Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity

Ando, Tomohiro and Bai, Jushan

University of Melbourne, Columbia University

30 June 2018

Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity

Tomohiro Ando [1] and Jushan Bai [2]

Abstract

This paper introduces a new procedure for analyzing the quantile co-movement of a large number of financial time series based on a large-scale panel data model with factor structures. The proposed method attempts to capture the unobservable heterogeneity of each of the financial time series based on sensitivity to explanatory variables and to the unobservable factor structure. In our model, the dimension of the common factor structure varies across quantiles, and the factor structure is allowed to be correlated with the explanatory variables. The proposed method allows for both cross-sectional and serial dependence, and heteroskedasticity, which are common in financial markets.

We propose new estimation procedures for both frequentist and Bayesian frameworks. Consistency and asymptotic normality of the proposed estimator are established. We also propose a new model selection criterion for determining the number of common factors together with theoretical support.

We apply the method to analyze the returns for over 6,000 international stocks from over 60 countries during the subprime crisis, European sovereign debt crisis, and subsequent period. The empirical analysis indicates that the common factor structure varies across quantiles. We find that the common factors for the quantiles and the common factors for the mean are different.

Keywords: Data-augmentation; Endogeneity; Heterogeneous panel; Quantile factor structure; Serial and cross-sectional correlations.

[1]Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.

[2]Department of Economics, Columbia University, jb3064@columbia.edu. 1019 International Affairs Building 420 West 118 Street New York, NY 10027 USA

1 Introduction

The goal of this paper is to develop a new statistical method for analyzing the quantile co-movement of a large number of financial time series and to empirically investigate the quantile co-movement structure of the global financial market. In the context of the arbitrage pricing theory (APT) of Ross (1976), the first theoretically grounded multifactor model in asset pricing (Goyal et al (2008)), the asset return variation of each security is explained by a linear combination of common factors, which captures the co-movements, plus the idiosyncratic return. A large body of asset pricing literature has considered models that explain expected returns or the mean structure (see, for example, Fama and French (2016), Griffin (2002), Hou et al. (2011) and references therein). However, little is known about the quantile co-movement structure of asset return distributions despite the fact that identifying the sources of co-movement is an important issue in asset pricing and risk management in finance. The chief obstacle to this investigation is that common factors that capture the quantile structure of asset returns may not be measurable/accessible in practice. Ideally, one would directly use measurable/accessible factors, such as Fama and French (1993)'s three factors. However, in reality, even for explaining expected asset returns, there is limited access to all common factors (Ando and Bai (2015)). Econometric methods for analyzing the quantile co-movement of a large number of financial time series and the effects of common factors on the asset return distribution, rather than the mean, remain limited.

The U.S. subprime crisis of 2007 led to massive declines in global financial markets, which subsequently affected economic activities worldwide. The Dow Jones Industrial Average hit the bottom in April 2009; the long-term interest rates of Euro zone countries started to increase at the end of 2009, the onset of the European sovereign debt crisis. From the perspective of governmental policy, regulators and asset management, it is important to understand the quantile co-movement structure during such extreme events. Similar to asset pricing studies that search for factors that explain the co-movement of expected returns in global stock markets (Fama and French (1998), Heston and Rouwenhorst (1994), Griffin

(2002), Hou et al. (2011)), it is worthwhile to identify the determinants of the quantile co-movement structure of the global financial market. By analyzing a large number of stock returns in the financial industry stock, this paper seeks to answer the following empirical questions.

1) Do the quantile common-factor structures that explain the asset-return distribution vary across quantiles?

2) Are the quantile common-factor structures symmetric in the sense that their structures in the lower tails and the upper tails are identical?

3) Are the quantile common-factor structures in the tails and at the mean different?

4) Are the co-movements of quantiles captured by the stock's listed exchange and industry?

5) Are there any special characteristics of the quantile structures of financial markets during the recent financial crisis and European sovereign debt crisis compared with the subsequent period?

To address these important but challenging empirical questions, we introduce a new heterogeneous panel quantile model with factor structures, in which a few unobservable factors may explain the co-movement of the asset return distributions in a large number of asset returns. Quantile regression methods have been previously used to model financial data (Engle and Manganelli (2004), Baur (2012), Baur et al. (2013), Chuang et al. (2009), Cappiello et al. (2014), Chen (2015), So and Chung (2015), Gerlach et al. (2016), Chen, Li and Nguyen (2017), Han et al. (2016)). In this paper, we consider large-scale panel data that consist of a large number of asset return time series. There is a growing number of studies on panel quantile models (see Koenker (2004), Abrevaya and Dahl (2008), Lamarche (2010), Kato et al. (2012), Harding and Lamarche (2014), and Chen et al. (2017), among others). In particular, we introduce a new panel data model with heterogeneous regression coefficients, which has many attractive features that are lacking in those used in the above studies. First, the heterogeneity of asset returns is captured by using heterogeneous regression coefficients and a factor error structure. Second, observable factors can be correlated with the unobservable common factors, factor loadings or both. Third, the unobservable common factors are allowed to vary across quantiles. The model is formulated without imposing any parametric family of distributions. We note that this is the first study that introduces and analyzes such a general model.

If we can ignore the unobservable common factor structures, then the estimator of the regression coefficient can be found by running the standard quantile regression approach equation by equation. However, the model allows for correlation between the unobservable effects and the explanatory variables. If the unobservable common factor structure exists and is ignored, which implies ignoring possible endogeneity, then the standard quantile regression approach produces biased results. Indeed, our simulation study indicates that failing to account for endogeneity increases the bias of the estimation. Even if the true regression coefficients are zero, a direct application of the principal component approach in the quantile panel model may yield inconsistent estimation of the unobservable factor structure (Chen et al. (2017)).

In the case of mean panel data models with factor structures, the inference procedure is well studied. Indeed, one can employ various estimation procedures, including Bai (2009) for homogeneous panels, Song (2013) for heterogeneous panels, and Ando and Bai (2015) for heterogeneous panels with shrinkage. However, these estimation procedures cannot be applied to the panel quantile models with interactive effects. Although some studies on non-linear panel studies are available (Freyberger (2015), Chen et al. (2014), and Fernandez-Val and Weidner (2016)), these studies focused on smoothed objective functions and homogeneous coefficients, and thus are not directly applicable to our settings. Furthermore, our model allows for a large number of parameters.

To overcome this issue, we propose new estimation procedures for both a frequentist and Bayesian framework. Our newly developed algorithm quickly searches the frequentist estimator. In a standard Bayesian quantile regression for cross-sectional data, Markov chain Monte Carlo (MCMC) is commonly employed. Assuming the asymmetric Laplace distribution for the error term, the MCMC posterior sampling procedure is well studied (Yu and Moyeed (2001), Geraci and Bottai (2007), Yue and Rue (2011)). However, these studies ignored the issue of "endogeneity", where the set of regressors are correlated with the error terms. Although Lancaster and Jun (2010) studied the Bayesian estimation of the quantile regression model with endogeneity, their study addresses cross-sectional data and a borrowed Bayesian exponentially tilted empirical likelihood framework. Obviously, these studies on MCMC for estimating the cross-sectional quantile regression model can not be easily extended to estimate panel quantile regression models with interactive fixed effects. There are no studies that consider the Bayesian MCMC procedure for estimating the panel quantile regression models with interactive fixed effects. This is the first study to investigate a data-augmentation approach to the analysis of panel quantile regression models with endogeneity. We develop a data-augmentation strategy without imposing any probability distributions on the error term. We demonstrate that our method will greatly simplify inference on the unobservable factor structure. Our Monte Carlo simulation study shows that the proposed procedure improves the estimation accuracy of the underlying quantile structures in the presence of interactive fixed effects.

In practical applications, the number of common factors should be determined. We note that previous studies (including Ahn and Horenstein (2013), Ando and Bai (2016, 2017), Bai and Ng (2002), Hallin and Liska (2007) and Onatski (2009)) cannot be applied directly because these methods were designed for panel "mean" regression models with factor structures instead of panel "quantile" models. As there are no studies that allow us to determine the number of common factors, this issue is not straightforward. We propose a new information criterion for selecting the number of common factors. Our simulation study indicates that the proposed information criterion is capable of selecting the true dimension of the common factors.

We make further theoretical contributions by developing an asymptotic theory for the proposed estimator. We establish the consistency and the asymptotic normality of the estimator. We also establish the model selection consistency of the information criterion in determining the number of common factors. In our asymptotic framework, the time series dimension and individual dimension are diverging. Due to the presence of the unobservable common factor structures and non-smoothness of the quantile loss function, the development of these results is non-trivial. Therefore, we need a novel strategy for the proof.

In summary, our contributions are as follows. First, a novel panel heterogeneous quantile model with a factor structure is introduced. Second, new estimation procedures are developed for the simultaneous estimation of heterogeneous regression coefficients and the factor structures. Third, the consistency and asymptotic normality of the proposed frequentist estimator are established. Fourth, a novel information criterion for determining the number of common factors is proposed together with theoretical supports. Finally, these new results are applied to investigate the quantile co-movement structure of global financial markets. In contrast to Ando and Bai (2017), which focused on the subprime crisis period, we compared the quantile structures of the subprime crisis period, European sovereign debt crisis period, and the subsequent period. As a result, a number of interesting findings are obtained. It is found that quantile common-factor structures in the tails and at the mean are different. Because Ando and Bai (2017)'s method is designed for exploring the mean structures, this paper complements Ando and Bai (2017) as it provides a useful tool for exploring the quantile structures.

The remainder of this paper is organized as follows. Section 2 introduces a new panel quantile model with a factor structure and its assumptions. Section 3 develops the parameter estimation procedures. Section 4 proposes the new information criterion for determining the number of common factors. Section 5 investigates the consistency and the asymptotic distribution of the estimator. Section 6 applies the procedure to the analysis of global stock market data. Concluding remarks are provided in Section 7. To save space, all technical proofs of the theoretical results and Monte Carlo simulations are provided in the online supplementary document.

2 Panel quantile regression with interactive fixed effects

2.1 The Model

Suppose that the response of an individual unit is measured over T time periods together with some observable regressors. For the i-th unit ($i = 1, ..., N$), at time t, its response y_{it} is observed together with a $(p+1)$-dimensional vector of observable regressors $\boldsymbol{x}_{it} = (1, x_{it,1}, ..., x_{it,p})'$. We consider the following model for the τ-th conditional quantile function of the response y_{it};

$$Q_{y_{it}}\left(\tau | \boldsymbol{x}_{it}, \boldsymbol{f}_{t,\tau}, \boldsymbol{\lambda}_{i,\tau}\right) \equiv \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} + \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T, \tag{1}$$

where $\boldsymbol{b}_{i,\tau} = (b_{i,0,\tau}, b_{i,1,\tau}, ..., b_{i,p,\tau})'$ is a $(p+1)$-dimensional vector of regression coefficients, where $\boldsymbol{f}_{t,\tau}$ is an $r_\tau \times 1$ vector of unobservable factors and $\boldsymbol{\lambda}_{i,\tau}$ represents the unobservable factor loadings. For notational simplicity, we often write $\eta_{it,\tau} = \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}$. This unobservable factor structure is known as the interactive effect in the econometric literature (e.g., Bai, 2009) and employed in asset pricing (e.g., Ando and Bai (2017)). This interactive effect provides a convenient way of modeling the cross-sectional dependence of asset returns. If we specify the dimension of the unobservable factor as $r_\tau = 1$ and the corresponding factor loading is constant over all possible τ, model (1) reduces to panel quantile regression models with individual fixed effects (Koenker (2004), Kato et al. (2012)). In contrast to these studies, model (1) allows the dimension of unobservable factor $\boldsymbol{f}_{t,\tau}$ to depend on quantile τ.

Here are some examples. The first example is an interactive effect model in the mean, $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i + \varepsilon_{it}$, where ε_{it} are independent over i, but are i.i.d over t. Let $q_i(\tau)$ be the τ-th quantile of ε_{it}, then,

$$Q_{y_{it}}\left(\tau | \boldsymbol{x}_{it}, \boldsymbol{f}_t, \boldsymbol{\lambda}_i\right) = \boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i + q_i(\tau).$$

We can absorb $q_i(\tau)$ into the coefficient of the constant regressor. Next, consider $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_i(u_{it}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i$, where u_{it} are i.i.d. $U(0,1)$, independent of $\{\boldsymbol{x}_{it}, \boldsymbol{f}_t, \boldsymbol{\lambda}_i\}$. Also, assume $\boldsymbol{x}_{it}'\boldsymbol{b}_i(u_{it})$ is increasing in u_{it}. Then,

$$Q_{y_{it}}\left(\tau | \boldsymbol{x}_{it}, \boldsymbol{f}_t, \boldsymbol{\lambda}_i\right) = \boldsymbol{x}_{it}'\boldsymbol{b}_i(\tau) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i.$$

The quantile literature treats $\boldsymbol{b}_i(\tau)$ as a non-random function of τ, and is interested in estimating this function. Finally, consider $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_i(u_{it}) + \boldsymbol{f}_t(u_{it})'\boldsymbol{\lambda}_i(u_{it})$, where u_{it} is $U(0,1)$, independent of $\boldsymbol{x}_{it}$. Also, assume that the right hand side of y_{it} is increasing in u_{it} (see Koenker (2005) for quantile regression models expressed as functions of uniform random variables), then

$$Q_{y_{it}}\left(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_t(\tau), \boldsymbol{\lambda}_i(\tau)\right) = \boldsymbol{x}_{it}'\boldsymbol{b}_i(\tau) + \boldsymbol{f}_t(\tau)'\boldsymbol{\lambda}_i(\tau).$$

Define the error term $\varepsilon_{it,\tau} \equiv y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}$, then $P(\varepsilon_{it,\tau} \leq 0|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,\tau}, \boldsymbol{\lambda}_{i,\tau}) = \tau$. Our purpose is to estimate the unknown conditional quantile function $Q_{y_{it}}\left(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,\tau}, \boldsymbol{\lambda}_{i,\tau}\right)$ based on the observations $\{(y_{it}, \boldsymbol{x}_{it});\ i = 1, \ldots, N,\ t = 1, \ldots, T\}$.

If one ignores the unobservable effects ($\eta_{it,\tau} = 0$), the quantile estimator of $\boldsymbol{b}_{i,\tau}$ ($i = 1, ..., N$) is found as the minimizer of the standard quantile loss function:

$$\ell_\tau(Y|X, B_\tau) = \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau}), \tag{2}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function, $Y \equiv \{y_{it}|i = 1, ..., N, t = 1, ..., T\}$, $X \equiv \{\boldsymbol{x}_{it}|i = 1, ..., N, t = 1, ..., T\}$ and $B_\tau = \{\boldsymbol{b}_{1,\tau}, ..., \boldsymbol{b}_{N,\tau}\}$.

However, if the factor structure exists and is ignored, the estimator of B_τ in the above is biased (see Section 5 and the Monte Carlo simulation result). This is because, as in Koenker (2004) and Kato et al. (2012), we allow for correlation between the unobservable effects $\eta_{it,\tau}$ and the observables $\boldsymbol{x}_{it}$ (existence of endogeneity). In such a case, we have to estimate the unknown parameters B_τ, $\Lambda_\tau = (\boldsymbol{\lambda}_{1,\tau}, ..., \boldsymbol{\lambda}_{N,\tau})'$, and $F_\tau = (\boldsymbol{f}_{1,\tau}, ..., \boldsymbol{f}_{T,\tau})'$ by simultaneously minimizing the following objective function

$$\ell_\tau(Y|X, B_\tau, F_\tau, \Lambda_\tau) = \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}) \tag{3}$$

A new parameter estimation procedure is proposed in Section 3. Because the likelihood function is nonlinear in the factor structure, inference on interactive fixed effects in model (1) is a challenging problem.

There are several past studies relating to our proposed model (1). Ando and Tsay (2011) consider a quantile regression model with factor-augmented predictors. In their study, the common factors that explain the quantile structure are also allowed to vary across quantiles

τ. In contrast to our study, however, their study is about the modeling of a quantile structure in the cross-sectional context. Harding and Lamarche (2014) consider a quantile regression model with interactive effects. In contrast to the heterogeneous regression coefficients in (1), their model only allows homogeneous regression coefficients. Moreover, in their setup, the common factor structure is not allowed to vary across quantiles. When we set the heterogeneous regression coefficients as $\boldsymbol{b}_{i,\tau} = \mathbf{0}$ for $i = 1, ..., N$, the model becomes similar to the quantile factor model of Chen et al. (2017) in the sense that the common factors vary across quantiles. We note that the parameter estimation procedure in model (1) becomes more complicated due to the existence of the term $\boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau}$. Furthermore, the investigation of asymptotic theory becomes challenging; the convergence rate for the estimated common factor structures and the convergence rate for the estimated regression coefficients depend on one another.

Remark 1 Past empirical studies (e.g., Nath and Brooks (2015), Ni et al. (2015)) indicated that a set of important (observable) common factors vary with τ. Thus, it is ideal to formulate model (1) so that unobservable common factors vary with quantile points τ because its dimension is often unknown for each τ in practice.

Before we propose the novel estimation procedures and the asymptotic theory, the set of assumptions on the model is clarified in the next section.

2.2 Assumptions

We first define some notations. Let $\|A\| = [tr(A'A)]^{1/2}$ be the usual norm of the matrix A, where "tr" denotes the trace of a square matrix. The equation $a_n = O(b_n)$ states that the deterministic sequence a_n is at most of order b_n; $c_n = O_p(d_n)$ states that the random variable c_n is at most of order d_n in terms of probability and $c_n = o_p(d_n)$ is of a smaller order in terms of probability. The true regression coefficient is denoted by $\boldsymbol{b}_{i,0,\tau}$. Further, $F_{0,\tau} = (\boldsymbol{f}_{1,0,\tau}, ..., \boldsymbol{f}_{T,0,\tau})'$ and $\Lambda_{0,\tau} = (\boldsymbol{\lambda}_{1,0,\tau}, ..., \boldsymbol{\lambda}_{N,0,\tau})'$ are the true common factor and its factor loadings. The set of regularity conditions that are imposed on the proposed model are as follows:

Assumption A: Common factors

Let $\mathcal{F}$ be a compact subset of R^{r_τ}. The common factors $\boldsymbol{f}_{t,0,\tau} \in \mathcal{F}$ satisfy $T^{-1}\sum_{t=1}^{T} \boldsymbol{f}_{t,0,\tau}\boldsymbol{f}_{t,0,\tau}{}' \to \Sigma_{F_\tau}$ as $T \to \infty$, where Σ_{F_τ} is an $r_\tau \times r_\tau$ positive definite matrix.

Assumption B: Factor loadings and regression coefficients

Let $\mathcal{B}$ and $\mathcal{L}$ be compact subsets R^{p+1} and R^{r_τ}. The regression coefficient $\boldsymbol{b}_{i,0,\tau}$ and the factor-loading for the common factors satisfy $\boldsymbol{b}_{i,0,\tau} \in \mathcal{B}$ and $\boldsymbol{\lambda}_{i,0,\tau} \in \mathcal{L}$. In addition, the factor-loading matrix $\Lambda_{0,\tau} = (\boldsymbol{\lambda}_{1,0,\tau}, \ldots, \boldsymbol{\lambda}_{N,0,\tau})'$ satisfies $N^{-1}\Lambda_{0,\tau}'\Lambda_{0,\tau}$ being a $r_\tau \times r_\tau$ positive definite matrix for all N.

Assumption C: Idiosyncratic error terms

(C1): The random variable $\varepsilon_{it,\tau} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0,\tau} - \boldsymbol{f}_{t,0,\tau}'\boldsymbol{\lambda}_{i,0,\tau}$ is independently distributed over i and t, conditional on X, $F_{0,\tau}$ and $\Lambda_{0,\tau}$. In addition, it satisfies $E[|\varepsilon_{it,\tau} - E[\varepsilon_{it,\tau}]|^K] < K!C_\varepsilon^K$ for $K \geq 1$ and positive constant $C_\varepsilon < \infty$.

(C2): The conditional density function of $\varepsilon_{it,\tau}$ given $(\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau})$, denoted as $g_{it}(\varepsilon_{it,\tau}|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau})$, is continuous. In addition, for any compact set $\mathcal{C}$, there exists a positive constant $\bar{g} > 0$ (depending on $\mathcal{C}$) such that $\inf_{c\in\mathcal{C}} g_{it}(c|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}) \geq \bar{g}$ for all i and t.

Assumption D: Predictors and design matrix

(D1): For a positive constant C_x, predictors satisfy $\sup_{it} \|\boldsymbol{x}_{it}\| < C_x < \infty$.

(D2): For each i and large T, there exist positive constants C_1 and C_2 such that

$$0 < C_1 < \lambda_{\min}(T^{-1}(X_i, F_{0,\tau})'(X_i, F_{0,\tau})) < \lambda_{\max}(T^{-1}(X_i, F_{0,\tau})'(X_i, F_{0,\tau})) < C_2 < \infty,$$

where $X_i = (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})'$, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest eigenvalue and largest eigenvalue of a matrix A, respectively.

(D3): Define $A_{i,\tau} = \frac{1}{T}X_i'M_{F_\tau}X_i$, $B_{i,\tau} = (\boldsymbol{\lambda}_{i,0,\tau}\boldsymbol{\lambda}_{i,0,\tau}') \otimes I_T$, $C_{i,\tau}' = \frac{1}{\sqrt{T}}\boldsymbol{\lambda}_{i,0,\tau}' \otimes (X_i'M_{F_\tau})$, $M_{F_\tau} = I - F_\tau(F_\tau'F_\tau)^{-1}F_\tau'$. Let $\mathcal{F}_\tau$ be the collection of F_τ such that $\mathcal{F}_\tau = \{F_\tau : F_\tau'F_\tau/T = I\}$. We assume

$$\inf_{F_\tau \in \mathcal{F}_\tau} \left[\frac{1}{N}\sum_{i=1}^{N} E_{i,\tau}(F_\tau)\right] \quad \text{is positive definite,} \tag{4}$$

where $E_{i,\tau}(F_\tau) = B_{i,\tau} - C'_{i,\tau} A^{-1}_{i,\tau} C_{i,\tau}$.

Remark 2 The full rank assumption in Assumptions A and B is necessary for the number of common factors to be r_τ. In Assumption C, heteroskedasticity is allowed. Although it is outside the scope of this paper, the errors can have serial correlation. In such a case, it will require more technical conditions such as those in Bai (2009). Assumption D imposes the regularity condition on design matrix X_i and common factor structure $F_{0,\tau}$. (D2) is the usual rank condition for identification. (D3) is similar to that used in Ando and Bai (2015) and Song (2013). Similar to Belloni and Chernozhukov (2011), Wang et al. (2012) and Tang et al. (2013), Sherwood and Wang (2016), the quantile function under a particular τ is focused rather than the entire quantile function over all possible τ. When the entire quantile function is the focus, there are studies that try to ensure the monotonicity of the quantile function as a function of τ, for example, He (1997), Bondell et al. (2010), Dette and Volgushev (2008), Chernozhukov et al. (2010) and Yuan et al. (2017). It is known that if non-crossing is required in linear quantile regression on an unbounded domain in any covariate direction, the quantile function results in the constant slope, location-shift model (Bondell et al. (2010)). Therefore, we consider a compact space for $\boldsymbol{x}_{it}$, $\boldsymbol{f}_{t,0,\tau}$, and $\boldsymbol{\lambda}_{i,0,\tau}$ in the assumptions.

3 Estimation

We begin by presenting the reason for which parameters B_τ, F_τ, and Λ_τ should be estimated simultaneously. One might consider a two-step procedure to estimate the model parameters. In the first step, the principal component analysis (see, e.g., Bai (2003), Connor and Korajzcyk (1986)) is applied to build the common factors. Plugging these estimated common factors into $\ell_\tau(Y|X, B_\tau, F_\tau, \Lambda_\tau)$ in (3), the second step jointly estimates the regression coefficients and factor loadings. However, as discussed in Chen et al. (2017) for pure factor models, when there exist common factors that affect the quantiles but not the means, the two-step procedure may result in inconsistent estimators due to the omission of important common factors in the second step purely because the principal component estimator cannot always expand the true common factor space for implementing the second step. Indeed, our empirical analysis indicates that there exist common factors that affect the quantiles but not the means. Therefore, the factor-augmented approach to quantile regression (Ando and Tsay (2011)) may not work, and the simultaneous direct minimization of $\ell_\tau(Y|X, B_\tau, F_\tau, \Lambda_\tau)$ in terms of B_τ, F_τ and Λ_τ is important. In Section 3.1, the frequentist estimation procedure is developed. We also propose the novel data-augmentation strategy in Section 3.2.

3.1 Frequentist estimation

Note that one cannot separately identify F_τ and Λ_τ without further restrictions because they enter the model in a multiplicative way. Following Bai and Ng (2013), we impose the following restrictions $F_\tau' F_\tau/T = I_{r_\tau}$ and $\Lambda_\tau' = (\Lambda_{1,\tau}', \Lambda_{2,\tau}')'$, with $\Lambda_{1,\tau}$ being an invertible lower triangular matrix. Bai and Ng (2013) demonstrate that this restriction will lead to the identification of F_τ and Λ_τ. We refer to Bai and Ng (2002, 2013) and Stock and Watson (2002)) for the identification of the principal component estimator for the mean panel data model. Then, the frequentist estimator is obtained by minimizing the following objective function $\ell_\tau(Y|X, B_\tau, F_\tau, \Lambda_\tau)$ in (3) under the restriction.

Given values of τ and the number of common factors r_τ, the following algorithm can be used to obtain the frequentist estimator.

Estimation algorithm:

Step 1. Initialize parameters $\hat{B}_\tau$, $\hat{F}_\tau$ and $\hat{\Lambda}_\tau$.

Step 2. Given $\hat{F}_\tau$, update $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{\boldsymbol{\lambda}}_{i,\tau}$ as the minimizer of $\sum_{t=1}^{T} \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \hat{\boldsymbol{f}}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau})$ for $i = 1, ..., N$.

Step 3. Given $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{\boldsymbol{\lambda}}_{i,\tau}$ $(i = 1, ..., N)$, update $\boldsymbol{f}_{t,\tau}$ as the minimizer of $\sum_{i=1}^{N} \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\hat{\boldsymbol{\lambda}}_{i,\tau})$ for $t = 1, ..., T$.

Step 4. Obtain a QR decomposition of $\hat{F}_\tau$ to yield $\hat{F}_\tau = \bar{Q}_\tau^F \bar{R}_\tau^F$, where $\bar{R}_\tau^F$ is an upper triangular matrix with positive diagonal elements and $\bar{Q}_\tau^F$ is an $T \times r_\tau$ orthogonal matrix such that $\bar{Q}_\tau^{F\prime} \bar{Q}_\tau^F = I_{r_\tau}$. Then, obtain a QR decomposition of $\bar{R}_\tau^F \hat{\Lambda}_\tau'$, to yield $\bar{R}_\tau^F \hat{\Lambda}_\tau' = \bar{Q}_\tau^\Lambda \bar{R}_\tau^\Lambda$. Here, $\bar{R}_\tau^\Lambda$ is an upper triangular matrix with positive diagonal elements, and $\bar{Q}_\tau^\Lambda$ is an $r_\tau \times r_\tau$ orthogonal matrix such that $\bar{Q}_\tau^{\Lambda\prime} \bar{Q}_\tau^\Lambda = I_{r_\tau}$. Update the common factor $\hat{F}_\tau$ and the factor loading matrix $\hat{\Lambda}_\tau$ as $\hat{F}_\tau = \sqrt{T}\bar{Q}_\tau^F \bar{Q}_\tau^\Lambda$ and $\hat{\Lambda}_\tau' = \bar{R}_\tau^\Lambda$, respectively.

Step 5. Repeat Step 2 $\sim$ Step 4 until convergence.

In Step 1, the initial values of parameters are prepared as follows. First, estimate $\boldsymbol{b}_{i,\tau}$ $(i = 1, ..., N)$ by minimizing $\ell_\tau(Y|X, B_\tau)$ in (2). Given $\hat{\boldsymbol{b}}_{i,\tau}$ $(i = 1, ..., N)$, define the variable $Z_\tau = (\boldsymbol{z}_{1,\tau}, \ldots, \boldsymbol{z}_{N,\tau})$ with $\boldsymbol{z}_{i,\tau} = \boldsymbol{y}_i - X_i \hat{\boldsymbol{b}}_{i,\tau}$. Obtain the principal components' estimate of $\hat{F}_\tau = (\hat{\boldsymbol{f}}_{1,\tau}, ..., \hat{\boldsymbol{f}}_{T,\tau})'$, subject to the normalization $F_\tau' F_\tau / T = I_{r_\tau}$, which is $\sqrt{T}$ times the eigenvectors corresponding to the r_τ largest eigenvalues of the $T \times T$ matrix $Z_\tau' Z_\tau$ (See Bai and Ng (2002)). Then, obtain $\hat{\boldsymbol{\lambda}}_{i,\tau}$ as the minimizer of $\sum_{t=1}^T \rho_\tau(y_{it} - \boldsymbol{x}_{it}' \hat{\boldsymbol{b}}_{i,\tau} - \hat{\boldsymbol{f}}_{t,\tau}' \boldsymbol{\lambda}_{i,\tau})$ for $i = 1, ..., N$. One can step the iteration based on $N^{-1} \sum_{i=1}^N \|\hat{\boldsymbol{b}}_{i,\tau}^{new} - \hat{\boldsymbol{b}}_{i,\tau}^{old}\|^2 + (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \{(\hat{\boldsymbol{f}}_{t,\tau}' \hat{\boldsymbol{\lambda}}_{i,\tau})^{new} - (\hat{\boldsymbol{f}}_{t,\tau}' \hat{\boldsymbol{\lambda}}_{i,\tau})^{old}\}^2 < \delta^2$ where δ^2 is a small value. Our simulation study found that the above algorithm converges quickly. In Section 4, we develop the asymptotic property of the frequentist estimator under large N and T.

Remark 3 This idea of Bai and Ng (2013) is employed in Step 4. We first note that the product $\hat{F}_\tau \hat{\Lambda}_\tau'$ remains the same even when the common factor and the factor loading matrix are rotated (see Bai and Ng (2013)). Regarding $\hat{F}_\tau$, we see that $\hat{F}_\tau' \hat{F}_\tau / T = (\sqrt{T}\bar{Q}_\tau^F \bar{Q}_\tau^\Lambda)'(\sqrt{T}\bar{Q}_\tau^F \bar{Q}_\tau^\Lambda)/T = \bar{Q}_\tau^{\Lambda\prime} \bar{Q}_\tau^\Lambda = I_{r_\tau}$. Furthermore, $\bar{R}_\tau^\Lambda$ is an upper triangular matrix with positive diagonal elements; thus, both $\hat{F}_\tau$ and $\hat{\Lambda}_\tau$ in Step 4 satisfy the restriction.

3.2 Data-augmentation approach for Bayesian inference

This section develops the data-augmentation approach for Bayesian inference. To implement the data-augmentation approach, we first define the pseudo likelihood

$$L(Y|X, B_\tau, F_\tau, \Lambda_\tau) = \exp\{-\ell_\tau(Y|X, B_\tau, F_\tau, \Lambda_\tau)\}$$

and specify the prior distribution of the parameters as $\pi(B_\tau, F_\tau, \Lambda_\tau)$. Similar to the frequentist estimator, the prior density $\pi(B_\tau, \Lambda_\tau)$ accommodates the following identification restriction: $F_\tau' F_\tau / T = I_{r_\tau}$ and $\Lambda_\tau' = (\Lambda_{1,\tau}', \Lambda_{2,\tau}')'$ with $\Lambda_{1,\tau}$ being an invertible lower triangular matrix.

Note that unlike previous studies on Bayesian inference in quantile regression that use the asymmetric Laplace distribution for the error component (Yu and Moyeed (2001), Geraci and Bottai (2007), Yue and Rue (2011)), we develop the data-augmentation strategy without directly imposing probability distributions. Then, the posterior density is $\pi(B_\tau, F_\tau, \Lambda_\tau|Y, X) \propto L(Y|X, F_\tau, \Lambda_\tau, B_\tau)\pi(B_\tau, F_\tau, \Lambda_\tau)$, which does not provide analytical density forms. Note that there is no easy method for sampling from their posterior distribution because the error distribution is unknown.

3.2.1 Data-augmentation strategy for F_τ

Because the common factor F_τ is subject to the normalization condition $F_\tau' F_\tau/T = I_{r_\tau}$ for identification purposes, F_τ belongs to a hyperball in T dimensions, and its support is restricted to being the Cartesian product of the T-dimensional hyperball. Furthermore, because of the orthogonality requirement, its support is then reduced to a Stiefel manifold S_{T,r_τ} of radius $\sqrt{T}$ (see, e.g., Khatri and Mardia (1977), Tsay and Ando (2012) and references therein). Thus, the prior of F_τ is a flat prior over the Stiefel manifold

$$\pi(F_\tau) = \frac{1}{C(T, r_\tau)} \cdot 1(F_\tau \in S_{T,r_\tau}), \tag{5}$$

where $1(\cdot)$ is the indicator function and $C(T,k) = 2^k \pi^{kT/2} T^{k(2T-k-1)/4} / \{\pi^{k(k-1)/4} \prod_{j=1}^k \Gamma\{(T-j+1)/2\}\}$ is the normalizing constant with $\Gamma(\cdot)$ being the Gamma function.

To derive the conditional posterior of F_τ, we use the following equality

$$\exp\left(-\{|\kappa| + (2\tau-1)\kappa\}\right) = \int_0^\infty \phi(\kappa|(1-2\tau)\omega, \omega)\exp(-2\tau(1-\tau)\omega)d\omega, \tag{6}$$

where $\phi(x|\mu,\omega)$ is the normal density, evaluated at x, for mean μ and variance ω (see, e.g., Polson and Scott (2013)). Using this result, the loss contribution of observation y_{it} can be expressed as

$$\begin{aligned} &\exp\left(-\{|y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}| + (2\tau-1)\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}\}\}\right) \\ \propto &\int_0^\infty \exp\left\{-\frac{\{z_{it,\tau} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}\}^2}{2\omega_{it,\tau}}\right\}\exp\{-2\tau(1-\tau)\omega_{it,\tau}\}d\omega_{it,\tau}, \end{aligned}$$

where $z_{it,\tau} = y_{it} - (1-2\tau)\omega_{it,\tau}$. Combining the terms from all observations yields the following expression for the conditional posterior of F_τ, given $\Omega_\tau \equiv \{\omega_{it,\tau} | i = 1, ..., N, t = 1, ..., T\}$:

$$
\begin{aligned}
& \pi(F_\tau | Y, X, B_\tau, \Lambda_\tau, \Omega_\tau) \\
\propto \;& \prod_{i=1}^{N} \prod_{t=1}^{T} \exp\left\{ -\frac{\{z_{it,\tau} - \boldsymbol{x}_{it}' \boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}' \boldsymbol{\lambda}_{i,\tau}\}^2}{2\omega_{it,\tau}} \right\} \times \pi(F_\tau) \\
\propto \;& \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} (\boldsymbol{z}_{i,\tau}^* - F_\tau \boldsymbol{\lambda}_{i,\tau})' \Omega_{i\tau}^{-1} (\boldsymbol{z}_{i,\tau}^* - F_\tau \boldsymbol{\lambda}_{i,\tau}) \right\} \times 1(F_\tau \in S_{T,r}),
\end{aligned} \tag{7}
$$

where $\Omega_{i,\tau} = \text{diag}\{\omega_{i1,\tau}, ..., \omega_{iT,\tau}\}$, $\boldsymbol{z}_{i,\tau}^* = (z_{i1,\tau}^*, ..., z_{iT,\tau}^*)$ with $z_{it,\tau}^* = z_{it,\tau} - \boldsymbol{x}_{it}' \boldsymbol{b}_{i,\tau}$.

Because the diagonal matrix $\Omega_{i,\tau}$ prevents the derivation of an analytical conditional posterior of F_τ, further analysis of the conditional posterior of F_τ in (7) is not straightforward. To generate the posterior sample of F_τ, we use the Metropolis-Hastings algorithm, which first draws a candidate parameter value F_τ^{new} from the proposal density $p(F_\tau)$. Then, this generated parameter value F_τ^{new} is accepted or rejected based on the acceptance probability

$$
\alpha = \min\left\{ 1, \frac{L(Y|X, F_\tau^{new}, \Lambda_\tau, B_\tau) \pi(B, F_\tau^{new}, \Lambda_\tau) / p(F_\tau^{new})}{L(Y|X, F_\tau^{old}, \Lambda_\tau, B_\tau) \pi(B_\tau, F_\tau^{old}, \Lambda_\tau) / p(F_\tau^{old})} \right\},
$$

where F_τ^{old} is the current state of F_τ.

In the practical implementation of the Metropolis-Hasting algorithm, we draw a new candidate F_τ^{new} from a proposal density. Here, a Bingham-von Mises-Fisher distribution with parameter $Z_\tau^* \Omega_{i,\tau}^{-1} \Lambda_\tau$ is employed for the proposal density. We refer to Hoff (2009) for generating a random matrix from Bingham-von Mises-Fisher distribution.

Remark 4 One can assume that $\varepsilon_{it,\tau} = y_{it} - \boldsymbol{x}_{it}' \boldsymbol{b}_{i,\tau} - \eta_{it,\tau}$ follows the asymmetric Laplace distribution or its variants, including Kozumi and Kobayashi (2011) and Yan and Kottas (2015). However, we show that it is possible to estimate the unknown parameters without imposing any probability distribution on $\varepsilon_{it,\tau}$.

3.2.2 Prior specification and posterior analysis of B_τ and Λ_τ

Here, we specify the prior densities on B_τ and Λ_τ and derive their conditional posterior distributions, given F_τ and Ω_τ. For simplicity of notation, we first express the loss contribution of observation y_{it} as

$$
\begin{aligned}
& \exp\left(-\{ |y_{it} - \boldsymbol{x}_{it}' \boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}' \boldsymbol{\lambda}_{i,\tau}| + (2\tau - 1)\{ y_{it} - \boldsymbol{x}_{it}' \boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}' \boldsymbol{\lambda}_{i,\tau} \} \} \right) \\
\propto \;& \int_0^\infty \exp\left\{ -\frac{\{z_{it,\tau} - \boldsymbol{v}_{it,\tau}' \boldsymbol{\gamma}_{i,\tau}\}^2}{2\omega_{it,\tau}} \right\} \exp\{-2\tau(1-\tau)\omega_{it,\tau}\} d\omega_{it,\tau},
\end{aligned}
$$

where $\boldsymbol{v}_{it,\tau} = (\boldsymbol{x}_{it}', \boldsymbol{f}_{t,\tau}')'$, and $\boldsymbol{\gamma}_{i,\tau} = (\boldsymbol{b}_{i,\tau}', \boldsymbol{\lambda}_{i,\tau}')'$.

We recall that the first r_τ factor loading vectors $\boldsymbol{\lambda}_{i,\tau}$ correspond to the invertible lower triangular matrix $\Lambda_{1,\tau}$, which comes from the identification restriction. If some elements of $\boldsymbol{\lambda}_{i,\tau}$ must be zero for identification purposes, we can ignore these elements in the estimation and denote the non-zero elements of $(\boldsymbol{b}_{i,\tau}', \boldsymbol{\lambda}_{i,\tau}')'$ as $\boldsymbol{\gamma}_{i,\tau}$. For these non-zero elements, we simply use the diffuse prior $\pi(\boldsymbol{\gamma}_{i,\tau}) \propto$ Const.. Then, the conditional posterior density of $\boldsymbol{\gamma}_{i,\tau}$ is

$$\pi(\boldsymbol{\gamma}_i | Y, X, F_\tau, B_{-i,\tau}, \Lambda_{-i,\tau}, \Omega_\tau) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{z}_{i,\tau} - W_{i,\tau}\boldsymbol{\gamma}_{i,\tau})'\Omega_{i,\tau}^{-1}(\boldsymbol{z}_{i,\tau} - W_{i,\tau}\boldsymbol{\gamma}_{i,\tau})\right\},$$

where $W_{i,\tau} = (X_i, F_\tau)$ is the design matrix, $B_{-i,\tau} = (\boldsymbol{b}_{1,\tau}, ..., \boldsymbol{b}_{i-1,\tau}, \boldsymbol{b}_{i+1,\tau}, ..., \boldsymbol{b}_{N,\tau})'$ and $\Lambda_{-i,\tau} = (\boldsymbol{\lambda}_{1,\tau}, ..., \boldsymbol{\lambda}_{i-1,\tau}, \boldsymbol{\lambda}_{i+1,\tau}, ..., \boldsymbol{\lambda}_N, \tau)'$. This implies that the conditional posterior density of $\boldsymbol{\gamma}_{i,\tau}$ is the multivariate normal density with mean $(W_{i,\tau}'\Omega_{i,\tau}^{-1}W_{i,\tau})^{-1}W_{i,\tau}'\Omega_{i,\tau}^{-1}\boldsymbol{z}_{i,\tau}$ and variance-covariance matrix $(W_{i,\tau}'\Omega_{i,\tau}^{-1}W_{i,\tau})^{-1}$.

Remark 5 When the dimension of $\boldsymbol{x}_{it}$ is large, one can also use a shrinkage prior on $\boldsymbol{\gamma}_{i,\tau}$, including the lasso prior (Park and Casella (2008)) and the adaptive lasso prior (Leng et al. (2014)). We can easily derive a conditional posterior distribution of $\boldsymbol{\gamma}_{i,\tau}$ based on these previous results. Or, one may also consider using a procedure for variable screening (He et al. (2013)) before applying our data augmentation procedure. However, these directions are outside the scope of this paper.

3.2.3 Prior specification and posterior analysis of $\omega_{it,\tau}$

We re-express the loss contribution of observation y_{it} as

$$\begin{aligned} &\exp\left(-\{|y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}|/2 + (\tau - 1/2)\{y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}\}\}\right) \\ \propto\ &\int_0^\infty \exp\left\{-\frac{\{a_{it,\tau} - (1-2\tau)\omega_{it,\tau}\}^2}{2\omega_{it,\tau}}\right\}\exp\{-2\tau(1-\tau)\omega_{it,\tau}\}d\omega_{it,\tau}, \end{aligned}$$

where $a_{it,\tau} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau}$.

Under the uniform prior $\omega_{it,\tau} \propto$ Const., the conditional posterior density of $\omega_{it,\tau}$ is

$$\pi(\omega_{it,\tau} | Y, X, B_\tau, F_\tau, \Lambda_\tau, \Omega_{-\omega_{it,\tau},\tau}) \propto \exp\left\{-\frac{a_{it,\tau}^2}{2\omega_{it,\tau}} - \frac{\omega_{it,\tau}}{2}\right\},$$

which is the generalized inverse-Gaussian distribution with parameter $(1, a_{it,\tau}, 1)$. Thus, we can easily draw a posterior sample of $\omega_{it,\tau}$ using the Gibbs sampler.

Note that this parameter $\omega_{it,\tau}$ is not included in the model (1). However, by introducing $\omega_{it,\tau}$ in our MCMC posterior sampling, we obtained the conditional posterior of B_τ and Λ_τ analytically. This allows us to use Gibbs sampling, which improves the efficiency of the MCMC algorithm.

3.2.4 Posterior sampling scheme

Due to the non-smoothness of the objective function, it was difficult to obtain the conditional posterior of B_τ and Λ_τ analytically without the data augmentation approach. Thanks to the data augmentation approach, we can analytically obtain the conditional posterior distributions of B_τ, Λ_τ and Ω_τ. Therefore, we can easily draw the posterior samples by implementing the Gibbs sampling algorithm. To draw F_τ, we can use the Metropolis-Hastings algorithm. We now summarize our data-augmentation strategy (given values of τ and the number of common factors r_τ) as follows.

Posterior sampling algorithm:

Step 1. Initialize the parameters.

Step 2. Sample F_τ from $\pi(F_\tau|Y, X, B_\tau, \Lambda_\tau, \Omega_\tau)$.

Step 3. Sample $\boldsymbol{\gamma}_{i,\tau}$ from $\pi(\boldsymbol{\gamma}_{i,\tau}|Y, X, F_\tau, B_{-i,\tau}, \Lambda_{-i,\tau}, \Omega_\tau)$ for $i = 1, ..., N$.

Step 4. Sample $\omega_{it,\tau}$ from $\pi(\omega_{it,\tau}|Y, X, F_\tau, B_\tau, \Lambda_\tau, \Omega_{-\omega_{it,\tau},\tau})$ for $i = 1, ..., N$ and $t = 1, ..., T$.

Step 5. Repeat Step 2 to Step 4 for a sufficiently large number of iterations.

To check MCMC convergence, several approaches were previously proposed (see, e.g., Robert and Casella (2004, Chapter 12)). In this paper, we follow Gerlach et al. (2011) by examining trace plots from the MCMC sampler (For more details, see the simulation study in the online supplementary documents). Because the number of parameters is large, a good starting point is helpful for implementing MCMC. Otherwise, the number of MCMC iterations, H, may need to be large so that the MCMC chain gets close to the samples from the posterior distribution. To avoid this computational burden, we use the frequentist estimator in Section 3.1. This allows us to start MCMC from the Bayesian maximum a posteriori.

The outcomes of the above algorithm can be regarded as a random sample from the joint posterior density function after a burn-in period (see, e.g., Ando (2010)). We then obtain a set of H posterior samples $\{B_\tau^{(k)}, F_\tau^{(k)}, \Lambda_\tau^{(k)}; k = 1, ..., H\}$, which can be employed for conducting Bayesian analysis numerically. For example, the Bayesian maximum a posteriori is approximately given as

$$\{\tilde{B}_\tau, \tilde{F}_\tau, \tilde{\Lambda}_\tau\} = \text{argmax}_{\{B_\tau^{(k)}, F_\tau^{(k)}, \Lambda_\tau^{(k)}\}; k=1,...,H} L(Y|X, B_\tau^{(k)}, F_\tau^{(k)}, \Lambda_\tau^{(k)}) \pi(B_\tau^{(k)}, F_\tau^{(k)}, \Lambda_\tau^{(k)}), \quad (8)$$

which is an approximated solution as the maximizer of $L(Y|X, B_\tau, F_\tau, \Lambda_\tau)\pi(B_\tau, F_\tau, \Lambda_\tau)$. In practice, the number of MCMC iterations is finite, and thus the Monte Carlo error (the difference between the exact maximizer and its approximated solution in (8)) exists. However, this Monte Carlo error will converge to zero as $H \to \infty$. Regarding the Monte Carlo error of MCMC, we refer to Doss et al. (2014), Jones (2004) and references therein.

3.3 Relationship between the frequentist estimator and Bayesian approach

In Section 3.2, the identification restrictions on F_τ and Λ_τ are accommodated in the prior distribution of F_τ and Λ_τ. Recall, also, that the diffuse prior was used for $\boldsymbol{\gamma}_\tau$. In this case, the Bayesian maximum a posteriori in (8) coincides with the frequentist estimator given in Section 3.1. The above priors were used for illustrating the core idea of our data-augmentation strategy. Even under a different prior specification (e.g., the lasso prior on $\boldsymbol{b}_{i,\tau}$), one can obtain the Bayesian maximum a posteriori in (8) by modifying our proposed posterior sampling algorithm. In this case, the Bayesian maximum posteriori in (8) no longer coincides with the frequentist estimator.

As an advantage of the Bayesian MCMC procedure, one can construct a Bayesian credible interval for the parameters even when N or T (or both) are small, while the asymptotic theory of the proposed frequentist estimator (See Section 5) relies on the large N and T. An online supplementary document compares the Bayesian estimators (the posterior mean, mode, median) and frequentist estimators under a small panel. While Bayesian estimators provided smaller estimation errors under a small panel, Bayesian estimators and the frequentist estimator are asymptotically equivalent. This is because the prior information is dominated by the pseudo likelihood $L(Y|X, F_\tau, \Lambda_\tau, B_\tau)$. The performance of the Bayesian estimators and the frequentist estimator became very similar as the panel size increased. Details are given in Section G.3 in the online supplementary document.

4 Model selection

In practice, we have to determine the dimension of the interactive effects, r_τ. Although several methods have been proposed to select the number of factors (e.g., Bai and Ng (2002), Amengual and Watson (2007), Hallin and Liška (2007), and Lam and Yao (2012)), these methods are fundamentally constructed for mean factor models, not quantile regression models.

One might think that cross-validation can be used for this purpose. However, as described in Ando and Bai (2018), it is not easy to apply cross-validation because of the existence of the factor structure. The reason is as follows. Consider leave-one-individual-out cross-validation. Based on the training sample, we can estimate the regression coefficients and the factor structures by using the algorithm in Section 3. However, it is not possible to obtain the factor loadings of the validation sample (deleted units) because the factor loadings are unit dependent. Instead, one may consider estimating the model based on the information observed up to time $t-1$ and then forecast the responses of each unit at time t. However, the factor structure at time t is not available to make a forecast. Thus, the pure cross-validation procedure is not easy to apply directly.

In this paper, we propose a novel information criterion. The number of common factors is selected by minimizing the following information criterion

$$IC_\tau(r) = \log\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_{i,\tau}(r) - \hat{\boldsymbol{f}}_{t,\tau}(r)'\hat{\boldsymbol{\lambda}}_{i,\tau}(r)\right)\right] + r \times q(N,T) \tag{9}$$

where $\hat{\boldsymbol{b}}_{i,\tau}(r), \hat{\boldsymbol{\lambda}}_{i,\tau}(r)$ and $\hat{\boldsymbol{f}}_{t,\tau}(r)$ is the estimated model parameters given the number of common factors r. The function $q(N,T)$ is a penalty on the dimension of interactive effects.

In numerical study, we specify the function as

$$q(N,T) = \log\left(\frac{NT}{N+T}\right)\left(\frac{N+T}{NT}\right). \tag{10}$$

The asymptotic performance of $IC_\tau(r)$ in (9) is investigated in the next section.

5 Asymptotic theory

Because our modeling procedure is new, it is ideal to investigate its theoretical properties. In this section, we first provide the consistency of the proposed estimator $\hat{F}_\tau$, $\hat{B}_\tau$ and $\hat{\Lambda}_\tau$. Here, the true parameter value $\{F_{0,\tau}, B_{0,\tau}, \Lambda_{0,\tau}\}$ is defined as the minimizer of the expected quantile loss function

$$\ell_{0,\tau}(Y|X, B_\tau, F_\tau, \Lambda_\tau) = E\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,\tau} - \boldsymbol{f}_{t,\tau}'\boldsymbol{\lambda}_{i,\tau})\right] \tag{11}$$

subject to the identification restriction on F_τ and Λ_τ. Here, the expectation is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$ conditional on X, $F_{0,\tau}$ and $\Lambda_{0,\tau}$. The following proposition provides the average consistency of $\hat{\boldsymbol{\gamma}}_{i,\tau} = (\hat{\boldsymbol{b}}_{i,\tau}', \hat{\boldsymbol{\lambda}}_{i,\tau}')'$ and $\hat{\boldsymbol{f}}_{t,\tau}$. In general, $\hat{\boldsymbol{f}}_{t,\tau}$ and $\hat{\boldsymbol{\lambda}}_{i,\tau}$ are estimating a rotation of the true factors and factor loadings, unless the latter satisfies the identification restrictions stated in the beginning of Section 3.1 (See also Bai and Ng (2013)). For simplicity of notation, we drop the rotation matrix.

Proposition 1 *Under Assumptions A–D, the following claims hold:*

$$N^{-1}\sum_{i=1}^{N}\|\hat{\boldsymbol{\gamma}}_{i,\tau} - \boldsymbol{\gamma}_{i,0,\tau}\|^2 = o_p(1),$$

$$T^{-1}\|\hat{F}_\tau - F_{0,\tau}\|^2 = T^{-1}\sum_{t=1}^{T}\|\hat{\boldsymbol{f}}_{t,\tau} - \boldsymbol{f}_{t,0,\tau}\|^2 = o_p(1),$$

where $\boldsymbol{\gamma}_{i,0,\tau} = (\boldsymbol{b}_{i,0,\tau}', \boldsymbol{\lambda}_{i,0,\tau}')'$.

The proof of Proposition 1 is given in an online supplementary document. The result $\|\hat{F}_\tau - F_{0,\tau}\|/T^{1/2} = o_p(1)$ implies that the space spanned by $F_{0,\tau}$ and the space spanned by the estimated factors $\hat{F}_\tau$ are asymptotically the same. We also prove the consistency of the estimators in the sense that $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{\boldsymbol{\lambda}}_{i,\tau}$ converge in probability to $\boldsymbol{b}_{i,0,\tau}$ $\boldsymbol{\lambda}_{i,0,\tau}$ and uniformly over $1 \le i \le N$. In addition, $\hat{\boldsymbol{f}}_{t,\tau}$ converges in probability to $\boldsymbol{f}_{t,0,\tau}$ uniformly over $1 \le t \le T$.

Theorem 1 *Suppose Assumptions A–D,* $\log(T)/N^{1/2} \to 0$ *and* $\log(N)/T^{1/2} \to 0$ *hold. Then,* $\hat{\boldsymbol{b}}_{i,\tau}$ *and* $\hat{\boldsymbol{\lambda}}_{i,\tau}$ *are consistent*

$$\max_{1 \le i \le N} \|\hat{\boldsymbol{b}}_{i,\tau} - \boldsymbol{b}_{i,0,\tau}\| = o_p(1), \tag{12}$$

$$\max_{1 \le i \le N} \|\hat{\boldsymbol{\lambda}}_{i,\tau} - \boldsymbol{\lambda}_{i,0,\tau}\| = o_p(1). \tag{13}$$

Moreover, the estimated common factor is consistent

$$\max_{1 \le t \le T} \|\hat{\boldsymbol{f}}_{t,\tau} - \boldsymbol{f}_{t,0,\tau}\| = o_p(1). \tag{14}$$

Theorem 2 shows that the asymptotic distribution of the estimated regression coefficients, $\hat{\boldsymbol{\gamma}}_{i,\tau} = (\hat{\boldsymbol{b}}'_{i,\tau}, \hat{\boldsymbol{\lambda}}'_{i,\tau})'$, is a multivariate normal distribution. Similarly, the asymptotic distribution of the estimated common factor $\hat{\boldsymbol{f}}_{t,\tau}$ is also a multivariate normal distribution.

Theorem 2 *Suppose that Assumptions A–D hold. Assume that* $T^{1/2}/N^{1-\gamma} \to 0$ *and* $N^{1/2}/T^{1-\gamma} \to 0$ *for a small* γ *satisfying* $1/16 < \gamma$. *Then, the asymptotic distribution of* $T^{1/2}(\hat{\boldsymbol{\gamma}}_{i,\tau} - \boldsymbol{\gamma}_{i,0,\tau})$ *is normal with mean zero and variance-covariance matrix*

$$\Sigma_{i,\tau} = \tau(1-\tau)\Gamma^{-1}_{i,0,\tau} V_{i,0,\tau} \Gamma^{-1}_{i,0,\tau}.$$

Here, $\Gamma_{i,0,\tau}$ *and* $V_{i,0,\tau}$ *are given as* $\Gamma_{i,0,\tau} \equiv \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} g_{it}(0|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}) \boldsymbol{z}_{it,0,\tau} \boldsymbol{z}'_{it,0,\tau}$ *and* $V_{i,0,\tau} \equiv \text{plim}_{T\to\infty} T^{-1} \sum_{t=1}^{T} \boldsymbol{z}_{it,0,\tau} \boldsymbol{z}'_{it,0,\tau}$ *with* $\boldsymbol{z}_{it,0,\tau} = (\boldsymbol{x}'_{it,\tau}, \boldsymbol{f}'_{t,0,\tau})'$. *Furthermore, the asymptotic distribution of* $N^{1/2}(\hat{\boldsymbol{f}}_{t,\tau} - \boldsymbol{f}_{t,0,\tau})$ *is normal with mean zero and variance-covariance matrix*

$$\Theta_{t,\tau} = \tau(1-\tau)\Psi^{-1}_{t,0,\tau} R_{0,\tau} \Psi^{-1}_{t,0,\tau}.$$

Here, $\Psi_{t,0,\tau}$ *and* $R_{0,\tau}$ *are given as* $\Psi_{t,0,\tau} \equiv \text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} g_{it}(0|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0,\tau}, \boldsymbol{\lambda}_{i,0,\tau}) \boldsymbol{\lambda}_{i,0,\tau} \boldsymbol{\lambda}'_{i,0,\tau}$ *and* $R_{0,\tau} \equiv \text{plim}_{N\to\infty} N^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_{i,0,\tau} \boldsymbol{\lambda}'_{i,0,\tau}$, *respectively.*

The proofs of Theorem 1 and Theorem 2 are given in the supplementary document. There are some studies on panel data models with factor structures, e.g., Bai (2009), Song (2013), Ando and Bai (2015). However, these results cannot be transfered to our setting directly because these methods were designed for panel "mean" regression models with factor structures instead of panel "quantile" models. Although Koenker (2004) and Kato et al.

(2012) investigated the asymptotic property of the panel quantile regression models, their results are derived from panel quantile regression models with "individual fixed effects". In contrast to these studies, the model (1) contains the factor structures and heterogeneous regression coefficients. In addition, the dimensions of the panel size N and T are diverging. Therefore, a novel proof is developed by addressing these technical difficulties. For the panel "mean" models, Song (2013) and Ando and Bai (2015) imposed $T^{1/2}/N \to 0$ and $N^{1/2}/T \to 0$ to obtain the asymptotic distribution. Because of the non-smoothness of the objective function and nonlinearity in terms of parameters, we need slightly stronger conditions on T and N.

Next, we provide a theoretical justification for the use of IC in (9), as none of the previous studies (e.g., Bai and Ng (2002), Amengual and Watson (2007), Hallin and Liška (2007), and Lam and Yao (2012)) have addressed the important question of how to determine the number of common factors in a panel quantile regression model with interactive fixed effects. Here, we provide a new solution to this issue and provide a theoretical justification for our proposed model selection criterion.

Theorem 3 *Suppose that assumptions in Theorem 2 hold. Under the model selection criterion (9) with penalty $q(N,T)$ that satisfies*

$$q(N,T) \to 0 \quad \text{and} \quad C_{NT}^{-1} \times q(N,T) \to \infty,$$

where $C_{NT} = \min\{N,T\}$, we have a consistent model selector of the true dimension of the interactive effects (i.e., the true number of common factors) $r_{0,\tau}$.

As shown in Bai and Ng (2002), the penalty function (10) satisfies the conditions in Theorem 3. One can also consider an alternative penalty function. However, this is outside the scope of this paper.

6 Empirical results

6.1 Data and model

We explore the quantile common factor structures of the global financial markets around the period of the subprime crisis, the period of the European sovereign debt crisis, and the subsequent period. Here, we analyze the stock returns of publicly traded firms and firms traded in over-the-counter trading markets for over 6,000 international stocks from over 100 financial markets. To investigate the impact of the subprime crisis and European sovereign debt crisis on the global financial industry, we analyze individual firms' stock returns belonging to the following industries: Banking, Life Insurance, Nonlife Insurance, Financial Services, and Real Estate Investment and Services. We collect all data from the Datastream database.

To study the dynamic characteristics of the global stock market, we analyze the following 3 periods.

Period 1: January 1, 2007, to April 31, 2009
Period 2: September 1, 2009, to December 31, 2012
Period 3: January 1, 2013, to March 31, 2015

Period 1 contains some key events during the subprime crisis, including the Chapter 11 bankruptcy of Lehman Brothers in September 2008. The Dow Jones Industrial Average then hit a bottom in the middle of 2009. Although the Dow Jones Industrial Average had been recovering stably since then, the long-term interest rates of Euro zone countries (including Greece, Portugal, and Ireland) started to increase by the end of 2009. In the middle of 2012, the long-term interest rates of the Greece government bond reached above 30%. After the announcement by the European Central Bank indicating free unlimited support for all Euro zone countries, the interest rate dropped by around 10% in December, 2012. Obviously, one could use a different specification for these sub-periods. However, similar results are obtained under a different sub-period specification. Stocks with missing returns and stocks with no variation are excluded from the sample used for analysis. The final samples for each period are summarized in Table 1 in an online supplementary document. Finally, because different financial markets do not have the same trading hours, it is common to use the rolling average, and the two-day returns of each of the firms are therefore employed for the returns (e.g., Forbes and Rigobon (2002), Ando and Bai (2017)). We consider the following panel quantile regression model with a factor structure:

$$\begin{aligned} Q_{y_{it}}\left(\tau|\boldsymbol{x}_{it}\right) &= \alpha_{i,\tau} + Mkt_t \times \beta_{Mkt,i,\tau} + HML_t \times \beta_{HML,i,\tau} + SMB_t \times \beta_{SMB,i,\tau} \\ &\quad + RMW_t \times \beta_{RMW,i,\tau} + CMA_t \times \beta_{CMA,i,\tau} + \sum_{k=1}^{11} LMkt_{t,k} \times \beta_{LMkt,k,\tau} + \boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}, \quad (15) \end{aligned}$$

where Mkt_t, HML_t, SMB_t, RMW_t and CMA_t are Fama/French global five factors at time t. Here, Mkt is the return on a region's value-weighted market portfolio minus the U.S. one-month T-bill rate, SMB (Small Minus Big) is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios, HML (High Minus Low) is the average return on the two value portfolios minus the average return on the two growth portfolios, RMW (Robust Minus Weak) is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios, and CMA (Conservative Minus Aggressive) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios. International tests of a five-factor asset pricing model are studied by Fama and French (2016). Further details of these factors and the historical data are obtained from the publicly available Fama/French data library[3]. A set of factors $\{LMk_{t,k},\ k = 1, ..., 11\}$ is the average return for 11 local stock exchange markets with more than 100 stocks in the dataset.

6.2 Result

6.2.1 Number of common factors

In this paper, we focus on both the upper and lower quantiles and consider $\tau = 0.05$ and $\tau = 0.95$. We apply the proposed model selection criterion, $IC(r_\tau)$ in (9), to select the number of common factors. The maximum number of common factors is set to twelve. For each period, the number of common factors is determined as the minimizer of the IC score. After we obtain the frequentist estimates, we also implemented the posterior sampling procedure described in Section 3.2. The total number of posterior samples is set at 3,000. As

[3]http : //mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

expected, the difference between the frequentist estimate and Bayesian maximum a posteriori in (8) was negligible. Hereafter, we thus report the results based on the frequentist estimator only.

The number of common factors detected is summarized in Table 4. In this table, the number of common factors for the $\tau = 0.05$ quantile in Period 1 is determined to be $r_\tau = 7$ because it achieved the smallest value of the proposed model-selection criterion, IC. This suggests that there are $r_\tau = 7$ common factors that explain the $\tau = 0.05$ quantile in Period 1. The table shows that the number of common factors in the $\tau = 0.95$ quantile is smaller than that in the $\tau = 0.05$ quantile in Period 1 and Period 2. This indicates that the $\tau = 0.05$ quantile exhibits greater variability due to the increase in the degree of complexity. As one of our referees suggested, this difference is partially due to differences in governments or regulations for the different stocks.

The number of common factors for the $\tau = 0.05$ quantile in Period 3 is smaller than that in other periods. This implies that the degree of market heterogeneity decreased in Period 3. In summary, the common factor structures that explain the asset return distribution vary across quantiles. Moreover, the common factor structures are not symmetric in the sense that the structures in the lower tails and the upper tails are different.

6.2.2 Common factors for the quantile and for the mean

To check whether the extracted common factors for the quantile function and the common factors for the mean are perfectly related, we implement canonical correlation analysis. Let $\hat{F}_\tau$ be the estimated common factors of dimension r_τ, which is determined by the $IC(r_\tau)$ score. Setting the dimension of the common factor for the mean to be identical to r_τ, we estimate the following asset pricing model with a common factor structure

$$\begin{aligned} y_{it} &= \alpha_i + Mkt_t \times \beta_{Mkt,i} + HML_t \times \beta_{HML,i} + SMB_t \times \beta_{SMB,i} + RMW_t \times \beta_{RMW,i} \\ &\quad + CMA_t \times \beta_{CMA,i} + \sum_{k=1}^{11} LMkt_{t,k} \times \beta_{LMkt,k,i} + \boldsymbol{f}_t' \boldsymbol{\lambda}_i + \varepsilon_{it}, \end{aligned}$$

by minimizing the least-squares objective function $\ell(B, F, \Lambda) = \sum_{i=1}^N \|\boldsymbol{y}_i - X_i \boldsymbol{\beta}_i - F\boldsymbol{\lambda}_i\|^2$ subject to the constraint $F'F/T = I_r$. Numerical optimization can be achieved by the iterative optimization of B, F, Λ based on the previous results (see, e.g., Bai (2009), Song (2013), Ando and Bai (2015), Wang (2017)).

Let $\hat{F}$ be the estimated common factors for the mean structure. Then, we apply the canonical correlation analysis for exploring the relationships between $\hat{F}$ and $\hat{F}_\tau$. To check whether all the columns of $\hat{F}$ and $\hat{F}_\tau$ are indeed related, we use the significance tests for canonical correlation analysis. Wilks' Lambda is used for this purpose. In determining the significant canonical correlation, the 5% significance level is used. The results are summarized in Table 5.

There is a certain degree of relatedness between the common factor for the quantile and that for the mean. Note that $1/2 = 50\%$ of $\tau = 95\%$ quantile common factors $\hat{F}_{0.95}$ in Period 1 are related to the estimated common factors for the mean structure $\hat{F}$. Thus, it should be noted that the common factor for the quantile and that for the mean are not always identical, as the statistically significant canonical correlation is smaller than the detected number of common factors for the quantiles. This implies that the two-step procedure to estimate the model parameters (in Section 3) will lead to inconsistent estimates of the regression coefficients and factor loadings. Therefore, our data-augmentation strategy is important for avoiding this issue.

6.2.3 Do the stock exchange and industry matter?

To explore the effects of stock exchanges and industries on individual stock returns, we apply a clustering approach to the estimated regression coefficients and factor loadings $\{(\hat{\boldsymbol{b}}_{i,\tau}', \hat{\boldsymbol{\lambda}}_{i,\tau}');\ i = 1, ..., N\}$ to create a set of groups based on the similarities in the sensitivity to the common factors. If the source of the sensitivity to the factors (both observables and unobservables) is solely attributable to stock exchanges, it is expected that the two-way table of the assigned group membership from the clustering approach against the stock exchanges will be diagonal. To save space, clustering results are provided in the online supplementary document. In short, the firm industry and the stock exchange on which a firm is listed are important factors to be considered. However, we also note that these nominal factors are insufficient to fully capture the underlying market structures.

6.2.4 Meaning of common factors and regional effects

Because the estimated common factors do not have an immediate economic interpretation, we explore the economic meanings of the estimated common factors. Here, we regress the estimated common factors on Fama and French's 5 factors (Global, North America, Europe and Asia Pacific). These factors and the historical data are obtained from the publicly available Fama/French data library. Because there are 5 factors (Mkt, SMB, HML, RMW and CMA) in each of the four regions, each of the estimated common factors is regressed on the set of 20 (Fama and French's 5 factors $\times$ 4 regions) variables. Mathematically, let $f_{jt,\tau}$ be the estimated value of the j-th common factor at time t and at the τ-th quantile; we then run the following regression $f_{jt,\tau} = \boldsymbol{z}_t'\boldsymbol{\gamma}_j + e_{jt}$, where $\boldsymbol{z}_t$ is the 20-dimensional vector of Fama and French's factors. Then, we conduct statistical significance tests of the least squares estimate $\hat{\boldsymbol{\gamma}}_j$.

To clearly demonstrate the regional effects, we calculate the following. We simply count the number of statistically significant Fama and French factors for each region. For example, in the lower quantile $\tau = 0.05$, there are $r_\tau = 7$ common factors in Period 1. For each of the factors, we run the following regression: $f_{jt,\tau} = \boldsymbol{z}_t'\boldsymbol{\gamma}_j + e_{jt}$ for $j = 1, 2, ..., r_\tau$. To investigate a connection to North America, we count the total sum of the number of significant Fama and French North America 5 factors across the seven ($r_\tau = 7$) regressions. Note that a particular Fama and French factor may be counted multiple times across the regressions. Because the number of common factors r_τ varies across quantiles and periods, we convert this count into percentage terms by dividing it by $5 \times r_\tau$, which is the upper bound of the count. For example, in the lower quantile $\tau = 0.05$ in Period 1, the total sum of the number of significant Fama and French North America 5 factors is divided by $5 \times r_\tau = 5 \times 7 = 35$. The same operation is performed for the others, Global, Europe and Asia Pacific.

Figure 3 summarizes the results. We see that the extracted common factors are less connected to Fama and French's global factors than regional factors. This implies that the extracted factors are more related to Fama and French's regional factors. We note that the Fama/French global five factors are included in the explanatory variables (15). Because the extracted common factors are still connected to the Fama and French's global factors, our model is useful in treating the endogeneity problem.

6.2.5 Discussion

The goal of our empirical analysis is to analyze the quantile co-movement of a large number of financial time series by investigating the quantile co-movement structure of the global financial market. We are interested primarily in the empirical questions described in the introduction. Here, we would like to provide a summary of our empirical findings.

Regarding the first question; "Do the quantile common factor structures that explain asset return distribution vary across quantiles?", we found that the number of common factors varies across quantiles. The number of common factors also varies over time. This is one of the reasons that understanding current financial market structures is important because the market structure changes over time.

We found empirical evidence to answer the second question. The common factor structure is not symmetric in the sense that the number of common factors in the lower tail τ is larger than that in the upper tail in Period 1 and Period 2. This implies that there is greater heterogeneity in the lower tail than in the upper tail in these two periods.

Third, Table 5 indicates that there is a fair degree of relatedness between the common factor for the quantile and that for the mean. However, they are not identical. Therefore, the two-step procedure described in Section 3 would lead to inconsistent parameter estimation results. To avoid such undesirable results, our initialization algorithm and data-augmentation strategy are important.

Fourth, the stock exchange on which a firm is listed is partially related to the extracted factor structures. These observable stock characteristics are not sufficient to explain the extracted factor structures. This implies that diversification based on the stock exchange on which a firm is listed is inadequate, as the common factor structures are not fully connected with these nominal classifications.

Finally, there are special characteristics of the quantiles of financial markets. Compared to Period 3, the number of common factors in Period 1 and Period 2 is larger. This implies that the heterogeneity of the global financial market decreased after Period 2. Figure 3 shows that the unobservable common factor structures are more connected to Asia Pacific regional factors during the subprime crisis (Period 1). This implies that the model (15) is missing some observable factors relating to Asia Pacific regional factors. In summary, the important common factors that govern asset return distributions vary across quantiles. These findings, derived from our general procedure, offer useful insights for institutional investors and regulators.

6.2.6 Robustness check

For the vectors of observable factors in (15), we use Fama and French (2016)'s global five factors and a set of average returns for 11 local stock exchange markets with more than 100 stocks in the dataset. It is also possible to implement the proposed modeling procedure under a different specification of the vectors of observable factors. For example, together with Fama and French (2016)'s global five factors, one can consider a set of average returns for local stock exchange markets with more than 50 stocks in the dataset. However, similar results are obtained under this specification.

The rolling two-day average of returns is used to cope with the differences in international market trading hours. Similar results can be obtained under daily returns instead of a two-day rolling average.

7 Conclusion

In this paper, we introduced a new panel quantile regression model with interactive fixed effects. The model has attractive features, including heterogeneous regression coefficients, unobservable common factors that vary across quantiles, and the ability to cope with endogeneity by allowing correlations between observable factors and unobservable factors and factor loadings.

To address endogeneity and a large number of parameters, we proposed frequentist and Bayesian data-augmented inference procedures. This allowed us to directly estimate the model parameters. Theoretical properties were established for the frequentist estimator. We also developed a new approach for selecting the number of common factors. Our empirical analysis delivered many insightful findings, which are of interest for investors and market regulators.

Acknowledgments The authors would like to thank the Co-editor, the Associate Editor and three anonymous reviewers for their constructive and helpful comments, which improved the quality of the paper considerably. We would also like to thank Lina Lu for helpful comments and the seminar participants at the University of Melbourne and the participants at the 1st International Conference on Econometrics and Statistics (EcoSta 2017) and the 30th Australasian Finance and Banking Conference. Ando's research is supported by a research grant from University of Melbourne, Melbourne Business School. Bai's research is supported by National Science Foundation, SES1658770.

References

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

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

Alhamzawi, R. and Yu, K. (2013). "Conjugate priors and variable selection for Bayesian quantile regression," *Computational Statistics & Data Analysis*, **64**, 209–219.

Ando, T. (2010). *Bayesian model selection and statistical modeling.* CRC Press, London.

Ando, T. and Bai, J. (2015). "Asset pricing with a general multifactor structure," *Journal of Financial Econometrics*, **13**, 556–604.

Ando, T. and Bai, J. (2016). "Panel data models with grouped factor structures under unknown group membership," *Journal of Applied Econometrics*, **136**, 163–191.

Ando, T. and Bai, J. (2017). "Clustering huge number of time series: A panel data approach with high-dimensional predictors and factor structures," *Journal of the American Statistical Association*, **112**, 1182–1198.

Ando, T. and Bai, J. (2018). "Selecting the regularization parameters in high-dimensional panel data models: consistency and efficiency," *Econometric Reviews*, **37**, 183–211.

Ando, T. and Tsay, R. S. (2011). "Quantile regression models with factor-augmented predictors and information criterion," *Econometrics Journal*, **14**, 1–24.

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

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

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

Bai, J. and Ng, S. (2013). "Principal components estimation and identification of static factors," *Journal of Econometrics*, **176**, 18–29.

Baur, D.G. (2013). "The structure and degree of dependence: a quantile regression approach," *Journal of Banking and Finance*, **37**, 786–798.

Baur, D.G., Dimpfl, T. and Jung, R.C. (2012). "Stock return autocorrelations revisited: a quantile regression approach," *Journal of Empirical Finance*, **19**, 254–265

Belloni, A., and Chernozhukov, V. (2011). "ℓ_1-penalized quantile regression in high-dimensional sparse models," *Annals of Statistics*, **39**, 82–130.

Bondell, H.D., Reich, B.J. and Wang, H. (2010). "Noncrossing quantile regression curve estimation," *Biometrika*, **97**, 825–838.

Cappiello, L., Gerard, L., Kadareja, A. and Manganelli, S. (2014). "Measuring Comovements by Regression Quantiles," *Journal of Financial Econometrics*, **12**, 645–678.

Chen, J. (2015). "Factor Instrumental Variable Quantile Regression," *Studies in Nonlinear Dynamics and Econometrics*, **19**, 71–92.

Chen, L., Dolado, J. and Gonzalo, J. (2017). "Quantile Factor Models," Working Paper.

Chen, M. Fernández-Val, I. and Weidner, M. (2014). Nonlinear Panel Models with Interactive Effects. Working Paper.

Chen, C.W.S. and Gerlach, R. (2013). "Semi-parametric quantile estimation for double threshold autoregressive models with heteroskedasticity," *Computational Statistics*, **28**, 1103–1131.

Chen, C.W.S., Li, M. and Nguyen, N.T.H., and Sriboonchitta, S. (2017). "On asymmetric market model with heteroskedasticity and quantile regression," *Computational Economics*, **49**, 155–174.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). "Quantile and Probability Curves Without Crossing," *Econometrica*, **78**, 1093–1125.

Chuang, C.C., Kuan, C.M. and Lin, H.Y. (2009). "Causality in quantiles and dynamic stock return-volume relations," *Journal of Banking and Finance*, **33**, 1351–1360.

Connor, G., and Korajzcyk, R. (1986). "Performance measurement with the arbitrage pricing theory: a new framework for analysis," *Journal of Financial Economics*, **15**, 373–394.

Dette, H. and Volgushev, S. (2008). "Non-crossing nonparametric estimates of quantile curves," *Journal of the Royal Statistical Society*, **B70**, 609–627.

Doss, C.R., Flegal, J.M., Jones, G.L. and Neath, R.C. "Markov chain Monte Carlo estimation of quantiles," *Electronic Journal of Statistics*, **8**, 2448–2478.

Engle, R.F., Manganelli, S. (2004). "CAViaR: conditional autoregressive value at risk by regression quantiles," *Journal of Business and Economic Statistics*, **22**, 367–381.

Fama, E. F., and French, K. R. (1993). "Common factors in the returns on stocks and bonds," *Journal of Financial Economics*, **33**, 3–56.

Fama, E. F., and French, K. R. (1998). "Value versus Growth: The International Evidence," *Journal of Finance*, **53**, 1975–1999.

Fama, E. F., and French, K. R. (2016). "International tests of a five-factor asset pricing model," *Journal of Financial Economics*, **123**, 441–463.

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

Forbes, K. J., and Rigobon, R. (2002). "No contagion, only interdependence: Measuring stock market comovements," *Journal of Finance*, **57**, 2223–2261.

Freyberger, J. (2015). "Asymptotic theory for differentiated products demand models with many markets," *Journal of Econometrics*, **185**, 162–181.

Geraci, M. and Bottai, M. (2007). "Quantile regression for longitudinal data using the asymmetric Laplace distribution," *Biostatistics*, **8**, 140–154.

Gerlach R, Chen, C.W.S. and Chan, N.Y.C. (2011). "Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets," *Journal of Business and Economic Statistics*, **29**, 481–492.

Gerlach, R., Chen, C.W.S. and Lin, E.M.H. (2016). "Bayesian Assessment of Dynamic Quantile Forecasts," *Journal of Forecasting*, **35**, 751–764.

Griffin, J. M. (2002). "Are the Fama and French Factors Global or Country Specific," *Review of Financial Studies*, **15**, 783–803.

Hallin, M., and R. Liška (2007). "The generalized dynamic factor model: determining the number of factors," *Journal of the American Statistical Association*, **102**, 603–617.

Han, H., Linton, O., Oka, T., Whang, Y.J. (2016). "The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series," *Journal of Econometrics*, **193**, 251–270.

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

He, X. (1997). "Quantile Curves Without Crossing," *American Statistician*, **51**, 186–192.

He, X., Wang, L. and Hong, H. G. (2013). "Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data," *Annals of Statistics*, **41**, 342–369.

Hoff. P.D. (2009). "Simulation of the matrix Bingham-von Mises-Fisher distribution, with applications to multivariate and relational data," *Journal of Computational and Graphical Statistics*, **18**, 438-456.

Hou, K., Karolyi, G.A., and Kho, B.-C. (2011). "What Factors Drive Global Stock Returns," *Review of Financial Studies*, **24**, 2527–2574.

Jones, G. L. (2004). "On the Markov chain central limit theorem," *Probability Surveys*, **1**, 299–320.

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

Khatri, C. G. and Mardia, K. V. (1977). "The von Mises.Fisher distribution in orientation statistics," *Journal of the Royal Statistical Society*, **B39**, 95–106.

Kozumi, H. and Kobayashi, G. (2011). "Gibbs sampling methods for Bayesian quantile regression," *Journal of Statistical Computation and Simulation*, **81**, 1565–1578.

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

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

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

Lancaster, T. and Jun, S.J. (2010). "Bayesian quantile regression methods," *Journal of Applied Econometrics*, **25**, 287–307.

Nath, H.B. and Brooks, R.D. (2015). "Assessing the idiosyncratic risk and stock returns relation in heteroskedasticity corrected predictive models using quantile regression," *International Review of Economics & Finance*, **38**, 94–111.

Ni, Z.X., Wang, D.Z. and Xue, W.J. (2015). "Investor sentiment and its nonlinear effect on stock returns – New evidence from the Chinese stock market based on panel quantile regression model," *Economic Modelling*, **50**, 266–274.

Onatski, A. (2009). "Testing hypotheses about the number of factors in large factor models," *Econometrica*, **77**, 1447–1479.

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

Polson, N.G. and Scott, J.G. (2013). "Data augmentation for non-Gaussian regression models using variance-mean mixtures," *Biometrika*, **100**, 459–471.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer Texts in Statistics.

Ross, S.A. (1976). "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, **13**, 341–360.

Shao, J. (1997). "An asymptotic theory for linear model selection," *Statistica Sinica*, **7**, 221–264.

Sherwood, B. and Wang, L. (2016). "Partially linear additive quantile regression in ultra-high dimension," *Annals of Statistics*, **44**, 288–317.

So, M. K. P. and Chung, R. S. W. (2015). "Statistical Inference of Conditional Quantiles in Nonlinear Time Series Models", *Journal of Econometrics*, **189**, 457–472.

Stock, J. H., and Watson, M. W. (2002). "Forecasting using principal components from a large number of observable factors," *Journal of the American Statistical Association*, **97**, 1167–1179.

Tang, Y., Song, X., Wang, H.J. and Zhu, Z. (2013). "Variable selection in high-dimensional quantile varying coefficient models," *Journal of Multivariate Analysis*, **122**, 115–132.

Tsay, R. and Ando, T. (2012). "Bayesian panel data analysis for exploring the impact of recent financial crisis on the U.S stock market," *Computational Statistics Data Analysis*, **56**, 3345–3365.

Wang, P. (2017). "Large dimensional factor models with a multi-level factor structure," Working Paper, HKUST.

Yan, Y., and Kottas, A. (2015). "A new family of error distributions for Bayesian quantile regression," Working Paper.

Yuan, Y., Chen, N. and Zhou, S. (2017). "Modelling Regression Quantile Process using Monotone B-splines," *Technometrics*, 59, 338–350.

Yue, Y. R. and Rue, H. (2011). "Bayesian inference for additive mixed quantile regression models," *Computational Statistics & Data Analysis*, **55**, 84–96.

Table 1: Estimated number of common factors r_τ for the quantiles $\tau = 0.05$ and 0.95. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

τ	Period 1	Period 2	Period 3	Period 4
0.05	7	6	3	10
0.95	2	2	4	8

Table 2: The result of canonical correlation analysis between the common factors for quantile τ and the common factors for the mean (See Section 6.2.2). To determine the significant canonical correlation, the 5% significance level is used. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

τ	Period 1	Period 2	Period 3	Period 4
0.05	6/7	5/6	3/3	10/10
0.95	1/2	2/2	4/4	7/8

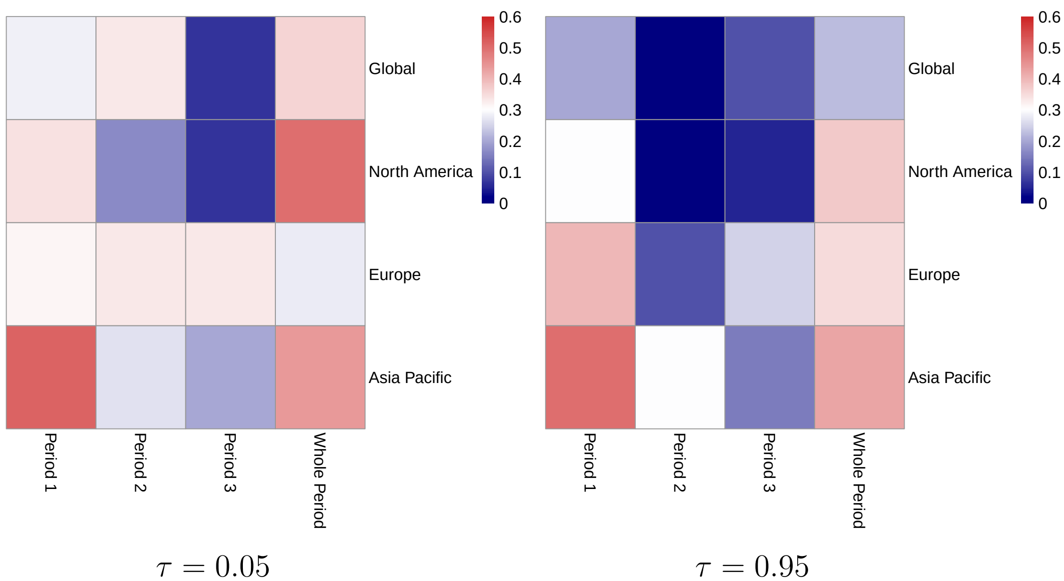


Figure 1: Link between the extracted common factors and Fama and French factors for each region (See Section 6.2.4). Each cell represents the fraction of statistical significant common factors explained by Fama and French factors. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

On-line supplement to Quantile co-movement in financial markets: A panel quantile model with unobserved heterogeneity

Tomohiro Ando [4] and Jushan Bai [5]

Abstract: Appendix A provides the proof of Proposition 1. Proof of Theorem 1 is given in Appendix B. To prove Theorem 2, we need the uniform convergence rates of estimated parameters. Appendix C provides this result. Theorem 2 and Theorem 3 are proved in Appendix D and E, respectively. In Appendix F, we provide additional information on the empirical analysis. Appendix G reports the Monte Carlo simulation results.

Notation For notational simplicity, we suppress the dependency of τ such that $\varepsilon_{it,\tau} = \varepsilon_{it}$, $\boldsymbol{b}_{i,\tau} = \boldsymbol{b}_i$, $\boldsymbol{\lambda}_{i,\tau} = \boldsymbol{\lambda}_i$, $\boldsymbol{\gamma}_{i,\tau} = (\boldsymbol{b}'_{i,\tau}, \boldsymbol{\lambda}'_{i,\tau})'$, $\boldsymbol{f}_{t,\tau} = \boldsymbol{f}_t$, $B_\tau = B$, $\Lambda_\tau = \Lambda$, $F_\tau = F$, etc throughout the proof. Also, we denote the true parameters as $\boldsymbol{b}_{i,0}$, $\boldsymbol{\lambda}_{i,0}$, $\boldsymbol{\gamma}_{i,0}$, $\boldsymbol{f}_{t,0}$, F_0, Λ_0 and B_0, etc.

A Proof of Proposition 1

The following lemma is used in the proof of Proposition 1.

Lemma 1 *(Lemma 2.2.10 of Van der Vaart and Wellner (1996)) Let X_1,...,X_n be arbitrary random variables that satisfy the tail bound:*

$$P(|X_i| > z) \leq 2\exp\left(-\frac{1}{2} \times \frac{z^2}{a+bz}\right)$$

for all z (and all i) and fixed $a, b > 0$. Then,

$$E\left|\max_{1\leq i\leq n} X_i\right| \leq C\left(b \times \log(n+1) + \sqrt{a \times \log(n+1)}\right)$$

for some positive constant C.

[4] Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.

[5] Department of Economics, Columbia University, jb3064@columbia.edu. 1019 International Affairs Building 420 West 118 Street New York, NY 10027 USA

Lemma 2 *(Lemma 2.2.11 of Van der Vaart and Wellner (1996)) Suppose that* $X_1,\ldots,X_n$ *be independent random variables with zero mean such that* $E|X_i|^m \leq m!M^{m-2}v_i/2$, *for every* $m \geq 2$ *(and all* i*) and some constants* M *and* v_i. *Then:*

$$P\left(\left|\sum_{i=1}^{n} X_i\right| \geq z\right) \leq 2\exp\left(-\frac{z^2}{v+Mz}\right),$$

for $v \geq v_1 + \cdots v_n$.

A.1 Proof of Proposition 1

We first show

$$\sup_{\boldsymbol{f}_t\in\mathcal{F},\boldsymbol{\lambda}_i\in\mathcal{L},\boldsymbol{b}_i\in\mathcal{B}}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\rho_\tau(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_i-\boldsymbol{f}_t'\boldsymbol{\lambda}_i)-E\left[\rho_\tau(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_i-\boldsymbol{f}_t'\boldsymbol{\lambda}_i)\right]\right\}\right| = o_p(1), \quad (1)$$

where $E[\cdot]$ is the expectation of y_{it} conditioned on X, F_0, Λ_0 and B_0. For any $e > 0$, $\mathcal{F}$, $\mathcal{L}$, and $\mathcal{B}$ are covered by $\cup_{\boldsymbol{f}\in\mathcal{F}}B_e(\boldsymbol{f})$, $\cup_{\boldsymbol{\lambda}\in\mathcal{L}}B_e(\boldsymbol{\lambda})$ and $\cup_{\boldsymbol{b}\in\mathcal{B}}B_e(\boldsymbol{b})$, respectively. Here $B_r(z)$ is a closed ball with center z and radius $r > 0$. Because of their compactness by Assumptions A and B, there exist finite positive integers C_F, C_Λ, C_B, and $\{\bar{\boldsymbol{f}}_1,\ldots,\bar{\boldsymbol{f}}_{C_F}\} \in \mathcal{F}$, $\{\bar{\boldsymbol{\lambda}}_1,\ldots,\bar{\boldsymbol{\lambda}}_{C_\Lambda}\} \in \mathcal{L}$, $\{\bar{\boldsymbol{b}}_1,\ldots,\bar{\boldsymbol{b}}_{C_B}\} \in \mathcal{B}$ such that

$$\mathcal{F} \subset \cup_{k=1}^{C_F}B_e(\bar{\boldsymbol{f}}_k), \quad \mathcal{L} \subset \cup_{k=1}^{C_\Lambda}B_e(\bar{\boldsymbol{\lambda}}_k), \quad \mathcal{B} \subset \cup_{k=1}^{C_B}B_e(\bar{\boldsymbol{b}}_k).$$

Define

$$\mathcal{J} \equiv \{\boldsymbol{x}'\boldsymbol{b}+\boldsymbol{f}'\boldsymbol{\lambda}: \ \boldsymbol{b}\in\mathcal{B}, \ \boldsymbol{f}\in\mathcal{F}, \ \boldsymbol{\lambda}\in\mathcal{L}\}.$$

The set $\mathcal{J}$ is covered by $\cup_{k=1}^{C_B}\cup_{u=1}^{C_F}\cup_{v=1}^{C_\Lambda}H(\bar{\boldsymbol{b}}_k,\bar{\boldsymbol{f}}_u,\bar{\boldsymbol{\lambda}}_v)$, where $H(\bar{\boldsymbol{b}}_k,\bar{\boldsymbol{f}}_u,\bar{\boldsymbol{\lambda}}_v) = \{\boldsymbol{x}'\boldsymbol{b}+\boldsymbol{f}'\boldsymbol{\lambda}; \ \boldsymbol{b}\in B_e(\bar{\boldsymbol{b}}_k), \ \boldsymbol{f}\in B_e(\bar{\boldsymbol{f}}_u), \ \boldsymbol{\lambda}\in B_e(\bar{\boldsymbol{\lambda}}_v)\}$. Suppose $\boldsymbol{x}'\boldsymbol{b}+\boldsymbol{f}'\boldsymbol{\lambda} \in H(\bar{\boldsymbol{b}}_k,\bar{\boldsymbol{f}}_u,\bar{\boldsymbol{\lambda}}_v)$. We then have

$$\begin{aligned}
& |\boldsymbol{x}'\boldsymbol{b}+\boldsymbol{f}'\boldsymbol{\lambda}-(\boldsymbol{x}'\bar{\boldsymbol{b}}_k+\bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v)| \\
\leq\ & \|\boldsymbol{x}'\boldsymbol{b}-\boldsymbol{x}'\bar{\boldsymbol{b}}_k\| + \|\boldsymbol{f}'\boldsymbol{\lambda}-\bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v\| \\
\leq\ & \|\boldsymbol{x}\|\times\|\boldsymbol{b}-\bar{\boldsymbol{b}}_k\| + \|\boldsymbol{f}-\bar{\boldsymbol{f}}_u\|\times\|\boldsymbol{\lambda}-\bar{\boldsymbol{\lambda}}_v\| + \|\bar{\boldsymbol{f}}_u\|\times\|\boldsymbol{\lambda}-\bar{\boldsymbol{\lambda}}_v\| + \|\boldsymbol{f}-\bar{\boldsymbol{f}}_u\|\times\|\bar{\boldsymbol{\lambda}}_v\| \\
\leq\ & eM + e^2 + eM + eM \\
\equiv\ & \delta,
\end{aligned}$$

where $M < \infty$ is an upper bound (in terms of norm) for the elements in $\mathcal{L}$, $\mathcal{F}$ and $\mathcal{X}$ (by compactness assumption). Note that $H(\bar{\boldsymbol{b}}_k,\bar{\boldsymbol{f}}_u,\bar{\boldsymbol{\lambda}}_v) \subset B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_s+\bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v)$. Thus,

$$\mathcal{J} \subset \cup_{k=1}^{C_B}\cup_{u=1}^{C_F}\cup_{v=1}^{C_\Lambda}B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_k+\bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v),$$

which implies that, for any $\delta > 0$, there exist finite integers C_B, C_F and C_Λ such that $\mathcal{J}$ is covered by the union of $C_B \times C_F \times C_\Lambda$ closed balls
$B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_1 + \bar{\boldsymbol{f}}_1'\bar{\boldsymbol{\lambda}}_1), ..., B_\delta(\boldsymbol{x}'\bar{\boldsymbol{b}}_{C_B} + \bar{\boldsymbol{f}}_{C_F}'\bar{\boldsymbol{\lambda}}_{C_\Lambda})$.
For each $\boldsymbol{x}_{it}$, $\boldsymbol{f}_t$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{b}_i$, we can identify $\bar{\boldsymbol{f}}_u$, $\bar{\boldsymbol{\lambda}}_v$ and $\bar{\boldsymbol{b}}_k$, such that $\boldsymbol{f}_t \in B_e(\bar{\boldsymbol{f}}_u)$, $\boldsymbol{\lambda}_i \in B_e(\bar{\boldsymbol{\lambda}}_v)$ and $\boldsymbol{b}_i \in B_e(\bar{\boldsymbol{b}}_k)$. Define $h_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i$ and

$$d_{it} := d_{it,k,u,v} = \boldsymbol{x}_{it}'\bar{\boldsymbol{b}}_k +' \bar{\boldsymbol{f}}_u'\bar{\boldsymbol{\lambda}}_v,$$

then $|h_{it} - d_{it}| \leq \delta$. Thus,

$$\begin{aligned}
& \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i) - E\left[\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)\right]\}\right| \\
= & \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it} - h_{it}) - E\left[\rho_\tau(y_{it} - h_{it})\right]\}\right| \\
\leq & \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it} - h_{it}) - \rho_\tau(y_{it} - d_{it})\}\right| \\
& + \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{E[\rho_\tau(y_{it} - h_{it})] - E[\rho_\tau(y_{it} - d_{it})]\}\right| \\
& + \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it} - d_{it}) - E[\rho_\tau(y_{it} - d_{it})]\}\right| \\
= & I_1 + I_2 + I_3.
\end{aligned}$$

Now,

$$\begin{aligned}
I_1 & \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|\rho_\tau(y_{it} - h_{it}) - \rho_\tau(y_{it} - d_{it})| \\
& \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{s_\tau(y_{it}, \delta) - E[s_\tau(y_{it}, \delta)]\} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E[s_\tau(y_{it}, \delta)],
\end{aligned}$$

where $s_\tau(y, \delta) = \sup_{a,b\in\mathcal{J},|a-b|\leq\delta}|\rho_\tau(y - a) - \rho_\tau(y - b)|$. The first term in the last line is $o_p(1)$ by a law of large numbers. By choosing a small enough δ, the second term can be made arbitrarily small. Similarly, I_2 can be made arbitrarily small by choosing a small enough δ. Finally, we consider I_3. Note that $d_{it} = d_{it,k,u,v}$, I_3 is uniformly bounded by

$$\begin{aligned}
& \sup_{k,u,v}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it} - d_{it,k,u,v}) - E\left[\rho_\tau(y_{it} - d_{it,k,u,v})\right]\}\right| \\
= & \max_{d_j\in\mathcal{S}}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{\rho_\tau(y_{it} - d_j) - E\left[\rho_\tau(y_{it} - d_j)\right]\}\right|,
\end{aligned}$$

where

$$\mathcal{S} \equiv \left\{\{d_{it,k,u,v}\}_{i=1,...,N,t=1,...,T} : \bar{\boldsymbol{f}}_u \in \{\bar{\boldsymbol{f}}_1, ..., \bar{\boldsymbol{f}}_{C_F}\}, \bar{\boldsymbol{\lambda}}_v \in \{\bar{\boldsymbol{\lambda}}_1, ..., \bar{\boldsymbol{\lambda}}_{C_\Lambda}\}, \bar{\boldsymbol{b}}_k \in \{\bar{\boldsymbol{b}}_1, ..., \bar{\boldsymbol{b}}_{C_B}\}\right\}$$

Note that there are some abuse of notations, but the idea should be clear. The maximum of number of elements of $\mathcal{S}$ is $C_B^N C_\Lambda^N C_F^T$ when k, u, v vary.

Note that

$$\begin{aligned}
& \sup_{h\in\mathcal{J}} \left|\rho_\tau(y_{it}-h) - E\left[\rho_\tau(y_{it}-h)\right]\right| \\
= \;& \sup_{h\in\mathcal{J}} \left|\rho_\tau(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i + \varepsilon_{it} - h) - E\left[\rho_\tau(\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i + \varepsilon_{it} - h)\right]\right| \\
\le \;& \sup_{h\in R} \left|\rho_\tau(\varepsilon_{it}-h) - E\left[\rho_\tau(\varepsilon_{it}-h)\right]\right| \\
\le \;& \left|\varepsilon_{it} - E\left[\varepsilon_{it}\right]\right| + E\left[\left|\varepsilon_{it} - E[\varepsilon_{it}]\right|\right],
\end{aligned}$$

where the last line can be obtained as follows. For any $X, z \in R$ with X being a random variable, we have $|\rho_\tau(z-h) - \rho_\tau(X-h)| \le \max\{\tau, 1-\tau\}|z-X| \le |z-X|$. This leads $|\rho_\tau(z-h) - E[\rho_\tau(X-h)]| \le E[|\rho_\tau(z-h) - \rho_\tau(X-h)|] \le E[|X-z|]$, which is further bounded by $|z - E[X]| + E[|X - E[X]|]$ (See Eq (13) in Bai (1998)). Thus, we have, for positive number $K(\ge 1)$,

$$\begin{aligned}
& E\left|\rho_\tau(y_{it}-d_j) - E\left[\rho_\tau(y_{it}-d_j)\right]\right|^K \\
\le \;& E\left[\Big|\left|\varepsilon_{it} - E[\varepsilon_{it}]\right| + E\left[\left|\varepsilon_{it} - E[\varepsilon_{it}]\right|\right]\Big|^K\right] \\
\le \;& 2^{K-1}\left[E\Big|\varepsilon_{it} - E[\varepsilon_{it}]\Big|^K + \Big(E\left|\varepsilon_{it} - E[\varepsilon_{it}]\right|\Big)^K\right].
\end{aligned}$$

From Assumption C such that $E[|\varepsilon_{it} - E[\varepsilon_{it}]|^K] \le K! C_\varepsilon^K$ for some finite constant C_ε, we obtain

$$E\left[\left|\rho_\tau(y_{it}-d_j) - E\left[\rho_\tau(y_{it}-d_j)\right]\right|^K\right] \le K! C_1^{K-2} C_2/2 \tag{2}$$

for all i and t. Here C_1 and C_2 are positive constant. Define

$$Z_{it} \equiv \rho_\tau(y_{it}-d_j) - E\left[\rho_\tau(y_{it}-d_j)\right].$$

It then follows from Lemma 2 that, for all $z \ge 0$,

$$P\left(\left|\sum_{i=1}^N \sum_{t=1}^T Z_{it}\right| \ge z\right) \le 2\exp\left(-\frac{z^2/2}{(NT)\times C_2 + C_1 z}\right). \tag{3}$$

Recall the cardinality of $\mathcal{S}$ is $|\mathcal{S}| \le {C_B}^N \times {C_F}^T \times {C_\Lambda}^N = O(C_{\max}^{2N+T})$ with $C_{\max} = \max\{C_B, C_F, C_\Lambda\} = O(1)$. From (3) and Lemma 1, we finally have

$$\begin{aligned}
& E\left[\max_{j\in\mathcal{S}} \frac{1}{NT}\left|\sum_{i=1}^N\sum_{t=1}^T \left\{\rho_\tau(y_{it}-d_j) - E\left[\rho_\tau(y_{it}-d_j)\right]\right\}\right|\right] \\
\le \;& C \times \frac{1}{NT}\left(C_1 \log(1+|\mathcal{S}|) + \sqrt{(NT)C_2}\{\log(1+|\mathcal{S}|)\}^{1/2}\right) \\
\le \;& C' \times \frac{1}{NT}\left[(2N+T)\log C_{\max} + (NT)^{1/2}(2N+T)^{1/2}\log C_{\max}\right] \\
= \;& O\left(\frac{1}{N^{1/2}} + \frac{1}{T^{1/2}}\right),
\end{aligned}$$

where C and C' are some positive constant. Therefore, we obtain the result in (1).
Recall that the estimator $\{\hat{B}, \hat{F}, \hat{\Lambda}\}$ is the minimizer of
$\ell_\tau(B, F, \Lambda) = \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)$. Then, for any given F, Λ and B, we have

$$\begin{aligned}
& U_\tau(B, F, \Lambda) \\
\equiv\; & \frac{1}{NT}\ell_\tau(B, F, \Lambda) - \frac{1}{NT}\ell_\tau(B_0, F_0, \Lambda_0) \\
=\; & \left[\frac{1}{NT}\ell_\tau(B, F, \Lambda) - \frac{1}{NT}E[\ell_\tau(B, F, \Lambda)]\right] + \left[\frac{1}{NT}\ell_\tau(B_0, F_0, \Lambda_0) - \frac{1}{NT}E[\ell_\tau(B_0, F_0, \Lambda_0)]\right] \\
& + \left[\frac{1}{NT}E[\ell_\tau(B, F, \Lambda)] - \frac{1}{NT}E[\ell_\tau(B_0, F_0, \Lambda_0)]\right] \\
=\; & J_1 + J_2 + J_3,
\end{aligned}$$

where $E[\ell_\tau(B, F, \Lambda)]$ is defined as $E[\ell_\tau(B, F, \Lambda)] \equiv \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T E[\rho_\tau(y_{it} - \boldsymbol{x}_{it}\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)]$.
From (1), we have

$$\sup_{\boldsymbol{f}_t \in \mathcal{F}, \boldsymbol{\lambda}_i \in \mathcal{L}, \boldsymbol{b}_i \in \mathcal{B}} |J_1| = o_p(1), \quad \text{and} \quad |J_2| = o_p(1).$$

About J_3, note that $\varepsilon_{it} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$, with conditional density $g_{it}(\cdot|x_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0})$.
By Taylor expansion, we obtain

$$\begin{aligned}
& E\left[\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i)\right] - E\left[\rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})\right] \\
=\; & E\left[\rho_\tau(\varepsilon_{it} - \{\boldsymbol{x}_{it}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\})\right] - E\left[\rho_\tau(\varepsilon_{it})\right] \\
=\; & g_{it}\left(\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i \middle| \boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}\right) \times \left(\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2 \\
\geq\; & \bar{g} \times \left(\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2,
\end{aligned}$$

where $\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i$ is between 0 and $\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$. By assumption, $\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i$ belongs to a compact set. This leads to $0 < \bar{g} \leq g_{it}(\boldsymbol{x}_{it}'\tilde{\boldsymbol{b}}_i + \tilde{\boldsymbol{f}}_t'\tilde{\boldsymbol{\lambda}}_i | x_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0})$ by Assumption C. Therefore, the last inequality was obtained. Then,

$$\begin{aligned}
J_3 &= \frac{1}{NT}E\left[\ell_\tau(B, F, \Lambda)\right] - \frac{1}{NT}E\left[\ell_\tau(B_0, F_0, \Lambda_0)\right] \\
&\geq \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \bar{g} \times \left(\boldsymbol{x}_{i,t}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)^2.
\end{aligned}$$

Without loss of generality, we assume that $\boldsymbol{b}_{i,0} = \boldsymbol{0}$, $i = 1, ..., N$ (for notational simplicity).
Note that the centered objective function satisfies

$$U_\tau(B_0, F_0, \Lambda_0) = 0,$$

where we used that the function $h(\boldsymbol{b}_i, \boldsymbol{f}_t, \boldsymbol{\lambda}_i) \equiv \boldsymbol{x}_{i,t}'\boldsymbol{b}_i + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$ evaluated at $\boldsymbol{b}_i = \boldsymbol{b}_{i,0}$, $\boldsymbol{f}_t = \boldsymbol{f}_{t,0}$ and $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{i,0}$ is zero. Note also that

$$U_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \leq U_\tau(B_0, F_0, \Lambda_0) = 0$$

by definition of $\{\hat{B}, \hat{F}, \hat{\Lambda}\}$. Therefore, we have

$$\begin{aligned} 0 &\geq U_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \\ &\geq \frac{\bar{g}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}'_{i,t}\hat{\boldsymbol{b}}_i + \hat{\boldsymbol{f}}'_t\hat{\boldsymbol{\lambda}}_i - \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0}\right)^2 + o_p(1). \end{aligned}$$

Combined with $\frac{\bar{g}}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}'_{i,t}\boldsymbol{b}_i + \boldsymbol{f}'_t\boldsymbol{\lambda}_i - \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0})^2 \geq 0$ for any $\boldsymbol{b}_i$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{f}_t$, it must be true that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{x}'_{i,t}\hat{\boldsymbol{b}}_i + \hat{\boldsymbol{f}}'_t\hat{\boldsymbol{\lambda}}_i - \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0}\right)^2 = \frac{1}{NT}\sum_{i=1}^{N}\left\|X_i\hat{\boldsymbol{b}}_i + \hat{F}\hat{\boldsymbol{\lambda}}_i - F_0\boldsymbol{\lambda}_{i,0}\right\|^2 = o_p(1). \tag{4}$$

Define $M_F = I - F(F'F)^{-1}F'$ and

$$\tilde{U}_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \equiv \frac{1}{NT}\sum_{i=1}^{N}\left\|M_{\hat{F}}\left(X_i\hat{\boldsymbol{b}}_i - F_0\boldsymbol{\lambda}_{i,0}\right)\right\|^2 \tag{5}$$

then

$$\tilde{U}_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \leq \frac{1}{NT}\sum_{i=1}^{N}\left\|X_i\hat{\boldsymbol{b}}_i + \hat{F}\hat{\lambda}_i - F_0\boldsymbol{\lambda}_{i,0}\right\|^2 = o_p(1),$$

by equation (4). This is because M_F is a projection matrix, $\|M_F Z\| \leq \|Z\|$ for any Z, and also $M_{\hat{F}}\hat{F} = 0$.

Now, by (5)

$$\begin{aligned} &\tilde{U}_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) \\ =\ &\frac{1}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{b}}'_i X'_i M_{\hat{F}} X_i\hat{\boldsymbol{b}}_i - \frac{2}{NT}\sum_{i=1}^{N}\hat{\boldsymbol{b}}'_i X'_i M_{\hat{F}} F_0\boldsymbol{\lambda}_{i,0} + \frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{\lambda}'_{i,0}F_0{}' M_{\hat{F}} F_0\boldsymbol{\lambda}_{i,0} \\ =\ &\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}'_i A_i\hat{\boldsymbol{b}}_i - \frac{2}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}'_i C_i\boldsymbol{\eta} + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}' B_i\boldsymbol{\eta}, \end{aligned}$$

where

$$A_i = \frac{1}{T}X'_i M_{\hat{F}} X_i, \quad B_i = (\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}'_{i,0}) \otimes I_T, \quad C'_i = \frac{1}{\sqrt{T}}\boldsymbol{\lambda}'_{i,0} \otimes (X'_i M_{\hat{F}}), \quad \boldsymbol{\eta} = \frac{1}{\sqrt{T}}\mathrm{vec}(M_{\hat{F}} F_0).$$

Completing the square,

$$\tilde{U}_\tau\left(\hat{B}, \hat{F}, \hat{\Lambda}\right) = \boldsymbol{\eta}'\left(\frac{1}{N}\sum_{i=1}^{N}E_i\right)\boldsymbol{\eta} + \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i + A_i^{-1}C_i\boldsymbol{\eta}\right)' A_i\left(\hat{\boldsymbol{b}}_i + A_i^{-1}C_i\boldsymbol{\eta}\right),$$

where $E_i = B_i - C'_i A_i^{-1} C_i$. Because each of the two terms is non-negative, this implies that

$$\boldsymbol{\eta}'\left(\frac{1}{N}\sum_{i=1}^{N}E_i\right)\boldsymbol{\eta} = o_p(1), \tag{6}$$

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{b}}_i + A_i^{-1}C_i\boldsymbol{\eta}\right)' A_i\left(\hat{\boldsymbol{b}}_i + A_i^{-1}C_i\boldsymbol{\eta}\right) = o_p(1). \tag{7}$$

From Assumption D, the matrix $N^{-1}\sum_{i=1}^{N}E_i$ is positive definite, and thus equation (6) implies that $\|\boldsymbol{\eta}\|^2 = o_p(1)$. This result implies that

$$\|M_{\hat{F}} - M_{F_0}\| = \|P_{\hat{F}} - P_{F_0}\| = o_p(1). \tag{8}$$

See Bai (2009, page 1265). That is, the space spanned by F_0 and the space spanned by the estimated factors $\hat{F}$ are asymptotically the same. Thus, we obtain $\|\hat{F} - F_0\|/\sqrt{T} = o_p(1)$. From $\|\boldsymbol{\eta}\|^2 = o_p(1)$, equation (7) implies that

$$o_p(1) = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i' A_{i,0}\hat{\boldsymbol{b}}_i + \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'(A_i - A_{i,0})\hat{\boldsymbol{b}}_i \geq (\rho_A + o_p(1))\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'\hat{\boldsymbol{b}}_i,$$

where $0 < \rho_A$ is the lower bound of the eigenvalues of $A_{i,0} = \frac{1}{T}X_i' M_{F_0} X_i$ $i = 1, ..., N$. Because assumption (D2), $\rho_A > 0$ exists. We also used the fact that $\|A_i - A_{i,0}\| = o_p(1)$, which is proved in Ando and Bai (2015, Theorem 1). The average consistency of $\hat{\boldsymbol{b}}_i$ follows from $\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{b}}_i'\hat{\boldsymbol{b}}_i = o_p(1)$ (recall we normalize $b_{i,0} = 0$)). The average consistency of $\hat{\boldsymbol{f}}_t$ and the average consistency of $\hat{\boldsymbol{b}}_i$ further imply the average consistency of $\hat{\boldsymbol{\lambda}}_i$ (see Ando and Bai (2015)). That is, $N^{-1}\sum_{i=1}^{N}\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|^2 = o_p(1)$. This completes the proof of Proposition 1.

B Proof of Theorem 1

Lemma 3 *(Theorem 2 of Hoeffding (1963)). Let X_1,...,X_n be independent random variables and bounded by the interval $[C_i, D_i]$. Then, for all $r > 0$, we have*

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - E\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]\right| \geq r\right) \leq \exp\left(-\frac{2n^2r^2}{\sum_{i=1}^{n}(D_i - C_i)^2}\right).$$

B.1 Proof of Theorem 1

We first prove the uniform consistency of $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{\lambda}}_i$. Let $\boldsymbol{\gamma}_i = (\boldsymbol{b}_i', \boldsymbol{\lambda}_i')'$. We now define the following loss

$$L_{NT,i}(\boldsymbol{\gamma}_i, F) \equiv \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right),$$

and its centered version

$$\begin{aligned}\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) &\equiv L_{NT,i}(\boldsymbol{\gamma}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\\ &= \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right) - \frac{1}{T}\sum_{t=1}^{T}\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)\end{aligned}$$

For each $\delta > 0$, we also define

$$B_{i,T}(\delta) \equiv \left\{\boldsymbol{\gamma}_i : \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq \delta\right\}$$

and

$$B_T(e) \equiv \left\{ F : \|F - F_0\|/T^{1/2} \le e \right\},$$

where $e > 0$ is small positive constant. It is enough to prove the consistency under the condition that the common factor F satisfies $\|F - F_0\|/T^{1/2} < e$, which is established in Proposition 1. We prove the the consistency of $\hat{\boldsymbol{b}}_i$ and $\hat{\boldsymbol{\lambda}}_i$ in the sense of
$\max_{1\le i\le N} \|\hat{\boldsymbol{b}}_i - \boldsymbol{b}_{i,0}\| = o_p(1)$ and $\max_{1\le i\le N} \|\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{i,0}\| = o_p(1)$.
Fix any $\delta > 0$. We can assume that the common factor F satisfies $\|F - F_0\|/T^{1/2} < e$. For each $\boldsymbol{\gamma}_i \notin B_{i,T}(\delta)$, define

$$\tilde{\boldsymbol{\gamma}}_i = s_i \boldsymbol{\gamma}_i + (1 - s_i)\boldsymbol{\gamma}_{i,0},$$

where $\boldsymbol{\gamma}_{i,0} = (\boldsymbol{b}'_{i,0}, \boldsymbol{\lambda}'_{i,0})'$ and $s_i = \delta/\{\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\|\} \in (0, 1)$. The convexity of the objective function, given F, leads

$$\begin{aligned} & s_i \left\{ L_{NT,i}(\boldsymbol{\gamma}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) \right\} \\ \ge\ & L_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) + L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F) \\ >\ & E\left[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)\right] + \left\{ \tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E\left[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)\right] \right\} - C \times e, \end{aligned} \tag{9}$$

where $E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)] = E[L_{NT,i}(\boldsymbol{\gamma}_i, F)] - E[L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0)]$ and the expectation $E[\cdot]$ is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$ conditional on X, F_0, Λ_0 and B_0. Here, we used

$$|L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F)| \le C \times T^{-1} \sum_{t=1}^{T} \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \le C \times \|F - F_0\|/T^{1/2} < C \times e$$

with C being a positive constant. Note that $F \in B_T(e)$. By choosing small e, this term is dominated by the term $E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] + \{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)]\}$. Thus, this term is negligible in our analysis below.
Let $\boldsymbol{\omega}_{it} = \{\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}\}$. Under $\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \le \delta$, the identity of Knight (1998) leads to

$$\begin{aligned} & \frac{1}{T}\sum_{t=1}^{T} E\left[\rho_\tau\left(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{b}_i - \boldsymbol{f}'_t\boldsymbol{\lambda}_i\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{b}_{i,0} - \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0}\right)\right] \\ =\ & \frac{1}{T}\sum_{t=1}^{T} \int_0^{\boldsymbol{x}'_{it}(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}'_{t,0}(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}) + (\boldsymbol{f}_t - \boldsymbol{f}_{t,0})'\boldsymbol{\lambda}_i} \left\{ G_{it}(s|\boldsymbol{\omega}_{it}) - \tau \right\} ds \\ =\ & \frac{1}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it}) \left\{ \boldsymbol{x}'_{it}\left(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}\right) + \boldsymbol{f}'_{t,0}\left(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}\right) + \left(\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\right)'\boldsymbol{\lambda}_i \right\}^2 \\ =\ & \frac{1}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it}) \left\{ \boldsymbol{x}'_{it}\left(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}\right) + \boldsymbol{f}'_{t,0}\left(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}\right) \right\}^2 + \frac{1}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it}) \left\{ \left(\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\right)'\boldsymbol{\lambda}_i \right\}^2 \\ & + \frac{2}{T}\sum_{t=1}^{T} g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it}) \left\{ \boldsymbol{x}'_{it}\left(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}\right) + \boldsymbol{f}'_{t,0}\left(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}\right) \right\} \left\{ \left(\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\right)'\boldsymbol{\lambda}_i \right\} \end{aligned} \tag{10}$$

where $\tilde{h}_{it}$ is between 0 and $\boldsymbol{x}_{it}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_{t,0}'(\boldsymbol{\lambda}_i - \boldsymbol{\lambda}_{i,0}) + (\boldsymbol{f}_t - \boldsymbol{f}_{t,0})'\boldsymbol{\lambda}_i$, $G_{it}(\cdot|\boldsymbol{\omega}_{it})$ is the conditional cumulative distribution of $\varepsilon_{it} = y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}$ conditional on $\boldsymbol{\omega}_{it}$. Because the ball $B_T(e)$ can be made arbitrary small, the first term (10) dominates the other two terms.

By assumptions, $\tilde{h}_{it}$ belongs to a compact set, which leads to $0 < \bar{g} \leq g_{it}(\tilde{h}_{it}|\boldsymbol{\omega}_{it})$. Together with Assumption (D2), there exists a positive constant e_δ such that for $0 < \delta_e \leq \delta$,

$$\frac{1}{T}\sum_{t=1}^{T} \bar{g}\left\{\boldsymbol{x}_{it}'\left(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}\right) + \boldsymbol{f}_{t,0}'\left(\boldsymbol{\lambda}_{i,0} - \boldsymbol{\lambda}_i\right)\right\}^2 > e_\delta \tag{11}$$

for all $1 \leq i \leq N$, for all $\|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| = \delta$. Note that $\tilde{\boldsymbol{\gamma}}_i$ is on the boundary of $B_{i,T}(\delta)$; i.e., $\partial B_{i,T}(\delta) \equiv \{\boldsymbol{\gamma}_i : \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| = \delta\}$. This implies that with probability approaching to 1,

$$E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] \geq \frac{e_\delta}{2}. \tag{12}$$

From (9), we obtain the following inclusion relation

$$\begin{aligned}
& \left\{\max_i \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| > \delta,\ F \in B_T(e)\right\} \\
= & \left\{\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| > \delta,\ \exists i,\ F \in B_T(e)\right\} \\
\subset & \left\{L_{NT,i}(\boldsymbol{\gamma}_i, F) \leq L_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0),\ 1 \leq \exists i \leq N, \exists \boldsymbol{\gamma}_i \notin B_{i,T}(\delta), F \in B_T(e)\right\} \\
\subset & \left\{\max_{1\leq i\leq N} \sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta),\ F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]\right| \geq \frac{e_\delta}{2}\right\},
\end{aligned}$$

where the following argument is used for the second inclusion. From (9), we have $s_i\left\{L_{NT,i}(\boldsymbol{\gamma}_i, F) - L_{NT,i}(\boldsymbol{\gamma}_{i0}, F_0)\right\} \geq E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] + \{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)]\}$. If $L_{NT,i}(\boldsymbol{\gamma}_i, F) \leq L_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0),\ 1 \leq \exists i \leq N, \exists \boldsymbol{\gamma}_i \notin B_{i,T}(\delta)$, this leads to $0 \geq E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] + \{\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)]\}$. From (12), $E[\tilde{L}_{NT,i}(\tilde{\boldsymbol{\gamma}}_i, F)] \geq e_\delta/2$, note that $\tilde{\boldsymbol{\gamma}}_i \in B_{i,T}(\delta)$. Thus, we should have $\max_{1\leq i\leq N}\sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),\ F\in B_T(e)} |\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]| \geq e_\delta/2$. The second inclusion is obtained. We show this event has small probability.

It suffices to show that for every $\varepsilon > 0$,

$$\lim_{N,T\to\infty} P\left\{\max_{1\leq i\leq N} \sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),\ F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]\right| \geq \varepsilon\right\} = 0.$$

It further suffices to prove that

$$\max_{1\leq i\leq N} P\left\{\sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),\ F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]\right| \geq \varepsilon\right\} = o(N^{-1}). \tag{13}$$

Let $h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) \equiv \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i) - \rho_\tau(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})$. Observe that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\bar{\boldsymbol{\gamma}}_i, \boldsymbol{f}_{t,0})| \leq C(\|\boldsymbol{x}_{it}\| + \|\boldsymbol{f}_{t,0}\|)\|\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}_i\| + C(\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|)$ and

$|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_{t,0}) - h(\bar{\boldsymbol{\gamma}}_i, \boldsymbol{f}_{t,0})| \leq C(\|\boldsymbol{x}_{it}\| + \|\boldsymbol{f}_{t,0}\|)\|\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}_i\|$ for some universal constant $C > 0$. Put $W_{it} \equiv C(\|\boldsymbol{x}_{it}\| + \|\boldsymbol{f}_{t,0}\|)$ and $\kappa \equiv \sup_{i,t} W_{it}$. Since $B_{i,T}(\delta)$ is compact subset, there exist C_Γ balls with centers $\bar{\boldsymbol{\gamma}}_k = (\bar{\boldsymbol{b}}_k', \bar{\boldsymbol{\lambda}}_k)'$, $k = 1, ..., C_\Gamma$ and radius $\varepsilon/(8\kappa)$ such that the collection of these balls covers $B_{i,T}(\delta)$. Note that C_Γ can be chosen such that $C_\Gamma(\varepsilon) = O(1/\varepsilon^{p+r})$ and $\varepsilon \to 0$. For each $\boldsymbol{\gamma}_i \in B_{i,T}(\delta)$, there is $\bar{\boldsymbol{\gamma}}_k$ such that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\bar{\boldsymbol{\gamma}}_k, \boldsymbol{f}_{t,0})| \leq W_{it}\varepsilon/(8\kappa)$. These investigations lead to

$$
\begin{aligned}
& |\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]| \\
\leq\ & |\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0)]| + \frac{C}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \\
\leq\ & |\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0) - E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0)]| + \frac{\varepsilon}{8\kappa} \times \frac{1}{NT}\left|\sum_{t=1}^{T}\sum_{i=1}^{N}\{W_{it} + E[W_{it}]\}\right| + \frac{C}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|,
\end{aligned}
$$

where the last term can be made arbitrarily small under large N and T because $T^{-1}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|^2 \leq e$, which implies $T^{-1}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \sqrt{\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|^2/T} \leq \sqrt{e}$. Therefore, we have

$$
\begin{aligned}
& P\left\{\sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta), F \in B_T(e)}\left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F_0)]\right| \geq \varepsilon\right\} \\
\leq\ & \sum_{k=1}^{C_\Gamma(\varepsilon)} P\left\{\left|\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0) - E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0)]\right| \geq \frac{\varepsilon}{3}\right\} \\
& + P\left\{\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\{W_{it} + E[W_{it}]\}\right| \geq \frac{8\kappa}{3}\right\} + P\left\{\left|\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\right| \geq \frac{\varepsilon}{3C}\right\},
\end{aligned}
$$

where the second term is 0 because $N^{-1}T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\{W_{it} + E[W_{it}]\}$ is smaller than 2κ. The last term can be made arbitrarily small under large N and T. Thus, we need to evaluate the first term.

Because of the uniform boundedness of $\boldsymbol{x}_{it}$ and $\boldsymbol{f}_t$, we have $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t)]| \leq M$ where M is bounded constant. From the independence property of ε_{it}, Lemma 3 leads

$$
\begin{aligned}
& P\left\{\left|\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0) - E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0)]\right| \geq \frac{\varepsilon}{3}\right\} \\
=\ & P\left\{\left|\frac{1}{T}\sum_{t=1}^{T}\left\{\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\bar{\boldsymbol{b}}_k - \boldsymbol{f}_{t,0}'\bar{\boldsymbol{\lambda}}_k\right) - E\left[\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\bar{\boldsymbol{b}}_k - \boldsymbol{f}_{t,0}'\bar{\boldsymbol{\lambda}}_k\right)\right]\right\}\right| \geq \frac{\varepsilon}{3}\right\} \\
\leq\ & \exp\left(\frac{-T^2\varepsilon^2}{18\sum_{t=1}^{T}M^2}\right),
\end{aligned}
$$

which leads

$$
\sum_{k=1}^{C_\Gamma(\varepsilon)} P\left\{\left|\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0) - E[\tilde{L}_{NT,i}(\bar{\boldsymbol{\gamma}}_k, F_0)]\right| \geq \frac{\varepsilon}{3}\right\}
$$

$$\begin{aligned}
&\leq C_\Gamma(\varepsilon) \times \exp\left(\frac{-T^2\varepsilon^2}{18\sum_{t=1}^T M^2}\right) \\
&= O\left(\varepsilon^{-p-r} \times \exp(-T)\right) \\
&= O\left(\exp\left[-T\left\{(1+(p+r)\frac{\log(\varepsilon)}{T}\right\}\right]\right) \\
&= o(N^{-2}), \qquad (14)
\end{aligned}$$

where the last line is obtained because $\log(N)/\sqrt{T} \to 0$ and $\log(\varepsilon)/T \to 0$. Thus, we have

$$\max_{1\leq i\leq N} P\left\{\sup_{\boldsymbol{\gamma}_i \in B_{i,T}(\delta), F\in B_T(e)} \left|\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\boldsymbol{\gamma}_i, F)]\right| \geq \varepsilon\right\} = o(N^{-1}),$$

which completes the proof of the uniform consistency of $\hat{\boldsymbol{\gamma}}_i$ for $i = 1, ..., N$.

Next, we prove that the estimated common factor is $\hat{\boldsymbol{f}}_t$ is uniformly consistent

$$\max_{1\leq t\leq T} \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| \to 0. \qquad (15)$$

Let $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', ..., \boldsymbol{\gamma}_N')'$. We define the following loss

$$S_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) \equiv \frac{1}{N}\sum_{i=1}^N \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right),$$

and its centered version

$$\begin{aligned}
\tilde{S}_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) &\equiv S_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) - S_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) \\
&= \frac{1}{N}\sum_{i=1}^N \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i - \boldsymbol{f}_t'\boldsymbol{\lambda}_i\right) - \frac{1}{N}\sum_{i=1}^N \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right),
\end{aligned}$$

where $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{1,0}', ..., \boldsymbol{\gamma}_{N,0}')'$. Fix any $\delta > 0$. For each $\boldsymbol{f}_t$ such that $\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| > \delta$, define $\tilde{\boldsymbol{f}}_t = s_t\boldsymbol{f}_t + (1-s_t)\boldsymbol{f}_{t,0}$ with $s_t = \delta/\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \in (0,1)$. Then, $\|\tilde{\boldsymbol{f}}_t - \boldsymbol{f}_t\| = \delta$. From (11), we obtain the inclusion relation

$$\begin{aligned}
&\left\{\max_t \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| > \delta\right\} \\
\equiv\ &\left\{\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| > \delta,\ \exists t\right\} \\
\subset\ &\left\{S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) \leq S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}),\ 1 \leq \exists t \leq T, \exists \boldsymbol{f}_t \text{ s.t. } \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| > \delta\right\} \\
\subset\ &\left\{\max_{1\leq t\leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\|\leq\delta} \left|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]\right| \geq \frac{e_\delta}{2}\right\},
\end{aligned}$$

where $E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t)] = E[S_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t)] - E[S_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})]$ and the expectation $E[\cdot]$ is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$

conditional on X, F_0 and Λ_0. The second inclusion is obtained as follows. Because of the convexity of the objective function, we have

$$\begin{aligned} & s_t \left\{ S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \right\} \\ \geq\ & S_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \\ =\ & \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \\ =\ & \left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t) \right] \right] + E\left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \right] \\ & - \left\{ \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) - E\left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \right] \right\} + E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t) \right]. \end{aligned}$$

Similar to (12), the fourth term in the last line is greater than or equal to e_δ for some $e_\delta > 0$. By consistency of $\hat{\boldsymbol{\gamma}}$, the second and third terms in the last line are $o_p(1)$. Thus, we have

$$\begin{aligned} & s_t \left\{ S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0}) \right\} \\ \geq\ & \left[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \tilde{\boldsymbol{f}}_t) - E\left[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \tilde{\boldsymbol{f}}_t) \right] \right] + e_\delta + o_p(1). \end{aligned}$$

If $S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) \leq S_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_{t,0})$ $1 \leq \exists t \leq T$ and $\exists \boldsymbol{f}_t$ s.t. $\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| > \delta$, under large N and T, we should have $\max_{1 \leq t \leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \delta} |\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)]| \geq e_\delta/2$. Note that $\|\tilde{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = \delta$. Therefore, the second inclusion is obtained.
Therefore, it suffices to show that for every $\varepsilon > 0$,

$$\lim_{N,T \to \infty} P \left\{ \max_{1 \leq t \leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \delta} \left| \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)] \right| \geq \varepsilon \right\} = 0.$$

It suffices to prove that

$$\lim_{N,T \to \infty} P \left\{ \max_{1 \leq t \leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \delta} \left| \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - E[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t)] \right| \geq \varepsilon \right\} = 0, \tag{16}$$

and

$$\lim_{N,T \to \infty} P \left\{ \max_{1 \leq t \leq T} \sup_{\|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq \delta} \left| E[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t)] - E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t)] \right| \geq \varepsilon \right\} = 0. \tag{17}$$

Since

$$\left| \tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}}, \boldsymbol{f}_t) - \tilde{S}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_t) \right| \leq C \times \left\{ \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{x}_{it}\| \|\hat{\boldsymbol{b}}_i - \boldsymbol{b}_{i,0}\| + \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{f}_t\| \|\hat{\boldsymbol{\lambda}}_i - \boldsymbol{\lambda}_{i,0}\| \right\},$$

and $\sup_{it} \|\boldsymbol{x}_{it}\| < \infty$ and $\sup_t \|\boldsymbol{f}_t\| < \infty$, consistency of $\hat{\boldsymbol{\gamma}}$ implies (17).

Finally, we prove (16). Because we already proved the consistency of $\hat{\boldsymbol{\gamma}}_i$ $(i = 1, ..., N)$, it is enough to show

$$\max_{1\le t\le T} P\left\{\sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),\|\boldsymbol{f}_t-\boldsymbol{f}_{t,0}\|\le\delta}\left|\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)]\right|\ge\varepsilon\right\}=o(T^{-1}). \qquad (18)$$

For each $\delta > 0$, we define

$$B_{t,N}(\delta) \equiv \left\{\boldsymbol{f}_t : \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \le \delta\right\}.$$

Observe that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\boldsymbol{\gamma}_i, \bar{\boldsymbol{f}}_t)| \le C\|\boldsymbol{\lambda}_i\| \times \|\bar{\boldsymbol{f}}_t - \boldsymbol{f}_t\|$ for some universal constant $C > 0$. Put $K_i \equiv C\|\boldsymbol{\lambda}_i\|$ and $\beta \equiv \sup_i K_i$. Since $B_{t,N}(\delta)$ is compact subset, there exist C_f balls with centers $\bar{\boldsymbol{f}}_k$, $k = 1, ..., C_f$ and radius ε/β such that the collection of these balls covers $B_{t,N}(\delta)$. Note that C_f can be chosen such that $C_f(\varepsilon) = O(1/\varepsilon^r)$ and $\varepsilon \to 0$. For each $\boldsymbol{f}_t \in B_{t,N}(\delta)$, there is $\bar{\boldsymbol{f}}_k$ such that $|h(\boldsymbol{\gamma}_i, \boldsymbol{f}_t) - h(\boldsymbol{\gamma}_i, \bar{\boldsymbol{f}}_k)| \le K_i\varepsilon/(8\beta)$. These investigations lead to

$$\begin{aligned}
& |\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)]| \\
\le\ & |\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_t)-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_t)]|+|\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)-\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_t)| \\
& +|E[\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)]-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_t)]| \\
\le\ & |\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_t)-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\boldsymbol{f}_t)]|+C\times\frac{1}{N}\sum_{i=1}^N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\| \\
\le\ & |\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\bar{\boldsymbol{f}}_k)-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma}_0,\bar{\boldsymbol{f}}_k)]|+\frac{\varepsilon}{8\beta}\left|\frac{1}{N}\sum_{i=1}^N\{K(\boldsymbol{\lambda}_i)+E[K(\boldsymbol{\lambda}_i)]\}\right|+\frac{C}{N}\sum_{i=1}^N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_0\|,
\end{aligned}$$

where C is universal constant. Therefore, we have

$$\begin{aligned}
& P\left\{\sup_{\boldsymbol{\gamma}_i\in B_{i,T}(\delta),F\in B_T(e)}\left|\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)-E[\tilde{S}_{NT,t}(\boldsymbol{\gamma},\boldsymbol{f}_t)]\right|\ge\varepsilon\right\} \\
\le\ & \sum_{k=1}^{C_\Gamma(\varepsilon)}P\left\{\left|\tilde{S}_{NT,t}(\bar{\boldsymbol{\gamma}}_k,F_0)-E[\tilde{S}_{NT,t}(\bar{\boldsymbol{\gamma}}_k,F_0)]\right|\ge\frac{\varepsilon}{3}\right\} \\
& +P\left\{\left|\frac{1}{N}\sum_{i=1}^N\{K_i+E[K_i]\}\right|\ge\frac{8\beta}{3}\right\}+P\left\{\left|\frac{1}{N}\sum_{i=1}^N\|\boldsymbol{\lambda}_i-\boldsymbol{\lambda}_{i,0}\|\right|\ge\frac{\varepsilon}{3C}\right\},
\end{aligned}$$

where the last term can be made arbitrarily small under large N and T. The second term is zero because $|\frac{1}{N}\sum_{i=1}^N\{K_i + E[K_i]\}| < 2\beta$. Because of the uniform boundedness of $\boldsymbol{x}_{it}$ and $\boldsymbol{f}_t$, we have $|h(\boldsymbol{\gamma}_{i,0}, \boldsymbol{f}_t) - E[h(\boldsymbol{\gamma}_{i,0}, \boldsymbol{f}_t)]| \le M$ where M is bounded constant. The first term is

$$\sum_{k=1}^{C_f(\varepsilon)}P\left\{\left|\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}},\bar{\boldsymbol{f}}_t)-E[\tilde{S}_{NT,t}(\hat{\boldsymbol{\gamma}},\bar{\boldsymbol{f}}_t)]\right|\ge\frac{\varepsilon}{3}\right\}$$

$$
\begin{aligned}
&\leq \; C_f(\varepsilon) \times \exp\left(\frac{-N^2\varepsilon^2}{18\sum_{i=1}^{N} M^2}\right) \\
&= \; O\left(\varepsilon^{-r} \times \exp(-N)\right) \\
&= \; O\left(\exp\left[-N\left\{1 + r\frac{\log(\varepsilon)}{N}\right\}\right]\right) \\
&= \; o(T^{-2})
\end{aligned} \tag{19}
$$

where Lemma 3 is applied to obtain the second inequality. Because $(\log T)/\sqrt{N} \to 0$, the last line is obtained. This implies (18), or equivalently, (16). This completes the proof of Theorem 1.

C Lemma 4

Lemma 4 *Under Assumptions of Theorem 2, the following results hold.*

$$
\max_{1\leq i\leq N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p\left(\frac{\log(N)}{T^{1/2}}\right), \tag{20}
$$

$$
\max_{1\leq t\leq T} \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = O_p\left(\frac{\log(T)}{N^{1/2}}\right). \tag{21}
$$

We note that these results are the preliminary convergence rates of $\hat{\boldsymbol{\gamma}}_i$ and $\hat{\boldsymbol{f}}_t$. The above results are used to prove Theorem 2. We first prepare lemmas which will be used in the proof of Lemma 4. The first lemma is due to Lemma 1 of Babu (1989), which is closely related to the Bernstein inequality.

Lemma 5 *Let X_i be a sequence of independent random variables with mean zero and $|X_i| < d$ for some $d > 0$. Let $V \geq \sum_{i=1}^{N} E[X_i^2]$. Then for all $0 < s < 1$ and $0 \leq a \leq V/(sd)$,*

$$
P\left(\left|\sum_{i=1}^{N} X_i\right| \geq a\right) \leq 2\exp(-a^2 s(1-s)/V).
$$

To state the next lemma, we introduce some notations. We let $\{\xi_t, t \geq 1\}$ be a stationary process taking values in a measurable space $(\mathcal{S}, \mathcal{Q})$. Here $\mathcal{S}$ is a Polish space and $\mathcal{Q}$ is a Borel σ-field. We denote $\mathcal{H}$ being a class of measurable functions on the measurable space $(S, \mathcal{Q})$. For a process $Z(h)$ defined on $(\mathcal{S}, \mathcal{Q})$, we define $\|Z(h)\|_{\mathcal{H}} \equiv \sup_{h\in\mathcal{H}} |Z(h)|$. The following lemma is a Bernstein type inequality for centered empirical processes (Talagrand (1996), Bousquet (2002)). The following Talagrand type inequality is due to Proposition B.2. of Kato et al. (2012).

Lemma 6 *Let $\mathcal{H}$ be a pointwise measurable class of functions on the measurable space $(\mathcal{S}, \mathcal{Q})$ uniformly bounded by some constant U. Suppose that, for any $h(\cdot) \in \mathcal{H}$, (i)*

$E[h(\xi_t)] = 0$, *(ii)* $\sup_{h\in\mathcal{H}}[h(\xi_t)^2] \le \sigma^2$. *For* $Z = \|\sum_{t=1}^T h(\xi_t)\|_{\mathcal{H}}$, *we have*

$$P\left(Z - E[Z] \ge \sqrt{2}s\left(T\sigma^2 + 2UE[Z]\right)^{1/2} + \frac{s^2U}{3}\right) \le \exp(-s^2),$$

for all $s > 0$.

C.1 Proof of Lemma 4

We define $\boldsymbol{z}_{it} = (\boldsymbol{x}_{it}', \boldsymbol{f}_t')'$, $\boldsymbol{z}_{it,0} = (\boldsymbol{x}_{it}', \boldsymbol{f}_{t,0}')'$, $\hat{\boldsymbol{z}}_{it} = (\boldsymbol{x}_{it}', \hat{\boldsymbol{f}}_t')'$, $\boldsymbol{\omega}_{it} = \{\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}\}$ and

$$\begin{aligned}
Q_{NT,i}(\boldsymbol{\gamma}_i, F) &\equiv \frac{1}{T}\sum_{t=1}^T \left(\tau - I(\varepsilon_{it} \le \boldsymbol{x}_{it}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})\right)\boldsymbol{z}_{it},\\
\overline{Q}_{NT,i}(\boldsymbol{\gamma}_i, F) &\equiv E[Q_{NT,i}(\boldsymbol{\gamma}_i, F)]\\
&= \frac{1}{T}\sum_{t=1}^T E\left[\left(\tau - G_{it}(\boldsymbol{x}_{it}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\middle|\boldsymbol{\omega}_{it}\right)\boldsymbol{z}_{it}\right],
\end{aligned}$$

where the expectation of $E[Q_{NT,i}(\boldsymbol{\gamma}_i, F)]$ is taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N,\ t = 1, ..., T\}$ conditional on X, F_0 and Λ_0, $G_{it}(\cdot|\boldsymbol{x}_{it})$ is the conditional cumulative distribution function of ε_{it}. Because of the computational property of the quantile regression estimator (Gutenbrunner and Jureckova (1992)), it is known that $\max_{1\le i\le N} |Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F})|$ is bounded by $O_p(T^{-1}\max_{1\le i\le N,\ 1\le t\le T}\|\boldsymbol{x}_{it}\|) + O_p(T^{-1}\max_{1\le t\le T}\|\hat{\boldsymbol{f}}_t\|) = O_p(T^{-1})$. Here, we used $\max_{1\le t\le T}\|\hat{\boldsymbol{f}}_t\| = O_p(1)$. We thus have

$$\begin{aligned}
O_p(T^{-1}) &= Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F})\\
&= Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) + \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F})\\
&\quad + \left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right\}. \qquad (22)
\end{aligned}$$

Expanding $\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F})$ at $(\boldsymbol{\gamma}_{i,0}, F_0)$, we obtain

$$\begin{aligned}
&\overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F})\\
=\ & -\frac{1}{T}\sum_{t=1}^T E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{z}_{it,0}'](\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) - \frac{1}{T}\sum_{t=1}^T E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}'](\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})\\
&+ o_p(\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|) + o_p\left(\frac{1}{T}\sum_{t=1}^T\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|\right), \qquad (23)
\end{aligned}$$

where we used the result of Theorem 1 such that $O_p(\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|^2) = o_p(\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|)$. It then follows from (22) and (23) that

$$\begin{aligned}
&\Gamma_i(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) + o_p(\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|)\\
=\ & Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) + \left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right\}\\
&- \left[\frac{1}{T}\sum_{t=1}^T E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}'](\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})\right] + O_p\left(\frac{1}{T}\right) + o_p\left(\frac{1}{T}\sum_{t=1}^T\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|\right), \qquad (24)
\end{aligned}$$

where $\Gamma_i = T^{-1}\sum_{t=1}^T E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{z}_{it,0}']$ and we used the result of Theorem 1 such that $O_p(T^{-1}\sum_{t=1}^T \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|^2) = o_p(T^{-1}\sum_{t=1}^T \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|)$.
Next, we define

$$V_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) \equiv \frac{1}{N}\sum_{i=1}^N \left(\tau - I(\varepsilon_{it} \le \boldsymbol{x}_{it}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})\right)\boldsymbol{\lambda}_i,$$

and

$$\overline{V}_{NT,t}(\boldsymbol{\gamma}, \boldsymbol{f}_t) \equiv E\left[\frac{1}{N}\sum_{i=1}^N \left(\tau - G_{it}(\boldsymbol{x}_{it}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}_t'\boldsymbol{\lambda}_i - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\Big|\boldsymbol{\omega}_{it})\right)\boldsymbol{\lambda}_i\right],$$

where $G_{it}(\cdot|\boldsymbol{\omega}_{it})$ is the true cumulative distribution function of ε_{it} conditioned on $\boldsymbol{\omega}_{it}$.
Noting that $\overline{V}_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) = 0$, the expansion of $\overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{f}}_t)$ at $(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})$ leads

$$\begin{aligned}
&\overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{f}}_t) \\
=\ & -\frac{1}{N}\sum_{i=1}^N E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}_{i,0}'](\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}) - \frac{1}{N}\sum_{i=1}^N E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}_{it,0}'](\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) \\
& + o_p\left(\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|\right) + o_p\left(\frac{1}{N}\sum_{i=1}^N \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|\right). \qquad (25)
\end{aligned}$$

Viewing the factor loadings $\boldsymbol{\lambda}_i$ as regressors and $\boldsymbol{f}_t$ as regression coefficient, the similar argument that derived the equation (24) leads

$$\begin{aligned}
& \hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0} + o_p(\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|) \\
=\ & \Psi_t^{-1}V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) - \Psi_t^{-1}\left\{V_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - \overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right\} \\
& + \Psi_t^{-1}\left(\frac{1}{N}\sum_{i=1}^N E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}_{it,0}'](\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})\right) + o_p\left(\frac{1}{N}\sum_{i=1}^N \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|\right) + O_p\left(\frac{1}{N}\right), \qquad (26)
\end{aligned}$$

where $\Psi_t = N^{-1}\sum_{i=1}^N E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{\lambda}_{i,0}']$.
Putting (26) into (24), we have

$$\begin{aligned}
& (\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) \\
=\ & \Gamma_i^{-1}Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) - \Gamma_i^{-1}\left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right\} \\
& - \frac{1}{T}\sum_{t=1}^T \Gamma_i^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}']\Psi_t^{-1}\left(\frac{1}{N}\sum_{j=1}^N E[g_{jt}(0|\boldsymbol{\omega}_{jt})\boldsymbol{\lambda}_{j,0}\boldsymbol{z}_{jt,0}'](\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0})\right) \\
& + \frac{1}{T}\sum_{t=1}^T \Gamma_i^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}']\Psi_t^{-1}V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) \\
& + \frac{1}{T}\sum_{t=1}^T \Gamma_i^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}_{i,0}']\Psi_t^{-1}\left\{V_{NT,t}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{f}}_t) - \overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{f}}_t) - V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right\} \\
& + O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + o_p(\|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\|), \qquad (27)
\end{aligned}$$

From now, we evaluate each of the terms in (27). First, the first term is

$$\max_i Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) = O_p\left(\frac{\log(N)}{T^{1/2}}\right). \tag{28}$$

To have (28), it suffices to show that, for any $u > 0$,

$$\max_{1\le i\le N} P\left(\left|Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right| > \frac{\log(N)}{T^{1/2}}u\right) = o\left(\frac{1}{N}\right),$$

which can be obtained by applying Lemma 5 with $a = 2\log(N)T^{1/2}$.

The second term in (27) is bounded by $o_p(T^{-1/2})$. To evaluate the second term $\Gamma_i^{-1}\{Q_{NT,i}(\hat{\gamma}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\gamma}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\}$ in (27), we apply Lemma 6. Define $h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) = I(u \le \boldsymbol{x}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0} - \boldsymbol{f}'_t\boldsymbol{\lambda}_i) - I(u \le 0)$. From the result of Theorem 1, we define $\mathcal{H} = \{h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) | \|\hat{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| < \kappa,\ \|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| < \kappa\}$ with $\kappa \to 0$. It is obvious that $E[h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]] = 0$ and that each element in $\mathcal{H}$ is bounded by 2. Also, $E[\{h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}^2] \le C \times \kappa^2$ for $h \in \mathcal{H}$. Put $Z_i = \|\sum_{t=1}^T\{h(\xi_{it}|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\xi_{it}|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\|_{\mathcal{H}}$ with $\xi_{it} = \varepsilon_{it}$ in Lemma 6. By Lemma 6, for all $s > 0$ with probability at least $1 - \exp(-s^2)$, we have

$$\frac{1}{T}Z_i \le \frac{1}{T}E[Z_i] + \frac{s\sqrt{2}}{T}\sqrt{T\kappa^2 + 4E[Z_i]} + \frac{2s^2}{3T}. \tag{29}$$

Because of the independence property of the idiosyncratic errors over i and t, we see that $T^{-1}E[Z_i] = o(T^{-1/2})$. We now take $s = \sqrt{2\log N}$ in (29). Then, it is seen that there exist a positive integer T_0 independent of i such that the right side on (29) is smaller than $o(T^{-1/2})$ for $T_0 < T$. This implies that $P(T^{-1}Z_i > T^{-1/2}) \le N^{-2}$. Therefore, the second term in (27) is bounded by $o_p(T^{-1/2})$.

Next, we show that the fourth term in (27) is $o_p(T^{-1/2})$. The fourth term of (27) satisfies

$$\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^T \Gamma_i^{-1}J_{it}\Psi_t^{-1}V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right] = \frac{1}{N\sqrt{T}}\sum_{i=1}^N\sum_{t=1}^T \Gamma_i^{-1}J_{it}\Psi_t^{-1}\left(\tau - I(\varepsilon_{it} \le 0)\right)\boldsymbol{\lambda}_i = o_p(1), \tag{30}$$

where $J_{it} = E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}]$. Because $\|\Gamma_i^{-1}J_{it}\Psi_t^{-1}\| < \infty$, it is enough to show that

$$\frac{1}{N\sqrt{T}}\sum_{i=1}^N\sum_{t=1}^T \left(\tau - I(\varepsilon_{it} \le 0)\right)\boldsymbol{\lambda}_i = o_p(1). \tag{31}$$

The expected value of its second moment is

$$\begin{aligned}
&\frac{1}{N^2T}E\left[\sum_{i,j=1}^N\sum_{t,s=1}^T \left(\tau - I(\varepsilon_{it} \le 0)\right)\boldsymbol{\lambda}_i\boldsymbol{\lambda}'_j\left(\tau - I(\varepsilon_{js} \le 0)\right)\right] \\
=\ &\frac{1}{N^2T}E\left[\sum_{i=1}^N\sum_{t,s=1}^T \left(\tau - I(\varepsilon_{it} \le 0)\right)\boldsymbol{\lambda}_i\boldsymbol{\lambda}'_i\left(\tau - I(\varepsilon_{is} \le 0)\right)\right] \\
=\ &\frac{1}{N^2T}E\left[\sum_{i=1}^N\sum_{t=1}^T \left(\tau - I(\varepsilon_{it} \le 0)\right)^2\boldsymbol{\lambda}_i\boldsymbol{\lambda}'_i\right],
\end{aligned}$$

which converges to zero. Here the first and second equality used the fact that the idiosyncratic errors are independent over i and t. Therefore, we obtain the claim (30).
We next evaluate the fifth term in (27), that is $O_p(T^{-1/4})$. This rate will be improved later on. Because of the consistency of $\{\hat{\boldsymbol{\gamma}}, \hat{F}\}$ and the uniform boundedness of $\|\Gamma_i^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\|$, $\boldsymbol{\lambda}_i$ and $\boldsymbol{\lambda}_{i,0}$, it suffices to show that

$$P\left\{\sup_{\boldsymbol{\gamma}\in B_\gamma(\kappa), F\in B_F(\kappa)}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right\|\geq\frac{1}{T^{1/4}}\right\} \quad (32)$$

converges to zero. Here $\zeta_{it}=\varepsilon_{it}$,
$h(u|\boldsymbol{x},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)=I(u\leq\boldsymbol{x}'(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+\boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0}-\boldsymbol{f}'_t\boldsymbol{\lambda}_i)-I(u\leq 0)$,
$B_\gamma(\kappa)\equiv\{\gamma;\ \|\boldsymbol{\gamma}_i-\boldsymbol{\gamma}_{i,0}\|\leq\kappa,\ i=1,...,N\}$, and $B_F(\kappa)\equiv\{F;\ \|\boldsymbol{f}_t-\boldsymbol{f}_{t,0}\|\leq\kappa,\ t=1,...,T\}$ with $\kappa\to 0$.
For some positive constant C_l, put $W_i\equiv C_l\times(\|\boldsymbol{\lambda}_i\|+\|\boldsymbol{\lambda}_{i,0}\|)$, $\alpha\equiv\sup_i E[W_i]$ and $\varepsilon=\frac{1}{T^{1/4}}$. There exist C_{γ_i} balls with centers $\bar{\boldsymbol{\gamma}}_{i_k}$, $k=1,...,C_{\gamma_i}$ and radius $\varepsilon/(8\alpha)$ such that the collection of these balls covers $B_{\gamma_i}(\kappa)\equiv\{\boldsymbol{\gamma}_i:\|\boldsymbol{\gamma}_i-\boldsymbol{\gamma}_{i,0}\|\leq\kappa\}$. Similarly, because $B_{f_t}(e)$ is compact subset, there exist C_{f_t} balls with centers $\bar{\boldsymbol{f}}_{t,j}$ $j=1,...,C_{f_j}$ and radius $\varepsilon/(8\alpha)$ such that the collection of these balls covers $B_{f_t}(e)$. We note that C_{γ_i} and C_{f_t} can be chosen such that $C_{\gamma_i}(\varepsilon)=O(1/\varepsilon^{p+r})$ and $C_{f_t}(\varepsilon)=O(1/\varepsilon^{r})$ with $\varepsilon\to 0$.
Because

$$\begin{aligned}&\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right|\\ \leq\ &\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right|+\left|\frac{\varepsilon}{8\alpha}\cdot\frac{1}{N}\sum_{i=1}^{N}\{W_i+E[W_i]\}\right|,\end{aligned}$$

we have

$$\begin{aligned}&P\left\{\sup_{\boldsymbol{\gamma}\in B_\Gamma(\delta),\ F\in B_F(e)}\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\right|\geq\frac{1}{T^{1/4}}\right\}\\ \leq\ &\sum_{k_1=1}^{C_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{C_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{C_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{C_{f,T}(\varepsilon)}P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})-E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right|\geq\frac{1}{2T^{1/4}}\right\}\\ &+P\left\{\left|\frac{\varepsilon}{8\alpha}\frac{1}{N}\sum_{i=1}^{N}\{W_i+E[W_i]\}\right|\geq\frac{1}{2T^{1/4}}\right\},\end{aligned}$$

where the second term is zero because $|\frac{1}{N}\sum_{i=1}^{N}\{W_i+E[W_i]\}|<2\alpha$ and $\varepsilon=\frac{1}{T^{1/4}}$ from its definition.
From the independence property of the idiosyncratic errors over t and the consistency of the estimated parameters,

$$E\left[\left\{\frac{1}{T}\sum_{t=1}^{T}h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)-\frac{1}{T}\sum_{t=1}^{T}E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\right\}^2\right]$$

$$
\begin{aligned}
&= \frac{1}{T^2}\sum_{t=1}^{T} E\left[\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}^2\right] \\
&= o(T^{-1})
\end{aligned}
$$

Therefore,

$$
\sum_{i=1}^{N} E\left\{T^{-1}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}^2 \le C \times N \times o(T^{-1})
$$

where C is some positive constant. Take
$Z_i = T^{-1}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]$ in Lemma 5, we then have

$$
\begin{aligned}
&P\left\{\frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \ge \frac{1}{2T^{1/4}}\right\} \\
= \; &P\left\{\left|\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - \frac{1}{T}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}\right| \ge \frac{N}{2T^{1/4}}\right\} \\
\le \; &\exp\left(-C \times \frac{N^2}{T^{1/2}\times N \times o(T^{-1})}\right)
\end{aligned}
$$

which leads

$$
\begin{aligned}
&\sum_{k_1=1}^{C_{\gamma_1}(\varepsilon)} \cdots \sum_{k_N=1}^{C_{\gamma_N}(\varepsilon)} \sum_{j_1=1}^{C_{f,1}(\varepsilon)} \cdots \sum_{j_T=1}^{C_{f,T}(\varepsilon)} P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \ge \frac{1}{2T^{1/4}}\right\} \\
&\le C_{\gamma_1}(\varepsilon)\times\cdots\times C_{\gamma_N}(\varepsilon)\times C_{f_1}(\varepsilon)\times\cdots\times C_{f_T}(\varepsilon)\times\exp\left(-C\times NT^{1/2}\right) \\
&= O\left(\varepsilon^{-N(p+r)}\times\varepsilon^{-Tr}\times\exp\left[-NT^{1/2}\right]\right) \\
&= O\left(\exp\left[-NT^{1/2}\left\{1-\frac{N(p+r)\log(\varepsilon)}{NT^{1/2}}-\frac{Tr\log(\varepsilon)}{NT^{1/2}}\right\}\right]\right) \\
&= o(1),
\end{aligned}
$$

where the last line is obtained by using $\log(\varepsilon)/T \to 0$, $\log(\varepsilon)/N \to 0$, $\sqrt{T}/N^{1-\gamma} \to 0$ and $\sqrt{N}/T^{1-\gamma} \to 0$. Here, $\gamma > 1/16$ is defined in Theorem 2. Therefore, the fifth term in (27) is $O_p(T^{-1/4})$.

Putting these results into (27), we have

$$
\begin{aligned}
&(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) \\
= \; &-\frac{1}{T}\sum_{t=1}^{T}\Gamma_i^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}]\Psi_t^{-1}\left(\frac{1}{N}\sum_{j=1}^{N}E[g_{jt}(0|\boldsymbol{\omega}_{jt})\boldsymbol{\lambda}_{j,0}\boldsymbol{z}'_{jt,0}](\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0})\right) + O_p\left(\frac{1}{T^{1/4}}\right),
\end{aligned}
$$

which leads the following expression

$$
(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \frac{1}{N}K(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + O_p(1/T^{1/4}),
$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1, ..., \hat{\boldsymbol{\gamma}}'_N)'$, $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{1,0}, ..., \boldsymbol{\gamma}'_{N,0})'$, and

$$K = \begin{pmatrix} K_{11} & K_{12} & \cdots & K_{1N} \\ K_{21} & K_{22} & \cdots & K_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ K_{N1} & K_{N2} & \cdots & K_{NN} \end{pmatrix}, \tag{33}$$

where $K_{ij} = T^{-1}\sum_{t=1}^{T} \Gamma_i^{-1} J_{it} \Psi_t^{-1} J'_{jt}$, $J_{it} = E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}]$. Because $(I - \frac{1}{N}K)$ is positive definite matrix, we obtain $\max_{1\leq i\leq N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(1/T^{1/4})$.
From $\max_{1\leq i\leq N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(1/T^{1/4})$, we further can improve the convergence rate of the fifth term in (27). We show that

$$P\left\{ \sup_{\boldsymbol{\gamma}\in B_\gamma(T^{-1/4}), F\in B_F(N^{-1/4})} \left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h_{it}(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h_{it}(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}\right\| \geq \frac{1}{T^{3/8}}\right\} \tag{34}$$

converges to zero. Here $\zeta_{it} = \varepsilon_{it}$, $\boldsymbol{\gamma}_i = (\boldsymbol{b}'_i, \boldsymbol{\lambda}'_i)'$,
$h(u|\boldsymbol{x}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) = I(u \leq \boldsymbol{x}'(\boldsymbol{b}_i - \boldsymbol{b}_{i,0}) + \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0} - \boldsymbol{f}'_t\boldsymbol{\lambda}_i) - I(u \leq 0)$,
$B_\gamma(T^{-1/4}) \equiv \{\gamma;\ \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq T^{-1/4},\ i = 1, ..., N\}$, and
$B_F(N^{-1/4}) \equiv \{F;\ \|\boldsymbol{f}_t - \boldsymbol{f}_{t,0}\| \leq N^{-1/4},\ t = 1, ..., T\}$.
For some positive constant C_l, put $M_i \equiv C_l \times (\|\boldsymbol{\lambda}_i\| + \|\boldsymbol{\lambda}_{i,0}\|)$, $\alpha \equiv \sup_i M_i$ and $\varepsilon = T^{-3/8}$.
There exist a positive constant C and $D_{\gamma_i} = [C \times (T^{2/3})^{p+r}]$ balls with centers $\bar{\boldsymbol{\gamma}}_{i_k}$, $k = 1, ..., D_{\gamma_i}$ and radius $T^{-3/8}/(3\alpha)$ such that the collection of these balls covers $B_{\gamma_i}(\kappa) \equiv \{\boldsymbol{\gamma}_i : \|\boldsymbol{\gamma}_i - \boldsymbol{\gamma}_{i,0}\| \leq T^{-1/4}\}$. Here $[a]$ denotes the maximum integer that does not exceed a. Similarly, because $B_{f_t}(N^{-1/4})$ is compact subset, there exist $D_{f_t} = [C \times (T^{3/8}/N^{1/4})^r]$ balls with centers $\bar{\boldsymbol{f}}_{t,j}$ $(j = 1, ..., D_{f_t})$ and radius $T^{-3/8}/(8\alpha)$ such that the collection of these balls covers $B_{f_t}(N^{-1/4})$.
Because

$$\begin{aligned} &\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}\right| \\ \leq\ &\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\}\right| + \left|\frac{T^{-3/8}}{8\alpha}\cdot\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{M_i + E[M_i]\}\right|, \end{aligned}$$

we have

$$\begin{aligned} &P\left\{ \sup_{\boldsymbol{\gamma}\in B_\Gamma(\delta),\ F\in B_F(e)} \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \boldsymbol{\gamma}_i, \boldsymbol{f}_t)]\}\right| \geq \frac{1}{T^{3/8}}\right\} \\ \leq\ &\sum_{k_1=1}^{D_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{D_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{D_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{D_{f,T}(\varepsilon)} P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it}, \bar{\boldsymbol{\gamma}}_{i_k}, \bar{\boldsymbol{f}}_{t_j})]\}\right| \geq \frac{1}{2T^{3/8}}\right\} \\ &+P\left\{\left|\frac{1}{N}\sum_{i=1}^{N}\{M_i + E[M_i]\}\right| \geq 4\alpha\right\}, \end{aligned}$$

where the second term is zero because $\frac{1}{N}\sum_{i=1}^{N}\{M_i + E[M_i]\} < 2\alpha$.
From the independence property of the idiosyncratic errors over t and the consistency of the estimated parameters,

$$\begin{aligned} & E\left[\left\{\frac{1}{T}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - \frac{1}{T}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\right\}^2\right] \\ = & \frac{1}{T^2}\sum_{t=1}^{T} E\left[\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}^2\right] \\ = & O\left(\frac{1}{T}\times\frac{T^{1/2}+N^{1/2}}{N^{1/2}T^{1/2}}\right), \end{aligned}$$

where we used the result of Lemma 4 such that
$E[\{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}^2] = (T^{1/2}+N^{1/2})/(N^{1/2}T^{1/2})$. Therefore,

$$\sum_{i=1}^{N} E\left\{T^{-1}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}^2 \le C\times\left(\frac{N^{1/2}(T^{1/2}+N^{1/2})}{T^{3/2}}\right),$$

where C is some positive constant. Take
$Z_i = T^{-1}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]$ in Lemma 5, we then have

$$\begin{aligned} & P\left\{\frac{1}{NT}\left|\sum_{i=1}^{N}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \ge \frac{1}{2T^{3/8}}\right\} \\ = & P\left\{\left|\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{t=1}^{T} h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - \frac{1}{T}\sum_{t=1}^{T} E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}\right| \ge \frac{N}{2T^{3/8}}\right\} \\ \le & \exp\left(-C\times\frac{N^2}{T^{3/4}}\times\frac{T^{3/2}}{N^{1/2}(T^{1/2}+N^{1/2})}\right), \end{aligned}$$

where C is some positive constant. This leads

$$\begin{aligned} & \sum_{k_1=1}^{D_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{D_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{D_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{D_{f,T}(\varepsilon)} P\left\{\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\right| \ge \frac{1}{2T^{3/8}}\right\} \\ \le & D_{\gamma_1}(\varepsilon)\times\cdots\times D_{\gamma_N}(\varepsilon)\times D_{f_1}(\varepsilon)\times\cdots\times D_{f_T}(\varepsilon)\times\exp\left(-C\times\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right) \\ = & O\left(\left(T^{2/3}\right)^{N(p+r)}\times\left(\frac{T^{3/8}}{N^{1/4}}\right)^{Tr}\times\exp\left[-\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right]\right) \\ = & O\Bigg(\exp\Bigg[-\left(-\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right)\Bigg\{1-\frac{N(N^{1/2}+T^{1/2})\log(T^{2/3})}{N^{3/2}T^{3/4}} \\ & \qquad -\frac{T(N^{1/2}+T^{1/2})\log(T^{3/8}/N^{1/4})}{N^{3/2}T^{3/4}}\Bigg\}\Bigg]\Bigg) \\ = & O\Bigg(\exp\Bigg[-\left(-\frac{N^{3/2}T^{3/4}}{N^{1/2}+T^{1/2}}\right)\Bigg\{1-\frac{\log(T)}{T^{3/4}}-\frac{\log(T)}{N^{1/2}T^{1/4}}-\frac{T^{1/4}\log(T^{3/8}/N^{1/4})}{N} \end{aligned}$$

$$
- \frac{T^{3/4}\log(T^{3/8}/N^{1/4})}{N^{3/2}}\Big\}\Big]\Big)
$$

$$
= o(1),
$$

where the last line is obtained by using $T^{1/2}/N \to 0$. Thus, the fifth term in (27) is bounded by $O_p(T^{-3/8})$. By repeating the argument that derived $\max_{1\le i\le N} \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(1/T^{1/4})$, we obtain the claim $\max_{1\le i\le N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(T^{-3/8})$. By using $\max_{1\le i\le N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(T^{-3/8})$, we can further strengthen this result to $\max_{1\le i\le N} \|\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}\| = O_p(T^{-7/16})$.

Finally, to obtain the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_i$, we show the fifth term in (27) is $o_p(T^{-1/2})$. It suffices to show that, for any $\eta > 0$,

$$
P\Bigg\{ \sup_{\boldsymbol{\gamma}\in B_\gamma(T^{-7/16}), F\in B_F(N^{-7/16})} \Bigg\| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \{h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h_{it}(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\Bigg\| \ge \frac{\eta}{T^{1/2}}\Bigg\} \quad (35)
$$

converges to zero. Here ζ_{it} and $h(u|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)$ are defined before, $B_\gamma(T^{-7/16})$ and $B_F(N^{-7/16})$ are defined in the proof of Lemma 4. By using the same argument, we have

$$
\begin{aligned}
& P\Bigg\{ \sup_{\boldsymbol{\gamma}\in B_\Gamma(T^{-7/16}),\ F\in B_F(N^{-7/16})} \Bigg| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \{h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\}\Bigg| \ge \frac{\eta}{T^{1/2}}\Bigg\} \\
\le & \sum_{k_1=1}^{E_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{E_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{E_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{E_{f,T}(\varepsilon)} P\Bigg\{ \Bigg| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\Bigg| \ge \frac{\eta}{2T^{1/2}}\Bigg\}
\end{aligned}
$$

and

$$
E\left[\left\{\frac{1}{T}\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t) - \frac{1}{T}\sum_{t=1}^T E[h(\zeta_{it}|\boldsymbol{x}_{it},\boldsymbol{\gamma}_i,\boldsymbol{f}_t)]\right\}^2\right] = O\left(\frac{1}{T}\times\frac{T^{7/8}+N^{7/8}}{N^{7/8}T^{7/8}}\right).
$$

Therefore,

$$
\sum_{i=1}^N E\left\{T^{-1}\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^T E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\right\}^2 \le C\times\left(\frac{N^{1/8}(T^{7/8}+N^{7/8})}{T^{15/8}}\right),
$$

where C is some positive constant. Again, we take Z_i as $Z_i = T^{-1}\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - T^{-1}\sum_{t=1}^T E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]$ in Lemma 5, we then have

$$
\begin{aligned}
& P\Bigg\{\frac{1}{NT}\Bigg|\sum_{i=1}^N\sum_{t=1}^T h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\Bigg| \ge \frac{\eta}{2T^{1/2}}\Bigg\} \\
\le & \exp\left(-C_\eta\cdot\frac{N^2}{T}\cdot\frac{T^{15/8}}{N^{1/8}(T^{7/8}+N^{7/8})}\right),
\end{aligned}
$$

where C_η is some positive constant. This leads

$$
\sum_{k_1=1}^{E_{\gamma_1}(\varepsilon)}\cdots\sum_{k_N=1}^{E_{\gamma_N}(\varepsilon)}\sum_{j_1=1}^{E_{f,1}(\varepsilon)}\cdots\sum_{j_T=1}^{E_{f,T}(\varepsilon)} P\Bigg\{ \Bigg| \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \{h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j}) - E[h(\zeta_{it}|\boldsymbol{x}_{it},\bar{\boldsymbol{\gamma}}_{i_k},\bar{\boldsymbol{f}}_{t_j})]\}\Bigg| \ge \frac{1}{2T^{1/2}}\Bigg\}
$$

$$
\begin{aligned}
&\leq D_{\gamma_1}(\varepsilon) \times \cdots \times D_{\gamma_N}(\varepsilon) \times D_{f_1}(\varepsilon) \times \cdots \times D_{f_T}(\varepsilon) \times \exp\left(-C_\eta \times \frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right) \\
&= O\left(\left(T^{1/16}\right)^{N(p+r)} \times \left(\frac{T^{1/2}}{N^{7/16}}\right)^{Tr} \times \exp\left[-\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right]\right) \\
&= O\Bigg(\exp\Bigg[-\left(\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right)\Bigg\{1-\frac{\{N(p+r)\}(N^{7/8}+T^{7/8})\log(T^{1/16})}{N^{15/8}T^{7/8}} \\
&\qquad -\frac{(Tr)(N^{7/8}+T^{7/8})\log(T^{1/2}/N^{7/16})}{N^{15/8}T^{7/8}}\Bigg\}\Bigg]\Bigg) \\
&= O\Bigg(\exp\Bigg[-\left(\frac{N^{15/8}T^{7/8}}{N^{7/8}+T^{7/8}}\right)\Bigg\{1-\frac{\log(T^{1/16})}{T^{7/8}}-\frac{\log(T^{1/16})}{N^{7/8}}-\frac{T^{1/8}\log(T^{1/2}/N^{7/16})}{N} \\
&\qquad -\frac{T\log(T^{1/2}/N^{7/16})}{N^{15/8}}\Bigg\}\Bigg]\Bigg) \\
&= o(1),
\end{aligned}
$$

where the last line is obtained by using $\sqrt{T}/N^{1-\gamma} \to 0$ for small value of γ $(1/16 < \gamma)$. Thus, the fifth term in (27) is bounded by $o_p(T^{-1/2})$. This completes the first claim of Lemma 4 in (20).

Next, we put (24) into (26)

$$
\begin{aligned}
&\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0} \\
= \; &\Psi_t^{-1}V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0}) - \Psi_t^{-1}\left\{V_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - \overline{V}_{NT,t}(\hat{\boldsymbol{\gamma}}_0, \hat{\boldsymbol{f}}_t) - V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right\} \\
&+\frac{1}{N}\sum_{i=1}^{N}\Psi_t^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}'_{it,0}]\Gamma_i^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{z}_{it,0}\boldsymbol{\lambda}'_{i,0}](\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})\right] \\
&-\frac{1}{N}\sum_{i=1}^{N}\Psi_t^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}'_{it,0}]\Gamma_i^{-1}Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) \\
&+\frac{1}{N}\sum_{i=1}^{N}\Psi_t^{-1}E[g_{it}(0|\boldsymbol{\omega}_{it})\boldsymbol{\lambda}_{i,0}\boldsymbol{z}'_{it,0}]\Gamma_i^{-1}\left\{Q_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\boldsymbol{\gamma}}_i, \hat{F}) - Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right\} \\
&+O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + o_p(\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\|). \qquad (36)
\end{aligned}
$$

By flipping the role of $\boldsymbol{f}_t$ and $\boldsymbol{\gamma}_i$, it is obvious that we can apply the same argument used to obtain (20). Thus, we have $\max_{1\leq t\leq T}\|\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}\| = O_p(\log(T)/N^{1/2})$. This completes the proof of Lemma 4.

D Proof of Theorem 2

We first study the asymptotic distribution of $\hat{\boldsymbol{\gamma}}_i$. Together with the analysis in the proof of Lemma 4, we can obtain the following expression.

$$
\sqrt{T}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) = \Gamma_i^{-1}Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0) - \frac{1}{T}\sum_{t=1}^{T}\Gamma_i^{-1}J_{it}\Psi_t^{-1}\left(\frac{1}{N}\sum_{j=1}^{N}J_{jt}(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j,0})\right) + o_p(1).
$$

This part of analysis is similar to Song (2013) and Ando and Bai (2015). We have

$$\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) = \frac{1}{N}\sqrt{T}K(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) + \sqrt{T}\boldsymbol{\eta} + o_p(1), \tag{37}$$

where $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1, ..., \hat{\boldsymbol{\gamma}}'_N)'$, $\boldsymbol{\gamma}_0 = (\boldsymbol{\gamma}_{1,0}, ..., \boldsymbol{\gamma}'_{N,0})'$, $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, ..., \boldsymbol{\eta}'_N)'$ with $\boldsymbol{\eta}_i = \Gamma_i^{-1} Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)$, and K is defined in (33). Solving (37) in terms of $\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)$, we have

$$\begin{aligned}\sqrt{T}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) &= \left(I - \frac{1}{N}K\right)^{-1}\left(\sqrt{T}\boldsymbol{\eta} + o_p(1)\right), \\ &= \left(I + \frac{1}{N}K + \frac{1}{N^2}K^2 + \frac{1}{N^3}K^3 + \cdots\right)\left(\sqrt{T}\boldsymbol{\eta} + o_p(1)\right).\end{aligned}$$

Note that we can ignore the higher order terms related to $\sqrt{T}N^{-\alpha}K\boldsymbol{\eta}$ as $o_p(1)$ due to the increasing order of $N^{-\alpha}$. Similar technique is also employed in Song (2013) and Ando and Bai (2015). Then, we have

$$\sqrt{T}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0}) = \Gamma_i^{-1}\left(T^{1/2}Q_{NT,i}(\boldsymbol{\gamma}_{i,0}, F_0)\right) + o_p(1).$$

We thus see that the asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})$ is normal with mean zero and variance-covariance matrix Σ_i.
It is obvious that we can again employ the similar argument that employed to derive the asymptotic distribution of $T^{1/2}(\hat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_{i,0})$. We then have

$$N^{1/2}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0}) = \Psi_t^{-1}\left(N^{1/2}V_{NT,t}(\boldsymbol{\gamma}_0, \boldsymbol{f}_{t,0})\right) + o_p(1),$$

which implies that the asymptotic distribution of $N^{1/2}(\hat{\boldsymbol{f}}_t - \boldsymbol{f}_{t,0})$ is normal with mean zero and variance-covariance matrix Θ_t. This completes the proof of Theorem 2.

E Proof of Theorem 3

We prove Theorem 3 by investigating the following two cases. Case 1: $r_0 < r$ and Case 2: $r < r_0$.

Case 1: $r_0 < r$

First consider the case $r_0 < r$ with r_0 being the true number of common factors. Because the number of common factors used in the model, r, is different from the true number of common factors, r_0, we first define the true factor structure for the panel quantile model with the dimension of interactive effects when $r \neq r_0$. Recall that the true quantile function $Q(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0})$ with the true dimension of the interactive effects r_0 is given as

$$Q(\tau|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}) = \boldsymbol{x}'_{it}\boldsymbol{b}_{i,0} + \boldsymbol{f}'_{t,0}\boldsymbol{\lambda}_{i,0},$$

where the dependency of τ is dropped from $\boldsymbol{b}_{i,0,\tau}$, $\boldsymbol{f}_{t,0,\tau}$ and $\boldsymbol{\lambda}_{i,0,\tau}$. The true parameters $\{B_0, \Lambda_0, F_0\}$ are minimizer of the loss function;

$$\begin{aligned}
& \bar{\ell}_{NT}(B,\Lambda,F) \\
\equiv \;& E\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_i-\boldsymbol{f}_t'\boldsymbol{\lambda}_i\right)-\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0}-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)\right] \\
= \;& E\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\boldsymbol{x}_{it}'(\boldsymbol{b}_i-\boldsymbol{b}_{i,0})+(\boldsymbol{f}_t'\boldsymbol{\lambda}_i-\boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})}\left\{G_{it}(s|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})-\tau\right\}ds\right],
\end{aligned}$$

which is zero at $\{B,\Lambda,F\}=\{B_0,\Lambda_0,F_0\}$. The expectation is taken with respect to the true conditional distribution of $\{y_{it}: i=1,...,N,\ t=1,...,T\}$ conditional on X, F_0, B_0 and Λ_0. Similar to Theorem 1 of Bai and Ng (2002), let H_r be an $r_0\times r$ matrix with $\text{rank}(H_r)=\min\{r,r_0\}$ and H_r^+ be the generalized inverse of H_r such that $H_rH_r^+=I_r$. Then, for $r_0<r$, the interactive effects in the true quantile function can be re-expressed as

$$Q(\tau|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})=\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0}+(H_r\boldsymbol{f}_{t,0})'(H_r^+\boldsymbol{\lambda}_{i,0}).$$

It is clear that these transformed true parameters $F_0(r)=(\boldsymbol{f}_{1,0}(r),...,\boldsymbol{f}_{T,0}(r))'$ and $\Lambda_0(r)=(\boldsymbol{\lambda}_{1,0}(r),...,\boldsymbol{\lambda}_{N,0}(r))'$ with $\boldsymbol{f}_{t,0}(r)\equiv H_r\boldsymbol{f}_{t,0}$ and $\boldsymbol{\lambda}_{i,0}(r)\equiv H_r^+\boldsymbol{\lambda}_{i,0}$ together with B_0, will let the loss $\bar{\ell}_{NT}(B,\Lambda,F)$ be zero when $r>r_0$. Therefore, we define $F_0(r)$ and $\Lambda_0(r)$ as the true factor structures when $r>r_0$.

Let $\hat{\boldsymbol{\gamma}}_i(r)=(\hat{\boldsymbol{b}}_i(r),\hat{\boldsymbol{\lambda}}_i(r))$ and $\hat{\boldsymbol{f}}_t(r)$ be the estimated model parameters under the number of common factors being r. Similar to the proofs of Theorem 1 and Theorem 2, we obtain $\max_{1\le i\le N}\|\hat{\boldsymbol{\gamma}}_i(r)-\boldsymbol{\gamma}_{i,0}(r)\|=O_p(\log(N)/\sqrt{T})$ and $\max_{1\le t\le T}\|\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r)\|=O_p(\log(T)/\sqrt{N})$. Here $\boldsymbol{\gamma}_{i,0}(r)=(\boldsymbol{b}_{i,0}',\boldsymbol{\lambda}_{i,0}(r)')'$. Note that $\boldsymbol{b}_{i,0}$ is the true parameter under the true model with the number of factors r_0.

Using Knight's identity $\rho_\tau(u-\nu)-\rho_\tau(u)=-\nu\psi_\tau(u)+\int_0^\nu(I(u\le s)-I(u\le 0))ds$ with $\psi_\tau(u)=\tau-I(u\le 0)$, we express

$$\begin{aligned}
& V(r) \\
\equiv \;& \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\left(y_{it}-\boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_i(r)-\hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)\right) \\
= \;& \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau\Big(\{y_{it}-\boldsymbol{x}_{it}'\boldsymbol{b}_{i,0}-\boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\}-\{\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+\hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\}\Big) \\
= \;& \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau(\varepsilon_{it})+\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+\hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\right)\psi(\varepsilon_{\tau,i}) \\
& +\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+\hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{f}_{t,0}(r)'\boldsymbol{\lambda}_{i,0}(r)\right)}(I(\varepsilon_{\tau,i}\le s)-I(\varepsilon_{\tau,i}\le 0))ds \\
= \;& \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\rho_\tau(\varepsilon_{it})+I_1(r)+I_2(r).
\end{aligned}$$

First, we investigate $I_2(r)$. For a notational simplicity, we denote $\min\{N,T\}$ as C_{NT}.

$$
\begin{aligned}
& (NT)\times I_2(r) \\
= & \sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)}(I(\varepsilon_{\tau,i}\le s)-I(\varepsilon_{\tau,i}\le 0))ds \\
= & \frac{(\log N)^{1/2}}{C_{NT}^{1/2}}\sum_{t=1}^{T}\sum_{i=1}^{N}\int_0^{\sqrt{C_{NT}/\log(N)}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)} \\
& \left(G_{it}\left(\varepsilon_{it}+\frac{s}{\sqrt{C_{NT}/\log N}}\middle|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0}\right)-G_{it}(\varepsilon_{it}|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})\right)ds \\
= & \frac{\log N}{C_{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\Bigg[\int_0^{\sqrt{C_{NT}/\log(N)}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)+\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)} \\
& g_{it}(0|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})sds+o_p(1)\Bigg] \\
= & \frac{\log N}{C_{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}\Bigg[\left(\frac{C_{NT}}{2\log N}\right)g_{it}(0|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})\Big(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})+(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r) \\
& +\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\Big)^2+o_p(1)\Bigg] \\
= & C\times\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})\right)^2+C\times\sum_{t=1}^{T}\sum_{i=1}^{N}\left((\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))'\hat{\boldsymbol{\lambda}}_i(r)\right)^2 \\
& +C\times\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\right)^2 \\
= & O_p(N)+O_p(T),
\end{aligned}
$$

where we used $N^{-1}\sum_{i=1}^{N}\|\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})\|^2=O_p(T^{-1})$,
$N^{-1}\sum_{i=1}^{N}\|\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\|^2=O_p(T^{-1})$, $\sum_{i=1}^{T}\|(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))\|^2/T=O_p(N^{-1})$, and $0<g_{it}(0|\boldsymbol{x}_{it},\boldsymbol{f}_{t,0},\boldsymbol{\lambda}_{i,0})<\infty$ for $i=1,...,N$, $t=1,...,T$.
Next, we evaluate the term $I_1(r)$. Noting that $\sum_{i=1}^{N}\|\boldsymbol{x}_{it}'(\hat{\boldsymbol{b}}_i(r)-\boldsymbol{b}_{i,0})\|^2/N=O_p(1/T)$, $\sum_{i=1}^{N}\|\boldsymbol{f}_{t,0}(r)'(\hat{\boldsymbol{\lambda}}_i(r)-\boldsymbol{\lambda}_{i,0}(r))\|^2/N=O_p(1/T)$ and $\sum_{t=1}^{T}\|(\hat{\boldsymbol{f}}_t(r)-\boldsymbol{f}_{t,0}(r))\|^2/T=O_p(1/N)$, we have

$$(NT)\times I_1(r)\le C\times O_p\left(C_{NT}^{-1/2}\right)\times\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\psi(\varepsilon_{\tau,i})\right)=O_p\left(\sqrt{NT}/C_{NT}^{1/2}\right).$$

Thus, we obtain

$$V(r)-V(r_0)=O_p\left(C_{NT}^{-1}\right).$$

Using the same argument of the proof of Corollary 1 in Bai and Ng (2002), for $r>r_0$, this implies that $V(r)/V(r_0)=1+O_p(1/C_{NT})$. Thus, $\log(V(r)/V(r_0))=O_p(C_{NT}^{-1})$. Because

$(r-r_0)q(N,T) \geq q(N,T)$, which converges to zero slower rate than $O_p(1/C_{NT})$, it follows that

$$P\left(IC(r) - IC(r_0) < 0\right) \leq P\left(O_p\left(1/C_{NT}\right) + q(N,T) < 0\right) \to 0.$$

This indicates that the probability that $IC(r)$ selects the number of common factors $r > r_0$ is asymptotically 0.

Case 2: $r < r_0$

Because $r < r_0$, an $T \times r$ common factor $F(r)$ can not span the true space spanned by the true common factor F_0 with dimension $T \times r_0$. Therefore, regardless of the values of model parameters $\{B(r), \Lambda(r), F(r)\}$ with r dimensional interactive effects, the following loss function can never be zero:

$$\begin{aligned}
&\bar{\ell}_{NT}(B(r), \Lambda(r), F(r)) \\
\equiv\ & \frac{1}{NT} E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_i(r) - \boldsymbol{f}_t(r)'\boldsymbol{\lambda}_i(r)\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\boldsymbol{b}_{i,0} - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0}\right)\right] \\
=\ & \frac{1}{NT} E\left[\sum_{t=1}^{T}\sum_{i=1}^{N} \int_0^{\boldsymbol{x}_{it}'(\boldsymbol{b}_i(r) - \boldsymbol{b}_{i,0}) + (\boldsymbol{f}_t(r)'\boldsymbol{\lambda}_i(r) - \boldsymbol{f}_{t,0}'\boldsymbol{\lambda}_{i,0})} \left\{G_{it}(s|\boldsymbol{x}_{it}, \boldsymbol{f}_{t,0}, \boldsymbol{\lambda}_{i,0}) - \tau\right\} ds\right].
\end{aligned}$$

From the investigation of (11), for some positive constant $C > 0$, not depending on N and T,

$$\lim_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_i(r) - \hat{\boldsymbol{f}}_t(r)'\hat{\boldsymbol{\lambda}}_i(r)\right) - \rho_\tau\left(y_{it} - \boldsymbol{x}_{it}'\hat{\boldsymbol{b}}_i(r_0) - \hat{\boldsymbol{f}}_t(r_0)'\hat{\boldsymbol{\lambda}}_i(r_0)\right)\right] > C,$$

where $\hat{\boldsymbol{b}}_i(r)$, $\hat{\boldsymbol{f}}_t(r)$ and $\hat{\boldsymbol{\lambda}}_i(r)$ are parameter estimates under the dimension of interactive effects r. Using the same argument of the proof of Corollary 1 in Bai and Ng (2002), we therefore have $V(r)/V(r_0) > 1 + c_0$ for some c_0 with large probability for all large N and T. This implies that $\log(V(r)/V(r_0)) \geq c_0'$ for some constant $0 < c_0'$, for large N and T. Because $q(N,T) \to 0$, we have $IC(r) - IC(r_0) > c_0 - (r_0 - r)q(N,T) \geq c_0''$ for some constant $0 < c_0''$, under large N and T with large probability. Thus

$$P\left(IC(r) - IC(r_0) < 0\right) \to 0.$$

This completes the proof of Theorem 3.

F Additional information on the empirical analysis

In Section 6, we analyzed the stock returns of publicly traded firms and firms traded in over-the-counter trading markets for over 6,000 international stocks from over 100 financial markets. The final samples for each period are summarized in Table 3.

To explore the effects of stock exchanges and industries on individual stock returns, Section 6.2.3 applied a clustering approach to the estimated regression coefficients and factor loadings $\{(\hat{\boldsymbol{b}}'_{i,\tau}, \hat{\boldsymbol{\lambda}}'_{i,\tau});\ i = 1, ..., N\}$ to create a set of groups based on the similarities in the sensitivity to the common factors. If the source of the sensitivity to the factors (both observables and unobservables) is solely attributable to stock exchanges, it is expected that the two-way table of the assigned group membership from the clustering approach against the stock exchanges will be diagonal. Note that the industry classifications and listed stock exchanges are known. Therefore, it is easy to create a two-way table of the assigned group membership against these classifications.

First, we investigate the effect of stock exchanges. There are 36 stock exchanges where more than 40 stocks are listed, and we consider these 36 stock exchange markets. Setting the number of clusters as 36, the clustering approach is applied to $\{(\hat{\boldsymbol{b}}'_{i,0.05}, \hat{\boldsymbol{\lambda}}'_{i,0.05});\ i = 1, ..., N\}$, the estimated regression coefficients and factor loadings for the lower tail. The left column in Figure 2 shows the distribution of the firms. An (i, j)-th element denotes the percentage of firms listed on stock exchange i such that they belong to the j-th group. Thus, each row represents the distribution of the firms listed on the same stock exchange. We can make the following observations. First, the degrees of similarity between Shanghai and Shenzhen are stable over the periods in both the upper and lower tails. Second, the New York Stock Exchange and NASDAQ tend to be very similar, while Non-NASDAQ OTC represents its dissimilarity to these two markets. The same observations can be seen from the right column of Figure 2, obtained from the upper tail's factor loadings $\{(\hat{\boldsymbol{b}}'_{i,0.95}, \hat{\boldsymbol{\lambda}}'_{i,0.95});\ i = 1, ..., N\}$. Thus, investors should consider such market characteristics although all three markets are located in the U.S.

A similar approach is applied to determine the effect of industry. To see the effect of industry (5 industries), we also set the number of clusters at 5. Figure 3 shows the distribution of the firms in each industry obtained from $\{(\hat{\boldsymbol{b}}'_{i,0.05}, \hat{\boldsymbol{\lambda}}'_{i,0.05});\ i = 1, ..., N\}$, the estimated regression coefficients and factor loadings for the lower tail. The i-th row represents the distribution of firms in industry i. Specifically, let $n_{i,j}$ denote the number of firms that belong to industry i and to group j. Then, the (i, j)-th element d_{ij} is calculated as $d_{ij} = n_{i,j}/\{\sum_{k=1}^{5} n_{i,k}\}$. Overall, there is one huge cluster that includes most of the firms from each of the 5 industries. Thus, investors regard these 5 industries as similar rather than treating them as very different groups. The same observations can be seen from Figure 3, obtained from the upper tail's $\{(\hat{\boldsymbol{b}}'_{i,0.95}, \hat{\boldsymbol{\lambda}}'_{i,0.95});\ i = 1, ..., N\}$. There seems to be other sources of variability in stock returns in addition to the industry effects.

In summary, the firm industry and the stock exchange on which a firm is listed are important factors to be considered. However, we also note that these nominal factors are insufficient to fully capture the underlying market structures.

G Simulation study

G.1 Performance of the frequentist estimator

To demonstrate the usefulness of the proposed estimation procedure, we conduct a Monte Carlo simulation study. Because the data-generating process and the model parameters are known, we can evaluate the performance of our approach. Here, we report the results for the challenging case in which the variables $x_{it,\tau}$ are correlated with the unobservable factor structures $\boldsymbol{f}_t'\boldsymbol{\lambda}_{i,\tau}$.

G.1.1 Data generating process

For the first data-generating process, we first generate the uniform independent random variable $u_{it} \sim U[0,1]$ for $i = 1, ..., N$ and $t = 1, ..., T$. Then, we generate the data from the following structure:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{b}_{i,u_{it}} + \boldsymbol{f}_{t,u_{it}}'\boldsymbol{\lambda}_{i,u_{it}} + \varepsilon_{it,u_{it}}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T,$$

where $\boldsymbol{x}_{it} = (x_{it,1}, ..., x_{it,p})'$ is a vector of regressors, the dimension of the common factor and the corresponding factor loading depends on the quantile u_{it}, and $\varepsilon_{it,u_{it}} = G^{-1}(u_{it})$. Here, $G(\cdot)$ is a cumulative distribution function of normal or Student-t distribution. The true quantile function of y_{it} at quantile point τ is

$$Q_{y_{it}}(\tau|\boldsymbol{x}_{it}\boldsymbol{f}_t(\tau), \boldsymbol{\lambda}_i(\tau)) = G^{-1}(\tau) + \boldsymbol{x}_{it}'\boldsymbol{b}_i(\tau) + \boldsymbol{f}_t(\tau)'\boldsymbol{\lambda}_i(\tau).$$

The quantile restriction $P(y_{it} - Q_{y_{it}}(\tau|\boldsymbol{\omega}_{it}) \leq 0) = \tau$ is satisfied.
We generate $T \times 5$ common factor matrix $F = (f_{tk})$ such that each element follows the uniform distribution over $[0,2]$. Using the generated u_{it}, we define the common factor for the i-th unit at time t as

$$\boldsymbol{f}_{t,u_{it}} = \begin{cases} (f_{t1}, f_{t2}, f_{t3})' & \text{if } u_{it} \leq 0.2 \\ (f_{t1}, f_{t2}, f_{t3}, f_{t4})' & \text{if } 0.2 < u_{it} \leq 0.8 \\ (f_{t1}, f_{t2}, f_{t3}, f_{t4}, f_{t5})' & \text{if } 0.8 < u_{it} \end{cases}.$$

Note that the dimension of the common factor may vary over i and t because $u_{it} \sim U[0,1]$ for $i = 1, ..., N$ and $t = 1, ..., T$. Given u_{it}, the corresponding factor-loading vector is specified as

$$\boldsymbol{\lambda}_{i,u_{it}} = \begin{cases} (\zeta_{i1} + 0.1u_{it}, \zeta_{i2} + 0.1u_{it}, \zeta_{i3} + 0.1u_{it})' & \text{if } u_{it} \leq 0.2 \\ (\zeta_{i1} + 0.1u_{it}, \zeta_{i2} + 0.1u_{it}, \zeta_{i3} + 0.1u_{it}, \zeta_{i4} + 0.1u_{it})' & \text{if } 0.2 < u_{it} \leq 0.8 \\ (\zeta_{i1} + 0.1u_{it}, \zeta_{i2} + 0.1u_{it}, \zeta_{i3} + 0.1u_{it}, \zeta_{i4} + 0.1u_{it}, \zeta_{i5} + 0.1u_{it})' & \text{if } 0.8 < u_{it} \end{cases},$$

where ζ_{ik} is generated from the uniform distribution over $[0,2]$ and then fixed over t. However, the quantile random variable u_{it} adds some variations over t because the factor loading depends on the quantile points.

Setting $p = 8$ for $i = 1, ..., N$, we generate the set of regressors as follows:

$$x_{it,1} = v_{it,1} + 0.02 f_{t1}^2 + 0.02 \zeta_{i1}^2, \quad x_{it,3} = v_{it,3} - 0.01 f_{t2}^2 + 0.02 \zeta_{i2}^2,$$
$$x_{it,5} = v_{it,5} - 0.01 f_{t3}^2 + 0.03 \zeta_{i3}^2, \quad x_{it,k} = v_{it,k} \quad (k \neq 1, 3, 5).$$

where $v_{it,k}$ is generated from the uniform distribution over $[0, 2]$. The k-th element of true parameter values of regression coefficients $\boldsymbol{b}_{i,0,u_{it}}$ are set to be

$$b_{ik,0,u_{it}} = \begin{cases} 1 + i/N + 0.1 u_{it} & \text{if } k = 2, 4, 6 \\ -1 + i/N + 0.1 u_{it} & \text{if } k \neq 2, 4, 6 \end{cases},$$

Similar to the factor loadings, the quantile random variable u_{it} adds some variations. Finally, cumulative distribution function of $\varepsilon_{it,u_{it}}$ is the normal distribution $N(0, 1)$.
The second data-generating process modifies the first data-generating process. We let the cumulative distribution function of $\varepsilon_{it,u_{it}}$ as the Student-t distribution with degrees of freedom 8. Thus, the error terms have a fat-tail property.

G.1.2 Results

We simulate a large panel with N individuals and T time periods. We consider various combinations of T and N. We base our estimate on the true number of factors and assess the robustness of the proposed strategy to endogeneity. The dimension of the interactive effects is set as its true dimension. For example, the dimension of the interactive effects is set as $r = 3$ when we estimate the $\tau = 5\%$ quantile structure,
The estimation results are averaged over 100 simulated data sets and reported in Table 4 $\sim$ Table 5. Tables show the mean squared error (MSE) between the true structure and the estimates

$$\text{MSE}_1 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2,$$
$$\text{MSE}_2 = \frac{1}{Np} \sum_{i=1}^{N} \|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2,$$

where $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $\boldsymbol{b}_{i,\tau,0}$ and the true τ-th quantile function $Q_{it}(\tau)$, respectively. These measures are computed for the estimators both with and without the factor structure. The panel quantile model without the factor structure is estimated by minimizing the standard loss function
$\ell_\tau(Y|X, B_\tau) = \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - \boldsymbol{x}_{it}' \boldsymbol{b}_{i,\tau})$. Table 4 $\sim$ Table 5 indicate that our estimator with the factor structure performs better than the standard approach. Similar results are also obtained under the second data generating process.

G.2 Performance of the model selection criterion

We investigate the performance of the proposed model selection criterion to select the dimension of the interactive effects. Using the two data generating processes in the previous section, we generate the dataset under the various combinations of N and T. We set the possible dimension of the interactive effects (i.e., the number of common factors) to range from 0 to 8. Calculating the scores of $IC(r)$ over all possible r, we can detect the number of r.

Table 6 reports the histogram of the selected number of common factors $\hat{r}$ over 200 simulation runs. As shown in the tables, the proposed criterion is capable of selecting the true number of factors. When the size of panel N and T increases, the procedure achieves better performance in terms of identifying the true dimension of the interactive effects.

G.3 Estimation under a small panel

Although we developed the asymptotic theory for the frequentist estimator, it was developed under large N and T. In this section, we compare the small sample properties of the frequentist estimator and the proposed Bayesian approach. We use the first data generating process described in Section G.1. We set total number of Markov chain Monte Carlo iterations to 3,000. If one aims to obtain the samples from the posterior distribution, the first iterations are usually discarded to ensure adequate dissipation of initial conditions, or burn-in. However, Geyer (2011) pointed out that Markov chain started anywhere near the center of the posterior distribution needs no burn-in. Because our frequentist estimator corresponds to the Bayesian maximum a posteriori estimator, our initial parameter value is already a good starting point for MCMC. Thus, burn-in period is not considered. We follow Gerlach et al. (2011) by examining trace plots from the MCMC sampler.

Figure 5 (a) shows the MCMC sampling path for the regression coefficient $b_{3,20,\tau}$ with $\tau = 0.05$, $N = T = 100$. We see that the sampling behaviors of each of MCMC sample are already stable from the beginning. Also, the generated posterior sample distributes around the true value of the regression coefficient. Figure 5 (b) compares of the asymptotic distribution of the frequentist estimator (based on Theorem 2) and the posterior density from Bayesian approach. From Theorem 2, we know that the asymptotic distribution of the frequentist estimator is normal distribution. Dashed blue line is the constructed asymptotic distribution from Theorem 2, Solid black line is the constructed posterior density from MCMC output. We see that the constructed posterior density has the wider tail than that from the asymptotic distribution. Thus, the Bayesian posterior credible interval is wider than that of the constructed 95% percent confidence interval of the frequentist estimator. This is commonly known because Bayesian approach takes account parameter uncertainty.

Next, we compare the performance of the following estimators: the frequentist estimator and the Bayesian estimators based on posterior mean, posterior mode and posterior median. We set the length of time series and the number of units as $T = 100, 300, 900$ and $N = 100$, respectively. We note that the similar results are obtained under the different data generating processes described in Section G.1 as well as the different quantile points τ. Figure 6 shows the boxplots of the average mean squared error between the true parameter vectors $\boldsymbol{b}_{i,\tau,0}$ and its estimates $\hat{\boldsymbol{b}}_{i,\tau}$ over i; $\text{MSE} = \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$. These results are obtained based on 200 repetitions at $\tau = 0.05$. Note that similar results are obtained under the different quantile points $\tau = 0.5$ and $\tau = 0.95$, and thus these results are omitted. We can make the following observations. First, MSE decreases as T increases. Second, the Bayesian estimator performs well in the sense that the median of MSEs are smaller than the frequantist estimator when the length of time series T is small. Although the computational time of our proposed Bayesian estimator is slower than the frequentist estimator, it provides better MSE than the others even when the panel size is small. Third, under $T = 900$, the performance of Bayesian estimators (the posterior mean, mode, median) and frequestist estimators became very similar. Because prior is dominated by the pseudo likelihood $L(Y|X, F_\tau, \Lambda_\tau, B_\tau)$, this property can be observed in the estimation results. Thus, Bayesian estimators (the posterior mean, mode, median) and frequestist estimators are asymptotically equivalent as long as the prior information is dominated by the pseudo likelihood $L(Y|X, F_\tau, \Lambda_\tau, B_\tau)$.

References

Ando, T., and Bai, J. (2015). "Asset pricing with a general multifactor structure," *Journal of Financial Econometrics*, **13**, 556–604.

Ando, T. and Bai, J. (2017). "Clustering huge number of time series: A panel data approach with high-dimensional predictors and factor structures," *Journal of the American Statistical Association*, **112**, 1182–1198.

Babu, G.J. (1989). "Strong representations for LAD estimators in linear models," *Probability Theory and Related Fields*, **83**, 547–558.

Bai, J. (1998). "Estimation of multiple-regime regressions with least absolutes deviation," *Journal of Statistical Planning and Inference*, **74**, 103–134.

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

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

Bousquet, O. (2002). “A bennett concentration inequality and its application to suprema of empirical processes,” *Comptes Rendus Mathematique*, **334**, 495-500.

Chen, L., Dolado, J. and Gonzalo, J. (2017). “Quantile Factor Models,” Working Paper.

Geyer, C.J. (2011). “Introduction to Markov Chain Monte Carlo,” In *Handbook of Markov Chain Monte Carlo* (ed. Gelman, A., Brooks, S., Jones, G. and Meng, X.-L.), CRC Press, 3–48.

Gerlach, R, Chen, C.W.S. and Chan, N.Y.C. (2011). “Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets,” *Journal of Business and Economic Statistics*, **29**, 481–492.

Gutenbrunner, C. and Jureckova, J. (1992). “Regression rank scores and regression quantiles,” *Annals of Statistics*, **20**, 305-330.

Hoeffding, W. (1963). “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, **58**, 13-30.

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

Knight, K. (1998). “Limiting distributions for L1 regression estimators under general conditions,” *Annals of Statistics*, **26**, 755-770.

Song, M. (2013). “Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications,” Unpublished manuscript, Department of Economics, Columbia University.

Talagrand, M. (1996). “New concentration inequalities in product spaces,” *Inventiones Mathematicae*, **126**, 505-563.

van der Vaart A. W. (1998). *Weak convergence and empirical processes.* Springer, New York.

Table 3: Distributions of the number of listed financial firms. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015).

	Period 1	Period 2	Period 3
Amman	68	68	67
Australian	99	98	98
Bangkok	86	85	85
Berlin	140	144	146
Borsa Istanbul	46	46	47
BSE Ltd	285	286	277
Copenhagen Stock Exchange	34	34	34
Dhaka	69	68	69
Euronext.liffe Paris	83	83	81
Frankfurt	531	536	529
Hong Kong	203	202	199
Indonesia	76	73	75
Johannesburg	40	39	40
Karachi	56	56	56
Korea Stock Exchange	53	52	53
Kuala Lumpur	99	99	99
Kuwait City	74	72	74
London	144	146	149
Milan	44	44	44
NASDAQ	374	375	370
National India	103	101	103
New York Stock Exchange (NYSE)	186	187	183
Non NASDAQ OTC	1370	1364	1289
OTC Bulletin Board	49	51	52
Philippine Stock Exchange	82	82	81
Santiago	42	40	41
Shanghai	73	73	73
Shenzen	55	55	54
Singapore	49	49	47
SIX Swiss	68	68	68
Stockholm	45	44	44
Stuttgart	63	61	63
Taiwan	61	61	60
Tel Aviv	107	105	106
Thailand	58	59	60
Tokyo Stock Exchange	177	175	177
Toronto	81	80	77
TSX Ventures	64	60	56
XETRA	83	83	82
Others	781	785	780

Table 4: First data-generating process. Comparison of the proposed estimator with factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$, and the standard estimator without factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$. The mean squared errors are defined as $\mathrm{MSE}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2$, $\mathrm{MSE}_2 = \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$, where $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $\boldsymbol{b}_{i,\tau,0}$ and the true τ-th quantile function $Q_{it}(\tau)$, respectively. Averages over 200 simulation data sets are reported. The second lines are corresponding to the standard deviation of the mean squared errors.

			With factor structure		Without factor structure	
τ	T	N	MSE_1	MSE_2	MSE_1	MSE_2
0.05	300	300	0.7140	0.2397	0.9571	0.3046
			0.0524	0.0137	0.0308	0.0123
	300	900	0.7004	0.2389	0.9558	0.3028
			0.0451	0.0088	0.0278	0.0075
	900	300	0.5649	0.0822	0.8701	0.1085
			0.0374	0.0039	0.0338	0.0045
	900	900	0.5063	0.0820	0.8689	0.1089
			0.0130	0.0024	0.0200	0.0027
0.5	300	300	0.3382	0.1279	0.6602	0.1589
			0.0116	0.0053	0.0174	0.0063
	300	900	0.3178	0.1305	0.6567	0.1586
			0.0117	0.0056	0.0139	0.0049
	900	300	0.3054	0.0745	0.6258	0.0828
			0.0090	0.0049	0.0083	0.0053
	900	900	0.2697	0.0766	0.6203	0.0822
			0.0049	0.0022	0.0120	0.0023
0.95	300	300	0.7438	0.2779	1.1611	0.4097
			0.0240	0.0140	0.0483	0.0185
	300	900	0.6892	0.2651	1.1526	0.4063
			0.0220	0.0075	0.0375	0.0105
	900	300	0.6171	0.0904	1.0466	0.1447
			0.0227	0.0041	0.0395	0.0061
	900	900	0.5506	0.0890	1.0272	0.1437
			0.0169	0.0032	0.0266	0.0037

Table 5: Second data-generating process. Comparison of the proposed estimator with factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$, and the standard estimator without factor structure $\boldsymbol{f}'_{t,\tau}\boldsymbol{\lambda}_{i,\tau}$. The mean squared errors are defined as $\text{MSE}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2$, $\text{MSE}_2 = \frac{1}{Np}\sum_{i=1}^{N}\|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$, where $\hat{\boldsymbol{b}}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $\boldsymbol{b}_{i,\tau,0}$ and the true τ-th quantile function $Q_{it}(\tau)$, respectively. Averages over 200 simulation data sets are reported. The second lines are corresponding to the standard deviation of the mean squared errors.

			With factor structure		Without factor structure	
τ	T	N	MSE_1	MSE_2	MSE_1	MSE_2
0.05	300	300	0.9283	0.3532	1.1423	0.4061
			0.0400	0.0140	0.0423	0.0144
	300	900	0.8691	0.3517	1.1262	0.4029
			0.0408	0.0121	0.0225	0.0106
	900	300	0.6723	0.1209	0.9908	0.1445
			0.0330	0.0042	0.0349	0.0061
	900	900	0.6012	0.1217	0.9994	0.1461
			0.0139	0.0028	0.0190	0.0036
0.5	300	300	0.3429	0.1338	0.6480	0.1665
			0.0110	0.0087	0.0191	0.0077
	300	900	0.3283	0.1375	0.6708	0.1681
			0.0091	0.0060	0.0162	0.0052
	900	300	0.3062	0.0735	0.6238	0.0822
			0.0146	0.0041	0.0149	0.0043
	900	900	0.2703	0.0761	0.6253	0.0838
			0.0041	0.0038	0.0118	0.0032
0.95	300	300	0.9736	0.4000	1.3840	0.5130
			0.0403	0.0172	0.0555	0.0208
	300	900	0.8970	0.3878	1.3884	0.5138
			0.0237	0.0081	0.0425	0.0114
	900	300	0.7750	0.1309	1.2211	0.1831
			0.0283	0.0053	0.0439	0.0078
	900	900	0.6833	0.1300	1.2190	0.1810
			0.0217	0.0041	0.0349	0.0055

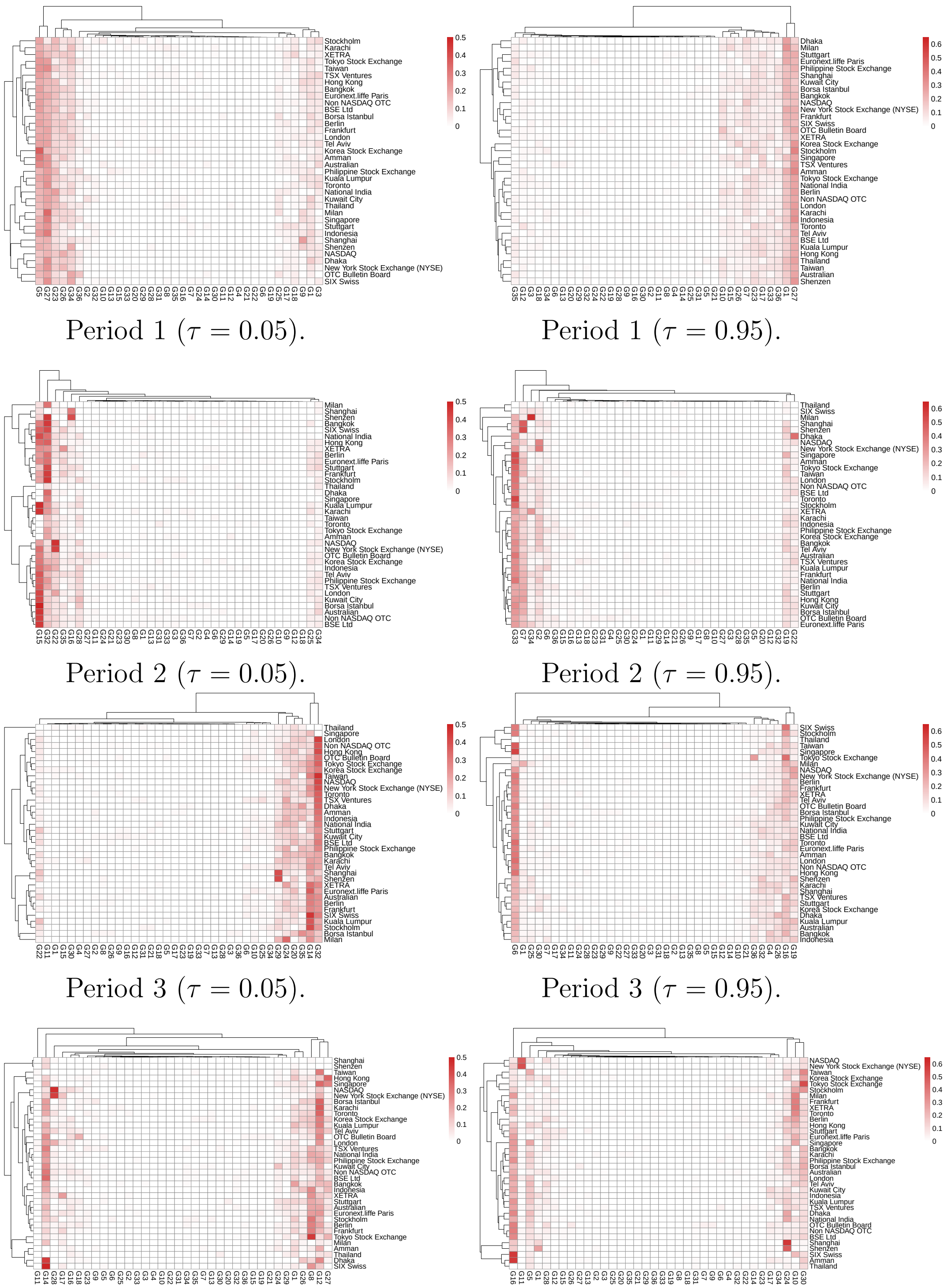
Table 6: The histogram of the selected number of common factors $\hat{r}$ over 100 simulation runs. The results are for $\tau = 5\%$ and $\tau = 95\%$ quantile points. The true number of common factors are $r_{0,\tau} = 3$ at $\tau = 5\%$ and $r_{0,\tau} = 5$ at $\tau = 95\%$.

First data generating process

		$\hat{r}$	0	1	2	3	4	5	6	7	8
τ		T	N								
0.05 ($r_{0,\tau} = 3$)	300	300	0	0	28	41	31	0	0	0	0
	300	900	0	0	0	46	54	0	0	0	0
	900	300	0	0	0	90	10	0	0	0	0
	900	900	0	0	0	96	4	0	0	0	0
0.95 ($r_{0,\tau} = 5$)	300	300	0	0	0	0	2	98	0	0	0
	300	900	0	0	0	0	0	100	0	0	0
	900	300	0	0	0	0	0	100	0	0	0
	900	900	0	0	0	0	0	100	0	0	0

Second data generating process

		$\hat{r}$	0	1	2	3	4	5	6	7	8
τ	T	N									
0.05 ($r_{0,\tau} = 3$)	300	300	0	10	22	38	30	0	0	0	0
	300	900	0	0	14	35	51	0	0	0	0
	900	300	0	0	8	82	10	0	0	0	0
	900	900	0	0	0	95	5	0	0	0	0
0.95 ($r_{0,\tau} = 5$)	300	300	0	0	0	16	29	55	0	0	0
	300	900	0	0	0	0	0	100	0	0	0
	900	300	0	0	0	0	0	100	0	0	0
	900	900	0	0	0	0	0	100	0	0	0



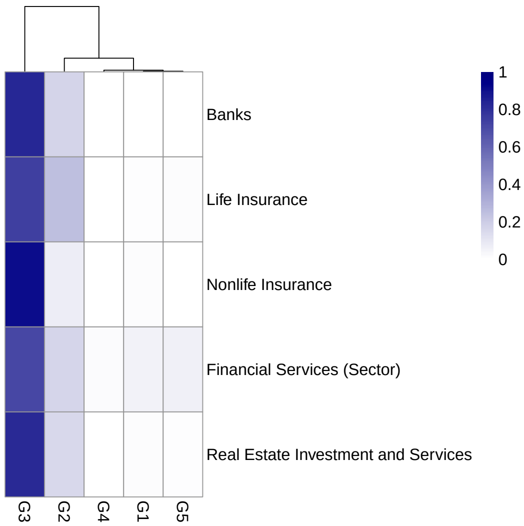
Period 1 ($\tau = 0.05$). Period 1 ($\tau = 0.95$).

Period 2 ($\tau = 0.05$). Period 2 ($\tau = 0.95$).

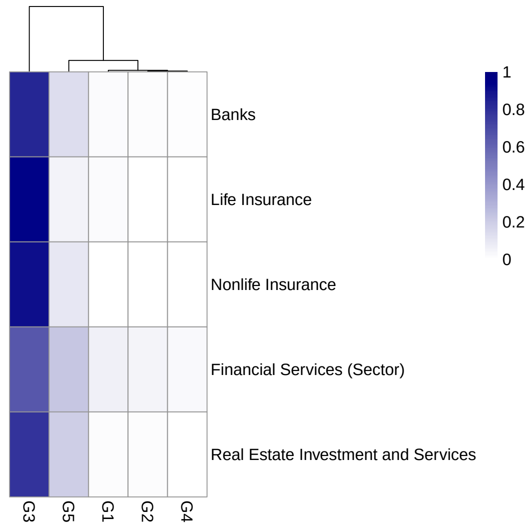
Period 3 ($\tau = 0.05$). Period 3 ($\tau = 0.95$).

Period 4 ($\tau = 0.05$). Period 4 ($\tau = 0.95$).

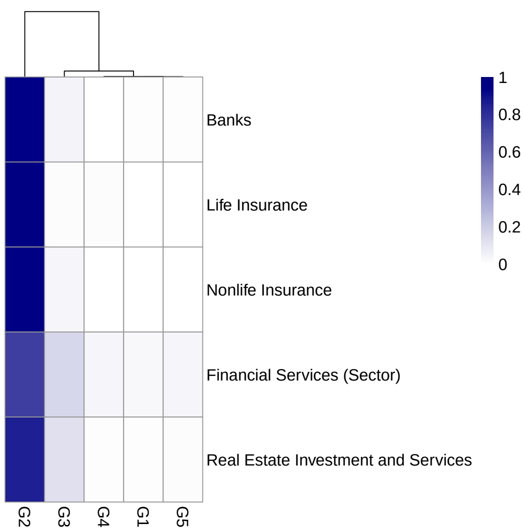
Figure 2: Distribution of firms in each of the stock exchanges (See Section 6.2.3). An (i, j)-th element denotes the percentage of firms in industry i such that they belong to the j-th group. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).



Period 1.



Period 2.



Period 3.

Period 4.

Figure 3: Lower quantile $\tau = 0.05$. Distribution of firms in each of the sectors. An (i, j)-th element denotes the percentage of firms in industry i such that they belong to the j-th group. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

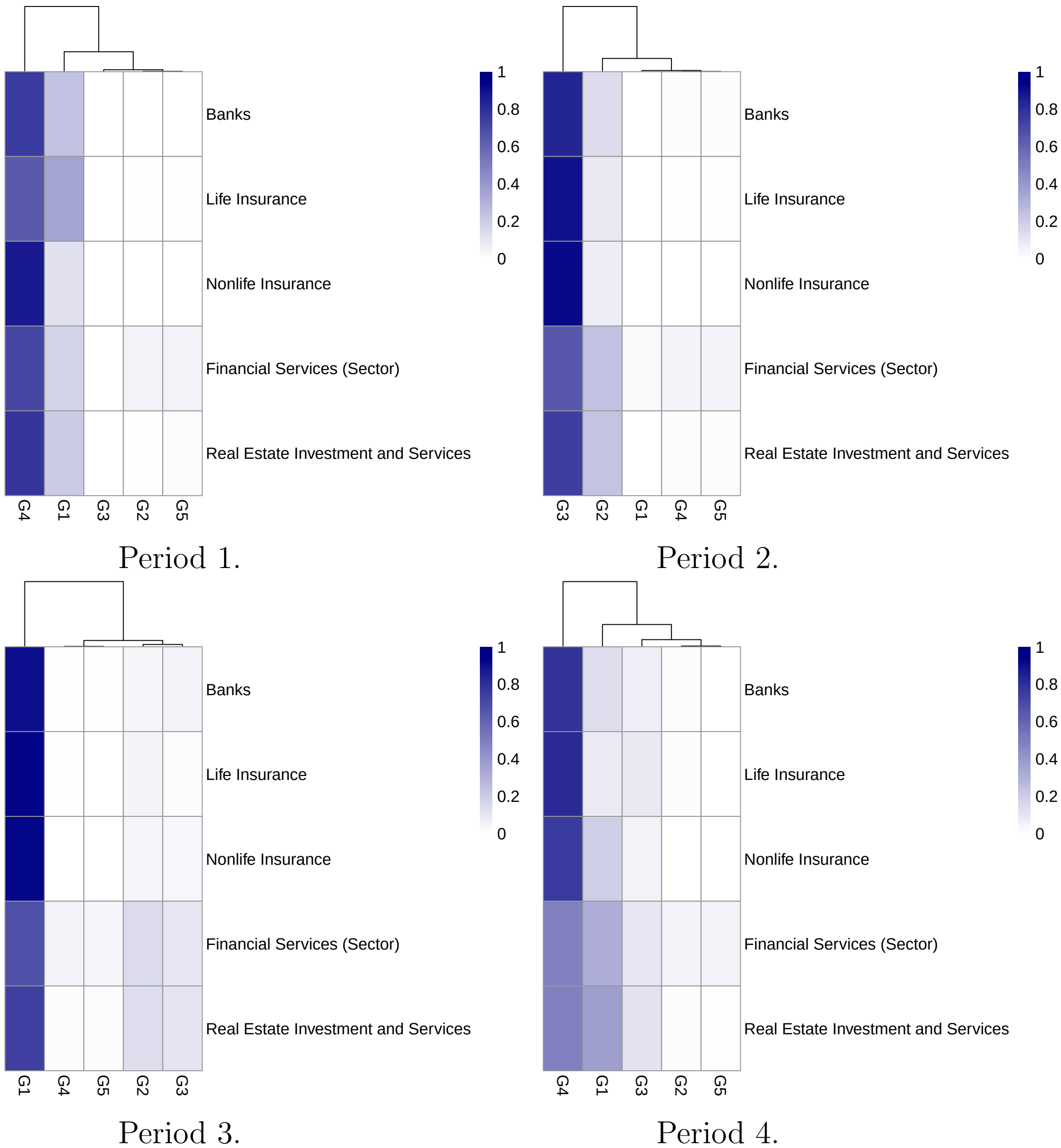
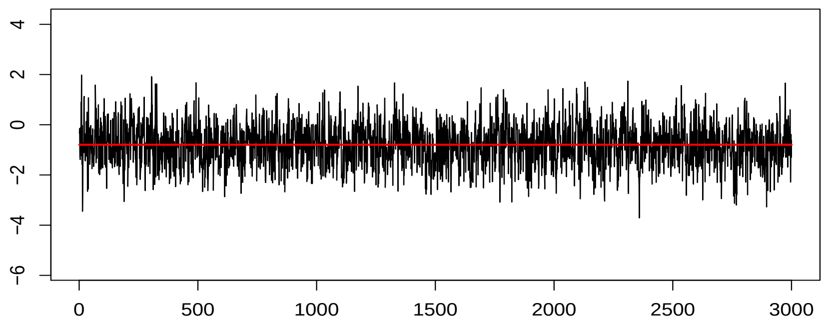
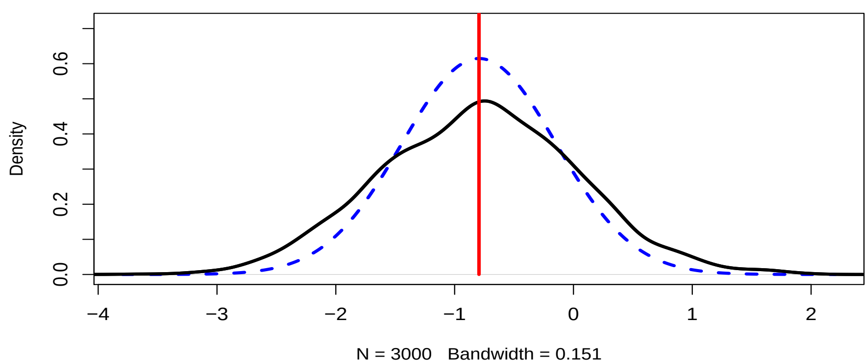


Figure 4: Upper quantile $\tau = 0.95$. Distribution of firms in each of the sectors. An (i, j)-th element denotes the percentage of firms in industry i such that they belong to the j-th group. Period 1 (January 1, 2007, to April 31, 2009); Period 2 (September 1, 2009, to December 31, 2012); Period 3 (January 1, 2013, to March 31, 2015), Period 4 (January 1, 2007, to March 31, 2015).

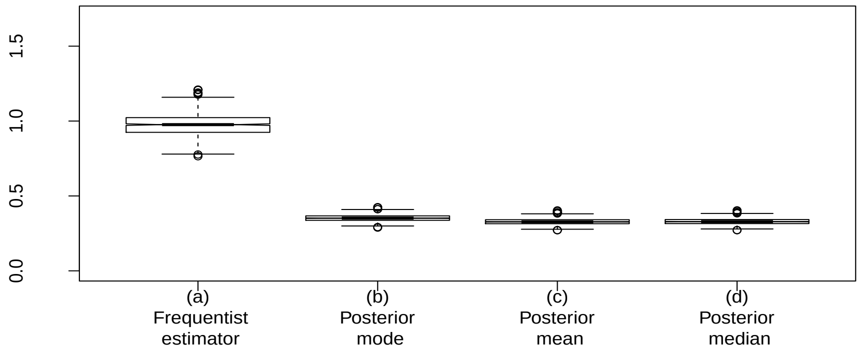


(a): Trance plot

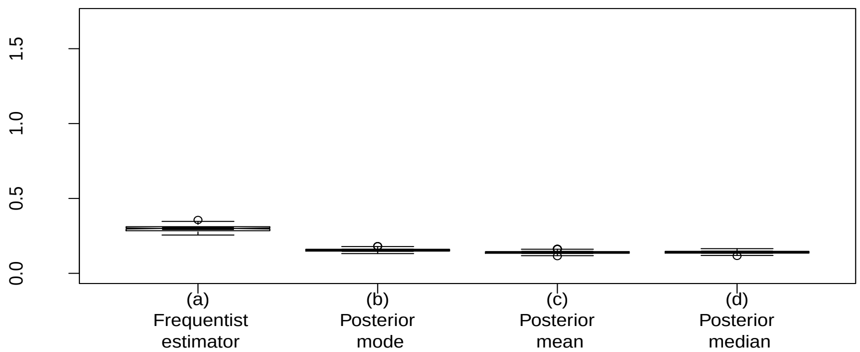


(b): Constructed distribution (from Theorem 2) for the frequestist estimator and posterior distribution

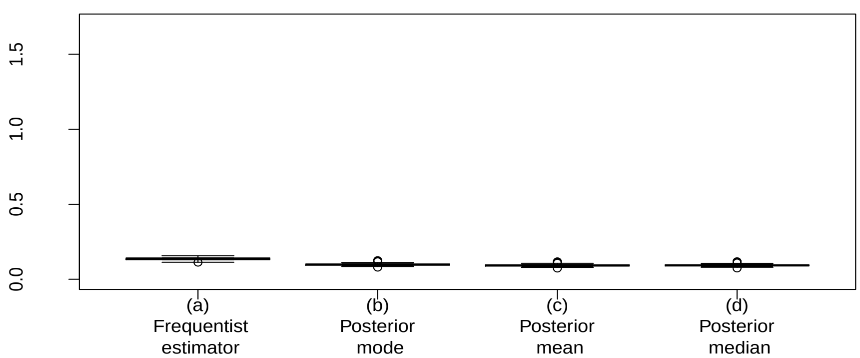
Figure 5: Summary of MCMC posterior sampling results with respect to the regression coefficient of $b_{3,20,\tau}$ at $\tau = 0.05$. A set of 3,000 samples were generated by the proposed data-augmentation algorithm. (a) Black line: trace plot of MCMC sample. Red line: true value of the regression coefficient. (b) Comparison of the constructed distribution (from Theorem 2) for the frequestist estimator and posterior distribution from MCMC for the regression coefficient. Solid black line: the constructed posterior density from MCMC output, Dashed blue line: the constructed asymptotic distribution from Theorem 2, Solid vertical line: true value of the regression coefficient.



(a): $T = 100$



(b): $T = 300$



(c): $T = 900$

Figure 6: Boxplot of the average mean squared errors: MSE $= \frac{1}{Np}\sum_{i=1}^{N} \|\boldsymbol{b}_{i,0,\tau} - \hat{\boldsymbol{b}}_{i,\tau}\|^2$, between the true parameter vector $\boldsymbol{b}_{i,\tau,0}$ and its estimate $\hat{\boldsymbol{b}}_{i,\tau}$. (a) Frequentist estimator denotes our proposed estimator given in Section 3.1 (b) $\sim$ (d) Bayesian estimators based on the proposed data-augmentation strategy in Section 3.2. (e) Without factor structure is based on the standard quantile regression that ignores the factor structures.