MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Binary Response Models for Heterogeneous Panel Data with Interactive Fixed Effects

Jiti Gao, Fei Liu and Bin Peng

December 2020

Working Paper 44/20

AACSB ACCREDITED EQUIS ACCREDITED AMBA ACCREDITED

Binary Response Models for Heterogeneous Panel Data with Interactive Fixed Effects

Jiti Gao[♯], Fei Liu[*], and Bin Peng[†]

[♯ †]Monash University and [*]Nankai University

December 6, 2020

Abstract

In this paper, we investigate binary response models for heterogeneous panel data with interactive fixed effects by allowing both the cross sectional dimension and the temporal dimension to diverge. From a practical point of view, the proposed framework can be applied to predict the probability of corporate failure, conduct credit rating analysis, etc. Theoretically and methodologically, we establish a link between a maximum likelihood estimation and a least squares approach, provide a simple information criterion to detect the number of factors, and achieve the asymptotic distributions accordingly. In addition, we conduct intensive simulations to examine the theoretical findings. In the empirical study, we focus on the sign prediction of stock returns, and then use the results of sign forecast to conduct portfolio analysis. By implementing rolling-window based out–of–sample forecasts, we show the finite–sample performance and demonstrate the practical relevance of the proposed model and estimation method.

Keywords: Binary Response, Heterogeneous Panel, Interactive Fixed Effects, Portfolio Analysis

JEL classification: C18, C23, G11

[♯] Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia. Email: Jiti.Gao@monash.edu

[*] School of Finance, Nankai University, Tianjin, China, 300190. Email: feiliu.econ@outlook.com

[†] Department of Econometrics and Business Statistics, Monash University, Caulfield East, VIC 3145, Australia. Email: Bin.Peng@monash.edu

1 Introduction

Varieties of binary response panel data models have been proposed and studied over the past a couple of decades, and earlier developments date back at least to Chamberlain (1984) and the references therein. The challenges in the previous studies often arise due to the identification issues caused by short time periods of data and non-closed form estimators (e.g., Manski, 1987; Chamberlain, 2010 among others). With the rise and availability of big and rich datasets, recent studies on binary response panel data models gradually shift focuses to the cases where both the cross-sectional dimension and the temporal dimension are allowed to diverge. An excellent review is given seen in Fernández-Val and Weidner (2018). Recently, an important strand of the literature is devoted to binary response models with interactive fixed effects (e.g., Boneva and Linton, 2017; Wang, 2019; Chen, Fernández-Val and Weidner, 2020). Within these studies, central questions are a) how to determine the number of factors ? and b) how to estimate the coefficients together with the factors and the factor loadings ?

For linear additive models, the aforementioned questions have been addressed well in the literature by utilizing different techniques. For example, using principal component analysis (PCA) and random matrix theory, Bai and Ng (2002), Lam and Yao (2012) and Ahn and Horenstein (2013) are able to detect the number of factors for large panel data using information criteria or eigenvalue ratio tests; Pesaran (2006) introduces a common correlated effects (CCE) estimator to take an advantage of a factor structure involving both dependent and independent variables. Meanwhile, Bai (2009), and Moon and Weidner (2015) develop alternative methods to estimate the coefficients together with the factors and the factor loadings; Li et al. (2020) and Huang et al. (2020) use the maximum likelihood method and the classifier-Lasso method to consider cases associated with heterogeneous coefficients respectively; and so forth.

However, for non-linear panel data models, especially for binary response panel data models associated with interactive fixed effects, there is limited progress, which is mainly due to the fact that a variety of tools adopted for linear additive models may no longer be directly applicable and useful. Below, we comment on the relevant literature. In Boneva and Linton (2017), the authors extend the CCE approach to a framework with binary responses, in which a key step is to estimate unobservable factors from the regressors. As a consequence, the approach requires an explicit structure of the regressors, and the usual limitation of a CCE type estimator occurs, e.g., the number of unobservable factors cannot be larger than the number of regressors (cf., Boneva and Linton, 2017, eq. 7). Wang (2019), and Chen, Fernández-Val and Weidner (2020) propose similar solutions to binary response panel data models, and the main difference is that the former does not include any regressors in the model. Thereby, we may regard Wang (2019), and Chen, Fernández-Val and Weidner (2020) as the binary response counterparts of Bai and

Ng (2002) and Bai (2009), respectively. Recently, Ando and Bai (2020), Ando and Lu (2020) and Chen, Dolado and Gonzalo (2020) bring attention to additive panel data models with interactive fixed effects, in which closed form estimators are less obvious. Specifically, quantile regressions are investigated in all three papers, of which only Ando and Lu (2020) include regressors in the investigation, while Ando and Bai (2020) propose a Bayesian approach.

In view of the relevant literature, we aim to contribute to binary response panel data models with interactive fixed effects by incorporating heterogeneous coefficients. From a modelling perspective, our setting is similar to Boneva and Linton (2017), but we do not require to impose a specific structure on the regressors, which allows us to avoid adding any restriction between the number of regressors and the number of unobservable factors. Our investigation establishes a link between a maximum likelihood estimation and a least squares approach. As a consequence, the identification restrictions provided in Bai (2009) and Moon and Weidner (2015) are readily to be applied to binary response models with very minor modifications. Our setting also improves those considered in Wang (2019), and Chen, Fernández-Val and Weidner (2020) by covering a broader family of distributions (see Appendix A for details). Meanwhile, we are able to estimate the unobservable factors and factor loadings consistently with asymptotic distributions. This may be considered as the binary response counterpart of those established in Bai and Ng (2013). In addition, we propose a simple information criterion to detect the number of factors for binary response panel data models. Last but not least, we conduct intensive numerical studies to examine the theoretical findings, and demonstrate the practical relevance.

From a practical perspective, the proposed framework can be relevant and applicable to to the following fields for instance. Predicting the probability of corporate failure has gained its attention since the seminal work of Altman (1968). Along this line of research, our paper provides a more generalized framework to extend those panel data driven studies (e.g., Caggiano et al., 2014). Similarly, our model and estimation method can also be applied to panel data based credit rating analysis (e.g., Jones et al., 2015). In an empirical study, we pay particular attention on portfolio analysis. More often than not, in order to calculate the optimal weights assigned to each stock, one adopts all stocks to construct a covariance matrix (e.g., Chen et al., 2019; Engle et al., 2019), which then naturally falls into the category of high dimensional covariance matrix estimation. Thereby, to boost the estimation accuracy, we see the increasing popularity of methods using rank reduction (Pelger and Xiong, 2019), penalization (Chen et al., 2019), or both (Fan et al., 2013) among others. It is worth emphasizing that by default the aforementioned techniques eventually include all stocks in practice, though some of the stocks may have relatively small weights compared to the others. As pointed out in Christoffersen and Diebold (2006) and Nyberg (2011), the sign of stock market returns may be predictable even if the returns themselves are not predictable. Therefore, an intuitive thought for portfolio analysis is to remove the stocks presenting high probabilities of having negative returns, and then construct the portfolio using those left. It is also in line with the motivation of big data econometrics (Varian, 2014), which is to implement certain type of "selection" in order to filter out the irrelevant information. Thus, we first use the proposed model and methodology to predict the sign of the return for each stock, and then use those having high probabilities of being positive to construct the portfolio. By implementing rolling-window out of sample forecasts, we demonstrate the practical relevance of our study.

In summary, the main contributions of this paper are as follows. (i) We propose a binary panel data model with both heterogeneous coefficients and interactive fixed effects; (ii) we develop a simultaneous estimation procedure to estimate the unknown coefficients and unobservable factors and factor loadings before we establish asymptotic distributions for the proposed estimators; and (iii) we demonstrate the practical relevance and general applicability of the proposed model and estimation by both simulated and real data examples.

The structure of this paper is as follows. Section 2 proposes our model, develops the methodology, and then establishes the associated asymptotic results. Section 3 provides an intensive simulation study to examine the finite–sample performance of the proposed model and estimation method. Section 4 discusses an empirical portfolio analysis. Section 5 concludes. Appendix A discusses a technical improvement presented by our work. The proofs are relegated to Appendix B of the paper.

Before proceeding, we introduce some notations that will be used repeatedly in the paper. $\|\cdot\|$ denotes the Euclidean norm of a vector or the Frobenius norm of a matrix; $O(1)$ always stands for a finite positive constant, and may be different at each appearance; $\to_P$ and $\to_D$ stand for convergence in probability and convergence in distribution respectively; $\Pr(A \mid B)$ represents the probability of the event A occurring conditional on the event B; for a matrix W with full column rank, let $P_W = W(W'W)^{-1}W'$.

2 The Model and Methodology

In this section, we present the model and the methodology, establish the associated asymptotic results, and further present an algorithm for numerical implementation in practice. First, we provide the basic setup in Section 2.1, and explain the difference between our work and some existing ones. In Section 2.2, we present the relevant asymptotic results. Section 2.3 provides a numerical implement for practical work.

2.1 The Setup

Specifically, we consider the next binary response panel data model with interactive fixed effects.

$$y_{it} = \begin{cases} 1, & x_{it}'\beta_{0i} + \gamma_{0i}' f_{0t} - \varepsilon_{it} \geq 0 \\ 0, & \text{otherwise} \end{cases}, \tag{2.1}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In the model (2.1), we observe the binary dependent variable y_{it} and the $d_\beta \times 1$ explanatory variables x_{it} with d_β being finite, and we also suppose that the probability density function (PDF) and the cumulative distribution function (CDF) of the error term ε_{it} are known. Specifically, we denote the PDF and the CDF as $g_\varepsilon(\cdot)$ and $G_\varepsilon(\cdot)$ respectively. Both the factor loading γ_{0i} and the factor f_{0t} are $d_f \times 1$ but unknown, where d_f is finite, and needs to be determined by the data. In what follows, we are interested in recovering (B_0, F_0, Γ_0), where $B_0 = (\beta_{01}, \ldots, \beta_{0N})'$, $F_0 = (f_{01}, \ldots, f_{0T})'$ and $\Gamma_0 = (\gamma_{01}, \ldots, \gamma_{0N})'$.

Remark 2.1. *From the modelling perspective, assuming that the distribution of ε_{it} is fully known does not loss any generality. Consider the Probit model as an example. An intuitive question might be "what happens when heteroskedasticity occurs, i.e., $\varepsilon_{it} \sim N(0, \sigma_i^2)$ with unknown σ_i". Then we can always rewrite the model as follows.*

$$y_{it} = \begin{cases} 1, & x_{it}'\beta_{0i}^* + \gamma_{0i}^{*\prime} f_{0t} - \varepsilon_{it}^* \geq 0 \\ 0, & \textit{otherwise} \end{cases}, \tag{2.2}$$

where $\beta_{0i}^* = \frac{\beta_{0i}}{\sigma_i}$, $\gamma_{0i}^* = \frac{\gamma_{0i}}{\sigma_i}$, and $\varepsilon_{it}^* = \frac{\varepsilon_{it}}{\sigma_i}$. The transferred model (2.2) indicates that we can only estimate the true parameters up to unknown constants σ_i's. Unless some further restrictions are imposed, we may never be able to identify σ_i's. Estimating some distribution specific parameters together with the interactive fixed effects is certainly an interesting topic, but is out of the scope of this paper.

Simple algebra shows that

$$\begin{aligned} \Pr(y_{it} = 1 \mid x_{it}, \gamma_{0i}, f_{0t}) &= G_\varepsilon(x_{it}'\beta_{0i} + \gamma_{0i}' f_{0t}), \\ \Pr(y_{it} = 0 \mid x_{it}, \gamma_{0i}, f_{0t}) &= 1 - G_\varepsilon(x_{it}'\beta_{0i} + \gamma_{0i}' f_{0t}), \end{aligned} \tag{2.3}$$

which immediately yields

$$E[y_{it} \mid x_{it}, \gamma_{0i}, f_{0t}] = G_\varepsilon(x_{it}'\beta_{0i} + \gamma_{0i}' f_{0t}). \tag{2.4}$$

Thus, the likelihood function is specified as follows.

$$L(B,F,\Gamma)=\prod_{i=1}^{N}\prod_{t=1}^{T}\left[1-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)\right]^{1-y_{it}}\cdot\left[G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)\right]^{y_{it}}, \tag{2.5}$$

where $B=(\beta_1,\ldots,\beta_N)'$, $F=(f_1,\ldots,f_T)'$ and $\Gamma=(\gamma_1,\ldots,\gamma_N)'$ are $N\times d_\beta$, $T\times d_{\max}$ and $N\times d_{\max}$ matrices respectively, and $d_{\max}$ ($\geq d_f$) is a user specified fixed large constant. Moreover, for the purpose of identification, we require

$$F\in\mathbf{F}=\left\{F\,\Big|\,\frac{1}{T}F'F=I_{d_{\max}}\right\}. \tag{2.6}$$

Finally, the log-likelihood function is defined below.

$$\log L(B,F,\Gamma)=\sum_{i=1}^{N}\sum_{t=1}^{T}\Big\{(1-y_{it})\log\left[1-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)\right] \\ +y_{it}\log G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)\Big\}, \tag{2.7}$$

in which $F\in\mathbf{F}$. The estimators are thus given by

$$(\widehat{B},\widehat{F},\widehat{\Gamma})=\underset{(B,F,\Gamma)}{\operatorname{argmax}}\log L(B,F,\Gamma), \tag{2.8}$$

where $\widehat{B}=(\widehat{\beta}_1,\ldots,\widehat{\beta}_N)'$, $\widehat{F}=(\widehat{f}_1,\ldots,\widehat{f}_T)'$ and $\widehat{\Gamma}=(\widehat{\gamma}_1,\ldots,\widehat{\gamma}_N)'$.

2.2 The Asymptotic Results

Before proceeding further, we first introduce some notations, and then provide the necessary assumptions with justifications to facilitate our development.

In what follows, we always let

$$\begin{aligned} z_{it}^0 &= x_{it}'\beta_{0i}+\gamma_{0i}'f_{0t}, \\ e_{it} &= \frac{1-y_{it}}{1-G_\varepsilon(z_{it}^0)}-\frac{y_{it}}{G_\varepsilon(z_{it}^0)} \end{aligned} \tag{2.9}$$

for notational simplicity. It is easy to see that

$$E[e_{it}]=0, \tag{2.10}$$

and one can regard e_{it} as a newly created residual term with mean 0, which is reflected in the theoretical development throughout the paper.

Assumption 1.

1. *There exists a set $\Xi_{NT}=[\Xi_{NT}^l,\Xi_{NT}^u]$ such that all z_{it}^0's belong to Ξ_{NT} with probability approaching to 1, and $0<G_\varepsilon(\Xi_{NT}^l)<G_\varepsilon(\Xi_{NT}^u)<1$.*

2. Let $\frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}E[|E[e_{it}e_{js}\,|\,w_{it}^0,w_{js}^0]|]=O(1)$, where $w_{it}^0=(x_{it},\gamma_{0i},f_{0t})$.

For Assumption 1.1, the range of Ξ_{NT} varies respect to the distribution considered. If a distribution is defined on $\mathbb{R}$ (say a normal distribution), we may allow the lower and upper bounds of Ξ_{NT} to diverge to negative and positive infinity respectively; if an exponential distribution is considered, we may let the lower bound converge to 0 and let the upper bound to diverge; etc. The design of Ξ_{NT} is not new in the literature. For instance, both Chen and Christensen (2015) and Li et al. (2016) use a similar technique to accommodate some unbounded supports of the regressors.

The current form of Assumption 1.2 allows for certain type of weak cross-sectional dependence and time series correlation. Assumption 1.2 essentially imposes a restriction on the second moment of the newly created error term e_{it}, which can be easily verified using independent and identically distributed (i.i.d.) conditions or certain mixing conditions (e.g., Assumption 1 of Dong et al., 2015 and detailed discussions therein).

Lemma 2.1. *Under Assumption 1, as* $(N,T)\to(\infty,\infty)$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(\widehat{z}_{it})-G_\varepsilon(z_{it}^0)]^2=O_P\left(\frac{1}{\sqrt{NT}}\right),$$

where z_{it}^0 *is defined in* (2.9), $\widehat{z}_{it}=x_{it}'\widehat{\beta}_i+\widehat{\gamma}_i'\widehat{f}_t$, *and* $\widehat{\beta}_i$, $\widehat{\gamma}_i$, *and* $\widehat{f}_t$ *are defined in* (2.8).

With very limited restrictions, Lemma 2.1 ensures the overall validity of the approach considered in this paper. It is noteworthy that Lemma 2.1 still holds even for a model without regressors:

$$y_{it}=\begin{cases}1, & \gamma_{0i}'f_{0t}-\varepsilon_{it}\geq 0\\ 0, & \text{otherwise}\end{cases}, \tag{2.11}$$

which is exactly the same as the model investigated in Wang (2019). Certainly, the maximum likelihood function should be adjusted in a very obvious manner, so we omit the details. Although a model like (2.11) is not the focus of this paper, we conjecture that Lemma 2.1 can help simplify the asymptotic development (and maybe assumptions) of Wang (2019) significantly.

In addition, Lemma 2.1 infers that a maximum likelihood estimation can reduce to a nonlinear least squares estimation more or less. Interestingly, if ε_{it} follows a uniform distribution, then the expression presented by Lemma 2.1 completely possesses a form of the least squares approach. As a result, most of the arguments made for the term $\widetilde{S}_{NT}(\beta,F)$ on pages 1264-1265 of Bai (2009) will apply. We refer interested readers to detailed discussions therein.

With Lemma 2.1 in hand, we further impose the next assumption in order to recover the coefficients and the factor structure.

Assumption 2.

1. *Suppose that there exists a sequence* $\{a_{NT}\}$ *satisfying that* $\inf_{w\in\Xi_{NT}} g_\varepsilon(w) \geq a_{NT} > 0$ *and* $a_{NT}^{-2}/\sqrt{NT} \to 0$*, in which* a_{NT} *may converge to 0 or be constant.*

2. *Suppose that the following limits exist given* $F \in \mathbf{F}$ *and* $\|B - B_0\|/\sqrt{N} \leq C$*, where* C *is a sufficiently large positive constant.*

   (a) $\sup_{(B,F)} \left| \frac{1}{NT} \sum_{i=1}^N \{Z_i' M_F Z_i - E[Z_i' M_F Z_i \,|\, F]\} \right| = o_P(1)$*, where* $Z_i = X_i(\beta_{0i} - \beta_i)$.

   (b) $\sup_{(B,F)} \left| \frac{1}{N\sqrt{T}} \sum_{i=1}^N \{\gamma_{0i} \otimes (M_F X_i) - E[\gamma_{0i} \otimes (M_F X_i) \,|\, F]\}(\beta_{0i} - \beta_i) \right| = o_P(1)$.

   (c) $\frac{1}{N}\Gamma_0'\Gamma_0 \to_P \Sigma_\gamma$ *and* $\frac{1}{T}F_0'F_0 \to_P \Sigma_f$.

3. *Suppose that* $0 < \inf_{F\in\mathbf{F}} \Omega(F)$*, where* $\Omega(F) = \mathrm{diag}\{\frac{1}{T}\Omega_{1T}(F), \ldots, \frac{1}{T}\Omega_{NT}(F)\}$*, and*

$$\Omega_{iT}(F) = E[X_i' M_F X_i \,|\, F] - E[\gamma_{0i} \otimes (M_F X_i) \,|\, F]'(\Sigma_\gamma \otimes I_T)^{-1} E[\gamma_{0i} \otimes (M_F X_i) \,|\, F].$$

Assumption 2.1 bounds the PDF from below, which is similar to the treatments of Hansen (2008), Li et al. (2012) among others. The first two conditions of Assumption 2.2 are similar to Assumption G of Li et al. (2020), wherein some detailed justifications are provided. It is noteworthy that they can be removed if one is willing to impose restrictions on the samples directly. See for example Assumption E of Ando and Lu (2020). Assumption 2.3 is the identification restriction accounting for the heterogeneous setting of the model which is in the same spirit of Assumption D of Ando and Bai (2017).

Using Assumption 2, we are able to consistently estimate the coefficients and the factor structure, which is formally stated in the next lemma.

Lemma 2.2. *Under Assumptions 1 and 2, as* $(N, T) \to (\infty, \infty)$,

1. $\frac{1}{N}\|\widehat{B} - B_0\|^2 = o_P(1)$;

2. $\frac{1}{NT}\|\widehat{F}\widehat{\Gamma}' - F_0\Gamma_0'\|^2 = o_P(1)$;

3. $\|P_{\widehat{F}} - P_{F_0}\|^2 = (d_{\max} - d_f) + o_P(1)$.

The first two results of Lemma 2.2 guarantee the consistency of the estimators in (2.8). In the third result, the reason of having the term $d_{\max} - d_f$ is that we have not selected the number of factors correctly. Provided the true number of factors can be detected, this result will reduce to $\|P_{\widehat{F}} - P_{F_0}\| = o_P(1)$, which then infers that the space spanned by F_0 can be recovered consistently. Note that the first two results of Lemma 2.2 still hold, even when we over select the number of factors. This finding for the nonlinear model (2.1) is consistent with some arguments made in Moon and Weidner (2015), wherein a linear model is considered.

That said, we move on to select the number of factors below. Specifically, in connection with Lemma 2.1, we define the next information criterion.

$$\mathrm{IC}(\mathsf{d}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - G_\varepsilon(\widehat{z}_{it}^{\mathsf{d}})\right]^2 + \mathsf{d}\cdot\frac{\xi_{NT}}{\sqrt{NT}}, \tag{2.12}$$

where $\widehat{z}_{it}^{\mathsf{d}}$'s are obtained using (2.8) by setting the number of factors as d, $\xi_{NT} \to \infty$ and $\frac{\xi_{NT}}{\sqrt{NT}} \to 0$. Then d_f is estimated by minimizing (2.12).

$$\widehat{\mathsf{d}} = \operatorname*{argmin}_{0\le \mathsf{d}\le d_{\max}} \mathrm{IC}(\mathsf{d}) \tag{2.13}$$

We summarize the asymptotic property of (2.13) in the first theorem of the paper below.

Theorem 2.1. *Under Assumptions 1 and 2, suppose further that $\xi_{NT} \to \infty$ and $\frac{\xi_{NT}}{\sqrt{NT}} \to 0$. As $(N,T) \to (\infty,\infty)$, $\Pr(\widehat{\mathsf{d}} = d_f) \to 1$.*

In the simulations of Section 3, we examine the finite sample performance of the above selection procedure. Having established Theorem 2.1, in what follows, we assume the number of factors has been correctly selected, and provide the rates of convergence and the necessary asymptotic normalities.

Assumption 3.

1. *Let $\max_{i\ge1,t\ge1} E\|x_{it}\|^{4+\delta} < \infty$, $\max_{i\ge1} E\|\gamma_{0i}\|^{4+\delta} < \infty$, and $\max_{t\ge1} E\|f_{0t}\|^{4+\delta} < \infty$, for some constant $\delta \ge 2$. Moreover, $F_0 \in \mathbf{F}$.*

2. *Suppose that*

   (a) *for $i = 1\ldots,N$, $\frac{1}{T}\sum_{t=1}^{T}\frac{[g_\varepsilon(z_{it}^0)]^2}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0 u_{it}^{0\prime} = \Sigma_{u,i} + O_P\left(\frac{1}{\sqrt{T}}\right)$, where $u_{it}^0 = (x_{it}', f_{0t}')'$ and $\Sigma_{u,i} > 0$;*

   (b) *for $t = 1,\ldots,T$, $\frac{1}{N}\sum_{i=1}^{N}\frac{[g_\varepsilon(z_{it}^0)]^2}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}\gamma_{0i}' = \Sigma_{\gamma,t} + O_P\left(\frac{1}{\sqrt{N}}\right)$ and $\Sigma_{\gamma,t} > 0$.*

3. *Suppose that $E[e_{it}\,|\,w_{it}^0, w_{js}^0] = 0$ and $|E[e_{it}e_{js}\,|\,w_{it}^0, w_{js}^0]| \le a_{ij,ts}$ almost surely, in which $a_{ij,ts}$ satisfies that*

   (a) *$\max_{i\ge1}\frac{1}{T}\sum_{t,s=1}^{T} a_{ii,ts} < \infty$;*

   (b) *$\max_{t\ge1}\frac{1}{N}\sum_{i,j=1}^{N} a_{ij,tt} < \infty$;*

   (c) *$\frac{1}{NT}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} a_{ij,ts} < \infty$.*

Assumption 3.1 imposes moments restrictions on x_{it}, γ_{0i} and f_{0t}, which are standard in the literature. The condition $F_0 \in \mathbf{F}$ is for the purpose of identification only. For example, for given F_0, we can always find a rotation matrix W such that

$$\frac{1}{T}W'F_0'F_0W = I_{d_f}. \tag{2.14}$$

As a consequence, we can write

$$\gamma_{0i}'f_{0t} = \gamma_{0i}'W^{-1} \cdot Wf_{0t}, \tag{2.15}$$

which infers that instead of having γ_{0i} and f_{0t} as true parameters, we in fact use $\gamma_{0i}'W^{-1}$ and Wf_{0t} as true parameters under Assumption 3.1. For linear models (e.g., Bai, 2009 among others), an unidentifiable rotational matrix usually kicks in through the PCA procedure, which gives some detailed expressions in many relevant studies. However, for the nonlinear models with nonclosed form estimators as in this paper, achieving an analytic form of such a rotation matrix seems to be impossible to the best of our knowledge. Similar issues can also be seen in Ando and Bai (2020) and Ando and Lu (2020) for example. Thus, we simply suppose that $F_0 \in \mathbf{F}$. Assumption 3.2 is standard in the literature. In Assumption 3.3, we adopt a mixing type of restriction on the newly created error term e_{it} in order to allow for weak cross-sectional dependence and time series correlation. The conditions can be replaced by restrictions similar to Assumption C of Bai (2009), however, it will create lengthy notations. Without loss of generality, we write Assumption 3.3 as it is.

Using these restrictions, we are ready to state the rates of convergence in the next lemma.

Lemma 2.3. *Under Assumptions 1-3, as* $(N,T) \to (\infty,\infty)$*,*

1. $\frac{1}{N}\|\widehat{\Theta} - \Theta_0\|^2 = O_P\left(\frac{1}{\min\{N,T\}}\right)$*, where* $\widehat{\Theta} = (\widehat{B},\widehat{\Gamma})$ *and* $\Theta_0 = (B_0,\Gamma_0)$*;*
2. $\frac{1}{T}\|\widehat{F} - F_0\|^2 = O_P\left(\frac{1}{\min\{N,T\}}\right)$*.*

Lemma 2.3 presents the rates of convergence of the estimators in (2.8), and also helps establish the asymptotic distributions below.

To close our theoretical investigation, we investigate the first order condition associated with the following presentations in order to establish asymptotic distributions.

$$\begin{aligned}
\frac{\partial \log L(\Theta,F)}{\partial \theta_i} &= \sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it})]g_\varepsilon(z_{it})}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}u_{it}, \\
\frac{\partial \log L(\Theta,F)}{\partial f_t} &= \sum_{i=1}^{N} \frac{[y_{it} - G_\varepsilon(z_{it})]g_\varepsilon(z_{it})}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}\gamma_i,
\end{aligned} \tag{2.16}$$

where $\Theta = (\theta_1,\ldots,\theta_N)'$, $\theta_i = (\beta_i',\gamma_i')'$ and $u_{it} = (x_{it}',f_t')'$. Using the above equations, we further introduce the following conditions on e_{it} to facilitate the development of the asymptotic distributions.

Assumption 4.

1. *Suppose that* $|E[e_{it}e_{js}e_{nl}e_{mk} \mid w_{it}^0, w_{js}^0, w_{nl}^0, w_{mk}^0]| \leq a_{ijnm,tslk}$ *almost surely, in which* $a_{ijnm,tslk}$ *satisfies that*

   (a) $\max_{i\geq 1} \frac{1}{T\max\{N,T\}} \sum_{j,n=1}^{N} \sum_{t,s=1}^{T} a_{ijin,ttss} < \infty$*;*

   (b) $\max_{t\geq 1} \frac{1}{N\max\{N,T\}} \sum_{i,j=1}^{N} \sum_{s,l=1}^{T} a_{ijij,ttsl} < \infty$*.*

Assumption 4 imposes additional bounds on the weak cross sectional dependence and time series correlation of among e_{it}'s, and is for the purpose of notational simplicity as well. The arguments made to Assumption 3.3 apply. Given the extra conditions, we are able to establish the asymptotic distributions in the next theorem.

Theorem 2.2. *Let Assumptions 1-4 hold. As* $(N,T) \to (\infty,\infty)$*,*

1. *if* $T/N^2 \to 0$ *and* $\frac{1}{\sqrt{T}} \frac{\partial \log L(B_0,F_0,\Gamma_0)}{\partial \theta_i} \to_D N(0,\Sigma_{\theta,i})$ *for each* $i = 1,\ldots,N$*, then* $\sqrt{T}(\widehat{\theta}_i - \theta_{0i}) \to_D N(0, \Sigma_{u,i}^{-1}\Sigma_{\theta,i}\Sigma_{u,i}^{-1})$*, where* $\widehat{\theta}$ *and* θ_{0i} *are defined in Lemma 2.2;*

2. *if* $N/T^2 \to 0$ *and* $\frac{1}{\sqrt{N}} \frac{\partial \log L(B_0,F_0,\Gamma_0)}{\partial f_t} \to_D N(0,\Sigma_{f,t})$ *for each* $t = 1,\ldots,T$*, then* $\sqrt{N}(\widehat{f}_t - f_{0t}) \to_D N(0, \Sigma_{\gamma,t}^{-1}\Sigma_{f,t}\Sigma_{\gamma,t}^{-1})$*.*

Theorem 2.2 infers that we can achieve the asymptotic normality for each component of the model (2.1) at a usual parametric rate. The conditions $T/N^2 \to 0$ and $N/T^2 \to 0$ in the body of the theorem are the same as those in Theorem 1 of Bai and Ng (2013), so one may regard the above theorem as the binary response counterpart of Theorem 1 of Bai and Ng (2013).

2.3 Numerical Implementation

Before proceeding to simulation, we provide an algorithm for numerical implementation in practice. A variety of algorithms have been proposed for the panel data models with interactive fixed effects, especially for the cases where the nonclosed form estimators are involved, e.g., Chen (2014), Ando and Bai (2017), Wang (2019), Chen, Fernández-Val and Weidner (2020), Ando and Lu (2020), just to name a few. Our numerical implementation largely follows from Chen, Fernández-Val and Weidner (2020) and Ando and Lu (2020) in particular. Specifically, for each given value of the number of factors, we implement the following procedure.

Step 0: Initial $\widehat{F}^{(0)}$ by using some random number generators, e.g., the standard norm distribution for each element of $\widehat{F}^{(0)}$. Implement the SVD decomposition on $\widehat{F}^{(0)}$, and update $\widehat{F}^{(0)}$ to ensure $\frac{1}{T}\widehat{F}^{(0)\prime}\widehat{F}^{(0)} = I$.

Step j: For Step j (≥ 1), obtain $\widehat{B}^{(j)} = (\widehat{\beta}_1^{(j)}, \ldots, \widehat{\beta}_N^{(j)})'$ and $\widehat{\Gamma}^{(j)} = (\widehat{\gamma}_1^{(j)}, \ldots, \widehat{\gamma}_N^{(j)})'$ by minimizing

$$\log L^{(j)}(\beta_i, \gamma_i) = \sum_{t=1}^{T} \Big\{(1 - y_{it}) \log \Big[1 - G_\varepsilon(x_{it}'\beta_i + \gamma_i'\widehat{f}_t^{(j-1)})\Big] \\ + y_{it} \log G_\varepsilon(x_{it}'\beta_i + \gamma_i'\widehat{f}_t^{(j-1)})\Big\}$$

over all $i \geq 1$, where $\widehat{F}^{(j-1)} = (\widehat{f}_1^{(j-1)}, \ldots, \widehat{f}_T^{(j-1)})'$ is obtained from the Step $j-1$. Then obtain $\widehat{F}^{(j)} = (\widehat{f}_1^{(j)}, \ldots, \widehat{f}_N^{(j)})'$ by minimizing

$$\log L^{(j)}(f_t) = \sum_{i=1}^{N} \Big\{(1 - y_{it}) \log \Big[1 - G_\varepsilon(x_{it}'\widehat{\beta}_i^{(j)} + \widehat{\gamma}_i^{(j)\prime} f_t)\Big] \\ + y_{it} \log G_\varepsilon(x_{it}'\widehat{\beta}_i^{(j)} + \widehat{\gamma}_i^{(j)\prime} f_t)\Big\}$$

over all $t \geq 1$. Finally, implement the SVD decomposition on $\widehat{F}^{(j)}$, and update $\widehat{F}^{(j)}$ to ensure $\frac{1}{T}\widehat{F}^{(j)\prime}\widehat{F}^{(j)} = I$.

Stop: Stop after reaching certain criterion, say, $\frac{1}{N}\|\widehat{B}^{(j)} - \widehat{B}^{(j-1)}\| \leq \epsilon$ in which ϵ is a sufficiently small number.

Note that for each given value of the number of factors, we can implement the above procedure, which then allows us to calculate the information criterion (2.12) to select the "optimal" number of factors.

Finally, we point out that there are a few studies aiming to justify the aforementioned algorithms theoretically. The early work probably dates back to the PhD thesis mentioned in Chen (2014), and recently, Liu et al. (2019) and Jiang et al. (2020) further provide theoretical justifications under different settings. We no longer pursue any theoretical results along this line of research, as it may lead to a different research paper.

3 Simulation

Below, we conduct simulations to examine the theoretical findings of Section 2. We specifically consider the following data generating process.

$$y_{it} = \begin{cases} 1, & x_{it}'\beta_{0i} + \gamma_{0i}'f_{0t} - \varepsilon_{it} \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

The factors and loadings are generated by $f_{0t,j} \sim U(-2.5, 2.5)$ and $\gamma_{0i,j} \sim U(0, 6)$, where $\gamma_{0i,j}$ and $f_{0t,j}$ stand for the j^{th} elements of γ_{0i} and f_{0t} respectively and $j = 1, \ldots, d_f$. In order to introduce correlation between the regressors and the factor structure, we let $x_{it,j} = N(0,1) + 0.5(|\gamma_{0i,1}| + |f_{0t,1}|)$, where $x_{it,j}$ stands for the j^{th} element of x_{it}, and $j = 1, \dots, d_\beta$. For the coefficients, let $\beta_{0i,j} = i/N$, where $i = 1, \dots, N$, $j = 1, \dots, d_\beta$, and $\beta_{0i,j}$ stands for the j^{th} element of β_{0i}. We consider two distributions for the error term:

Case 1 – a light tailed distribution: $\varepsilon_{it} \sim N(0,1)$, where $N(0,1)$ is the standard normal distribution;

Case 2 – a heavy tailed distribution: $\varepsilon_{it} \sim \text{Logistic}(0,1)$, where 0 and 1 are the values of location and scale parameters respectively. Thus, the variance of ε_{it} is $\pi^2/3$, which is much flatter than Case 1.

For each generated dataset, we first select the number of factors using the information criterion defined in (2.12), and then conduct the estimation. We repeat the above procedure M times, and report the following values to evaluate the finite sample performance:

$$P_c = \frac{1}{M}\sum_{j=1}^{M} I(\widehat{\mathsf{d}}_j = d_f), \quad P_u = \frac{1}{M}\sum_{j=1}^{M} I(\widehat{\mathsf{d}}_j < d_f), \quad P_o = \frac{1}{M}\sum_{j=1}^{M} I(\widehat{\mathsf{d}}_j > d_f),$$

$$\text{RMSE}_{B_0} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{N}\|\widehat{B}_j - B_0\|^2}, \quad \text{RMSE}_{F_0} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\|P_{\widehat{F}_j} - P_{F_0}\|^2}, \tag{3.1}$$

where $\widehat{\mathsf{d}}_j$, $\widehat{B}_j$ and $\widehat{F}_j$ stand for the estimated number of factors, the estimated value of $B_0 = (\beta_{01}, \dots, \beta_{0N})'$, and the estimated value of $F_0 = (F_{01}, \dots, F_{0T})'$ at the j^{th} iteration respectively. It is obvious that P_c, P_u and P_o measure the probabilities of correctly, under and over selecting the number of factors respectively. As shown in Section 2, $\widehat{F}_j$ yields a consistent estimation of F_0 up to a rotation matrix, so we measure the distance between $P_{\widehat{F}_j}$ and P_{F_0} in (3.1).

We let, $d_\beta = 2$, $d_f = 2$, $N, T \in \{50, 100, 150\}$, and $M = 500$.[1] Moreover, let ξ_{NT} of (2.12) be $\log\sqrt{N+T}$ throughout the numerical studies without loss of generality. The results are summarized in Table 1 and Table 2 below. First, we point out that the newly proposed methodology works well regardless the tail behaviour of the error term ε_{it}, because the values associated with Case 1 and Case 2 are roughly same across both tables. Second, in Table 1, the values of P_c go up to 1, as the sample sizes increase. It is noteworthy that when the sample sizes are relatively small, the information criterion tends to under select the number of factors, which has a clear impact on the values of RMSE_{F_0} presented in Table 2. Third, in Table 2, the values of RMSE_{B_0} and RMSE_{F_0} decrease as the sample sizes go up, which is expected.

[1] Due to the restrictions of computational power, we no longer explore larger values of M. In practice, heavier tails require longer time to compute, which should be expected.

Table 1: The percentages of correctly, under and over identifying the number of factors. Specifically, P_c, P_u and P_o measure the probabilities of correctly, under and over selecting the number of factors respectively.

		P_c			P_u			P_o		
	$N \setminus T$	50	100	150	50	100	150	50	100	150
Case 1	50	0.246	0.742	0.914	0.754	0.258	0.086	0.000	0.000	0.000
	100	0.622	0.954	1.000	0.378	0.046	0.000	0.000	0.000	0.000
	150	0.806	0.980	0.992	0.194	0.012	0.000	0.000	0.008	0.008
Case 2	50	0.236	0.748	0.934	0.764	0.252	0.066	0.000	0.000	0.000
	100	0.696	0.992	1.000	0.304	0.008	0.000	0.000	0.000	0.000
	150	0.804	1.000	1.000	0.196	0.000	0.000	0.000	0.000	0.000

Table 2: The root mean square errors associated with the estimates of B_0 and F_0. RMSE_{B_0} and RMSE_{F_0} are defined in (3.1).

		RMSE_{B_0}			RMSE_{F_0}		
	$N \setminus T$	50	100	150	50	100	150
Case 1	50	0.646	0.575	0.533	1.021	0.848	0.786
	100	0.639	0.560	0.517	0.881	0.725	0.684
	150	0.637	0.557	0.516	0.809	0.677	0.656
Case 2	50	0.648	0.549	0.495	1.028	0.885	0.823
	100	0.645	0.533	0.469	0.855	0.696	0.670
	150	0.640	0.527	0.456	0.770	0.616	0.598

4 Case Study

In this section, we apply the model and methodology to portfolio analysis using daily returns data of S&P 500 stocks. In both economics and finance disciplines, a vast amount of efforts have been devoted to portfolio analysis and management. Among them, a fundamental one is to select the optimal weights associated with each stock when constructing a portfolio. Mathematically, it is realized by the next minimisation problem.

$$\min_{w} \; w'\Sigma_p w \quad \text{subject to} \quad w'1_N = 1, \tag{4.1}$$

where 1_N is a $N \times 1$ vector of ones, Σ_p is a positive definite matrix and is usually decided by the data of stock returns, and w includes the weights assigned to each stock. The analytic solution to the above minimisation problem is

$$w = \frac{\Sigma_p^{-1} 1_N}{1_N' \Sigma_p^{-1} 1_N}. \tag{4.2}$$

More often than not, one adopts all stocks to construct Σ_p (e.g., Chen et al., 2019; Engle et al., 2019), which naturally falls into a category of high dimensional matrix estimation.

Thereby, to boost the estimation accuracy, we see the increasing popularity of methods using rank reduction (Pelger and Xiong, 2019), penalization (Chen et al., 2019), or both (Fan et al., 2013) among others. By default, the aforementioned techniques account for all stocks in practice, though some of the stocks may have relatively small weights compared to the others. As pointed out in Christoffersen and Diebold (2006) and Nyberg (2011), the sign of stock market returns may be predictable even if the returns themselves are not predictable. Therefore, an intuitive thought for portfolio analysis is to remove the stocks presenting high probabilities of having negative returns, and then construct the portfolio using those left. It is also in line with the motivation of big data econometrics (Varian, 2014), which is to implement certain type of "selection" in order to filter out the irrelevant information. Motivated by this line of research, we use the proposed framework to predict the sign of the return for each stock, and then use those having high probabilities of being positive to construct the portfolio.

4.1 Data

The stock prices data are collected from `https://www.kaggle.com` over the time period between 2 January 2008 and 31 December 2018. After removing the companies which have missing stock returns during the whole time period, we end up with 319 stocks ($N = 319$). We adopt the log-normalised CBOE volatility index (VIX) (`http://www.cboe.com`) as the only regressor, which is widely viewed as a good indicator of market sentiment (e.g., Christoffersen and Diebold, 2006; Pelger and Xiong, 2019). Furthermore, we collect risk-free interest (RFI) data from the U.S. Department of the Treasury (`https://www.treasury.gov`) to construct the Sharpe ratio later on in order to evaluate the performance of the proposed method.

4.2 Empirical Analysis

Below, we conduct a rolling-window analysis, and focus on the out-of-sample forecast. For each window, we estimate Σ_p with the sample information on the most recent 505 trading days (roughly two years) using the next model.

$$y_{i,t+1} = \begin{cases} 1, & x_{it}\beta_{0i} + \gamma_{0i}' f_{0t} - \varepsilon_{it} \geq 0 \\ 0, & \text{otherwise} \end{cases},$$

where 1 and 0 stand for positive and non-positive returns respectively. We always count the first available trading day of each rolling-window as 0, and use $t = 0, \ldots, 503$ to implement estimation. x_{it} includes the value of VIX at day t, and $y_{i,t+1}$ is the sign of stock i at day $t+1$. For each estimation, we first select the number of factors, and then estimate the coefficient of VIX together with the interactive fixed effects. Afterwards, we bring the value of VIX at

$t = 504$ in the estimated model to forecast the probabilities of the stocks having positive returns at $t = 505$. We keep the stocks with estimated probabilities of having positive returns greater than or equal to 0.5 to construct Σ_p of (4.1), and then calculate the weight vector w using (4.2). If all estimated probabilities are less than 0.5, then record $w = 0$ (i.e., no transaction made for the day). Finally, we calculate the weighted average of returns for $t = 505$. We repeat the above forecasting process from the first available window till the end, and consider the Probit model for the error term. The results of using other distributions (e.g., t-distribution and Logit model) for the error terms are quite similar, so we focus on the results of Probit model below. Note that the above procedure admits short selling (i.e., elements of w can be negative), and also assumes that there is no transaction cost, which is consistent with many existing works of the literature (e.g., Chen et al., 2019 and references therein). As a comparison, we also consider a model with fixed effects only.

In what follows, we report (1). the average of weighted returns across the entire available period (Mean), which is annualized by multiplying 252; (2). the standard deviation of weighted returns across the entire available period (Std), which annualized by multiplying $\sqrt{252}$; (3). the information ratio defined as the ratio of Mean to Std (IR); and (4) the Sharp ratio defined as the mean of returns minus RFI normalised by Std (SR). The annualized Mean and Std are consistent with those defined in Section 6.2 of Engle et al. (2019). We refer interested readers to their paper for more relevant discussions. Moreover, since a strand of literature on portfolio analysis is interested in estimation from a factor model with a predetermined number of factors (Pelger and Xiong, 2019), we report the results for the cases where the number of factors are fixed as $1, \ldots, 5$ respectively. In addition, we report the results of the case where IC of (2.12) is used to select the optimal number of factors for each estimation. The results are summarized in Table 3.

Table 3: The results of the out-of-sample forecasts. "FE" and "IFE" refer to the results associated with the models with fixed effects and the models with interactive fixed effects respectively. For IFE models, "Optimal" refers to the results when the number of factors is selected using the information criterion (2.12) in each estimation.

	Factor No.	Mean (%)	Std	IR	SR
IFE	1	14.92	11.56	1.29	1.26
	2	15.98	13.82	1.16	1.13
	3	13.38	13.40	1.00	0.97
	4	13.61	12.82	1.06	1.03
	5	13.84	12.61	1.10	1.07
	Optimal	16.11	12.33	1.31	1.28
FE		8.56	10.27	0.83	0.80

We note that, under the newly proposed framework, one does not need to make transactions in every single trading day, which is the main difference between our work and the existing literature. In general, we are looking for a strategy, which generates a small value of Std, but large values of Mean, IR and SR. The FE model yields the smallest Std, which in a sense should be expected due to simplicity of the model. It also produces the minimum values of Mean, IR, and SR respectively. Overall, the IFE models outperform the FE model regarding the criteria Mean, IR and SR. Among them, the strategy that updates the number of factors in each estimation (i.e., "Optimal") has the best performance.

5 Conclusion

In this paper, we investigate binary response models for heterogeneous panel data with interactive fixed effects by allowing both the cross sectional dimension and the time dimension to diverge. From the modelling perspective, our setting is similar to Boneva and Linton (2017), but we do not require a specific structure on the regressors, which allows us to avoid putting any restriction between the number of regressors and the number of unobservable factors. Our investigation establishes a link between a maximum likelihood estimation and a least squares approach. As a consequence, the identification restrictions provided in Bai (2009) and Moon and Weidner (2015) are readily to be applied to binary response models with very minor modifications. Our setting also improves those considered in Wang (2019) and Chen, Fernández-Val and Weidner (2020) by covering a broader family of distributions (See Appendix A for details), and we further establish asymptotic distributions for the unobservable factors and their loadings, which can be considered as the binary response counterpart of those established in Bai and Ng (2013). In addition, we provide a simple information criterion to detect the number of factors for binary response panel data models. Last but not least, we conduct intensive numerical studies to examine the finite sample performance of the newly proposed model and methodology, and demonstrate the practical relevance.

From the practical perspective, the framework can be applied to predict the probability of corporate failure (Caggiano et al., 2014), conduct credit rating analysis (Jones et al., 2015), etc. Following Christoffersen and Diebold (2006) and Nyberg (2011), in the empirical study, we focus on the sign prediction of stock returns, and then use the results of sign forecast to to conduct portfolio analysis. By implementing rolling-window out of sample forecasts, we demonstrate the practical relevance of the paper.

References

Ahn, S. C. and Horenstein, A. R. (2013), 'Eigenvalue ratio test for the number of factors', *Econometrica* **81**(3), 1203–1227.

Altman, E. I. (1968), 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *Journal of Finance* **23**(4), 589–609.

Ando, T. and Bai, J. (2017), 'Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures', *Journal of the American Statistical Association* **112**(519), 1182–1198.

Ando, T. and Bai, J. (2020), 'Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity', *Journal of the American Statistical Association* **115**(529), 266–279.

Ando, T. and Lu, L. (2020), A spatial panel quantile model with unobserved heterogeneity. Working paper available at "https://ssrn.com/abstract=3516306".

Bai, J. (2009), 'Panel data models with interactive fixed effects', *Econometrica* **77**(4), 1229–1279.

Bai, J. and Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Bai, J. and Ng, S. (2013), 'Principal components estimation and identification of static factors', *Journal of Econometrics* **176**(1), 18–99.

Boneva, L. and Linton, O. (2017), 'A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance', *Journal of Applied Econometrics* **32**(7), 1226–1243.

Caggiano, G., Calice, P. and Leonida, L. (2014), 'Early warning systems and systemic banking crises in low income countries: A multinomial logit approach', *Journal of Banking & Finance* **47**, 258 – 269.

Chamberlain, G. (1984), Panel data, *in* Z. Griliches and M. D. Intriligator, eds, 'Handbook of Econometrics', 1 edn, Vol. 2, Elsevier, chapter 22, pp. 1247–1318.

Chamberlain, G. (2010), 'Binary response models for panel data: Identification and information', *Econometrica* **78**(1), 159–168.

Chen, J., Li, D. and Linton, O. (2019), 'A new semiparametric estimation approach for large dynamic covariance matrices with multiple conditioning variables', *Journal of Econometrics* **212**(1), 155 – 176.

Chen, L., Dolado, J. J. and Gonzalo, J. (2020), 'Quantile factor models', *Econometrica* p. forthcoming.

Chen, M. (2014), Estimation of nonlinear panel models with multiple unobserved effects. Warwick Economics Research Paper Series No. 1120.

Chen, M., Fernández-Val, I. and Weidner, M. (2020), 'Nonlinear factor models for network and panel data', *Journal of Econometrics* p. forthcoming.

Chen, X. and Christensen, T. (2015), ‘Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions’, *Journal of Econometrics* **188**(2), 447–465.

Christoffersen, P. F. and Diebold, F. X. (2006), ‘Financial asset returns, direction-of-change forecasting, and volatility dynamics’, *Management Science* **52**(8), 1273–1287.

Dong, C., Gao, J. and Peng, B. (2015), ‘Semiparametric single-index panel data model with cross-sectional dependence’, *Journal of Econometrics* **188**(1), 301–312.

Engle, R. F., Ledoit, O. and Wolf, M. (2019), ‘Large dynamic covariance matrices’, *Journal of Business & Economic Statistics* **37**(2), 363–375.

Fan, J., Liao, Y. and Mincheva, M. (2013), ‘Large covariance estimation by thresholding principal orthogonal complements’, *Journal of the Royal Statistical Society: Series B* **75**(4), 603 – 680.

Fernández-Val, I. and Weidner, M. (2018), ‘Fixed effects estimation of large-t panel data models’, *Annual Review of Economics* **10**(1), 109–138.

Hansen, B. E. (2008), ‘Uniform convergence rates for kernel estimation with dependent data’, *Econometric Theory* **24**(3), 726–748.

Huang, W., Jin, S., Phillips, P. C. B. and Su, L. (2020), ‘Nonstationary panel models with latent group structures and cross-section dependence’, *Journal of Econometrics* p. forthcoming.

Jiang, B., Yang, Y., Gao, J. and Hsiao, C. (2020), ‘Recursive estimation in large panel data models: Theory and practice’, *Journal of Econometrics* p. forthcoming.

Jones, S., Johnstone, D. and Wilson, R. (2015), ‘An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes’, *Journal of Banking & Finance* **56**, 72 – 85.

Lam, C. and Yao, Q. (2012), ‘Factor modeling for high-dimensional time series: Inference for the number of factors’, *Annals of Statistics* **40**(2), 694–726.

Li, D., Lu, Z. and Linton, O. (2012), ‘Local linear fitting under near epoch dependence: Uniform consistency with convergence rates’, *Econometric Theory* **28**(5), 935–958.

Li, D., Tjøstheim, D. and Gao, J. (2016), ‘Estimation in nonlinear regression with Harris recurrent Markov chains’, *Annals of Statistics* **44**(5), 1957–1987.

Li, K., Cui, G. and Lu, L. (2020), ‘Efficient estimation of heterogeneous coefficients in panel data models with common shocks’, *Journal of Econometrics* **216**(2), 327 – 353.

Liu, F., Gao, J. and Yang, Y. (2019), Nonparametric estimation in panel data models with heterogeneity and time-varyingness. https://ssrn.com/abstract=3214046.

Manski, C. F. (1987), ‘Semiparametric analysis of random effects linear models from binary panel data’, *Econometrica* **55**(2), 357–362.

Moon, H. R. and Weidner, M. (2015), ‘Linear regression for panel with unknown number of factors as interactive fixed effects’, *Econometrica* **83**(4), 1543–1579.

Nyberg, H. (2011), 'Forecasting the direction of the us stock market with dynamic binary probit models', *International Journal of Forecasting* **27**(2), 561 – 578.

Pelger, M. and Xiong, R. (2019), State-varying factor models of large dimensions. arXiv:1807.02248.

Pesaran, M. H. (2006), 'Estimation and inference in large heterogeneous panels with a multifactor error structure', *Econometrica* **74**(4), 967–1012.

Varian, H. R. (2014), 'Big data: New tricks for econometrics', *Journal of Economic Perspectives* **28**(2), 3–28.

Wang, F. (2019), Maximum likelihood estimation and inference for high dimensional nonlinear factor models with application to factor-augmented regressions. Working paper available at "https://ssrn.com/abstract=3376794".

Appendix A

In this Appendix, we explain how our settings improve those of of Wang (2019) and Chen, Fernández-Val and Weidner (2020). The model (2.1) considered above is closely related to Wang (2019) and Chen, Fernández-Val and Weidner (2020). Both papers are established on the next fundamental assumption

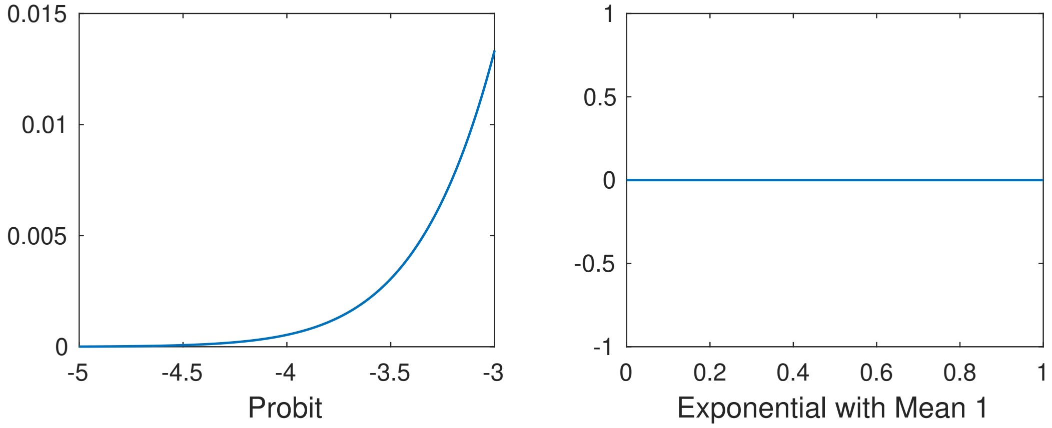
$$0 < b_L \leq -\frac{\partial^2 l_{it}(w)}{\partial w^2} \leq b_U < \infty \tag{A.1}$$

at each (i,t), where $l_{it}(w) = (1-y_{it})\log[1-G_\varepsilon(w)] + y_{it}\log G_\varepsilon(w)$ under the context of this paper. Simple algebra shows that

$$\begin{aligned}\frac{\partial^2 l_{it}(w)}{\partial w^2} &= -(1-y_{it})\frac{g_\varepsilon^{(1)}(w)(1-G_\varepsilon(w)) + [g_\varepsilon(w)]^2}{(1-G_\varepsilon(w))^2} \\ &\quad + y_{it}\frac{g_\varepsilon^{(1)}(w)G_\varepsilon(w) - [g_\varepsilon(w)]^2}{[G_\varepsilon(w)]^2}.\end{aligned} \tag{A.2}$$

For simplicity, we let $y_{it} = 0$, and consider the values of $-\frac{\partial^2 l_{it}(w)}{\partial w^2}$ when the error term takes the standard normal distribution (Probit) and exponential distribution with mean 1 respectively. Then we can generate the next figure.

Figure 1: Plots of $-\frac{\partial^2 l_{it}(w)}{\partial w^2}$ when $y_{it} = 0$.



As shown in Figure 1, neither of the sub-figures satisfies (A.1), so it motivates us to present those settings and results in Section 2.

Appendix B

In the appendix, we provide the proofs for the asymptotic results of the main text. $O(1)$ always stands for a constant, and may be different at each appearance. Recall in the main text, we have denoted the following notations $\theta_i = (\beta_i', \gamma_i')'$, $\theta_{0i} = (\beta_{0i}', \gamma_{0i})$, $\widehat{\theta}_i = (\widehat{\beta}_i', \widehat{\gamma}_i')'$, $\Theta = (\theta_1, \cdots, \theta_N)'$, $\Theta_0 = (\theta_{01}, \cdots, \theta_{0N})'$. $u_{it} = (x_{it}', f_t')'$, $u_{it}^0 = (x_{it}', f_{0t}')'$ and $\widehat{u}_{it} = (x_{it}', \widehat{f}_t')'$, which will be repeated used throughout the development below.

Proof of Lemma 2.1:

(1). For notational simplicity, let $z_{it} = x_{it}'\beta_i + \gamma_i' f_t$ and $\Delta z_{it} = z_{it} - z_{it}^0$, where z_{it}^0 is defined in the beginning of Section 2.2. Note that under Assumption 1, we need to investigate the minimization of (2.8) under the constraint that z_{it}'s $\in \Xi_{NT}$. Moreover, note that provided $0 < x, x_0 < 1$, we have the following two expressions by the Taylor expansion:

$$\log x = \log x_0 + (x - x_0)\frac{1}{x_0} - (x - x_0)^2 \frac{1}{2(x^*)^2}, \tag{B.1}$$

$$\log(1 - x) = \log(1 - x_0) - (x - x_0)\frac{1}{1 - x_0} - (x - x_0)^2 \frac{1}{2(1 - x^\dagger)^2}, \tag{B.2}$$

where both x^* and $x^\dagger$ lie between x and x_0.

We are now ready to start our investigation. By (B.1) and (B.2), write

$$\begin{aligned} &\log L(\beta_0, F_0, \Gamma_0) - \log L(\beta, F, \Gamma) \\ = &-\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T (1 - y_{it}) \left\{\log\left[1 - G_\varepsilon(z_{it})\right] - \log\left[1 - G_\varepsilon(z_{it}^0)\right]\right\} \\ &-\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T y_{it} \left\{\log G_\varepsilon(z_{it}) - \log G_\varepsilon(z_{it}^0)\right\} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)] \cdot \frac{1-y_{it}}{1-G_\varepsilon(z_{it}^0)} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 \cdot \frac{1-y_{it}}{2(1-G_{it}^\dagger)^2} \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)] \cdot \frac{y_{it}}{G_\varepsilon(z_{it}^0)} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 \cdot \frac{y_{it}}{2(G_{it}^*)^2} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)] \cdot \left[\frac{1-y_{it}}{1-G_\varepsilon(z_{it}^0)} - \frac{y_{it}}{G_\varepsilon(z_{it}^0)}\right] \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 \cdot \left[\frac{1-y_{it}}{2(1-G_{it}^\dagger)^2} + \frac{y_{it}}{2(G_{it}^*)^2}\right] \\
&:= \mathbb{L}_{1NT} + \mathbb{L}_{2NT}, \qquad \text{(B.3)}
\end{aligned}
$$

where both G_{it}^* and $G_{it}^\dagger$ lie between $G_\varepsilon(z_{it})$ and $G_\varepsilon(z_{it}^0)$, and the definitions of $\mathbb{L}_{1NT}$ and $\mathbb{L}_{2NT}$ are obvious.

We then consider $\mathbb{L}_{1NT}$ and $\mathbb{L}_{2NT}$ respectively, and start with $\mathbb{L}_{1NT}$. Recall that we have defined e_{it} right above Assumption 2. Then consider

$$
\begin{aligned}
& E\left[\max_{z_{it}\text{'s}}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_\varepsilon(z_{it})e_{it}\right|\right]^2 \\
\le\ & E\left[\max_{z_{it}\text{'s},\ z_{js}\text{'s}}\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}G_\varepsilon(z_{it})G_\varepsilon(z_{js})e_{it}e_{js}\right] \\
\le\ & E\left[\max_{z_{it}\text{'s},\ z_{js}\text{'s}}\frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}G_\varepsilon(z_{it})G_\varepsilon(z_{js}) \cdot |E[e_{it}e_{js} \mid z_{it}^0, z_{js}^0]|\right] \\
\le\ & \frac{1}{N^2T^2}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}E[|E[e_{it}e_{js} \mid z_{it}^0, z_{js}^0]|] = O\left(\frac{1}{NT}\right),
\end{aligned}
$$

in which the third inequality follows from the fact that $G_\varepsilon(\cdot) \le 1$ uniformly, and the last equality follows from Assumption 1. Thus, it is easy to know that

$$
|\mathbb{L}_{1NT}| = O_P\left(\frac{1}{\sqrt{NT}}\right). \qquad \text{(B.4)}
$$

We next investigate $\mathbb{L}_{2NT}$. Write

$$
\begin{aligned}
\mathbb{L}_2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 \cdot \left[\frac{1-y_{it}}{2(1-G_{it}^\dagger)^2} + \frac{y_{it}}{2(G_{it}^*)^2}\right] \\
&\ge \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 \cdot \left\{\frac{1-y_{it}}{4[1+(G_{it}^\dagger)^2]} + \frac{y_{it}}{2(G_{it}^*)^2}\right\}
\end{aligned}
$$

$$\geq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 \cdot \left\{\frac{1-y_{it}}{4\cdot 2} + \frac{y_{it}}{2}\right\}$$
$$\geq \frac{1}{8}\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2,$$

where the first inequality follows from $\frac{1}{(a+b)^2} \geq \frac{1}{2a^2+2b^2}$ because of $(a+b)^2 \leq 2a^2+2b^2$, the second inequality follows from the fact that G_{it}^* and $G_{it}^\dagger$ lie between $G_\varepsilon(z_{it})$ and $G_\varepsilon(z_{it}^0)$, and the third inequality follows from that $\frac{1-y_{it}}{4\cdot 2} + \frac{y_{it}}{2} \geq \frac{1}{8}$ because of y_{it} taking the value of 1 or 0 only.

By the fact that $0 \geq \log L(B_0, F_0, \Gamma_0) - \log L(\widehat{B}, \widehat{F}, \widehat{\Gamma})$, and (B.3) and (B.4), we now can conclude that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]^2 = O_P\left(\frac{1}{\sqrt{NT}}\right),$$

which completes the proof for the first result of this lemma. ■

Proof of Lemma 2.2:

(1). Again, let $z_{it} = x_{it}'\beta_i + \gamma_i' f_t$. Then we consider that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(z_{it}) - G_\varepsilon(z_{it}^0)]^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[g_\varepsilon(z_{it}^\dagger)\Delta z_{it}]^2$$
$$= \frac{1}{NT}\sum_{i=1}^{N}[\mathscr{G}_i X_i(\beta_i - \beta_{0i}) + \mathscr{G}_i(F\gamma_i - F_0\gamma_{0i})]'[\mathscr{G}_i X_i(\beta_i - \beta_{0i}) + \mathscr{G}_i(F\gamma_i - F_0\gamma_{0i})]$$
$$\geq \frac{a_{NT}^2}{NT}\sum_{i=1}^{N}[X_i(\beta_i - \beta_{0i}) + (F\gamma_i - F_0\gamma_{0i})]'[X_i(\beta_i - \beta_{0i}) + (F\gamma_i - F_0\gamma_{0i})]$$
$$\geq \frac{a_{NT}^2}{NT}\sum_{i=1}^{N}[X_i(\beta_{0i} - \beta_i) + F_0\gamma_{0i}]' M_F [X_i(\beta_{0i} - \beta_i) + F_0\gamma_{0i}]$$
$$= \frac{a_{NT}^2}{NT}\sum_{i=1}^{N}\left[(\beta_{0i} - \beta_i)' A_i(\beta_{0i} - \beta_i) + \eta' B_i \eta + (\beta_{0i} - \beta_i)' C_i' \eta\right] + o_P(a_{NT}^2) \tag{B.5}$$

where $\mathscr{G}_i = \text{diag}\left\{g_\varepsilon(z_{i1}^\dagger), \ldots, g_\varepsilon(z_{iT}^\dagger)\right\}$ with $z_{it}^\dagger$ lying between z_{it} and z_{it}^0 for each (i,t), $A_i = E[X_i' M_F X_i \,|\, F]$, $B_i = E[\gamma_{0i}\gamma_{0i}'] \otimes I_T$, $C_i = E[\gamma_{0i} \otimes (M_F X_i) \,|\, F]$, $\eta = \text{vec}(M_F F_0)$, and the first inequality follows from Assumption 2.1.

Note that expect the extra term a_{NT}^2, the right hand side of (B.5) has the identical form as in Bai (2009, pp. 1265-1266) for each i. In connection Lemma 2.1 and Assumption 2, we are readily to conclude that $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\beta}_i - \beta_{0i}\|^2 = o_P(1)$.

(2). After establishing the second result, we rewrite (B.5) as follows.

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]^2$$
$$\geq \frac{a_{NT}^2}{NT}\sum_{i=1}^{N}\left[X_i(\widehat{\beta}_i - \beta_{0i}) + (\widehat{F}\widehat{\gamma}_i - F_0\gamma_{0i})\right]'\left[X_i(\widehat{\beta}_i - \beta_{0i}) + (\widehat{F}\widehat{\gamma}_i - F_0\gamma_{0i})\right]$$

$$= \frac{a_{NT}^2}{NT}\sum_{i=1}^N(\widehat{\beta}_i-\beta_{0i})'X_i'X_i(\widehat{\beta}_i-\beta_{0i})+\frac{a_{NT}^2}{NT}\sum_{i=1}^N(\widehat{F}\widehat{\gamma}_i-F_0\gamma_{0i})'(\widehat{F}\widehat{\gamma}_i-F_0\gamma_{0i})$$
$$+\frac{2a_{NT}^2}{NT}\sum_{i=1}^N(\widehat{\beta}_i-\beta_{0i})'X_i'(\widehat{F}\widehat{\gamma}_i-F_0\gamma_{0i}), \tag{B.6}$$

which immediately yields that

$$\frac{1}{NT}\sum_{i=1}^N\|\widehat{F}\widehat{\gamma}_i-F_0\gamma_{0i}\|^2=o_P(1). \tag{B.7}$$

The second result then follows.

(3). By (B.5) and the second result of this lemma, we obtain that

$$o_P(1)=\frac{1}{NT}\sum_{i=1}^N\gamma_{0i}F_0'M_{\widehat{F}}F_0\gamma_{0i}=\text{trace}\left\{\frac{F_0'M_{\widehat{F}}F_0}{T}\cdot\frac{\Gamma_0'\Gamma_0}{N}\right\},$$

which in connection with $\frac{\Gamma_0'\Gamma_0}{N}\to_P\Sigma_\gamma$ of Assumption 2 yields that

$$o_P(1)=\text{trace}\left\{\frac{F_0'M_{\widehat{F}}F_0}{T}\right\}=\text{trace}\left\{\frac{F_0'F_0}{T}-\frac{F_0'\widehat{F}}{T}\cdot\frac{\widehat{F}'F_0}{T}\right\}.$$

By $\frac{F_0'F_0}{T}\to_P\Sigma_f$ of Assumption 2, we can further write

$$o_P(1)=\text{trace}\left\{I_{d_f}-\frac{F_0'\widehat{F}}{T}\cdot\frac{\widehat{F}'F_0}{T}\left(\frac{F_0'F_0}{T}\right)^{-1}\right\}$$
$$=\text{trace}\left\{I_{d_f}-\frac{\widehat{F}'P_{F_0}\widehat{F}}{T}\right\}. \tag{B.8}$$

Note that it is easy to show that

$$\left\|P_{\widehat{F}}-P_{F_0}\right\|^2=\text{tr}\left[(P_{\widehat{F}}-P_{F_0})^2\right]=\text{tr}\left[P_{\widehat{F}}-P_{\widehat{F}}P_{F_0}-P_{F_0}P_{\widehat{F}}+P_{F_0}\right]$$
$$=\text{tr}\left[I_{d_{\max}}\right]-2\cdot\text{tr}\left[P_{\widehat{F}}P_{F_0}\right]+\text{tr}\left[I_{d_f}\right]$$
$$=(d_{\max}-d_f)+2\cdot\text{tr}[I_{d_f}-\widehat{F}'P_{F_0}\widehat{F}/T], \tag{B.9}$$

which in connection (B.8) yields the third result. The proof is now complete. ■

Proof of Theorem 2.1:

Without loss of generality, we consider the next two cases:

Case 1 (over selection): $\mathsf{d}=d_f+1$;

Case 2 (under selection): $\mathsf{d}=d_f-1$.

Start with Case 1. Note that

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[y_{it}-G_\varepsilon(z_{it}^0)+G_\varepsilon(z_{it}^0)-G_\varepsilon(\widehat{z}_{it}^{\mathsf{d}})\right]^2$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - G_\varepsilon(z_{it}^0)\right]^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[G_\varepsilon(z_{it}^0) - G_\varepsilon(\widehat{z}_{it}^{\mathsf{d}})\right]^2$$
$$+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - G_\varepsilon(z_{it}^0)\right]\cdot\left[G_\varepsilon(z_{it}^0) - G_\varepsilon(\widehat{z}_{it}^{\mathsf{d}})\right]. \tag{B.10}$$

By Lemma 2.1, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[G_\varepsilon(z_{it}^0) - G_\varepsilon(\widehat{z}_{it}^{\mathsf{d}})\right]^2 = O_P\left(\frac{1}{\sqrt{NT}}\right)$$

and similar to (B.4), we can show that uniformly in z_{it}^{d}'s

$$\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - G_\varepsilon(z_{it}^0)\right]\cdot\left[G_\varepsilon(z_{it}^0) - G_\varepsilon(z_{it}^{\mathsf{d}})\right]\right| = O_P\left(\frac{1}{\sqrt{NT}}\right).$$

Thus, simple algebra shows that

$$\mathrm{IC}(\mathsf{d}) - \mathrm{IC}(d_f) = O_P\left(\frac{1}{\sqrt{NT}}\right) + (\mathsf{d} - d_f)\cdot\frac{\xi_{NT}}{\sqrt{NT}} > 0 \tag{B.11}$$

with probability 1, given $\xi_{NT} \to \infty$ and $\frac{\xi_{NT}}{\sqrt{NT}}\ to 0$.

Next, consider Case 2. Note that under Case 2, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[G_\varepsilon(z_{it}^0) - G_\varepsilon(\widehat{z}_{it}^{\mathsf{d}})\right]^2$ is no longer negligible. Otherwise, (B.9) would give $\left\|P_{\widehat{F}} - P_{F_0}\right\|^2 < 0$ with probability approaching one, which is apparently not feasible.

Based on the above development, the result follows. ■

Having established Theorem 2.1, in what follows, we assume the number of factors has been correctly selected. Therefore, we will no longer use $\widehat{d}_{\max}$ by slightly abusing the notations when no misunderstanding arises.

Proof of Lemma 2.3:

Before proceeding further, recall those notations defined at the beginning of the appendix, and recall that $z_{it} = x_{it}'\beta_i + \gamma_i' f_t$, $z_{it}^0 = x_{it}'\beta_{0i} + \gamma_{0i}' f_{0t}$ and $\widehat{z}_{it} = x_{it}'\widehat{\beta}_i + \widehat{\gamma}_i'\widehat{f}_t$.

(1). In the following proofs, we first find expressions for $\widehat{\theta}_i - \theta_{0i}$ and $\widehat{f}_t - f_{0t}$ from the first-order conditions. After deriving their leading terms, we can then compute the rates of convergence for $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\theta}_i - \theta_{0i}\|^2$ and $\frac{1}{T}\sum_{t=1}^{T}\|\widehat{f}_t - f_{0t}\|$.

We proceed with the first derivatives of the log-likelihood function with respect to β_i, f_t and γ_i:

$$\begin{aligned}\frac{\partial \log L(B,F,\Gamma)}{\partial \beta_i} &= \sum_{t=1}^{T}\left\{-(1-y_{it})\frac{g_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)}{1 - G_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)} + y_{it}\frac{g_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)}{G_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)}\right\}x_{it} \\ &= \sum_{t=1}^{T}\left\{\frac{[y_{it} - G_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)]g_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)}{[1 - G_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)]G_\varepsilon(x_{it}'\beta_i + \gamma_i' f_t)}\right\}x_{it},\end{aligned}$$

$$\frac{\partial \log L(B,F,\Gamma)}{\partial f_t} = \sum_{i=1}^{N}\left\{-(1-y_{it})\frac{g_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}{1-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}+y_{it}\frac{g_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}{G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}\right\}\gamma_i$$
$$= \sum_{i=1}^{N}\left\{\frac{[y_{it}-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)]g_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}{[1-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)]G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}\right\}\gamma_i,$$

$$\frac{\partial \log L(B,F,\Gamma)}{\partial \gamma_i} = \sum_{t=1}^{T}\left\{-(1-y_{it})\frac{g_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}{1-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}+y_{it}\frac{g_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}{G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}\right\}f_t$$
$$= \sum_{t=1}^{T}\left\{\frac{[y_{it}-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)]g_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}{[1-G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)]G_\varepsilon(x_{it}'\beta_i+\gamma_i'f_t)}\right\}f_t. \tag{B.12}$$

Using the notations defined at the beginning of the appendix, we can rewrite (B.12) as

$$\frac{\partial \log L(\Theta,F)}{\partial \theta_i} = \sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it})]g_\varepsilon(z_{it})}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}u_{it},$$
$$\frac{\partial \log L(\Theta,F)}{\partial f_t} = \sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(z_{it})]g_\varepsilon(z_{it})}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}\gamma_i. \tag{B.13}$$

Furthermore, we can obtain the second derivatives of the log-likelihood functions:

$$\frac{\partial^2 \log L(\Theta,F)}{\partial \theta_i\partial\theta_i'} = -\sum_{t=1}^{T}\left\{\frac{[g_\varepsilon(z_{it})]^2}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}\right\}u_{it}u_{it}',$$
$$+\sum_{t=1}^{T}\left\{\frac{[y_{it}-G_\varepsilon(z_{it})][g_\varepsilon^{(1)}(z_{it})G_\varepsilon(z_{it})(1-G_\varepsilon(z_{it}))+[g_\varepsilon(z_{it})]^2(1-2G_\varepsilon(z_{it}))]}{[1-G_\varepsilon(z_{it})]^2[G_\varepsilon(z_{it})]^2}\right\}u_{it}u_{it}',$$

$$\frac{\partial^2 \log L(\Theta,F)}{\partial f_t\partial f_t'} = -\sum_{i=1}^{N}\left\{\frac{[g_\varepsilon(z_{it})]^2}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}\right\}\gamma_i\gamma_i',$$
$$+\sum_{i=1}^{N}\left\{\frac{[y_{it}-G_\varepsilon(z_{it})][g_\varepsilon^{(1)}(z_{it})G_\varepsilon(z_{it})(1-G_\varepsilon(z_{it}))+[g_\varepsilon(z_{it})]^2(1-2G_\varepsilon(z_{it}))]}{[1-G_\varepsilon(z_{it})]^2[G_\varepsilon(z_{it})]^2}\right\}\gamma_i\gamma_i',$$

$$\frac{\partial^2 \log L(\Theta,F)}{\partial \theta_i\partial f_t'} = -\frac{[g_\varepsilon(z_{it})]^2}{[1-G_\varepsilon(z_{it})]G_\varepsilon(z_{it})}u_{it}\gamma_i',$$
$$+\frac{[y_{it}-G_\varepsilon(z_{it})][g_\varepsilon^{(1)}(z_{it})G_\varepsilon(z_{it})(1-G_\varepsilon(z_{it}))+[g_\varepsilon(z_{it})]^2(1-2G_\varepsilon(z_{it}))]}{[1-G_\varepsilon(z_{it})]^2[G_\varepsilon(z_{it})]^2}u_{it}\gamma_i'.$$

From the first-order condition for θ_i, we take the Taylor expansion at (B_0, F_0, Γ_0),

$$0 = \frac{1}{T}\frac{\partial \log L(\widehat{\Theta},\widehat{F})}{\partial \theta_i}$$
$$= \frac{1}{T}\frac{\partial \log L(\Theta_0,F_0)}{\partial \theta_i} + \frac{1}{T}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial \theta_i\partial\theta_i'}(\widehat{\theta}_i-\theta_{0i}) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial \theta_i\partial f_t'}(\widehat{f}_t - f_{0t})$$
$$+O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right) + O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right).$$

Then we can write that

$$
\begin{aligned}
\widehat{\theta}_i - \theta_{0i} = & -\left(\frac{1}{T}\frac{\partial^2 \log L(\Theta_0, F_0)}{\partial\theta_i\partial\theta_i'}\right)^{-1} \cdot \frac{1}{T}\frac{\partial \log L(\Theta_0, F_0)}{\partial\theta_i} \\
& -\left(\frac{1}{T}\frac{\partial^2 \log L(\Theta_0, F_0)}{\partial\theta_i\partial\theta_i'}\right)^{-1} \cdot \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log L(\Theta_0, F_0)}{\partial\theta_i\partial f_t'}(\widehat{f}_t - f_{0t}) \\
& + O_P\left(\frac{1}{N}\|\widehat{\Theta} - \Theta_0\|^2\right) + O_P\left(\frac{1}{T}\|\widehat{F} - F_0\|^2\right).
\end{aligned}
$$

To derive the rate of convergence for $\widehat{\theta}_i - \theta_{0i}$, we first show that the inverse matrix $\left(\frac{1}{T}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial\theta_i\partial\theta_i'}\right)^{-1}$ will converge to a finite and positive-definite matrix. By Assumption 3, we have

$$
\begin{aligned}
\frac{1}{T}\frac{\partial^2 L(\Theta_0, F_0)}{\partial\theta_i\partial\theta_i'} = & \Sigma_{u,i} + \frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]^2[G_\varepsilon(z_{it}^0)]^2} \cdot u_{it}^0 u_{it}^{0\prime} \\
& + \frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it} - G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2[1 - 2G_\varepsilon(z_{it}^0)]}{[1 - G_\varepsilon(z_{it}^0)]^2[G_\varepsilon(z_{it}^0)]^2} \cdot u_{it}^0 u_{it}^{0\prime} \\
& + O_P\left(\frac{1}{\sqrt{T}}\right). \qquad \text{(B.14)}
\end{aligned}
$$

In the following arguments, we show the rates of convergence for the second and third terms on the right hand side of (B.14) by computing their first and second moments. Note that for the second term on the right hand side of (B.14), by the law of iterated expectations and the fact $E[y_{it}|x_{it}, \gamma_{0i}, f_{0t}] = G_\varepsilon(z_{it}^0)$,

$$
E\left[\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]^2[G_\varepsilon(z_{it}^0)]^2} \cdot u_{it}^0 u_{it}^{0\prime}\right] = 0. \qquad \text{(B.15)}
$$

For its second moment, we first note that $\max_{z_it}\{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)\}^{-1} = O(1)$ by Assumption 1. In addition, since $g_\varepsilon^{(1)}(\cdot)$ is uniformly bounded and by Assumption 3,

$$
\begin{aligned}
& E\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]^2[G_\varepsilon(z_{it}^0)]^2} \cdot u_{it}^0 u_{it}^{0\prime}\right\|^2 \\
\le & O(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\|u_{it}^0\|^2 \cdot \|u_{is}^0\|^2 \cdot |E[e_{it}e_{is} \,|\, w_{it}^0, w_{is}^0]|] \\
\le & O(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\|u_{it}^0\|^2 \cdot \|u_{is}^0\|^2] \cdot \alpha_{ii,ts} \\
\le & O(1)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\alpha_{ii,ts} = O\left(\frac{1}{T}\right), \qquad \text{(B.16)}
\end{aligned}
$$

where the last inequality holds by CauchySchwarz inequality and Assumption 3:

$$
E[\|u_{it}^0\|^2\|u_{is}^0\|^2] \le \left\{E[\|u_{it}^0\|^4] \cdot E[\|u_{is}^0\|^4]\right\}^{\frac{1}{2}} = O(1).
$$

Therefore, by (B.15) and (B.16),

$$
\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]^2[G_\varepsilon(z_{it}^0)]^2} \cdot u_{it}^0 u_{it}^{0\prime} = O_P\left(\frac{1}{\sqrt{T}}\right). \qquad \text{(B.17)}
$$

Analogously to (B.17), for the third term on the right hand side of (B.14),

$$\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_{\varepsilon}(z_{it}^{0})][g_{\varepsilon}(z_{it}^{0})]^{2}[1-2G_{\varepsilon}(z_{it}^{0})]}{[1-G_{\varepsilon}(z_{it}^{0})]^{2}[G_{\varepsilon}(z_{it}^{0})]^{2}}\cdot u_{it}^{0}u_{it}^{0\prime}=O_{P}\left(\frac{1}{\sqrt{T}}\right). \tag{B.18}$$

By (B.17) and (B.18), we can conclude that

$$\frac{1}{T}\frac{\partial^{2}L(\Theta_{0},F_{0})}{\partial\theta_{i}\partial\theta_{i}^{\prime}}=\Sigma_{u,i}+O_{P}\left(\frac{1}{\sqrt{T}}\right). \tag{B.19}$$

Jointly with the condition that $\Sigma_{u,i}$ is invertible in Assumption 3, it immediately yields that

$$\begin{aligned}\widehat{\theta}_{i}-\theta_{0i}=&-\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\frac{\partial\log L(\Theta_{0},F_{0})}{\partial\theta_{i}}-\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^{2}\log L(\Theta_{0},F_{0})}{\partial\theta_{i}\partial f_{t}^{\prime}}(\widehat{f}_{t}-f_{0t})\\&+O_{P}\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_{0}\|^{2}\right)+O_{P}\left(\frac{1}{T}\|\widehat{F}-F_{0}\|^{2}\right)+O_{P}\left(\frac{1}{\sqrt{T}}\right).\end{aligned} \tag{B.20}$$

In (B.20), we have derived the leading terms in $\widehat{\theta}_{i}-\theta_{0i}$. Now we proceed with FOC for $\widehat{f}_{t}$. By the Taylor expansion at (B_{0},F_{0},Γ_{0}), we have

$$\begin{aligned}0=&\frac{1}{N}\frac{\partial\log L(\widehat{\Theta},\widehat{F})}{\partial f_{t}}\\=&\frac{1}{N}\frac{\partial\log L(\Theta_{0},F_{0})}{\partial f_{t}}+\frac{1}{N}\frac{\partial^{2}\log L(\Theta_{0},F_{0})}{\partial f_{t}\partial f_{t}^{\prime}}(\widehat{f}_{t}-f_{0t})+\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2}L(\Theta_{0},F_{0})}{\partial f_{t}\partial\theta_{i}^{\prime}}(\widehat{\theta}_{i}-\theta_{0i})\\&+O_{P}\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_{0}\|^{2}\right)+O_{P}\left(\frac{1}{T}\|\widehat{F}-F_{0}\|^{2}\right).\end{aligned}$$

For the second derivative of the log-likelihood function with respect to f_{t}, analogously to (B.19) and by Assumption 3, we obtain that

$$\frac{1}{N}\frac{\partial^{2}\log L(\Theta_{0},F_{0})}{\partial f_{t}\partial f_{t}^{\prime}}=\Sigma_{\gamma,t}+O_{P}\left(\frac{1}{\sqrt{N}}\right).$$

Therefore, we can further write $\widehat{f}_{t}-f_{0t}$ as

$$\begin{aligned}\widehat{f}_{t}-f_{0t}=&-\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\frac{\partial\log L(\Theta_{0},F_{0})}{\partial f_{t}}-\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^{2}\log L(\Theta_{0},F_{0})}{\partial f_{t}\partial\theta_{i}^{\prime}}(\widehat{\theta}_{i}-\theta_{0i})\\&+O_{P}\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_{0}\|^{2}\right)+O_{P}\left(\frac{1}{T}\|\widehat{F}-F_{0}\|^{2}\right)+O_{P}\left(\frac{1}{\sqrt{N}}\right).\end{aligned} \tag{B.21}$$

From (B.20), we know that the leading term in $\widehat{\theta}_{i}-\theta_{0i}$ can be expressed by $\widehat{f}_{t}-f_{0t}$. On the other hand, $\widehat{\theta}_{i}-\theta_{0i}$ constitutes the leading term in $\widehat{f}_{t}-f_{0t}$, as can be seen from (B.21).

Substituting (B.21) into (B.20), the second term on the right hand side of (B.20) becomes

$$\begin{aligned}&-\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^{2}\log L(\Theta_{0},F_{0})}{\partial\theta_{i}\partial f_{t}^{\prime}}(\widehat{f}_{t}-f_{0t})\\=&\ \Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^{2}\log L(\Theta_{0},F_{0})}{\partial\theta_{i}\partial f_{t}^{\prime}}\cdot\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\frac{\partial\log L(\Theta_{0},F_{0})}{\partial f_{t}}\end{aligned}$$

$$+\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\log L(\Theta_0,F_0)}{\partial\theta_i\partial f_t'}\cdot\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{\partial^2\log L(\Theta_0,F_0)}{\partial f_t\partial\theta_j'}(\widehat{\theta}_j-\theta_{0j})$$
$$+O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)+O_P\left(\frac{1}{\sqrt{N}}\right). \tag{B.22}$$

We now proceed with the first term on the right hand side of (B.22). By (B.13),

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}E\left\|\frac{1}{N}\frac{\partial\log L(\Theta_0,F_0)}{\partial f_t}\right\|^2 &= \frac{1}{T}\sum_{t=1}^{T}E\left\|\frac{1}{N}\sum_{i=1}^{N}g_\varepsilon(z_{it}^0)\gamma_{0i}e_{it}\right\|^2\\ &\leq O(1)\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E[\|\gamma_{0i}\|\cdot\|\gamma_{0j}\|\cdot|E[e_{it}e_{jt}|w_{it}^0,w_{jt}^0]|].\end{aligned} \tag{B.23}$$

By Assumption 3 and CauchySchwarz inequality,

$$\begin{aligned}&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E[\|\gamma_{0i}\|\cdot\|\gamma_{0j}\|\cdot|E[e_{it}e_{jt}|w_{it}^0,w_{jt}^0]|]\\ \leq\ &\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}E[\|\gamma_{0i}\|\cdot\|\gamma_{0j}\|]\cdot\alpha_{ij,tt}\\ \leq\ &\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\left\{E[\|\gamma_{0i}\|^2]\right\}^{\frac{1}{2}}\cdot\left\{E[\|\gamma_{0j}\|^2]\right\}^{\frac{1}{2}}\cdot\alpha_{ij,tt}\\ \leq\ &O(1)\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\alpha_{ij,tt}=O\left(\frac{1}{N}\right).\end{aligned} \tag{B.24}$$

By (B.23) and (B.24), we have

$$\frac{1}{T}\sum_{t=1}^{T}E\left\|\frac{1}{N}\frac{\partial\log L(\Theta_0,F_0)}{\partial f_t}\right\|^2=O\left(\frac{1}{N}\right). \tag{B.25}$$

With (B.25), we are able to show the convergence of the first term on the right hand side of (B.22). By CauchySchwarz inequality and (B.25),

$$\begin{aligned}&\left\|\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\log L(\Theta_0,F_0)}{\partial\theta_i\partial f_t'}\cdot\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\frac{\partial\log L(\Theta_0,F_0)}{\partial f_t}\right\|\\ &\leq O(1)\frac{1}{T}\left(\sum_{t=1}^{T}\left\|\frac{\partial^2\log L(\Theta_0,F_0)}{\partial\theta_i\partial f_t'}\right\|^2\right)^{\frac{1}{2}}\cdot\left(\sum_{t=1}^{T}\left\|\frac{1}{N}\frac{\partial\log L(\Theta_0,F_0)}{\partial f_t}\right\|^2\right)^{\frac{1}{2}}\\ &=O_P\left(\frac{1}{\sqrt{N}}\right).\end{aligned} \tag{B.26}$$

Now we proceed with the second term on the right hand side of (B.22). Following analogous arguments to the derivation of (B.26),

$$
\begin{aligned}
&\Sigma_{u,i}^{-1} \cdot \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \log L(\Theta_0, F_0)}{\partial \theta_i \partial f_t'} \cdot \Sigma_{\gamma,t}^{-1} \cdot \frac{1}{N} \sum_{j=1}^{N} \frac{\partial^2 \log L(\Theta_0, F_0)}{\partial f_t \partial \theta_j'} (\widehat{\theta}_j - \theta_{0j}) \\
&= \Sigma_{u,i}^{-1} \cdot \frac{1}{T} \sum_{t=1}^{T} \frac{[g_\varepsilon(z_{it}^0)]^2}{[1 - G_\varepsilon(z_{it}^0)] G_\varepsilon(z_{it}^0)} u_{it}^0 \gamma_{0i}' \cdot \Sigma_{\gamma,t}^{-1} \cdot \frac{1}{N} \sum_{j=1}^{N} \frac{[g_\varepsilon(z_{jt}^0)]^2}{[1 - G_\varepsilon(z_{jt}^0)] G_\varepsilon(z_{jt}^0)} \gamma_{0j} u_{jt}^{0\prime} \cdot (\widehat{\theta}_j - \theta_{0j}) \\
&+ O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right).
\end{aligned}
$$

Then we can further simplify (B.22) as

$$
\begin{aligned}
&-\Sigma_{u,i}^{-1} \cdot \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \log L(\Theta_0, F_0)}{\partial \theta_i \partial f_t'} (\widehat{f}_t - f_{0t}) \\
&= \frac{1}{N} \sum_{j=1}^{N} D_{ij} (\widehat{\theta}_j - \theta_{0j}) + O_P\left(\frac{1}{N} \|\widehat{\Theta} - \Theta_0\|^2\right) \\
&+ O_P\left(\frac{1}{T} \|\widehat{F} - F_0\|^2\right) + O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right),
\end{aligned}
\tag{B.27}
$$

where

$$
D_{ij} = \frac{1}{T} \sum_{t=1}^{T} \frac{[g_\varepsilon(z_{it}^0)]^2}{[1 - G_\varepsilon(z_{it}^0)] G_\varepsilon(z_{it}^0)} u_{it}^0 \gamma_{0i}' \cdot \Sigma_{\gamma,t}^{-1} \cdot \frac{[g_\varepsilon(z_{jt}^0)]^2}{[1 - G_\varepsilon(z_{jt}^0)] G_\varepsilon(z_{jt}^0)} \gamma_{0j} u_{jt}^{0\prime} \cdot .
$$

We obtain that $\frac{1}{N} \sum_{j=1}^{N} D_{ij} (\widehat{\theta}_j - \theta_{0j})$ as the leading term in the second term on the right hand side of (B.22), while the small terms in it are bounded by $O_P\left(\frac{1}{N} \|\widehat{\Theta} - \Theta_0\|^2\right)$, $O_P\left(\frac{1}{T} \|\widehat{F} - F_0\|^2\right)$ and $O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right)$. In addition, recall that we have shown that the first term on the right hand side of (B.22) is bounded by $O_P\left(\frac{1}{\sqrt{N}}\right)$ in (B.26). Therefore, we can rewrite (B.20) as

$$
\begin{aligned}
\widehat{\theta}_i - \theta_{0i} = &-\Sigma_{u,i}^{-1} \cdot \frac{1}{T} \frac{\partial \log L(\Theta_0, F_0)}{\partial \theta_i} + \frac{1}{N} \sum_{j=1}^{N} D_{ij} (\widehat{\theta}_j - \theta_{0j}) \\
&+ O_P\left(\frac{1}{N} \|\widehat{\Theta} - \Theta_0\|^2\right) + O_P\left(\frac{1}{T} \|\widehat{F} - F_0\|^2\right) + O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right).
\end{aligned}
$$

Following an argument similar to those for Eq. (41) of the supplementary document of Ando and Lu (2020), we can further write

$$
\begin{aligned}
\widehat{\theta}_i - \theta_{0i} = &-\Sigma_{u,i}^{-1} \cdot \frac{1}{T} \frac{\partial \log L(\Theta_0, F_0)}{\partial \theta_i} + O_P\left(\frac{1}{N} \|\widehat{\Theta} - \Theta_0\|^2\right) \\
&+ O_P\left(\frac{1}{T} \|\widehat{F} - F_0\|^2\right) + O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right).
\end{aligned}
\tag{B.28}
$$

Note that

$$
E\left[\frac{1}{T} \frac{\partial \log L(\Theta_0, F_0)}{\partial \theta_i}\right] = E\left[\frac{1}{T} \sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)] g_\varepsilon(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)] G_\varepsilon(z_{it}^0)} u_{it}^0\right] = 0, \tag{B.29}
$$

and

$$
\begin{aligned}
E\left\|\frac{1}{T}\frac{\partial \log L(\Theta_0,F_0)}{\partial \theta_i}\right\|^2 &= E\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\right\|^2 \\
&\le \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\|u_{it}^0\|\cdot\|u_{is}^0\|\cdot|E[e_{it}e_{is}\,|\,w_{it}^0,w_{is}^0]|] \\
&= O\left(\frac{1}{T}\right),
\end{aligned} \tag{B.30}
$$

where the second equality follows from Assumption 3. Therefore, by (B.29) and (B.30),

$$
\frac{1}{T}\frac{\partial \log L(\Theta_0,F_0)}{\partial \theta_i} = O_P\left(\frac{1}{\sqrt{T}}\right), \tag{B.31}
$$

which in connection with (B.28) immediately yields that

$$
\begin{aligned}
\widehat{\theta}_i-\theta_{0i} &= -\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\frac{\partial \log L(\Theta_0,F_0)}{\partial \theta_i}+O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right) \\
&\quad +O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)+O_P\left(\frac{1}{\min\{\sqrt{N},\sqrt{T}\}}\right) \\
&= O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right) \\
&\quad +O_P\left(\frac{1}{\min\{\sqrt{N},\sqrt{T}\}}\right).
\end{aligned} \tag{B.32}
$$

Substituting (B.20) into (B.21), the second term of (B.21) gives that

$$
\begin{aligned}
&-\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial f_t\partial \theta_i'}(\widehat{\theta}_i-\theta_{0i}) \\
&= \Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial f_t\partial \theta_i'}\cdot\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\frac{\partial \log L(\Theta_0,F_0)}{\partial \theta_i} \\
&\quad +\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial f_t\partial \theta_i'}\cdot\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial \theta_i\partial f_t'}(\widehat{f}_t-f_{0t}) \\
&\quad +O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)+O_P\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}
$$

Similar to the development of (B.32), we can obtain that

$$
\begin{aligned}
\widehat{f}_t-f_{0t} &= -\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\frac{\partial \log L(\Theta_0,F_0)}{\partial f_t}+O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right) \\
&\quad +O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)+O_P\left(\frac{1}{\min\{\sqrt{N},\sqrt{T}\}}\right).
\end{aligned} \tag{B.33}
$$

Note that

$$
E\left[\frac{1}{N}\frac{\partial \log L(\Theta_0,F_0)}{\partial f_t}\right] = E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}\right] = 0, \tag{B.34}
$$

and

$$
E\left\|\frac{1}{N}\frac{\partial \log L(\Theta_0,F_0)}{\partial f_t}\right\|^2 = E\left\|\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}\right\|^2
$$

$$
\begin{aligned}
&\leq O(1)\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\|\gamma_{0i}\|\cdot\|\gamma_{0j}\|\cdot|E[e_{it}e_{jt}\,|\,w_{it}^0,w_{ij}^0]|\right]\\
&\leq O(1)\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\|\gamma_{0i}\|\cdot\|\gamma_{0j}\|\right]\cdot\alpha_{ij,tt}\\
&=O\left(\frac{1}{N}\right),
\end{aligned}
\tag{B.35}
$$

where the second equality follows from Assumption 3 and CauchySchwarz inequality:

$$
\begin{aligned}
E\left[\|\gamma_{0i}\|\cdot\|\gamma_{0j}\|\right] &\leq \left\{E\left[\|\gamma_{0i}\|^2\right]\right\}^{\frac{1}{2}}\cdot\left\{E\left[\|\gamma_{0j}\|^2\right]\right\}^{\frac{1}{2}}\\
&=O(1).
\end{aligned}
$$

Then, by (B.34) and (B.35),

$$
\frac{1}{N}\frac{\partial\log L(\Theta_0,F_0)}{\partial f_t}=O_P\left(\frac{1}{\sqrt{N}}\right),
\tag{B.36}
$$

which in connection with (B.33) yields that

$$
\begin{aligned}
\widehat{f}_t-f_{0t}&=-\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\frac{\partial\log L(\Theta_0,F_0)}{\partial f_t}+O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)\\
&\quad+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)+O_P\left(\frac{1}{\min\{\sqrt{N},\sqrt{T}\}}\right)\\
&=O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)\\
&\quad+O_P\left(\frac{1}{\min\{\sqrt{N},\sqrt{T}\}}\right).
\end{aligned}
\tag{B.37}
$$

At current stage, we have shown that $\widehat{\theta}_i-\theta_{0i}$ and $\widehat{f}_t-f_{0t}$ are both bounded by $O_P\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)$, $O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)$ and $O_P\left(\frac{1}{\min\{\sqrt{N},\sqrt{T}\}}\right)$.

We now revisit the rates of convergence of $\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2$ and $\frac{1}{T}\|\widehat{F}-F_0\|^2$ from (B.20) and (B.21).

For the first term in (B.20), we can show that

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left\|\Sigma_{u,i}^{-1}\left(\frac{1}{T}\frac{\partial\log L(\Theta_0,F_0)}{\partial\theta_i}\right)\right\|^2 &\leq O(1)\frac{1}{NT^2}\sum_{i=1}^{N}\left\|\frac{\partial\log L(\Theta_0,F_0)}{\partial\theta_i}\right\|^2\\
&=O_P\left(\frac{1}{T}\right),
\end{aligned}
\tag{B.38}
$$

where the last step follows from (B.30).

For the second term in (B.20), write

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left\|\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\log L(\Theta_0,F_0)}{\partial\theta_i\partial f_t'}(\widehat{f}_t-f_{0t})\right\|^2\\
&\leq O(1)\frac{1}{NT^2}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}\frac{\partial^2\log L(\Theta_0,F_0)}{\partial\theta_i\partial f_t'}(\widehat{f}_t-f_{0t})\right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&= O_P\left(\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)^2\right) + O_P\left(\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)^2\right) \\
&\quad + O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.39)}
\end{aligned}
$$

Thus, by (B.20), (B.38) and (B.39), we have

$$
\begin{aligned}
\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2 &= O_P\left(\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)^2\right) + O_P\left(\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)^2\right) \\
&\quad + O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.40)}
\end{aligned}
$$

We then proceed with (B.21). For the first term in (B.21), we can show that

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left\|\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\frac{\partial \log L(\Theta_0,F_0)}{\partial f_t}\right\|^2 &\leq O(1)\frac{1}{N^2T}\sum_{t=1}^{T}\left\|\frac{\partial \log L(\Theta_0,F_0)}{\partial f_t}\right\|^2 \\
&= O_P\left(\frac{1}{N}\right), \qquad \text{(B.41)}
\end{aligned}
$$

where the last step follows from (B.25).

For the second term in (B.21), write

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\|\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial f_t \partial \theta_i'}(\widehat{\theta}_i-\theta_{0i})\right\|^2 \\
&\leq C\frac{1}{N^2T}\sum_{t=1}^{T}\left\|\sum_{i=1}^{N}\frac{\partial^2 \log L(\Theta_0,F_0)}{\partial f_t \partial \theta_i'}(\widehat{\theta}_i-\theta_{0i})\right\|^2 \\
&= O_P\left(\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)^2\right) + O_P\left(\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)^2\right) \\
&\quad + O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.42)}
\end{aligned}
$$

By (B.21), (B.41) and (B.42),

$$
\begin{aligned}
\frac{1}{T}\|\widehat{F}-F_0\|^2 &= O_P\left(\left(\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2\right)^2\right) + O_P\left(\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)^2\right) \\
&\quad + O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.43)}
\end{aligned}
$$

Based on the above development, using (B.32), (B.37), (B.40), (B.43) and the fact that $\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2 = o_P(1)$ and $\frac{1}{T}\|\widehat{F}-F_0\|^2$, which is implied by Lemma 2.2, we can now conclude that

$$
\begin{aligned}
\frac{1}{N}\|\widehat{\Theta}-\Theta_0\|^2 &= O_P\left(\frac{1}{\min\{N,T\}}\right), \\
\frac{1}{T}\|\widehat{F}-F_0\|^2 &= O_P\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned}
$$

Moreover, we also can conclude that

$$\widehat{\theta}_i - \theta_{0i} = O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right) \text{ for } i = 1, \ldots, N, \tag{B.44}$$

and

$$\widehat{f}_t - f_{0t} = O_P\left(\frac{1}{\min\{\sqrt{N}, \sqrt{T}\}}\right) \text{ for } t = 1, \ldots, T. \tag{B.45}$$

The proof is now complete. ■

Proof of Lemma 2.2:

(1) With additional conditions on the weak cross sectional dependence and time series correlation on error terms in Assumption 4, we can establish a $\sqrt{T}$-consistency for $\widehat{\beta}_i$ and $\sqrt{N}$-consistency for $\widehat{f}_t$ in this lemma. To begin with, recall that we have the following first derivatives of log-likelihood functions:

$$\begin{aligned}
\frac{\partial \log L(\Theta, F)}{\partial \theta_i} &= \sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it})]g_\varepsilon(z_{it})}{[1 - G_\varepsilon(z_{it})]G_\varepsilon(z_{it})} u_{it}, \\
\frac{\partial \log L(\Theta, F)}{\partial f_t} &= \sum_{i=1}^{N} \frac{[y_{it} - G_\varepsilon(z_{it})]g_\varepsilon(z_{it})}{[1 - G_\varepsilon(z_{it})]G_\varepsilon(z_{it})} \gamma_i.
\end{aligned}$$

We first derive the leading terms in $\widehat{\beta}_i - \beta_{0i}$ from $\frac{\partial \log L(\Theta,F)}{\partial \theta_i}$. For $\frac{\partial \log L(\Theta,F)}{\partial \theta_i}$, the first order condition implies that

$$\begin{aligned}
0 &= \frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(\widehat{z}_{it})]\, g_\varepsilon(\widehat{z}_{it})}{[1 - G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it})} \widehat{u}_{it} \\
&= \frac{1}{T}\sum_{t=1}^{T} \frac{\left[y_{it} - G_\varepsilon(z_{it}^0)\right]\, g_\varepsilon(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 \\
&\quad + \frac{1}{T}\sum_{t=1}^{T} \left\{ \frac{[y_{it} - G_\varepsilon(\widehat{z}_{it})]\, g_\varepsilon(\widehat{z}_{it})}{[1 - G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it})} \widehat{u}_{it} - \frac{\left[y_{it} - G_\varepsilon(z_{it}^0)\right]\, g_\varepsilon(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 \right\} \\
&:= A_{1Ti} + A_{2Ti}.
\end{aligned} \tag{B.46}$$

A_{1Ti} only depends on the statistical behaviours of z_{it}^0 and u_{it}^0. We now proceed with A_{2Ti}. For A_{2Ti}, we have

$$\begin{aligned}
A_{2Ti} &= \frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{2,it}^{*} + \frac{1}{T}\sum_{t=1}^{T} (a_{it}^{\dagger-1} - a_{it}^{-1}) a_{2,it}^{*}, \\
&:= A_{3Ti} + A_{4Ti},
\end{aligned} \tag{B.47}$$

where

$$\begin{aligned}
a_{it} &= [1 - G_\varepsilon(z_{it}^0)]^2 [G_\varepsilon(z_{it}^0)]^2, \\
a_{it}^{\dagger} &= [1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)[1 - G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it}), \\
a_{2,it}^{*} &= [y_{it} - G_\varepsilon(\widehat{z}_{it})]g_\varepsilon(\widehat{z}_{it})[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)\widehat{u}_{it}
\end{aligned}$$

$$-[y_{it} - G_\varepsilon(z_{it}^0)]g(z_{it}^0)[1 - G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it})u_{it}^0.$$

Among them, a_{it} is a function of real value z_{it}^0, therefore we are interested in the convergence of $a_{2,it}^*$ and $a_{it}^\dagger$. We start our investigation by looking at $a_{2,it}^*$, and write

$$\begin{aligned}
a_{2,it}^* &= -[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0 \\
&\quad +[y_{it} - G_\varepsilon(z_{it}^0)][g_\varepsilon(\widehat{z}_{it}) - g_\varepsilon(z_{it}^0)][1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0 \\
&\quad +[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0 \\
&\quad -[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)][G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]u_{it}^0 \\
&\quad +[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)(\widehat{u}_{it} - u_{it}^0) \\
&\quad -[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)][g_\varepsilon(\widehat{z}_{it}) - g_\varepsilon(z_{it}^0)][1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0 \\
&\quad -[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)(\widehat{u}_{it} - u_{it}^0) \\
&\quad +[y_{it} - G_\varepsilon(z_{it}^0)][g_\varepsilon(\widehat{z}_{it}) - g_\varepsilon(z_{it}^0)][1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)(\widehat{u}_{it} - u_{it}^0) \\
&\quad +[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)]^2u_{it}^0 \\
&\quad -[G_\varepsilon(\widehat{z}_{it}) - G_\varepsilon(z_{it}^0)][g_\varepsilon(\widehat{z}_{it}) - g_\varepsilon(z_{it}^0)][1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)(\widehat{u}_{it} - u_{it}^0) \\
&:= a_{2,it}^{(1)*} + \cdots + a_{2,it}^{(10)*},
\end{aligned}$$

where the definitions of $a_{2,it}^{(1)*}$ to $a_{2,it}^{(10)*}$ are obvious. Below, we examine the terms on the right hand side one by one.

For $a_{2,it}^{(1)*}$, by the Taylor expansion, write

$$\begin{aligned}
a_{2,it}^{(1)*} &= -g_\varepsilon^2(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it} - z_{it}^0) \\
&\quad -\frac{1}{2}g_\varepsilon^{(1)}(\dot{z}_{it})g_\varepsilon(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it} - z_{it}^0)^2 \\
&:= a_{21,it}^{(1)*} + a_{22,it}^{(1)*},
\end{aligned}$$

where $\dot{z}_{it}$ lies between $\widehat{z}_{it}$ and z_{it}^0, and the definitions of $a_{21,it}^{(1)*}$ and $a_{22,it}^{(1)*}$ are obvious.

Note that for $a_{21,it}^{(1)*}$, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T a_{it}^{-1}a_{21,it}^{(1)*} &= -\frac{1}{T}\sum_{t=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0(\widehat{z}_{it} - z_{it}^0) \\
&= -\frac{1}{T}\sum_{t=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0x_{it}'(\widehat{\beta}_i - \beta_{0i}) \\
&\quad -\frac{1}{T}\sum_{t=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0(\widehat{\gamma}_i'\widehat{f}_t - \gamma_{0i}'f_{0t}) \\
&= -\frac{1}{T}\sum_{t=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
&\quad -\frac{1}{T}\sum_{t=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t})
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0(\widehat{\gamma}_i-\gamma_{0i})'(\widehat{f}_t-f_{0t})\\
=&-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0u_{it}^{0\prime}(\widehat{\theta}_i-\theta_{0i})\\
&-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t})\\
&+O_P\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned}
$$

where the last equality holds, because

$$
\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0(\widehat{\gamma}_i-\gamma_{0i})'(\widehat{f}_t-f_{0t})\right\|\\
&\leq\|\widehat{\gamma}_i-\gamma_{0i}\|\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\right\|^2\right\}^{\frac{1}{2}}\cdot\left\{\frac{1}{T}\sum_{t=1}^{T}\left\|\widehat{f}_t-f_{0t}\right\|^2\right\}^{\frac{1}{2}}\\
&=O_P\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned}
$$

in which we have used the Cauchy-Schwarz inequality and Lemma 2.3. Furthermore, by Lemma 2.3 and (B.44), we have

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0u_{it}^{0\prime}(\widehat{\theta}_i-\theta_{0i})\\
&=\Sigma_{u,i}(\widehat{\theta}_i-\theta_{0i})+O_P\left(\frac{1}{\sqrt{T}\min\{\sqrt{N},\sqrt{T}\}}\right).
\end{aligned}
$$

Therefore, we obtain that

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}a_{2,it}^{-1}a_{21,it}^{(1)*}=&-\Sigma_{u,i}(\widehat{\theta}_i-\theta_{0i})-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t})\\
&+O_P\left(\frac{1}{\min\{N,T\}}\right).
\end{aligned}
\tag{B.48}
$$

For $a_{22,it}^{(1)*}$, by Lemma 2.3 and (B.44), it is clear to see that

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{22,it}^{(1)*}&=-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^{(1)}(\dot{z}_{it})g_\varepsilon(z_{it}^0)}{2[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0(\widehat{z}_{it}-z_{it}^0)^2\\
&=O_P(\|\widehat{\theta}_i-\theta_{0i}\|^2)+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right)\\
&=O_P\left(\frac{1}{\min\{N,T\}}\right).
\end{aligned}
\tag{B.49}
$$

By (B.48) and (B.49),

$$
\frac{1}{T}\sum_{t=1}^{T}a_{2,it}^{-1}a_{2,it}^{(1)*}=-\Sigma_{u,i}(\widehat{\theta}_i-\theta_{0i})-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t})
$$

$$+O_P\left(\frac{1}{\min\{N,T\}}\right). \tag{B.50}$$

After obtaining the leading term in $a_{2,it}^{(1)*}$, we proceed with $a_{2,it}^{(2)*}$. For $a_{2,it}^{(2)*}$, by the Taylor expansion, we have

$$\begin{aligned}
a_{2,it}^{(2)*} &= [y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it} - z_{it}^0) \\
&\quad + \frac{1}{2}[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(2)}(\ddot{z}_{it})[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it} - z_{it}^0)^2 \\
&:= a_{21,it}^{(2)*} + a_{22,it}^{(2)*},
\end{aligned}$$

where $\ddot{z}_{it}$ lies between $\widehat{z}_{it}$ and z_{it}^0, and the definitions of $a_{21,it}^{(2)*}$ are $a_{22,it}^{(2)*}$ are obvious.

For $a_{21,it}^{(2)*}$, write

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T a_{it}^{-1} a_{21,it}^{(2)*} &= \frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0(\widehat{z}_{it} - z_{it}^0) \\
&= \frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
&\quad + \frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 \gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
&\quad + \frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 (\widehat{\gamma}_i - \gamma_{0i})'(\widehat{f}_t - f_{0t}).
\end{aligned} \tag{B.51}$$

Recall that $e_{it} = \frac{y_{it} - G_\varepsilon(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}$. For the first term in (B.51), note that

$$E\left[\frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 u_{it}^{0\prime}\right] = 0. \tag{B.52}$$

In addition, we have

$$\begin{aligned}
&E\left\|\frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 u_{it}^{0\prime}\right\|^2 \\
&\le O(1)\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T E[\|u_{it}\|^2 \cdot \|u_{is}^0\|^2 \cdot |E[e_{it}e_{is} \,|\, w_{it}^0, w_{is}^0]] \\
&= O\left(\frac{1}{T}\right),
\end{aligned} \tag{B.53}$$

where the equality follows from Assumption 3. By (B.52) and (B.53),

$$\frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 u_{it}^{0\prime} = O_P\left(\frac{1}{\sqrt{T}}\right), \tag{B.54}$$

which in connection with (B.44) yields that

$$\frac{1}{T}\sum_{t=1}^T \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) = O_P\left(\frac{1}{\sqrt{T}\min\{\sqrt{N},\sqrt{T}\}}\right).$$

Therefore, we have shown that the first term in (B.51) is $o_P(\frac{1}{\sqrt{T}})$. For the third term in (B.51), by Lemma 2.3, (B.44) and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\left\| \frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 (\widehat{\gamma}_i - \gamma_{0i})'(\widehat{f}_t - f_{0t}) \right\| \\
&\leq \left\{ \frac{1}{T}\sum_{t=1}^{T} \left\| \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 \right\|^2 \right\}^{\frac{1}{2}} \cdot \|\widehat{\gamma}_i - \gamma_{0i}\| \cdot \left\{ \frac{1}{T}\sum_{t=1}^{T} \|\widehat{f}_t - f_{0t}\|^2 \right\}^{\frac{1}{2}} \\
&= O_P\left(\frac{1}{\min\{N, T\}}\right).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{21,it}^{(2)*} = &\frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 \gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
&+ O_P\left(\frac{1}{\min\{N, T\}}\right). \qquad \text{(B.55)}
\end{aligned}$$

For $a_{22,it}^{(2)*}$, similar to (B.49), we can obtain that

$$\frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{22,it}^{(2)*} = O_P\left(\frac{1}{\min\{N, T\}}\right). \qquad \text{(B.56)}$$

By (B.55) and (B.56),

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{2,it}^{(2)*} = &\frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} u_{it}^0 \gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
&+ O_P\left(\frac{1}{\min\{N, T\}}\right). \qquad \text{(B.57)}
\end{aligned}$$

At current stage, we have obtained the leading terms in $a_{2,it}^{(1)*}$ and $a_{2,it}^{(2)*}$. Following the argument analogously to that for these two terms, we can derive the leading terms for $a_{2,it}^{(3)*}$ and $a_{2,it}^{(4)*}$. Therefore, we omit the proofs and provide the results directly:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{2,it}^{(3)*} = &\frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2}{[1 - G_\varepsilon(z_{it}^0)]^2 G_\varepsilon(z_{it}^0)} u_{it}^0 \gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
&+ O_P\left(\frac{1}{\min\{N, T\}}\right), \\
\frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{2,it}^{(4)*} = &-\frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2}{[1 - G_\varepsilon(z_{it}^0)][G_\varepsilon(z_{it}^0)]^2} u_{it}^0 \gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
&+ O_P\left(\frac{1}{\min\{N, T\}}\right). \qquad \text{(B.58)}
\end{aligned}$$

For $a_{2,it}^{(5)*}$, recall that we have $u_{it}^0 = (x_{it}', f_{0t}')'$ and $\widehat{u}_{it} = (x_{it}', \widehat{f}_t')'$, we obtain that

$$\frac{1}{T}\sum_{t=1}^{T} a_{it}^{-1} a_{2,it}^{(5)*} = \frac{1}{T}\sum_{t=1}^{T} \frac{[y_{it} - G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1 - G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)} (\widehat{u}_{it} - u_{it}^0)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}(0'_{d_\beta},(\widehat{f}_t-f_{0t})')'. \tag{B.59}$$

For $a_{2,it}^{(6)}*$, by the Taylor expansion, write

$$\begin{aligned} a_{2,it}^{(6)*} &= -[G_\varepsilon(\widehat{z}_{it})-G_\varepsilon(z_{it}^0)][g_\varepsilon(\widehat{z}_{it})-g_\varepsilon(z_{it}^0)][1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0 \\ &= -g_\varepsilon(z_{it}^\dagger)g_\varepsilon^{(1)}(z_{it}^\ddagger)[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it}-z_{it}^0)^2, \end{aligned}$$

where $z_{it}^\dagger$ and $z_{it}^\ddagger$ lie between $\widehat{z}_{it}$ and z_{it}^0. Then by Lemma 2.3 and (B.45), we obtain that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{2,it}^{(6)*} &= -\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon(z_{it}^\dagger)g_\varepsilon^{(1)}(z_{it}^\ddagger)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0(\widehat{z}_{it}-z_{it}^0)^2 \\ &= O_P(\|\widehat{\theta}_i-\theta_{0i}\|^2)+O_P\left(\frac{1}{T}\|\widehat{F}-F_0\|^2\right) \\ &= O_P\left(\frac{1}{\min\{N,T\}}\right). \end{aligned} \tag{B.60}$$

The derivations for the terms with $a_{2,it}^{(7)*}$, $a_{2,it}^{(8)*}$, $a_{2,it}^{(9)*}$ and $a_{2,it}^{(10)*}$ are analogously and one can easily shows it by the Taylor expansion, Lemma 2.3, (B.44) and (B.45). Therefore, the detailed proofs for these terms are omitted and we list the results directly here:

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{2,it}^{(7)*} &= O_P\left(\frac{1}{\min\{N,T\}}\right), \\ \frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{2,it}^{(8)*} &= O_P\left(\frac{1}{\min\{N,T\}}\right), \\ \frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{2,it}^{(9)*} &= O_P\left(\frac{1}{\min\{N,T\}}\right), \\ \frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{2,it}^{(10)*} &= O_P\left(\frac{1}{\min\{N,T\}}\right). \end{aligned} \tag{B.61}$$

We have finished all the derivations for these ten terms in $a_{2,it}^*$ and we are ready to combine the leading terms in them. By (B.50), (B.57), (B.58), (B.59), (B.60) and (B.61), we have

$$\begin{aligned} A_{3Ti} &= \frac{1}{T}\sum_{t=1}^{T}a_{it}^{-1}a_{2,it}^* \\ &= -\Sigma_{u,i}(\widehat{\theta}_i-\theta_{0i})-\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t}) \\ &\quad +\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t}) \\ &\quad +\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2}{[1-G_\varepsilon(z_{it}^0)]^2G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t}) \\ &\quad -\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2}{[1-G_\varepsilon(z_{it}^0)][G_\varepsilon(z_{it}^0)]^2}u_{it}^0\gamma_{0i}'(\widehat{f}_t-f_{0t}) \end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{T}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}(0'_{d_\beta},(\widehat{f}_t-f_{0t})')' \\
&+O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.62)}
\end{aligned}
$$

We then proceed with A_{4Ti}, note that

$$
\begin{aligned}
a_{it}^\dagger - a_{it} &= [1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)\left\{[1-G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it}) - [1-G_\varepsilon(z_{it}^0)][G_\varepsilon(z_{it}^0)]\right\} \\
&= -[1-G_\varepsilon(z_{it}^0)][G_\varepsilon(z_{it}^0)]^2[G(\widehat{z}_{it})-G(z_{it}^0)] \\
&\quad +[1-G_\varepsilon(z_{it}^0)]^2G_\varepsilon(z_{it}^0)[G(\widehat{z}_{it})-G(z_{it}^0)] \\
&\quad -[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)[G(\widehat{z}_{it})-G(z_{it}^0)]^2.
\end{aligned}
$$

Then by the Taylor expansion and Lemma 2.3,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}(a_{it}^{\dagger-1}-a_{it}^{-1})a_{2,it}^{(1)*} \\
=& -\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}a_{it}^{-1}(a_{it}^\dagger-a_{it})a_{2,it}^{(1)*} \\
=& -\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}[G_\varepsilon(\widehat{z}_{it})-G_\varepsilon(z_{it}^0)]^2G_\varepsilon(z_{it}^0)g_\varepsilon(z_{it}^0)u_{it}^0 \\
&+\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}[G_\varepsilon(\widehat{z}_{it})-G_\varepsilon(z_{it}^0)]^2[1-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)u_{it}^0 \\
&-\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}[G_\varepsilon(\widehat{z}_{it})-G_\varepsilon(z_{it}^0)]^3g_\varepsilon(z_{it}^0)u_{it}^0+O_P\left(\frac{1}{\min\{N,T\}}\right) \\
=& -\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}G_\varepsilon(z_{it}^0)[g_\varepsilon(z_{it}^\dagger)]^2g_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it}-z_{it}^0)^2 \\
&+\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}[1-G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^\dagger)]^2g_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it}-z_{it}^0)^2 \\
&-\frac{1}{T}\sum_{t=1}^{T}a_{it}^{\dagger-1}[g_\varepsilon(z_{it}^\dagger)]^3g_\varepsilon(z_{it}^0)u_{it}^0(\widehat{z}_{it}-z_{it}^0)^3+O_P\left(\frac{1}{\min\{N,T\}}\right) \\
=& \,O_P\left(\frac{1}{\min\{N,T\}}\right).
\end{aligned}
$$

Following analogous arguments, we can show that the rest terms in $T^{-1}\sum_{t=1}^{T}(a_{it}^{\dagger-1}-a_{it}^{-1})a_{2,it}^*$ is bounded by probability of the order $O_P(1/\min\{N,T\})$. Therefore,

$$
A_{4Ti} = O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.63)}
$$

Finishing the discussions on A_{3Ti} and A_{4Ti}, we have derived the leading terms in A_{2Ti}, all of which depend on the convergence of $\widehat{f}_t - f_{0t}$. Since A_{1Ti} only contains real values which contributes to the CLT, we leave it for further discussions. By (B.46), (B.47), (B.50) and (B.63), we have

$$
\begin{aligned}
\widehat{\theta}_i - \theta_{0i} \;=\; & \Sigma_{u,i}^{-1} A_{1Ti} - \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
& + \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
& + \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2}{[1-G_\varepsilon(z_{it}^0)]^2G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
& - \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)][g_\varepsilon(z_{it}^0)]^2}{[1-G_\varepsilon(z_{it}^0)][G_\varepsilon(z_{it}^0)]^2}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t}) \\
& + \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}(0_{d_\beta}', (\widehat{f}_t - f_{0t})')' \\
& + O_P\left(\frac{1}{\min\{N,T\}}\right) \\
:= \; & A_{5Ti} + \cdots + A_{10Ti} + O_P\left(\frac{1}{\min\{N,T\}}\right). \qquad \text{(B.64)}
\end{aligned}
$$

We now proceed with the derivation of $\widehat{f}_t - f_{0t}$. From the first order condition in (B.13), we have

$$
\begin{aligned}
0 \;=\; & \frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(\widehat{z}_{it})]g_\varepsilon(\widehat{z}_{it})}{[1-G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it})}\widehat{\gamma}_i \\
= \; & \frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i} \\
& + \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{[y_{it}-G_\varepsilon(\widehat{z}_{it})]g_\varepsilon(\widehat{z}_{it})}{[1-G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it})}\widehat{\gamma}_i - \frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}\right\} \\
:= \; & B_{1Nt} + B_{2Nt}, \qquad \text{(B.65)}
\end{aligned}
$$

where the definitions of B_{1NT} and B_{2NT} are obvious.

For B_{1Nt}, there is no estimated value involved in it. Therefore, we are interested in the leading term of B_{2Nt}. Recall that

$$
\begin{aligned}
a_{it} &= [1-G_\varepsilon(z_{it}^0)]^2[G_\varepsilon(z_{it}^0)]^2, \\
a_{it}^\dagger &= [1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)[1-G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it}).
\end{aligned}
$$

The, for B_{2Nt}, we write

$$
\begin{aligned}
B_{2Nt} \;=\; & \frac{1}{N}\sum_{i=1}^{N}a_{it}^{-1}b_{2,it}^* + \frac{1}{N}\sum_{i=1}^{N}(a_{it}^{\dagger-1} - a_{it}^{-1})b_{2,it}^*, \\
:= \; & B_{3Nt} + B_{4Nt}, \qquad \text{(B.66)}
\end{aligned}
$$

where

$$
\begin{aligned}
b_{2,it}^* = \; & [y_{it}-G_\varepsilon(\widehat{z}_{it})]g_\varepsilon(\widehat{z}_{it})[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)\widehat{\gamma}_i \\
& - [y_{it}-G_\varepsilon(z_{it}^0)]g(z_{it}^0)[1-G_\varepsilon(\widehat{z}_{it})]G_\varepsilon(\widehat{z}_{it})\gamma_{0i}.
\end{aligned}
$$

Following developments similar to the derivations of (B.62) and (B.63), we can show that

$$\begin{aligned}
B_{3Nt} = & -\Sigma_{\gamma,t}(\widehat{f}_t - f_{0t}) - \frac{1}{N}\sum_{i=1}^{N}\frac{g_{\varepsilon}^2(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& +\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)]g_{\varepsilon}^{(1)}(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& +\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)][g_{\varepsilon}(z_{it}^0)]^2}{[1-G_{\varepsilon}(z_{it}^0)]^2G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& -\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)][g_{\varepsilon}(z_{it}^0)]^2}{[1-G_{\varepsilon}(z_{it}^0)][G_{\varepsilon}(z_{it}^0)]^2}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& +\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)]g_{\varepsilon}(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}(\widehat{\gamma}_i - \gamma_{0i}) \\
& +O_P\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned} \tag{B.67}$$

and

$$B_{4Nt} = O_P\left(\frac{1}{\min\{N,T\}}\right). \tag{B.68}$$

Then we are ready to combine the leading terms in $\widehat{f}_t - f_{0t}$. By (B.65), (B.67) and (B.68),

$$\begin{aligned}
\widehat{f}_t - f_{0t} = & \Sigma_{\gamma,t}^{-1}B_{1Nt} - \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{g_{\varepsilon}^2(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& +\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)]g_{\varepsilon}^{(1)}(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& +\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)][g_{\varepsilon}(z_{it}^0)]^2}{[1-G_{\varepsilon}(z_{it}^0)]^2G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)][g_{\varepsilon}(z_{it}^0)]^2}{[1-G_{\varepsilon}(z_{it}^0)][G_{\varepsilon}(z_{it}^0)]^2}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
& +\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^0)]g_{\varepsilon}(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}(\widehat{\gamma}_i - \gamma_{0i}) \\
& +O_P\left(\frac{1}{\min\{N,T\}}\right) \\
:= & B_{5Nt} + \cdots + B_{10Nt} + O_P\left(\frac{1}{\min\{N,T\}}\right).
\end{aligned} \tag{B.69}$$

From (B.69), we show that the leading terms in $\widehat{f}_t - f_{0t}$ can be expressed by $\widehat{\theta}_i - \theta_{0i}$. Recall that the leading terms in $\widehat{\theta}_i - \theta_{0i}$ are related to $\widehat{f}_t - f_{0t}$ as well by (B.64).

Substituting (B.64) into the second term (B_{6Nt}) in (B.69), it becomes

$$\begin{aligned}
B_{6Nt} = & -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{g_{\varepsilon}^2(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_i - \theta_{0i}) \\
= & -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{g_{\varepsilon}^2(z_{it}^0)}{[1-G_{\varepsilon}(z_{it}^0)]G_{\varepsilon}(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}(A_{5Ti} + \cdots + A_{10Ti})
\end{aligned}$$

$$
\begin{aligned}
&+O_P\left(\frac{1}{\min\{N,T\}}\right) \\
&:= B_{6Nt}^{(1)} + \cdots + B_{6Nt}^{(6)} + O_P\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned}
$$

where $A_{5Ti}, \ldots, A_{10Ti}$ are defined in (B.64).

For $B_{6Nt}^{(1)}$,

$$
\begin{aligned}
B_{6Nt}^{(1)} &= -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}A_{5Ti} \\
&= -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}A_{1Ti} \\
&= -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{s=1}^{T}\frac{\left[y_{is}-G_\varepsilon(z_{is}^0)\right]g_\varepsilon(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}u_{is}^0 \\
&= -\frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot\frac{\left[y_{is}-G_\varepsilon(z_{is}^0)\right]g_\varepsilon(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}\cdot\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0.
\end{aligned}
$$

By the Law of iterated expectations and Assumption 3,

$$
\begin{aligned}
E[B_{6Nt}^{(1)}] &= \frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}E\left[E[e_{is}|w_{is}^0,w_{it}^0]\cdot\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot g_\varepsilon(z_{is}^0)\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\right] \\
&= 0. \qquad \text{(B.70)}
\end{aligned}
$$

For its second moment, we first note that $\max_{z_it}\{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)\}^{-1} = O(1)$ by Assumption 1. In addition, since $g_\varepsilon^{(1)}(\cdot)$ is uniformly bounded and by Assumption 3,

$$
\begin{aligned}
&E[\|B_{6Nt}^{(1)}\|^2] \\
\leq\ & O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{l=1}^{T}E\left[|E[e_{is}e_{jl}|w_{it}^0,w_{is}^0,w_{jt}^0,w_{jl}^0]\,\|\gamma_{0i}\|\,\|\gamma_{0j}\|\,\|u_{it}^0\|\,\|u_{jt}^0\|\,\|u_{is}^0\|\,\|u_{jl}^0\|\right] \\
\leq\ & O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{l=1}^{T}\alpha_{ij,sl}E\left[\|\gamma_{0i}\|\,\|\gamma_{0j}\|\,\|u_{it}^0\|\,\|u_{jt}^0\|\,\|u_{is}^0\|\,\|u_{jl}^0\|\right] \\
=\ & O\left(\frac{1}{NT}\right), \qquad \text{(B.71)}
\end{aligned}
$$

where we can show the last equality holds by using Cauchy-Schwarz inequality successively:

$$
\begin{aligned}
&E\left[\|\gamma_{0i}\|\,\|\gamma_{0j}\|\,\|u_{it}^0\|\,\|u_{jt}^0\|\,\|u_{is}^0\|\,\|u_{jl}^0\|\right] \\
\leq\ & E\left[\|\gamma_{0i}\|^2\,\|u_{it}^0\|^2\,\|u_{jt}^0\|^2\right]^{\frac{1}{2}}\cdot E\left[\|\gamma_{0j}\|^2\,\|u_{is}^0\|^2\,\|u_{jl}^0\|^2\right]^{\frac{1}{2}} \\
\leq\ & E\left[\|\gamma_{0i}\|^4\,\|u_{it}^0\|^2\right]^{\frac{1}{4}}\cdot E\left[\|u_{it}^0\|^2\,\|u_{jt}^0\|^4\right]^{\frac{1}{4}}\cdot E\left[\|\gamma_{0j}\|^4\,\|u_{is}^0\|^2\right]^{\frac{1}{4}}\cdot E\left[\|u_{is}^0\|^2\,\|u_{jl}^0\|^4\right]^{\frac{1}{4}} \\
\leq\ & E\left[\|\gamma_{0i}\|^{4+\delta}\right]^{\frac{1}{4+\delta}}\cdot E\left[\|\gamma_{0j}\|^{4+\delta}\right]^{\frac{1}{4+\delta}}\cdot E\left[\|u_{it}^0\|^{\frac{2(4+\delta)}{\delta}}\right]^{\frac{\delta}{2(4+\delta)}}\cdot E\left[\|u_{jt}^0\|^{4+\delta}\right]^{\frac{1}{4+\delta}}
\end{aligned}
$$

$$
\begin{aligned}
&\times E\left[\left\|u_{is}^0\right\|^{\frac{2(4+\delta)}{\delta}}\right]^{\frac{\delta}{2(4+\delta)}} \cdot E\left[\left\|u_{jl}^0\right\|^{4+\delta}\right]^{\frac{1}{4+\delta}} \\
=\ & O(1).
\end{aligned}
$$

By (B.70) and (B.71),

$$
B_{6Nt}^{(1)} = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{B.72}
$$

After having proved that $B_{6Nt}^{(1)}$ is bounded in probability of the order $O_p\left(\frac{1}{\sqrt{NT}}\right)$, we proceed with $B_{6Nt}^{(2)}$. For $B_{6Nt}^{(2)}$,

$$
\begin{aligned}
& B_{6Nt}^{(2)} \\
=\ & -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}A_{6Ti} \\
=\ & \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{s=1}^T \frac{g_\varepsilon^2(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}u_{is}^0\gamma_{0i}'(\widehat{f}_s - f_{0s}) \\
=\ & \frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N\sum_{s=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\frac{g_\varepsilon^2(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\gamma_{0i}'(\widehat{f}_s - f_{0s}) \\
:=\ & \frac{1}{T}\sum_{s=1}^T \Delta_{f,ts}(\widehat{f}_s - f_{0s}),
\end{aligned} \tag{B.73}
$$

where

$$
\Delta_{f,ts} = \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot\frac{g_\varepsilon^2(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}\cdot\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\gamma_{0i}'.
$$

For $B_{6Nt}^{(3)}$,

$$
\begin{aligned}
& B_{6Nt}^{(3)} \\
=& -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}A_{7Ti} \\
=& -\frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}\frac{1}{T}\Sigma_{u,i}^{-1}\sum_{s=1}^T \frac{[y_{is}-G_\varepsilon(z_{is}^0)]g_\varepsilon^{(1)}(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}u_{is}^0\gamma_{0i}'(\widehat{f}_s - f_{0s}) \\
=& -\frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^N\sum_{s=1}^T \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\frac{[y_{is}-G_\varepsilon(z_{is}^0)]g_\varepsilon^{(1)}(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\gamma_{0i}'(\widehat{f}_s - f_{0s})
\end{aligned}
$$

Analogously to (B.54) and by Assumption 3, we can show that

$$
\begin{aligned}
& \sum_{s=1}^T\left\|\sum_{i=1}^N \frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot\frac{[y_{is}-G_\varepsilon(z_{is}^0)]g_\varepsilon^{(1)}(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}\cdot\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\gamma_{0i}'\right\|^2 \\
=\ & O_P\left(NT\right).
\end{aligned} \tag{B.74}
$$

By (B.74), Lemma 2.3 and Cauchy-Schwarz inequality

$$
\begin{aligned}
&\left\|B_{6Nt}^{(3)}\right\| \\
&\leq O(1)\frac{1}{NT}\left\{\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\frac{g_{\varepsilon}^{2}(z_{it}^{0})}{[1-G_{\varepsilon}(z_{it}^{0})]G_{\varepsilon}(z_{it}^{0})}\frac{[y_{is}-G_{\varepsilon}(z_{is}^{0})]g_{\varepsilon}^{(1)}(z_{is}^{0})}{[1-G_{\varepsilon}(z_{is}^{0})]G_{\varepsilon}(z_{is}^{0})}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^{0}\gamma_{0i}^{\prime}\right\|^{2}\right\}^{\frac{1}{2}} \\
&\quad\cdot\left\{\sum_{s=1}^{T}\|\widehat{f}_{s}-f_{0s}\|^{2}\right\}^{\frac{1}{2}} \\
&= O_{P}\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right). \qquad\qquad (B.75)
\end{aligned}
$$

Analogously to (B.75), we have

$$
\begin{aligned}
B_{6Nt}^{(4)} &= O_{P}\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right), \\
B_{6Nt}^{(5)} &= O_{P}\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right), \\
B_{6Nt}^{(6)} &= O_{P}\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right). \qquad\qquad (B.76)
\end{aligned}
$$

Then we are ready to write the leading terms in B_{6Nt}. By (B.72), (B.73), (B.75), and (B.76)),

$$
B_{6Nt}=\frac{1}{T}\sum_{s=1}^{T}\Delta_{f,ts}(\widehat{f}_{s}-f_{0s})+O_{P}\left(\frac{1}{\min\{N,T\}}\right). \qquad (B.77)
$$

Now we proceed with B_{7Nt}. Substituting (B.64) into the third term (B_{7Nt}) in (B.69), it becomes

$$
\begin{aligned}
B_{7Nt} &= \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^{0})]g_{\varepsilon}^{(1)}(z_{it}^{0})}{[1-G_{\varepsilon}(z_{it}^{0})]G_{\varepsilon}(z_{it}^{0})}\gamma_{0i}u_{it}^{0\prime}(\widehat{\theta}_{i}-\theta_{0i}) \\
&= \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^{0})]g_{\varepsilon}^{(1)}(z_{it}^{0})}{[1-G_{\varepsilon}(z_{it}^{0})]G_{\varepsilon}(z_{it}^{0})}\gamma_{0i}u_{it}^{0\prime}(A_{5Ti}+\cdots+A_{10Ti}) \\
&\quad +O_{P}\left(\frac{1}{\min\{N,T\}}\right) \\
&:= B_{7Nt}^{(1)}+\cdots+B_{7Nt}^{(6)}+O_{P}\left(\frac{1}{\min\{N,T\}}\right),
\end{aligned}
$$

where $A_{5Ti},\ldots,A_{10Ti}$ are defined in (B.64).

For $B_{7Nt}^{(1)}$, we have

$$
\begin{aligned}
B_{7Nt}^{(1)} &= \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^{0})]g_{\varepsilon}^{(1)}(z_{it}^{0})}{[1-G_{\varepsilon}(z_{it}^{0})]G_{\varepsilon}(z_{it}^{0})}\gamma_{0i}u_{it}^{0\prime}A_{5Ti} \\
&= \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_{\varepsilon}(z_{it}^{0})]g_{\varepsilon}^{(1)}(z_{it}^{0})}{[1-G_{\varepsilon}(z_{it}^{0})]G_{\varepsilon}(z_{it}^{0})}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}\cdot\frac{1}{T}\sum_{s=1}^{T}\frac{\left[y_{is}-G_{\varepsilon}(z_{is}^{0})\right]g_{\varepsilon}(z_{is}^{0})}{[1-G_{\varepsilon}(z_{is}^{0})]G_{\varepsilon}(z_{is}^{0})}u_{is}^{0} \\
&= \frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}\left\{\frac{[y_{it}-G_{\varepsilon}(z_{it}^{0})]g_{\varepsilon}^{(1)}(z_{it}^{0})}{[1-G_{\varepsilon}(z_{it}^{0})]G_{\varepsilon}(z_{it}^{0})}\right\}\left\{\frac{\left[y_{is}-G_{\varepsilon}(z_{is}^{0})\right]g_{\varepsilon}(z_{is}^{0})}{[1-G_{\varepsilon}(z_{is}^{0})]G_{\varepsilon}(z_{is}^{0})}\right\}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^{0}.
\end{aligned}
$$

We check the first and second moments of $B_{7Nt}^{(1)}$ to show its convergence. By the law of iterated expectations and Assumption 3:

$$
\begin{aligned}
\|E[B_{7Nt}^{(1)}]\| &= \frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}E[E[e_{it}e_{is}|w_{it}^0,w_{is}^0]g_\varepsilon^{(1)}(z_{it}^0)g_\varepsilon(z_{is}^0)\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0] \\
&\le O(1)\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}E[|E[e_{it}e_{is}|w_{it}^0,w_{is}^0]|\,\|\gamma_{0i}\|\,\|u_{it}^0\|\,\|u_{is}^0\|] \\
&\le O(1)\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\alpha_{ii,ts}E\left[\|\gamma_{0i}\|\,\|u_{it}^0\|\,\|u_{is}^0\|\right] \\
&= O\left(\frac{1}{T}\right),
\end{aligned} \tag{B.78}
$$

where the last equality holds because $E\left[\|\gamma_{0i}\|\,\|u_{it}^0\|\,\|u_{is}^0\|\right] < \infty$ which can be proved by Cauchy-Schwarz inequality and Assumption 3. Additionally, we use the fact $\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{T}\alpha_{ii,ts} < \infty$ for fixed t which is directly implied by the conditions in Assumption 3. It can be justified easily when we consider independent samples or α-mixing process. For its second moment, by Assumption 4,

$$
\begin{aligned}
&E\left[\left\|B_{7Nt}^{(1)}\right\|^2\right] \\
\le\ & O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{l=1}^{T}E[E[e_{it}e_{is}e_{jt}e_{jl}|w_{it}^0w_{is}^0w_{jt}^0w_{jl}^0]\,\|\gamma_{0i}\|\,\|\gamma_{0j}\|\,\|u_{it}^0\|\,\|u_{jt}^0\|\,\|u_{is}^0\|\,\|u_{jl}^0\|] \\
\le\ & O(1)\frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{l=1}^{T}\alpha_{ijij,ttsl}E[\|\gamma_{0i}\|\,\|\gamma_{0j}\|\,\|u_{it}^0\|\,\|u_{jt}^0\|\,\|u_{is}^0\|\,\|u_{jl}^0\|] \\
=\ & O_P\left(\frac{1}{T\min\{N,T\}}\right).
\end{aligned} \tag{B.79}
$$

By (B.78) and (B.79),

$$
B_{7Nt}^{(1)} = O_P\left(\frac{1}{\sqrt{T}\min\{\sqrt{N},\sqrt{T}\}}\right). \tag{B.80}
$$

For $B_{7Nt}^{(2)}$, note that

$$
\begin{aligned}
B_{7Nt}^{(2)} &= \frac{1}{N}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\gamma_{0i}u_{it}^{0\prime}A_{6Ti} \\
&= -\frac{1}{NT}\Sigma_{\gamma,t}^{-1}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\frac{g_\varepsilon^2(z_{is}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{is}^0)}\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\gamma_{0i}'(\widehat{f}_s-f_{0s}).
\end{aligned}
$$

By Lemma 2.3, Assumption 3 and Cauchy-Schwarz inequality, write

$$
\begin{aligned}
\|B_{7Nt}^{(2)}\| &\le O(1)\frac{1}{NT}\left\{\sum_{s=1}^{T}\|\widehat{f}_s-f_{0s}\|^2\right\}^{\frac{1}{2}} \\
&\quad\cdot\left\{\sum_{s=1}^{T}\left\|\sum_{i=1}^{N}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot\frac{g_\varepsilon^2(z_{is}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{is}^0)}\cdot\gamma_{0i}u_{it}^{0\prime}\Sigma_{u,i}^{-1}u_{is}^0\gamma_{0i}'\right\|^2\right\}^{\frac{1}{2}} \\
&= O_P\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right).
\end{aligned} \tag{B.81}
$$

Analogously, we can show that

$$
\begin{aligned}
B_{7Nt}^{(3)} &= O_P\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right),\\
B_{7Nt}^{(4)} &= O_P\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right),\\
B_{7Nt}^{(5)} &= O_P\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right),\\
B_{7Nt}^{(6)} &= O_P\left(\frac{1}{\sqrt{N}\min\{\sqrt{N},\sqrt{T}\}}\right).
\end{aligned} \tag{B.82}
$$

We have showed the rates of convergence for all the terms in B_{7Nt}. By (B.80), (B.81) and (B.82),

$$
\|B_{7Nt}\| = O_P\left(\frac{1}{\min\{N,T\}}\right). \tag{B.83}
$$

Analogously to (B.83), we can show the convergence of B_{8Nt}, B_{9Nt} and B_{10Nt}. We have

$$
\begin{aligned}
\|B_{8Nt}\| &= O_P\left(\frac{1}{\min\{N,T\}}\right),\\
\|B_{9Nt}\| &= O_P\left(\frac{1}{\min\{N,T\}}\right),\\
\|B_{10Nt}\| &= O_P\left(\frac{1}{\min\{N,T\}}\right).
\end{aligned} \tag{B.84}
$$

Having derived the leading terms in B_{6Nt} and shown B_{7Nt}, B_{8Nt}, B_{9Nt}, B_{10Nt} are bounded in probability of the order $O_P\left(\frac{1}{\min\{N,T\}}\right)$, we are ready to write the leading terms in $\widehat{f}_t - f_{0t}$. By (B.77), (B.83), and (B.84),

$$
\widehat{f}_t - f_{0t} = \Sigma_{\gamma,t}^{-1}B_{1Nt} + \frac{1}{T}\sum_{s=1}^{T}\Delta_{f,ts}(\widehat{f}_s - f_{0s}) + O_P\left(\frac{1}{\min\{N,T\}}\right). \tag{B.85}
$$

Following the argument which is analogous to that for (B.72),

$$
\begin{aligned}
\frac{1}{T}\sum_{s=1}^{T}\Delta_{f,ts}\Sigma_{\gamma,t}^{-1}B_{1Ns} &= \frac{1}{T}\sum_{s=1}^{T}\Delta_{f,ts}\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{[y_{is}-G_\varepsilon(z_{is}^0)]g_\varepsilon(z_{is}^0)}{[1-G_\varepsilon(z_{is}^0)]G_\varepsilon(z_{is}^0)}\gamma_{0i}\\
&= O_P\left(\frac{1}{\sqrt{NT}}\right).
\end{aligned}
$$

Following a development similar to those for Equation (41) in the supplement document of Ando and Lu (2020), we can show that

$$
\frac{1}{T}\sum_{s=1}^{T}\Delta_{f,ts}(\widehat{f}_s - f_{0s}) = O_P\left(\frac{1}{\sqrt{NT}}\right).
$$

Therefore,

$$
\widehat{f}_t - f_{0t} = \Sigma_{\gamma,t}^{-1}B_{1Nt} + O_P\left(\frac{1}{\min\{N,T\}}\right). \tag{B.86}
$$

In B.86, we show that the leading term in $\widehat{f}_t - f_{0t}$ is given by $\Sigma_{\gamma,t}^{-1} B_{1Nt}$ which will generate the central limit theory. The other terms in $\widehat{f}_t - f_{0t}$ have been shown to be bounded in probability of the order $O_P\left(\frac{1}{\min\{N,T\}}\right)$.

We now proceed with $\widehat{\theta}_i - \theta_{0i}$. Recall we have A_{5Ti}, A_{6Ti}, ... ,A_{10Ti} as the leading terms in $\widehat{\theta}_i - \theta_{0i}$ as shown in (B.64). Substituting (B.86) into the second term (A_{6Ti}) in (B.64), it becomes

$$\begin{aligned} A_{6Ti} &= -\frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t}) \\ &= -\frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'\Sigma_{\gamma,t}^{-1}B_{1Nt} + O_P\left(\frac{1}{\min\{N,T\}}\right). \end{aligned}$$

Then we have

$$\begin{aligned} A_{6Ti} &= -\frac{1}{T}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{[y_{jt}-G_\varepsilon(z_{jt}^0)]g_\varepsilon(z_{jt}^0)}{[1-G_\varepsilon(z_{jt}^0)]G_\varepsilon(z_{jt}^0)}\gamma_{0j} \\ &\quad + O_P\left(\frac{1}{\min\{N,T\}}\right) \\ &= -\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\frac{g_\varepsilon^2(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot\frac{[y_{jt}-G_\varepsilon(z_{jt}^0)]g_\varepsilon(z_{jt}^0)}{[1-G_\varepsilon(z_{jt}^0)]G_\varepsilon(z_{jt}^0)}\cdot u_{it}^0\gamma_{0i}'\Sigma_{\gamma,t}^{-1}\gamma_{0j} \\ &\quad + O_P\left(\frac{1}{\min\{N,T\}}\right) \\ &= O_P\left(\frac{1}{\sqrt{NT}}\right) + O_P\left(\frac{1}{\min\{N,T\}}\right) \\ &= O_P\left(\frac{1}{\min\{N,T\}}\right), \end{aligned} \tag{B.87}$$

where the second last equality holds analogously to (B.72) using Assumption 3.

Substituting (B.86) into the third term (A_{7Ti}) in (B.64), it becomes

$$\begin{aligned} A_{7Ti} &= \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'(\widehat{f}_t - f_{0t}) \\ &= \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'\Sigma_{\gamma,t}^{-1}B_{1Nt} + O_P\left(\frac{1}{\min\{N,T\}}\right). \end{aligned}$$

Then we have

$$\begin{aligned} A_{7Ti} &= \frac{1}{T}\Sigma_{u,i}^{-1}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}u_{it}^0\gamma_{0i}'\Sigma_{\gamma,t}^{-1}\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{[y_{jt}-G_\varepsilon(z_{jt}^0)]g_\varepsilon(z_{jt}^0)}{[1-G_\varepsilon(z_{jt}^0)]G_\varepsilon(z_{jt}^0)}\gamma_{0j} \\ &\quad + O_P\left(\frac{1}{\min\{N,T\}}\right) \\ &= \frac{1}{NT}\Sigma_{u,i}^{-1}\sum_{j=1}^{N}\sum_{t=1}^{T}\frac{[y_{it}-G_\varepsilon(z_{it}^0)]g_\varepsilon^{(1)}(z_{it}^0)}{[1-G_\varepsilon(z_{it}^0)]G_\varepsilon(z_{it}^0)}\cdot\frac{[y_{jt}-G_\varepsilon(z_{jt}^0)]g_\varepsilon(z_{jt}^0)}{[1-G_\varepsilon(z_{jt}^0)]G_\varepsilon(z_{jt}^0)}\cdot u_{it}^0\gamma_{0i}'\Sigma_{\gamma,t}^{-1}\gamma_{0j} \\ &\quad + O_P\left(\frac{1}{\min\{N,T\}}\right) \\ &= O_P\left(\frac{1}{\sqrt{T}\min\{\sqrt{N},\sqrt{T}\}}\right) + O_P\left(\frac{1}{\min\{N,T\}}\right) \end{aligned}$$

$$= O_P\left(\frac{1}{\min\{N,T\}}\right), \tag{B.88}$$

where the second last equality can be proved by arguments similar to those for (B.80).

Analogously to (B.88), we can show that

$$\begin{aligned} A_{8Ti} &= O_P\left(\frac{1}{\min\{N,T\}}\right), \\ A_{9Ti} &= O_P\left(\frac{1}{\min\{N,T\}}\right), \\ A_{10Ti} &= O_P\left(\frac{1}{\min\{N,T\}}\right). \end{aligned} \tag{B.89}$$

We have shown that A_{6Ti}, A_{7Ti}, ..., A_{10Ti} are all bounded in probability of the order $O_P\left(\frac{1}{\min\{N,T\}}\right)$. We are ready to write the leading terms in $\widehat{\theta}_i - \theta_{0i}$. By (B.87), (B.88) and (B.89),

$$\widehat{\theta}_i - \theta_{0i} = \Sigma_{u,i}^{-1} A_{1Ti} + O_P\left(\frac{1}{\min\{N,T\}}\right). \tag{B.90}$$

By (B.86), (B.90) and the conditions in the body of this theorem, the proof is complete. ■