PHBS WORKING PAPER SERIES

Two-Step Estimation of Quantile Panel Data Models with Interactive Fixed Effects

Liang Chen
Peking University

January 2021

Working Paper 20210103

Abstract

This paper considers the estimation of panel data models with interactive fixed effects where the idiosyncratic errors are subject to conditional quantile restrictions. I propose a two-step estimator for the coeffcient of the observed regressors that is easy to implement in practice. In the first step, the principal component analysis is applied to the cross-sectional averages of the regressors to estimate the latent factors. In the second step, the smoothed quantile regression is used to estimate the coeffcient of the observed regressors and the factor loadings jointly. The consistency and asymptotic normality of the estimator are established under large N,T asymptotics. It is found that the asymptotic distribution of the estimator suffers from asymptotic biases, and I show how to correct the biases using both analytical and split-panel jackknife bias corrections. Simulation studies confirm that the proposed estimator performs well with moderate sample sizes.

Keywords: Panel data, interactive fixed effects, smoothed quantile regressions, principal compo-nent analysis, analytical bias correction, split-panel jackknife.
JEL Classification: C14, C31, C33

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

Two-Step Estimation of Quantile Panel Data Models with Interactive Fixed Effects

Liang Chen*

Peking University HSBC Business School

January 5, 2021

Abstract

This paper considers the estimation of panel data models with interactive fixed effects where the idiosyncratic errors are subject to conditional quantile restrictions. I propose a two-step estimator for the coefficient of the observed regressors that is easy to implement in practice. In the first step, the principal component analysis is applied to the cross-sectional averages of the regressors to estimate the latent factors. In the second step, the smoothed quantile regression is used to estimate the coefficient of the observed regressors and the factor loadings jointly. The consistency and asymptotic normality of the estimator are established under large N, T asymptotics. It is found that the asymptotic distribution of the estimator suffers from asymptotic biases, and I show how to correct the biases using both analytical and split-panel jackknife bias corrections. Simulation studies confirm that the proposed estimator performs well with moderate sample sizes.

Keywords: panel data, interactive fixed effects, smoothed quantile regressions, principal component analysis, analytical bias correction, split-panel jackknife.

JEL codes: C14, C31, C33.

*No.2199 Lishui Road, Shenzhen, Guangdong 518055, China. Email: chenliang@phbs.pku.edu.cn. I am indebted to Liangjun Su and three anonymous referees for their constructive inputs which have greatly improved the paper. Financial support from the National Natural Science Foundation of China (Grant No.71703089) is gratefully acknowledged.

1 Introduction

This paper considers panel data models with interactive fixed effects, where the unobserved errors has a latent factor model structure. The assumption of interactive fixed effects has been adopted in a lot of recent studies — see Pesaran (2006), Bai (2009), Moon and Weidner (2015), and Lu and Su (2016) among many others. This assumption is general enough to nest the standard panel data models with only individual fixed effects and models with additive individual and time effects. It also allows the unobserved factors (or common shocks) to affect the dependent variables with different intensities that are measured by the individual-specific factor loadings. Moreover, the latent factor structure has become an important tool to characterize cross-sectional dependence in panel data models — see Chudik and Pesaran (2015) for an excellent review. Yet, most of the existing studies focus on linear models where the idiosyncratic errors are subject to conditional mean restrictions, and the main object of interest is the coefficient that represents the partial effect of the regressors on the conditional mean of the dependent variable. In this paper, I consider panel data models with interactive effects where the conditional mean restrictions are replaced by conditional quantile restrictions. In such models, the coefficient of the regresors measures the partial quantile effect, providing a more complete picture of how the regressors affect the distributions of the dependent variables.

In this paper I adopt the popular common correlated effects (CCE, hereafter) framework pioneered by Pesaran (2006). In this framework, the regressors are assumed to be driven by the same latent factors that affect the dependent variables, allowing the the space of the common factors to be approximated by the cross-sectional averages of the observed variables. Compared with the approach that estimates the coefficient and fixed effects jointly, the CCE approach has two main advantages that are particular valuable for the quantile panel models: first, the computation of the CCE estimator is easy, because given the estimated factors, the coefficient of the regressors and the factor loadings can be simply estimated by treating the estimated factors as known. Second, the asymptotic properties of the estimators are much easier to derive since the estimated factors have a relatively simple expansion.

Like the CCE estimator, the proposed estimation method in this paper contains two steps. However, both of the steps differ from the standard CCE method that is widely used the for linear and quantile panel data models in existing studies. In the first step, to avoid the *degenerated-regressors* problem of the standard CCE method (see Karabiyik et al. 2017 and Remark 1 below), I apply the principal component analysis (PCA, hereafter) to the cross-sectional averages of the regressors to estimate the common factors. In the second step, inspired by Galvao and Kato (2016), the smoothed quantile regression (SQR, hereafter) instead of the standard quantile regression is used to estimate the coefficient of the regressors and the factor loadings jointly, treating the estimated factors from the first step as given. The main motivation of making such modifications in both steps of the standard CCE estimator is to facilitate the asymptotic analysis of the proposed estimator.

In the "large N, small T" framework[1], the identification and estimation of quantile panel data models are very challenging even when there are only individual effects (see Arellano and Bonhomme 2016 and Graham et al. 2018 for example). When there are interactive effects in quantile panel models, there remains the open question of whether the parameter of interest can be point identified (see Chen 2015 for a result of set identification). Thus, in this paper, I follow Fernández-Val and Weidner (2016) and Chen et al. (2020) and consider the "large N, large T" framework where the realizations of the factors and factor loadings are treated as non-random fixed parameters, and the main contribution of this paper is that I establish the asymptotic properties of the proposed estimator in this context. In particular, under some regularity conditions, I show that the proposed two-step estimator for the coefficient of the regressors is $\sqrt{NT}$-consistent, asymptotically normally distributed, and it has a leading bias of order $T^{-1} + N^{-1}$. More importantly, I derive the analytical expression of the leading bias term, providing the basis of analytical bias-correction and a heuristic justification for the use of the split-panel jackknife (SPJ, hereafter) bias correction in practice. The Bahadur representation of my two-step estimator extends the similar representations of the estimators for linear panel data models (see Bai 2009) and nonlinear panel data models with smooth object functions (see Chen et al. 2020) to quantile panel models. To the best of my knowledge, this is the first result of this kind in the literature.

Related Literature

This paper paper is related to the large and growing literature on quantile regressions for panel models. Abrevaya and Dahl (2008), Rosen (2012), Arellano and Bonhomme (2016), Graham et al. (2018), and Cai et al. (2018) considered identification and estimation of quantile effects with fixed T. In the large T framework, Canay (2011) and Chen and Huo (2020) proposed two-step estimation methods, Koenker (2004), Lamarche (2010) and Galvao and Montes-Rojas (2010) proposed penalized quantile regressions for panel models, Galvao (2011) considered quantile regressions of dynamic panels, Kato et al. (2012), Galvao and Kato (2016) and Galvao et al. (2020) focused on the asymptotic distributions of quantile regressions and smoothed quantile regressions, Galvao et al. (2013) studied censored quantile regressions for panel data, Yoon and Galvao (2020) considered the robust estimation of the covariance matrix, Chen (2019) studied the nonparametric estimation of quantile panel models.

All the studies mentioned above only considered models with individual effects. Quantile panel models with interactive fixed effects were first studied by Harding and Lamarche (2014), and more recently by Chen et al. (2019), Belloni et al. (2019), Feng (2019), Harding et al. (2020),

[1] Throughout this paper, I use N and T to denote the numbers of cross-sectional and time-series observations respectively.

Ando and Bai (2020), Ma et al. (2020).

As in this paper, Harding and Lamarche (2014) and Harding et al. (2020) also adopted the CCE framework. However, unlike my two-step estimator, they proposed to use the standard CCE estimator where the cross-sectional averages of the regressors and the dependent variables are used as the proxies of the unobserved factors, and in the second step they use standard quantile regressions instead of SQR to estimate the coefficient of the regressors. More importantly, their asymptotic results are quite different in nature from the main conclusion of this paper. In particular, Harding and Lamarche (2014) showed that the CCE estimator has no asymptotic bias, while Harding et al. (2020) proved that the CCE estimator of the common slope parameter is $\sqrt{NT}$-consistent, with a leading bias term of approximate order $T^{-3/4}$, but they didn't give the analytical expression of the bias[2].

Chen et al. (2019) and Ando and Bai (2020) both proposed iterative procedure to estimate the quantile factor and factor loadings jointly. Chen et al. (2019)'s model has no observed regressors and they mainly focused on the asymptotic properties of the estimated factors and factor loadings. The model of Ando and Bai (2020) contains observed regressors but they assumed the coefficients to be heterogenous across individuals. As a consequence, their estimators of the heterogenous coefficients converges at the rate of $\sqrt{N}$ and are free of asymptotic biases. Moreover, Ando and Bai (2020)'s asymptotic analysis requires all the finite moments of the idiosyncratic errors to be bounded, while in this paper I only need the density functions of the idiosyncratic errors to exist and to be sufficiently smooth (i.e., continuously differentiable). Ma et al. (2020) considered a model that is similar to the quantile factor models of Chen et al. (2019) except that they assumed the factor loadings to be smooth functions of observed (and time-invariant) individual characteristics.

One potential problem of the methods proposed by Chen et al. (2019) and Ando and Bai (2020) is that their computational algorithm does not necessarily converge to the global minimum because their object function is not convex. To solve this problem, Belloni et al. (2019) and Feng (2019) added to the object function a nuclear-norm penalty term that is widely used in the matrix completion literature, resulting in a new object function that is convex in the parameters. However, the convergence rates of their estimators are much slower than $\sqrt{NT}$ in general due to the regularization bias, and the asymptotic distributions of their estimators are not derived.

Structure of the Paper

The rest of the paper is organized as follows. Section 2 introduces the model and the new two-step estimator. Second 3 establishes the consistency and the asymptotic distribution of the estimator, and discusses how to correct the asymptotic bias, how to estimate the asymptotic

[2] Harding et al. (2020) also considered heterogenous slopes and showed that the CCE estimators are $\sqrt{N}$-consistent.

covariance matrix, and how to choose the tuning parameters in practice. Monte Carlo simulations are used to evaluate the finite sample performance of the proposed estimator and the effectiveness of the alternative bias-correction methods. Finally, Section 5 concludes. The proofs of all the theorems are relegated to the online appendix.

Notations

Through out the paper, $\mathsf{Q}_Y[\tau|X = x]$ denotes the conditional $\tau-$quantile of Y of given $X = x$, $\|A\|$ denotes the Frobenius norm of matrix A, and $\mathsf{Tr}(\cdot)$ denotes the trace of a square matrix. For two sequences of non-decreasing real numbers $\{a_j\}$ and $\{b_j\}$, $a_j \asymp b_j$ means that there exists $0 < c_1 < c_2 < \infty$ such that $c_1 < a_j/b_j < c_2$ for all large j.

2 The Model and The Estimator

2.1 The Model

For some $\tau \in (0, 1)$, consider the model:

$$Y_{it} = \beta_0(\tau)' X_{it} + \lambda_i(\tau)' f_t + u_{it} \quad \text{for } i = 1, \ldots, N; t = 1, \ldots, T, \tag{1}$$

where $(Y_{it}, X_{it}) \in \mathbb{R} \times \mathbb{R}^p$ is the vector of observed variables for individual i at time t, $\lambda_i(\tau) \in \mathbb{R}^r$ and $f_t \in \mathbb{R}^r$ are the unobserved factor loadings (or individual effects) and common factors (or time effects), respectively. The idiosyncratic error u_{it} is assumed to satisfy the following conditional quantile restriction almost surely:

$$\mathsf{Q}_{u_{it}}[\tau|X_{it}, \lambda_i(\tau), f_t] = 0. \tag{2}$$

Given the above restriction, we have $\mathsf{Q}_{Y_{it}}[\tau|X_{it}, \lambda_i(\tau), f_t] = \beta_0(\tau)' X_{it} + \lambda_i(\tau)' f_t$. Thus, our main object of interest is $\beta_0(\tau)$, i.e., the marginal quantile effect of the regressors X_{it} conditional on the factors and factor loadings.

In addition, following the literature on common correlated effects (CCE) estimation of panel data models (see Pesaran 2006 and Karabiyik et al. 2017), I assume that the regressors are driven by the common factors f_t, i.e., the dynamics of X_{it} is captured by the following factor model structure:

$$X_{it} = \Gamma_i f_t + e_{it}, \quad \text{for } i = 1, \ldots, N; t = 1, \ldots, T, \tag{3}$$

where $\Gamma_i \in \mathbb{R}^{p \times r}$ is a matrix of constants, and $e_{it} \in \mathbb{R}^p$ is a vector of random errors.

The main reason of adopting the CCE framework in this paper is that it allows us to estimate $\beta_0(\tau)$ in a simple two-step procedure that will be defined below. The benefits of employing the two-step estimation approach are twofold: first, under some standard assumptions, the factors can be consistently estimated in the first step using the regressors, which greatly simplifies the asymptotic analysis of the estimator in the second step; second, the low computational cost of the two-step estimator makes it appealing to empirical researchers.

In comparison, in an alternative framework where the relationship between the regressors and the factors is left unspecified (such as Bai 2009 and Ando and Bai 2020), the coefficients for the regressors, the factors and the factor loadings are usually estimated jointly. One the one hand, such "joint estimators" are computationally intensive since they involve iterations between the factors and the factor loadings. On the other hand, the asymptotic properties of such "joint estimators" are much more difficult to establish in the context of quantile regressions. Ando and Bai (2020) consider heterogeneous panels where the estimator of each individual's coefficient converges at $\sqrt{N}$ rate. In a homogeneous panel, the estimator for the coefficient converges at the much fast $\sqrt{NT}$ rate, making it much more challenging to derive the asymptotic distribution of the estimator because many higher order terms that have been ignored in Ando and Bai (2020)'s analysis will become relevant. Moreover, note that the asymptotic analysis of Chen et al. (2020) for nonlinear panel data models with interactive fixed effects, which is already very involved, does not apply to these "joint estimators" since the parameters in the quantile models are defined through non-smooth moment conditions.[3]

2.2 The Two-Step Estimator

For the moment, assume that the number of factors r is known (Section 3.1 below discusses how to consistently estimate r) and that $p \geq r$. Define $\bar{X}_t = N^{-1}\sum_{i=1}^{N} X_{it}$, and $\hat{\Sigma}_{\bar{X}} = T^{-1}\sum_{t=1}^{T} \bar{X}_t \bar{X}_t'$. Moreover, let $K(z) = 1 - \int_{-1}^{z} k(x)dx$, where $k(\cdot)$ is a symmetric continuous kernel function with support $[-1, 1]$ and h is a bandwidth parameter. Then the two-step estimator of $\beta_0(\tau)$ is defined as follows:

Step 1: $\hat{f}_t = \hat{\Psi}'\bar{X}_t$, where $\hat{\Psi} \in \mathbb{R}^{p\times r}$ is the matrix of eigenvectors associated with largest r eigenvalues of $\hat{\Sigma}_{\bar{X}}$.

Step 2: $\hat{\beta}(\tau)$ is defined as:

$$[\hat{\beta}(\tau), \hat{\Lambda}(\tau)] = \underset{\beta\in\mathcal{B},\lambda_i\in\mathcal{A}}{\arg\min} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tau - K\left(\frac{Y_{it} - \beta' X_{it} - \lambda_i'\hat{f}_t}{h}\right)\right](Y_{it} - \beta' X_{it} - \lambda_i'\hat{f}_t), \quad (4)$$

where $\hat{\Lambda}(\tau) = [\hat{\lambda}_1(\tau), \ldots, \hat{\lambda}_N(\tau)]'$.

Define $l(u) = (\tau - K(u/h))u$ and $L(\beta, \Lambda) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} l(Y_{it} - \beta' X_{it} - \lambda_i'\hat{f}_t)$, Step

[3]One can smooth the object function in quantile regressions like I do in this paper, but some important assumptions of Chen et al. (2020) (such as the boundness of the derivatives of the object function) can not be satisfied by the smoothed check function.

2 of the estimation procedure can be effectively solved by the gradient descent algorithm as follows:

Step 2.1: Choose the initial value of the parameter: $(\beta^{(0)}, \Lambda^{(0)})$.

Step 2.2: For $j = 0$, set $s_j = 1$; for $j \geq 1$, define $L_j^{\beta} = -(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} l^{(1)}(Y_{it} - \beta^{(j)\prime} X_{it} - \lambda_i^{(j)\prime} \hat{f}_t) X_{it}$, $L_j^{\lambda_i} = -(NT)^{-1} \sum_{t=1}^{T} l^{(1)}(Y_{it} - \beta^{(j)\prime} X_{it} - \lambda_i^{(j)\prime} \hat{f}_t) \hat{f}_t$, where $l^{(1)}(u) = \partial l(u)/\partial u$. Set[4]

$$s_j = \frac{\left| (\beta^{(j)} - \beta^{(j-1)})'(L_j^{\beta} - L_{j-1}^{\beta}) + \sum_{i=1}^{N} (\lambda_i^{(j)} - \lambda_i^{(j-1)})'(L_j^{\lambda_i} - L_{j-1}^{\lambda_i}) \right|}{\left\| L_j^{\beta} - L_{j-1}^{\beta} \right\|^2 + \sum_{i=1}^{N} \left\| L_j^{\lambda_i} - L_{j-1}^{\lambda_i} \right\|^2}.$$

Step 2.3: Update the parameters by

$$\beta^{(j+1)} = \beta^{(j)} - s_j \cdot L_j^{\beta} \quad \text{and} \quad \lambda_i^{(j+1)} = \lambda_i^{(j)} - s_j \cdot L_j^{\lambda_i}.$$

Step 2.4: Iterate Step 2.2 and Step 2.3 until the object function converges.

Since the object function $L(\beta, \Lambda)$ is not convex in (β, Λ), there is no guarantee that the gradient descent algorithm above is able to find the global minimum. Thus, choosing a good initial value for the parameter is essential. In practice, I recommend using the following estimator as the initial value of the parameter:

$$[\tilde{\beta}(\tau), \tilde{\Lambda}(\tau)] = \underset{\beta \in \mathcal{B}, \lambda_i \in \mathcal{A}}{\arg\min} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(Y_{it} - \beta' X_{it} - \lambda_i' \hat{f}_t), \tag{5}$$

where $\rho_\tau = (\tau - \mathbf{1}\{u \leq 0\})u$ is the check function. Under Assumptions 1 and 2 of the next section, it can be shown that $\tilde{\beta}(\tau)$ is a consistent estimator of $\beta_0(\tau)$.[5]

Remark 1 *The way I estimate the unknown factors in Step 1 is different from the standard CCE method that uses $\bar{X}_t$ and $\bar{Y}_t = N^{-1} \sum_{i=1}^{N} Y_{it}$ as the proxies of f_t. A problem with the CCE approach, as pointed out by Karabiyik et al. (2017), is that the second moment matrix of the estimated factors is asymptotically singular when $p+1 > r$, known as the problem of "degenerated regressors". This problem results in two possible complications for nonlinear panel data models: first, the asymptotic property of the CCE estimator is more challenging to establish and there might be extra biases due to the degenerated regressors (see Theorem 3 of Karabiyik et al. 2017); Second, since the nonlinear models usually requires nonlinear optimization algorithms to obtain the estimator, it is difficult to find the (local) minimum with degenerated regressors. My approach*

[4] This method of choosing the step size is known as the Barzilai-Borwein method.

[5] The proof of this claim is essentially the same as the proof of Theorem 1, and it is therefore omitted. In fact, the consistency of $\tilde{\beta}(\tau)$ does not require $\mathcal{B}$ to be compact thanks to the convexity of the check function — see Kato et al. (2012).

avoids this problem because it will be shown in the next section that the second moment matrix of the estimated factors is asymptotically full rank as long as $p \geq r$.

Remark 2 *A natural question that arises is why not just use the estimator given in (5) in Step 2. The main reason is that it is difficult to work out the analytical expression of the asymptotic bias of $\tilde{\beta}(\tau)$ due to the non-smoothness of the check function — see Kato et al. (2012) for a detailed discussion. The use of SQR in Step 2 is inspired by Galvao and Kato (2016), who derived the asymptotic bias of the fixed effects estimator for quantile panel data models with only individual effects. Similar ideas has been explored by Amemiya (1982) and Horowitz (1998), but for different objectives.*

Remark 3 *At $\tau = 0.5$, the models (1) to (3) can be viewed as a variant of the model of Pesaran (2006) where the assumption that u_{it} has conditional mean 0 is replaced by the assumption that u_{it} has conditional median 0. Accordingly, my two-step estimator at $\tau = 0.5$ can be viewed as the least absolute deviation (LAD) counterpart of the CCE estimator. As will be shown below, the advantage of the LAD estimator is that I only need restrictions on the conditional density of u_{it} to establish its consistency and asymptotic normality, making it more robust to outliers and heavy-tailed distributions. The robustness of the proposed estimator against heavy-tailed distributions is examined through Monte Carlo simulations in Section 4.2.*

3 Asymptotic Results

Suppose that we have a panel of observations $\{(Y_{it}, X_{it}), i = 1, \ldots, N, t = 1, \ldots, T\}$ generated from (1) and (3), where the realized values of the individual and time effects are $\Lambda_0(\tau) = [\lambda_{01}(\tau), \ldots, \lambda_{0N}(\tau)]'$ and $F_0 = [f_{01}, \ldots, f_{0T}]'$. In the section, following the literature on nonlinear panel data models, I adopt a fixed effects approach by treating $\Lambda_0(\tau)$ and F_0 as fixed (nuisance) parameters. Thus, given $\Lambda_0(\tau)$ and F_0, my model can be written as

$$Y_{it} = \beta_0(\tau)' X_{it} + \lambda_{0i}(\tau)' f_{0t} + u_{it}, \quad \mathsf{Q}_{u_{it}}[\tau | X_{it}] = \tau, \quad \text{and } X_{it} = \Gamma_i f_{0t} + e_{it}.$$

Alternatively, all the assumptions and asymptotic results in this section can be understood as being conditional on $\Lambda(\tau) = \Lambda_0(\tau)$ and $F = F_0$. Moreover, to simplify the notations, I suppress the dependence of $\lambda_{0i}(\tau)$ on τ throughout this section.

3.1 The Number of Factors

In the previous section I assume that r is known, which is rarely that case in most empirical applications. Thus, in this subsection I propose a consistent estimator of r.

Note that

$$\hat{\Sigma}_{\bar{X}} = \frac{1}{T}\sum_{t=1}^{T}\bar{e}_t\bar{e}_t' + \bar{\Gamma}\cdot\frac{1}{T}\sum_{t=1}^{T}f_{0t}\bar{e}_t' + \frac{1}{T}\sum_{t=1}^{T}\bar{e}_t f_{0t}'\bar{\Gamma}' + \bar{\Gamma}\hat{\Sigma}_{f_0}\bar{\Gamma}',$$

where $\hat{\Sigma}_{f_0} = T^{-1}\sum_{t=1}^{T} f_{0t}f_{0t}'$, $\bar{\Gamma} = N^{-1}\sum_{i=1}^{N}\Gamma_i$ and $\bar{e}_t = N^{-1}\sum_{i=1}^{N} e_{it}$. If we assume that $\{e_{it}, i = 1, \ldots, N\}$ is weakly dependent for each t, the first three terms on the right-hand side of the above equation can be shown to be $o_P(1)$. Moreover, if both $\bar{\Gamma}$ and $\hat{\Sigma}_{f_0}$ have full rank, then $\hat{\Sigma}_{\bar{X}}$ converges in probability to a matrix with rank r. This observation motivates the following estimator of r.

Let $\hat{\rho}_1 \geq \hat{\rho}_2 \cdots \geq \hat{\rho}_p$ be the eigenvalues of $\hat{\Sigma}_{\bar{X}}$, and let $\mathbb{P}_{NT}$ be a sequence of non-negative constants. Then the estimator of r is defined as

$$\hat{r} = \sum_{j=1}^{p}\mathbf{1}\{\hat{\rho}_j > \mathbb{P}_{NT}\}.$$

In order to prove the consistency of $\hat{r}$, I impose the following conditions:

Assumption 1 *Let $M > 0$ be a generic bounded constant.*
(i) $p \geq r$.
(ii) $\|f_{0t}\| \leq M$ for all t. There exists $\Sigma_{f_0} \in \mathbb{R}^{r\times r}$ and $\Gamma_0 \in \mathbb{R}^{p\times r}$ such that $\|\hat{\Sigma}_{f_0} - \Sigma_{f_0}\| = O(T^{-1/2})$, $\|\bar{\Gamma} - \Gamma_0\| = O(N^{-1/2})$, and $\text{rank}(\Sigma_{f_0}) = \text{rank}(\Gamma_0) = r$.
(iii) $\mathbb{E}[e_{it}] = 0$ for i, t, and $\mathbb{E}\|N^{-1/2}\sum_{i=1}^{N} e_{it}\|^2 \leq M$ for all t.

The conditions that $p \geq r$ and $\text{rank}(\Gamma_0) = r$ are standard in the literature of CCE estimation to ensure that the space of the common factors can be approximated by the cross-sectional averages of the regressors. Condition (ii) implies $\Gamma_0\Sigma_{f_0}\Gamma_0'$ has full rank. It is also worth noting that I only require weak cross-sectional correlations of e_{it} through condition (iii), and the temporal correlations of e_{it} are left unrestricted.

Then it can be shown that:

Proposition 1 *Under Assumption 1, $P[\hat{r} = r] \to 1$ as $N, T \to \infty$ if $\mathbb{P}_{NT} \to 0$ and $\mathbb{P}_{NT} \cdot \min(\sqrt{N}, \sqrt{T}) \to \infty$.*

Given the above result, the number of factors r can be treated as known in the subsequent analysis regarding the asymptotic properties of $\hat{\beta}(\tau)$ (see footnote 5 of Bai 2003).

3.2 Consistency

let $\Psi_0 \in \mathbb{R}^{p\times r}$ be the matrix of eigenvectors associated with the r distinct positive eigenvalues of $\Gamma_0\Sigma_{f_0}\Gamma_0'$, and define $\mathbf{H}_0 = \Psi_0'\Gamma_0$, $\tilde{f}_{0t} = \mathbf{H}_0 f_{0t}$, $\tilde{\lambda}_{0i} = (\mathbf{H}_0')^{-1}\lambda_{0i}$. Note that $\mathbf{H}_0$ is a full rank matrix[6]. In addition, define $V_{it} = [X'_{it}, \tilde{f}'_{0t}]'$, let $\mathsf{f}_{it}(\cdot|x)$ denote the conditional density of u_{it} given $X_{it} = x$, and let $\varrho_{i,T}$ denote the smallest eigenvalue of $T^{-1}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})V_{it}V'_{it}]$.

To derive the consistency of the estimator, I impose the following conditions in addition to Assumption 1:

Assumption 2 *Let $M > 0$ be a generic bounded constant and let $m \geq 1$ be a positive integer.*
(i) The r positive eigenvalues of $\Gamma_0\Sigma_{f_0}\Gamma'_0$ are distinct.
(ii) $\beta_0(\tau) \in \mathcal{B}$, $\tilde{\lambda}_{0i} \in \mathcal{A}$ for all i, and $\mathcal{A}, \mathcal{B}$ are compact.
(iii) There exists $\underline{\varrho} > 0$ such that $N^{-1}\sum_{i=1}^{N}\varrho_{i,T} > \underline{\varrho}$ for all large N and T. $\mathsf{f}_{it}^{(1)}(c|x) = \partial\mathsf{f}_{it}(c|x)/\partial c$ exists and $\max_{i,t}|\mathsf{f}_{it}^{(1)}(c|x)| < M$ uniformly over (c, x).
(iv) For each i, the sequence $\{(X_{it}, u_{it}), i = 1, \ldots, N\}$ is α-mixing with coefficients $\alpha_i(j)$ satisfying that $\max_{1\leq i\leq N}\alpha_i(j) \leq M \cdot \alpha^j$ for some $0 < \alpha < 1$.
(v) There exists $\gamma > 0$ such that $\mathbb{E}\|X_{it}\|^{2m+\gamma} < M$ for all i, t.
(vi) As $N, T \to \infty$, $h \to 0$ and $N/T^m \to 0$.

Before presenting the consistency result, I briefly comment on the conditions in Assumption 2.

Condition (i) allows the use of perturbation theory for the eigenvectors of $\Gamma_0\Sigma_{f_0}\Gamma'_0$, which is important for the result that $\hat{f}_t$ converges to $\mathbf{H}_0 f_{0t}$. Similar condition has been imposed in the study of PCA estimators for approximate factor models (see Assumption G of Bai 2003).

Condition (ii) requires the parameter spaces to be compact. The compactness of $\mathcal{B}$ is needed because the smoothed check function is no longer convex in (β, λ_i) given (X_{it}, f_t), and the compactness of $\mathcal{A}$ helps to bound the impact of the estimation errors of $\hat{f}_t$ on the object function.

Condition (iii) is similar to the standard identification condition in quantile regressions. The main difference here is that I have to take into account the common factors. Note that it allows $\varrho_{i,T}$, the smallest eigenvalue of $T^{-1}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})V_{it}V'_{it}]$, to be 0 for some i as long as $N^{-1}\sum_{i=1}^{N}\varrho_{i,T}$ is bounded below by a positive constant. But it will fail if $X_i = [X_{i1}, \ldots, X_{iT}]'$ and $F_0 = [f_{01}, \ldots, f_{0t}]'$ span the same space for all i, e.g., $e_{it} = 0$ for all i, t.

Condition (iv) is also standard in the literature (see Assumption D.1 of Kato et al. 2012). The strong mixing condition is used to derive moments bounds in order to apply law of large numbers and central limit theorems. It is commonly employed in nonlinear panel data models because the mixing property is nicely preserved by nonlinear transformations. However, I don't assume stationarity because for the factor loadings to be quantile dependent, u_{it} should be allowed to depend on the factors. Therefore, conditional on f_{0t}, it become necessary to allow the distribution of u_{it} to change across t. Moreover, conditional on f_{0t}, the mean of X_{it} is given by $\Gamma_i f_{0t}$, thus the distribution of X_{it} should also be time dependent.

[6] see the proof of Lemma 1.

Condition (v) and (vi) reflect a trade-off between the moments of X_{it} and the required relative size of T compare to N. The existence of higher moments of X_{it} allows for less restrictive conditions on the size of T. In particular, if $m = 1$ and $\mathbb{E}\|X_{it}\|^{2+\gamma} < M$, we need $N/T \to 0$ — a very strong condition that is hard to satisfy in most empirical applications. However, if $m = 2$ and $\mathbb{E}\|X_{it}\|^{4+\gamma} < M$, only $N/T^2 \to 0$ is needed. Moreover, if it is assumed that $\|X_{it}\| \leq M$ for all i, t almost surely, condition (vi) can be relaxed to $\log N/\sqrt{T} \to 0$ (see Proposition 3.1 of Galvao and Kato 2016).

Besides the above conditions imposed in Assumption 2, it is worth mentioning that the cross-sectional dependence of (X_{it}, u_{it}) is not explicitly restricted. While the cross-sectional dependence of X_{it} is implicitly controlled by Assumption 1(iii), no such restriction is needed for u_{it}. The intuition is that in the large-T asymptotic framework, given $\{f_{0t}\}$, $\beta_0(\tau)$ can be consistently estimated from observations of any individual i. Thus, for consistency I only need weak dependence of u_{it} on the time dimension, and the weak cross-sectional dependence of X_{it} is only needed to ensure that the space of $\{f_{0t}\}$ can be well approximated by $\{\hat{f}_t\}$.

Last but not least, note that unlike Ando and Bai (2020), I don't impose any moment restrictions on u_{it}, making my estimation procedure robust to outliers and heavy-tailed distributions. Moreover, compared with the procedures that estimate the factor and factor loadings jointly (such as Chen et al. 2019), I don't need any rank condition on the factor loading matrix, which means that some of the factor loadings in model (1) can be 0. In other words, there can be some factors that affect X_{it} but not Y_{it}.

The following theorem establishes the consistency of $\hat{\beta}(\tau)$ for any given $\tau \in (0, 1)$.

Theorem 1 *Under Assumptions 1 and 2, $\hat{\beta}(\tau)$ is weakly consistent, i.e., $\|\hat{\beta}(\tau) - \beta_0(\tau)\| = o_P(1)$ for any $\tau \in (0, 1)$.*

3.3 Asymptotic Distribution

Let $\mathsf{f}_{it}(\cdot)$ be the density function of u_{it}, $\mathsf{f}_{it}(\cdot|x_{it})$ be the conditional density of u_{it} given $X_{it} = x_{it}$, and $\mathsf{f}_{i,ts}(\cdot, \cdot|x_{it}, x_{is})$ be the joint density of (u_{it}, u_{is}) given $(X_{it}, X_{is}) = (x_{it}, x_{is})$. Moreover, let $\mathsf{f}_{it}^{(j)}(c) = \partial^j \mathsf{f}_{it}(c)/\partial c^j$, $\mathsf{f}_{it}^{(j)}(c|x_{it}) = \partial^j \mathsf{f}_{it}(c|x_{it})/\partial c^j$, $\mathsf{f}_{i,ts}^{(j,k)}(c_1, c_2|x_{it}, x_{is}) = \partial^{j+k} \mathsf{f}_{it}(c|x_{it})/\partial c_1^j \partial c_2^k$. In particular, let $\mathsf{f}_{it}^{(0)}(c) = \mathsf{f}_{it}(c)$ and $\mathsf{f}_{it}^{(0)}(c|x_{it}) = \mathsf{f}_{it}(c|x_{it})$.

In addition, define

$$\underbrace{\Xi_i}_{p\times r} = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}]f_{0t}', \quad \underbrace{\Omega_i}_{r\times r} = \frac{1}{T}\sum_{t=1}^{T} \mathsf{f}_{it}(0) f_{0t} f_{0t}', \quad \underbrace{\Phi_i}_{p\times r} = \Xi_i \Omega_i^{-1},$$

$$\underbrace{Z_{it}}_{p\times 1} = X_{it} - \Xi_i \Omega_i^{-1} f_{0t}, \quad \underbrace{\Delta}_{r\times r} = \lim_{N,T\to\infty} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}Z_{it}'].$$

To derive the asymptotic distribution of $\hat{\beta}(\tau)$, I need the following conditions:

Assumption 3 *Let $M > 0$ a generic bounded constant, and let $q \geq 8$ be an even integer.*
(i) $\beta_0(\tau)$ is an interior point of $\mathcal{B}$ and $\{\tilde{\lambda}_{01}, \ldots, \tilde{\lambda}_{0N}\}$ are interior points of $\mathcal{A}$.
(ii) $\{\Omega_i, i = 1, \ldots, N\}$ are all invertible for large T and Δ is invertible.
(iii) $\max_{i,t} \|X_{it}\| < M$ almost surely.
(iv) Define $X_i^T = (X_{i1}, \ldots, X_{iT})$ and $u_i^T = (u_{i1}, \ldots, u_{iT})$. $\{(X_i^T, u_i^T), i = 1, \ldots, T\}$ are independent across i.
(v) $\mathsf{f}_{it}(c|x_{it})$ is q times continuously differentiable with respect to c and $\mathsf{f}_{i,ts}(c_1, c_2|x_{it}, x_{is})$ is q times continuously differentiable with respect to (c_1, c_2); $\left|\mathsf{f}_{it}^{(j)}(c|x_{it})\right| \leq M$ uniformly over (c, x_{it}) for all $j = 0, \ldots, q$; $\left|\mathsf{f}_{i,ts}^{(j,0)}(c_1, c_2|x_{it}, x_{is})\right| \leq M$ and $\left|\mathsf{f}_{i,ts}^{(0,j)}(c_1, c_2|x_{it}, x_{is})\right| \leq M$ uniformly over $(c_1, c_2, x_{it}, x_{is})$ for all $j = 0, \ldots, q$.
(vi) $\int_{-1}^{1} k(u)du = 1$, $\int_{-1}^{1} k(u)u^j du = 0$ for $j = 1, \ldots, q-1$ and $\int_{-1}^{1} k(u)u^q du \neq 0$.
(vii) $N/T \to \kappa^2 > 0$ as $N, T \to \infty$. $h \asymp T^{-c}$ and $1/q < c < 1/6$.

Remark 4 *The conditions of Assumption 3 above are very similar to the assumptions imposed in Galvao and Kato (2016). Thus, I refer to Galvao and Kato (2016) for the details of these conditions. The only difference is that Galvao and Kato (2016) requires $q \geq 4$ and $1/q < c < 1/3$ while I need the stronger conditions that $q \geq 8$ and $1/q < c < 1/6$. More specifically, due to the presence of the interactive effects, Lemma B.2 of Galvao and Kato (2016) can not be used to show that the remaining terms in the expansion of $\hat{\beta}(\tau) - \beta_0(\tau)$ is $o_P(T^{-1})$. Instead, to bound the higher order terms, I combine the uniform convergence rates of $\hat{\lambda}_i$ and $\hat{f}_t$ and the fact that the third order derivative of the object function is uniformly bounded (up to a positive constant) by $1/h^2$ — this is why I need a much larger h and therefore a much smaller c.*

Next, define:

$$\underbrace{\mathbf{A}_t}_{p\times r} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}]\lambda_{0i}', \quad \underbrace{\mathbf{B}_{t,k}}_{r\times r} = \frac{1}{N}\sum_{i=1}^{N}\mathsf{f}_{it}(0)\lambda_{0i}\Phi_{i,k},$$

$$\underbrace{\mathbf{C}_{i,k}}_{r\times r} = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]f_{0t}f_{0t}', \quad \underbrace{\mathbf{D}_{t,k}}_{r\times r} = -\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]\lambda_{0i}\lambda_{0i}'.$$

I need to impose some extra conditions to make sure that the asymptotic biases of $\hat{\beta}(\tau)$ are well defined.

Assumption 4 *Define:*

$$\omega_{T,i}^{(1)} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big] f_{0t}'\Omega_i^{-1}f_{0t},$$

$$\omega_{T,i}^{(2)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\left(\tau\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}] - \mathbb{E}\left[\int_{\infty}^{0}\mathsf{f}_{i,ts}(0,v|X_{it},X_{is})dv\cdot Z_{it}\right]\right) f_{0t}'\Omega_i^{-1}f_{0s},$$

$$\omega_{T,i,k}^{(3)} = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T} f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0t},$$

$$\omega_{T,i,k}^{(4)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\left\{\mathbb{E}[\mathbf{1}\{u_{it}\leq 0, u_{is}\leq 0\}] - \tau^2\right\} f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0s},$$

and assume that the following limits exist:

$$b_1 = -(\tau-0.5)\cdot\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(1)} - \lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(2)},$$

$$b_{2,k} = 0.5\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i,k}^{(3)} + 0.5\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i,k}^{(4)},$$

$$d_1 = -\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it}],$$

$$d_{2,k} = 0.5\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathsf{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\cdot\Psi_0(\mathbf{H}_0')^{-1}\left(2\mathbf{B}_{t,k}+\mathbf{D}_{t,k}\right)(\mathbf{H}_0)^{-1}\Psi_0'\right\}.$$

The following theorem gives the asymptotic distribution of $\hat{\beta}(\tau)$.

Theorem 2 *Suppose that Assumptions 1 to 4 hold, then as $N, T\to\infty$,*

$$\sqrt{NT}\left[\hat{\beta}(\tau)-\beta_0(\tau)\right]\xrightarrow{d}\mathcal{N}(\Delta^{-1}(\kappa b+\kappa^{-1}d),\Delta^{-1}\mathbf{V}\Delta^{-1}),$$

where $b = b_1 + b_2$, $d = d_1 + d_2$, $b_2 = [b_{2,1},\dots b_{2,p}]'$, $d_2 = [d_{2,1},\dots d_{2,p}]'$, $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$,

$$\mathbf{V}_1 = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[W_{it}W_{it}'],\quad \mathbf{V}_2 = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}[W_{it}W_{is}'],$$

and

$$W_{it} = [\tau - \mathbf{1}\{u_{it}\leq 0\}]Z_{it} - \mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'e_{it}.$$

Remark 5 *In the proof of Theorem 2, I show that $\hat{\beta}(\tau)-\beta_0(\tau)$ has the following Bahadur* *representation:*

$$\Delta(\hat{\beta}(\tau) - \beta_0(\tau)) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}^* + \frac{b}{T} + \frac{d}{N} + o_P(T^{-1}),$$

where $W_{it}^* = l^{(1)}(u_{it})Z_{it} - \mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0' e_{it}$, *and* $l^{(1)}(u) = \tau - K(u/h) + k(u/h)u/h$. *The first bias term* b/T *is caused by the estimation of* Λ_0 *and the second bias term* d/N *originates from the estimation of* F_0. *In nonlinear (probit, logit) panel data models with interactive effects, Chen et al. (2020) is the first to establish a similar Bahadur representation for the fixed-effects estimator. Similar to my result, the biases of their estimator, which are generally non-zero except for some special cases, arise from the estimation of the fixed effects. This is in contrast to linear panel data models with interactive effects, where the fixed-effects estimator of the slope parameter has a similar Bahadur representation (see Theorem 3 of Bai 2009), but the bias term* b/T *is due to cross-sectional correlation and heteroskedasticity and the bias term* d/T *is caused by serial correlation and heteroskedasticity.*

3.3.1 Some Special Cases

(a) Observed factors

First, in some applications, the common factors are observed (e.g., money suppy). In this case, we don't need to estimate F_0 from the first step. As a consequence, the asymptotic distribution of $\hat{\beta}(\tau)$ will not be affected by the estimation errors of the factors. In particular, it can be shown that $d_1 = d_2 = 0$ and that $W_{it}^* = [\tau - \mathbf{1}\{u_{it} \leq 0\}]Z_{it}$. Thus,

$$\sqrt{NT}\left[\hat{\beta}(\tau) - \beta_0(\tau)\right] \xrightarrow{d} \mathcal{N}(\kappa\Delta^{-1}b, \Delta^{-1}\mathbf{V}\Delta^{-1})$$

where

$$\mathbf{V} = \tau(1-\tau)\cdot \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[Z_{it}Z_{it}'] +$$
$$\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}\left[\left(\tau^2 - \mathsf{F}_{it}(0|X_{it},X_{is}) - \mathsf{F}_{is}(0|X_{it},X_{is}) + \mathsf{F}_{i,ts}(0,0|X_{it},X_{is})\right)Z_{it}Z_{is}'\right],$$

and $\mathsf{F}_{it}(0|X_{it},X_{is}) = \mathbb{E}[\mathbf{1}\{u_{it}\leq 0\}|X_{it},X_{is}]$, $\mathsf{F}_{i,ts}(0,0|X_{it},X_{is}) = \mathbb{E}[\mathbf{1}\{u_{it}\leq 0, u_{is}\leq 0\}|X_{it},X_{is}]$.

(b) Only individual effects

If we further assume that $r = 1$, $f_{0t} = 1$ for all t, and $\{(X_{it}, u_{it}), t = 1, \ldots, T\}$ is stationary for each i, then $\Xi_i = \mathbb{E}[\mathsf{f}_i(0|X_{it})X_{it}]$, $\Omega_i = \mathsf{f}_i(0)$, $Z_{it} = X_{it} - \mathbb{E}[\mathsf{f}_i(0|X_{it})X_{it}]/\mathsf{f}_i(0)$,

$$\omega_{T,i}^{(1)} = \mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]/\mathsf{f}_i(0) = 0,$$

$$\omega_{T,i}^{(2)} = \sum_{1\le|k|\le T-1}\left(1-\frac{|k|}{T}\right)\left(\tau\mathbb{E}[\mathsf{f}_i(0|X_{it})Z_{it}] - \mathbb{E}\left[\int_{\infty}^{0}\mathsf{f}_{i,t,t+k}(0,v|X_{it},X_{i,t+k})dv\cdot Z_{it}\right]\right)/\mathsf{f}_i(0),$$

$$\omega_{T,i}^{(3)} = -\frac{\tau(1-\tau)}{\mathsf{f}_i(0)^2}\mathbb{E}[\mathsf{f}_i^{(1)}(0|X_{it})Z_{it}],$$

$$\omega_{T,i}^{(4)} = -\sum_{1\le|k|\le T-1}\left(1-\frac{|k|}{T}\right)\left\{\mathbb{E}[\mathbf{1}\{u_{it}\le 0, u_{i,t+k}\le 0\}] - \tau^2\right\}\cdot\mathbb{E}[\mathsf{f}_i^{(1)}(0|X_{it})Z_{it}]/\mathsf{f}_i(0)^2.$$

Therefore, the asymptotic distribution of $\hat{\beta}(\tau) - \beta_0(\tau)$ is identical to the one given by Theorem 3.2 of Galvao and Kato (2016).

(c) No time-series dependences

When there are no time-series dependences, i.e., $\{(X_{it}, u_{it}), t = 1, \dots, T\}$ is independent across t for each i, it is easy to see that $\omega_{T,i}^{(2)} = 0$, $\omega_{T,i,k}^{(4)} = 0$, and $\mathbf{V}_2 = 0$. Thus, we have

$$b_1 = -(\tau - 0.5)\cdot\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]f_{0t}'\Omega_i^{-1}f_{0t},$$

$$b_{2,k} = \frac{\tau(1-\tau)}{2}\cdot\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0t},$$

and

$$\mathbf{V} = \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[W_{it}W_{it}'] = \tau(1-\tau)\cdot\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[Z_{it}Z_{it}']$$
$$+\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'\mathbb{E}[e_{it}e_{it}']\Psi_0(\mathbf{H}_0')^{-1}\mathbf{A}_t'.$$

3.4 Bias Correction

3.4.1 Analytical Bias Correction

Theorem 2 above provides the basis of analytical bias correction for $\hat{\beta}(\tau)$. Suppose that $\hat{\Delta}$, $\hat{b}$ and $\hat{d}$ are consistent estimators of Δ, b, d respectively, and define

$$\hat{\beta}_{abc}(\tau) = \hat{\beta}(\tau) - \hat{\Delta}^{-1}\left(\frac{\hat{b}}{T} + \frac{\hat{d}}{N}\right).$$

Then it follows easily from Theorem 2 that the bias corrected estimator $\hat{\beta}_{ABC}(\tau)$ will have an asymptotic normal distribution that is centered around 0, i.e.,

$$\sqrt{NT}\left[\hat{\beta}_{abc}(\tau) - \beta_0(\tau)\right] \xrightarrow{d} \mathcal{N}(0, \Delta^{-1}\mathbf{V}\Delta^{-1}). \tag{6}$$

To construct consistent estimators of Δ, b, d, let $\{\hat{e}_{it}\}$ be the OLS residuals of regressing $\{X_{it}\}$ on $\{\hat{f}_t\}$, define $l(u) = (\tau - K(u/h))u$, $l^{(j)}(u) = \partial^j l(u)/\partial u^j$, $\hat{u}_{it} = Y_{it} - \hat{\beta}(\tau)' X_{it} - \hat{\lambda}_i' \hat{f}_t$, and

$$\hat{\Xi}_i = \frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it}) X_{it}\hat{f}_t', \quad \hat{\Omega}_i = \frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{f}_t\hat{f}_t', \quad \hat{\Phi}_i = \hat{\Xi}_i\hat{\Omega}_i^{-1},$$

$$\hat{Z}_{it} = X_{it} - \hat{\Xi}_i\hat{\Omega}_i^{-1}\hat{f}_t, \quad \hat{\Delta} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{Z}_{it}',$$

$$\hat{\mathbf{B}}_{t,k} = \frac{1}{N}\sum_{i=1}^{N} l^{(2)}(\hat{u}_{it})\hat{\lambda}_i\hat{\Phi}_{i,k} \quad, \hat{\mathbf{C}}_{i,k} = \frac{1}{T}\sum_{t=1}^{T} l^{(3)}(\hat{u}_{it})\hat{Z}_{it,k}\hat{f}_t\hat{f}_t', \quad \hat{\mathbf{D}}_{t,k} = \frac{1}{N}\sum_{i=1}^{N} l^{(3)}(\hat{u}_{it})\hat{Z}_{it,k}\hat{\lambda}_i\hat{\lambda}_i'.$$

$$\hat{\omega}_{T,i}^{(1)} = \frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t, \quad \hat{\omega}_{T,i,k}^{(3)} = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_t,$$

$$\hat{\omega}_{T,i}^{(2)} = \frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L} l^{(2)}(\hat{u}_{it})\hat{Z}_{it} l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s + \frac{1}{T}\sum_{t=L+1}^{T}\sum_{s=t-L}^{t-1} l^{(2)}(\hat{u}_{it})\hat{Z}_{it} l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s,$$

$$\omega_{T,i,k}^{(4)} = \frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L} l^{(1)}(\hat{u}_{it}) l^{(1)}(\hat{u}_{is})\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_s + \frac{1}{T}\sum_{t=L+1}^{T}\sum_{s=t-L}^{t-1} l^{(1)}(\hat{u}_{it}) l^{(1)}(\hat{u}_{is})\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_s,$$

$$\hat{b}_1 = -(\tau - 0.5)\cdot\frac{1}{N}\sum_{i=1}^{N}\hat{\omega}_{T,i}^{(1)} - \frac{1}{N}\sum_{i=1}^{N}\hat{\omega}_{T,i}^{(2)}, \quad \hat{b}_{2,k} = 0.5\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\omega}_{T,i,k}^{(3)} + \hat{\omega}_{T,i,k}^{(4)}\right),$$

$$\hat{d}_1 = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'\hat{\Psi}'\hat{e}_{it}, \quad \hat{d}_{2,k} = 0.5\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}'\hat{\Psi}\left(2\hat{\mathbf{B}}_{t,k} + \hat{\mathbf{D}}_{t,k}\right)\hat{\Psi}'\hat{e}_{it}.$$

Given the above definitions, the estimators for b and d are given by $\hat{b} = \hat{b}_1 + \hat{b}_2$ and $\hat{d} = \hat{d}_1 + \hat{d}_2$ respectively, where $\hat{b}_2 = [\hat{b}_{2,1}, \ldots, \hat{b}_{2,p}]'$ and $\hat{d}_2 = [\hat{d}_{2,1}, \ldots, \hat{d}_{2,p}]'$. The following result confirms the validity of the proposed analytical bias correction.

Theorem 3 *Let $\nu < 1/6 - c$ be a positive constant where c is defined in Assumption 3(vii). Then under Assumptions 1 to 4, $\hat{\Delta} = \Delta + o_P(1)$, $\hat{b} = b + o_P(1)$, $\hat{d} = d + o_P(1)$ and therefore (6) holds if $L \to \infty$ and $L/(T^{0.5-\nu}h^3) \to 0$ as $N, T \to \infty$.*

3.4.2 Jackknife Bias Correction

Following Dhaene and Jochmans (2015), Fernández-Val and Weidner (2016) and Chen et al. (2020), an alternative method to correct the leading bias of $\hat{\beta}(\tau)$ is the SPJ.

For a given τ, let $\hat{\beta}^{(1)}_{N,T/2}(\tau)$ be the two-step estimator, defined as in (4), using the subsample $i = 1, \ldots, N; t = 1, \ldots, T/2$, and let $\hat{\beta}^{(2)}_{N,T/2}(\tau)$ be the two-step estimator using the subsample $i = 1, \ldots, N; t = T/2 + 1, \ldots, T$. Similarly, define $\hat{\beta}^{(1)}_{N/2,T}(\tau)$ as the two-step estimator using the subsample $i = 1, \ldots, N/2; t = 1, \ldots, T$, and $\hat{\beta}^{(2)}_{N/2,T}(\tau)$ as the two-step estimator using the subsample $i = N/2 + 1, \ldots, N$ and $t = 1, \ldots, T$. Then the bias-corrected estimator using the SPJ is defined as

$$\hat{\beta}_{spj}(\tau) = 3\hat{\beta}(\tau) - \frac{1}{2}\left[\hat{\beta}^{(1)}_{N,T/2}(\tau) + \hat{\beta}^{(2)}_{N,T/2}(\tau)\right] - \frac{1}{2}\left[\hat{\beta}^{(1)}_{N/2,T}(\tau) + \hat{\beta}^{(2)}_{N/2,T}(\tau)\right]. \tag{7}$$

The computation of this estimator is almost as easy as the original two-step estimator $\hat{\beta}(\tau)$.

The main intuition of the SPJ estimator is that if the underlying distributions of the data are stable across i and t, the term $0.5(\hat{\beta}^{(1)}_{N,T/2}(\tau) + \hat{\beta}^{(2)}_{N,T/2}(\tau)) - \hat{\beta}(\tau)$ is a good estimate of b/T, and the term $0.5(\hat{\beta}^{(1)}_{N/2,T}(\tau) + \hat{\beta}^{(2)}_{N/2,T}(\tau)) - \hat{\beta}(\tau)$ is a good estimate of d/N. In models with only individual effects, the asymptotic bias of the fixed-effects estimator is determined by the distribution of (X_{it}, u_{it}). Thus, the formal justification of the SPJ only requires the sequence $\{(X_{it}, u_{it}), t = 1, 2, \ldots\}$ to be stationary for each i (see Dhaene and Jochmans 2015 and Galvao and Kato 2016). However, in models with interactive effects, the asymptotic biases are also affected by Λ_0 and F_0. Thus, to justify the use of the SPJ, we also need some kind of of conditions to ensure that the distributions of $f_1, \ldots, f_T$ are stable across t and that the distributions of $\lambda_1, \ldots, \lambda_N$ are stable across i. On one hand, such assumptions involve the unconditional distributions of Λ and F; on the other hand, my asymptotic theory is established conditional on Λ_0 and F_0 (realizations of Λ and F) — this gap makes it difficult to rigorously prove the validity of the SPJ estimator. I leave this important but challenging question for future research, and the finite sample performance of the SPJ estimator is evaluated in the next section using Monte Carlo simulations.

3.5 Estimating the Variance

The previous subsection gives a consistent estimator of Δ. Thus, it remains to construct a consistent estimator of $\mathbf{V}$. Define

$$\hat{\mathbf{A}}_t = \frac{1}{N}\sum_{i=1}^{N} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i', \quad \hat{W}_{it} = l^{(1)}(\hat{u}_{it})\hat{Z}_{it} - \hat{\mathbf{A}}_t\hat{\Psi}'\hat{e}_{it}, \quad \hat{\mathbf{V}}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{W}_{it}\hat{W}_{it}',$$

$$\hat{\mathbf{V}}_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\hat{W}_{it}\hat{W}_{is}' + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1+L}^{T}\sum_{s=t-L}^{t-1}\hat{W}_{it}\hat{W}_{is}',$$

and $\hat{\mathbf{V}} = \hat{\mathbf{V}}_1 + \hat{\mathbf{V}}_2$. The following result establishes the consistency of $\hat{\mathbf{V}}$.

Theorem 4 *Let L satisfy the condition of Theorem 3. Then under Assumptions 1 to 4,* $\hat{\mathbf{V}} = \mathbf{V} + o_P(1)$.

3.6 The Choice of Tuning Parameters in Practice

The implementation of the proposed estimation procedure in practice involves choosing the kernel function $k(\cdot)$, the bandwidth parameter h, and the truncation parameter L in the HAC-type estimators of the biases and variance.

First, Assumption 3 requires $k(\cdot)$ to be (at least) an eighth-order kernel function. Thus, I recommend using the following kernel function (see Muller 1984):

$$k(z) = \mathbf{1}\{|z| \leq 1\} \cdot \frac{3465}{8192}\left(7 - 105z^2 + 462z^4 - 858z^6 + 715z^8 - 221z^{10}\right).$$

Second, if one chooses the eighth-order kernel function above, Assumption 3 requires that $h \asymp T^{-c}$ and $1/8 < c < 1/6$. Thus, when N is about the size of T in practice, a possible choice is $h = 1.5(NT)^{-1/14}$, which is the one I use in all the simulations in the next section. Note that even when there are only individual effects, the optimal bandwidth choice in SQR still remains an open question (see Galvao and Kato 2016). Thus, I would like to leave the important but challenging question for future research.

Finally, the choice of L in dynamic models is a more delicate issue, especially in the current context. As pointed out by Galvao and Kato (2016), the standard theory for the HAC estimator of covariance matrix in models with smooth object functions does not apply to the quantile panel data models. Even though Theorem 3 and Theorem 4 require that L goes to infinity as N, T get large, the simulation results in the next section support the proposal of Galvao and Kato (2016) and Hahn and Kuersteiner (2011) to use $L = 1$ in practice, especially when T is not large. Thus, following the literature and based on my simulation results (see the next section), I recommend choosing $L = 1$ in practice as a rule of thumb.

4 Finite Sample Performance

To evaluate the finite sample performance of the proposed estimators, the following data generating process (DGP) is employed:

$$Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it},$$

where $[\beta_1, \beta_2, \beta_3] = [1, 1, 1]$, $\alpha_i \sim$ i.i.d $\mathcal{N}(0, 1)$, $\gamma_i \sim$ i.i.d $\mathcal{N}(0, 1)$, $f_t \sim$ i.i.d $\mathcal{N}(0, 1)$, $X_{it,1} \sim$ i.i.d $\chi^2(1) + 1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d $\mathcal{N}(1, 1)$. Since the asymptotic results are conditional on the fixed effects, only $X_{it,1}, e_{2,it}, e_{3,it}, \epsilon_{it}$ vary across repetitions. The distributions of $e_{2,it}, e_{3,it}, \epsilon_{it}$ are specified in each subsection below. Throughout this section, the kernel function $k(\cdot)$ and the bandwidth parameter h are chosen as mentioned in Section 3.6.

In this DGP, there are two common factors: 1 and f_t and the factor loading is given by $\lambda_i = [\alpha_i, \gamma_i]'$. Section 3.1 below examines the estimation of r, while Section 3.2 and Section 3.3 focus on the estimation of the coefficient of $X_{it,1}$, which varies across different quantiles.

4.1 The Number of Factors

The performance of the estimator for the number of factor depends crucially on the properties of $e_{j,it}$. Following Bai and Ng (2002), I consider the following DGP for $e_{j,it}$:

$$e_{j,it} = \gamma e_{j,it-1} + \nu_{j,it} + \zeta \cdot \sum_{l=i-m, l\neq i}^{i+m} \nu_{j,lt}$$

where $\nu_{j,it} \sim$ i.i.d $\mathcal{N}(0, 1)$ for $j = 2, 3$. The parameter γ controls the serial dependence and the parameters ζ, m determine the cross-sectional dependence. The following models are considered in the simulations:

Q1: i.i.d errors: $\gamma = \zeta = 0$.
Q2: serial dependence: $\gamma = 0.2$, $\zeta = 0$.
Q3: cross-sectional dependence: $\gamma = 0$, $\zeta = 0.2$, $m = 5$.
Q4: serial and cross-sectional dependence: $\gamma = 0.2$, $\zeta = 0.2$, $m = 5$.

Recall that the estimator for the number of factors is defined as the number of eigenvalues of $\hat{\Sigma}_{\bar{X}}$ that is larger than $\mathbb{P}_{NT}$. Note that Proposition 1 requires that $\mathbb{P}_{NT} = (\min\{N, T\})^{-c}$ for some $0 < c < 1/2$. Thus, in the simulations I choose $c = 1/3$. Tabel 1 reports the frequencies of choosing the right number of factors (denoted as $\hat{P}[\hat{r} = 2]$) and the mean number of estimated factors (denoted as mean[$\hat{r}$]) from 1000 repetitions for $N, T \in \{20, 50, 100, 200\}$. It can be seen that the proposed method choose the right number of factors in all models with very high precision as long as $\min[N, T] \geq 50$.

4.2 Static Models

In this subsection, $e_{2,it}, e_{3,it}$ are generated as i.i.d standard normal random variables, and I consider two different specifications for the distribution of ϵ_{it}:

M1: $\epsilon_{it} \sim$ i.i.d $\mathcal{N}(0, 1)$.
M2: $\epsilon_{it} \sim$ i.i.d $\mathcal{T}(3)$, where $\mathcal{T}(3)$ denotes the student's t distribution with 3 degrees of freedom.

The main object of interest is the quantile coefficients of $X_{1,it}$ at $\tau = 0.25, 0.9$, and the following three estimators are considered:

$\hat{\beta}(\tau)$: the two-step estimator using SQR.
$\hat{\beta}_{abc}(\tau)$: the bias-corrected two-step estimator using analytical bias correction.
$\hat{\beta}_{spj}(\tau)$: the bias-corrected two-step estimator using the SPJ.

The kernel function and the bandwidth parameter are chosen as mentioned in Section 3.6. Given the excellent performance of the estimated number of factors in the previous subsection I treat the true number of factors as known. The simulation results from 500 repetitions are reported in Table 2, where column 3 to column 5 report the biases the estimators, column 6 to column 8 report the standard deviations, and the last three columns report the coverage rates of the confidence intervals with 95% nominal levels. Note that the DGP in this subsection has no serial correlations. Thus, when constructing the analytical-bias-correction estimators and the confidence intervals, the biases are estiamted by setting $\hat{\omega}^{(2)}_{T,i} = \hat{\omega}^{(4)}_{T,i} = 0$, and the covariance matrices are estimated using the formula given in Section 3.5 with $\hat{\mathbf{V}}_2 = 0$.

There are four main takeaways from the simulation results. First, the biases and the standard deviations of the estimators are larger when the distributions of the idiosyncratic errors have heavier tails (normal v.s. student's t distibutions) and when the quantile of interest is further away from the median ($\tau = 0.9$ v.s. $\tau = 0.25$). This is true for both the original two-step estimators and the bias-corrected estimators.

Second, it is clear from the results that the biases of the estimators decrease either as N increases while T is fixed, or as T increases while N is fixed. This confirms the existence of a leading bias term whose size depends on both N and T, as I have established in Theorem 2. Such results are in contrast with the findings in quantile panel modes with only individual effects, where the leading bias term is approximately of order T^{-1} and thus the biases decrease only when T increases.

Third, for the analytical bias correction to have good performance, the number of time series observations (T) needs to be at least 100. On the other hand, the SPJ perform much better when $T = 50$ because there is no need to estimate those complex objects (such as the inverse of the density functions) when constructing the estimators of the biases.

Last but not least, it can be seen that both analytical and the SPJ bias corrections can significantly reduce the biases of the two-step estimator, as predicted by my theoretical results. However, the reduction of biases comes at the cost of inflating the standard deviations — this is especially noticeable for the analytical bias correction when $T = 50$. As a consequence, the coverage rates of the confidence intervals based on the bias-corrected estimators are in general lower than those based on the original two-step estimators. Therefore, different from the usual suggestion of applying bias correction technique to the fixed-effects estimator of nonlinear panel data models (including quantile panel data models) to improve finite sample performance, for the models considered in this paper the important lesson we can learn is that bias correction can be harmful and it is actually better to use the original estimator (without bias-correction) to achieve better finite sample performance.

4.3 Dynamic Models

In this subsection, I consider dynamic models where ϵ_{it} are generated as autoregressive processes:

$$\epsilon_{it} = \rho \cdot \epsilon_{i,t-1} + \sqrt{1-\rho^2} \cdot \nu_{it}, \text{ where } \nu_{it} \sim \text{i.i.d } \mathcal{N}(0,1)$$

As in the previous subsection, $e_{2,it}$ and $e_{3,it}$ are i.i.d standard normal variables. Now, $\hat{\omega}^{(2)}_{T,i}$, $\hat{\omega}^{(4)}_{T,i}$ and $\hat{\mathbf{V}}_2$ are estimated by the formulas given in Section 3.4 and 3.5. As discussed in Section 3.6, I focus on the choice of $L = 1, 2$. The results with weak serial correlation ($\rho = 0.2$) and moderate serial correlation ($\rho = 0.5$) are reported in Table 3 and Table 4 respectively.

In general, except for a few cases where the standard deviation of $\hat{\beta}_{abc}$ is extremely large, which usually happen when $T = 50$, the results are very similar to those reported in Table 2 for the static cases. In particular, changing the truncation parameter L from 1 to 2 does not significantly improve the finite sample performance of the estimators. This is also true if I allow L to increase with sample sizes (more simulation results are available upon request). Thus, as mentioned in Section 3.6, $L = 1$ is recommended as a rule of thumb for practitioners.

5 Conclusions

Estimating the coefficients of the regressors and the interactive fixed effects jointly in a quantile panel model is not only computationally difficult but also theoretically challenging to derive the asymptotic properties of the estimators, mainly due to the fact that the object function is non-smooth and non-convex. In this paper, I propose a two-step estimator that is easy to implement in practice. Because I use smoothed quantile regressions in the second step, the derivation of the asymptotic distribution and the asymptotic biases of the estimator is possible. The asymptotic distribution provides a formal justification for the use of analytical bias correction and a heuristic argument for the use of the SPJ to correct the asymptotic biases, and the simulation results confirm that both bias-correction methods can effectively reduce the biases with moderate sample sizes. However, it should be cautioned that the bias correction methods inevitably inflate the standard deviations of the estimators, and result in confidence intervals with lower coverage rate than the estimators without bias correction. Finally, even though I have provided conditions with regard to the sizes of the bandwidth parameter in SQR and the truncation parameter in the HAC-type estimators of the bias and variance, there remains the important but challenging question of how to choose these parameters optimally in a data-dependent manner. Such an interesting question is left for future research.

Table 1: The Number of Factors.

(N,T)	**Q1**		**Q2**		**Q3**		**Q4**	
	$\hat{P}[\hat{r}=2]$	mean$[\hat{r}]$	$\hat{P}[\hat{r}=2]$	mean$[\hat{r}]$	$\hat{P}[\hat{r}=2]$	mean$[\hat{r}]$	$\hat{P}[\hat{r}=2]$	mean$[\hat{r}]$
(20,20)	0.997	1.997	0.997	1.997	0.996	1.996	0.997	1.997
(20,50)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(20,100)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(20,200)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(50,20)	0.002	1.002	0.003	1.003	0.064	1.064	0.075	1.075
(50,50)	1.000	2.000	1.000	2.000	1.000	2.000	0.999	1.999
(50,100)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(50,200)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(100,20)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(100,50)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(100,100)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(100,200)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,20)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,50)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,100)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000
(200,200)	1.000	2.000	1.000	2.000	1.000	2.000	1.000	2.000

Note: 1000 repetitions. DGP: $f_t \sim$ i.i.d $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d $\chi^2(1)+1$, and $X_{it,2} = \theta_{2i}+\eta_{2i}f_t+e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i}f_t + e_{3,it}$, where $e_{j,it} = \gamma e_{j,it-1} + \nu_{j,it} + \zeta \cdot \sum_{l=i-m,l\neq i}^{i+m} \nu_{j,lt}$, $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d $\mathcal{N}(1,1)$, $\nu_{2,it}, \nu_{3,it} \sim$ i.i.d $\mathcal{N}(0,1)$. **Q1:** $\gamma = \zeta = 0$; **Q2:** $\gamma = 0.2$, $\zeta = 0$; **Q3:** $\gamma = 0$, $\zeta = 0.2$, $m = 5$; **Q4:** $\gamma = 0.2$, $\zeta = 0.2$, $m = 5$. The above table reports the frequencies of choosing the right number of factors (denoted as $\hat{P}[\hat{r} = r]$) and the mean number of estimated factors (denoted as mean$[\hat{r}]$).

Table 2: Static Models.

M1		Bias			Std			Coverage Rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.024	0.004	0.005	0.086	0.484	0.111	0.908	0.596	0.852
	(50,100)	0.016	0.005	-0.011	0.062	0.081	0.076	0.910	0.860	0.854
	(50,200)	0.007	-0.002	-0.001	0.043	0.049	0.050	0.934	0.910	0.906
	(100,50)	0.019	-0.014	0.001	0.062	0.413	0.074	0.908	0.582	0.898
	(100,100)	0.009	0.005	0.000	0.043	0.063	0.051	0.924	0.814	0.890
	(100,200)	0.003	0.001	-0.001	0.031	0.035	0.034	0.946	0.926	0.920
	(200,50)	0.016	0.017	-0.001	0.043	0.484	0.050	0.924	0.484	0.906
	(200,100)	0.007	0.005	-0.001	0.030	0.050	0.035	0.926	0.820	0.908
	(200,200)	0.004	0.002	0.001	0.022	0.024	0.024	0.946	0.920	0.916
0.9	(50,50)	-0.051	0.004	0.017	0.110	0.812	0.151	0.874	0.676	0.812
	(50,100)	-0.030	-0.022	0.001	0.076	0.097	0.100	0.874	0.808	0.798
	(50,200)	-0.014	-0.010	-0.001	0.054	0.061	0.067	0.914	0.874	0.834
	(100,50)	-0.049	-0.039	0.019	0.074	0.218	0.091	0.864	0.664	0.860
	(100,100)	-0.026	-0.021	0.003	0.055	0.066	0.065	0.884	0.834	0.850
	(100,200)	-0.010	-0.003	0.004	0.038	0.043	0.045	0.898	0.868	0.846
	(200,50)	-0.048	-0.046	-0.013	0.057	0.128	0.069	0.796	0.640	0.866
	(200,100)	-0.021	-0.017	0.004	0.037	0.045	0.044	0.864	0.814	0.862
	(200,200)	-0.012	-0.007	0.001	0.026	0.028	0.030	0.884	0.882	0.880
M2		Bias			Std			Coverage Rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.033	0.077	0.004	0.110	1.502	0.140	0.874	0.604	0.814
	(50,100)	0.020	0.010	0.000	0.075	0.095	0.090	0.884	0.844	0.842
	(50,200)	0.010	-0.000	-0.000	0.053	0.062	0.061	0.928	0.902	0.898
	(100,50)	0.028	0.022	-0.002	0.077	0.341	0.092	0.896	0.606	0.882
	(100,100)	0.012	0.005	-0.004	0.053	0.078	0.061	0.934	0.818	0.902
	(100,200)	0.005	0.001	-0.003	0.036	0.042	0.040	0.950	0.902	0.928
	(200,50)	0.027	0.023	-0.001	0.055	0.214	0.064	0.904	0.522	0.904
	(200,100)	0.010	0.007	-0.004	0.037	0.056	0.042	0.924	0.788	0.902
	(200,200)	0.007	0.002	-0.000	0.026	0.029	0.029	0.946	0.916	0.918
0.9	(50,50)	-0.113	-0.259	-0.006	0.180	3.739	0.250	0.806	0.668	0.782
	(50,100)	-0.061	-0.070	-0.004	0.122	0.183	0.151	0.806	0.750	0.782
	(50,200)	-0.028	-0.027	0.001	0.097	0.111	0.117	0.814	0.784	0.760
	(100,50)	-0.115	-0.113	0.001	0.121	0.266	0.161	0.778	0.656	0.812
	(100,100)	-0.055	-0.050	-0.000	0.095	0.109	0.116	0.768	0.742	0.778
	(100,200)	-0.029	-0.021	0.001	0.066	0.076	0.078	0.808	0.786	0.782
	(200,50)	-0.109	-0.111	0.012	0.094	0.182	0.117	0.688	0.594	0.810
	(200,100)	-0.055	-0.051	0.001	0.067	0.080	0.079	0.726	0.688	0.772
	(200,200)	-0.027	-0.021	0.001	0.044	0.048	0.050	0.786	0.782	0.802

Note: 500 repetitions. DGP: $Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it}$, where $[\beta_1, \beta_2, \beta_3] = [1,1,1]$, $\alpha_i \sim$ i.i.d $\mathcal{N}(0,1)$, $\gamma_i \sim$ i.i.d $\mathcal{N}(0,1)$, $f_t \sim$ i.i.d $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d $\chi^2(1) + 1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d $\mathcal{N}(1,1)$, $e_{2,it}, e_{3,it} \sim$ i.i.d $\mathcal{N}(0,1)$. **M1**: $\epsilon_{it} \sim$ i.i.d $\mathcal{N}(0,1)$; **M2**: $\epsilon_{it} \sim$ i.i.d $\mathcal{T}(3)$.

Table 3: Dynamic Models with $\rho = 0.2$.

$L = 1$		Bias			Std			Coverage Rate (95%)		
τ	(N, T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.024	0.004	0.000	0.083	1.244	0.105	0.914	0.594	0.878
	(50,100)	0.016	0.003	0.001	0.061	0.088	0.073	0.904	0.832	0.864
	(50,200)	0.009	0.001	-0.000	0.046	0.055	0.053	0.916	0.866	0.890
	(100,50)	0.018	0.008	-0.005	0.061	0.655	0.073	0.940	0.548	0.916
	(100,100)	0.012	0.008	0.002	0.041	0.073	0.049	0.948	0.822	0.918
	(100,200)	0.003	-0.001	-0.002	0.029	0.034	0.033	0.948	0.914	0.918
	(200,50)	0.018	0.031	-0.003	0.044	0.236	0.052	0.922	0.564	0.902
	(200,100)	0.010	0.004	0.000	0.030	0.044	0.034	0.944	0.830	0.904
	(200,200)	0.004	0.001	-0.001	0.021	0.024	0.023	0.940	0.912	0.918
0.9	(50,50)	-0.060	-14.137	0.013	0.113	315.845	0.153	0.840	0.604	0.794
	(50,100)	-0.030	-0.026	0.003	0.075	0.106	0.097	0.858	0.770	0.794
	(50,200)	-0.009	-0.004	0.008	0.051	0.062	0.062	0.926	0.874	0.880
	(100,50)	-0.050	-0.056	0.015	0.081	0.238	0.105	0.838	0.626	0.806
	(100,100)	-0.026	-0.020	0.002	0.056	0.074	0.068	0.842	0.778	0.830
	(100,200)	-0.013	-0.006	0.001	0.038	0.044	0.043	0.906	0.876	0.884
	(200,50)	0.049	-0.046	0.020	0.055	0.117	0.069	0.820	0.648	0.840
	(200,100)	-0.027	-0.021	0.027	0.041	0.052	0.048	0.794	0.746	0.836
	(200,200)	-0.013	-0.007	0.001	0.026	0.029	0.030	0.880	0.890	0.870
$L = 2$		Bias			Std			Coverage Rate (95%)		
τ	(N, T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.024	-0.011	0.000	0.083	1.037	0.105	0.914	0.614	0.878
	(50,100)	0.016	0.003	0.001	0.061	0.088	0.073	0.902	0.838	0.862
	(50,200)	0.009	0.001	-0.000	0.046	0.055	0.053	0.918	0.862	0.890
	(100,50)	0.018	0.005	-0.005	0.061	0.600	0.073	0.940	0.554	0.918
	(100,100)	0.012	0.009	0.002	0.041	0.074	0.049	0.948	0.814	0.918
	(100,200)	0.003	-0.001	-0.002	0.029	0.034	0.033	0.950	0.916	0.920
	(200,50)	0.018	0.034	-0.003	0.044	0.262	0.052	0.926	0.560	0.898
	(200,100)	0.010	0.004	0.000	0.030	0.045	0.034	0.942	0.836	0.904
	(200,200)	0.004	0.001	-0.001	0.021	0.024	0.023	0.942	0.914	0.918
0.9	(50,50)	-0.060	-12.934	0.013	0.113	288.685	0.153	0.838	0.604	0.790
	(50,100)	-0.003	-0.026	0.003	0.075	0.106	0.097	0.858	0.766	0.800
	(50,200)	-0.009	-0.004	0.008	0.051	0.062	0.062	0.930	0.874	0.882
	(100,50)	-0.050	-0.057	0.015	0.081	0.243	0.105	0.840	0.614	0.802
	(100,100)	-0.026	-0.021	0.002	0.056	0.074	0.068	0.842	0.778	0.830
	(100,200)	-0.013	-0.006	0.001	0.038	0.044	0.043	0.908	0.880	0.886
	(200,50)	-0.049	-0.045	0.020	0.055	0.117	0.069	0.820	0.658	0.838
	(200,100)	-0.027	-0.021	0.003	0.041	0.052	0.048	0.796	0.750	0.836
	(200,200)	-0.013	-0.007	0.001	0.026	0.029	0.030	0.880	0.890	0.868

Note: 500 repetitions. DGP: $Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it}$, where $[\beta_1, \beta_2, \beta_3] = [1, 1, 1]$, $\alpha_i \sim$ i.i.d $\mathcal{N}(0, 1)$, $\gamma_i \sim$ i.i.d $\mathcal{N}(0, 1)$, $f_t \sim$ i.i.d $\mathcal{N}(0, 1)$, $X_{it,1} \sim$ i.i.d $\chi^2(1) + 1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d $\mathcal{N}(1, 1)$, $e_{2,it}, e_{3,it} \sim$ i.i.d $\mathcal{N}(0, 1)$. $\epsilon_{it} = \rho \cdot \epsilon_{i,t-1} + \sqrt{1 - \rho^2} \cdot \nu_{it}$, where $\nu_{it} \sim$ i.i.d $\mathcal{N}(0, 1)$ and $\rho = 0.2$.

Table 4: Dynamic Models with $\rho = 0.5$.

$L=1$		Bias			Std			Coverage Rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.026	-0.010	-0.007	0.087	1.141	0.112	0.904	0.544	0.854
	(50,100)	0.021	0.009	0.002	0.063	0.088	0.073	0.924	0.844	0.886
	(50,200)	0.012	0.005	0.004	0.047	0.056	0.052	0.922	0.880	0.898
	(100,50)	0.024	0.005	-0.001	0.062	0.419	0.076	0.930	0.542	0.894
	(100,100)	0.017	0.009	0.004	0.042	0.077	0.050	0.940	0.776	0.916
	(100,200)	0.005	0.001	-0.001	0.031	0.039	0.036	0.948	0.914	0.916
	(200,50)	0.024	0.013	-0.002	0.045	0.290	0.053	0.908	0.508	0.894
	(200,100)	0.012	0.006	-0.000	0.032	0.050	0.037	0.932	0.794	0.902
	(200,200)	0.006	0.003	0.000	0.023	0.027	0.025	0.934	0.880	0.914
0.9	(50,50)	-0.076	-0.072	0.013	0.110	0.473	0.154	0.824	0.582	0.798
	(50,100)	-0.040	-0.033	-0.001	0.074	0.099	0.092	0.848	0.788	0.810
	(50,200)	-0.012	-0.006	0.005	0.052	0.064	0.062	0.916	0.882	0.886
	(100,50)	-0.063	-0.050	0.015	0.082	0.414	0.107	0.814	0.612	0.814
	(100,100)	-0.035	-0.030	0.000	0.058	0.083	0.071	0.822	0.774	0.808
	(100,200)	-0.018	-0.009	0.002	0.039	0.044	0.045	0.876	0.868	0.870
	(200,50)	-0.065	-0.065	0.011	0.055	0.151	0.071	0.728	0.584	0.836
	(200,100)	-0.032	-0.024	0.004	0.040	0.048	0.049	0.786	0.792	0.840
	(200,200)	-0.015	-0.008	0.002	0.028	0.031	0.031	0.854	0.852	0.856
$L=2$		Bias			Std			Coverage Rate (95%)		
τ	(N,T)	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$	$\hat{\beta}(\tau)$	$\hat{\beta}_{abc}(\tau)$	$\hat{\beta}_{spj}(\tau)$
0.25	(50,50)	0.026	-0.001	-0.007	0.087	1.120	0.112	0.906	0.522	0.862
	(50,100)	0.021	0.008	0.002	0.063	0.091	0.073	0.930	0.842	0.890
	(50,200)	0.012	0.005	0.004	0.047	0.057	0.052	0.926	0.886	0.902
	(100,50)	0.024	0.006	-0.001	0.062	0.411	0.076	0.938	0.520	0.898
	(100,100)	0.017	0.008	0.004	0.042	0.083	0.050	0.940	0.766	0.918
	(100,200)	0.005	0.001	-0.001	0.031	0.040	0.036	0.952	0.914	0.916
	(200,50)	0.024	0.010	-0.002	0.045	0.312	0.053	0.914	0.508	0.898
	(200,100)	0.012	0.005	-0.000	0.032	0.052	0.037	0.934	0.776	0.912
	(200,200)	0.006	0.003	0.000	0.023	0.027	0.025	0.936	0.876	0.916
0.9	(50,50)	-0.076	-0.073	0.013	0.110	0.474	0.154	0.828	0.574	0.804
	(50,100)	-0.040	-0.033	-0.001	0.074	0.101	0.092	0.850	0.780	0.814
	(50,200)	-0.012	-0.005	0.005	0.052	0.064	0.062	0.877	0.916	0.901
	(100,50)	-0.063	-0.049	0.015	0.082	0.461	0.107	0.818	0.616	0.808
	(100,100)	-0.035	-0.030	0.000	0.058	0.084	0.071	0.824	0.762	0.816
	(100,200)	-0.018	-0.009	0.002	0.039	0.044	0.045	0.878	0.864	0.870
	(200,50)	-0.065	-0.063	0.011	0.055	0.161	0.071	0.736	0.562	0.844
	(200,100)	-0.032	-0.024	0.004	0.040	0.049	0.049	0.790	0.808	0.844
	(200,200)	-0.015	-0.008	0.002	0.028	0.031	0.031	0.860	0.856	0.858

Note: 500 repetitions. DGP: $Y_{it} = \beta_1 X_{it,1} + \beta_2 X_{it,2} + \beta_3 X_{it,3} + \alpha_i + \gamma_i f_t + X_{it,1} \cdot \epsilon_{it}$, where $[\beta_1, \beta_2, \beta_3] = [1,1,1]$, $\alpha_i \sim$ i.i.d $\mathcal{N}(0,1)$, $\gamma_i \sim$ i.i.d $\mathcal{N}(0,1)$, $f_t \sim$ i.i.d $\mathcal{N}(0,1)$, $X_{it,1} \sim$ i.i.d $\chi^2(1) + 1$, and $X_{it,2} = \theta_{2i} + \eta_{2i} f_t + e_{2,it}$, $X_{it,3} = \theta_{3i} + \eta_{3i} f_t + e_{3,it}$, where $\theta_{2i}, \theta_{3i}, \eta_{2i}, \eta_{3i} \sim$ i.i.d $\mathcal{N}(1,1)$, $e_{2,it}, e_{3,it} \sim$ i.i.d $\mathcal{N}(0,1)$. $\epsilon_{it} = \rho \cdot \epsilon_{i,t-1} + \sqrt{1-\rho^2} \cdot \nu_{it}$, where $\nu_{it} \sim$ i.i.d $\mathcal{N}(0,1)$ and $\rho = 0.5$.

References

Abrevaya, J. and C. M. Dahl (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics 26*(4), 379–397.

Amemiya, T. (1982). Two stage least absolute deviations estimators. *Econometrica*, 689–711.

Ando, T. and J. Bai (2020). Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity. *Journal of the American Statistical Association 115*(529), 266–279.

Arellano, M. and S. Bonhomme (2016). Nonlinear panel data estimation via quantile regressions. *The Econometrics Journal 19*(3), C61–C94.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

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

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

Belloni, A., M. Chen, and O. H. M. Padilla (2019). High dimensional latent panel quantile regression with an application to asset pricing. working paper.

Cai, Z., L. Chen, and Y. Fang (2018). A semiparametric quantile panel data model with an application to estimating the growth effect of FDI. *Journal of Econometrics 206*(2), 531–553.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal 14*(3), 368–386.

Chen, L. (2015). Set identification of panel data models with interactive effects via quantile restrictions. *Economics Letters 137*, 36 – 40.

Chen, L. (2019). Nonparametric quantile regressions for panel data models with large T. working paper.

Chen, L., J. Dolado, and J. Gonzalo (2019). Quantile factor models. working paper.

Chen, L. and Y. Huo (2020). A simple estimator for quantile panel data models using smoothed quantile regressions. *The Econometrics Journal (forthcoming).*

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

Chudik, A. and H. Pesaran (2015). Large panel data models with cross-sectional dependence. In *The Oxford Handbook of Panel Data.*

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies 82*(3), 991–1030.

Feng, J. (2019). Regularized quantile regression with interactive fixed effects. working paper.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics 192*(1), 291 – 312.

Galvao, A. F. (2011). Quantile regression for dynamic panel data with fixed effects. *Journal of Econometrics 164*(1), 142–157.

Galvao, A. F., J. Gu, and S. Volgushev (2020). On the unbiased asymptotic normality of quantile regression with fixed effects. *Journal of Econometrics (forthcoming).*

Galvao, A. F. and K. Kato (2016). Smoothed quantile regression for panel data. *Journal of Econometrics 193*(1), 92–112.

Galvao, A. F., C. Lamarche, and L. R. Lima (2013). Estimation of censored quantile regression for panel data with fixed effects. *Journal of the American Statistical Association 108*(503), 1075–1089.

Galvao, A. F. and G. V. Montes-Rojas (2010). Penalized quantile regression for dynamic panel data. *Journal of Statistical Planning and Inference 140*(11), 3476–3497.

Graham, B. S., J. Hahn, A. Poirier, and J. L. Powell (2018). A quantile correlated random coefficients panel data model. *Journal of Econometrics 206*(2), 305–335.

Hahn, J. and G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory 27*(6), 1152–1191.

Harding, M. and C. Lamarche (2014). Estimating and testing a quantile regression model with interactive effects. *Journal of Econometrics 178*, 101–113.

Harding, M., C. Lamarche, and M. H. Pesaran (2020). Common correlated effects estimation of heterogeneous dynamic panel quantile regression models. *Journal of Applied Econometrics 35*(3), 294–314.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 1327–1351.

Karabiyik, H., S. Reese, and J. Westerlund (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics 197*(1), 60 – 64.

Kato, K., A. F. Galvao, and G. V. Montes-Rojas (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics 170*(1), 76–91.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91*(1), 74–89.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics 157*(2), 396–408.

Lu, X. and L. Su (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics 190*(1), 148–175.

Ma, S., O. Linton, and J. Gao (2020). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics (forthcoming).*

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

Muller, H. G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics 12*(2), 766–774.

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

Rosen, A. M. (2012). Set identification via quantile restrictions in short panels. *Journal of Econometrics 166*(1), 127–137.

Yoon, J. and A. F. Galvao (2020). Cluster robust covariance matrix estimation in panel quantile regression with individual fixed effects. *Quantitative Economics 11*(2), 579–608.

Online Appendix to "Two-Step Estimation of Quantile Panel Data Models with Interactive Fixed Effects"

Liang Chen

Peking University HSBC Business School, No.2199 Lishui Road, Shenzhen, Guangdong 518055, China
Email: chenliang@phbs.pku.edu.cn.

January 5, 2021

A Proof of the Main Resutls

Definitions and Notations

For any random variable W_i or W_t, $W_i = \bar{O}_P(1)$ means that $\max_{1\leq i\leq N}\|W_i\| = O_P(1)$, and $W_t = \bar{O}_P(1)$ means that $\max_{1\leq t\leq T}\|W_t\| = O_P(1)$. $\bar{o}_P(1)$ is defined similarly. For notational simplicity, we suppress the dependence of $\lambda_{0i}(\tau)$, $\hat{\beta}(\tau)$ and $\beta_0(\tau)$ on τ. Let $M > 0$ denote a generic bounded constant that does not depend on N or T.

Define:

$$\mathbb{S}_{NT}(\beta,\Lambda,F) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\beta,\lambda_i,f_t) \quad \mathbb{S}^*_{NT}(\beta,\Lambda,F) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{it}(\beta,\lambda_i,f_t)$$

$$\mathbb{S}_{i,T}(\beta,\lambda_i,F) = \frac{1}{T}\sum_{t=1}^{T} l_{it}(\beta,\lambda_i,f_t) \quad \mathbb{S}^*_{i,T}(\beta,\lambda_i,F) = \frac{1}{T}\sum_{t=1}^{T}\rho_{it}(\beta,\lambda_i,f_t)$$

where $l(u) = [\tau - K(u/h)]u$, $\rho_\tau(u)$ is the check function, and

$$l_{it}(\beta,\lambda_i,f_t) = l(Y_{it} - \beta' X_{it} - \lambda_i' f_t), \quad \rho_{it}(\beta,\lambda_i,f_t) = \rho_\tau(Y_{it} - \beta' X_{it} - \lambda_i' f_t).$$

For any random function $L(\beta,\Lambda,F)$ and fixed (β,Λ,F), define $\bar{L}(\beta,\Lambda,F) = \mathbb{E}[L(\beta,\Lambda,F)]$ and $\tilde{L}(\beta,\Lambda,F) = L(\beta,\Lambda,F) - \bar{L}(\beta,\Lambda,F)$. Let $l^{(j)}(u)$ denote the jth order derivative of l, i.e.,

$$l^{(1)}(u) = \tau - K(u/h) + k(u/h)u/h, \quad l^{(2)}(u) = 2k(u/h)1/h + k^{(1)}(u/h)u/h^2,$$

$$l^{(3)}(u) = 3k^{(1)}(u/h)1/h^2 + k^{(2)}(u/h)u/h^3, \quad l^{(4)}(u) = 4k^{(2)}(u/h)1/h^3 + k^{(3)}(u/h)u/h^4.$$

Let $l_{it}^{(j)}(\beta, \lambda_i, f_t) = l^{(j)}(Y_{it} - \beta' X_{it} - \lambda_i' f_t)$ for $j = 1, \ldots, 4$, and their arguments are dropped when evaluated at $(\beta_0, \lambda_{0i}, f_{0t})$.

Finally, define $\tilde{\lambda}_{0i} = (\mathbf{H}_0')^{-1}\lambda_{0i}$, $\tilde{f}_{0t} = \mathbf{H}_0 f_{0t}$, $\tilde{\Lambda}_0 = [\tilde{\lambda}_{01}, \ldots, \tilde{\lambda}_{0N}]'$, and $\tilde{F}_0 = [\tilde{f}_{01}, \ldots, \tilde{f}_{0T}]'$.

A.1 Proof of Proposition 1:

Proof. Note that

$$\begin{aligned}
&\|\hat{\Sigma}_{\bar{x}} - \Gamma_0 \Sigma_{f_0} \Gamma_0'\| \\
=& \left\| \bar{\Gamma} \hat{\Sigma}_{f_0} \bar{\Gamma}' - \Gamma_0 \Sigma_{f_0} \Gamma_0' + \bar{\Gamma} \cdot \frac{1}{T} \sum_{t=1}^{T} f_{0t} \bar{e}_t' + \frac{1}{T} \sum_{t=1}^{T} \bar{e}_t f_{0t}' \cdot \bar{\Gamma}' + \frac{1}{T} \sum_{t=1}^{T} \bar{e}_t \bar{e}_t' \right\| \\
\leq& \left\| \bar{\Gamma} \hat{\Sigma}_{f_0} \bar{\Gamma}' - \Gamma_0 \Sigma_{f_0} \Gamma_0' \right\| + 2 \left\| \bar{\Gamma} \cdot \frac{1}{T} \sum_{t=1}^{T} f_{0t} \bar{e}_t' \right\| + \left\| \frac{1}{T} \sum_{t=1}^{T} \bar{e}_t \bar{e}_t' \right\|.
\end{aligned}$$

First, by Assumption 1(ii), $\left\| \bar{\Gamma} \hat{\Sigma}_{f_0} \bar{\Gamma}' - \Gamma_0 \Sigma_{f_0} \Gamma_0' \right\| = O(N^{-1/2} + T^{-1/2})$. Second, by Assumption 1(iii) and the Cauchy-Schwarz inequality, we have

$$\left\| \frac{1}{T} \sum_{t=1}^{T} f_{0t} \bar{e}_t' \right\| \leq \frac{1}{\sqrt{N}} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \|f_{0t}\|^2} \sqrt{\frac{1}{T} \sum_{t=1}^{T} \|\sqrt{N} \bar{e}_t\|^2} = O_P(N^{-1/2})$$

and

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \bar{e}_t \bar{e}_t' \right\| \leq \frac{1}{N} \cdot \frac{1}{T} \sum_{t=1}^{T} \|\sqrt{N} \bar{e}_t\|^2 = O_P(N^{-1}).$$

It then follows that $\|\hat{\Sigma}_{\bar{x}} - \Gamma_0 \Sigma_{f_0} \Gamma_0'\| = O_P(N^{-1/2} + T^{-1/2})$. Third, by matrix perturbation theory (Hoffman-Wielandt inequality) and the fact that $\Gamma_0 \Sigma_{f_0} \Gamma_0'$ is a matrix with rank r (Assumption 1(ii)), it can be concluded that $\hat{\rho}_1, \ldots, \hat{\rho}_r$ converge in probability to some positive constants, while $\hat{\rho}_{r+1}, \ldots, \hat{\rho}_k$ are all $O_P(N^{-1/2} + T^{-1/2})$. Thus, it follows that

$$P[\hat{r} \neq r] \leq P[\hat{r} < r] + P[\hat{r} > r] \leq P[\hat{\rho}_r < \mathbb{P}_{NT}] + P[\hat{\rho}_{r+1} \geq \mathbb{P}_{NT}] \to 0,$$

and the desired result follows. □

A.2 Proof of Theorem 1

Lemma 1. *Define* $\hat{\mathbf{H}} = \hat{\Psi}' \bar{\Gamma}$. *Under Assumptions 1 and 2, (i)* $\hat{f}_t = \hat{\mathbf{H}} f_{0t} + \hat{\Psi}' \bar{e}_t$*; (ii)* $\hat{\Psi} \xrightarrow{p} \Psi_0$ *and* $\hat{\mathbf{H}} \xrightarrow{p} \mathbf{H}_0 = \Psi_0' \Gamma_0$*; (iii)* $\hat{\mathbf{H}}$ *is invertible with probability approaching 1.*

Proof. Result (i) follows directly from $\bar{X}_t = \bar{\Gamma} f_{0t} + \bar{e}_t$ and $\hat{f}_t = \hat{\Psi}' \bar{X}_t$. Given Assumption 2(i), it follows from the Bauer-Fike Theorem and $\|\hat{\Sigma}_{\bar{x}} - \Gamma_0 \Sigma_{f_0} \Gamma_0'\| = o_P(1)$ (see the proof of Proposition 1) that $\|\hat{\Psi} - \Psi_0\| = o_P(1)$. Thus, result (ii) follows since $\|\bar{\Gamma} - \Gamma_0\| = o(1)$ by Assumption 1(ii). Finally, suppose that $\text{rank}(\mathbf{H}_0) < r$, and let $\mathcal{D}$ be the diagonal matrix with the eigenvalues of $\Gamma_0 \Sigma_{f_0} \Gamma_0'$ as the diagonal elements, then $\text{rank}(\mathcal{D}) = \text{rank}(\Psi_0'\Gamma_0\Sigma_{f_0}\Gamma_0'\Psi_0) = \text{rank}(\mathbf{H}_0\Sigma_{f_0}\mathbf{H}_0') \le \text{rank}(\mathbf{H}_0) < r$, which contradicts with Assumption 1(ii). Thus, we have $\text{rank}(\mathbf{H}_0) = r$ and result (iii) follows from $\hat{\mathbf{H}} = \mathbf{H}_0 + o_P(1)$. □

Proof of Theorem 1

Proof. **Step 1:** By Lemma 1, we have

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\|^2 \le \frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t - \hat{\mathbf{H}}f_{0t}\|^2 + M\|\mathbf{H}_0 - \hat{\mathbf{H}}\|^2 \le \frac{r}{NT}\sum_{t=1}^{T}\|\sqrt{N}\bar{e}_t\|^2 + o_P(1) = O_P\left(\frac{1}{N}\right) + o_P(1).$$

Step 2: Adding and subtracting terms, we can write

$$\mathbb{S}_{NT}(\beta,\Lambda,F) = (\mathbb{S}_{NT}(\beta,\Lambda,F) - \mathbb{S}^*_{NT}(\beta,\Lambda,F)) + (\mathbb{S}^*_{NT}(\beta,\Lambda,F) - \mathbb{S}^*_{NT}(\beta,\Lambda,\tilde{F}_0)) + \mathbb{S}^*_{NT}(\beta,\Lambda,\tilde{F}_0).$$

By the definition of the estimators, $\mathbb{S}_{NT}(\hat{\beta},\hat{\Lambda},\hat{F}) \le \mathbb{S}_{NT}(\beta_0,\tilde{\Lambda}_0,\hat{F})$. Thus, we have

$$\begin{aligned}\mathbb{S}^*_{NT}(\hat{\beta},\hat{\Lambda},\tilde{F}_0) - \mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0) \le{}& [\mathbb{S}_{NT}(\beta_0,\tilde{\Lambda}_0,\hat{F}) - \mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\hat{F})] + [\mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\hat{F}) - \mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0)]\\ &- [\mathbb{S}_{NT}(\hat{\beta},\hat{\Lambda},\hat{F}) - \mathbb{S}^*_{NT}(\hat{\beta},\hat{\Lambda},\hat{F})] - [\mathbb{S}^*_{NT}(\hat{\beta},\hat{\Lambda},\hat{F}) - \mathbb{S}^*_{NT}(\hat{\beta},\hat{\Lambda},\tilde{F}_0)]. \quad (A.1)\end{aligned}$$

Step 3: Let δ be a positive number close to 0. Define $B_{\delta,i} = \{\beta \in \mathcal{B}, \lambda_i \in \mathcal{A} : \|\beta - \beta_0\|_1 + \|\lambda_i - \tilde{\lambda}_{0i}\|_1 \le \delta\}$. Consider any $(\beta,\lambda_i) \in B^C_{\delta,i}$. Let $m = \|\beta-\beta_0\|_1 + \|\lambda_i - \tilde{\lambda}_{0i}\|_1 > \delta$, then $(\bar{\beta},\bar{\lambda}_i) = (\beta,\lambda_i)\delta/m + (\beta_0,\tilde{\lambda}_{0i})(1-\delta/m)$ is on the boundary of $B_{\delta,i}$. Note that given X_{it} and f_t, the check function ρ_{it} is convex in (β,λ_i). Thus,

$$\delta/m \cdot \rho_{it}(\beta,\lambda_i,\tilde{f}_{0t}) + (1-\delta/m)\cdot\rho_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t}) \ge \rho_{it}(\bar{\beta},\bar{\lambda}_i,\tilde{f}_{0t}),$$

and it follows that

$$\rho_{it}(\beta,\lambda_i,\tilde{f}_{0t}) - \rho_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t}) \ge m/\delta\cdot\left[\rho_{it}(\bar{\beta},\bar{\lambda}_i,\tilde{f}_{0t}) - \rho_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t})\right]$$

and

$$\mathbb{S}^*_{NT}(\beta,\Lambda,\tilde{F}_0) - \mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0) \gtrsim \mathbb{S}^*_{NT}(\bar{\beta},\bar{\Lambda},\tilde{F}_0) - \mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0)$$

if $(\beta,\lambda_i) \in B^C_{\delta,i}$ for all i.

Write

$$\mathbb{S}^*_{NT}(\bar{\beta},\bar{\Lambda},\tilde{F}_0) - \mathbb{S}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0) = \bar{\mathbb{S}}^*_{NT}(\bar{\beta},\bar{\Lambda},\tilde{F}_0) - \bar{\mathbb{S}}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0) + \tilde{\mathbb{S}}^*_{NT}(\bar{\beta},\bar{\Lambda},\tilde{F}_0) - \tilde{\mathbb{S}}^*_{NT}(\beta_0,\tilde{\Lambda}_0,\tilde{F}_0).$$

First, by Taylor expansion $\bar\rho_{it}(\beta,\lambda_i,\tilde f_{0t})$ around $(\beta_0,\tilde\lambda_{0i})$ of we have

$$\bar{\mathbb{S}}^*_{NT}(\bar\beta,\bar\Lambda,\tilde F_0)-\bar{\mathbb{S}}^*_{NT}(\beta_0,\tilde\Lambda_0,\tilde F_0)=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\bar\rho_{it}(\bar\beta,\bar\lambda_i,\tilde f_{0t})-\bar\rho_{it}(\beta_0,\tilde\lambda_{0i},\tilde f_{0t})$$
$$=\frac{1}{N}\sum_{i=1}^N[(\bar\beta-\beta_0)',(\bar\lambda_i-\lambda_{0i})]\cdot\left(\frac{1}{T}\sum_{t=1}^T\mathbb{E}[\mathsf{f}_{it}(0|X_{it})V_{it}V_{it}']\right)\cdot[(\bar\beta-\beta_0)',(\bar\lambda_i-\lambda_{0i})]'+o(\delta^2)$$
$$\geq\delta^2\cdot\frac{1}{N}\sum_{i=1}^N\varrho_{i,T}+o(\delta^2)\geq\underline{\varrho}\delta^2 \quad \text{(A.2)}$$

by Assumption 2(iii). Second, we have

$$\left|\tilde{\mathbb{S}}^*_{NT}(\bar\beta,\bar\Lambda,\tilde F_0)-\tilde{\mathbb{S}}^*_{NT}(\beta_0,\tilde\Lambda_0,\tilde F_0)\right|\leq\max_{1\leq i\leq N}\sup_{(\beta,\lambda_i)\in B_{\delta,i}}\left|\frac{1}{T}\sum_{t=1}^T\left[\tilde\rho_{it}(\beta,\lambda_i,\tilde f_{0t})-\tilde\rho_{it}(\beta_0,\tilde\lambda_{0i},\tilde f_{0t})\right]\right| \quad \text{(A.3)}$$

Step 4: $\|\hat\beta-\beta_0\|_1>\delta$ implies that $(\hat\beta,\hat\lambda_i)\in B^C_{\delta,i}$ for all i. It then follows from (A.1), (A.2) and (A.3) that for small $\delta>0$, there exists an $\epsilon>0$ (depending on δ) such that

$$P[\|\hat\beta-\beta_0\|_1>\delta]\leq P\left[\sup_{\beta,\Lambda,F}\left|\mathbb{S}_{NT}(\beta,\Lambda,F)-\mathbb{S}^*_{NT}(\beta,\Lambda,F)\right|>1/3\epsilon\right]$$
$$+P\left[\sup_{\beta,\lambda_i\in\mathcal{A}}\left|\mathbb{S}^*_{NT}(\beta,\Lambda,\hat F)-\mathbb{S}^*_{NT}(\beta,\Lambda,\tilde F_0)\right|>1/3\epsilon\right]$$
$$+P\left[\max_{1\leq i\leq N}\sup_{(\beta,\lambda_i)\in B_{\delta,i}}\left|\frac{1}{T}\sum_{t=1}^T\left[\tilde\rho_{it}(\beta,\lambda_i,\tilde f_{0t})-\tilde\rho_{it}(\beta_0,\tilde\lambda_{0i},\tilde f_{0t})\right]\right|>1/3\epsilon\right]. \quad \text{(A.4)}$$

The first term on the right-hand side of (A.4) is $o(1)$ because it is easy to show that[1]:

$$\sup_{\beta,\Lambda,F}\left|\mathbb{S}_{NT}(\beta,\Lambda,F)-\mathbb{S}^*_{NT}(\beta,\Lambda,F)\right|\lesssim h$$

and $h\to0$ as $N,T\to\infty$. The second term on the right-hand side of (A.4) is $o(1)$ since by the result of Step 1 and Assumption 2(ii),

$$\sup_{\beta,\lambda_i\in\mathcal{A}}\left|\mathbb{S}^*_{NT}(\beta,\Lambda,\hat F)-\mathbb{S}^*_{NT}(\beta,\Lambda,F_0)\right|\leq\max_{1\leq i\leq N}\sup_{\beta,\lambda_i\in\mathcal{A}}\left|\frac{1}{T}\sum_{t=1}^T\rho_{it}(\beta,\lambda_i,\hat f_t)-\frac{1}{T}\sum_{t=1}^T\rho_{it}(\beta,\lambda_i,\tilde f_{0t})\right|$$
$$\lesssim\sqrt{\frac{1}{T}\sum_{t=1}^T\|\hat f_t-\tilde f_{0t}\|^2}=o_P(1).$$

Finally, for the consistency of $\hat\beta$, it remains to show that the third term on the right-hand side of (A.4)

[1] Note that $|l(u)-\rho_\tau(u)|=|(\tau-\mathbf{1}\{u\leq0\})u-(\tau-K(u/h))u|\leq|u|\cdot|\mathbf{1}\{u\leq0\}-K(u/h)|\lesssim|u|\cdot\mathbf{1}\{|u|\leq h\}\lesssim h$.

is $o(1)$. By the union bound, it suffices to show that for all i

$$P\left[\sup_{(\beta,\lambda_i)\in B_{\delta,i}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\rho}_{it}(\beta,\lambda_i,\tilde{f}_{0t})-\tilde{\rho}_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t})\right]\right|>1/3\epsilon\right]=o(N^{-1}). \tag{A.5}$$

Step 5: Write $\theta_i=(\beta',\lambda_i')'$, and $\theta_{0i}=(\beta_0',\tilde{\lambda}_{0i}')'$. Define $\Delta_{it}(\theta_i)=\rho_{it}(\beta,\lambda_i,\tilde{f}_{0t})-\rho_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t})$. Note that there exists $C_1,C_2>0$ such that $|\Delta_{it}(\theta_i^a)-\Delta_{it}(\theta_i^b)|\leq C_1\cdot\|\beta^a-\beta^b\|\cdot\|X_{it}\|+C_2\cdot\|\lambda_i^a-\lambda_i^b\|$. Suppose that there exists $M_X>0$ such that $\mathbb{E}\|X_{it}\|\leq M_X$ for all i,t (see Assumption 2(v)).

Since $B_{\delta,i}$ is compact, for any $\eta>0$, there exists a positive integer L and a maximal set of points $\theta_i^{(1)},\ldots,\theta_i^{(L)}$ in $B_{\delta,i}$ such that $\|\theta_i^{(k)}-\theta_i^{(j)}\|\geq\eta$ for any $k\neq j$. For any $\theta_i\in B_{\delta,i}$, let $\theta_i^*=\{\theta_i^{(j)}:1\leq j\leq L,\|\theta_i-\theta_i^{(j)}\|\leq\eta\}$. Then,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\rho}_{it}(\beta,\lambda_i,\tilde{f}_{0t})-\tilde{\rho}_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t})\right]&=\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i)-\mathbb{E}(\Delta_{it}(\theta_i))\right]\\&=\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^*)-\mathbb{E}(\Delta_{it}(\theta_i^*))\right]+\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i)-\Delta_{it}(\theta_i^*)-\mathbb{E}(\Delta_{it}(\theta_i)-\Delta_{it}(\theta_i^*))\right],\end{aligned}$$

and

$$\begin{aligned}\sup_{(\beta,\lambda_i)\in B_{\delta,i}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\rho}_{it}(\beta,\lambda_i,\tilde{f}_{0t})-\tilde{\rho}_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t})\right]\right|&\leq\max_{1\leq j\leq L}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^{(j)})-\mathbb{E}(\Delta_{it}(\theta_i^{(j)}))\right]\right|\\&+\sup_{\|\theta^a-\theta^b\|\leq\eta}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^a)-\Delta_{it}(\theta_i^b)-\mathbb{E}(\Delta_{it}(\theta_i^a)-\Delta_{it}(\theta_i^b))\right]\right|\end{aligned}$$

Note that

$$\begin{aligned}&\sup_{\|\theta^a-\theta^b\|\leq\eta}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^a)-\Delta_{it}(\theta_i^b)-\mathbb{E}(\Delta_{it}(\theta_i^a)-\Delta_{it}(\theta_i^b))\right]\right|\\&\qquad\leq C_1\eta\left(\frac{1}{T}\sum_{t=1}^{T}(\|X_{it}\|-\mathbb{E}\|X_{it}\|)\right)+2(C_2+C_1M_X)\eta,\end{aligned}$$

it then follows from the previous two inequalities that

$$\begin{aligned}&P\left[\sup_{(\beta,\lambda_i)\in B_{\delta,i}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\rho}_{it}(\beta,\lambda_i,\tilde{f}_{0t})-\tilde{\rho}_{it}(\beta_0,\tilde{\lambda}_{0i},\tilde{f}_{0t})\right]\right|>1/3\epsilon\right]\\&\qquad\leq\sum_{j=1}^{L}P\left[\left|\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^{(j)})-\mathbb{E}(\Delta_{it}(\theta_i^{(j)}))\right]\right|\geq1/9\epsilon\right]\\&\qquad+P\left[C_1\eta\left|\frac{1}{T}\sum_{t=1}^{T}(\|X_{it}\|-\mathbb{E}\|X_{it}\|)\right|\geq1/9\epsilon\right]+P\left[2(C_2+C_1M_X)\eta\geq1/9\epsilon\right].\end{aligned}\tag{A.6}$$

First, choosing $\eta < \epsilon/(18(C_2 + C_1 M_X))$, the last term on the right-hand side of (A.6) is 0.

Second, for any $\theta_i \in B_{\delta,i}$, $\mathbb{E}|\Delta_{it}(\theta_i)|^{2m+\gamma} \leq M \cdot \mathbb{E}\|X_{it}\|^{2m+\gamma} + O(1) < \infty$ by Assumption 2(v). Thus, by Assumption 2(iv) and Theorem 3 of Yoshihara (1978) we have

$$\mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^{(j)}) - \mathbb{E}(\Delta_{it}(\theta_i^{(j)}))\right]\right|^{2m} \leq M,$$

and by Markov's inequality,

$$P\left[\left|\frac{1}{T}\sum_{t=1}^{T}\left[\Delta_{it}(\theta_i^{(j)}) - \mathbb{E}(\Delta_{it}(\theta_i^{(j)}))\right]\right| \geq 1/9\epsilon\right] = O(T^{-m}).$$

Finally, we can show that the second term on the right-hand side of (A.6) is $O(T^{-m})$ in a similar way. Thus, (A.5) follows since $N/T^m \to 0$ by Assumption 2(vi). This completes the proof. □

A.3 Proof of Theorem 2

To simplify the notations, write $\check{f}_{0t} = \hat{\mathbf{H}} f_{0t}$ and $\check{\lambda}_{0i} = (\hat{\mathbf{H}}')^{-1}\lambda_{0i}$.

Lemma 2. *Under Assumptions 1 to 4,*
(i) $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\| = O_P(N^{-1/2})$, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^2 = O_P(N^{-1})$, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^3 = O_P(N^{-3/2})$, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^4 = O_P(N^{-2})$.
(ii) $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\| = O_P(N^{-1/2})$, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\|^2 = O_P(N^{-1})$, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\|^3 = O_P(N^{-3/2})$, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\|^4 = O_P(N^{-2})$.
(iii) $\max_{1\leq t\leq T}\|\hat{f}_t - \check{f}_{0t}\| = O_P(\sqrt{\log T}/\sqrt{N})$ *and* $\max_{1\leq t\leq T}\|\hat{f}_t - \tilde{f}_{0t}\| = O_P(\sqrt{\log T}/\sqrt{N})$.

Proof. By the properties of L_p norms in the Euclidean space, it suffices to prove that $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^4 = O_P(N^{-2})$ and $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\|^4 = O_P(N^{-2})$. Note that Assumption 3(iv) implies that $\{e_{1t}, \ldots, e_{Nt}\}$ is independent across i and $\|e_{it}\| < M$ for all i, t. Thus,

$$\mathbb{E}\|\sqrt{N}\bar{e}_t\|^4 = \mathbb{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\right\|^4 \leq \left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\|e_{it}\|^2\right)^2 + \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\|e_{it}\|^4 = O(1),$$

and by Lemma 1

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^4 \leq \|\hat{\Psi}\|^4 \cdot \frac{1}{N^2 T}\sum_{t=1}^{T}\|\sqrt{N}\bar{e}_t\|^4 = O_P(N^{-2}).$$

Moreover, $T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \tilde{f}_{0t}\|^4 \leq T^{-1}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^4 + C\|\hat{\mathbf{H}} - \mathbf{H}_0\|^4$. Then result (ii) follows if we can show that $\|\hat{\mathbf{H}} - \mathbf{H}_0\| = O_P(N^{-1/2})$.

By definition, $\|\hat{\mathbf{H}} - \mathbf{H}_0\| \leq O_P(\|\hat{\Psi} - \Psi_0\|) + O_P(\|\bar{\Gamma} - \Gamma_0\|)$. By the proof of Proposition 1 and Lemma 1 we have $\|\hat{\Psi} - \Psi_0\| \lesssim \|\hat{\Sigma}_{\bar{x}} - \Gamma_0\Sigma_{f_0}\Gamma_0'\| = O_P(N^{-1/2} + T^{-1/2})$. Then the result (ii) follows from Assumption 1(ii) and the fact that $N \asymp T$.

Finally, note that $\max_{1\leq t\leq T}\|\hat{f}_t - \check{f}_{0t}\| \leq O_P(1) \cdot \max_{1\leq t\leq T}\|\bar{e}_t\|$. For $1 \leq h \leq r$, by the Hoeffding's inequality, we have $P[\sqrt{N}|\bar{e}_{th}| \geq c] \leq \exp(-c^2/C)$ for some constant C. Thus, it follows from Lemma 2.2.1 and Lemma 2.2.2 of van der Vaart and Wellner (1996) that $\mathbb{E}[\max_{1\leq t\leq T}\sqrt{N}|\bar{e}_{th}|] = O(\sqrt{\log T})$. Thus, result (iii) follows since $\|\hat{\mathbf{H}} - \mathbf{H}_0\| = o_P(T^{-1/2})$. □

Lemma 3. *Under Assumptions 1 to 4,* $\max_{1\leq i\leq N}\|\hat{\lambda}_i - \check{\lambda}_{0i}\| = o_P(1)$.

Proof. Recall that $\tilde{\lambda}_{0i} = (\mathbf{H}_0')^{-1}\lambda_{0i}$, $\tilde{f}_{0t} = \mathbf{H}_0 f_{0t}$. By the definition of the estimators we have $\mathbb{S}_{i,T}(\hat{\beta}, \hat{\lambda}_i, \hat{F}) \leq \mathbb{S}_{i,T}(\hat{\beta}, \tilde{\lambda}_{0i}, \hat{F})$ for each i. Note that

$$\mathbb{S}_{i,T}(\beta, \lambda_i, F) = \mathbb{S}^*_{i,T}(\beta_0, \lambda_i, \tilde{F}_0) + (\mathbb{S}_{i,T}(\beta, \lambda_i, F) - \mathbb{S}^*_{i,T}(\beta, \lambda_i, F)) + (\mathbb{S}^*_{i,T}(\beta, \lambda_i, F) - \mathbb{S}^*_{i,T}(\beta_0, \lambda_i, \tilde{F}_0)).$$

Thus, $\mathbb{S}_{i,T}(\hat{\beta}, \hat{\lambda}_i, \hat{F}) \leq \mathbb{S}_{i,T}(\hat{\beta}, \tilde{\lambda}_{0i}, \hat{F})$ implies that

$$\begin{aligned}\mathbb{S}^*_{i,T}(\beta_0, \hat{\lambda}_i, \tilde{F}_0) - \mathbb{S}^*_{i,T}(\beta_0, \tilde{\lambda}_{0i}, \tilde{F}_0) \leq (\mathbb{S}_{i,T}(\hat{\beta}, \tilde{\lambda}_{0i}, \hat{F}) - \mathbb{S}^*_{i,T}(\hat{\beta}, \tilde{\lambda}_{0i}, \hat{F})) + (\mathbb{S}^*_{i,T}(\hat{\beta}, \tilde{\lambda}_{0i}, \hat{F}) - \mathbb{S}^*_{i,T}(\beta_0, \tilde{\lambda}_{0i}, \tilde{F}_0)) \\ - (\mathbb{S}_{i,T}(\hat{\beta}, \hat{\lambda}_{0i}, \hat{F}) - \mathbb{S}^*_{i,T}(\hat{\beta}, \hat{\lambda}_{0i}, \hat{F})) - (\mathbb{S}^*_{i,T}(\hat{\beta}, \hat{\lambda}_{0i}, \hat{F}) - \mathbb{S}^*_{i,T}(\beta_0, \hat{\lambda}_{0i}, \tilde{F}_0)). \quad \text{(A.7)}\end{aligned}$$

Similar to the proof of Theorem 1, for small $\delta > 0$, define $B_{\delta,i} = \{\lambda_i \in \mathcal{A} : \|\lambda_i - \tilde{\lambda}_{0i}\| \leq \delta\}$. For any $\lambda_i \in B^C_{\delta,i}$, let $m = \|\lambda_i - \tilde{\lambda}_{0i}\| > \delta$. Then $\bar{\lambda}_i = \lambda_i \cdot \delta/m + \tilde{\lambda}_{0i} \cdot (1 - \delta/m)$ is on the boundary of $B_{\delta,i}$, i.e., $\|\bar{\lambda}_i - \tilde{\lambda}_{0i}\| = \delta$. Given β_0 and $\tilde{f}_{0t}$, the check function is convex in λ_i, thus

$$\delta/m \cdot \rho_{it}(\beta_0, \lambda_i, \tilde{f}_{0t}) + (1 - \delta/m) \cdot \rho_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t}) \geq \rho_{it}(\beta_0, \bar{\lambda}_i, \tilde{f}_{0t}),$$

and it follows that

$$\rho_{it}(\beta_0, \lambda_i, \tilde{f}_{0t}) - \rho_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t}) \geq m/\delta \cdot \left(\rho_{it}(\beta_0, \bar{\lambda}_i, \tilde{f}_{0t}) - \rho_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t})\right).$$

Note that $\rho_{it}(\beta_0, \bar{\lambda}_i, \tilde{f}_{0t}) - \rho_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t}) = \bar{\rho}_{it}(\beta_0, \bar{\lambda}_i, \tilde{f}_{0t}) - \bar{\rho}_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t}) + \tilde{\rho}_{it}(\beta_0, \bar{\lambda}_i, \tilde{f}_{0t}) - \tilde{\rho}_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t})$, and

$$\bar{\rho}_{it}(\beta_0, \bar{\lambda}_i, \tilde{f}_{0t}) - \bar{\rho}_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t}) \geq (\bar{\lambda}_i - \tilde{\lambda}_{0i})' \left(\mathsf{f}_{it}(0)\tilde{f}_{0t}\tilde{f}'_{0t}\right)(\bar{\lambda}_i - \tilde{\lambda}_{0i}) + o(\delta^2).$$

Thus, if $\|\hat{\lambda}_i - \tilde{\lambda}_{0i}\| > \delta$, by Assumption 3(ii)

$$\begin{aligned}\mathbb{S}^*_{i,T}(\beta_0, \hat{\lambda}_i, \tilde{F}_0) - \mathbb{S}^*_{i,T}(\beta_0, \tilde{\lambda}_{0i}, \tilde{F}_0) &= \frac{1}{T}\sum_{t=1}^{T}\left[\rho_{it}(\beta_0, \hat{\lambda}_i, \tilde{f}_{0t}) - \rho_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t})\right] \\ \geq (\bar{\hat{\lambda}}_i - \tilde{\lambda}_{0i})'\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{f}_{it}(0)\tilde{f}_{0t}\tilde{f}'_{0t}\right)(\bar{\hat{\lambda}}_i - \tilde{\lambda}_{0i}) &+ o(\delta^2) + m/\delta \cdot \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\rho}_{it}(\beta_0, \bar{\hat{\lambda}}_i, \tilde{f}_{0t}) - \tilde{\rho}_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t})\right] \\ &\geq C\delta^2 + m/\delta \cdot \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\rho}_{it}(\beta_0, \bar{\hat{\lambda}}_i, \tilde{f}_{0t}) - \tilde{\rho}_{it}(\beta_0, \tilde{\lambda}_{0i}, \tilde{f}_{0t})\right],\end{aligned}$$

where $\bar{\hat{\lambda}}_i$ is between $\tilde{\lambda}_{0i}$ and $\hat{\lambda}_i$ and is on the boundary of $B_{\delta,i}$. Thus, it follows from (A.7) that there exists some $\epsilon > 0$ (depending on δ) such that

$$P\left[\max_{1\le i\le N}\|\hat{\lambda}_i-\tilde{\lambda}_{0i}\|>\delta\right]\le P\left[\max_{1\le i\le N}\sup_{\beta,\lambda_i,F}\left|\mathbb{S}_{i,T}(\beta,\lambda_i,F)-\mathbb{S}^*_{i,T}(\beta,\lambda_i,F)\right|>\epsilon\right]$$
$$+P\left[\max_{1\le i\le N}\sup_{\lambda_i\in\mathcal{A}}\left|\mathbb{S}^*_{i,T}(\hat{\beta},\lambda_i,\hat{F})-\mathbb{S}^*_{i,T}(\beta_0,\lambda_i,\tilde{F}_0)\right|>\epsilon\right]$$
$$+P\left[\max_{1\le i\le N}\sup_{\lambda_i\in B_{\delta,i}}\left|\tilde{\mathbb{S}}^*_{i,T}(\beta_0,\lambda_i,\tilde{F}_0)-\tilde{\mathbb{S}}^*_{i,T}(\beta_0,\lambda_i,\tilde{F}_0)\right|>\epsilon\right]. \quad (A.8)$$

Similar to the proof of Theorem 1, it can be shown that the first and last term on the right-hand side of (A.8) are both $o(1)$. It remains to show that the second term is $o(1)$.

By the property of the check function, we have

$$\max_{1\le i\le N}\sup_{\lambda_i\in\mathcal{A}}\left|\mathbb{S}^*_{i,T}(\hat{\beta},\lambda_i,\hat{F})-\mathbb{S}^*_{i,T}(\beta_0,\lambda_i,\tilde{F}_0)\right|\lesssim\|\hat{\beta}-\beta_0\|\cdot\left(\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^T\|X_{it}\|\right)+\frac{1}{T}\sum_{t=1}^T\|\hat{f}_t-\tilde{f}_{0t}\|.$$

It then follows from Assumption 3(iii), $\|\hat{\beta}-\beta_0\|=o_P(1)$ and Lemma 2(ii) that

$$\max_{1\le i\le N}\sup_{\lambda_i\in\mathcal{A}}\left|\mathbb{S}^*_{i,T}(\hat{\beta},\lambda_i,\hat{F})-\mathbb{S}^*_{i,T}(\beta_0,\lambda_i,\tilde{F}_0)\right|=o_P(1).$$

This implies that the second term on the right-hand side of (A.8) is $o(1)$. The desired result follows by noting that $\max_i\|\tilde{\lambda}_{0i}-\check{\lambda}_{0i}\|\le O_P(1)\cdot\|\hat{\mathbf{H}}-\mathbf{H}_0\|=o_P(1)$. □

Lemma 4. *Under Assumptions 1 to 4,* $\|\hat{\beta}-\beta_0\|=o_P(\max_i\|\hat{\lambda}_i-\check{\lambda}_{0i}\|)+o_P(1/\sqrt{T})$.

Proof. **Step 1:** Define the following notations:

$$\underbrace{S^\beta(\beta,\Lambda,F)}_{p\times 1}=\partial\mathbb{S}_{NT}(\beta,\Lambda,F)/\partial\beta,\quad \underbrace{S^\lambda(\beta,\Lambda,F)}_{Nr\times 1}=\partial\mathbb{S}_{NT}(\beta,\Lambda,F)/\partial\Lambda,$$

$$\underbrace{S^{\beta\beta'}(\beta,\Lambda,F)}_{p\times p}=\partial^2\mathbb{S}_{NT}(\beta,\Lambda,F)/\partial\beta\partial\beta',\quad \underbrace{S^{\beta\lambda'}(\beta,\Lambda,F)}_{p\times Nr}=\partial^2\mathbb{S}_{NT}(\beta,\Lambda,F)/\partial\beta\partial\Lambda'.$$

The other functions such as $S^{\beta f'}(\beta,\Lambda,F)$, $S^{\lambda\lambda'}(\beta,\Lambda,F)$, $S^{\lambda f'}(\beta,\Lambda,F)$, $S^{\beta\beta' f_{th}}(\beta,\Lambda,F)$ are defined in a similar fashion. The arguments of these functions are dropped when they are evaluated at $(\beta,\Lambda,F)=(\beta_0,\check{\Lambda}_0,\check{F}_0)$, where $\check{\Lambda}_0=(\check{\lambda}_{01},\ldots,\check{\lambda}_{0N})'$, $\check{F}_0=(\check{f}_{01},\ldots,\check{f}_{0T})'$ (recall that $\check{f}_{0t}=\hat{\mathbf{H}}f_{0t}$ and $\check{\lambda}_{0i}=(\hat{\mathbf{H}}')^{-1}\lambda_{0i}$).

Expanding $S^\beta_{NT}(\hat{\beta},\hat{\Lambda},\hat{F})$ and $S^\lambda_{NT}(\hat{\beta},\hat{\Lambda},\hat{F})$ around $(\beta_0,\check{\Lambda}_0,\check{F}_0)$ up to the third order gives:

$$0=S^\beta(\hat{\beta},\hat{\Lambda},\hat{F})=S^\beta+S^{\beta\beta'}(\hat{\beta}-\beta_0)+S^{\beta\lambda'}(\hat{\Lambda}-\check{\Lambda}_0)+S^{\beta f'}(\hat{F}-\check{F}_0)+1/2R^\beta(\beta^*,\Lambda^*,F^*), \quad (A.9)$$

$$0=S^\lambda(\hat{\beta},\hat{\Lambda},\hat{F})=S^\lambda+S^{\lambda\beta'}(\hat{\beta}-\beta_0)+S^{\lambda\lambda'}(\hat{\Lambda}-\check{\Lambda}_0)+S^{\lambda f'}(\hat{F}-\check{F}_0)+1/2R^\lambda(\beta^*,\Lambda^*,F^*), \quad (A.10)$$

where $(\beta^*, \Lambda^*, F^*)$ is between $(\beta_0, \check{\Lambda}_0, \check{F}_0)$ and $(\hat{\beta}, \hat{\Lambda}, \hat{F})$, and

$$\begin{aligned} &R^{\beta}(\beta^*, \Lambda^*, F^*) = \\ &\quad \sum_{k=1}^{p} S_*^{\beta\beta'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\beta} - \beta_0) + \sum_{k=1}^{p} S_*^{\beta\lambda'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\Lambda} - \check{\Lambda}_0) + \sum_{k=1}^{p} S_*^{\beta f'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{F} - \check{F}_0) \\ &+ \sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta\beta'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{\beta} - \beta_0) + \sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta\lambda'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{\Lambda} - \check{\Lambda}_0) + \sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{F} - \check{F}_0) \\ &+ \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\beta\beta' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{\beta} - \beta_0) + \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\beta\lambda' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{\Lambda} - \check{\Lambda}_0) + \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\beta f' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{F} - \check{F}_0), \end{aligned} \tag{A.11}$$

$$\begin{aligned} &R^{\lambda}(\beta^*, \Lambda^*, F^*) = \\ &\quad \sum_{k=1}^{p} S_*^{\lambda\beta'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\beta} - \beta_0) + \sum_{k=1}^{p} S_*^{\lambda\lambda'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\Lambda} - \check{\Lambda}_0) + \sum_{k=1}^{p} S_*^{\lambda f'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{F} - \check{F}_0) \\ &+ \sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\lambda\beta'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{\beta} - \beta_0) + \sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\lambda\lambda'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{\Lambda} - \check{\Lambda}_0) + \sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\lambda f'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{F} - \check{F}_0) \\ &+ \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\lambda\beta' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{\beta} - \beta_0) + \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\lambda\lambda' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{\Lambda} - \check{\Lambda}_0) + \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\lambda f' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{F} - \check{F}_0), \end{aligned} \tag{A.12}$$

where the asterisk in the subscript of the functions means that these functions are evaluated at $(\beta^*, \Lambda^*, F^*)$.

Define $\tilde{S}^{\lambda\lambda'} = S^{\lambda\lambda'} - \bar{S}^{\lambda\lambda'}$, $\tilde{S}^{\beta\lambda'} = S^{\beta\lambda'} - \bar{S}^{\beta\lambda'}$, where $\bar{S}^{\beta\lambda'} = N^{-1}(\tilde{\Xi}_1, \ldots, \tilde{\Xi}_N)$, $\bar{S}^{\lambda\lambda'} = N^{-1}\operatorname{diag}(\tilde{\Omega}_1, \ldots, \tilde{\Omega}_N)$, and $\tilde{\Xi}_i = \Xi_i \hat{\mathbf{H}}'$, $\tilde{\Omega}_i = \hat{\mathbf{H}}\Omega_i\hat{\mathbf{H}}'$. Recall that

$$\Xi_i = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}]f_{0t}', \quad \Omega_i = \frac{1}{T}\sum_{t=1}^{T} \mathsf{f}_{it}(0)f_{0t}f_{0t}'.$$

Then (A.9) can be written as

$$0 = S^{\beta} + S^{\beta\beta'}(\hat{\beta} - \beta_0) + \bar{S}^{\beta\lambda'}(\hat{\Lambda} - \check{\Lambda}_0) + \tilde{S}^{\beta\lambda'}(\hat{\Lambda} - \check{\Lambda}_0) + S^{\beta f'}(\hat{F} - \check{F}_0) + 1/2R^{\beta}(\beta^*, \Lambda^*, F^*), \tag{A.13}$$

and (A.10) can be written as

$$0 = S^{\lambda} + S^{\lambda\beta'}(\hat{\beta} - \beta_0) + \bar{S}^{\lambda\lambda'}(\hat{\Lambda} - \check{\Lambda}_0) + \tilde{S}^{\lambda\lambda'}(\hat{\Lambda} - \check{\Lambda}_0) + S^{\lambda f'}(\hat{F} - \check{F}_0) + 1/2R^{\lambda}(\beta^*, \Lambda^*, F^*). \tag{A.14}$$

Plugging (A.14) into (A.13) gives

$$
\begin{aligned}
&[S^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda\beta'}](\hat{\beta} - \beta_0) = \\
&- \left[S^{\beta} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda}\right] - \left[S^{\beta f'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda f'}\right](\hat{F} - \check{F}_0) - \left[\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}\right](\hat{\Lambda} - \check{\Lambda}_0) \\
&\qquad - 1/2\left[R^{\beta}(\beta^*, \Lambda^*, F^*) - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}R^{\lambda}(\beta^*, \Lambda^*, F^*)\right]. \qquad (A.15)
\end{aligned}
$$

Step 2: The term $S^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda\beta'}$ can be written as

$$\bar{S}^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\bar{S}^{\lambda\beta'} + \tilde{S}^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\beta'},$$

where $\bar{S}^{\beta\beta'} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}X_{it}']$, $\bar{S}^{\lambda\beta'} = (\bar{S}^{\beta\lambda'})'$. Note that

$$\bar{S}^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\bar{S}^{\lambda\beta'} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}X_{it}' - \Xi_i\Omega_i^{-1}\Xi_i'] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}Z_{it}'].$$

Next, we show that $\tilde{S}^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\beta'} = o_P(1)$. Write

$$\tilde{S}^{\beta\beta'} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(l_{it}^{(2)}X_{it}X_{it}' - \mathbb{E}[l_{it}^{(2)}X_{it}X_{it}']\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbb{E}[l_{it}^{(2)}X_{it}X_{it}'] - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}X_{it}']\right)$$

where the second term on the right-side of the above equation is $O(h^q) = o(1)$ by Lemma S1, and for the first term, by Assumption 3(iv) we have

$$\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(l_{it}^{(2)}X_{it}X_{it}' - \mathbb{E}[l_{it}^{(2)}X_{it}X_{it}']\right)\right\|^2 \leq \frac{1}{Th^2}\cdot\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h\left(l_{it}^{(2)}X_{it}X_{it}' - \mathbb{E}[l_{it}^{(2)}X_{it}X_{it}']\right)\right\|^2.$$

By Lemma S1 and Assumption 3(iii), $\|h\cdot l_{it}^{(2)}X_{it}X_{it}'\| \leq M$ almost surely. Thus, it follows from the mixing property (Assumption 2(iv)), the fact that $Th^2 \to \infty$ (Assumption 3(vii)) and Theorem 3 of Yoshihara (1978) that the right-hand side of the above inequality is $o(1)$. Thus, we have $\tilde{S}^{\beta\beta'} = o_P(1)$. Similarly, we can show that $\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\beta'} = o_P(1)$. Therefore, it follows that

$$S^{\beta\beta'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda\beta'} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}Z_{it}'] + o_P(1) = \Delta + o_P(1). \qquad (A.16)$$

Step 3: $S^{\beta} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda}$ can be written as

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}l_{it}^{(1)}Z_{it} = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(l_{it}^{(1)}Z_{it} - \mathbb{E}[l_{it}^{(1)}Z_{it}]\right) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[l_{it}^{(1)}Z_{it}].$$

By Lemma S1 and Assumption 3(vii), $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[l_{it}^{(1)}Z_{it}] = o(h^q) = o(T^{-1/2})$. Similar to the proof of (A.53) below, the first term on the right-hand side of the above equation is $O_P(1/\sqrt{NT}) =$

$o_P(T^{-1/2})$. Thus, we have

$$S^{\beta} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda} = o_P(1/\sqrt{T}). \tag{A.17}$$

Step 4: By Lemma 8 below

$$[S^{\beta f'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda f'}](\hat{F} - \check{F}_0) = o_P(1/\sqrt{T}). \tag{A.18}$$

Step 5: Now consider the term: $\left[\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}\right](\hat{\Lambda} - \check{\Lambda}_0)$. Write

$$\tilde{S}^{\beta\lambda'}(\hat{\Lambda} - \check{\Lambda}_0) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} X_{it} f_{0t}' - \Xi_i\right)\hat{\mathbf{H}}(\hat{\lambda}_i - \check{\lambda}_{0i}).$$

By the Cauchy-Schwarz inequality, we have

$$\left\|\tilde{S}^{\beta\lambda'}(\hat{\Lambda} - \check{\Lambda}_0)\right\| \lesssim \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} X_{it} f_{0t}' - \Xi_i\right\|^2} \cdot \|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}.$$

Note that by Lemma S1,

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} X_{it} f_{0t}' - \Xi_i\right\|^2 &= \mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(2)} X_{it} - \mathbb{E}[l_{it}^{(2)} X_{it}]\right) f_{0t}' + \frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}[l_{it}^{(2)} X_{it}] - \mathbb{E}[\mathsf{f}_{it}(0|X_{it}) X_{it}]\right) f_{0t}'\right\|^2 \\ &\leq \frac{1}{h^2}\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T} h\left(l_{it}^{(2)} X_{it} - \mathbb{E}[l_{it}^{(2)} X_{it}]\right) f_{0t}'\right\|^2 + o(1)\end{aligned}$$

and by the mixing property and Theorem 3 of Yoshihara (1978) the first term on the right hand side of the above inequality is $O(1/(Th^2)) = o(1)$. Thus, it follows that

$$\left\|\tilde{S}^{\beta\lambda'}(\hat{\Lambda} - \check{\Lambda}_0)\right\| = o_P(1) \cdot \|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}.$$

Similarly, we can show that $\|\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}(\hat{\Lambda} - \check{\Lambda}_0)\| = o_P(1) \cdot \|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}$, and conclude that

$$\left\|\left(\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}\right)(\hat{\Lambda} - \check{\Lambda}_0)\right\| = o_P(1) \cdot \|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}. \tag{A.19}$$

Step 6: We will show that:

$$R^{\beta}(\beta^*, \Lambda^*, F^*) = o_P(\|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}) + o_P(\|\hat{\beta} - \beta_0\|) + o_P(1/\sqrt{T}), \tag{A.20}$$

$$\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}R^{\lambda}(\beta^*, \Lambda^*, F^*) = o_P(\|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}) + o_P(\|\hat{\beta} - \beta_0\|) + o_P(1/\sqrt{T}). \tag{A.21}$$

To save space, we focus on (A.20), which follows from Results 1 to 9 below. The proof of (A.21) is similar.

Result 1: $S_*^{\beta\beta'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\beta} - \beta_0) = o_P(\|\hat{\beta} - \beta_0\|)$.

Observe that:

$$S_*^{\beta\beta'\beta_k}(\hat{\beta}_k-\beta_{0k})(\hat{\beta}-\beta_0)=-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it}'X_{it,k}(\hat{\beta}_k-\beta_{0k})(\hat{\beta}-\beta_0),$$

so

$$\|S_*^{\beta\beta'\beta_k}(\hat{\beta}_k-\beta_{0k})(\hat{\beta}-\beta_0)\|\leq\|\hat{\beta}-\beta_0\|^2\cdot\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it}'X_{it,k}\right\|.$$

Expanding $l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)$ around $(\beta^*,\lambda_i^*,\tilde{f}_{0t})$ gives

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it}'X_{it,k}\\&=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}l_{it}^{(4)}(\beta^*,\lambda_i^*,f_t^{**})X_{it}X_{it}'X_{it,k}(\lambda_i^*)'(f_t^*-\tilde{f}_{0t})+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}]|_{\beta=\beta^*,\lambda_i=\lambda_i^*}\\&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}-\mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}]|_{\beta=\beta^*,\lambda_i=\lambda_i^*}\right].\end{aligned}\tag{A.22}$$

where f_t^{**} is between f_t^* and $\check{f}_{0t}$. By Lemma S1, the second term on the right hand side of (A.22) is $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}^{(1)}(\cdot|X_{it})X_{it}X_{it}'X_{it,k}]+\bar{O}(h^{q-1})=O(1)$, and the first term is bounded by

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|f_t^*-\tilde{f}_{0t}\|^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[l_{it}^{(4)}(\beta^*,\lambda_i^*,f_t^{**})]^2\cdot\|X_{it}\|^6\cdot\|\lambda_i^*\|^2}=O_P\left(1/\sqrt{Th^6}\right)=o_P(1),$$

since we have $|l_{it}^{(4)}(\cdot)|\lesssim 1/h^3$ by Lemma S1 and $T^{-1}\sum_{t=1}^{T}\|f_t^*-\tilde{f}_{0t}\|^2\leq T^{-1}\sum_{t=1}^{T}\|\hat{f}_t-\tilde{f}_{0t}\|^2=O_P(N^{-1})=O_P(T^{-1})$ by Lemma 2 and Assumption 3(vii). Finally, with probability approaching 1, the last term on the right hand side of (A.22) is bounded by $N^{-1}\sum_{i=1}^{N}\mathcal{Z}_i$, where

$$\mathcal{Z}_i=\sup_{(\beta,\lambda_i)\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}-\mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}]\right]\right\|$$

and $B_{\delta,i}$ is a neighbourhood of $(\beta_0,\tilde{\lambda}_{0i})$. Then Result 1 follows if we can show that $\max_{1\leq i\leq N}\mathbb{E}[\mathcal{Z}_i]<\infty$.

For any $\epsilon>0$, let $\theta_i^{(1)},\ldots,\theta_i^{(L)}$ be a maximal set of points in $B_{\delta,i}$ such that $\|\theta_i^{(j)}-\theta_i^{(l)}\|\geq\epsilon$ for any $j\neq l$. It is well know that L, the packing number of a Euclidean ball, is bounded (up to a positive constant that only depends on $p+r$) by $(1/\epsilon)^{p+r}$. Thus, we have

$$\mathbb{E}\mathcal{Z}_i\leq\frac{1}{\sqrt{Th^4}}\cdot\mathbb{E}\left[\max_{1\leq j\leq L}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h^2\left[l_{it}^{(3)}(\beta^{(j)},\lambda_i^{(j)},\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}-\mathbb{E}[l_{it}^{(3)}(\beta^{(j)},\lambda_i^{(j)},\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}]\right]\right\|\right]+O(\epsilon/h^3),$$

where we have used that fact that $|l_{it}^{(3)}(\beta^a,\lambda_i^a,\tilde{f}_{0t})-l_{it}^{(3)}(\beta^b,\lambda_i^b,\tilde{f}_{0t})|\lesssim(\|\beta^a-\beta^b\|+\|\lambda_i^a-\lambda_i^b\|)/h^3$. Note that $h^2l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})X_{it}X_{it}'X_{it,k}$ is uniformly bounded (almost surely) by Lemma S1 and Assumption 3(iii). Thus, by Lemma 2.2.2 of van der Vaart and Wellner (1996) and Theorem 3 of Yoshihara (1978), for any $J\geq 2$, the first term on right hand side of the above inequality is bounded (up to a positive constant) by $L^{1/J}/\sqrt{Th^4}$. Choosing $\epsilon = 1/\sqrt{T}$, we have

$$\mathbb{E}\mathcal{Z}_i \leq C\left(\frac{T^{\frac{p+r}{2J}}}{\sqrt{Th^4}} + \frac{1}{\sqrt{Th^6}}\right)$$

for some positive constant C. Since J is arbitrary, we can choose $J > (p+r)/(2c)$ (c is defined in Assumption 3(iii)) it follows that $T^{\frac{p+r}{2J}}/\sqrt{Th^4} = o(\sqrt{Th^6})$. Then from Assumption 3(iii) we have $\max_{1\leq i\leq N}\mathbb{E}[\mathcal{Z}_i] < \infty$ and the desired result follows.

Result 2: $S_*^{\beta\lambda'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\Lambda} - \check{\Lambda}_0) = o_P(\|\hat{\beta} - \beta_0\|)$.

We have

$$\begin{aligned}\|S_*^{\beta\lambda'\beta_k}(\hat{\beta}_k - \beta_{0k})(\hat{\Lambda} - \check{\Lambda}_0)\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^*)X_{it}X_{it,k}f_t^{*'}\right)(\hat{\lambda}_i - \check{\lambda}_{0i})(\hat{\beta} - \beta_0)\right\| \\ &\leq \|\hat{\beta} - \beta_0\| \cdot \max_{i\leq N}\|\hat{\lambda}_i - \check{\lambda}_{0i}\| \cdot \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^*)X_{it}X_{it,k}f_t^{*'}\right\|.\end{aligned}$$

Note that

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^*)X_{it}X_{it,k}f_t^{*'} &= \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, \tilde{f}_{0t})X_{it}X_{it,k}\tilde{f}_{0t}' + \\ &+ \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^{**})X_{it}X_{it,k}(f_t^* - \tilde{f}_{0t})' + \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(4)}(\beta^*, \lambda_i^*, f_t^{**})X_{it}X_{it,k}\tilde{f}_{0t}'(\lambda_i^*)'(f_t^* - \tilde{f}_{0t}),\end{aligned}$$

where f_t^{**} is between f_t^* and $\tilde{f}_{0t}$. Using Lemma S1, Lemma 2, Assumption 3(iii) and the Cauchy-Schwarz inequality, we can show that the last two terms on right-hand side of the above inequality is $\bar{O}_P(1/\sqrt{Nh^6}) = \bar{O}_P(1/\sqrt{Th^6}) = \bar{o}_P(1)$. For the first term on right-hand side of the above inequality, with probability approaching 1, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, \tilde{f}_{0t})X_{it}X_{it,k}\tilde{f}_{0t}'\right\| &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\mathbb{E}\left[l_{it}^{(3)}(\beta, \lambda_i, \tilde{f}_{0t})X_{it}X_{it,k}\tilde{f}_{0t}'\right]|_{\beta=\beta^*, \lambda_i=\lambda_i^*}\right\| \\ &+ \frac{1}{N}\sum_{i=1}^{N}\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(3)}(\beta, \lambda_i, \tilde{f}_{0t})X_{it}X_{it,k}\tilde{f}_{0t}' - \mathbb{E}\left[l_{it}^{(3)}(\beta, \lambda_i, \tilde{f}_{0t})X_{it}X_{it,k}\tilde{f}_{0t}'\right]\right)\right\|\end{aligned}$$

The first term on the right-hand side of the above inequality is $O(1)$ by Lemma S1. Similar to the proof of Result 1, The second term on the right-hand side of the above inequality can be shown to be $o_P(1)$. Thus, we have

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^*)X_{it}X_{it,k}f_t^{*'}\right\| = O_P(1),$$

and the result follows from uniform consistency of $\hat{\lambda}_i$.

Result 3: $S_*^{\beta f'\beta_k}(\hat\beta_k-\beta_{0k})(\hat F-\check F_0)=o_P(\|\hat\beta-\beta_0\|)$.

Note that by Lemma 2,

$$\begin{aligned}
& \|S_*^{\beta f'\beta_k}(\hat\beta_k-\beta_{0k})(\hat F-\check F_0)\| \\
\leq\ & \|\hat\beta-\beta_0\|\cdot\left\|\frac{1}{T}\sum_{t=1}^T\left(\frac{1}{N}\sum_{i=1}^N l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it,k}\lambda_i^{*'}\right)(\hat f_t-\check f_{0t})\right\| \\
\leq\ & \|\hat\beta-\beta_0\|\cdot\sqrt{\frac{1}{T}\sum_{t=1}^T\|\hat f_t-\check f_{0t}\|^2\cdot\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T[l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)]^2\cdot\|X_{it}\|^4\cdot\|\lambda_i^*\|^2} \\
=\ & \|\hat\beta-\beta_0\|\cdot O_P\left(1/\sqrt{Nh^4}\right)=o_P(\|\hat\beta-\beta_0\|),
\end{aligned}$$

because by Lemma S1, $h^2 l_{it}^{(3)}(\beta,\lambda_i,f_t)$ is uniformly bounded, and $Nh^4\to\infty$ by Assumption 3(vii).

Result 4: $\sum_{i=1}^N\sum_{h=1}^r S_*^{\beta\beta'\lambda_{ih}}(\hat\lambda_{ih}-\check\lambda_{0i,h})(\hat\beta-\beta_0)=o_P(\|\hat\beta-\beta_0\|)$.

Observe that for each $h\leq r$,

$$\sum_{i=1}^N S_*^{\beta\beta'\lambda_{ih}}(\hat\lambda_{ih}-\check\lambda_{0i,h})(\hat\beta-\beta_0)=-\frac{1}{N}\sum_{i=1}^N(\hat\lambda_{ih}-\check\lambda_{0i,h})\left(\frac{1}{T}\sum_{t=1}^T l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it}'f_{th}^*\right)(\hat\beta-\beta_0),$$

so

$$\left\|\sum_{i=1}^N S_*^{\beta\beta'\lambda_{ih}}(\hat\lambda_{ih}-\check\lambda_{0i,h})(\hat\beta-\beta_0)\right\|\leq\|\hat\beta-\beta_0\|\cdot\max_{i\leq N}\|\hat\lambda_i-\check\lambda_{0i}\|\cdot\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it}'f_{th}^*\right\|,$$

which can be shown to be $o_P(\|\hat\beta-\beta_0\|)$, similar to the proof of Result 2.

Result 5: $\sum_{t=1}^T\sum_{h=1}^r S_*^{\beta\beta' f_{th}}(\hat f_{th}-\check f_{0t,h})(\hat\beta-\beta_0)=o_P(\|\hat\beta-\beta_0\|)$.

The proof is similar to Result 3. For each $h\leq r$, we have

$$\sum_{t=1}^T S_*^{\beta\beta' f_{th}}(\hat f_{th}-\check f_{0t,h})=-\frac{1}{T}\sum_{t=1}^T(\hat f_{th}-\check f_{0t,h})\left(\frac{1}{N}\sum_{i=1}^N l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}X_{it}'\lambda_{ih}^*\right),$$

so

$$\left\|\sum_{t=1}^T S_*^{\beta\beta' f_{th}}(\hat f_{th}-\check f_{0t,h})\right\|\leq\sqrt{\frac{1}{T}\sum_{t=1}^T\|\hat f_t-\check f_{0t}\|^2\cdot\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T[l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)]^2\cdot\|X_{it}\|^4\cdot\|\lambda_i^*\|^2},$$

which is $O_P(1/\sqrt{Th^4})$. So Result 5 follows.

Result 6: $\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0,ih})(\hat{F}-\check{F}_0) = o_P(\|\hat{\Lambda}-\check{\Lambda}_0\|/\sqrt{N})$.

Write:

$$\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{F}-\check{F}_0) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[l_{it}^{(2)}(\beta^*,\lambda_i^*,f_t^*)X_{it}-l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(\lambda_i^*)'f_t^*](\hat{\lambda}_i-\check{\lambda}_{0i})'(\hat{f}_t-\check{f}_{0t}).$$

Thus, by the Cauchy-Schwarz inequality, Lemma 2 and Lemma S1

$$\begin{aligned}
&\left\|\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{F}-\check{F}_0)\right\| \le \|\hat{\Lambda}-\check{\Lambda}_0\|/\sqrt{N}\cdot \\
&\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t-\check{f}_{0t}\|^2\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|[l_{it}^{(2)}(\beta^*,\lambda_i^*,f_t^*)X_{it}-l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(\lambda_i^*)'f_t^*]\right\|^2} \\
= \;& \|\hat{\Lambda}-\check{\Lambda}_0\|/\sqrt{N}\cdot O_P(1/\sqrt{Nh^4}),
\end{aligned}$$

so the result follows by Assumption 3(vii).

Result 7: $\sum_{t=T}^{N}\sum_{h=1}^{r} S_*^{\beta\lambda' f_{th}}(\hat{f}_{th}-\check{f}_{0t,h})(\hat{\Lambda}-\check{\Lambda}_0) = o_P(\|\hat{\Lambda}-\check{\Lambda}_0\|/\sqrt{N})$.

The proof is similar to the proof the Result 6.

Result 8: $\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta\lambda'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0,ih})(\hat{\Lambda}-\check{\Lambda}_0) = o_P(\|\hat{\Lambda}-\check{\Lambda}_0\|/\sqrt{N})$.

Note that for each $h \le r$, we have

$$\sum_{i=1}^{N} S_*^{\beta\lambda'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{\Lambda}-\check{\Lambda}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(f_t^*)'f_{th}^*(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{\lambda}_i-\check{\lambda}_{0i}),$$

by Lemma 3

$$\begin{aligned}
\left\|\sum_{i=1}^{N} S_*^{\beta\lambda'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{\Lambda}-\check{\Lambda}_0)\right\| &\le \max_{1\le N}\|\hat{\lambda}_i-\check{\lambda}_{0i}\|\cdot\frac{1}{N}\sum_{i=1}^{N}\left(\|\hat{\lambda}_i-\check{\lambda}_{0i}\|\cdot\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}f_t^* f_{th}^*\right\|\right) \\
&\le o_P(1)\cdot\|\hat{\Lambda}-\check{\Lambda}_0\|/\sqrt{N}\cdot\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(f_t^*)'f_{th}^*\right\|^2}.
\end{aligned}$$

Thus, it remains to show that

$$\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(f_t^*)'f_{th}^*\right\|^2 = O_P(1). \tag{A.23}$$

First, write

$$
\begin{aligned}
l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(f_t^*)'f_{th}^* &= l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(\tilde f_{0t})'\tilde f_{0,th} + l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}\left((f_t^*)'f_{th}^* - (\tilde f_{0t})'\tilde f_{0,th}\right)\\
&= l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde f_{0t})X_{it}(\tilde f_{0t})'\tilde f_{0,th} - l_{it}^{(4)}(\beta^*,\lambda_i^*,f_t^{**})(\lambda_i^*)'(f_t^*-\tilde f_{0t})X_{it}(\tilde f_{0t})'\tilde f_{0,th}\\
&\qquad + l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}\left((f_t^*)'f_{th}^* - (\tilde f_{0t})'\tilde f_{0,th}\right),
\end{aligned}
$$

thus, by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(f_t^*)'f_{th}^*\right\|^2 &\le \frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde f_{0t})X_{it}(\tilde f_{0t})'\tilde f_{0,th}\right\|^2\\
&+\frac{1}{T}\sum_{t=1}^T\|(f_t^*)'f_{th}^*-(\tilde f_{0t})'\tilde f_{0,th}\|^2\cdot\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}\|^2\\
&+\frac{1}{T}\sum_{t=1}^T\|f_t^*-\tilde f_{0t}\|^2\cdot\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|l_{it}^{(4)}(\beta^*,\lambda_i^*,f_t^{**})(\lambda_i^*)X_{it}(\tilde f_{0t})'\tilde f_{0,th}\|^2.
\end{aligned}
$$

The last two terms on the right-hand side of the above inequality are both $o_P(1)$ by Lemma 2 and Lemma S1. For the first term on the right-hand side of the above inequality, by Assumption 3(iii), with probability approaching 1,

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{t=1}^T l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde f_{0t})X_{it}(\tilde f_{0t})'\tilde f_{0,th}\right\|^2 &\le \max_{i,t}\left\|\mathbb{E}[g_{it}(\beta,\lambda_i)|_{\beta=\beta^*,\lambda_i=\lambda_i^*}]\right\|^2+\\
&\frac{1}{N}\sum_{i=1}^N\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^T\left(g_{it}(\beta,\lambda_i)-\mathbb{E}[g_{it}(\beta,\lambda_i)]\right)\right\|^2,
\end{aligned}
$$

where $g_{it}(\beta,\lambda_i)=l_{it}^{(3)}(\beta,\lambda_i,\tilde f_{0t})X_{it}(\tilde f_{0t})'\tilde f_{0,th}$. The first term on the right-hand side of the above inequality is $O(1)$ by Lemma S1. Thus, to prove (A.23), it suffices to show that

$$
\max_{1\le i\le N}\mathbb{E}\left[\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^T\left(g_{it}(\beta,\lambda_i)-\mathbb{E}[g_{it}(\beta,\lambda_i)]\right)\right\|^2\right]=O(1).
$$

Similar to the proof of Result 1, for any $\epsilon>0$, let $\theta_i^{(1)},\dots,\theta_i^{(L)}$ be a maximal set of points in $B_{\delta,i}$ such that $\|\theta_i^{(j)}-\theta_i^{(l)}\|\ge\epsilon$ for any $j\ne l$. Thus, for some constants $C_1,C_2>0$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^T\left(g_{it}(\beta,\lambda_i)-\mathbb{E}[g_{it}(\beta,\lambda_i)]\right)\right\|^2\right]\\
&\le C_1\frac{1}{Th^4}\mathbb{E}\left[\max_{1\le j\le L}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T h^2\left(g_{it}(\beta^{(j)},\lambda_i^{(j)})-\mathbb{E}[g_{it}(\beta^{(j)},\lambda_i^{(j)})]\right)\right\|^2\right]+C_2\epsilon^2/h^6=O\left(\frac{L^{1/J}}{Th^4}+\epsilon^2/h^6\right)
\end{aligned}
$$

for any $J \geq 1$. Choosing $\epsilon = 1/\sqrt{T}$ and $J > (p+r)/(4c)$, then we have

$$O\left(\frac{L^{1/J}}{Th^4} + \epsilon^2/h^6\right) = O\left(\frac{T^{(p+r)/(2J)}}{Th^4}\right) + O\left(\frac{1}{Th^6}\right) = o\left(\frac{T^{2c}}{Th^4}\right) + o(1) = o(1).$$

This completes the proof of Result 8.

Result 9: $\sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\beta f' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{F} - \check{F}_0) = o_P(1/\sqrt{T})$.

For each $h \leq r$, we have

$$\sum_{t=1}^{T} S_*^{\beta f' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{F} - \check{F}_0) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^*)X_{it}(\lambda_i^*)'\lambda_{ih}^*(\hat{f}_{th} - \check{f}_{0t,h})(\hat{f}_t - \check{f}_{0t}),$$

so by Assumption 3(iii), Lemma 2 and Lemma S1,

$$\left\|\sum_{t=1}^{T} S_*^{\beta f' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{F} - \check{F}_0)\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\left(\|\hat{f}_t - \check{f}_{0t}\|^2 \cdot \left\|\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(\beta^*, \lambda_i^*, f_t^*)X_{it}(\lambda_i^*)'\lambda_{ih}^*\right\|\right)$$

$$\lesssim \frac{1}{h^2}\cdot\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t - \check{f}_{0t}\|^2 = O_P\Big(1/Nh^2\Big).$$

Then the result follows since $O_P\left(1/Nh^2\right) = O_P\left(\frac{1}{\sqrt{N}\cdot\sqrt{N}h^2}\right)$ and Assumption 3(vii) implies that $\sqrt{N}h^2 \to \infty$.

Step 7: It follows from (A.15) to (A.21) that:

$$\Delta(\hat{\beta} - \beta_0) = o_P(\|\hat{\beta} - \beta_0\|) + o_P(\|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N}) + o_P(1/\sqrt{T}),$$

then the desired result follows from the assumption that Δ is positive definite and the fact that $\|\hat{\Lambda} - \check{\Lambda}_0\|/\sqrt{N} \leq \max_i \|\hat{\lambda}_i - \check{\lambda}_{0i}\|$. □

Lemma 5. *Under Assumptions 1 to 4, there exists* $0 < \nu < 1/6 - c$ *(where* c *is defined in Assumption 3(vii)) such that* $\max_{1\leq i\leq N}\|\hat{\lambda}_i - \check{\lambda}_{0i}\| = O_P(1/T^{1/2-\nu})$.

Proof. Expanding the first order condition: $T^{-1}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{\beta},\hat{\lambda}_i,\hat{f}_t)\hat{f}_t = 0$ gives:

$$\tilde{\Omega}_i(\hat{\lambda}_i-\check{\lambda}_{0i}) = \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}\check{f}_{0t}+\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t-\check{f}_{0t})-\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_t[X_{it}'(\hat{\beta}-\beta_0)]-\left(\tilde{\Omega}_i-\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\check{f}_{0t}\check{f}_{0t}'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})$$
$$-\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}(\hat{f}_t-\check{f}_{0t})\check{f}_{0t}'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})-\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_t(\hat{f}_t-\check{f}_{0t})'\right)\check{\lambda}_{0i}+0.5\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[X_{it}'(\hat{\beta}-\beta_0)]^2$$
$$+\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[X_{it}'(\hat{\beta}-\beta_0)]\cdot[(f_t^*)'(\hat{\lambda}_i-\check{\lambda}_{0i})]+\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[X_{it}'(\hat{\beta}-\beta_0)]\cdot[(\lambda_i^*)'(\hat{f}_t-\check{f}_{0t})]$$
$$+\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[(\lambda_i^*)'(\hat{f}_t-\check{f}_{0t})]\cdot[(f_t^*)'(\hat{\lambda}_i-\check{\lambda}_{0i})]+0.5\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[(\lambda_i^*)'(\hat{f}_t-\check{f}_{0t})]^2+0.5\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[(f_t^*)'(\hat{\lambda}_i-\check{\lambda}_{0i})]^2.$$

Step 1: Let M be a generic bounded constant. By Lemma S1,

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}\check{f}_{0t}\right\| \le \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} \tilde{l}_{it}^{(1)}f_{0t}\right\|\cdot\|\hat{\mathbf{H}}\|+O(h^q).$$

Since, $\{l_{it}^{(1)}f_{0t}\}$ is uniformly bounded, by the mixing property of u_{it} and Theorem 3 of Yoshihara (1978), for any $J\ge 2$ we have $\mathbb{E}\left\|T^{-1/2}\sum_{t=1}^{T}\tilde{l}_{it}^{(1)}f_{0t}\right\|^J < M$ and it follows from Lemma 2.2.2 of van der Vaart and Wellner (1996) that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} \tilde{l}_{it}^{(1)}f_{0t}\right\| = O_P(N^{1/J}/\sqrt{T}) = O_P(1/T^{1/2-1/J}).$$

Choosing J large enough such that $\nu = 1/J < 1/6 - c$, we have

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}\check{f}_{0t}\right\| = O_P(1/T^{1/2-\nu}) \tag{A.24}$$

since $O(h^q)=o(T^{-1})$ by Assumption 3(vii).

Step 2: By Lemma S1, $l_{it}^{(1)}$ is uniformly bounded, so it follows from Lemma 2 that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t-\check{f}_{0t})\right\| \le O_P(1)\cdot\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t-\check{f}_{0t}\| = O_P(N^{-1/2}) = O_P(T^{-1/2}). \tag{A.25}$$

Step 3: Note that:

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_t[X_{it}'(\hat{\beta}-\beta_0)]\right\| \le \|\hat{\beta}-\beta_0\|\cdot\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_tX_{it}'\right\|$$
$$\le \|\hat{\beta}-\beta_0\|\cdot\|\hat{\mathbf{H}}\|\cdot\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}f_{0t}X_{it}'\right\| + \|\hat{\beta}-\beta_0\|\cdot\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}(\hat{f}_t-\check{f}_{0t})X_{it}'\right\|.$$

First, by Assumption 3(iii), Lemma 2 and Lemma S1, we have

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}(\hat{f}_t-\check{f}_{0t})X_{it}'\right\| \le O_P(h^{-1})\cdot\frac{1}{T}\sum_{t=1}^{T}\|\hat{f}_t-\check{f}_{0t}\| = O_P(1/(\sqrt{N}h)) = o_P(1).$$

Second, by Lemma S1,

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t}X_{it}'\right\| \le \frac{1}{\sqrt{Th^2}}\cdot\max_{1\le i\le N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\left(l_{it}^{(2)} f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(2)} f_{0t}X_{it}']\right)\right\| + O(1).$$

Since $\|hl_{it}^{(2)} f_{0t}X_{it}'\|$ is uniformly bounded by Assumption 3(iii) and Lemma S1, it follows from Theorem 3 of Yoshihara (1978) that

$$\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\left(l_{it}^{(2)} f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(2)} f_{0t}X_{it}']\right)\right\|^4 < M,$$

and then it follows from Lemma 2.2.2 of van der Vaart and Wellner (1996) that

$$\frac{1}{\sqrt{Th^2}}\cdot\max_{1\le i\le N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\left(l_{it}^{(2)} f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(2)} f_{0t}X_{it}']\right)\right\| = O_P(N^{1/4}/\sqrt{Th^2}) = o_P(1).$$

Thus, it can be concluded that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_t[X_{it}'(\hat{\beta}-\beta_0)]\right\| = O_P(\|\hat{\beta}-\beta_0\|). \tag{A.26}$$

Step 4: By the definition of $\tilde{\Omega}_i$, we have

$$\max_{1\le i\le N}\left\|\left(\tilde{\Omega}_i - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\check{f}_{0t}\check{f}_{0t}'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})\right\| \le \|\hat{\mathbf{H}}\|^2\cdot\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(2)}-\mathsf{f}_{it}(0))f_{0t}f_{0t}'\right\|\cdot\max_{1\le i\le N}\|\hat{\lambda}_i-\check{\lambda}_{0i}\|.$$

Similar to the proof of Step 3, we can show that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(2)}-\mathsf{f}_{it}(0))f_{0t}f_{0t}'\right\| = o_P(1),$$

thus it follows that

$$\max_{1\le i\le N}\left\|\left(\tilde{\Omega}_i - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\check{f}_{0t}\check{f}_{0t}'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})\right\| = o_P\left(\max_i\|\hat{\lambda}_i-\check{\lambda}_{0i}\|\right). \tag{A.27}$$

Step 5: By Lemma 2, Lemma S1 and Assumption 3(vii) we have

$$\max_{1\le i\le N}\left\|\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}(\hat{f}_t-\check{f}_{0t})\check{f}_{0t}'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})\right\| \le \|\hat{\mathbf{H}}\|\cdot O_P(1/\sqrt{Nh^2})\cdot \max_{1\le i\le N}\|\hat{\lambda}_i-\check{\lambda}_{0i}\| = o_P\left(\max_i \|\hat{\lambda}_i-\check{\lambda}_{0i}\|\right). \tag{A.28}$$

Step 6: Note that

$$\begin{aligned}\max_{1\le i\le N}\left\|\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_t(\hat{f}_t-\check{f}_{0t})'\right)\check{\lambda}_{0i}\right\| &\le O_P(1)\cdot\|\hat{\mathbf{H}}\|\cdot \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t}(\hat{f}_t-\check{f}_{0t})'\right\| \\ &\quad + O_P(1)\cdot \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}(\hat{f}_t-\check{f}_{0t})(\hat{f}_t-\check{f}_{0t})'\right\|.\end{aligned}$$

The second term on the right-hand side of the above inequality is $O_P(1/(Th))$ by Lemma S1 and Lemma 2. Next, we have

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t}(\hat{f}_t-\check{f}_{0t})'\right\| \lesssim \max_{1\le t\le T}\|\hat{f}_t-\check{f}_{0t}\|\cdot \max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left|l_{it}^{(2)}\right|,$$

and

$$\begin{aligned}\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left|l_{it}^{(2)}\right| &\lesssim 1/h\cdot \max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\{|u_{it}|\le h\} \\ &\le \max_{i,t} P[|u_{it}|\le h]/h + h^{-1}\cdot\max_{i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\{|u_{it}|\le h\}-P(|u_{it}|\le h)\right]\right|.\end{aligned}$$

It is easy to see that $\max_{i,t} P[|u_{it}|\le h]=O(h)$. Moreover, similar to the proof of Step 3, we can show that

$$h^{-1}\cdot\max_{i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\{|u_{it}|\le h\}-P(|u_{it}|\le h)\right]\right| = o_P(1).$$

Therefore,

$$\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left|l_{it}^{(2)}\right| = O_P(1),$$

and by Lemma 2

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t}(\hat{f}_t-\check{f}_{0t})'\right\| = O_P(\sqrt{\log T}/\sqrt{N}).$$

Thus, we can conclude that

$$\max_{1\le i\le N}\left\|\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\hat{f}_t(\hat{f}_t-\check{f}_{0t})'\right)\check{\lambda}_{0i}\right\| = O_P(\sqrt{\log T}/\sqrt{N}) + O_P(1/(Th)) = O_P(\sqrt{\log T}/\sqrt{T}). \tag{A.29}$$

Step 7: By the consistency of $\hat{\beta}$,

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t\cdot[X_{it}'(\hat{\beta}-\beta_0)]^2\right\| \le \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t X_{it}'\right\|\cdot\|\hat{\beta}-\beta_0\|^2$$
$$= o_P(\|\hat{\beta}-\beta_0\|)\cdot\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t X_{it}'\right\|.$$

Write

$$\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t X_{it}' = \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde{f}_{0t})\tilde{f}_{0t}X_{it}' + \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(4)}(\beta^*,\lambda_i^*,f_t^{**})\tilde{f}_{0t}X_{it}'\cdot[(f_t^*-\tilde{f}_{0t})\lambda_i^*] + \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)(\hat{f}_t-\tilde{f}_{0t})X_{it}'.$$

It then follows from Lemma 2 and Lemma S1 that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t X_{it}'\right\| \lesssim \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde{f}_{0t})f_{0t}X_{it}'\right\| + O_P(1/\sqrt{Th^6}).$$

Let $B_{\delta,i}$ be a neighbourhood of $(\beta_0,\tilde{\lambda}_{0i})$, then by the uniform consistency of $\hat{\lambda}_i$, with probability approaching 1, we have

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*,\lambda_i^*,\tilde{f}_{0t})f_{0t}X_{it}'\right\| \le \max_{i,t}\sup_{\theta_i\in B_{\delta,i}}\left\|\mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}']\right\|$$
$$+\max_{1\le i\le N}\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}']\right)\right\|.$$

The first term on the right-hand side of the above inequality is $O(1)$ by Lemma S1. Next, for each i, let $\theta_i^{(1)},\ldots,\theta_i^{(L_i)}$ be a maximal set of points in $B_{\delta,i}$ such that $\max_i\|\theta_i^{(j)}-\theta_i^{(k)}\|\le\epsilon$ for some small $\epsilon>0$. It follows that

$$\max_{1\le i\le N}\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}']\right)\right\| \le$$
$$\frac{1}{\sqrt{T}h^2}\cdot\max_{1\le i\le N}\max_{1\le j\le L_i}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h^2\left(l_{it}^{(3)}(\beta^{(j)},\lambda_i^{(j)},\tilde{f}_{0t})f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(3)}(\beta^{(j)},\lambda_i^{(j)},\tilde{f}_{0t})f_{0t}X_{it}']\right)\right\| + O(\epsilon/h^3).$$

Note $\max_i L_i \lesssim \bar{L} = (1/\epsilon)^{p+r}$, and by Theorem 3 of Yoshihara (1978) for any $J\ge 2$,

$$\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h^2\left(l_{it}^{(3)}(\beta^{(j)},\lambda_i^{(j)},\tilde{f}_{0t})f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(3)}(\beta^{(j)},\lambda_i^{(j)},\tilde{f}_{0t})f_{0t}X_{it}']\right)\right\|^J < M.$$

Thus, by Lemma 2.2.2 of van der Vaart and Wellner (1996), we have

$$\max_{1\le i\le N}\sup_{\theta_i\in B_{\delta,i}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}' - \mathbb{E}[l_{it}^{(3)}(\beta,\lambda_i,\tilde{f}_{0t})f_{0t}X_{it}']\right)\right\| = O_P\left(\frac{(N\bar{L})^{1/J}}{\sqrt{Th^4}}\right) + O(\epsilon/h^3).$$

Choosing $\epsilon = 1/\sqrt{T}$ and $J > (p+r+2)/(2c)$, we have

$$O_P\left(\frac{(N\bar{L})^{1/J}}{\sqrt{Th^4}}\right) + O(\epsilon/h^3) = O_P(1/\sqrt{Th^6}) = o_P(1).$$

Therefore, we can conclude that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\beta^*, \lambda_i^*, \tilde{f}_{0t}) f_{0t} X_{it}'\right\| = O_P(1)$$

and that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t \cdot [X_{it}'(\hat{\beta}-\beta_0)]^2\right\| = o_P(\|\hat{\beta}-\beta_0\|). \tag{A.30}$$

Step 8: Similar to the proof of Step 7, we can show that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t \cdot [X_{it}'(\hat{\beta}-\beta_0)] \cdot [(f_t^*)'(\hat{\lambda}_i - \check{\lambda}_{0i})]\right\| = o_P(\|\hat{\beta}-\beta_0\|), \tag{A.31}$$

and

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t \cdot [(f_t^*)'(\hat{\lambda}_i - \check{\lambda}_{0i})]^2\right\| = O_P\left(\max_i \|\hat{\lambda}_i - \lambda_{0i}\|^2\right), \tag{A.32}$$

Step 9: From Lemma 2 and Lemma S1 it follows easily that:

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t \cdot [X_{it}'(\hat{\beta}-\beta_0)] \cdot [(\lambda_i^*)'(\hat{f}_t - \check{f}_{0t})]\right\| = o_P(\|\hat{\beta}-\beta_0\|), \tag{A.33}$$

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t \cdot [(\lambda_i^*)'(\hat{f}_t - \check{f}_{0t})] \cdot [(f_t^*)'(\hat{\lambda}_i - \check{\lambda}_{0i})]\right\| = O_P(1/\sqrt{Th^4}) \cdot O_P\left(\max_i \|\hat{\lambda}_i - \lambda_{0i}\|\right), \tag{A.34}$$

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\hat{f}_t \cdot [(\lambda_i^*)'(\hat{f}_t - \check{f}_{0t})]^2\right\| = O_P(1/(Th^2)). \tag{A.35}$$

Finally, since $\max_i \|\tilde{\Omega}_i^{-1}\| \le \|\hat{\mathbf{H}}^{-1}\|^2 \cdot \max_i \|\Omega_i^{-1}\| = O_P(1)$ by Assumption 3 (ii), it follows from (A.24) to (A.35) that:

$$\max_{1\le i\le N}\|\hat{\lambda}_i - \lambda_{0i}\| = O_P(\|\hat{\beta}-\beta_0\|) + o_P\left(\max_i \|\hat{\lambda}_i - \lambda_{0i}\|\right) + O_P(1/T^{1/2-\nu}), \tag{A.36}$$

then the desired result follows from (A.36) and Lemma 4. □

Lemma 6. *Under Assumptions 1 to 4,*

$$\tilde{\Omega}_i(\hat{\lambda}_i - \check{\lambda}_{0i}) = \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}\check{f}_{0t} + \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t - \check{f}_{0t}) - \left(\tilde{\Omega}_i - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\check{f}_{0t}\check{f}_{0t}'\right)(\hat{\lambda}_i - \check{\lambda}_{0i})$$
$$- \left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}\check{f}_{0t}(\hat{f}_t - \check{f}_{0t})'\right)\check{\lambda}_{0i} + \bar{O}_P(\|\hat{\beta} - \beta_0\|) + \bar{O}_P(1/(T^{1-\nu}h^2)) + \bar{O}_P(1/T^{1-2\nu}). \quad \text{(A.37)}$$

Proof. The result follows immediately from the proof of Lemma 5. □

From (A.15) and the proof of Lemma 4 we have:

$$\Delta(\hat{\beta} - \beta_0) = o_P(\|\hat{\beta} - \beta_0\|) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(1)} Z_{it}$$
$$- \left[S^{\beta f'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda f'}\right](\hat{F} - \check{F}_0) - \left[\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}\right](\hat{\Lambda} - \check{\Lambda}_0)$$
$$- 1/2\sum_{i=1}^{N}\sum_{h=1}^{r}(S_*^{\beta\lambda'\lambda_{ih}} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda\lambda'\lambda_{ih}})(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{\Lambda} - \check{\Lambda}_0)$$
$$- 1/2\sum_{i=1}^{N}\sum_{h=1}^{r}(S_*^{\beta f'\lambda_{ih}} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda f'\lambda_{ih}})(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{F} - \check{F}_0)$$
$$- 1/2\sum_{t=1}^{T}\sum_{h=1}^{r}(S_*^{\beta\lambda' f_{th}} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda\lambda' f_{th}})(\hat{f}_{th} - \check{f}_{0t,h})(\hat{\Lambda} - \check{\Lambda}_0)$$
$$- 1/2\sum_{t=1}^{T}\sum_{h=1}^{r}(S_*^{\beta f' f_{th}} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda f' f_{th}})(\hat{f}_{th} - \check{f}_{0t,h})(\hat{F} - \check{F}_0). \quad \text{(A.38)}$$

In the next 5 lemmas, we analyze each term on the right-hand side of (A.38).

Lemma 7. *Under Assumptions 1 to 4,*

$$[\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}](\hat{\Lambda} - \check{\Lambda}_0) = -\frac{b_1}{T} + o_P(T^{-1}) + o_P(\|\hat{\beta} - \beta_0\|).$$

Proof. **Step 1**: Note that we can write:

$$[\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}](\hat{\Lambda} - \tilde{\Lambda}_0) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\hat{\mathbf{H}}(\hat{\lambda}_i - \check{\lambda}_{0i}).$$

Plugging in the result of Lemma 6, we have

$$
\begin{aligned}
&[\tilde{S}^{\beta\lambda'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\tilde{S}^{\lambda\lambda'}](\hat{\Lambda} - \check{\Lambda}_0) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right) + \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}(\hat{\mathbf{H}}')^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t - \check{f}_{0t})\right) \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t}(\hat{f}_t - \check{f}_{0t})'\right)\tilde{\lambda}_{0i} \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\Omega_i - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t} f_{0t}'\right)(\hat{\mathbf{H}}')^{-1}(\hat{\lambda}_i - \check{\lambda}_{0i}) \\
&\qquad + \left[O_P(\|\hat{\beta} - \beta_0\|) + O_P(1/(T^{1-\nu}h^2)) + O_P(1/T^{1-2\nu})\right] \cdot O_P\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right\|\right). \quad \text{(A.39)}
\end{aligned}
$$

Step 2: By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}(\hat{\mathbf{H}}')^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t - \check{f}_{0t})\right)\right\| \\
&\qquad \lesssim \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right\|^2} \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t - \check{f}_{0t})\right\|^2}.
\end{aligned}
$$

First, by Theorem 3 of Yoshihara (1978),

$$
\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right\|^2 = \frac{1}{Th^2}\cdot\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\cdot l_{it}^{(2)} Z_{it} f_{0t}'\right\|^2 = O_P(1/(Th^2)).
$$

Second, by Lemma 1, we have

$$
\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t - \check{f}_{0t})\right\|^2 \le \left\|\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} l_{it}^{(1)} e_{jt}\right\|^2 \cdot \left\|\hat{\Psi}\right\|^2 \le O_P(1)\cdot\left\|\frac{1}{NT}\sum_{j\neq i}^{N}\sum_{t=1}^{T} l_{it}^{(1)} e_{jt}\right\|^2 + O_P(1/N^2)\cdot\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)} e_{it}\right\|^2.
$$

For simplicity, consider the case where $p = 1$. Then by the mixing property we have

$$
\begin{aligned}
\mathbb{E}\left\|\frac{1}{NT}\sum_{j\neq i}^{N}\sum_{t=1}^{T} l_{it}^{(1)} e_{jt}\right\|^2 &= \frac{1}{N^2T^2}\sum_{j\neq i}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[l_{it}^{(1)} l_{is}^{(1)}]\cdot\mathbb{E}[e_{jt}e_{js}] \\
&= \frac{1}{NT}\cdot\frac{1}{N}\sum_{j\neq i}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[l_{it}^{(1)} l_{is}^{(1)}]\cdot\mathbb{E}[e_{jt}e_{js}]\right) = O_P((NT)^{-1}) = O_P(1/T^2).
\end{aligned}
$$

There, it follows that

$$
\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}(\hat{\mathbf{H}}')^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)}(\hat{f}_t - \check{f}_{0t})\right)\right\| = O_P\left(\frac{1}{T}\cdot\frac{1}{\sqrt{Th^2}}\right) = o_P(T^{-1}). \quad \text{(A.40)}
$$

Similarly, it can be shown that

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t}(\hat{f}_t - \check{f}_{0t})'\right)\tilde{\lambda}_{0i}\right\| = O_P\left(\frac{1}{T}\cdot\frac{1}{\sqrt{Th^4}}\right) = o_P(T^{-1}). \quad (A.41)$$

Step 3: By the Cauchy-Schwarz inequality and Lemma 5,

$$\begin{aligned}
&\left\| \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\Omega_i - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t} f_{0t}'\right)(\hat{\mathbf{H}}')^{-1}(\hat{\lambda}_i - \check{\lambda}_{0i})\right\| \\
&\lesssim \max_{1\le i\le N}\|\hat{\lambda}_i - \check{\lambda}_{0i}\| \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right\|^2} \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(2)} - \mathsf{f}_{it}(0)) f_{0t} f_{0t}'\right\|^2} \\
&= O_P(1/T^{1/2-\nu}) \cdot O_P(1/\sqrt{Th^2}) \cdot \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(2)} - \mathsf{f}_{it}(0)) f_{0t} f_{0t}'\right\|^2}.
\end{aligned}$$

Moreover, similar to the proof of the previous step, we can show that

$$\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(2)} - \mathsf{f}_{it}(0)) f_{0t} f_{0t}'\right\|^2 = O_P(1/(Th^2)).$$

Thus, by Assumption 3(vii)

$$\left\| \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\Omega_i - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} f_{0t} f_{0t}'\right)(\hat{\mathbf{H}}')^{-1}(\hat{\lambda}_i - \check{\lambda}_{0i})\right\| = O_P\left(\frac{1}{T}\cdot\frac{1}{T^{0.5-2c-\nu}}\right) = o_P\left(\frac{1}{T}\right) \quad (A.42)$$

because $0.5 - 2c - \nu > 0$.

Step 4: By Lemma S1 and Assumption 3(vii), we can write

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right) &= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} \tilde{l}_{it}^{(1)} f_{0t}\right) + o_P(h^q) \\
&\frac{1}{T}\cdot\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \tilde{l}_{it}^{(1)} f_{0t}\right) + o_P(T^{-1}). \quad (A.43)
\end{aligned}$$

First, by Lemma S1 and the mixing property, we have

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}l_{it}^{(2)}Z_{it}f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{l}_{it}^{(1)}f_{0t}\right)\right]=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[l_{it}^{(2)}\tilde{l}_{is}^{(1)}Z_{it}f_{0t}'\Omega_i^{-1}f_{0s}]\right)$$

$$=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[l_{it}^{(2)}\tilde{l}_{it}^{(1)}Z_{it}f_{0t}'\Omega_i^{-1}f_{0t}]\right)+\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}\mathbb{E}[l_{it}^{(2)}\tilde{l}_{is}^{(1)}Z_{it}f_{0t}'\Omega_i^{-1}f_{0s}]\right)$$

$$=(\tau-0.5)\cdot\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(1)}+\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(2)}+O(Th^q)=-b_1+o(1),$$

where

$$\omega_{T,i}^{(1)}=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]f_{0t}'\Omega_i^{-1}f_{0t},$$

$$\omega_{T,i}^{(2)}=\frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}\left(\tau\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}]-\mathbb{E}\left[\int_{\infty}^{0}\mathsf{f}_{i,ts}(0,v|X_{it},X_{is})dv\cdot Z_{it}\right]\right)f_{0t}'\Omega_i^{-1}f_{0s},$$

and

$$b_1=-(\tau-0.5)\cdot\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(1)}-\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(2)}.$$

Second, similar to the proof the Lemma A6 of Galvao and Kato (2016), we can show that

$$\mathrm{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}l_{it}^{(2)}Z_{it}f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{l}_{it}^{(1)}f_{0t}\right)\right)=o(1).$$

Thus, we have

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}l_{it}^{(2)}Z_{it}f_{0t}'\right)\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{l}_{it}^{(1)}f_{0t}\right)=-b_1+o_P(1),$$

and the desired result follows from (A.39) to (A.43) □

Lemma 8. *Under Assumptions 1 to 4,*

$$[S^{\beta f'}-\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda f'}](\hat{F}-\check{F}_0)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'e_{it}-\frac{d_1}{N}+o_P(T^{-1}).$$

Proof. **Step 1:** Write

$$
\begin{aligned}
& [S^{\beta f'} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda f'}](\hat{F} - \check{F}_0) \\
= & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(2)} X_{it}\lambda_{0i}'\right)(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) - \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\Phi_i(\hat{\mathbf{H}})^{-1}(l_{it}^{(2)}\check{f}_{0t}\check{\lambda}_{0i}' - l_{it}^{(1)})\right)(\hat{f}_t - \check{f}_{0t}) \\
= & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}[l_{it}^{(2)} X_{it}\lambda_{0i}' - l_{it}^{(2)}\Phi_i f_{0t}\lambda_{0i}' + l_{it}^{(1)}\Phi_i]\right)(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) \\
= & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}[l_{it}^{(2)} Z_{it}\lambda_{0i}' + l_{it}^{(1)}\Phi_i]\right)(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) \\
= & \frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_t(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) + \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\{(l_{it}^{(2)} Z_{it} - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}])\lambda_{0i}' + l_{it}^{(1)}\Phi_i\right\}\right)(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}).
\end{aligned}
$$

Step 2: By Lemma 1, we have $\hat{f}_t - \check{f}_{0t} = \hat{\Psi}'\bar{e}_t$, thus,

$$
\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_t(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0' e_{it} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t\left[(\hat{\mathbf{H}})^{-1}\hat{\Psi}' - (\mathbf{H}_0)^{-1}\Psi_0'\right]e_{it}.
$$

It is easy to show that

$$
\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t\left[(\hat{\mathbf{H}})^{-1}\hat{\Psi}' - (\mathbf{H}_0)^{-1}\Psi_0'\right]e_{it}\right\| \le O_P((NT)^{-1/2})\cdot\left(\|\hat{\mathbf{H}} - \mathbf{H}_0\| + \|\hat{\Psi} - \Psi_0\|\right) = o_P(T^{-1}).
$$

Therefore,

$$
\frac{1}{T}\sum_{t=1}^{T}\mathbf{A}_t(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0' e_{it} + o_P(T^{-1}). \tag{A.44}
$$

Step 3: Since $\|\hat{\mathbf{H}} - \mathbf{H}_0\| = O_P(T^{-1/2})$ and $\|\hat{\Psi} - \Psi_0\| = O_P(T^{-1/2})$, we have

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}\left\{(l_{it}^{(2)} Z_{it} - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}])\lambda_{0i}' + l_{it}^{(1)}\Phi_i\right\}\right)(\hat{\mathbf{H}})^{-1}(\hat{f}_t - \check{f}_{0t}) \\
& = (1+o_P(1))\cdot\frac{1}{N}\cdot\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(l_{it}^{(2)} Z_{it} - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}])\lambda_{0i}' + l_{it}^{(1)}\Phi_i\right\}\right)(\mathbf{H}_0)^{-1}\Psi_0'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\right).
\end{aligned}
\tag{A.45}
$$

Next, by Lemma S1 and Assumption 3(iv), it can be shown that

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(l_{it}^{(2)}Z_{it}-\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}])\lambda_{0i}'+l_{it}^{(1)}\Phi_i\right\}\right)(\mathbf{H}_0)^{-1}\Psi_0'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\right)\right]$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\left(l_{it}^{(2)}Z_{it}\lambda_{0i}'+l_{it}^{(1)}\Phi_i\right)(\mathbf{H}_0)^{-1}\Psi_0'e_{it}\right]$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it}]+o_P(1)=-d_1+o_P(1). \quad \text{(A.46)}$$

Step 4: Define

$$\mathcal{Z}_t=\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(l_{it}^{(2)}Z_{it}-\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}])\lambda_{0i}'+l_{it}^{(1)}\Phi_i\right\}\right)(\mathbf{H}_0)^{-1}\Psi_0'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\right).$$

To complete the proof, it remains to show that

$$\left\|\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{Z}_t\right)\right\|=o(1).$$

Note that

$$\left\|\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{Z}_t\right)\right\|\le\frac{1}{Th^2}\cdot\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h(\mathcal{Z}_t-\mathbb{E}[\mathcal{Z}_t)]\right\|^2.$$

By Assumption 3(iv), $\{\mathcal{Z}_1,\ldots,\mathcal{Z}_T\}$ is $\alpha-$mixing. Thus, by Theorem 3 of Yoshihara (1978) it suffices to show that $\mathbb{E}\|h\mathcal{Z}_t\|^4<\infty$ for all t. By the Cauchy-Schwarz inequality,

$$\mathbb{E}\|h\mathcal{Z}_t\|^4\lesssim\sqrt{\mathbb{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}h\left\{(l_{it}^{(2)}Z_{it}-\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}])\lambda_{0i}'+l_{it}^{(1)}\Phi_i\right\}\right\|^8}\cdot\sqrt{\mathbb{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}e_{it}\right\|^8}=O(1).$$

Thus, we have

$$\left\|\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{Z}_t\right)\right\|=O\left(\frac{1}{Th^2}\right)=o(1).$$

This completes the proof. □

Lemma 9. *Under Assumptions 1 to 4,*

$$\sum_{i=1}^{N}\sum_{h=1}^{r}(S_*^{\beta\lambda'\lambda_{ih}}-\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda\lambda'\lambda_{ih}})(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{\Lambda}-\check{\Lambda}_0)=-\frac{2b_2}{T}+o_P(T^{-1})+o_P(\|\hat{\beta}-\beta_0\|).$$

Proof. Let Φ_{ik} be the kth row of $\Phi_i=\Xi_i\Omega_i^{-1}$. The kth element of $-\sum_{i=1}^{N}\sum_{h=1}^{r}(S_*^{\beta\lambda'\lambda_{ih}}-\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda\lambda'\lambda_{ih}})(\hat{\lambda}_{ih}-$

$\check{\lambda}_{0i,h})(\hat{\Lambda} - \check{\Lambda}_0)$ can be written as

$$\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{r} l_{it}^{(3)}(*)(X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^{*'}f_{th}^*(\hat{\lambda}_{ih} - \check{\lambda}_{0i,h})(\hat{\lambda}_i - \check{\lambda}_{0i}) \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{\lambda}_i - \check{\lambda}_{0i})' l_{it}^{(3)}(*)(X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^* f_t^{*'}(\hat{\lambda}_i - \check{\lambda}_{0i}) \\
= \; & \frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i - \check{\lambda}_{0i})'\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)(X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^* f_t^{*'}\right)(\hat{\lambda}_i - \check{\lambda}_{0i}).
\end{aligned}$$

Step 1: Note that

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)(X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^* f_t^{*'} \\
& \quad = \hat{\mathbf{H}}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)} Z_{it,k} f_{0t} f_{0t}'\right)\hat{\mathbf{H}}' + \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)(X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}F_t^*)(f_t^* f_t^{*'} - \hat{\mathbf{H}} f_{0t} f_{0t}' \hat{\mathbf{H}}) \\
& \qquad + \hat{\mathbf{H}}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)\Phi_{ik}(f_{0t} - (\hat{\mathbf{H}})^{-1}f_t^*)f_{0t}f_{0t}'\right)\hat{\mathbf{H}}' + \hat{\mathbf{H}}\left(\frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(3)}(*) - l_{it}^{(3)})Z_{it,k}f_{0t}f_{0t}'\right)\hat{\mathbf{H}}'.
\end{aligned}$$

Thus, by Lemma 2, Lemma 5 and Lemma S1 we have

$$\begin{aligned}
& \Bigg\| \frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i - \check{\lambda}_{0i})'\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(*)(X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^* f_t^{*'}\right)(\hat{\lambda}_i - \check{\lambda}_{0i}) \\
& \qquad\qquad - \frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i - \check{\lambda}_{0i})'\hat{\mathbf{H}}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)} Z_{it,k} f_{0t} f_{0t}'\right)\hat{\mathbf{H}}'(\hat{\lambda}_i - \check{\lambda}_{0i}) \Bigg\| \\
& = O_P(T^{1-2\nu})\cdot O_P(\sqrt{\log T}/\sqrt{Th^6}) + O_P(T^{1-2\nu})\cdot O_P(1/(T^{0.5-\nu}h^3)) = O_P\left(\frac{1}{T}\cdot\frac{1}{T^{0.5-3c-3\nu}}\right) = o_P(T^{-1}),
\end{aligned}$$

where we have used the fact that

$$|l_{it}^{(3)}(*) - l_{it}^{(3)}| \lesssim \left(\|\hat{\beta} - \beta_0\| + \|\hat{f}_t - \check{f}_{0t}\| + \|\hat{\lambda}_i - \check{\lambda}_{0i}\|\right)/h^3,$$

and thus

$$\begin{aligned}
\max_i \left\| \frac{1}{T}\sum_{t=1}^{T}(l_{it}^{(3)}(*) - l_{it}^{(3)})Z_{it,k}f_{0t}f_{0t}' \right\| & \lesssim \|\hat{\beta} - \beta_0\|/h^3 + \max_t \|\hat{f}_t - \check{f}_{0t}\|/h^3 + \max_i \|\hat{\lambda}_i - \check{\lambda}_{0i}\|/h^3 \\
& = O_P(\sqrt{\log T}/\sqrt{Th^6}) + O_P(1/(T^{0.5-\nu}h^3)).
\end{aligned}$$

Step 2: Recall that

$$\mathbf{C}_{i,k} = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]f_{0t}f_{0t}'.$$

First, by Lemma 5,

$$\left\|\frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i-\check{\lambda}_{0i})'\hat{\mathbf{H}}\left(\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(3)}Z_{it,k}f_{0t}f_{0t}'\right)\hat{\mathbf{H}}'(\hat{\lambda}_i-\check{\lambda}_{0i})-\frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i-\check{\lambda}_{0i})'\hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'(\hat{\lambda}_i-\check{\lambda}_{0i})\right\|$$
$$\leq O_P(T^{1-2\nu})\cdot\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(3)}Z_{it,k}f_{0t}f_{0t}'-\mathbf{C}_{i,k}\right\|.$$

Second, by Lemma S1 and Theorem 3 of Yoshihara (1978),

$$\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(3)}Z_{it,k}f_{0t}f_{0t}'-\mathbf{C}_{i,k}\right\|\leq\frac{1}{\sqrt{Th^4}}\sqrt{\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}h^2(l_{it}^{(3)}Z_{it,k}-\mathbb{E}[l_{it}^{(3)}Z_{it,k}])f_{0t}f_{0t}'\right\|^2}+o(T^{-1})=O(1/\sqrt{Th^4}).$$

Thus, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{r}l_{it}^{(3)}(*)(X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^{*\prime}f_{th}^*(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{\lambda}_i-\check{\lambda}_{0i})$$
$$=\frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i-\check{\lambda}_{0i})'\hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'(\hat{\lambda}_i-\check{\lambda}_{0i})+o_P(T^{-1})\quad\text{(A.47)}$$

because $O_P(T^{1-2\nu}\cdot T^{-1/2}h^{-2})=o_P(T^{-1})$.

Step 3: By Lemma 6 we can write

$$\hat{\lambda}_i-\check{\lambda}_{0i}=\tilde{\Omega}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(1)}\check{f}_{0t}+g_i+\bar{O}_P(\|\hat{\beta}-\beta_0\|)+\bar{O}_P(1/(T^{1-\nu}h^2))+\bar{O}_P(1/T^{1-2\nu}).$$

where

$$g_i=\tilde{\Omega}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(1)}(\hat{f}_t-\check{f}_{0t})-\tilde{\Omega}_i^{-1}\left(\tilde{\Omega}_i-\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(2)}\check{f}_{0t}\check{f}_{0t}'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})-\tilde{\Omega}_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}l_{it}^{(2)}\check{f}_{0t}(\hat{f}_t-\check{f}_{0t})'\right)\check{\lambda}_{0i}.$$

By the proof of Lemma 7, it can be shown that

$$\frac{1}{N}\sum_{i=1}^{N}\|g_i\|=O_P\left(\frac{1}{T^{1-\nu}h}\right).$$

Thus,

$$\left\| \frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i - \check{\lambda}_{0i})'\hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'(\hat{\lambda}_i - \check{\lambda}_{0i}) - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}'\right)\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right)\right\|$$

$$= O_P\left(\max_{1\le i\le N}\|\hat{\lambda}_i - \check{\lambda}_{0i}\|\right)\cdot\frac{1}{N}\sum_{i=1}^{N}\|g_i\| + o_P(\|\hat{\beta}-\beta_0\|) + O_P\left(\max_{1\le i\le N}\|\hat{\lambda}_i - \check{\lambda}_{0i}\|\right)\cdot O_P(1/(T^{1-\nu}h^2) + 1/T^{1-2\nu})$$

$$= O_P\left(\frac{1}{T^{1.5-2\nu}h^2} + \frac{1}{T^{1.5-3\nu}}\right) + o_P(\|\hat{\beta}-\beta_0\|) = o_P(T^{-1}) + o_P(\|\hat{\beta}-\beta_0\|).$$

because $0.5 - 2c - 2\nu > 0$ and $0.5 - 3\nu > 0$, and we can write

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i - \check{\lambda}_{0i})'\hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'(\hat{\lambda}_i - \check{\lambda}_{0i})$$
$$= \frac{1}{T}\cdot\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}'\right)\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right) + o_P(T^{-1}) + o_P(\|\hat{\beta}-\beta_0\|). \quad \text{(A.48)}$$

Step 4: First, by Lemma S1,

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}'\right)\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right)\right]$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left(l_{it}^{(1)}\right)^2 f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0t} + \frac{1}{N}\sum_{i=1}^{N}\cdot\frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}\left[l_{it}^{(1)}l_{is}^{(1)}\right] f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0s}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(3)} + \frac{1}{N}\sum_{i=1}^{N}\omega_{T,i}^{(4)} + o(1) = 2b_2 + o(1),$$

where

$$\omega_{T,i,k}^{(3)} = \tau(1-\tau)\cdot\frac{1}{T}\sum_{t=1}^{T} f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0t}$$

$$\omega_{T,i,k}^{(4)} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\{\mathbb{E}[\mathbf{1}\{u_{it}\le 0, u_{is}\le 0\}] - \tau^2\} f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0s}$$

and

$$b_{2,k} = 0.5\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i,k}^{(3)} + 0.5\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\omega_{T,i,k}^{(4)}.$$

Step 5: Finally, note that by Theorem 3 of Yoshihara (1978),

$$\left\| \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}'\right)\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right)\right]\right\|$$
$$\lesssim \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\left\|\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}'\right)\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}\right)\right\|^2 \lesssim \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} l_{it}^{(1)} f_{0t}'\right\|^4$$
$$= O(N^{-1}) = o(1),$$

it then follows from (A.47) and (A.48) that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{h=1}^{r} l_{it}^{(3)}(*)(X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*)f_t^{*'}f_{th}^*(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{\lambda}_i-\check{\lambda}_{0i}) = -2b_{2,k}+o_P(T^{-1})+o_P(\|\hat{\beta}-\beta_0\|),$$

and the desired result follows. □

Lemma 10. *Under Assumptions 1 to 4,*

$$\sum_{t=1}^{T}\sum_{h=1}^{r}(S_*^{\beta f' f_{th}} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda f' f_{th}})(\hat{f}_{th}-\check{f}_{0t,h})(\hat{F}-\check{F}_0) = -\frac{2d_2}{N} + o_P(T^{-1}).$$

Proof. **Step 1:**

Note that the kth element of $-\sum_{t=1}^{T}\sum_{h=1}^{r}(S_*^{\beta f' f_{th}} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S_*^{\lambda f' f_{th}})(\hat{f}_{th}-\check{f}_{0t,h})(\hat{F}-\check{F}_0)$ can be written as:

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T}\sum_{h=1}^{r}\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*](\lambda_i^*)'\lambda_{ih}^*\right)(\hat{f}_{th}-\check{f}_{0t,h})(\hat{f}_t-\check{f}_{0t}) \\ = & \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'}\right)(\hat{f}_t-\check{f}_{0t}) \\ & +2\cdot\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(2)}(*)\lambda_i^*\Phi_{ik}(\hat{\mathbf{H}})^{-1}\right)(\hat{f}_t-\check{f}_{0t}). \end{aligned}$$

Step 2:

First, by Lemma 2,

$$\left\|\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'}\right)(\hat{f}_t-\check{f}_{0t}) - \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}(\hat{f}_t-\check{f}_{0t})\right\|$$
$$\le O_P(\log T/T)\cdot\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'} - (\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\right\|.$$

Second, write

$$\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'} = \frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)} Z_{it,k}\check{\lambda}_{0i}(\check{\lambda}_{0i})'+\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}[\Phi_{ik}f_{0t}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\check{\lambda}_{0i}(\check{\lambda}_{0i})'$$

$$+\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\left(\lambda_i^*(\lambda_i^*)'-\check{\lambda}_{0i}(\check{\lambda}_{0i})'\right)+\frac{1}{N}\sum_{i=1}^{N}\left(l_{it}^{(3)}(*)-l_{it}^{(3)}\right)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\check{\lambda}_{0i}(\check{\lambda}_{0i})',$$

it then follows that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'}-(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\right\| \leq$$
$$O_P(1)\cdot\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)} Z_{it,k}\lambda_{0i}\lambda_{0i}'-\mathbf{D}_{t,k}\right\|+O_P\left(\max_t\|\hat{f}_t-\check{f}_{0t}\|/h^3\right)+O_P\left(\max_i\|\hat{\lambda}_i-\check{\lambda}_{0i}\|/h^3\right),$$

where we have used Lemma S1, Lemma 4, and the fact that

$$|l_{it}^{(3)}(*)-l_{it}^{(3)}| \lesssim \left(\|\hat{\beta}-\beta_0\|+\|\hat{f}_t-\check{f}_{0t}\|+\|\hat{\lambda}_i-\check{\lambda}_{0i}\|\right)/h^3.$$

Similar to the proof of Lemma 9, it can be shown that

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)} Z_{it,k}\lambda_{0i}\lambda_{0i}'-\mathbf{D}_{t,k}\right\| = O_P(1/\sqrt{Nh^4}).$$

Therefore, by Lemma 2 and Lemma 5,

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k}-\Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'}\right)(\hat{f}_t-\check{f}_{0t}) = \frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}(\hat{f}_t-\check{f}_{0t})$$
$$+O_P\left(\frac{\log T}{T}\right)\left[O_P\left(\frac{1}{\sqrt{Nh^4}}\right)+O_P\left(\frac{\log T}{\sqrt{Nh^6}}\right)+O_P\left(\frac{1}{T^{0.5-\nu}h^3}\right)\right]$$
$$=\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}(\hat{f}_t-\check{f}_{0t})+o_P(T^{-1}). \quad \text{(A.49)}$$

Step 3: By Lemma 2 we can write

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t-\check{f}_{0t})'(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}(\hat{f}_t-\check{f}_{0t}) = \frac{1}{N}\cdot\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)'\hat{\Psi}(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\hat{\Psi}'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)$$
$$=\frac{1}{N}\cdot\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)'\Psi_0(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)+\frac{1}{N}\cdot o_P\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right\|^2\right)$$
$$=\frac{1}{N}\cdot\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)'\Psi_0(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)+o_P(N^{-1}).$$

Define

$$\mathcal{Z}_t = \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right)' \Psi_0(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0' \left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} e_{it}\right).$$

First, it is easy to show that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\mathcal{Z}_t] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\Psi_0(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\}.$$

Second, we have

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{Z}_t\right) = \frac{1}{T}\cdot\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathcal{Z}_t - \mathbb{E}[\mathcal{Z}_t]\right)\right|^2.$$

Since the process $\{\mathcal{Z}_1,\ldots,\mathcal{Z}_T\}$ is $\alpha-$mixing, it then follows from $\mathbb{E}|\mathcal{Z}_t|^4 \lesssim \mathbb{E}\|N^{-1/2}\sum_{i=1}^{N} e_{it}\|^8 < \infty$ that

$$\mathbb{E}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\mathcal{Z}_t - \mathbb{E}[\mathcal{Z}_t]\right)\right|^2 = O(1)$$

and thus

$$\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T}\mathcal{Z}_t\right) = O(T^{-1}) = o(1).$$

Therefore, we have

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - \check{f}_{0t})'\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(3)}(*)[X_{it,k} - \Phi_{ik}(\hat{\mathbf{H}})^{-1}f_t^*]\lambda_i^*\lambda_i^{*'}\right)(\hat{f}_t - \check{f}_{0t})$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\Psi_0(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\} + o_P(T^{-1}). \quad \text{(A.50)}$$

Step 4: Similarly, we can show that:

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - \check{f}_{0t})'\left(\frac{1}{N}\sum_{i=1}^{N} l_{it}^{(2)}(*)\lambda_i^*\Phi_{ik}(\hat{\mathbf{H}})^{-1}\right)(\hat{f}_t - \check{f}_{0t}) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\} + o_P(T^{-1}).$$

This completes the proof. □

Lemma 11. *Under Assumptions 1 to 4,*

$$\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0,ih})(\hat{F} - \check{F}_0) = o_P(T^{-1}), \quad \sum_{t=1}^{T}\sum_{h=1}^{r} S_*^{\beta\lambda' f_{th}}(\hat{f}_{th} - \check{f}_{0t,h})(\hat{\Lambda} - \check{\Lambda}_0) = o_P(T^{-1}),$$

$$\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\lambda f'\lambda_{ih}}(\hat{\lambda}_{ih} - \check{\lambda}_{0,ih})(\hat{F} - \check{F}_0) = o_P(T^{-1}),$$

$$\bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}\sum_{t=T}^{N}\sum_{h=1}^{r}S_*^{\lambda\lambda' f_{th}}(\hat{f}_{th}-\tilde{f}_{0t,h})(\hat{\Lambda}-\check{\Lambda}_0)=o_P(T^{-1}).$$

Proof. To save space, we only prove the first result. The proof of the other results are similar. Write

$$\sum_{i=1}^{N}\sum_{h=1}^{r}S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{F}-\check{F}_0)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[l_{it}^{(2)}(\beta^*,\lambda_i^*,f_t^*)X_{it}-l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(\lambda_i^*)'f_t^*](\hat{\lambda}_i-\check{\lambda}_{0i})'(\hat{f}_t-\check{f}_{0t}).$$

Let $R_{it}(*)=l_{it}^{(2)}(\beta^*,\lambda_i^*,f_t^*)X_{it}-l_{it}^{(3)}(\beta^*,\lambda_i^*,f_t^*)X_{it}(\lambda_i^*)'f_t^*$, and $R_{it}=l_{it}^{(2)}X_{it}-l_{it}^{(3)}X_{it}\lambda_{0i}'f_{0t}$, then by Lemma 1 and Lemma 4 we have

$$\sum_{i=1}^{N}\sum_{h=1}^{r}S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{F}-\check{F}_0)=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}R_{it}(*)(\hat{f}_t-\check{f}_{0t})'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})$$
$$=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}R_{it}(\hat{f}_t-\check{f}_{0t})'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})+O_P\left(\max_i\left\|\hat{\lambda}_i-\check{\lambda}_{0i}\right\|\right)\cdot O_P\left(\max_t\left\|\hat{f}_t-\check{f}_{0t}\right\|\right)\cdot O_P\left(\max_{i,t}\|R_{it}(*)-R_{it}\|\right).$$

Similar to the proof of Lemma 10, it can be shown that

$$\max_{i,t}\|R_{it}(*)-R_{it}\|\lesssim\left(\|\hat{\beta}-\beta_0\|+\|\hat{f}_t-\check{f}_{0t}\|+\|\hat{\lambda}_i-\check{\lambda}_{0i}\|\right)/h^3=O_P(1/(T^{0.5-\nu}h^3)),$$

thus, by Lemma 2 and Lemma 6,

$$O_P\left(\max_i\left\|\hat{\lambda}_i-\check{\lambda}_{0i}\right\|\right)\cdot O_P\left(\max_t\left\|\hat{f}_t-\check{f}_{0t}\right\|\right)\cdot O_P\left(\max_{i,t}\|R_{it}(*)-R_{it}\|\right)$$
$$=O_P\left(\frac{\sqrt{\log T}}{\sqrt{T}}\cdot\frac{1}{T^{0.5-\nu}}\cdot\frac{1}{T^{0.5-\nu-3c}}\right)=o_P(T^{-1})$$

because $0.5-2\nu-3c>0$. Therefore,

$$\sum_{i=1}^{N}\sum_{h=1}^{r}S_*^{\beta f'\lambda_{ih}}(\hat{\lambda}_{ih}-\check{\lambda}_{0i,h})(\hat{F}-\check{F}_0)=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}R_{it}(\hat{f}_t-\check{f}_{0t})'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})+o_P(T^{-1}).\qquad\text{(A.51)}$$

Note that

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}R_{it}(\hat{f}_t-\check{f}_{0t})'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})\right\|\leq\max_i\left\|\hat{\lambda}_i-\check{\lambda}_{0i}\right\|\cdot\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}R_{it}(\hat{f}_t-\check{f}_{0t})'\right\|.$$

Similar to Step 2 of the proof of Lemma 7, we can show that

$$\mathbb{E}\left\|\frac{1}{T}\sum_{t=1}^{T}R_{it}(\hat{F}_t-\tilde{F}_{0t})'\right\|=O(T^{-1}h^{-2}),$$

it then follows from Lemma 6 that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}R_{it}(\hat{f}_t-\check{f}_{0t})'\right)(\hat{\lambda}_i-\check{\lambda}_{0i})=O_P\left(\frac{1}{T^{3/2-\nu}h^2}\right)=o_P(T^{-1})\qquad\text{(A.52)}$$

because $0.5 - \nu - 2c > 0$. Thus, it follows from (A.51) and (A.52) that

$$\sum_{i=1}^{N}\sum_{h=1}^{r} S_*^{\beta f' \lambda_{ih}}(\hat{\lambda}_{ih} - \tilde{\lambda}_{0,ih})(\hat{F} - \tilde{F}_0) = o_P(T^{-1}).$$

The proofs of the other results are similar and thus are omitted. □

Proof of Theorem 2

Proof. It follows from (A.38) and Lemma 7 to Lemma 11 that

$$\Delta(\hat{\beta}-\beta_0) = -\left[S^{\beta} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda}\right] - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0' e_{it} + \frac{b}{T} + \frac{d}{N} + o_P(T^{-1}) + o_P(\|\hat{\beta}-\beta_0\|).$$

Since

$$S^{\beta} - \bar{S}^{\beta\lambda'}(\bar{S}^{\lambda\lambda'})^{-1}S^{\lambda} = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(1)} Z_{it},$$

it then follows from Assumption 3(vii) that

$$\sqrt{NT}(\hat{\beta}-\beta_0) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta^{-1}(l_{it}^{(1)} Z_{it} - \mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0' e_{it}) + \Delta^{-1}(\kappa b + \kappa^{-1} d) + o_P(1) + o_P(\sqrt{NT}\|\hat{\beta}-\beta_0\|).$$

Define $W_{it}^* = l_{it}^{(1)} Z_{it} - \mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0' e_{it}$ and $\bar{W}_i^* = T^{-1/2}\sum_{t=1}^{T} W_{it}^*$, it remains to show that

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} W_{it}^* = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{W}_i^* \xrightarrow{d} \mathcal{N}(0, \mathbf{V}). \tag{A.53}$$

First, by Lemma S1, we have $\mathbb{E}[\bar{W}_i^*] = O(h^q) = o(T^{-1})$, thus

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{W}_i^* = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\bar{W}_i^* - \mathbb{E}[\bar{W}_i^*]\right) + o(1). \tag{A.54}$$

Second, by the mixing property and Theorem 3 of Yoshihara (1978)

$$\mathbb{E}\left\|\bar{W}_i^* - \mathbb{E}[\bar{W}_i^*]\right\|^3 = \mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(W_{it}^* - \mathbb{E}[W_{it}^*])\right\|^3 < \infty.$$

Third, since $\bar{W}_1^*, \ldots, \bar{W}_N^*$ are independent, it follows from Lyapunov's CLT that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\bar{W}_i^* - \mathbb{E}[\bar{W}_i^*]\right) \xrightarrow{d} \mathcal{N}(0, \mathbf{V}^*) \tag{A.55}$$

where

$$\mathbf{V}^* = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[(\bar{W}_i^* - \mathbb{E}[\bar{W}_i^*])(\bar{W}_i^* - \mathbb{E}[\bar{W}_i^*])'\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[(W_{it}^* - \mathbb{E}[W_{it}^*])(W_{it}^* - \mathbb{E}[W_{it}^*])'\right]$$
$$+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}\left[(W_{it}^* - \mathbb{E}[W_{it}^*])(W_{is}^* - \mathbb{E}[W_{is}^*])'\right].$$

Finally, by Lemma S1 it is easy to show that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[(W_{it}^* - \mathbb{E}[W_{it}^*])(W_{it}^* - \mathbb{E}[W_{it}^*])'\right] = \tau(1-\tau)\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[Z_{it}Z_{it}']$$
$$+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'\mathbb{E}[e_{it}e_{it}']\Psi_0(\mathbf{H}_0')^{-1}\mathbf{A}_t' + o(1) = \mathbf{V}_1 + o(1),$$

and

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}\left[(W_{it}^* - \mathbb{E}[W_{it}^*])(W_{is}^* - \mathbb{E}[W_{is}^*])'\right]$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}[(\tau^2 - \mathbb{E}[\mathbf{1}\{u_{it}\leq 0\}|X_{it},X_{is}] - \mathbb{E}[\mathbf{1}\{u_{is}\leq 0\}|X_{it},X_{is}] + \mathbb{E}[\mathbf{1}\{u_{it}\leq 0\},u_{is}\leq 0\}|X_{it},X_{is}])Z_{it}Z_{is}']$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbb{E}[(\tau - \mathbb{E}[\mathbf{1}\{u_{it}\leq 0\}|X_{it},X_{is}])Z_{it}e_{is}]\Psi_0(\mathbf{H}_0')^{-1}\mathbf{A}_s'$$
$$- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'\mathbb{E}[e_{it}Z_{is}'(\tau - \mathbb{E}[\mathbf{1}\{u_{is}\leq 0\}|X_{it},X_{is}])]$$
$$+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s\neq t}^{T}\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'\mathbb{E}[e_{it}e_{is}]\Psi_0(\mathbf{H}_0')^{-1}\mathbf{A}_s' + o(1) = \mathbf{V}_2 + o(1)$$

Thus, we have $\mathbf{V}^* = \mathbf{V} + o(1)$, and (A.53) follows from (A.54) and (A.55). This completes the proof. □

A.4 Proof of Theorem 3

Lemma 12. *Under Assumptions 1 to 4, we have*
(i) $\max_{1\leq i\leq N}\|\hat{\Xi}_i - \tilde{\Xi}_i\| = O_P(1/(T^{0.5-\nu}h))$, $\max_{1\leq i\leq N}\|\hat{\Omega}_i - \tilde{\Omega}_i\| = O_P(1/(T^{0.5-\nu}h))$, $\max_{1\leq i\leq N}\|\hat{\Phi}_i - \Phi_i\hat{\mathbf{H}}^{-1}\| = O_P(1/(T^{0.5-\nu}h))$, $\max_{i,t}\|\hat{Z}_{it} - Z_{it}\| = O_P(1/(T^{0.5-\nu}h))$.
(ii) $\max_{i,t}\|\hat{e}_{it} - e_{it}\| = O_P(\sqrt{\log N}/\sqrt{T})$.

Proof. **Step 1:** Adding the subtracting terms, we have

$$\hat{\Xi}_i - \tilde{\Xi}_i = \frac{1}{T}\sum_{t=1}^{T}l^{(2)}(\hat{u}_{it})X_{it}\hat{f}_t' - \frac{1}{T}\sum_{t=1}^{T}l_{it}^{(2)}X_{it}f_{0t}'\hat{\mathbf{H}}' + \frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(2)}X_{it} - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}]\right)f_{0t}'\cdot\hat{\mathbf{H}}'.$$

First, by Theorem 3 of Yoshihara (1978)

$$\mathbb{E}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} h\left(l_{it}^{(2)}X_{it} - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}]\right)f_{0t}'\right\|^{J} < M$$

for any $J \geq 2$. Choosing $J \geq 1/\nu$, it then follows from Lemma 2.2.2 of van der Vaart and Wellner (1996) that

$$\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(2)}X_{it} - \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}]\right)f_{0t}'\right\| = O_P\left(\frac{N^{1/J}}{\sqrt{T}h}\right) = O_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

Second,

$$\left\|\frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})X_{it}\hat{f}_t' - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}X_{it}f_{0t}'\hat{\mathbf{H}}'\right\| \leq \left\|\frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})X_{it}(\hat{f}_t - \check{f}_{0t})'\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T}\left(l^{(2)}(\hat{u}_{it}) - l_{it}^{(2)}\right)X_{it}f_{0t}'\right\|\cdot\left\|\hat{\mathbf{H}}'\right\|.$$

The first term on the right-hand side of the above inequality is $O_P(1/\sqrt{Nh^2})$ by Lemma 2. For the second term, using Taylor expansion and Lemma 5 we can write

$$\frac{1}{T}\sum_{t=1}^{T}\left(l^{(2)}(\hat{u}_{it}) - l_{it}^{(2)}\right)X_{it}f_{0t}' = -\frac{1}{T}\sum_{t=1}^{T}\left\{l_{it}^{(3)}[X_{it}'(\hat{\beta}-\beta_0)]X_{it}f_{0t}' + l_{it}^{(3)}[\tilde{f}_{0t}'(\hat{\lambda}_i - \tilde{\lambda}_{0i})]X_{it}f_{0t}' + l_{it}^{(3)}[\tilde{\lambda}_{0i}'(\hat{f}_t - \tilde{f}_{0t})]X_{it}f_{0t}'\right\} + \bar{O}_P\left(1/(T^{1-2\nu}h^3)\right).$$

By Theorem 2,

$$\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}[X_{it}'(\hat{\beta}-\beta_0)]X_{it}f_{0t}'\right\| = O_P\left(\frac{1}{\sqrt{NT}}\cdot\frac{1}{h^2}\right) = o_P\left(\frac{1}{\sqrt{T}h}\right).$$

Next, by Lemma 2 and Lemma 5,

$$\begin{aligned}\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}[\tilde{f}_{0t}'(\hat{\lambda}_i - \tilde{\lambda}_{0i})]X_{it}f_{0t}'\right\| &\leq \max_{1\leq i\leq N}\max_{1\leq k\leq r}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}X_{it}f_{0t}'\tilde{f}_{0t,k}\right\|\cdot\max_{1\leq i\leq N}\|\hat{\lambda}_i - \tilde{\lambda}_{0i}\| \\ &= O_P\left(\frac{1}{T^{0.5-\nu}}\right)\cdot\max_{1\leq i\leq N}\max_{1\leq k\leq r}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}X_{it}f_{0t}'\tilde{f}_{0t,k}\right\|.\end{aligned}$$

Similar to the proof of Lemma 4, it can be shown that

$$\max_{1\leq i\leq N}\max_{1\leq k\leq r}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}X_{it}f_{0t}'\tilde{f}_{0t,k}\right\| = O_P(1).$$

Thus, we have

$$\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}[\tilde{f}_{0t}'(\hat{\lambda}_i - \tilde{\lambda}_{0i})]X_{it}f_{0t}'\right\| = O_P\left(\frac{1}{T^{0.5-\nu}}\right) = o_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

Similarly,

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}[\tilde{\lambda}_{0i}'(\hat{f}_t-\tilde{f}_{0t})]X_{it}f_{0t}'\right\| \le \max_{1\le k\le r}\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}(\hat{f}_{t,k}-\tilde{f}_{0t,k})X_{it}f_{0t}'\tilde{\lambda}_{0i,k}\right\|$$
$$\lesssim \max_{1\le t\le T}\|\hat{f}_{0t}-\tilde{f}_{0t}\|\cdot\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}|l_{it}^{(3)}|\cdot\|X_{it}f_{0t}'\|.$$

Since $\mathbb{E}[|l_{it}^{(3)}|\cdot\|X_{it}f_{0t}'\|]=\bar{O}(h^{-1})$ by Lemma S1, and it can be shown that

$$\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\left\{|l_{it}^{(3)}|\cdot\|X_{it}f_{0t}'\|-\mathbb{E}[|l_{it}^{(3)}|\cdot\|X_{it}f_{0t}'\|]\right\}=O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right)=o_P(1),$$

and it follows from Lemma 2 that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l_{it}^{(3)}[\tilde{\lambda}_{0i}'(\hat{f}_t-\tilde{f}_{0t})]X_{it}f_{0t}'\right\|=O_P(\sqrt{\log T}/\sqrt{N})\cdot O_P(h^{-1})=o_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

Combining all the above results, we have

$$\max_{1\le i\le N}\|\hat{\Xi}_i-\tilde{\Xi}_i\|=O_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

It can be shown in a similar way that

$$\max_{1\le i\le N}\|\hat{\Omega}_i-\tilde{\Omega}_i\|=O_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

Moreover, by the definition of $\hat{\Phi}_i$ and Φ_i,

$$\max_{1\le i\le N}\|\hat{\Phi}_i-\Phi_i\hat{\mathbf{H}}^{-1}\|=\max_{1\le i\le N}\left\|\hat{\Xi}_i\hat{\Omega}_i^{-1}-\Xi_i\Omega_i^{-1}\hat{\mathbf{H}}^{-1}\right\|=\max_{1\le i\le N}\left\|\hat{\Xi}_i\hat{\Omega}_i^{-1}-\tilde{\Xi}_i\tilde{\Omega}_i^{-1}\right\|,$$

thus, it follows from the above results and the fact that $\tilde{\Omega}_i$ is positive definite with probability approaching 1 that

$$\max_{1\le i\le N}\|\hat{\Phi}_i-\Phi_i\hat{\mathbf{H}}^{-1}\|=O_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

Finally, by the definitions of Z_{it} and $\hat{Z}_{it}$ and the results above,

$$\max_{i,t}\left\|\hat{Z}_{it}-Z_{it}\right\|\lesssim\max_i\left\|\tilde{\Xi}_i-\hat{\Xi}_i\right\|+\max_i\left\|\tilde{\Omega}_i^{-1}-\hat{\Omega}_i^{-1}\right\|+\max_t\left\|\hat{f}_t-\check{f}_{0t}\right\|=O_P\left(\frac{1}{T^{0.5-\nu}h}\right).$$

Step 2: Define $X_i=(X_{i1},\ldots,X_{iT})'$, and $\hat{\Gamma}_i'=(\hat{F}'\hat{F})^{-1}\hat{F}'X_i$. Then we have $\hat{e}_{it}=X_{it}-\hat{\Gamma}_i\hat{f}_t$, and $e_{it}=X_{it}-\Gamma_i f_{0t}=X_{it}-\Gamma_i\hat{\mathbf{H}}^{-1}\check{f}_{0t}$. It follows that

$$\max_{i,t}\|\hat{e}_{it}-e_{it}\|\lesssim\max_i\|\hat{\Gamma}_i-\Gamma_i\hat{\mathbf{H}}^{-1}\|+\max_t\|\hat{f}_t-\check{f}_{0t}\|.$$

Write

$$\hat{\Gamma}_i' = \left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t\hat{f}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t X_{it}'\right) = \left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t\hat{f}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t f_{0t}'\right)\Gamma_i' + \left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t\hat{f}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t e_{it}'\right).$$

By Lemma 2, we have

$$\left(\frac{1}{T}\sum_{t=1}^T \hat{f}_t\hat{f}_t'\right)^{-1} = (\hat{\mathbf{H}}')^{-1}\left(\frac{1}{T}\sum_{t=1}^T f_{0t}f_{0t}'\right)^{-1}\hat{\mathbf{H}}^{-1} + O_P(N^{-1/2})$$

and

$$\frac{1}{T}\sum_{t=1}^T \hat{f}_t f_{0t}' = \hat{\mathbf{H}}\cdot\frac{1}{T}\sum_{t=1}^T f_{0t}f_{0t}' + O_P(N^{-1/2}).$$

Thus, it follows that

$$\max_i \|\hat{\Gamma}_i - \Gamma_i\hat{\mathbf{H}}^{-1}\| \leq O_P(N^{-1/2}) + \max_i \left\|\frac{1}{T}\sum_{t=1}^T \hat{f}_t e_{it}'\right\|.$$

Next, by the proof of Lemma 2,

$$\max_i \left\|\frac{1}{T}\sum_{t=1}^T \hat{f}_t e_{it}'\right\| \leq \max_i \left\|\frac{1}{T}\sum_{t=1}^T f_{0t} e_{it}'\right\| + O_P(N^{-1/2}) = O_P(\sqrt{\log N}/\sqrt{T}) + O_P(N^{-1/2}).$$

It then follows from the above results and Lemma 2 that $\max_{i,t}\|\hat{e}_{it} - e_{it}\| = O_P(\sqrt{\log N}/\sqrt{T})$. □

Proof of Theorem 3

Proof. **Step 1:** We first show that $\hat{\Delta} = \Delta + o_P(1)$.

Note that by the definition of Δ and Lemma S1,

$$\|\hat{\Delta} - \Delta\| \leq \left\|\hat{\Delta} - \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T l_{it}^{(2)} Z_{it}Z_{it}\right\| + \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left\{l_{it}^{(2)} Z_{it}Z_{it} - \mathbb{E}[l_{it}^{(2)} Z_{it}Z_{it}]\right\}\right\| + o(1).$$

Similar to Step 2 in the proof of Lemma 4, it can be shown that

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \left\{l_{it}^{(2)} Z_{it}Z_{it} - \mathbb{E}[l_{it}^{(2)} Z_{it}Z_{it}]\right\} = o_P(1).$$

Thus, it remains to show that

$$\left\|\hat{\Delta} - \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T l_{it}^{(2)} Z_{it}Z_{it}\right\| = o_P(1).$$

By the definition of $\hat{\Delta}$, we have

$$\left\|\hat{\Delta} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it} Z_{it}\right\| \lesssim \max_{i,t}\left\|l^{(2)}(\hat{u}_{it}) - l_{it}^{(2)}\right\| + \max_{i,t}\left\|\hat{Z}_{it} - Z_{it}\right\|/h.$$

First, by Lemma 2 and Lemma 5,

$$\begin{aligned}\max_{i,t}\left\|l^{(2)}(\hat{u}_{it}) - l_{it}^{(2)}\right\| &\lesssim \left(\|\hat{\beta} - \beta_0\| + \max_i\left\|\hat{\lambda}_i - \lambda_{0i}\right\| + \max_t\left\|\hat{f}_t - f_{0t}\right\|\right)/h^2 \\ &= O_P(\sqrt{\log T}/\sqrt{Nh^4}) + O_P(1/(T^{0.5-\nu}h^2)) = o_P(1)\end{aligned}$$

because $0.5 - \nu - 2c > 0$. Second, by Lemma 12

$$\max_{i,t}\left\|\hat{Z}_{it} - Z_{it}\right\|/h = O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right) = o_P(1)$$

because $0.5 - 2c - \nu > 0$. Therefore, the desired result follows.

Step 2: Next, we show that $\hat{b} = b + o_P(1)$.

By definitions, it suffices to show that

$$\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(1)} - \omega_{T,i}^{(1)}\right\| = o_P(1), \quad \max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(2)} - \omega_{T,i}^{(2)}\right\| = o_P(1),$$

and

$$\max_{1\le i\le N}\left\|\hat{\omega}_{T,i,k}^{(3)} - \omega_{T,i,k}^{(3)}\right\| = o_P(1), \quad \max_{1\le i\le N}\left\|\hat{\omega}_{T,i,k}^{(4)} - \omega_{T,i,k}^{(4)}\right\| = o_P(1)$$

for $k \le r$.

First, similar to the proof of the previous step, we have

$$\begin{aligned}&\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(1)} - \omega_{T,i}^{(1)}\right\| = \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]f_{0t}'\Omega_i^{-1}f_{0t}\right\| \\ &\le \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - \frac{1}{T}\sum_{t=1}^{T} l_{it}^{(2)}Z_{it}\cdot f_{0t}'\Omega_i^{-1}f_{0t}\right\| + \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(2)}Z_{it} - \mathbb{E}\big[\mathsf{f}_{it}(0|X_{it})Z_{it}\big]\right)f_{0t}'\Omega_i^{-1}f_{0t}\right\| \\ &\lesssim \max_{i,t}\left\|l^{(2)}(\hat{u}_{it}) - l_{it}^{(2)}\right\| + \max_{i,t}\left\|\hat{Z}_{it} - Z_{it}\right\|/h + \max_{i,t}\left\|\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - f_{0t}'\Omega_i^{-1}f_{0t}\right\|/h + O_P\left(\frac{1}{T^{0.5-\nu}h}\right) \\ &= \max_{i,t}\left\|\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - f_{0t}'\Omega_i^{-1}f_{0t}\right\|/h + o_P(1).\end{aligned}$$

Note that by Lemma 2 and Lemma 12,

$$\begin{aligned}&\max_{i,t}\left\|\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - f_{0t}'\Omega_i^{-1}f_{0t}\right\|/h = \max_{i,t}\left\|\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - f_{0t}'\hat{\mathbf{H}}'(\hat{\mathbf{H}}')^{-1}\Omega_i^{-1}(\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}f_{0t}\right\|/h \\ &= \max_{i,t}\left\|\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_t - \check{f}_{0t}'\tilde{\Omega}_i^{-1}\check{f}_{0t}\right\|/h \lesssim \max_t\|\hat{f}_t - \check{f}_{0t}\|/h + \max_i\|\hat{\Omega}_i^{-1} - \tilde{\Omega}_i^{-1}\|/h = O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right) = o_P(1)\end{aligned}$$

because $0.5 - \nu - 2c > 0$. Thus, it follows that

$$\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(1)} - \omega_{T,i}^{(1)}\right\| = o_P(1).$$

Second, by Lemma 12

$$\begin{aligned}
\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(3)} - \omega_{T,i}^{(3)}\right\| &= \tau(1-\tau)\cdot\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_t - \frac{1}{T}\sum_{t=1}^{T}f_{0t}'\Omega_i^{-1}\mathbf{C}_{i,k}\Omega_i^{-1}f_{0t}\right\| \\
&\lesssim \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t'\hat{\Omega}_i^{-1}\hat{\mathbf{C}}_{i,k}\hat{\Omega}_i^{-1}\hat{f}_t - \frac{1}{T}\sum_{t=1}^{T}\check{f}_{0t}'\tilde{\Omega}_i^{-1}\cdot\hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'\cdot\tilde{\Omega}_i^{-1}\check{f}_{0t}\right\| \\
\le \max_t\|\hat{f}_t - \check{f}_{0t}\|/h^2 + \max_i\|\hat{\Omega}_i^{-1} - \tilde{\Omega}_i^{-1}\|/h^2 &+ \max_i\left\|\hat{\mathbf{C}}_{i,k} - \hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'\right\| = \max_i\left\|\hat{\mathbf{C}}_{i,k} - \hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'\right\| + o_P(1).
\end{aligned}$$

By the definitions of $\hat{\mathbf{C}}_{i,k}$ and $\mathbf{C}_{i,k}$, it follows from Lemma 12 that

$$\begin{aligned}
\max_i\left\|\hat{\mathbf{C}}_{i,k} - \hat{\mathbf{H}}\mathbf{C}_{i,k}\hat{\mathbf{H}}'\right\| &= \max_i\left\|\frac{1}{T}\sum_{t=1}^{T}l^{(3)}(\hat{u}_{it})\hat{Z}_{it,k}\hat{f}_t\hat{f}_t' - \frac{1}{T}\sum_{t=1}^{T}-\mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]\hat{\mathbf{H}}f_{0t}f_{0t}'\hat{\mathbf{H}}'\right\| \\
\le \max_i\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l^{(3)}(\hat{u}_{it})\hat{Z}_{it,k}\hat{f}_t\hat{f}_t' - l_{it}^{(3)}Z_{it,k}\check{f}_{0t}\check{f}_{0t}'\right)\right\| &+ \max_i\left\|\frac{1}{T}\sum_{t=1}^{T}\left(l_{it}^{(3)}Z_{it,k} + \mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})Z_{it,k}]\right)f_{0t}f_{0t}'\right\|\cdot\|\hat{\mathbf{H}}\|^2 \\
&\lesssim \max_t\|\hat{f}_t - \check{f}_{0t}\|/h^2 + \max_{i,t}\|\hat{u}_{it} - u_{it}\|/h^3 + \max_{i,t}\|\hat{Z}_{it} - Z_{it}\|/h^2 + O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right) \\
&= O_P\left(\frac{\sqrt{\log T}}{\sqrt{Nh^6}}\right) + O_P\left(\frac{1}{T^{0.5-\nu}h^3}\right) = o_P(1)
\end{aligned}$$

because $0.5 - \nu - 3c > 0$. Thus, we have

$$\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(3)} - \omega_{T,i}^{(3)}\right\| = o_P(1).$$

Third, by the proof Lemma 7 and Assumption 2(iv), it can be shown that

$$\omega_{T,i}^{(2)} = \frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\mathbb{E}\left[l_{it}^{(2)}Z_{it}l_{it}^{(1)}\right]\cdot f_{0t}'\Omega_i^{-1}f_{0s} + \frac{1}{T}\sum_{t=L+1}^{T}\sum_{s=t-L}^{t-1}\mathbb{E}\left[l_{it}^{(2)}Z_{it}l_{is}^{(1)}\right]\cdot f_{0t}'\Omega_i^{-1}f_{0s} + \bar{O}_P(Th^q) + O(\alpha^L).$$

Since $0 < \alpha < 1$ and $L \to \infty$ as $T \to \infty$, by the definition of $\hat{\omega}_{T,i}^{(2)}$, it follows that

$$\begin{aligned}
\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(2)} - \omega_{T,i}^{(2)}\right\| &\le \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{l^{(2)}(\hat{u}_{it})\hat{Z}_{it}l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s - \mathbb{E}\left[l_{it}^{(2)}Z_{it}l_{is}^{(1)}\right]\cdot f_{0t}'\Omega_i^{-1}f_{0s}\right\}\right\| \\
&+ \max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=L+1}^{T}\sum_{s=t-L}^{t-1}\left\{l^{(2)}(\hat{u}_{it})\hat{Z}_{it}l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s - \mathbb{E}\left[l_{it}^{(2)}Z_{it}l_{is}^{(1)}\right]\cdot f_{0t}'\Omega_i^{-1}f_{0s}\right\}\right\| + o_P(1).
\end{aligned}$$

Note that

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{l^{(2)}(\hat{u}_{it})\hat{Z}_{it}l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s-\mathbb{E}\left[l_{it}^{(2)}Z_{it}l_{it}^{(1)}\right]\cdot f_{0t}'\Omega_i^{-1}f_{0s}\right\}\right\|\le$$
$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{l^{(2)}(\hat{u}_{it})\hat{Z}_{it}l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s-l_{it}^{(2)}Z_{it}l_{it}^{(1)}\cdot f_{0t}'\Omega_i^{-1}f_{0s}\right\}\right\|$$
$$+\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{l_{it}^{(2)}Z_{it}l_{it}^{(1)}-\mathbb{E}\left[l_{it}^{(2)}Z_{it}l_{it}^{(1)}\right]\right\}\cdot f_{0t}'\Omega_i^{-1}f_{0s}\right\|$$

It can be shown that the second term on the right-hand side of the above inequality is $O_P(L/(T^{0.5-\nu}h))=o_P(1)$. For the first term, similar to the proof for $\hat{\omega}_{T,i}^{(1)}$ we have

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{l^{(2)}(\hat{u}_{it})\hat{Z}_{it}l^{(1)}(\hat{u}_{is})\cdot\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s-l_{it}^{(2)}Z_{it}l_{it}^{(1)}\cdot f_{0t}'\Omega_i^{-1}f_{0s}\right\}\right\|$$
$$\lesssim L\left(\max_{i,t}\left\|l^{(2)}(\hat{u}_{it})-l_{it}^{(2)}\right\|+\max_{i,t}\left\|l^{(1)}(\hat{u}_{it})-l_{it}^{(1)}\right\|/h+\max_{i,t}\left\|\hat{Z}_{it}-Z_{it}\right\|/h+\max_{t,s}\left\|\hat{f}_t'\hat{\Omega}_i^{-1}\hat{f}_s-f_{0t}'\Omega_i^{-1}f_{0s}\right\|/h\right)$$
$$=O_P\left(\frac{L}{T^{0.5-\nu}h^2}\right)=o_P(1).$$

Thus, we have

$$\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(2)}-\omega_{T,i}^{(2)}\right\|=o_P(1).$$

Finally, we show can that

$$\max_{1\le i\le N}\left\|\hat{\omega}_{T,i}^{(4)}-\omega_{T,i}^{(4)}\right\|=O_P\left(\frac{L}{T^{0.5-\nu}h^3}\right)=o_P(1)$$

in a similar way. Thus, we can conclude that $\hat{b}=b+o_P(1)$.

Step 3: Finally, we show that $\hat{d}=d+o_P(1)$, which follows from $\hat{d}_1=d_1+o_P(1)$ and $\hat{d}_{2,k}=d_{2,k}+o_P(1)$ for all $k\le r$.

First, by the definitions of d_1 and $\hat{d}_1$, we need to show that

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'\hat{\Psi}'\hat{e}_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it}]\right\|=o_P(1).$$

We have

$$\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it}] \right\|$$

$$\leq \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it} \right\|$$

$$+ \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it}] \right\|.$$

It is easy to see that the second term on the right-hand side of the above inequality is $O_P(1/\sqrt{NTh^2}) + O_P(h^q) = o_P(1)$. For the first term, by Lemma 2, Lemma 5 and Lemma 12 we have

$$\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}^{(2)} Z_{it}\lambda_{0i}'(\mathbf{H}_0)^{-1}\Psi_0'e_{it} \right\|$$

$$\leq \max_{i,t}\|l^{(2)}(\hat{u}_{it}) - l_{it}^{(2)}\| + \max_{i,t}\|\hat{Z}_{it} - Z_{it}\|/h + \max_{i}\|\hat{\lambda}_i - \tilde{\lambda}_i\|/h + \|\hat{\Psi} - \Psi_0\|/h + \max_{i,t}\|\hat{e}_{it} - e_{it}\|/h$$

$$= O_P\left(\frac{\sqrt{\log T}}{\sqrt{Nh^4}}\right) + O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right) = o_P(1).$$

Therefore, we can conclude that $\hat{d}_1 = d_1 + o_P(1)$.

Second, by the definitions of $d_{2,k}$ and $\hat{d}_{2,k}$, we need to show that

$$\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{e}_{it}'\hat{\Psi}\hat{\mathbf{B}}_{t,k}\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\cdot\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\} \right\| = o_P(1),$$

and

$$\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{e}_{it}'\hat{\Psi}\hat{\mathbf{D}}_{t,k}\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\cdot\Psi_0(\mathbf{H}_0')^{-1}\mathbf{D}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\} \right\| = o_P(1).$$

To save space, we only prove the first result. The proof of the second result is similar. Note that

$$\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{e}_{it}'\hat{\Psi}\hat{\mathbf{B}}_{t,k}\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\cdot\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\} \right\|$$

$$\leq \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{e}_{it}'\hat{\Psi}\hat{\mathbf{B}}_{t,k}\hat{\Psi}'\hat{e}_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}'\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'e_{it} \right\| +$$

$$\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}'\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'e_{it} - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{Tr}\left\{\mathbb{E}[e_{it}e_{it}']\cdot\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'\right\} \right\|.$$

It is easy to see that the second term on the right-hand side of the above inequality is $O_P(1/\sqrt{NT})$. For the first term, we have

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{e}_{it}'\hat{\Psi}\hat{\mathbf{B}}_{t,k}\hat{\Psi}'\hat{e}_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{it}'\Psi_0(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\Psi_0'e_{it}\right\|$$
$$\lesssim \max_{i,t}\|\hat{e}_{it}-e_{it}\|/h+\|\hat{\Psi}-\Psi_0\|/h+\max_{t}\left\|\hat{\mathbf{B}}_{t,k}-(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\right\|$$
$$=\max_{t}\left\|\hat{\mathbf{B}}_{t,k}-(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\right\|+O_P(\sqrt{\log N}/\sqrt{Th^2}).$$

By the definitions of $\hat{\mathbf{B}}_{t,k}$ and $\mathbf{B}_{t,k}$, it follows that

$$\max_{t}\left\|\hat{\mathbf{B}}_{t,k}-(\mathbf{H}_0')^{-1}\mathbf{B}_{t,k}(\mathbf{H}_0)^{-1}\right\|=\max_{t}\left\|\frac{1}{N}\sum_{i=1}^{N}l^{(2)}(\hat{u}_{it})\hat{\lambda}_i\hat{\Phi}_{i,k}-\frac{1}{N}\sum_{i=1}^{N}\mathsf{f}_{it}(0)(\mathbf{H}_0')^{-1}\lambda_{0i}\Phi_{i,k}(\mathbf{H}_0)^{-1}\right\|$$
$$\leq \max_{t}\left\|\frac{1}{N}\sum_{i=1}^{N}l^{(2)}(\hat{u}_{it})\hat{\lambda}_i\hat{\Phi}_{i,k}-\frac{1}{N}\sum_{i=1}^{N}l_{it}^{(2)}\tilde{\lambda}_{0i}\Phi_{i,k}(\mathbf{H}_0)^{-1}\right\|+\max_{t}\left\|\frac{1}{N}\sum_{i=1}^{N}[l_{it}^{(2)}-\mathsf{f}_{it}(0)]\tilde{\lambda}_{0i}\Phi_{i,k}\right\|\cdot\|\mathbf{H}_0^{-1}\|$$
$$\lesssim \max_{i,t}\|l^{(2)}(\hat{u}_{it})-l_{it}^{(2)}\|+\max_{i}\|\hat{\lambda}_i-\tilde{\lambda}_i\|/h+\max_{i}\|\hat{\Phi}_i-\Phi_i\hat{\mathbf{H}}^{-1}\|/h+\|\mathbf{H}_0-\mathbf{H}^{-1}\|/h+O_P(1/(T^{0.5-\nu}h))$$
$$=O_P\left(\frac{\sqrt{\log T}}{\sqrt{Nh^4}}\right)+O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right)=o_P(1).$$

Combining the all above results gives that $\hat{d}_{2,k}=d_{2,k}+o_P(1)$ for all $k\leq r$. This completes the proof of Theorem 3. □

A.5 Proof of Theorem 4

Proof. First, similar to the proof of Theorem 3, we have

$$\max_{1\leq t\leq T}\left\|\hat{\mathbf{A}}_t-\mathbf{A}_t\mathbf{H}_0^{-1}\right\|=\max_{1\leq t\leq T}\left\|\frac{1}{N}\sum_{i=1}^{N}l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'-\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}]\lambda_{0i}'\mathbf{H}_0^{-1}\right\|$$
$$\leq \max_{1\leq t\leq T}\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{l^{(2)}(\hat{u}_{it})\hat{Z}_{it}\hat{\lambda}_i'-l_{it}^{(2)}Z_{it}\tilde{\lambda}_{0i}'\right\}\right\|+\max_{1\leq t\leq T}\left\|\frac{1}{N}\sum_{i=1}^{N}\left\{l_{it}^{(2)}Z_{it}-\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}]\right\}\lambda_{0i}'\right\|\cdot\|\mathbf{H}_0^{-1}\|$$
$$\lesssim \max_{i,t}\|l^{(2)}(\hat{u}_{it})-l_{it}^{(2)}\|+\max_{i,t}\|\hat{Z}_{it}-Z_{it}\|/h+\max_{i}\|\hat{\lambda}_i-\tilde{\lambda}_i\|/h+O_P\left(\frac{1}{T^{0.5-\nu}h}\right)=O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right).$$

Second, it follows that

$$\max_{i,t}\|\hat{W}_{it}-W_{it}^*\|=\max_{i,t}\left\|l^{(1)}(\hat{u}_{it})\hat{Z}_{it}-\hat{\mathbf{A}}_t\hat{\Psi}'\hat{e}_{it}-[l_{it}^{(1)}Z_{it}-\mathbf{A}_t(\mathbf{H}_0)^{-1}\Psi_0'e_{it}]\right\|\lesssim\max_{i,t}\left\|l^{(1)}(\hat{u}_{it})-l_{it}^{(1)}\right\|$$
$$+\max_{i,t}\left\|\hat{Z}_{it}-Z_{it}\right\|+\max_{t}\left\|\hat{\mathbf{A}}_t-\mathbf{A}_t\mathbf{H}_0^{-1}\right\|+\|\hat{\Psi}-\Psi_0\|+\max_{i,t}\|\hat{e}_{it}-e_{it}\|=O_P\left(\frac{1}{T^{0.5-\nu}h^2}\right).$$

Third, similar to the proof of Theorem 3, we have

$$\begin{aligned}
\left\|\hat{\mathbf{V}}_2 - \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\sum_{s\neq t}^T \mathbb{E}[W_{it}^* W_{is}^{*'}]\right\| &\le \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{\hat{W}_{it}\hat{W}_{is}' - \mathbb{E}[W_{it}^*W_{is}^{*'}]\right\}\right\| \\
&\quad + \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1+L}^{T}\sum_{s=t-L}^{t-1}\left\{\hat{W}_{it}\hat{W}_{is}' - \mathbb{E}[W_{it}^*W_{is}^{*'}]\right\}\right\| + o_P(1) \\
&\le \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{\hat{W}_{it}\hat{W}_{is}' - W_{it}^*W_{is}^{*'}\right\}\right\| + \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^{T-L}\sum_{s=t+1}^{t+L}\left\{W_{it}^*W_{is}^{*'} - \mathbb{E}[W_{it}^*W_{is}^{*'}]\right\}\right\| \\
&\quad + \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1+L}^{T}\sum_{s=t-L}^{t-1}\left\{\hat{W}_{it}\hat{W}_{is}' - W_{it}^*W_{is}^{*'}\right\}\right\| + \left\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1+L}^{T}\sum_{s=t-L}^{t-1}\left\{W_{it}^*W_{is}^{*'} - \mathbb{E}[W_{it}^*W_{is}^{*'}]\right\}\right\| \\
&\lesssim L\cdot\max_{i,t}\|\hat{W}_{it} - W_{it}^*\| + O_P(L/\sqrt{NT}) = O_P\left(\frac{L}{T^{0.5-\nu}h^2}\right) = o_P(1).
\end{aligned}$$

By the proof of Theorem 2, it follows that $\hat{\mathbf{V}}_2 = \mathbf{V}_2 + o_P(1)$.

Finally, we can show that $\hat{\mathbf{V}}_1 = \mathbf{V}_1 + o_P(1)$ in a similar way. This completes the proof. □

B Some Auxiliary Lemmas

Lemma S1. *Let M be a generic bounded constant. Under Assumptions 1 to 3, it can be shown that (i) $\sup_{u\in\mathbb{R}} l^{(j)}(u)\cdot h^{j-1} \le M$ for $j = 1,\ldots,4$;*
(ii)

$$\mathbb{E}[l_{it}^{(1)}X_{it}] = O(h^q),\ \mathbb{E}[l_{it}^{(1)}] = O(h^q),\ \mathbb{E}[l_{it}^{(2)}X_{it}] = \mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}] + O(h^q),\ \mathbb{E}[l_{it}^{(2)}] = \mathsf{f}_{it}(0) + O(h^q),$$

$$\mathbb{E}[l_{it}^{(3)}X_{it}] = -\mathbb{E}[\mathsf{f}_{it}^{(1)}(0|X_{it})X_{it}] + O(h^{q-1}),\quad \mathbb{E}[l_{it}^{(2)}l_{it}^{(1)}X_{it}] = (\tau - 0.5)\cdot\mathbb{E}[\mathsf{f}_{it}(0|X_{it})X_{it}] + o(1),$$

$$\mathbb{E}[l_{it}^{(2)}l_{is}^{(1)}X_{it}] = \tau\mathbb{E}[\mathsf{f}_{it}(0|X_{it})Z_{it}] - \mathbb{E}\left[\int_\infty^0 \mathsf{f}_{i,ts}(0,v|X_{it},X_{is})dv\cdot Z_{it}\right] + o(1),$$

$$\mathbb{E}\left[\left(l_{it}^{(1)}\right)^2\right] = \tau(1-\tau) + o(1),\ \mathbb{E}\left[l_{it}^{(1)}l_{is}^{(1)}\right] = \mathbb{E}[\mathbf{1}\{u_{it}\le 0, u_{is}\le 0\}] - \tau^2 + o(1),$$

$$\mathbb{E}\left[\left(l_{it}^{(1)}\right)^2 Z_{it}Z_{it}'\right] = \tau(1-\tau)\cdot\mathbb{E}[Z_{it}Z_{it}'] + o(1).$$

(iii) $\max_{i,t}\sup_{\beta\in\mathcal{B},\lambda_i\in\mathcal{A}}\mathbb{E}[l^{(3)}(Y_{it} - \beta' X_{it} - \lambda_i\tilde{f}_{0t})|X_{it}] \le M$ almost surely.

Proof. The proof of the above results follow from standard calculations of nonparametric kernel estimators, and can be found in Horowitz (1998) or Galvao and Kato (2016). Thus, it is omitted. □

References

Galvao, A. F. and K. Kato (2016). Smoothed quantile regression for panel data. *Journal of Econometrics 193*(1), 92–112.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica*, 1327–1351.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.

Yoshihara, K.-I. (1978). Moment inequalities for mixing sequences. *Kodai Mathematical Journal 1*(2), 316–328.