - X_i \beta^{(k+1)})' \right],$$

CM-step 3: Given β and γ_t , the parameter α is updated via

$$\alpha^{(k+1)} = T^{-1}(\hat{Y}^{(k)} - X\beta^{(k+1)})\gamma^{(k+1)},$$

(4) Iterate until convergence.

The CM-step 2 calculates the R largest eigenvector of the matrix in brackets, arranged in decreasing order. It imposes the normalizations of Condition 1 by using eigenvectors. An alternative estimation procedure based on a QR decomposition that does not impose Condition 1(ii) is also proposed below.

Definition 4.3. The QR-based decomposition EM procedure for estimating a panel probit model with multi-dimensional interactive fixed effects is defined by the following:

(1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X'_{it}\beta^{(k)} + (\alpha_i^{(k)})'\gamma_t^{(k)}$,

(2) **E-step:** Calculate

$$\begin{aligned} \hat{Y}_{it}^{(k)} &: = E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)}) / \{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)}))\}, \end{aligned}$$

(3) **M-step:** This contains three conditional maximization (CM) steps

CM-step 1: Given α_i and γ_t , the parameter β is updated via

$$\beta^{(k+1)} = \left(\sum_{i=1}^N X_i' X_i \right)^{-1} \left\{ \sum_{i=1}^N X_i' (\hat{Y}_i^{(k)} - \alpha_i^{(k)} \gamma^{(k)}) \right\},$$

CM-step 2: Given β and α_i , the parameter γ is updated via

$$\gamma^{(k+1)} = (\hat{Y}^{(k)} - X\beta^{(k+1)})' \alpha^{(k)} ((\alpha^{(k)})' \alpha^{(k)})^{-1}.$$

Compute the QR decomposition $\gamma^{(k+1)} = \tilde{\gamma}^{(k+1)} R_M$ and replace $\gamma^{(k+1)}$ by $\tilde{\gamma}^{(k+1)}$,

CM-step 3: Given β and $\tilde{\gamma}$, the parameter α is updated via

$$\alpha^{(k+1)} = (\hat{Y}^{(k)} - X\beta^{(k+1)})\tilde{\gamma}^{(k+1)},$$

(4) Iterate until convergence.

Through the iterations, the columns of the updated values of γ are made orthonormal via the QR decomposition (imposing normalization, but other decomposition methods can also

be used), i.e., $(\tilde{\gamma}^{(k+1)})'\tilde{\gamma}^{(k+1)}$ is orthonormal (I_R). The QR decomposition is often used to solve the linear least squares problem, and is the basis for a particular eigenvalue algorithm. With additional restrictions, such as a full rank condition on γ and a sign restriction on R_M , the QR decomposition method can achieve unique values of α and γ .

Note that the orthogonalization does not alter the convergence property. Let $\gamma^{(k+1)}$ be the optimizer before orthogonalization. Then $S(\beta, \gamma^{(k+1)}, \alpha^{(k)}) \leq S(\beta, \gamma^{(k)}, \alpha^{(k)})$. Let $\gamma^{(k+1)} = \tilde{\gamma}^{(k+1)}R_M$ be the QR decomposition of $\gamma^{(k+1)}$, and let $\tilde{\alpha}^{(k)} = \alpha^{(k)}R_M'$. Then $\tilde{\alpha}^{(k)}(\tilde{\gamma}^{(k+1)})' = \alpha^{(k)}(\gamma^{(k+1)})'$, so $S(\beta, \tilde{\gamma}^{(k+1)}, \tilde{\alpha}^{(k)}) = S(\beta, \gamma^{(k+1)}, \alpha^{(k)})$, and, consequently, $S(\beta, \tilde{\gamma}^{(k+1)}, \tilde{\alpha}^{(k)}) \leq S(\beta, \gamma^{(k)}, \alpha^{(k)})$.

4.2.1 Consistency

In general, the consistency proof contains two steps as shown in the proof for PPIF. The first step involves the consistency of the conditional expectation, and the second checks the assumptions needed for the consistency of the “linearized” model.

Assumption 3. (*Bounded second-order derivative*) $\partial_{\pi^2}\mathcal{L}_{NT}(\beta, \pi) \geq b_{\min}$.

Lemma 4.1. *Under Assumption 3 and Assumption 1(i), (ii), and (iv), $\hat{\beta}_{IF-EM} = \beta^0 + o_p(1)$.*

Proof: See Appendix C.2.

5 Simulations

This section reports evidence on the finite sample behavior of fixed effects estimators in static models with strictly exogenous regressors. This includes several cases: no unobserved effects, individual effects, additive individual and time effects, and interactive individual and time effects. I analyze the performance of the generalized least square (GLS) method using the **R**-package **glm**, which is available on CRAN, and the fixed effects EM-type estimators in terms of bias and inference accuracy based on their asymptotic distribution. I also analyze the performance of the uncorrected and bias-corrected interactive fixed effects EM-type estimators in terms of bias and inference accuracy. In particular, I compute the biases, standard deviations, and root mean squared errors (RMSE) of the estimators, the ratio of averaged standard errors to the simulation standard deviations (SE/SD); and the empirical coverages of confidence intervals with 95% nominal value ($p; .95$). All results are based on 500 replications.

The data generating processes are:

- DGP-1: $Y_{it} = \mathbf{1}\{X_{it}\beta + \varepsilon_{it} > 0\}$, $(i = 1, \dots, N; t = 1, \dots, T)$,
- DGP-2: $Y_{it} = \mathbf{1}\{X_{it}\beta + \alpha_i + \varepsilon_{it} > 0\}$, $(i = 1, \dots, N; t = 1, \dots, T)$,
- DGP-3: $Y_{it} = \mathbf{1}\{X_{it}\beta + \alpha_i + \gamma_t + \varepsilon_{it} > 0\}$, $(i = 1, \dots, N; t = 1, \dots, T)$,
- DGP-4: $Y_{it} = \mathbf{1}\{X_{it}\beta + \alpha_i\gamma_t + \varepsilon_{it} > 0\}$, $i = 1, \dots, N; t = 1, \dots, T$,

where $\beta = 1$, $\alpha_i \sim N(0, 1)$, $\gamma_t \sim N(0, 1)$, and $X_{it} \sim N(0, 1)$ are strictly exogenous with respect to ε_{it} with $\varepsilon_{it} \sim N(0, 1)$.

Throughout, “No FE” refers to the probit without fixed effects; “FE i” refers to the probit with individual fixed effects; “FE 2” refers to the probit with additive individual and time fixed effects; “IF” refers to the probit with interactive fixed effects; “glm” refers to the GLS estimator in \mathbf{R} , while “EM” refers to the fixed effects EM-type estimators proposed. For interactive fixed effects, I also implement the bias correction procedure proposed here; “BC-IF” refers to the bias-corrected estimator. All the results are reported in percentages of the true parameter value.

The simulation results are summarized in Table 1 for $N = 100$ and $T = 8, 12, 20$, and in Table 2 for $N=52$ and $T = 14, 26, 52$. They show that in all the cases analyzed EM has smaller biases and variances and compares favorably to **glm**. For example, for the case with additive individual and time effects, when $N = 100$ and $T = 12$, the bias for **glm** is 21%, whereas the EM estimator is only 11%. Even for the case without unobserved effects, when $N = 100$ and $T = 20$, the bias for **glm** is 0.52%, whereas the EM estimator is only 0.11%. In terms of RMSE, for the case of individual effects, when $N = 52$ and $T = 14$, the RMSE for **glm** is 16%, whereas for the EM estimator it is 15%. When there is a bias, the results also show that it is of the same order of magnitude as the standard deviation for the uncorrected EM and **glm** estimator, and this causes severe undercoverage of the confidence intervals. The analytical bias correction removes the bias without increasing dispersion and produces substantial improvements in terms of RMSE and coverage probabilities. For example, the analytical bias correction reduces the RMSE by more than 4% and increases coverage by around 20% in the $N = 100$ and $T = 12$ case.

6 Empirical example

6.1 A gravity equation and the extensive margins of trade

Understanding how different trade barriers influence trade flows is key when one wants to study the impact of distance, trade agreements, and other trade frictions. See Helpman et al. (2008); Bernard et al. (2007); Charbonneau (2012). For my application, I use the same data set as in Helpman et al. (2008), which consists of information on who trades with whom for a large set of countries.

I illustrate the estimation and difference when including differing degrees of fixed effects, namely the cases with no fixed effects, only individual fixed effects, additive individual and time fixed effects, and interactive fixed effects. The fixed effects are importer and exporter fixed effects for a single year, the year 1986. I obtain a balanced panel of 158 countries that account for the majority of world trade. The probability of country j exporting to country i is

$$Prob[Trade_{ij} = 1|X_{ij}, g(\alpha_i, \gamma_j)] = \Phi(X'_{ij}\beta + g(\alpha_i, \gamma_j)).$$

Here X_{ij} contains D_{ij} , representing the distance between country i 's and country j 's most populated cities; $Border_{ij}$, a dummy that takes the value 1 if i and j share a border; $Legal_{ij}$, a dummy that takes the value 1 if the two countries have the same legal system; $Language_{ij}$, a dummy that takes the value 1 if i and j have the same official language; $Colony_{ij}$, a dummy that takes the value 1 if i and j were ever in a colonial relationship; $Currency_{ij}$, a dummy that takes the value of 1 if the two countries use the same currency; RTA_{ij} , a dummy that takes the value 1 if i and j are in a regional trade agreement; and, finally, α_i and γ_j , respectively representing importer and exporter fixed effects.

The results of the effects of trade barriers are summarized in Table 3. After accounting for exporter fixed effects the effect of a common currency decreases in magnitude from about -0.45 to -0.16. This suggests that excluding exporter effects may overstate the decrease in the likelihood of trade when trading partners share a common currency. The changes of magnitude on language and region suggest that excluding exporter effects may understate the importance of having the same language and the same religion. Similarly, the magnitude changes of distance, from about -0.19 to -0.29, suggesting that excluding exporter effects may understate the importance of distance. Importantly, the magnitude of the coefficient for border changes from 0.16 to -0.03 suggests overstating the importance of sharing a border.

Note also that the effect of free trade agreements is rather robust to the inclusion or complete omission of fixed effects. This suggests that perhaps the effect of a free trade agreement on the likelihood of trade between a pair of countries does not depend on the exact trade network of those countries; FTAs appear to increase the likelihood of trade regardless of which fixed effects are included.

7 Conclusion

This paper presents an EM type method of estimating nonlinear panel data models with multiple unobserved effects, allowing for interactions between the unobserved individual and time specific effects. The method can be applied to models with individual effects, additive individual and time effects, interactive effects and other general functional form of unobserved effects. In finite-sample simulations, the method outperform the existing generalized least square methods for the models with individual effects and additive individual and time effects in terms of both bias and variance. Furthermore, I derive the asymptotic distribution of the proposed EM estimator for the panel probit model with interactive fixed effects. Analytical bias corrections are developed to deal with the incidental parameter problem for both the estimates of the coefficients and its associated average partial effects. Simulations demonstrate the correction works well in reducing the bias and root mean squared error and improves coverage rates. Finally for purpose of illustration, I use the example of international trade networks demonstrating that misspecifying the fixed effects model can over or understate the importance of certain factors on the likelihood of trade. A wide range of future theoretical and empirical work can build upon the results of this paper. For example, sample selection models with interactive effects or models with strategic interactions, such as binary game models, could benefit from and build on the approach proposed here.

References

- Ahn, S. C., Lee, Y. H., Schmidt, P., 2001. Gmm estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics* 101 (2), 219–255.
- Ahn, S. C., Lee, Y. H., Schmidt, P., 2013. Panel data models with multiple time-varying individual effects. *Journal of Econometrics* 174 (1), 1–14.
- Andrews, D. W., 2005. Cross-section regression with common shocks. *Econometrica* 73 (5), 1551–1585.

- Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: Some recent developments, r. blundell, wk newey, and t. persson eds., advances in economics and econometrics: Theory and applications. In: Ninth World Congress. Vol. 3.
- Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71 (1), 135–171.
- Bai, J., 2009a. Likelihood approach to small t dynamic panel models with interactive effects. Unpublished working paper.
- Bai, J., 2009b. Panel data models with interactive fixed effects. *Econometrica* 77 (4), 1229–1279.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70 (1), 191–221.
- Bernard, A. B., Jensen, J. B., Redding, S. J., Schott, P. K., 2007. Firms in international trade. Tech. rep., National Bureau of Economic Research.
- Bester, C. A., Hansen, C., 2009. Identification of marginal effects in a nonparametric correlated random effects model. *Journal of Business & Economic Statistics* 27 (2).
- Bonhomme, S., 2012. Functional differencing. *Econometrica* 80 (4), 1337–1385.
- Carro, J. M., 2007. Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics* 140 (2), 503–528.
- Chamberlain, G., 1980. Analysis of covariance with qualitative data. *The Review of Economic Studies* 47 (1), 225–238.
- Charbonneau, K., 2012. Multiple fixed effects in nonlinear panel data models. Unpublished manuscript.
- Chen, M., Fernández-Val, I., Weidner, M., 2013. Interactive fixed effects in nonlinear panel data models with large n, t.
- Cunha, F., Heckman, J. J., 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43 (4), 738–782.
- Cunha, F., Heckman, J. J., Schennach, S. M., 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78 (3), 883–931.

- Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.
- Dhaene, G., Jochmans, K., 2010. Split-panel jackknife estimation of fixed-effect models.
- Elrod, T., Keane, M. P., 1995. A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, 1–16.
- Fernandez-Val, I., 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150 (1), 71–85.
- Fernández-Val, I., Vella, F., 2011. Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics* 163 (2), 144–162.
- Fernández-Val, I., Weidner, M., 2014. Individual and time effects in nonlinear panel data models with large n , t .
- Freyberger, J., 2012. Nonparametric panel data models with interactive fixed effects. Tech. rep., Working Paper.
- Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalised residuals. *Journal of Econometrics* 34 (1), 5–32.
- Greene, W. H., 2003. *Econometric analysis*. Pearson Education India.
- Greene, W. H., 2004. The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal* 7(1), 98–119.
- Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and t are large. *Econometrica* 70 (4), 1639–1657.
- Hahn, J., Kuersteiner, G., 2011. Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27 (6), 1152.
- Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72 (4), 1295–1319.
- Heckman, J. J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics* 24 (3), 411–482.

- Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics* 123 (2), 441–487.
- Holtz-Eakin, D., Newey, W., Rosen, H., 1988. Estimating vector autoregressions with panel data. *Econometrica: Journal of the Econometric Society*, 1371–1395.
- Laird, N. M., 1985. Missing information principle. *Encyclopedia of Statistical Sciences*.
- McLeish, D., 1974. Dependent central limit theorems and invariance principles. *the Annals of Probability*, 620–628.
- Meng, X.-L., Rubin, D. B., 1993. Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika* 80 (2), 267–278.
- Moon, H., Weidner, M., 2010a. Dynamic linear panel regression models with interactive fixed effects. Manuscript. Unpublished manuscript, University of Southern California.
- Moon, H., Weidner, M., 2010b. Linear regression for panel with unknown number of factors as interactive fixed effects. Unpublished manuscript, University of Southern California.
- Moon, H. R., Shum, M., Weidner, M., 2014. Estimation of random coefficients logit demand models with interactive fixed effects. Tech. rep., cemmap working paper, Centre for Microdata Methods and Practice.
- Neyman, J., Scott, E., 1948. Consistent estimates based on partially consistent observations. *Econometrica*, 1–32.
- Orchard, T., Woodbury, M. A., et al., 1972. A missing information principle: theory and applications. In: *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*. The Regents of the University of California.
- Pan, J., 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of financial economics* 63 (1), 3–50.
- Pastorello, S., Patilea, V., Renault, E., 2003. Iterative and recursive estimation in structural nonadaptive models. *Journal of Business & Economic Statistics* 21 (4), 449–509.
- Pesaran, M. H., 2006. Estimation and inference in large heterogeneous panels with a multi-factor error structure. *Econometrica* 74 (4), 967–1012.

A Results of Section 2

A.1 Panel probit with additive individual and time effects

In this setting, I consider the following model

$$\begin{aligned} Y_{it}^* &= X_{it}'\beta + \alpha_i + \gamma_t + \varepsilon_{it}, \\ Y_{it} &= \mathbf{1}\{Y_{it}^* \geq 0\}, \end{aligned} \quad (15)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. Here, Y_{it} is a scalar outcome variable of interest, X_{it} is a vector of explanatory variables, β is a finite-dimensional parameter vector, the variables α_i and γ_t are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks respectively.

Similarly to Section (2.1), I model the conditional distribution of Y_{it} using the single-index specification

$$P(Y_{it} = 1 | X_{it}, \beta, \alpha_i, \gamma_t) = \Phi(X_{it}\beta + \alpha_i + \gamma_t),$$

and for estimation I adopt a fixed effects approach treating the unobserved individual and time effects as parameters to be estimated. I collect all these effects in the vector $\phi_{NT} = (\alpha_1, \dots, \alpha_N, \gamma_1, \dots, \gamma_N)'$. The true values of the parameters are denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, \dots, \alpha_N^0, \gamma_1^0, \dots, \gamma_T^0)'$. Other quantities of interest involve averages over the data and unobserved effects

$$\delta_{NT}^0 = \mathbb{E}_\phi[\Delta_{NT}(\beta^0, \phi_{NT}^0)], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \beta, \alpha_i, \gamma_t), \quad (16)$$

and examples of partial effects (Δ) are the following:

Example A.1. (Average partial effects) If $X_{it,k}$, the k -th element of X_{it} , is binary, its partial effect for model (15) on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \beta, \alpha_i + \gamma_t) = \Phi(\beta_k + X_{it,-k}'\beta_{-k} + \alpha_i + \gamma_t) - \Phi(X_{it,-k}'\beta_{-k} + \alpha_i + \gamma_t), \quad (17)$$

where β_k is the k -th element of β , and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the k -th element. If $X_{it,k}$ is continuous, for model (15) the partial effects of $X_{it,k}$ on the conditional probability of Y_{it} is

$$\Delta(X_{it}, \alpha_i, \gamma_t) = \beta_k \phi_f(X_{it}'\beta + \alpha_i + \gamma_t), \quad (18)$$

where $\phi_f(\cdot)$ is the derivative of Φ .

Definition A.1. The fixed effect EM estimator for panel probit with additive fixed effects is defined by

- (1) Given initial $(\beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)})$, denote $\mu_{it}^{(k)} = X'_{it}\beta^{(k)} + \alpha_i^{(k)} + \gamma_t^{(k)}$,
- (2) **E-step:** Calculate

$$\begin{aligned}\hat{Y}_{it}^{(k)} &: = E[Y_{it}^* | Y_{it}, X_{it}, \beta^{(k)}, \alpha_i^{(k)}, \gamma_t^{(k)}] \\ &= \mu_{it}^{(k)} + (Y_{it} - \Phi(\mu_{it}^{(k)})) \cdot \phi_f(\mu_{it}^{(k)}) / \{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)}))\},\end{aligned}$$

- (3) **M-step:** This contains three conditional maximization steps

CM-step 1: Given α_i and γ_t , the parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it} X'_{it} \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it} (\hat{Y}_{it}^{(k)} - \alpha_i^{(k)} - \gamma_t^{(k)}) \right\},$$

CM-step 2: Given β and γ_t , the parameter α_i can be updated by

$$\alpha_i^{(k+1)} = \frac{1}{T} \sum_{t=1}^T (\hat{Y}_{it}^{(k)} - X'_{it}\beta^{(k+1)} - \gamma_t^{(k)}),$$

CM-step 3: Given β and α_i , the parameter γ_t can be updated by

$$\gamma_t^{(k+1)} = \frac{1}{N} \sum_{i=1}^N (\hat{Y}_{it}^{(k)} - X'_{it}\beta^{(k+1)} - \alpha_i^{(k+1)})$$

- (4) Iterate until convergence.

Note that the CM-step 2 and CM-step 3 here are just the average over time and individual using $\hat{Y}_{it}^{(k)}$ as surrogate for Y_{it}^* .

A.2 Proof of Proposition 2.1

By second-order Taylor expansion, for any two arguments θ_1 and θ_2 ,

$$-\log F(\theta_1) = -\log F(\theta_2) - \frac{\partial \log F(\theta_2)}{\partial \theta_2} (\theta_1 - \theta_2) - \frac{1}{2} \frac{\partial^2 \log F(\theta)}{\partial^2 \theta} \Big|_{\theta^*} (\theta_1 - \theta_2)^2.$$

Denote $h(\theta) = -\frac{\partial \log F(\theta)}{\partial \theta}$. Using the fact that $-\log F(q_{it}z_{it})$ is strictly convex on $(0, 1)$ for logit and probit, and simple calculation shows $0 < -\frac{\partial^2 \log F(\theta)}{\partial^2 \theta}|_{\theta^*} < 1$, one has

$$-\log F(\theta_1) \leq -\log F(\theta_2) + h(\theta_2)(\theta_1 - \theta_2) + \frac{1}{2}(\theta_1 - \theta_2)^2,$$

by completing the square, this can be written as

$$-\log F(\theta_1) \leq -\log F(\theta_2) + \frac{1}{2}(\theta_1 - \theta_2 + h(\theta_2))^2 - \frac{1}{2}h^2(\theta_2).$$

Now substitute $q_{it}(X'_{it}\beta + \alpha'_i\gamma_t)$ for θ_1 and $q_{it}(X'_{it}\tilde{\beta} + \tilde{\alpha}'_i\tilde{\gamma}_t)$ for θ_2 , one has

$$\begin{aligned} -\log F(q_{it}(X'_{it}\beta + \alpha'_i\gamma_t)) &\leq -\log F(q_{it}(X'_{it}\tilde{\beta} + \tilde{\alpha}'_i\tilde{\gamma}_t)) - \frac{1}{2}h^2(q_{it}(X'_{it}\tilde{\beta} + \tilde{\alpha}'_i\tilde{\gamma}_t)) \\ &\quad + \frac{1}{2}((X'_{it}\beta + \alpha'_i\gamma_t) - (X'_{it}\tilde{\beta} + \tilde{\alpha}'_i\tilde{\gamma}_t) + q_{it}h(q_{it}(X'_{it}\tilde{\beta} + \tilde{\alpha}'_i\tilde{\gamma}_t)))^2 \end{aligned}$$

sum over i and t to obtain the required results.

B Proofs of Section 3

B.1 Proof of Consistency for $\hat{\beta}_{PPIF}$

The proof contains two steps. In Step 1, I show the estimated index \tilde{z}_{it} is a good approximation to z_{it} with some structural error (the generalized residuals). In Step 2, I show the structural error satisfies the assumption in Bai (2009b) for linear panel data models with interactive fixed effects. With a little abuse of notation, in this section I use $\hat{\beta}$ to denote $\hat{\beta}_{PPIF}$ which is the estimate of the EM procedure for panel probit models.

Step 1. Denote $q_{it} = 2Y_{it} - 1$. I prove the consistence directly from the likelihood function

$$\ell_{it}(\beta, \alpha_i, \gamma_t) = \log \Phi(q_{it}(X'_{it}\beta + \alpha_i\gamma_t)), \quad \mathcal{L}_{NT} = \frac{1}{NT} \sum_{i,t} \ell_{it} = \sum_{i,t} \log \Phi(q_{it}z_{it}),$$

for any θ_1 and θ_2 , the following is an upper bound for the negative log-likelihood:

$$\begin{aligned} -\log\Phi(\theta_1) &\leq -\log\Phi(\theta_2) - \frac{\phi_f(\theta_2)}{\Phi(\theta_2)}(\theta_1 - \theta_2) + \frac{1}{2}(\theta_1 - \theta_2)^2 \\ &= -\log\Phi(\theta_2) + \frac{1}{2}(\theta_1 - \theta_2 - \frac{\phi_f(\theta_2)}{\Phi(\theta_2)})^2 - \frac{1}{2}(\frac{\phi_f(\theta_2)}{\Phi(\theta_2)})^2, \end{aligned}$$

where $\phi_f(\cdot)$ is the Gaussian density. Substitute $q_{it}z_{it}$ for θ_1 and $q_{it}\tilde{z}_{it}$ for θ_2 , then

$$-\log\Phi(q_{it}z_{it}) \leq -\log\Phi(q_{it}\tilde{z}_{it}) + \frac{1}{2}(z_{it} - \tilde{z}_{it} + q_{it}\frac{\phi_f(q_{it}\tilde{z}_{it})}{\Phi(q_{it}\tilde{z}_{it})})^2 - \frac{1}{2}(\frac{\phi_f(q_{it}\tilde{z}_{it})}{\Phi(q_{it}\tilde{z}_{it})})^2. \quad (19)$$

Note, from the proof here, one can also infer using $\tilde{z}_{it} = z_{it} + q_{it}\frac{\phi_f(q_{it}\tilde{z}_{it})}{\Phi(q_{it}\tilde{z}_{it})} = z_{it} + \frac{Y_{it} - \Phi(z_{it})}{\Phi(z_{it})(1 - \Phi(z_{it}))}\phi_f(q_{it}z_{it})$ is a good next step approximation, as the quadratic loss is a surrogate for the Bernoulli log-likelihood function.

Step 2. Denote the structural error (generalized residual) as $e_{it} = \frac{Y_{it} - \phi_f(z_{it})}{\Phi(z_{it})\Phi(z_{it})}\phi_f(q_{it}z_{it})$. One has $\mathbb{E}_\phi[e_{it}] = 0$. Since the estimated parameters minimize the objective function, with equation (19) one has

$$0 \geq \mathcal{L}_{NT}(\beta^0, \phi^0) - \mathcal{L}_{NT}(\hat{\beta}, \hat{\phi}) \geq \frac{1}{2NT} \sum_{i,t} [(z_{it}^0 - \hat{z}_{it} + e_{it})^2 - e_{it}^2]$$

The consistency proof for $\hat{\beta}$ is equivalent to that for the linear regression model with interactive fixed effects. In matrix notation, as in Section 4, the above inequality would be

$$\begin{aligned} \frac{1}{NT} \text{Tr}(e'e) &\geq \frac{1}{NT} \text{Tr}[(X'(\hat{\beta} - \beta^0) + \hat{\alpha}\hat{\gamma} - \alpha^0\gamma^0 - e)'(X'(\hat{\beta} - \beta^0) + \hat{\alpha}\hat{\gamma}' - \alpha^0\gamma^0 - e)] \\ &\geq \frac{1}{NT} \text{Tr}[(X'(\hat{\beta} - \beta^0) - e)'M_{(\hat{\alpha}, \alpha^0)}(X'(\hat{\beta} - \beta^0) - e)] \end{aligned}$$

where $M_{(\hat{\alpha}, \alpha^0)} = \mathbf{1}_T - (\hat{\alpha}, \alpha^0)[(\hat{\alpha}, \alpha^0)'(\hat{\alpha}, \alpha^0)]^{-1}(\hat{\alpha}, \alpha^0)'$ is the projector that projects orthogonal to $(\hat{\alpha}, \alpha^0)$.

With Assumption 1 (iv), which says that no linear combination of the regressors converges to zero, even after projecting any factor loading α , one has $\frac{1}{NT}\text{Tr}(Xe'e) = o_p(1)$, and $\mathbb{E}[e_{it}] = 0$. One can also check that $\|e\| = o_p(\sqrt{NT})$. The assumption $\frac{1}{NT}\text{Tr}(XX') = O_p(1)$ is

satisfied from the distributional assumption on the regressors above. One then has

$$\begin{aligned} \left| \frac{1}{NT} \text{Tr}(e' M_{(\hat{\alpha}, \alpha^0)} X_k) \right| &\leq \frac{1}{NT} |\text{Tr}(e' X_k)| + \frac{1}{NT} |\text{Tr}(e' P_{(\hat{\alpha}, \alpha^0)} X_k)| \\ &\leq o_p(1) + \frac{2}{NT} \|e\| \|X_k\| = o_p(1). \end{aligned}$$

Under these, one has

$$0 \geq c \|\hat{\beta} - \beta\| + o_p(\|\hat{\beta} - \beta^0\|) + o_p(1),$$

from which it is concluded that $\hat{\beta} = \beta^0 + o_p(1)$.

B.2 Proofs of Theorems 3.1 and 3.2

In the section, I suppress the dependence on NT of all the sequences of functions and parameters to lighten the notation, e.g. I write \mathcal{L} for \mathcal{L}_{NT} and ϕ for ϕ_{NT} . It is also convenient to introduce some notation that will be extensively used in the analysis. Let

$$\mathcal{S}(\beta, \phi) = \partial_\phi \mathcal{L}(\beta, \phi) \quad \mathcal{H}(\beta, \phi) = -\partial_{\phi\phi'} \mathcal{L}(\beta, \phi),$$

where $\partial_x f$ denotes the partial derivative of f with respect to x , and additional subscripts denote higher-order partial derivatives. I refer to the $\dim \phi$ -vector $\mathcal{S}(\beta, \phi)$ as the incidental parameter score, and to the $\dim \phi \times \dim \phi$ matrix $\mathcal{H}(\beta, \phi)$ as the incidental parameter Hessian. I omit the argument of the functions when they are evaluated at the true parameter values (β^0, ϕ^0) , e.g. $\mathcal{H} = \mathcal{H}(\beta^0, \phi^0)$. I use a bar to indicate expectations, e.g. $\partial_\beta \bar{\mathcal{L}} = \mathbb{E}[\partial_\beta \mathcal{L}]$, and a tilde to denote that the variables are in deviation with respect to their expectations, e.g. $\partial_\beta \tilde{\mathcal{L}} = \partial_\beta \mathcal{L} - \partial_\beta \bar{\mathcal{L}}$. For $c \geq 0$, I define the sets $\mathcal{B}(c, \beta^0) = \{\beta : \|\beta - \beta^0\|_\infty \leq c\}$, and $\mathcal{B}_q(c, \beta^0, \phi^0) = \{(\beta, \phi) : \|\beta - \beta^0\| < c, \|\phi - \phi^0\|_q < c\}$, which are closed balls of radius c around the true parameters β^0 and (β^0, ϕ^0) , respectively, under the L_2 norm and L_q -norm.

Analogous to Ξ_{it} defined in Eq (13), I define

$$\Lambda_{it} = -\frac{1}{NT} \sum_{j=1}^N \sum_{\tau=1}^T (\bar{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \gamma_\tau^0 \gamma_t^0 + \bar{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1} \alpha_j^0 \gamma_t^0 + \bar{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \gamma_\tau^0 \alpha_i^0 + \bar{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \alpha_j^0 \alpha_i^0) \partial_\pi \ell_{j\tau}$$

and analogous to $D_\beta \ell_{it}$ defined in the main text I also define $D_\beta \Delta_{it} = \partial_\beta \Delta_{it} - \partial_\pi \Delta_{it} \Xi_{it}$.

With a little abuse of notation, in this section I use $\hat{\beta}$ to denote $\hat{\beta}_{PPIF}$ which is the estimate of the EM procedure for panel probit models.

A close look at the iterative EM procedure yields

$$\begin{aligned}\hat{\beta}^{(k+1)} &= \left(\sum_{i,t} X_{it} X'_{it}\right)^{-1} \sum_{i,t} X_{it} (\hat{Y}_{it}^{(k)} - \hat{\alpha}_i^{(k)} \gamma_t^{(k)}) \\ &= \beta^{(k)} + (X'X)^{-1} \partial_{\beta} \mathcal{L}(\beta^{(k)}, \hat{\phi}(\beta^{(k)})),\end{aligned}\tag{20}$$

which depends on the score of the profile likelihood function.

For $r \geq 0$, define the sets $\mathcal{B}(r, \beta^0) = \{\beta : \|\beta - \beta^0\| \leq r\}$, and $\mathcal{B}_q(r, \phi^0) = \{\phi : \|\phi - \phi^0\|_q \leq r\}$, which are closed balls of radius r around the true parameter values β^0 and ϕ^0 , respectively.

Before going to the proof of Theorems 3.1 and 3.2, I first introduce two lemmas that will be used.

Lemma B.1. *(Asymptotic expansions of $\hat{\beta}$). Let Assumption 1 hold. Then*

$$\sqrt{NT}(\hat{\beta} - \beta^0) = \bar{W}_{\infty}^{-1} U + o_p(1),$$

where $U = U^{(0)} + U^{(1)}$, $\bar{W}_{\infty} := \lim_{N,T \rightarrow \infty} \bar{W}$ exists with $\bar{W}_{\infty} > 0$, and

$$\begin{aligned}\bar{W} &= -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\mathbb{E}_{\phi}(\partial_{\beta\beta'} \ell_{it}) + \mathbb{E}_{\phi}(-\partial_{\pi^2} \ell_{it}) \Xi_{it} \Xi'_{it}], \\ U^{(0)} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T D_{\beta} \ell_{it}, \\ U^{(1)} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \{-\Lambda_{it} [D_{\beta\pi} \ell_{it} - \mathbb{E}(D_{\beta\pi} \ell_{it})] + \frac{1}{2} \Lambda_{it}^2 \mathbb{E}(D_{\beta\pi^2} \ell_{it})\}.\end{aligned}$$

Proof. The proof follows from using Theorem B.1 of Fernández-Val and Weidner (2014) and applying Lemma D.1. From Theorem B.1 of Fernández-Val and Weidner (2014),

$$\sqrt{NT} \partial_{\beta} \mathcal{L}(\beta, \hat{\phi}(\beta)) = U - \bar{W} \sqrt{NT} (\beta - \beta^0) + R(\beta),$$

with

$$\bar{W} = -(\partial_{\beta\beta'} \bar{\mathcal{L}} + [\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} [\partial_{\phi\beta'} \bar{\mathcal{L}}]),$$

hence applying Lemma D.1 (ii) yields

$$\bar{W} = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T [\mathbb{E}_{\phi}(\partial_{\beta\beta'} \ell_{it}) + \mathbb{E}_{\phi}(-\partial_{\pi^2} \ell_{it}) \Xi_{it} \Xi'_{it}].\tag{21}$$

Similarly, applying Theorem B.1 of Fernández-Val and Weidner (2014) yields

$$U^{(0)} = \sqrt{NT}(\partial_\beta \mathcal{L} + [\partial_{\beta\phi'} \tilde{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathcal{S}),$$

$$\begin{aligned} U^{(1)} &= \sqrt{NT}([\partial_{\beta\phi'} \tilde{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathcal{S} - [\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S}) \\ &\quad + \sqrt{NT} \sum_{g=1}^{\dim \phi} (\partial_{\beta\phi' \phi_g} \bar{\mathcal{L}} + [\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} [\partial_{\phi\phi' \phi_g} \bar{\mathcal{L}}]) [\bar{\mathcal{H}}^{-1} \mathcal{S}] [\bar{\mathcal{H}}^{-1} \mathcal{S}]_g / 2. \end{aligned}$$

By using Lemma D.1 (i),

$$U^{(0)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T (\partial_\beta \ell_{it} - \Xi_{it} \partial_\pi \ell_{it}) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T D_\beta \ell_{it}. \quad (22)$$

Decompose $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$U^{(1a)} = \sqrt{NT}([\partial_{\beta\phi'} \tilde{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathcal{S} - [\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S}),$$

and

$$U^{(1b)} = \sqrt{NT} \sum_{g=1}^{\dim \phi} (\partial_{\beta\phi' \phi_g} \bar{\mathcal{L}} + [\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} [\partial_{\phi\phi' \phi_g} \bar{\mathcal{L}}]) [\bar{\mathcal{H}}^{-1} \mathcal{S}] [\bar{\mathcal{H}}^{-1} \mathcal{S}]_g / 2.$$

By using Lemma D.1 (i) and (iii),

$$U^{(1a)} = -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \Lambda_{it} (\partial_{\beta\pi} \tilde{\ell}_{it} + \Xi_{it} \partial_{\pi^2} \tilde{\ell}_{it}) = -\frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \Lambda_{it} [D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi(D_{\beta\pi} \ell_{it})],$$

and

$$U^{(1b)} = \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \Lambda_{it}^2 [\mathbb{E}_\phi(\partial_{\beta\pi^2} \ell_{it}) + [\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathbb{E}_\phi(\partial_\phi \partial_{\pi^2} \ell_{it})],$$

where for each i, t it is the case that $\partial_\phi \partial_{\pi^2} \ell_{it}$ is a $\dim \phi$ -vector, which can be written as

$$\partial_\phi \partial_{\pi^2} \ell_{it} = \begin{pmatrix} A 1_T \\ A' 1_N \end{pmatrix} \text{ for an } N \times T \text{ matrix } A \text{ with elements } A_{j\tau} = \partial_{\pi^3} \ell_{j\tau} \text{ if } j = i \text{ and } \tau = t,$$

and $A_{j\tau} = 0$ otherwise. Thus, again applying Lemma D.1(i) yields $[\partial_{\beta\phi'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \partial_\phi \partial_{\pi^2} \ell_{it} =$

$-\sum_{j,\tau} \Xi_{j\tau} \delta_{(i=j)} \delta_{(t=\tau)} \partial_{\pi^3} \ell_{it} = -\Xi_{it} \partial_{\pi^3} \ell_{it}$. Therefore

$$U^{(1b)} = \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \Lambda_{it}^2 \mathbb{E}(\partial_{\beta\pi^2} \ell_{it} - \Xi_{it} \partial_{\pi^3} \ell_{it}) = \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \Lambda_{it}^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it}),$$

hence

$$U^{(1)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \{-\Lambda_{it} [D_{\beta\pi} \ell_{it} - \mathbb{E}(D_{\beta\pi} \ell_{it})] + \frac{1}{2} \Lambda_{it}^2 \mathbb{E}(D_{\beta\pi^2} \ell_{it})\}. \quad (23)$$

□

Lemma B.2. (Asymptotic expansion of $\hat{\delta}$). Let Assumptions 1 and 2 hold and let $\|\hat{\beta} - \beta^0\| = O_p((NT)^{-1/2}) = o_p(r_\beta)$. Then

$$\sqrt{NT}(\hat{\delta} - \delta) = V_\Delta^{(0)} + V_\Delta^{(1)} + o_p(1),$$

where

$$\begin{aligned} V_\Delta^{(0)} &= \left[\frac{1}{NT} \sum_{i,t} \mathbb{E}_\phi(D_{\beta} \Delta_{it}) \right]' \bar{W}_\infty^{-1} U^{(0)} - \frac{1}{\sqrt{NT}} \sum_{i,t} \mathbb{E}_\phi(\Psi_{it}) \partial_\pi \ell_{it}, \\ V_\Delta^{(1)} &= \left[\frac{1}{NT} \sum_{i,t} \mathbb{E}_\phi(D_{\beta} \Delta_{it}) \right]' \bar{W}_\infty^{-1} U^{(1)} + \frac{1}{\sqrt{NT}} \sum_{i,t} \Lambda_{it} [\Psi_{it} \partial_{\pi^2} \ell_{it} - \mathbb{E}_\phi(\Psi_{it}) \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})] \\ &\quad + \frac{1}{2\sqrt{NT}} \sum_{i,t} \Lambda_{it}^2 [\mathbb{E}_\phi(\partial_{\pi^2} \ell_{it}) - \mathbb{E}_\phi(\partial_{\pi^3} \ell_{it}) \mathbb{E}_\phi(\Psi_{it})]. \end{aligned}$$

Proof. The proof follows from using Theorem B.4 of Fernández-Val and Weidner (2014) and applying Lemma D.1. Theorem B.4 of Fernández-Val and Weidner (2014) implies

$$\hat{\delta} - \delta = [\partial_{\beta'} \bar{\Delta} + (\partial_{\phi'} \bar{\Delta}) \bar{\mathcal{H}}^{-1} (\partial_{\phi\beta'} \bar{\mathcal{L}})] (\hat{\beta} - \beta^0) + U_\Delta^{(0)} + U_\Delta^{(1)} + o_p(1/\sqrt{NT}), \quad (24)$$

with

$$\begin{aligned} U_\Delta^{(0)} &= (\partial_{\phi'} \bar{\Delta}) \bar{\mathcal{H}}^{-1} \mathcal{S}, \\ U_\Delta^{(1)} &= (\partial_{\phi'} \tilde{\Delta}) \bar{\mathcal{H}}^{-1} \mathcal{S} - (\partial_{\phi} \bar{\Delta}) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &\quad + \frac{1}{2} \mathcal{S}' \bar{\mathcal{H}}^{-1} [\partial_{\phi\phi'} \bar{\Delta} + \sum_{g=1}^{\dim \phi} [\partial_{\phi\phi'\phi_g} \bar{\mathcal{L}}] [\bar{\mathcal{H}}^{-1} (\partial_{\phi} \bar{\Delta})]_g] \bar{\mathcal{H}}^{-1} \mathcal{S}. \end{aligned}$$

By using Lemma D.1,

$$\sqrt{NT}U_{\Delta}^{(0)} = -\frac{1}{\sqrt{NT}}\sum_{i,t}\mathbb{E}_{\phi}(\Psi_{it})\partial_{\pi}\ell_{it}, \quad (25)$$

$$\begin{aligned} \sqrt{NT}U_{\Delta}^{(1)} &= \frac{1}{\sqrt{NT}}\sum_{i,t}\Lambda_{it}[\Psi_{it}\partial_{\pi^2}\ell_{it} - \mathbb{E}_{\phi}(\Psi_{it})\mathbb{E}_{\phi}(\partial_{\pi^2}\ell_{it})] \\ &\quad + \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^2[\mathbb{E}_{\phi}(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_{\phi}(\partial_{\pi^3}\ell_{it})\mathbb{E}_{\phi}(\Psi_{it})]. \end{aligned} \quad (26)$$

From the proof of Lemma B.1 and the following proof of Theorem 3.1, it follows that $\sqrt{NT}(\hat{\beta} - \beta^0) = \bar{W}_{\infty}^{-1}U + o_p(1) = O_p(1)$, by Lemma D.1,

$$\sqrt{NT}[\partial_{\beta'}\bar{\Delta} + (\partial_{\phi'}\bar{\Delta})\bar{\mathcal{H}}^{-1}(\partial_{\phi\beta'}\bar{\mathcal{L}})](\hat{\beta} - \beta^0) = \left[\frac{1}{NT}\sum_{i,t}\mathbb{E}_{\phi}(D_{\beta}\Delta_{it})\right]'\bar{W}_{\infty}^{-1}(U^{(0)} + U^{(1)}) + o_p(1). \quad (27)$$

Combining equations 24, 25, 26 and 27 gives the result. \square

B.2.1 Proof of Asymptotics for $\hat{\beta}_{PPIF}$

I characterize the asymptotic distribution of $\hat{\beta}$ from the limit average Hessian \bar{W}_{∞} and the limiting distribution of the approximated score U . Next two steps are to get the eventual result.

Step 1 shows $U^{(0)} \xrightarrow{d} N(0, \bar{W}_{\infty})$. In the likelihood setting $\mathbb{E}\partial_{\beta}\mathcal{L} = 0$, $\mathbb{E}\mathcal{S} = 0$, and, by the Bartlett identities $\mathbb{E}(\partial_{\beta}\mathcal{L}\partial_{\beta'}\mathcal{L}) = -\frac{1}{NT}\partial_{\beta\beta'}\bar{\mathcal{L}}$, $\mathbb{E}(\partial_{\beta}\mathcal{L}\mathcal{S}') = -\frac{1}{NT}\partial_{\beta\phi'}\bar{\mathcal{L}}$, and $\mathbb{E}(\mathcal{S}\mathcal{S}') = \frac{1}{NT}\bar{\mathcal{H}}$. Denote $v = ((\alpha^0)', -(\gamma^0)')'$, $\mathcal{S}'v = 0$ and $\partial_{\beta\phi'}\bar{\mathcal{L}}v = 0$.

From the definitions $\bar{W} = -(\partial_{\beta\beta'}\bar{\mathcal{L}} + [\partial_{\beta\phi'}\bar{\mathcal{L}}]\bar{\mathcal{H}}^{-1}[\partial_{\phi\beta'}\bar{\mathcal{L}}])$ and $U^{(0)} = \sqrt{NT}(\partial_{\beta}\mathcal{L} + [\partial_{\beta\phi'}\bar{\mathcal{L}}]\bar{\mathcal{H}}^{-1}\mathcal{S})$,

$$\mathbb{E}(U^{(0)}) = 0, \quad Var(U^{(0)}) = \bar{W} \quad (28)$$

which implies $\lim_{N,T \rightarrow \infty} Var(U^{(0)}) = \bar{W}_{\infty}$.

According to Lemma B.1

$$U^{(0)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T D_{\beta}\ell_{it}, \quad (29)$$

where $D_\beta \ell_{it} := \partial_\beta \ell_{it} - \partial_\pi \ell_{it} \Xi_{it}$ is a martingale difference sequence for each i and independent across i , conditional on ϕ . Applying Theorem 2.3 in McLeish (1974) yields

$$U^{(0)} \xrightarrow{d} N[0, \lim_{N,T \rightarrow \infty} \text{Var}(U^{(0)})] \sim N(0, \bar{W}_\infty) \quad (30)$$

Step 2 shows that $U^{(1)} \rightarrow_P \kappa \bar{B}_\infty + \kappa^{-1} \bar{D}_\infty$. Since $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$U^{(1a)} = -\frac{1}{\sqrt{NT}} \sum_{i,t} \Lambda_{it} [D_{\beta\pi} \ell_{it} - E_\phi(D_{\beta\pi} \ell_{it})]$$

and

$$U^{(1b)} = \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \Lambda_{it}^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it})$$

Plugging-in the definition of Λ_{it} , I decompose $U^{(1a)} = U^{(1a,1)} + U^{(1a,2)} + U^{(1a,3)} + U^{(1a,4)}$, where

$$U^{(1a,1)} = \frac{1}{(NT)^{3/2}} \sum_{i,j} \bar{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \left(\sum_\tau \partial_\pi \ell_{j\tau} \gamma_\tau^0 \right) \sum_t (D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi D_{\beta\pi} \ell_{it}) \gamma_t^0,$$

$$U^{(1a,2)} = \frac{1}{(NT)^{3/2}} \sum_{t,j} \bar{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \left(\sum_\tau \partial_\pi \ell_{j\tau} \gamma_\tau^0 \right) \sum_i (D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi D_{\beta\pi} \ell_{it}) \alpha_i^0,$$

$$U^{(1a,3)} = \frac{1}{(NT)^{3/2}} \sum_{i,\tau} \bar{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1} \left(\sum_j \partial_\pi \ell_{j\tau} \alpha_j^0 \right) \sum_t (D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi D_{\beta\pi} \ell_{it}) \gamma_t^0,$$

$$U^{(1a,4)} = \frac{1}{(NT)^{3/2}} \sum_{t,\tau} \bar{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \left(\sum_j \partial_\pi \ell_{j\tau} \alpha_j^0 \right) \sum_i (D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi D_{\beta\pi} \ell_{it}) \alpha_i^0.$$

By the Cauchy-Schwarz inequality applied to the sum over t in $U^{(1a,2)}$,

$$\begin{aligned} (U^{(1a,2)})^2 &\leq \frac{1}{(NT)^3} \left[\sum_t \left(\sum_{j,\tau} \bar{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \partial_\pi \ell_{j\tau} \gamma_\tau^0 \right)^2 \right] \left[\sum_t \left(\sum_i (D_{\beta\pi} \ell_{it} - \mathbb{E} D_{\beta\pi} \ell_{it}) \alpha_i^0 \right)^2 \right] \\ &= \frac{1}{(NT)^3} \left[\sum_t O_p(NT) \right] \left[\sum_t O_p(N) \right] = O_p(1/N) = o_p(1) \end{aligned}$$

Using that both $\bar{\mathcal{H}}_{(\gamma\alpha)}^{-1} \partial_\pi \ell_{j\tau} \gamma_\tau^0$ and $(D_{\beta\pi} \ell_{it} - \mathbb{E} D_{\beta\pi} \ell_{it}) \alpha_i^0$ are mean zero, independent across

i .

Therefore, $U^{(1a,2)} = o_p(1)$. Analogously $U^{(1a,3)} = o_p(1)$.

According to Lemma B.5, it is the case that $\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} = -\text{diag}[(\frac{1}{NT} \sum_{t=1}^T \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it}(\gamma_t^0)^2))^{-1}] + O_p(1)$. Analogously to the proof of $U^{(1a,2)}$, the $O_p(1)$ part of $\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}$ has an asymptotically negligible contribution to $U^{(1a,1)}$. Thus,

$$\begin{aligned} U^{(1a,1)} &= \frac{1}{(NT)^{3/2}} \sum_{i,j} \overline{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \left(\sum_{\tau} \partial_{\pi} \ell_{j\tau} \gamma_{\tau}^0 \right) \sum_t (D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi D_{\beta\pi} \ell_{it}) \gamma_t^0 \\ &= -\frac{1}{(NT)^{1/2}} \sum_i \frac{(\sum_{\tau} \partial_{\pi} \ell_{i\tau} \gamma_{\tau}^0) \sum_t (D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi D_{\beta\pi} \ell_{it}) \gamma_t^0}{\sum_{t=1}^T \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it}(\gamma_t^0)^2)} + o_p(1) \end{aligned}$$

previous assumptions guarantee that $\mathbb{E}_\phi[(U_i^{(1a,1)})^2] = \mathcal{O}_p(1)$, uniformly over i . Note that both the denominator and the numerator of $U_i^{(1a,1)}$ are of order T . For the denominator this is obvious because of the sum over T . For the numerator there are two sums over T , but both $\partial_{\pi} \ell_{i\tau} \gamma_{\tau}^0$ and $(D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi(D_{\beta\pi} \ell_{it})) \gamma_t^0$ are mean zero weakly correlated processes, the sum over which is of order \sqrt{T} each. By applying the WLLN over i , $\frac{1}{N} \sum_i U_i^{(1a,1)} = \frac{1}{N} \mathbb{E}_\phi U_i^{(1a)} + o_p(1)$, and therefore

$$U^{(1a,1)} = \underbrace{-\sqrt{\frac{N}{T}} \frac{1}{N} \sum_{i=1}^N \frac{\sum_{t=1}^T \sum_{\tau=t}^T \mathbb{E}_\phi(\partial_{\pi} \ell_{it} D_{\beta\pi} \ell_{i\tau} \gamma_t^0 \gamma_{\tau}^0)}{\sum_{t=1}^T \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it}(\gamma_t^0)^2)}}_{\equiv \sqrt{\frac{N}{T}} \overline{B}^{(1)}} + o_p(1).$$

Here, I use that $\mathbb{E}_\phi(\partial_{\pi} \ell_{it} D_{\beta\pi} \ell_{i\tau}) = 0$ for $t > \tau$. Analogously,

$$U^{(1a,4)} = \underbrace{-\sqrt{\frac{T}{N}} \frac{1}{T} \sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}_\phi(\partial_{\pi} \ell_{it} D_{\beta\pi} \ell_{it}(\alpha_i^0)^2)}{\sum_{i=1}^N \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it}(\alpha_i^0)^2)}}_{\equiv \sqrt{\frac{T}{N}} \overline{D}^{(1)}} + o_p(1).$$

hence $U^{(1a)} = \kappa \overline{B}^{(1)} + \kappa^{-1} \overline{D}^{(1)} + o_p(1)$.

Next, I analyze $U^{(1b)}$. I decompose $\Lambda_{it} = \Lambda_{it}^{(1)} + \Lambda_{it}^{(2)} + \Lambda_{it}^{(3)} + \Lambda_{it}^{(4)}$, where

$$\begin{aligned}\Lambda_{it}^{(1)} &= -\frac{1}{NT} \sum_{j=1}^N \overline{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} \gamma_t^0 \sum_{\tau=1}^T \partial_\pi \ell_{j\tau} \gamma_\tau^0, & \Lambda_{it}^{(2)} &= -\frac{1}{NT} \sum_{j=1}^N \overline{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} \alpha_i^0 \sum_{\tau=1}^T \partial_\pi \ell_{j\tau} \gamma_\tau^0, \\ \Lambda_{it}^{(3)} &= -\frac{1}{NT} \sum_{\tau=1}^T \overline{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1} \gamma_t^0 \sum_{j=1}^N \partial_\pi \ell_{j\tau} \alpha_j^0, & \Lambda_{it}^{(4)} &= -\frac{1}{NT} \sum_{\tau=1}^T \overline{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1} \alpha_i^0 \sum_{j=1}^N \partial_\pi \ell_{j\tau} \alpha_j^0.\end{aligned}$$

This decomposition of Λ_{it} includes the following decomposition of $U^{(1b)}$

$$U^{(1b)} = \sum_{p,q=1}^4 U^{(1b,p,q)}, \quad U^{(1b,p,q)} = \frac{1}{2\sqrt{NT}} \sum_{i,t} \Lambda_{it}^{(p)} \Lambda_{it}^{(q)} \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it}).$$

Due to symmetry $U^{(1b,p,q)} = U^{(1b,q,p)}$ this is a decomposition into 10 distinct terms. Consider $U^{(1b,1,2)}$,

$$U^{(1b,1,2)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N U_i^{(1b,1,2)}, \text{ with}$$

$$U_i^{(1b,1,2)} = \frac{1}{2T} \sum_{t=1}^T \gamma_t^0 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it}) \frac{1}{N^2} \sum_{j_1, j_2=1}^N \overline{\mathcal{H}}_{(\alpha\alpha)ij_1}^{-1} \overline{\mathcal{H}}_{(\gamma\alpha)tj_2}^{-1} \alpha_i^0 \left(\frac{1}{\sqrt{T}} \sum_{\tau=1}^T \partial_\pi \ell_{j_1\tau} \gamma_\tau^0 \right) \left(\frac{1}{\sqrt{T}} \sum_{\tau=1}^T \partial_\pi \ell_{j_2\tau} \gamma_\tau^0 \right).$$

Using $\mathbb{E}_\phi(\sum_t \partial_\pi \ell_{it} \gamma_t^0) = 0$, $\mathbb{E}_\phi(\sum_t \partial_\pi \ell_{it} \gamma_t^0 \sum_j \partial_{j\tau} \gamma_\tau^0)$ for $i \neq j$, and the properties of the inverse expected Hessian from Theorem B.5 one finds $\mathbb{E}_\phi[U_i^{(1b,1,2)}] = O_p(1/N)$, uniformly over i , and $\mathbb{E}_\phi[(U_i^{(1b,1,2)})^2] = O_p(1)$, uniformly over i , and $\mathbb{E}_\phi[U_i^{(1b,1,2)} U_j^{(1b,1,2)}] = O_p(1/N)$, uniformly over $i \neq j$. This implies that $\mathbb{E}_\phi U^{(1b,1,2)} = O_p(1/N)$, and $\mathbb{E}_\phi[(U^{(1b,1,2)} - \mathbb{E}_\phi U^{(1b,1,2)})^2] = O_p(1/\sqrt{N})$, and therefore $U^{(1b,1,2)} = o_p(1)$. By similar arguments one obtains $U^{(1b,p,q)} = o_p(1)$ for all combinations of $p, q = 1, 2, 3, 4$, except for $p = q = 1$ and $p = q = 4$.

For $p = q = 1$, $U^{(1b,1,1)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^N U_i^{(1b,1,1)}$, and

$$U_i^{(1b,1,1)} = \frac{1}{2T} \sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it}) \frac{1}{N^2} \sum_{j_1, j_2=1}^N \overline{\mathcal{H}}_{(\alpha\alpha)ij_1}^{-1} \overline{\mathcal{H}}_{(\alpha\alpha)ij_2}^{-1} \left(\frac{1}{\sqrt{T}} \sum_{\tau=1}^T \partial_\pi \ell_{j_1\tau} \gamma_\tau^0 \right) \left(\frac{1}{\sqrt{T}} \sum_{\tau=1}^T \partial_\pi \ell_{j_2\tau} \gamma_\tau^0 \right).$$

Analogous to the result for $U^{(1b,1,2)}$ one finds $\mathbb{E}_\phi[(U^{(1b,1,1)} - \mathbb{E}_\phi U^{(1b,1,1)})^2] = O_p(1/\sqrt{N})$, and therefore $U^{(1b,1,1)} = \mathbb{E}_\phi U^{(1b,1,1)} + o_p(1)$.

Furthermore,

$$\begin{aligned}
\mathbb{E}_\phi U^{(1b,1,1)} &= \frac{1}{2\sqrt{NT}} \sum_{i=1}^N \frac{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it}) \mathbb{E}_\phi[(\partial_\pi \ell_{it} \gamma_t^0)^2]}{[\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})]^2} + o(1) \\
&= -\underbrace{\sqrt{\frac{N}{T}} \frac{1}{2N} \sum_{i=1}^N \frac{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it})}{\sum_{t=1}^T (\gamma_t^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})}}_{\equiv \sqrt{\frac{N}{T}} \bar{B}^{(2)}} + o(1),
\end{aligned}$$

analogously,

$$U^{(1b,4,4)} = \mathbb{E}_\phi U^{(1b,4,4)} + o_p(1) = -\underbrace{\sqrt{\frac{T}{N}} \frac{1}{2T} \sum_{t=1}^T \frac{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it})}{\sum_{i=1}^N (\alpha_i^0)^2 \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it})}}_{\equiv \sqrt{\frac{T}{N}} \bar{D}^{(2)}} + o_p(1),$$

thus $U^{(1b)} = \kappa \bar{B}^{(2)} + \kappa^{-1} \bar{D}^{(2)} + o_p(1)$.

Since $\bar{B}_\infty = \lim_{N,T \rightarrow \infty} [\bar{B}^{(1)} + \bar{B}^{(2)}]$ and $\bar{D}_\infty = \lim_{N,T \rightarrow \infty} [\bar{D}^{(1)} + \bar{D}^{(2)}]$, then $U^{(1)} = \kappa \bar{B}_\infty + \kappa^{-1} \bar{D}_\infty + o_p(1)$.

I have shown $U^{(0)} \xrightarrow{d} N(0, \bar{W}_\infty)$, and $U^{(1)} \xrightarrow{p} \kappa \bar{B}_\infty + \kappa^{-1} \bar{D}_\infty$. Using this and Lemma B.1 I obtain

$$\sqrt{NT}(\hat{\beta} - \beta^0) \xrightarrow{d} \bar{W}_\infty^{-1} N(\kappa \bar{B}_\infty + \kappa^{-1} \bar{D}_\infty, \bar{W}_\infty).$$

B.2.2 Proof of asymptotic distribution of APE

I consider the case of scalar Δ_{it} to simplify the notation. Decompose

$$\sqrt{NT}(\hat{\delta} - \delta_{NT}^0 - \bar{B}_\infty^\delta/T - \bar{D}_\infty^\delta/N) = \sqrt{NT}(\delta - \delta_{NT}^0) + \sqrt{NT}(\hat{\delta} - \delta - \bar{B}_\infty^\delta/T - \bar{D}_\infty^\delta/N).$$

Part (1): Limit of $\sqrt{NT}(\hat{\delta} - \delta - \bar{B}_\infty^\delta/T - \bar{D}_\infty^\delta/N)$. An argument analogous to the proof of 3.1 using Lemma B.2 yields

$$\sqrt{NT}(\hat{\delta} - \delta) \xrightarrow{d} N(\kappa \bar{B}_\infty^\delta + \kappa^{-1} \bar{D}_\infty^\delta, \bar{V}_\infty^{\delta(1)}),$$

where $\bar{V}_\infty^{\delta(1)} = \mathbb{E}\{(NT)^{-1} \sum_{i,t} \mathbb{E}_\phi[\Gamma_{it}^2]\}$, for the expressions of \bar{B}_∞^δ , \bar{D}_∞^δ , and Γ_{it} given in the

statement of the theorem. Then, by Mann-Wald theorem

$$\sqrt{NT}(\hat{\delta} - \delta - \bar{B}_\infty^\delta/T - \bar{D}_\infty^\delta/N) \xrightarrow{d} N(0, \bar{V}_\infty^{\delta(1)}).$$

Part (2): Limit of $\sqrt{NT}(\delta - \delta_{NT}^0)$. Here I show that $\sqrt{NT}(\delta - \delta_{NT}^0) \xrightarrow{d} N(0, \bar{V}_\infty^{\delta(2)})$ and characterize the asymptotic variance $\bar{V}_\infty^{\delta(2)}$. I characterize $\bar{V}_\infty^{\delta(2)}$ as $\bar{V}_\infty^{\delta(2)} = \mathbb{E}\{NT\mathbb{E}[(\delta - \delta_{NT}^0)^2]\}$, because $\mathbb{E}[\delta - \delta_{NT}^0] = 0$. Note, the rate \sqrt{NT} is determined through $\mathbb{E}[(\delta - \delta_{NT}^0)^2]$, where

$$\mathbb{E}[(\delta - \delta_{NT}^0)^2] = \mathbb{E}\left[\left(\frac{1}{NT} \sum_{i,t} \tilde{\Delta}_{it}\right)^2\right] = \frac{1}{N^2T^2} \sum_{i,j,t,s} \mathbb{E}[\tilde{\Delta}_{it}\tilde{\Delta}_{js}], \quad (31)$$

for $\tilde{\Delta}_{it} = \Delta_{it} - \mathbb{E}(\Delta_{it})$. The order of $\mathbb{E}[(\delta - \delta_{NT}^0)^2]$ is equal to the number of terms of the sums in equation (31) that are nonzero, which is determined by the sample properties of $\{(X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T\}$. Under Assumption 2(i)

$$\mathbb{E}[(\delta - \delta_{NT}^0)^2] = \frac{1}{N^2T^2} \sum_{i,t,s} \mathbb{E}[\tilde{\Delta}_{it}\tilde{\Delta}_{is}] = O(N^{-1}),$$

because $\{\tilde{\Delta}_{it} : 1 \leq i \leq N; 1 \leq t \leq T\}$ is independent across i and α -mixing across t .

#Part(3): Limit of $\sqrt{NT}(\hat{\delta} - \delta_{NT}^0 - T^{-1}\bar{B}_\infty^\delta - N^{-1}\bar{D}_\infty^\delta)$. The conclusion of the Theorems follows because $(\delta - \delta_{NT}^0)$ and $(\hat{\delta} - \delta - T^{-1}\bar{B}_\infty^\delta - N^{-1}\bar{D}_\infty^\delta)$ are asymptotically independent and $\bar{V}_\infty^\delta = \bar{V}_\infty^{\delta(2)} + \bar{V}_\infty^{\delta(1)}$.

B.3 Proofs of Theorems 3.3 and 3.4

I start by stating a lemma that is going to be used for this section. It corresponds to Lemma C.2 of Fernández-Val and Weidner (2014) and the proof is omitted for brevity.

Lemma B.3. *Let $G(\beta, \phi) := \frac{1}{N(T-j)} \sum_{i,t \geq j+1} g(X_{it}, X_{i,t-j}, \beta, \alpha_i \gamma_t, \alpha_i \gamma_{t-j})$ for $0 \leq j < T$, and $\mathcal{B}_\varepsilon^0$ be a subset of $\mathbb{R}^{\dim \beta + 2}$ that contains an ε -neighborhood of $(\beta, \pi_{it}^0, \pi_{i,t-j}^0)$ for all i, t, j, N, T , and for some $\varepsilon > 0$. Assume that $(\beta, \pi_1, \pi_2) \rightarrow g_{itj}(\beta, \pi_1, \pi_2) := g(X_{it}, X_{i,t-j}, \beta, \pi_1, \pi_2)$ is Lipschitz continuous over $\mathcal{B}_\varepsilon^0$ a.s., i.e. $|g_{itj}(\beta_1, \pi_{11}, \pi_{21}) - g_{itj}(\beta_0, \pi_{10}, \pi_{20})| \leq M_{itj} \|(\beta_1, \pi_{11}, \pi_{21}) - (\beta_0, \pi_{10}, \pi_{20})\|$ for all $(\beta_1, \pi_{11}, \pi_{21}) \in \mathcal{B}_\varepsilon^0$, $(\beta_0, \pi_{10}, \pi_{20}) \in \mathcal{B}_\varepsilon^0$, and some $M_{itj} = O_p(1)$ for all i, t, j, N, T . Let $(\hat{\beta}, \hat{\phi})$ be an estimator of (β, ϕ) such that $\|\hat{\beta} - \beta^0\| \xrightarrow{p} 0$ and $\|\hat{\phi} - \phi^0\|_\infty \xrightarrow{p} 0$. Then,*

$$G(\hat{\beta}, \hat{\phi}) \xrightarrow{p} \mathbb{E}[G(\beta^0, \phi^0)],$$

provided that the limit exists.

This lemma shows the consistency of the estimators of averages of the data and parameters. I will use this result to show the validity of the analytical bias corrections and the consistency of the variance estimators.

B.3.1 Proof of Theorem 3.3

I separate the proof in two parts corresponding to the two statements of the theorem.

Part I: Proof of $\hat{W} \xrightarrow{p} \bar{W}_\infty$. The asymptotic variance and its estimators can be expressed as $\bar{W}_\infty = \mathbb{E}[W(\beta^0, \phi^0)]$ and $\widehat{W} = W(\hat{\beta}, \hat{\phi})$, where $W(\beta, \phi)$ has a first order representation as a continuously differentiable transformation of terms that have the form of $G(\beta, \phi)$ in Lemma B.3. The result then follows by the continuous mapping theorem noting that $\|\hat{\beta} - \beta^0\| \xrightarrow{p} 0$ and $\|\hat{\phi} - \phi^0\|_\infty \xrightarrow{p} 0$.

Part II: Proof of $\sqrt{NT}(\tilde{\beta}^A - \beta^0) \xrightarrow{d} N(0, \bar{W}_\infty^{-1})$. I show that $\hat{B} \xrightarrow{p} \bar{B}_\infty$ and $\hat{D} \xrightarrow{p} \bar{D}_\infty$. These asymptotic biases and their fixed effects estimators are either time-series averages of fractions of cross-sectional averages, or vice versa. The nesting of the averages makes the analysis a bit more cumbersome than the analysis of \widehat{W} , but the results follows by similar standard arguments, also using that $L \rightarrow \infty$ and $L/T \rightarrow 0$ guarantee that the trimmed estimator in \hat{B} is also consistent for the spectral expectations; see Lemma 6 in Hahn and Kuersteiner (2011).

B.3.2 Proof of Theorem 3.4

I separate the proof into two parts corresponding to the two statements of the theorem.

Part I: $\hat{V}^\delta \xrightarrow{p} \bar{V}_\infty^\delta$. \bar{V}_∞^δ and \hat{V}^δ have a similar structure to \bar{W}_∞ and \widehat{W} in part I of the proof of Theorem 3.3, so that the consistency follows by an analogous argument.

Part II: $\sqrt{NT}(\tilde{\delta}^A - \delta_{NT}^0) \xrightarrow{d} N(0, \bar{V}_\infty^\delta)$. As in the proof of Theorem 3.2, I decompose

$$\sqrt{NT}(\tilde{\delta}^A - \delta_{NT}^0) = \sqrt{NT}(\delta - \delta_{NT}^0) + \sqrt{NT}(\tilde{\delta}^A - \delta).$$

Then, by Mann-Wald theorem,

$$\sqrt{NT}(\tilde{\delta}^A - \delta) = \sqrt{NT}(\hat{\delta} - \hat{B}^\delta/T - \hat{D}^\delta/N - \delta) \xrightarrow{d} N(0, \bar{V}_\infty^{\delta(1)}),$$

provided that $\hat{B}^\delta \xrightarrow{p} \bar{B}_\infty^\delta$ and $\hat{D}^\delta \xrightarrow{p} \bar{D}_\infty^\delta$, and $\sqrt{NT}\delta - \delta_{NT}^0 \xrightarrow{d} N(0, \bar{V}_\infty^{\delta(2)})$, where $\bar{V}_\infty^{\delta(1)}$ and $\bar{V}_\infty^{\delta(2)}$ are defined as in the proof of Theorem 3.2. The statement thus follows by using a

similar argument to part II of the proof of Theorem 3.3 to show the consistency of \hat{B}^δ and \hat{D}^δ , and because $(\delta - \delta_{NT}^0)$ and $(\tilde{\delta}^A - \delta)$ are asymptotically independent, and $\bar{V}_\infty^\delta = \bar{V}^{\delta(2)} + \bar{V}^{\delta(1)}$.

B.4 Properties of the Inversed Expected Incidental Parameter Hessian

The following two lemmas would be used in the proof of asymptotic distributions of β and δ .

Lemma B.4. *Let Assumption 1 hold, then $\|\bar{\mathcal{H}}_{(\alpha\alpha)}^{-1}\bar{\mathcal{H}}_{(\alpha\gamma)}\|_\infty < 1 - \frac{b_{\min}}{b_{\max}}$, and $\|\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1}\bar{\mathcal{H}}_{(\gamma\alpha)}\|_\infty < 1 - \frac{b_{\min}}{b_{\max}}$.*

Proof. Let $h_{it} = \mathbb{E}(-\partial_{\pi^2}\ell_{it})$, Assumption 1 guarantees that $b_{\min} \leq h_{it} \leq b_{\max}$, therefore

$$\begin{aligned} \|\bar{\mathcal{H}}_{(\alpha\alpha)}^{-1}\bar{\mathcal{H}}_{(\alpha\gamma)}\|_\infty &= \max_i \frac{\sum_t |\alpha_i^0 \gamma_t^0 h_{it}|}{\sum_t (\gamma_t^0)^2 h_{it}} = 1 - \max_i \frac{\sum_t ((\gamma_t^0)^2 - |\alpha_i^0 \gamma_t^0|) h_{it}}{\sum_t (\gamma_t^0)^2 h_{it}} \\ &\leq 1 - \frac{\|\gamma^0\|^2 - \min_i |\alpha_i^0| \|\gamma^0\|_1}{\|\gamma^0\|^2} \frac{b_{\min}}{b_{\max}} \end{aligned}$$

similar,

$$\begin{aligned} \|\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1}\bar{\mathcal{H}}_{(\gamma\alpha)}\|_\infty &= \max_t \frac{\sum_i |\alpha_i^0 \gamma_t^0 h_{it}|}{\sum_i (\alpha_i^0)^2 h_{it}} = 1 - \max_t \frac{\sum_i ((\alpha_i^0)^2 - |\alpha_i^0 \gamma_t^0|) h_{it}}{\sum_i (\alpha_i^0)^2 h_{it}} \\ &\leq 1 - \frac{\|\alpha^0\|^2 - \min_t |\gamma_t^0| \|\alpha^0\|_1}{\|\alpha^0\|^2} \frac{b_{\min}}{b_{\max}} \end{aligned}$$

Since $\|\alpha^0\|^2 \geq \frac{1}{N} \|\alpha^0\|_1^2$, as long as $\frac{1}{N} \|\alpha^0\|_1 \geq \min_t |\gamma_t^0|$, $\|\bar{\mathcal{H}}_{(\alpha\alpha)}^{-1}\bar{\mathcal{H}}_{(\alpha\gamma)}\|_\infty \leq 1 - \frac{b_{\min}}{b_{\max}}$; similarly since $\|\gamma^0\|^2 \geq \frac{1}{T} \|\gamma^0\|_1^2$, as long as $\frac{1}{T} \|\gamma^0\|_1 \geq \min_i |\alpha_i^0|$, $\|\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1}\bar{\mathcal{H}}_{(\gamma\alpha)}\|_\infty \leq 1 - \frac{b_{\min}}{b_{\max}}$. \square

Lemma B.5. *Under Assumption 1,*

$$\|\bar{\mathcal{H}}^{-1} - \text{diag}(\bar{\mathcal{H}}_{(\alpha\alpha)}, \bar{\mathcal{H}}_{(\gamma\gamma)})^{-1}\|_{\max} = O_p(1).$$

Proof. By the inversion formula for partitioned matrices

$$\bar{\mathcal{H}}^{-1} = \begin{pmatrix} A & -A\bar{\mathcal{H}}_{(\alpha\gamma)}\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1} \\ -\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1}\bar{\mathcal{H}}_{(\gamma\alpha)}A & \bar{\mathcal{H}}_{(\gamma\gamma)}^{-1} + \bar{\mathcal{H}}_{(\gamma\gamma)}^{-1}\bar{\mathcal{H}}_{(\gamma\alpha)}A\bar{\mathcal{H}}_{(\alpha\alpha)}\bar{\mathcal{H}}_{(\gamma\gamma)}^{-1} \end{pmatrix},$$

with

$$\begin{aligned} A &\equiv (\overline{\mathcal{H}}_{(\alpha\alpha)} - \overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^{-1} = \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}(\mathbf{I} - \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^{-1} \\ &= \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \sum_{n=0}^{\infty} (\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^n. \end{aligned}$$

Define

$$B \equiv \sum_{n=1}^{\infty} (\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^n,$$

then $A = \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} + \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}B$. By using the matrix norm property that $\|AB\|_{\max} \leq \|A\|_{\infty}\|B\|_{\max}$ and Lemma B.4

$$\begin{aligned} \|B\|_{\max} &\leq \sum_{n=1}^{\infty} (\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)})^n \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty} \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty} \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\max} \\ &\leq \left[\sum_{n=1}^{\infty} \left(1 - \frac{b_{\min}}{b_{\max}}\right)^{2n} \right] T \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty} \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty} \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max}^2 = O(N^{-1}). \end{aligned}$$

From this I obtain

$$\|A\|_{\infty} \leq \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty} + N \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty} \|B\|_{\max} = O(N).$$

From the different blocks of

$$\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1} = \begin{pmatrix} A - \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} & -A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \\ -\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)}A & \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)}A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \end{pmatrix}$$

it can be seen that

$$\begin{aligned} \|A - \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\max} &= \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}B\|_{\max} \leq \|\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1}\|_{\infty} \|B\|_{\max} = O_p(1), \\ \|-A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\max} &\leq \|A\|_{\infty} \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty} = O_p(1) \\ \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)}A\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\max} &\leq \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty}^2 \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\infty} \|A\|_{\infty} \|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \\ &\leq N \|\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\|_{\infty}^2 \|A\|_{\infty} \|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\max}^2 = O_p(1) \end{aligned}$$

Having the bound $O_p(1)$ for the max-norm of each block of the matrix yields also the same bound for the max-norm of the matrix itself, as desired. \square

This result establishes that $\overline{\mathcal{H}}^{-1}$ can be uniformly approximated by a diagonal matrix, which is given by the inverse of the diagonal terms of $\overline{\mathcal{H}}$. The diagonal elements of $\text{diag}(\overline{\mathcal{H}}_{(\alpha\alpha)}, \overline{\mathcal{H}}_{(\gamma\gamma)})^{-1}$ are of order N and T respectively, hence the order of difference established by the lemma is relatively small.

With this result, $\|\overline{\mathcal{H}}^{-1}\|_\infty \leq \|\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1}\|_\infty + \|\overline{\mathcal{D}}^{-1}\|_\infty \leq (N + T)\|\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1}\|_{\max} + \|\overline{\mathcal{D}}^{-1}\|_\infty = O_p(N)$ which can be used to verify the assumption in the proof of Theorem B.1 of Fernández-Val and Weidner (2014).

C Proof of Section 4

C.1 Compare with existing methods

C.1.1 Proof of Proposition 4.1

The proof is mainly for the case without unobserved effects, but similarly argument can be used to the proof of other cases.

The model looks $Y_{it} = \mathbf{1}\{X'_{it}\beta + \varepsilon_{it} \geq 0\}$, and ε_{it} is normally distributed with variance 1. When estimating the structural parameter of probit using MLE,

$$\beta \in \arg \max_{\beta \in \Theta} \mathcal{L}_{NT} = \sum_{i,t} \ell_{it} = \sum_{i,t} Y_{it} \log \Phi(X'_{it}\beta) + (1 - Y_{it}) \log(1 - \Phi(X'_{it}\beta)),$$

and then the score of β is

$$\sum_{i,t} X_{it} \underbrace{\left\{ Y_{it} \frac{\phi_f(X'_{it}\beta)}{\Phi(X'_{it}\beta)} - (1 - Y_{it}) \frac{\phi_f(X'_{it}\beta)}{1 - \Phi(X'_{it}\beta)} \right\}}_{\tilde{g}_{it}(\beta)} = 0 \Leftrightarrow \sum_{i,t} X_{it} \left\{ \frac{Y_{it} - \Phi(X'_{it}\beta)}{\Phi(X'_{it}\beta)(1 - \Phi(X'_{it}\beta))} \phi_f(X'_{it}\beta) \right\} = 0,$$

which relates to the generalized residuals part of EM,

$$\begin{aligned} \hat{Y}_{it} &= X_{it}\beta + \underbrace{Y_{it} \cdot \phi_f(X_{it}\beta)/\Phi(X_{it}\beta) - (1 - Y_{it}) \cdot \phi_f(X_{it}\beta)/\{1 - \Phi(X_{it}\beta)\}}_{g_{it}(\beta)}, \\ &= X_{it}\beta + (Y_{it} - \Phi(X_{it}\beta)) \cdot \phi_f(X_{it}\beta)/\{\Phi(X_{it}\beta)(1 - \Phi(X_{it}\beta))\}, \end{aligned}$$

and

$$\beta = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it} X'_{it} \right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X'_{it} \hat{Y}_{it} \right\}.$$

Denote $\mu_{it}^{(k)} = X'_{it}\beta^{(k)}$, the score function is of β is zero, i.e. the unique fixed-point property, means that,

$$\sum_{i=1}^N \sum_{t=1}^T X'_{it} ((Y_{it} - \Phi(X'_{it}\beta)) \cdot \phi_f(X'_{it}\beta) / \{\Phi(X'_{it}\beta)(1 - \Phi(X'_{it}\beta))\}) = 0 \Rightarrow \beta^{(k)} = \beta^0,$$

this is due to the identification condition that

$$E^0[g_{it}(\beta^0)|X_{it}] = E^0[E[\varepsilon_{it}|Y_{it}, X_{it}, \beta^0]|X_{it}] = E^0[\varepsilon_{it}|X_{it}] = 0.$$

By central limit theory for the score

$$\sqrt{NT}E[\nabla_{\beta}l_{it}] = \sqrt{NT}E[\sum_{i,t} X_{it}g_{it}(\beta)] \xrightarrow{d} N(0, E\frac{\phi_{it}^2}{\Phi_{it}(1 - \Phi_{it})}X_{it}X'_{it}),$$

with $Var(\sum_{i,t} X_{it}\tilde{g}_{it}(\beta)) = Var(\sum_{i,t} X_{it}\frac{Y_{it}-\Phi(X_{it}\beta)}{\Phi(X_{it}\beta)(1-\Phi(X_{it}\beta))}\phi_f(X'_{it}\beta))$. Since $Var(Y_{it}-\Phi(X'_{it}\beta)|X_{it}) = \Phi(X'_{it}\beta)(1 - \Phi(X'_{it}\beta))$,

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, [E\frac{\phi_{it}^2}{\Phi_{it}(1 - \Phi_{it})}X_{it}X'_{it}]^{-1})$$

for both EM and MLE.

C.1.2 Proof of Proposition 4.2

This is to show the difference between the proposed fixed effects EM-type estimator and the Newton's method as described in Greene (2003).

From the **E**-step, one has $\hat{Y}_{it}^{(k)} = X'_{it}\beta^{(k)} + \alpha_i^{(k)} + \underbrace{\frac{Y_{it} - \Phi(\mu_{it}^{(k)})}{\Phi(\mu_{it}^{(k)})(1 - \Phi(\mu_{it}^{(k)}))}\phi_{it}(\mu_{it}^{(k)})}_{g_{it}^{(k)}}$.

For fixed effects EM-type estimator, given α_i , parameter β can be updated by

$$\beta^{(k+1)} = \left(\sum_{i=1}^N \sum_{t=1}^T X_{it}X'_{it}\right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it}(\hat{Y}_{it}^{(k)} - \alpha_i^{(k)}) \right\} = \beta^{(k)} + \underbrace{\left(\sum_{i=1}^N \sum_{t=1}^T X_{it}X'_{it}\right)^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T X_{it}g_{it}^{(k)} \right\}}_{\Delta_{\beta EM}^{(k)}},$$

hence α_i can be updated by

$$\alpha_i^{(k+1)} = \frac{1}{T} \sum_{t=1}^T (\hat{Y}_{it}^{(k)} - X'_{it} \beta^{(k+1)}) = \alpha_i^{(k)} + g_{ii}^{(k)} - \frac{1}{T} \sum_{t=1}^T X'_{it} \Delta_{\beta_{EM}}^{(k)}.$$

For Newton's method as described in Greene (2003) Chapter 21

$$\beta^{(k+1)} = \beta^{(k)} - \left\{ \sum_{i=1}^N \sum_{t=1}^T h_{it} (X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i)' \right\}^{-1} \left\{ \sum_{i=1}^N \sum_{t=1}^T g_{it}^{(k)} (X_{it} - \bar{X}_i) \right\} = \beta^{(k)} + \Delta_{\beta_{NR}}^{(k)},$$

and

$$\alpha_i^{(k+1)} = \alpha_i^{(k)} - g_{ii}^{(k)} / h_{ii}^{(k)} - \bar{X}'_i \Delta_{\beta_{NR}}^{(k)},$$

here $h_{it} = g'_{it} = \frac{\phi_f(z_{it} q_{it})}{\Phi(z_{it} q_{it})} - \left(\frac{\phi_f(z_{it} q_{it})}{\Phi(z_{it} q_{it})} \right)^2$, $z_{it} = X'_{it} \beta + \alpha_i$, $q_{it} = 1 - 2Y_{it}$, $h_{ii} = \sum_{t=1}^T h_{it}$, and $g_{ii} = \sum_{t=1}^T g_{it}$. The sign difference is due to that h_{it} is negative for all values of $z_{it} q_{it}$.

C.2 Proof of Consistency for general $\hat{\beta}$

In general, the consistency proof will contain two steps as shown in the proof of PPIF.

Denote $z_{it} = X'_{it} \beta + \alpha_i \gamma_t$, under the bounded from below of the second order derivatives assumption

$$\forall y \in \mathcal{Y}, z \in \mathcal{Z} : b_{min} < \partial_{z^2} \mathcal{L}(y, z),$$

also assume that \mathcal{Z} is convex, i.e. since $\mathcal{Z} \subset \mathbb{R}$ it is an interval (either open or closed). From this it follows that for all $z_1, z_2 \in \mathcal{Z}$ one has

$$\begin{aligned} \mathcal{L}(y, z_1) - \mathcal{L}(y, z_2) &= [\partial_z \mathcal{L}(y, z_1)](z_1 - z_2) + \frac{1}{2} [\partial_{z^2} \mathcal{L}(y, \tilde{z})](z_1 - z_2)^2 \\ &\geq [\partial_z \mathcal{L}(y, z_1)](z_1 - z_2) + \frac{b_{min}}{2} (z_1 - z_2)^2 \\ &= \frac{b_{min}}{2} (z_1 - z_2 + \frac{1}{b_{min}} [\partial_z \mathcal{L}(y, z_1)])^2 - \frac{1}{2b_{min}} [\partial_z \mathcal{L}(y, z_1)]^2, \end{aligned}$$

where $z_1 \leq \tilde{z} \leq z_2$. Define $\hat{z}_{it} = z_{it}(\hat{\beta}, \hat{\alpha}_i, \hat{\gamma}_t)$, and $e_{it} = \frac{1}{b_{min}} [\partial_z \mathcal{L}_{it}]$. Note that $\mathbb{E}(e_{it}) = 0$. Since the estimated parameters minimize the objective function, observe that

$$\begin{aligned}
0 &\geq \mathcal{L}_{NT}(\beta^0, \phi^0) - \mathcal{L}_{NT}(\hat{\beta}, \hat{\phi}) = \frac{1}{NT} \sum_{i,t} [\mathcal{L}_{it}(z_{it}^0) - \mathcal{L}_{it}(\hat{z}_{it})] \\
&\geq \frac{b_{min}}{2NT} \sum_{i,t} [(z_{it}^0 - \hat{z}_{it} + e_{it})^2 - e_{it}^2] = \frac{b_{min}}{2NT} \sum_{i,t} \{[X'_{it}(\hat{\beta} - \beta^0) + \hat{\alpha}_i \hat{\gamma}_t - \alpha_i^0 \gamma_t^0 - e_{it}]^2 - e_{it}^2\}.
\end{aligned}$$

Once the last inequality is obtained, the consistency proof for $\hat{\beta}$ is equivalent to that for the linear regression model with interactive fixed effects. In matrix notation, the above inequality reads

$$\begin{aligned}
\frac{1}{NT} Tr(e'e) &\geq \frac{1}{NT} Tr[(X'(\hat{\beta} - \beta^0) + \hat{\alpha} \hat{\gamma} - \alpha^0 \gamma^0 - e)'(X'(\hat{\beta} - \beta^0) + \hat{\alpha} \hat{\gamma} - \alpha^0 \gamma^0 - e)] \\
&\geq \frac{1}{NT} Tr[(X'(\hat{\beta} - \beta^0) - e)'M_{(\hat{\alpha}, \alpha^0)}(X'(\hat{\beta} - \beta^0) - e)]
\end{aligned}$$

where $M_{(\hat{\alpha}, \alpha^0)} = \mathbf{1}_T - (\hat{\alpha}, \alpha^0)[(\hat{\alpha}, \alpha^0)'(\hat{\alpha}, \alpha^0)]^{-1}(\hat{\alpha}, \alpha^0)'$ is the projector that projects orthogonal to $(\hat{\alpha}, \alpha^0)$.

The assumptions on the panel model already guarantee that $\frac{1}{NT} Tr(Xe') = o_P(1)$. One can furthermore show that $\|e\| = o_P(\sqrt{NT})$, also the assumption $\frac{1}{NT} Tr(XX') = O_p(1)$ is satisfied from the distribution assumption on the regressors above. Then,

$$\begin{aligned}
\left| \frac{1}{NT} Tr(e' M_{(\hat{\alpha}, \alpha^0)} X_k) \right| &\leq \frac{1}{NT} |Tr(e' X_k)| + \frac{1}{NT} |Tr(e' P_{(\hat{\alpha}, \alpha^0)} X_k)| \\
&\leq o_p(1) + \frac{2}{NT} \|e\| \|X_k\| = o_p(1).
\end{aligned}$$

Under these, one has

$$0 \geq c \|\hat{\beta} - \beta\| + o_p(\|\hat{\beta} - \beta^0\|) + o_p(1)$$

from which $\hat{\beta} = \beta^0 + o_p(1)$.

D Some useful algebraic results

For any $N \times T$ matrix A , define the $N \times T$ matrix $\mathbb{P}A$ as follows

$$(\mathbb{P}A)_{it} = \alpha_i^0 \gamma_t^0 (\alpha_i^* + \gamma_t^*), \quad (\alpha^*, \gamma^*) \in \arg \min_{\alpha_i, \gamma_t} \sum_{i,t} \mathbb{E}(-\partial_{\pi^2} \ell_{it}) (A_{it} - \alpha_i^0 \gamma_t^0 (\alpha_i + \gamma_t))^2.$$

Here, the minimization is over $\alpha \in \mathbb{R}^N$ and $\gamma \in \mathbb{R}^T$, and \mathbb{P} is the projection operator. It is a linear projection, i.e. $\mathbb{P}\mathbb{P} = \mathbb{P}$. It is also convenient to define

$$\tilde{\mathbb{P}}A = \mathbb{P}\tilde{A}, \quad \text{where} \quad \tilde{A}_{it} = \frac{A_{it}}{\mathbb{E}(-\partial_{\pi^2} \ell_{it})}. \quad (32)$$

$\tilde{\mathbb{P}}$ is a linear operator, but not a projection. Note that Ξ and Λ defined before can be written as $\Xi_k = \tilde{\mathbb{P}}B_k$ and $\Lambda = \tilde{\mathbb{P}}C$, where $C_{it} = -\partial_{\pi} \ell_{it}$ and $B_{k,it} = -\mathbb{E}_{\phi}(\partial_{\beta_k \pi} \ell_{it})$, for $k = 1, \dots, \dim \beta$.

¹⁰

The linear operator $\tilde{\mathbb{P}}$ is closely related to the projection operator \mathbb{P} . The following lemma shows how in the context of panel probit model some expressions that regularly appear in the general expansions can conveniently be expressed by using the operator $\tilde{\mathbb{P}}$.

Lemma D.1. *Let A , B and C be $N \times T$ matrices, and let the expected incidental parameter Hessian $\bar{\mathcal{H}}$ be invertible. Define the $N + T$ vectors \mathcal{A} and \mathcal{B} and the $(N + T) \times (N + T)$ matrix \mathcal{C} as follows*

$$\mathcal{A} = \frac{1}{NT} \begin{pmatrix} A\gamma^0 \\ A'\alpha^0 \end{pmatrix}, \quad \mathcal{B} = \frac{1}{NT} \begin{pmatrix} B\gamma^0 \\ B'\alpha^0 \end{pmatrix},$$

and

$$\mathcal{C} = \frac{1}{NT} \begin{pmatrix} \text{diag}(C(\gamma^0 \circ \gamma^0)) & C \circ (\alpha^0 (\gamma^0)') \\ (C \circ (\alpha^0 (\gamma^0)'))' & \text{diag}(C'(\alpha^0 \circ \alpha^0)) \end{pmatrix}$$

where \circ denotes the Hadamard product, i.e., element-by-element product. Then

- (i) $\mathcal{A}'\bar{\mathcal{H}}^{-1}\mathcal{B} = \frac{1}{NT} \sum_{i,t} (\tilde{\mathbb{P}}A_{it})B_{it} = \frac{1}{NT} \sum_{i,t} (\tilde{P}B)_{it}A_{it}$,
- (ii) $\mathcal{A}'\bar{\mathcal{H}}^{-1}\mathcal{B} = \frac{1}{NT} \sum_{i,t} \mathbb{E}(-\partial_{\pi^2} \ell_{it}) (\tilde{\mathbb{P}}A)_{it} (\tilde{\mathbb{P}}B)_{it}$,
- (iii) $\mathcal{A}'\bar{\mathcal{H}}^{-1}\mathcal{C}\bar{\mathcal{H}}^{-1}\mathcal{B} = \frac{1}{NT} \sum_{i,t} (\tilde{\mathbb{P}}A)_{it} C_{it} (\tilde{\mathbb{P}}B)_{it}$.

¹⁰ B_k and Ξ_k are $N \times T$ matrices with entries $B_{k,it}$ and $\Xi_{k,it}$ respectively, while B_{it} and Ξ_{it} are $\dim \beta$ -vectors with entries $B_{k,it}$ and $\Xi_{k,it}$.

Proof. Let $\alpha_i^0 \gamma_t^0 (\tilde{\alpha}_i^* + \tilde{\gamma}_t^*) = (\mathbb{P}\tilde{A})_{it} = (\tilde{\mathbb{P}}A)_{it}$, with \tilde{A} as defined in eq (32). The FOC of the minimization problem in the definition of $(\mathbb{P}\tilde{A})_{it}$ can be written as $\overline{\mathcal{H}} \begin{pmatrix} \alpha^0 \circ \tilde{\alpha}^* \\ \gamma^0 \circ \tilde{\gamma}^* \end{pmatrix} = \mathcal{A}$.

One solution to this is $\begin{pmatrix} \alpha^0 \circ \tilde{\alpha}^* \\ \gamma^0 \circ \tilde{\gamma}^* \end{pmatrix} = \overline{\mathcal{H}}^{-1} \mathcal{A}$. Therefore,

$$\mathcal{A}' \overline{\mathcal{H}}^{-1} \mathcal{B} = \begin{pmatrix} \alpha^0 \circ \tilde{\alpha}^* \\ \gamma^0 \circ \tilde{\gamma}^* \end{pmatrix}' \mathcal{B} = \frac{1}{NT} \sum_{i,t} \alpha_i^0 \gamma_t^0 (\tilde{\alpha}_i^* + \tilde{\gamma}_t^*) B_{it} = \frac{1}{NT} \sum_{i,t} (\tilde{\mathbb{P}}A)_{it} B_{it}.$$

This is the first equality of the Statement (i) in the lemma. The second equality of Statement (i) follows by symmetry. Statement (ii) is a special case of Statement (iii) with $\mathcal{C} = \overline{\mathcal{H}}$, so Statement (iii) needs to be proved.

Let $\alpha_i^0 \gamma_t^0 (\alpha_i^* + \gamma_t^*) = (\mathbb{P}\tilde{B})_{it} = (\tilde{\mathbb{P}}B)_{it}$, where $\tilde{B}_{it} = \frac{B_{it}}{\mathbb{E}(-\partial_{\pi_2} \ell_{it})}$. Analogous to the above, choose $\begin{pmatrix} \alpha^0 \circ \alpha^* \\ \gamma^0 \circ \gamma^* \end{pmatrix} = \overline{\mathcal{H}}^{-1} \mathcal{B}$ as one solution to the minimization problem. Then

$$\mathcal{A}' \overline{\mathcal{H}}^{-1} \mathcal{C} \overline{\mathcal{H}}^{-1} \mathcal{B} = \frac{1}{NT} \sum_{i,t} (\alpha_i^0 \gamma_t^0)^2 [\tilde{\alpha}_i^* C_{it} \alpha_i^* + \tilde{\gamma}_t^* C_{it} \alpha_i^* + \tilde{\alpha}_i^* C_{it} \gamma_t^* + \tilde{\gamma}_t^* C_{it} \gamma_t^*] = \sum_{i,t} (\tilde{\mathbb{P}}A)_{it} C_{it} (\tilde{\mathbb{P}}B)_{it}$$

□

Table 1: Finite Sample Properties of Static Probit Estimators, N=100

| Model | Estimator | Bias | Std.Dev. | RMSE | SE/SD | P;95 |
|--------|-----------|-------|----------|-------|-------|------|
| T=8 | | | | | | |
| No FE | EM | 0.26 | 7.48 | 7.49 | 1.03 | 0.97 |
| | glm | 0.69 | 7.59 | 7.61 | 1.02 | 0.96 |
| FE i | EM | 20.74 | 10.37 | 23.18 | 0.73 | 0.29 |
| | glm | 22.38 | 11.73 | 25.26 | 0.85 | 0.39 |
| Add-FE | EM | 20.73 | 9.24 | 22.69 | 0.86 | 0.28 |
| | glm | 29.21 | 13.95 | 32.36 | 0.83 | 0.32 |
| IF | | 8.95 | 10.08 | 13.47 | 0.72 | 0.69 |
| | BC-IF | -4.69 | 8.91 | 10.06 | 0.81 | 0.84 |
| T=12 | | | | | | |
| No FE | EM | -0.10 | 6.01 | 6.02 | 1.04 | 0.96 |
| | glm | 0.31 | 6.09 | 6.09 | 1.03 | 0.96 |
| FE i | EM | 12.53 | 7.61 | 14.65 | 0.79 | 0.45 |
| | glm | 13.43 | 8.11 | 15.68 | 0.89 | 0.53 |
| Add-FE | EM | 10.88 | 6.62 | 12.73 | 0.99 | 0.64 |
| | glm | 20.81 | 10.20 | 23.17 | 0.89 | 0.38 |
| IF | | 7.64 | 6.94 | 10.32 | 0.83 | 0.73 |
| | BC-IF | -0.45 | 6.42 | 6.43 | 0.9 | 0.92 |
| T=20 | | | | | | |
| No FE | EM | 0.11 | 4.93 | 4.94 | 0.98 | 0.94 |
| | glm | 0.52 | 5.00 | 5.02 | 0.97 | 0.95 |
| FE i | EM | 6.44 | 5.22 | 8.28 | 0.85 | 0.67 |
| | glm | 7.20 | 5.50 | 9.06 | 0.95 | 0.70 |
| Add-FE | EM | 3.56 | 4.60 | 5.82 | 1.02 | 0.89 |
| | glm | 10.88 | 6.57 | 12.71 | 0.93 | 0.60 |
| IF | | 4.03 | 4.86 | 6.31 | 0.90 | 0.83 |
| | BC-IF | -0.99 | 4.62 | 4.72 | 0.95 | 0.94 |

Notes: All the entries are in percentage of the true parameter value. 500 replications.

Table 2: Finite Sample Properties of Static Probit Estimators, N=52

| Model | Estimator | Bias | Std.Dev. | RMSE | SE/SD | P;:95 |
|--------|-----------|-------|----------|-------|-------|-------|
| T=14 | | | | | | |
| No FE | EM | -0.02 | 7.83 | 7.84 | 1.03 | 0.94 |
| | glm | 0.43 | 7.97 | 7.98 | 1.01 | 0.95 |
| FE i | EM | 11.3 | 9.55 | 14.79 | 0.81 | 0.68 |
| | glm | 12.47 | 10.53 | 16.31 | 0.9 | 0.77 |
| Add-FE | EM | 2.92 | 7.74 | 8.27 | 1.02 | 0.94 |
| | glm | 24.05 | 15.28 | 28.48 | 0.8 | 0.53 |
| IF | | 4.8 | 9.28 | 10.44 | 0.79 | 0.83 |
| | BC-IF | -3.56 | 8.52 | 9.22 | 0.86 | 0.87 |
| T=26 | | | | | | |
| No FE | EM | -0.13 | 5.92 | 5.92 | 0.99 | 0.94 |
| | glm | 0.27 | 5.99 | 5.99 | 0.99 | 0.94 |
| FE i | EM | 4.88 | 6 | 7.73 | 0.88 | 0.85 |
| | glm | 5.33 | 6.21 | 8.17 | 0.98 | 0.89 |
| Add-FE | EM | 0.53 | 5.63 | 5.65 | 1 | 0.95 |
| | glm | 10.94 | 8.08 | 13.59 | 0.93 | 0.7 |
| IF | | 3.43 | 6.28 | 7.16 | 0.85 | 0.87 |
| | BC-IF | -1.3 | 5.96 | 6.09 | 0.9 | 0.92 |
| T=52 | | | | | | |
| No FE | EM | -0.18 | 4.22 | 4.22 | 0.98 | 0.95 |
| | glm | 0.22 | 4.27 | 4.27 | 0.98 | 0.95 |
| FE i | EM | 2.2 | 4.07 | 4.62 | 0.91 | 0.89 |
| | glm | 2.48 | 4.2 | 4.88 | 1 | 0.92 |
| Add-FE | EM | 1.21 | 3.97 | 4.15 | 1 | 0.94 |
| | glm | 6.99 | 5.17 | 8.69 | 0.96 | 0.71 |
| IF | | 1.5 | 3.91 | 4.18 | 0.96 | 0.91 |
| | BC-IF | -1.48 | 3.78 | 4.05 | 0.99 | 0.94 |

Notes: All the entries are in percentage of the true parameter value. 500 replications.

Table 3: Coefficients of Static Probit Model for Trade

| | (1) | (2) | (3) | (4) |
|------------------|--------|--------|--------|--------|
| Distance | -0.185 | -0.177 | -0.294 | -0.297 |
| Border | 0.161 | 0.152 | -0.027 | -0.041 |
| Island | -0.175 | -0.178 | -0.153 | -0.16 |
| Landlock | -0.357 | -0.358 | -0.471 | -0.474 |
| Legal | -0.308 | -0.309 | -0.208 | -0.212 |
| Language | 0.08 | 0.079 | 0.166 | 0.173 |
| Colony | 2.222 | 2.245 | 2.06 | 1.962 |
| Currency | -0.446 | -0.449 | -0.158 | -0.19 |
| FTA | 1.685 | 1.629 | 1.645 | 1.648 |
| Religion | 0.2 | 0.191 | 0.367 | 0.36 |
| Importer effects | | Yes | Yes | Yes |
| Exporter effects | | | Yes | Yes |
| Interactive | | | | Yes |