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.

1Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.
2Department 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 het-
erogeneous 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), Cap-
piello 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 unobserv-
able 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 distri-
bution 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 (\( i = 1, \ldots, N \)), at time \( t \), its response \( y_{it} \) is observed together with a \((p + 1)\)-dimensional vector of observable regressors \( x_{it} = (1, x_{it,1}, \ldots, x_{it,p})' \). We consider the following model for the \( \tau \)-th conditional quantile function of the response \( y_{it} \);

\[
Q_{y_{it}} (\tau | x_{it}, f_{t,\tau}, \lambda_{i,\tau}) \equiv x_{it}'b_{i,\tau} + f_{t,\tau}'\lambda_{i,\tau}, \quad i = 1, \ldots, N, \ t = 1, \ldots, T, \tag{1}
\]

where \( b_{i,\tau} = (b_{i,0,\tau}, b_{i,1,\tau}, \ldots, b_{i,p,\tau})' \) is a \((p + 1)\)-dimensional vector of regression coefficients, where \( f_{t,\tau} \) is an \( r_{\tau} \times 1 \) vector of unobservable factors and \( \lambda_{i,\tau} \) represents the unobservable factor loadings. For notational simplicity, we often write \( \eta_{it,\tau} = f_{t,\tau}'\lambda_{i,\tau} \). 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 \( \tau \), 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 \( f_{t,\tau} \) to depend on quantile \( \tau \).

Here are some examples. The first example is an interactive effect model in the mean, \( y_{it} = x_{it}'b_{i} + f_{i}'\lambda_{i} + \varepsilon_{it} \), where \( \varepsilon_{it} \) are independent over \( i \), but are i.i.d over \( t \). Let \( q_{i}(\tau) \) be the \( \tau \)-th quantile of \( \varepsilon_{it} \), then,

\[
Q_{y_{it}} (\tau | x_{it}, f_{t}, \lambda_{i}) = x_{it}'b_{i} + f_{i}'\lambda_{i} + q_{i}(\tau).
\]

We can absorb \( q_{i}(\tau) \) into the coefficient of the constant regressor. Next, consider \( y_{it} = x_{it}'b_{i}(u_{it}) + f_{i}'\lambda_{i} \), where \( u_{it} \) are i.i.d \( U(0, 1) \), independent of \( \{x_{it}, f_{t}, \lambda_{i}\} \). Also, assume \( x_{it}'b_{i}(u_{it}) \) is increasing in \( u_{it} \). Then,

\[
Q_{y_{it}} (\tau | x_{it}, f_{t}, \lambda_{i}) = x_{it}'b_{i}(\tau) + f_{i}'\lambda_{i}.
\]
The quantile literature treats $b_i(\tau)$ as a non-random function of $\tau$, and is interested in estimating this function. Finally, consider $y_{it} = x_{it}'b_i(u_{it}) + f_t(u_{it})'\lambda_i(u_{it})$, where $u_{it}$ is $U(0,1)$, independent of $x_{it}$. 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

$$Q_{y_{it}}(\tau | x_{it}, f_t(\tau), \lambda_i(\tau)) = x_{it}'b_i(\tau) + f_t(\tau)'\lambda_i(\tau).$$

Define the error term $\varepsilon_{it,\tau} \equiv y_{it} - x_{it}'b_{i,\tau} - f_{t,\tau}'\lambda_{i,\tau}$, then $P(\varepsilon_{it,\tau} \leq 0 | x_{it}, f_{t,\tau}, \lambda_{i,\tau}) = \tau$. Our purpose is to estimate the unknown conditional quantile function $Q_{y_{it}}(\tau | x_{it}, f_{t,\tau}, \lambda_{i,\tau})$ based on the observations $\{(y_{it}, x_{it}); i = 1, \ldots, N, t = 1, \ldots, T\}$.

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

$$\ell_{\tau}(Y | X, B_{\tau}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau}(y_{it} - x_{it}'b_{i,\tau}),$$

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

However, if the factor structure exists and is ignored, the estimator of $B_{\tau}$ 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 $x_{it}$ (existence of endogeneity). In such a case, we have to estimate the unknown parameters $B_{\tau}, \Lambda_{\tau} = (\lambda_{1,\tau}, \ldots, \lambda_{N,\tau})'$, and $F_{\tau} = (f_{1,\tau}, \ldots, f_{T,\tau})'$ by simultaneously minimizing the following objective function

$$\ell_{\tau}(Y | X, B_{\tau}, F_{\tau}, \Lambda_{\tau}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau}(y_{it} - x_{it}'b_{i,\tau} - f_{t,\tau}'\lambda_{i,\tau})$$

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.
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 $b_{i,\tau} = 0$ 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 $\mathbf{x}'_n b_{i,\tau}$. 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 $\tau$. Thus, it is ideal to formulate model (1) so that unobservable common factors vary with quantile points $\tau$ because its dimension is often unknown for each $\tau$ 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 $b_{i,0,\tau}$. Further, $F_{0,\tau} = (f_{1,0,\tau}, ..., f_{T,0,\tau})'$ and $\Lambda_{0,\tau} = (\lambda_{1,0,\tau}, ..., \lambda_{N,0,\tau})'$ 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 \( F \) be a compact subset of \( \mathbb{R}^r \). The common factors \( f_{t,0,r} \in F \) satisfy \( T^{-1} \sum_{t=1}^{T} f_{t,0,r} f_{t,0,r}' \to \Sigma_F \) as \( T \to \infty \), where \( \Sigma_F \) is an \( r \times r \) positive definite matrix.

Assumption B: Factor loadings and regression coefficients

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

Assumption C: Idiosyncratic error terms

(C1): The random variable \( \varepsilon_{it,r} = y_{it} - x_{it}'b_{i,0,r} - f_{t,0,r}'\lambda_{i,0,r} \) is independently distributed over \( i \) and \( t \), conditional on \( X, F_{0,r} \) and \( \Lambda_{0,r} \). In addition, it satisfies \( E[|\varepsilon_{it,r} - E[\varepsilon_{it,r}]|^K] \leq K! C_\varepsilon^K \) for \( K \geq 1 \) and positive constant \( C_\varepsilon < \infty \).

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

Assumption D: Predictors and design matrix

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

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

\[
0 < C_1 \leq \lambda_{\min}(T^{-1}(X_i, F_{0,r})'(X_i, F_{0,r})) \leq \lambda_{\max}(T^{-1}(X_i, F_{0,r})'(X_i, F_{0,r})) < C_2 < \infty,
\]

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

(D3): Define \( A_{i,r} = \frac{1}{T} X_i'M_{F,r} X_i, B_{i,r} = (\lambda_{i,0,r} \lambda_{0,r}') \otimes I_T, C_{i,r} = \frac{1}{\sqrt{T}} \lambda_{i,0,r} \otimes (X_i'M_{F,r}), M_{F,r} = I - F_{F,r}^{-1} F_{F,r} \). Let \( \mathcal{F}_r \) be the collection of \( F_r \) such that \( \mathcal{F}_r = \{ F_r : F_r F_r/T = I \} \).

We assume

\[
\inf_{F_r \in \mathcal{F}_r} \left[ \frac{1}{N} \sum_{i=1}^{N} E_{i,r}(F_r) \right] \text{ is positive definite,} \tag{4}
\]
where \( E_{i,\tau}(F_{\tau}) = B_{i,\tau} - C_{i,\tau}'A_{i,\tau}^{-1}C_{i,\tau} \).

**Remark 2** The full rank assumption in Assumptions A and B is necessary for the number of common factors to be \( r_\tau \). 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 \( \tau \) is focused rather than the entire quantile function over all possible \( \tau \). 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 \( \tau \), 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 \( x_{it}, f_{t,0,\tau}, \) and \( \lambda_{i,0,\tau} \) in the assumptions.

### 3 Estimation

We begin by presenting the reason for which parameters \( B_{\tau}, F_{\tau}, \) and \( \Lambda_{\tau} \) 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 Korajczcyk (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 $B_\tau$, $F_\tau$ and $\Lambda_\tau$ 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 $F_\tau$ and $\Lambda_\tau$ 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 $F_\tau$ and $\Lambda_\tau$. 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 $\tau$ and the number of common factors $r_\tau$, 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 $\hat{b}_{i,\tau}$ and $\hat{\lambda}_{i,\tau}$ as the minimizer of $\sum_{t=1}^{T} \rho_\tau(y_{it} - x_{it}' \hat{b}_{i,\tau} - \hat{f}_{t,\tau}' \hat{\lambda}_{i,\tau})$ for $i = 1, ..., N$.

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

Step 4. Obtain a QR decomposition of $\hat{F}_\tau$ to yield $\hat{F}_\tau = \hat{Q}^F \hat{R}^F_\tau$, where $\hat{R}^F_\tau$ is an upper triangular matrix with positive diagonal elements and $\hat{Q}^F_\tau$ is an $T \times r_\tau$ orthogonal matrix such that $\hat{Q}^F_\tau' \hat{Q}^F_\tau = I_{r_\tau}$. Then, obtain a QR decomposition of $\hat{R}^F_\tau \hat{\Lambda}'_\tau$, to yield $\hat{R}^F_\tau \hat{\Lambda}'_\tau = \hat{Q}^\Lambda \hat{R}^\Lambda_\tau$. Here, $\hat{R}^\Lambda_\tau$ is an upper triangular matrix with positive diagonal elements, and $\hat{Q}^\Lambda_\tau$ is an $r_\tau \times r_\tau$ orthogonal matrix such that $\hat{Q}^\Lambda_\tau \hat{Q}^\Lambda_\tau = I_{r_\tau}$. 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 ∼ Step 4 until convergence.

In Step 1, the initial values of parameters are prepared as follows. First, estimate $b_{i,\tau}$ ($i = 1, ..., N$) by minimizing $\ell_\tau(Y|X,B_\tau)$ in (2). Given $\hat{b}_{i,\tau}$ ($i = 1, ..., N$), define the variable $Z_\tau = (z_{1,\tau}, ..., z_{N,\tau})$ with $z_{i,\tau} = y_i - X_i \hat{b}_{i,\tau}$. Obtain the principal components’ estimate of $\hat{F}_\tau = (\hat{f}_{1,\tau}, ..., \hat{f}_{T,\tau})'$, subject to the normalization $F_\tau'F_\tau/T = I_{r_\tau}$, which is $\sqrt{T}$ times the eigenvectors corresponding to the $r_\tau$ largest eigenvalues of the $T \times T$ matrix $Z_\tau'Z_\tau$ (See Bai and Ng (2002)). Then, obtain $\hat{\lambda}_{i,\tau}$ as the minimizer of $\sum_{t=1}^T \rho_\tau(y_{it} - x_{it}'\hat{b}_{i,\tau} - \hat{f}_{t,\tau}'\hat{\lambda}_{i,\tau})$ for $i = 1, ..., N$. One can stop the iteration based on $N^{-1}\sum_{i=1}^N \|\hat{b}_{i,\tau}^{\text{new}} - \hat{b}_{i,\tau}^{\text{old}}\|^2 + (NT)^{-1}\sum_{i=1}^N \sum_{t=1}^T \{(\hat{f}_{t,\tau}'\hat{\lambda}_{i,\tau})^{\text{new}} - (\hat{f}_{t,\tau}'\hat{\lambda}_{i,\tau})^{\text{old}}\}^2 < \delta^2$ where $\delta^2$ 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 $\hat{F}_\tau'\hat{F}_\tau/T = (\sqrt{T}\bar{Q}_\tau^F \bar{Q}_\tau^\Lambda)' / (\sqrt{T}\bar{Q}_\tau^F \bar{Q}_\tau^\Lambda) = Q_\tau^N Q_\tau^\Lambda = I_{r_\tau}$. Furthermore, $R_\tau^\Lambda$ 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 \( F_\tau \)

Because the common factor \( F_\tau \) is subject to the normalization condition \( F_\tau F_\tau' / T = I_r \) for identification purposes, \( F_\tau \) 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 \( S_{T,r} \) of radius \( \sqrt{T} \) (see, e.g., Khatri and Mardia (1977), Tsay and Ando (2012) and references therein). Thus, the prior of \( F_\tau \) is a flat prior over the Stiefel manifold

\[
\pi(F_\tau) = \frac{1}{C(T, r_\tau)} \cdot 1(F_\tau \in S_{T,r_\tau}),
\]

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 \( F_\tau \), we use the following equality

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

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

\[
\exp \left( -\{ |y_{it} - x_{it}' b_i,\tau - f_{t,\tau}' \lambda_{i,\tau}| + (2\tau - 1)\{ y_{it} - x_{it}' b_i,\tau - f_{t,\tau}' \lambda_{i,\tau} \} \} \right)
\]

\[
\propto \int_0^\infty \exp \left\{ -\left( \frac{z_{it,\tau} - x_{it}' b_i,\tau - f_{t,\tau}' \lambda_{i,\tau}}{2\omega_{it,\tau}} \right)^2 \right\} \exp\{-2\tau(1-\tau)\omega_{it,\tau}\} d\omega_{it,\tau},
\]

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 $F_\tau$, given $\Omega_\tau \equiv \{\omega_{it,\tau} | i = 1, ..., N, t = 1, ..., T\}$:

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

(7)

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

Because the diagonal matrix $\Omega_{i,\tau}$ prevents the derivation of an analytical conditional posterior of $F_\tau$, further analysis of the conditional posterior of $F_\tau$ in (7) is not straightforward.

To generate the posterior sample of $F_\tau$, we use the Metropolis-Hastings algorithm, which first draws a candidate parameter value $F_{\tau}^{\text{new}}$ from the proposal density $p(F_\tau)$. Then, this generated parameter value $F_{\tau}^{\text{new}}$ is accepted or rejected based on the acceptance probability

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

where $F_{\tau}^{\text{old}}$ is the current state of $F_\tau$.

In the practical implementation of the Metropolis-Hasting algorithm, we draw a new candidate $F_{\tau}^{\text{new}}$ from a proposal density. Here, a Bingham-von Mises-Fisher distribution with parameter $Z^*_{i,\tau}\Omega_{i,\tau}^{-1}\Lambda_{\tau}$ 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 $\varepsilon_{it,\tau} = y_{it} - x'_{it}b_{i,\tau} - \eta_{it,\tau}$ 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 $B_\tau$ and $\Lambda_{\tau}$

Here, we specify the prior densities on $B_\tau$ and $\Lambda_{\tau}$ and derive their conditional posterior distributions, given $F_\tau$ and $\Omega_\tau$. For simplicity of notation, we first express the loss contribution of observation $y_{it}$ as

$$
\exp \left\{ - \{y_{it} - x'_{it}b_{i,\tau} - f'_{t,\tau}\Lambda_{i,\tau} \} + (2\tau - 1) \{y_{it} - x'_{it}b_{i,\tau} - f'_{t,\tau}\Lambda_{i,\tau} \} \right\}
\propto \int_{0}^{\infty} \exp \left\{ - \frac{(z_{it,\tau} - \gamma_{it,\tau}b_{i,\tau} - f'_{t,\tau}\Lambda_{i,\tau})^2}{2\omega_{it,\tau}} \right\} \exp(-2\tau(1 - \tau)\omega_{it,\tau}) d\omega_{it,\tau},
$$

15
where \( \mathbf{v}_{it,\tau} = (\mathbf{x}'_{it}, \mathbf{f}'_{t,\tau})' \), and \( \gamma_{i,\tau} = (b'_{i,\tau}, \mathbf{X}'_{i,\tau})' \).

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

\[
\pi(\gamma_{i,\tau} | Y, X, F_\tau, B_{-i,\tau}, \Lambda_{-i,\tau}, \Omega_{\tau}) \propto \exp \left\{ \frac{1}{2}(z_{i,\tau} - W_{i,\tau} \gamma_{i,\tau})' \Omega_{\tau}^{-1}(z_{i,\tau} - W_{i,\tau} \gamma_{i,\tau}) \right\},
\]

where \( W_{i,\tau} = (X_i, F_\tau) \) is the design matrix, \( B_{-i,\tau} = (b_{1,\tau}, \ldots, b_{i-1,\tau}, b_{i+1,\tau}, \ldots, b_{N,\tau})' \) and \( \Lambda_{-i,\tau} = (\mathbf{\lambda}_{1,\tau}, \ldots, \mathbf{\lambda}_{i-1,\tau}, \mathbf{\lambda}_{i+1,\tau}, \ldots, \mathbf{\lambda}_{N,\tau})' \). This implies that the conditional posterior density of \( \gamma_{i,\tau} \) is the multivariate normal density with mean \( (W_{i,\tau}' \Omega_{i,\tau}^{-1}W_{i,\tau})^{-1}W_{i,\tau}' \Omega_{i,\tau}^{-1}z_{i,\tau} \) and variance-covariance matrix \( (W_{i,\tau}' \Omega_{i,\tau}^{-1}W_{i,\tau})^{-1} \).

Remark 5 When the dimension of \( \mathbf{x}_{it} \) is large, one can also use a shrinkage prior on \( \gamma_{i,\tau} \), 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 \( \gamma_{i,\tau} \) 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

\[
\exp \left\{ -\{ |y_{it} - \mathbf{x}'_{it} b_{i,\tau} - \mathbf{f}'_{t,\tau} \mathbf{\lambda}_{i,\tau} | / 2 + (\tau - 1/2) \{ y_{it} - \mathbf{x}'_{it} b_{i,\tau} - \mathbf{f}'_{t,\tau} \mathbf{\lambda}_{i,\tau} \} \} \right\} \\
\propto \int_0^\infty \exp \left\{ -\left( \frac{a_{i,\tau} - (1 - 2\tau) \omega_{i,\tau}}{2\omega_{i,\tau}} \right)^2 \right\} \exp\{-2\tau(1 - \tau)\omega_{i,\tau}\} d\omega_{i,\tau},
\]

where \( a_{i,\tau} = y_{it} - \mathbf{x}'_{it} b_{i,\tau} - \mathbf{f}'_{t,\tau} \mathbf{\lambda}_{i,\tau} \).

Under the uniform prior \( \omega_{i,\tau} \propto \text{Const.} \), the conditional posterior density of \( \omega_{i,\tau} \) is

\[
\pi(\omega_{i,\tau} | Y, X, F_\tau, B_\tau, \Lambda_\tau, \Omega_{-\omega_{i,\tau}}) \propto \exp \left\{ -\frac{a_{i,\tau}^2}{2\omega_{i,\tau}} - \frac{\omega_{i,\tau}}{2} \right\},
\]
which is the generalized inverse-Gaussian distribution with parameter \((1, a_{it}, 1)\). 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 \(B_\tau\) and \(\Lambda_\tau\) 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 \(B_\tau\) and \(\Lambda_\tau\) analytically without the data augmentation approach. Thanks to the data augmentation approach, we can analytically obtain the conditional posterior distributions of \(B_\tau\), \(\Lambda_\tau\) and \(\Omega_\tau\). Therefore, we can easily draw the posterior samples by implementing the Gibbs sampling algorithm. To draw \(F_\tau\), we can use the Metropolis-Hastings algorithm.

We now summarize our data-augmentation strategy (given values of \(\tau\) and the number of common factors \(r_\tau\)) as follows.

**Posterior sampling algorithm:**

1. **Step 1.** Initialize the parameters.
2. **Step 2.** Sample \(F_\tau\) from \(\pi(F_\tau|Y, X, B_\tau, \Lambda_\tau, \Omega_\tau)\).
3. **Step 3.** Sample \(\gamma_{i,\tau}\) from \(\pi(\gamma_{i,\tau}|Y, X, F_\tau, B_{-i,\tau}, \Lambda_{-i,\tau}, \Omega_\tau)\) for \(i = 1, \ldots, N\).
4. **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 \(i = 1, \ldots, N\) and \(t = 1, \ldots, T\).
5. **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 \( \{B^{(k)}_\tau, F^{(k)}_\tau, \Lambda^{(k)}_\tau; k = 1, \ldots, H\} \), which can be employed for conducting Bayesian analysis numerically. For example, the Bayesian maximum a posteriori is approximately given as

\[
\{\tilde{B}_\tau, \tilde{F}_\tau, \tilde{\Lambda}_\tau\} = \arg\max_{\{B^{(k)}_\tau, F^{(k)}_\tau, \Lambda^{(k)}_\tau; k = 1, \ldots, H\}} L(Y|X, B^{(k)}_\tau, F^{(k)}_\tau, \Lambda^{(k)}_\tau) \pi(B^{(k)}_\tau, F^{(k)}_\tau, \Lambda^{(k)}_\tau),
\]

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 $F_\tau$ and $\Lambda_\tau$ are accommodated in the prior distribution of $F_\tau$ and $\Lambda_\tau$. Recall, also, that the diffuse prior was used for $\gamma_\tau$. 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 $b_{i,\tau}$), 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, $r_\tau$. 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

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

(9)

where $\hat{b}_{i,\tau}(r), \hat{\lambda}_{i,\tau}(r)$ and $\hat{f}_{t,\tau}(r)$ 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
\[ q(N,T) = \log \left( \frac{NT}{N+T} \right) \left( \frac{N+T}{NT} \right). \] (10)

The asymptotic performance of \( IC_r(r) \) 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 \( \hat{F}_r, \hat{B}_r \) and \( \hat{\Lambda}_r \). Here, the true parameter value \( \{F_{0,r}, B_{0,r}, \Lambda_{0,r}\} \) is defined as the minimizer of the expected quantile loss function
\[ \ell_{0,r}(Y|X, B_r, F_r, \Lambda_r) = E \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_r(y_{it} - x'_{it}b_{i,\tau} - f'_{t,\tau}\lambda_{i,\tau}) \right] \] (11)
subject to the identification restriction on \( F_r \) and \( \Lambda_r \). Here, the expectation is taken with respect to the true conditional distribution of \( \{y_{it} : i = 1, ..., N, \ t = 1, ..., T\} \) conditional on \( X, F_{0,r} \) and \( \Lambda_{0,r} \). The following proposition provides the average consistency of \( \hat{\gamma}_{i,\tau} = (\hat{b}'_{i,\tau}, \hat{\lambda}'_{i,\tau})' \) and \( \hat{f}_{t,\tau} \). In general, \( \hat{f}_{t,\tau} \) and \( \hat{\lambda}_{i,\tau} \) 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:

\[
N^{-1} \sum_{i=1}^{N} \| \hat{\gamma}_{i,\tau} - \gamma_{i,0,\tau} \|^2 = o_p(1),
\]
\[
T^{-1} \| \hat{F}_r - F_{0,r} \|^2 = T^{-1} \sum_{t=1}^{T} \| \hat{f}_{t,\tau} - f_{t,0,\tau} \|^2 = o_p(1),
\]

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

The proof of Proposition 1 is given in an online supplementary document. The result \( \| \hat{F}_r - F_{0,r} \|/T^{1/2} = o_p(1) \) implies that the space spanned by \( F_{0,r} \) and the space spanned by the estimated factors \( \hat{F}_r \) are asymptotically the same. We also prove the consistency of the estimators in the sense that \( \hat{b}_{i,\tau} \) and \( \hat{\lambda}_{i,\tau} \) converge in probability to \( b_{i,0,\tau}, \lambda_{i,0,\tau} \) and uniformly over \( 1 \leq i \leq N \). In addition, \( \hat{f}_{t,\tau} \) converges in probability to \( f_{t,0,\tau} \) uniformly over \( 1 \leq t \leq T \).
Theorem 1 Suppose Assumptions A–D, \( \log(T)/N^{1/2} \to 0 \) and \( \log(N)/T^{1/2} \to 0 \) hold. Then, \( \hat{b}_{i,\tau} \) and \( \hat{\lambda}_{i,\tau} \) are consistent

\[
\max_{1 \leq i \leq N} \| \hat{b}_{i,\tau} - b_{i,0,\tau} \| = o_p(1),
\]

(12)

\[
\max_{1 \leq i \leq N} \| \hat{\lambda}_{i,\tau} - \lambda_{i,0,\tau} \| = o_p(1).
\]

(13)

Moreover, the estimated common factor is consistent

\[
\max_{1 \leq t \leq T} \| \hat{f}_{t,\tau} - f_{t,0,\tau} \| = o_p(1).
\]

(14)

Theorem 2 shows that the asymptotic distribution of the estimated regression coefficients, \( \hat{\gamma}_{i,\tau} = (\hat{b}_{i,\tau}', \hat{\lambda}_{i,\tau}')' \), is a multivariate normal distribution. Similarly, the asymptotic distribution of the estimated common factor \( \hat{f}_{t,\tau} \) 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 \( \gamma \) satisfying \( 1/16 < \gamma \). Then, the asymptotic distribution of \( T^{1/2}(\hat{\gamma}_{i,\tau} - \gamma_{i,0,\tau}) \) is normal with mean zero and variance-covariance matrix

\[
\Sigma_{i,\tau} = \tau(1-\tau)\Gamma_{i,0,\tau}^{-1}V_{i,0,\tau}\Gamma_{i,0,\tau}^{-1}.
\]

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

\[
\Theta_{t,\tau} = \tau(1-\tau)\Psi_{t,0,\tau}^{-1}R_{0,\tau}\Psi_{t,0,\tau}^{-1}.
\]

Here, \( \Psi_{t,0,\tau} \) and \( R_{0,\tau} \) are given as \( \Psi_{t,0,\tau} \equiv \text{plim}_{N \to \infty} N^{-1} \sum_{i=1}^{N} g_{it}(0|x_{it}, f_{t,0,\tau}, \lambda_{i,0,\tau})\lambda_{i,0,\tau}\lambda_{i,0,\tau}' \) and \( R_{0,\tau} \equiv \text{plim}_{N \to \infty} N^{-1} \sum_{i=1}^{N} \lambda_{i,0,\tau}\lambda_{i,0,\tau}' \), 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 transferred 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

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

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:

\[
Q_{yt} (\tau | \mathbf{x}_{it}) = \alpha_{i, \tau} + Mkt_t \times \beta_{Mkt,i, \tau} + HML_t \times \beta_{HML,i, \tau} + SMB_t \times \beta_{SMB,i, \tau} \\
+RMW_t \times \beta_{RMW,i, \tau} + CMA_t \times \beta_{CMA,i, \tau} + \sum_{k=1}^{11} LMkt_{t,k} \times \beta_{LMkt,k,i, \tau} + \mathbf{f}'_{t, \tau} \lambda_{i, \tau},
\]

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 \( \{LMkt_{t,k}, \ k = 1, \ldots, 11\} \) 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 $r_\tau$, which is determined by the $IC(r_\tau)$ score. Setting the dimension of the common factor for the mean to be identical to $r_\tau$, we estimate the following asset pricing model with a common factor structure

$$
y_{it} = \alpha_i + Mkt_t \times \beta_{Mkt,i} + HML_t \times \beta_{HML,i} + SMB_t \times \beta_{SMB,i} + RMW_t \times \beta_{RMW,i} + \sum_{k=1}^{11} LMkt_{t,k} \times \beta_{LMkt,k,i} + f'\lambda_i + \epsilon_{it},
$$

by minimizing the least-squares objective function $\ell(B, F, \Lambda) = \sum_{i=1}^{N} \|y_i - X_i\beta_i - F\lambda_i\|^2$ subject to the constraint $F'F/T = I_r$. Numerical optimization can be achieved by the
iterative optimization of $B$, $F$, $\Lambda$ 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 $\{ (\hat{b}_{i,\tau}, \hat{\lambda}_{i,\tau}^\prime); \ i = 1, \ldots, N \}$ 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 $\tau$-th quantile; we then run the following regression $f_{jt,\tau} = z_t'\gamma_j + e_{jt}$, where $z_t$ is the 20-dimensional vector of Fama and French’s factors. Then, we conduct statistical significance tests of the least squares estimate $\hat{\gamma}_j$.

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: $f_{jt,\tau} = z_t'\gamma_j + e_{jt}$ 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 $r_\tau$ 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**


Table 1: Estimated number of common factors $r_{\tau}$ 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).

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>0.95</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 2: The result of canonical correlation analysis between the common factors for quantile $\tau$ 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).

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>6/7</td>
<td>5/6</td>
<td>3/3</td>
<td>10/10</td>
</tr>
<tr>
<td>0.95</td>
<td>1/2</td>
<td>2/2</td>
<td>4/4</td>
<td>7/8</td>
</tr>
</tbody>
</table>

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 ε_{it,τ} = ε_{it}, b_{i,τ} = b_i, λ_{i,τ} = λ_i, γ_{i,τ} = (b_i', λ_i')', f_{t,τ} = f_t, B_{τ} = B, Λ_{τ} = Λ, F_{τ} = F, etc throughout the proof. Also, we denote the true parameters as b_{i,0}, λ_{i,0}, γ_{i,0}, 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 X₁,...,Xₙ 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,

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

for some positive constant C.

---

⁴Melbourne Business School, Melbourne University, T.Ando@mbs.edu. 200 Leicester Street, Carlton, Victoria 3053, Australia.
⁵Department 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:

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

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

### A.1 Proof of Proposition 1

We first show

$$\sup_{f_i \in F, \lambda_i \in \mathcal{L}, b_i \in B} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - x_i' b_i - f_i' \lambda_i) - E[\rho_r(y_{it} - x_i' b_i - f_i' \lambda_i)] \right\} \right| = o_p(1), \quad (1)$$

where $E[\cdot]$ is the expectation of $y_{it}$ conditioned on $X$, $F_0$, $\Lambda_0$ and $B_0$. For any $e > 0$, $F$, $\mathcal{L}$, and $B$ are covered by $\bigcup_{f \in F} B_e(f)$, $\bigcup_{\lambda \in \mathcal{L}} B_e(\lambda)$ and $\bigcup_{b \in B} B_e(b)$, respectively. Here $B_e(z)$ 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 $C_F, C_\Lambda, C_B$, and $\{ f_1, \ldots, f_{C_F} \} \in F$, $\{ \lambda_1, \ldots, \lambda_{C_\Lambda} \} \in \mathcal{L}$, $\{ b_1, \ldots, b_{C_B} \} \in B$ such that

$$F \subset \bigcup_{k=1}^{C_F} B_e(f_k), \quad \mathcal{L} \subset \bigcup_{k=1}^{C_\Lambda} B_e(\lambda_k), \quad B \subset \bigcup_{k=1}^{C_B} B_e(b_k).$$

Define

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

The set $\mathcal{J}$ is covered by $\bigcup_{k=1}^{C_B} \bigcup_{u=1}^{C_F} \bigcup_{v=1}^{C_\Lambda} H(b_k, f_u, \lambda_v)$, where

$$H(b_k, f_u, \lambda_v) = \{ x' b + f' \lambda : b \in B_e(b_k), f \in B_e(f_u), \lambda \in B_e(\lambda_v) \}.$$

Suppose $x' b + f' \lambda \in H(b_k, f_u, \lambda_v)$. We then have

$$|x' b + f' \lambda - (x' b_k + f'_u \lambda_v)|$$

$$\leq \| x' b - x' b_k \| + \| f' \lambda - f'_u \lambda_v \|$$

$$\leq \| x \| \times \| b - b_k \| + \| f - f_u \| \times \| \lambda - \lambda_v \| + \| f_u \| \times \| \lambda - \lambda_v \| + \| f - f_u \| \times \| \lambda_v \|$$

$$\leq eM + e^2 + eM + eM$$

$$\equiv \delta,$$

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

$$\mathcal{J} \subset \bigcup_{k=1}^{C_B} \bigcup_{u=1}^{C_F} \bigcup_{v=1}^{C_\Lambda} B_\delta(x' b_k + f'_u \lambda_v),$$

37
which implies that, for any $\delta > 0$, there exist finite integers $C_B, C_F$ and $C_A$ such that $\mathcal{J}$ is covered by the union of $C_B \times C_F \times C_A$ closed balls $B_\delta(x'b_1 + f'_1\lambda_1), ..., B_\delta(x'b_{C_B} + f'_{C_F}\lambda_{C_A})$.

For each $x_i, f_i, \lambda_i$ and $b_i$, we can identify $\bar{f}_u, \bar{\lambda}_u$ and $\bar{b}_k$, such that $f_i \in B_{\epsilon}(\bar{f}_u)$, $\lambda_i \in B_{\epsilon}(\lambda_u)$ and $b_i \in B_{\epsilon}(b_k)$. Define $h_{it} = x'_itb_i + f'_i\lambda_i$ and

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

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

$$
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - x'_itb_i - f'_i\lambda_i) - E[\rho_r(y_{it} - x'_itb_i - f'_i\lambda_i)] \right\} 
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - h_{it}) - E[\rho_r(y_{it} - h_{it})] \right\} 
\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - h_{it}) - \rho_r(y_{it} - d_{it}) \right\} 
+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ E[\rho_r(y_{it} - h_{it})] - E[\rho_r(y_{it} - d_{it})] \right\} 
+ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - d_{it}) - E[\rho_r(y_{it} - d_{it})] \right\} 
= I_1 + I_2 + I_3.
$$

Now,

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

where $s_{\tau}(y, \delta) = \sup_{a,b \in J, |a-b| \leq \delta} |\rho_r(y-a) - \rho_r(y-b)|$. The first term in the last line is $o_p(1)$ by a law of large numbers. By choosing a small enough $\delta$, the second term can be made arbitrarily small. Similarly, $I_2$ can be made arbitrarily small by choosing a small enough $\delta$.

Finally, we consider $I_3$. Note that $d_{it} = d_{it,k,u,v}$, $I_3$ is uniformly bounded by

$$\sup_{k,u,v} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - d_{it,k,u,v}) - E[\rho_r(y_{it} - d_{it,k,u,v})] \right\} 
= \max_{d_j \in S} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_r(y_{it} - d_j) - E[\rho_r(y_{it} - d_j)] \right\},$$

where

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

38
Note that there are some abuse of notations, but the idea should be clear. The maximum of number of elements of $\mathcal{S}$ is $C_B^N C_F^T C_A^N$ when $k, u, v$ vary.

Note that

$$\sup_{h \in J} |\rho_r(y_{it} - h) - E[\rho_r(y_{it} - h)]|$$

$$= \sup_{h \in J} |\rho_r(x_{it}'b_i + f'_i \lambda_i + \varepsilon_{it} - h) - E[\rho_r(x_{it}'b_i + f'_i \lambda_i + \varepsilon_{it} - h)]|$$

$$\leq \sup_{h \in \mathbb{R}} |\rho_r(\varepsilon_{it} - h) - E[\rho_r(\varepsilon_{it} - h)]|$$

$$\leq |\varepsilon_{it} - E[\varepsilon_{it}]| + E[|\varepsilon_{it} - E[\varepsilon_{it}]|],$$

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

$$E[\rho_r(y_{it} - d_j) - E[\rho_r(y_{it} - d_j)]|^K$$

$$\leq E\left[|\varepsilon_{it} - E[\varepsilon_{it}]| + E[|\varepsilon_{it} - E[\varepsilon_{it}]|]\right]^K$$

$$\leq 2^{K-1}\left(E|\varepsilon_{it} - E[\varepsilon_{it}]|^K + \left(E|\varepsilon_{it} - E[\varepsilon_{it}]|\right)^K\right).$$

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

$$E\left[|\rho_r(y_{it} - d_j) - E[\rho_r(y_{it} - d_j)]|^K\right] \leq K!C_1^K C_2^{K-2} 2^K$$

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

$$Z_{it} \equiv \rho_r(y_{it} - d_j) - E[\rho_r(y_{it} - d_j)].$$

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

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

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

$$E\left[\max_{j \in \mathcal{S}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{\rho_r(y_{it} - d_j) - E[\rho_r(y_{it} - d_j)]\}\right]$$

$$\leq C \times \frac{1}{NT} \left(C_1 \log(1 + |\mathcal{S}|) + \sqrt{(NT) C_2 \log(1 + |\mathcal{S}|)}\right)^{1/2}$$

$$\leq C' \times \frac{1}{NT} \left[(2N + T) \log C_{\max} + (NT)^{1/2}(2N + T)^{1/2} \log C_{\max}\right]$$

$$= O\left(\frac{1}{N^{1/2}} + \frac{1}{T^{1/2}}\right),$$

39
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

$$
\ell_r(B, F, \Lambda) = \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_r(y_{it} - x_{it}^\prime b_i - f_{t,0}^\prime \lambda_i) .
$$

Then, for any given $F$, $\Lambda$ and $B$, we have

\[
U_r (B, F, \Lambda) = \frac{1}{NT} \ell_r (B, F, \Lambda) - \frac{1}{NT} \ell_r (B_0, F_0, \Lambda_0) \\
= \left[ \frac{1}{NT} \ell_r (B, F, \Lambda) - \frac{1}{NT} E[\ell_r (B, F, \Lambda)] \right] + \left[ \frac{1}{NT} \ell_r (B_0, F_0, \Lambda_0) - \frac{1}{NT} E[\ell_r (B_0, F_0, \Lambda_0)] \right] \\
+ \frac{1}{NT} E[\ell_r (B, F, \Lambda)] - \frac{1}{NT} E[\ell_r (B_0, F_0, \Lambda_0)] \\
= J_1 + J_2 + J_3,
\]

where $E[\ell_r (B, F, \Lambda)]$ is defined as $E[\ell_r (B, F, \Lambda)] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E[\rho_r(y_{it} - x_{it} b_i - f_{t,0}^\prime \lambda_i)]$.

From (1), we have

$$
\sup_{f_t \in F, \lambda_i \in C : b_i \in B} |J_1| = o_p(1), \quad \text{and} \quad |J_2| = o_p(1).
$$

About $J_3$, note that $\varepsilon_{it} = y_{it} - x_{it}^\prime b_{i,0} - f_{t,0}^\prime \lambda_{i,0}$, with conditional density $g_{it}(\cdot | x_{it}, f_{t,0}, \lambda_{i,0})$.

By Taylor expansion, we obtain

$$
E[\rho_r(y_{it} - x_{it}^\prime b_i - f_{t,0}^\prime \lambda_i)] - E[\rho_r(y_{it} - x_{it}^\prime b_{i,0} - f_{t,0}^\prime \lambda_{i,0})] \\
= E[\rho_r(\varepsilon_{it} - \{x_{it}^\prime b_i + f_{t,0}^\prime \lambda_i - x_{it}^\prime b_{i,0} - f_{t,0}^\prime \lambda_{i,0}\})] - E[\rho_r(\varepsilon_{it})] \\
= g_{it}(x_{it}^\prime \tilde{b}_i + f_{t,0}^\prime \tilde{\lambda}_i | x_{it}, f_{t,0}, \lambda_{i,0}) \times (x_{it}^\prime (b_{i,0} - b_i) + f_{t,0}^\prime \lambda_{i,0} - f_{t,0}^\prime \lambda_i)^2 \\
\geq \tilde{g} \times (x_{it}^\prime (b_{i,0} - b_i) + f_{t,0}^\prime \lambda_{i,0} - f_{t,0}^\prime \lambda_i)^2,
$$

where $x_{it}^\prime \tilde{b}_i + f_{t,0}^\prime \tilde{\lambda}_i$ is between 0 and $x_{it}^\prime (b_{i,0} - b_i) + f_{t,0}^\prime \lambda_{i,0} - f_{t,0}^\prime \lambda_i$. By assumption, $x_{it}^\prime \tilde{b}_i + f_{t,0}^\prime \tilde{\lambda}_i$ belongs to a compact set. This leads to $0 < \tilde{g} \leq g_{it}(x_{it}^\prime \tilde{b}_i + f_{t,0}^\prime \tilde{\lambda}_i | x_{it}, f_{t,0}, \lambda_{i,0})$ by Assumption C. Therefore, the last inequality was obtained. Then,

$$
J_3 = \frac{1}{NT} E[\ell_r (B, F, \Lambda)] - \frac{1}{NT} E[\ell_r (B_0, F_0, \Lambda_0)] \\
\geq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{g} \times (x_{it}^\prime (b_{i,0} - b_i) + f_{t,0}^\prime \lambda_{i,0} - f_{t,0}^\prime \lambda_i)^2 .
$$

Without loss of generality, we assume that $b_{i,0} = 0, i = 1, \ldots, N$ (for notational simplicity).

Note that the centered objective function satisfies

$$
U_r (B_0, F_0, \Lambda_0) = 0,
$$

where we used that the function $h(b_i, f_t, \lambda_i) \equiv x_{it}^\prime b_i + f_{t,0}^\prime \lambda_i - f_{t,0}^\prime \lambda_{i,0}$ evaluated at $b_i = b_{i,0}, f_t = f_{t,0}$ and $\lambda_i = \lambda_{i,0}$ is zero. Note also that

$$
U_r (\hat{B}, \hat{F}, \hat{\Lambda}) \leq U_r (B_0, F_0, \Lambda_0) = 0
$$

40
by definition of \( \{ \hat{B}, \hat{F}, \hat{\Lambda} \} \). Therefore, we have

\[
0 \geq U_r(\hat{B}, \hat{F}, \hat{\Lambda}) \geq \frac{\vartheta}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( x_{i,t} \hat{b}_i + f'_t \hat{\lambda}_i - f'_{t,0} \lambda_{i,0} \right)^2 + o_p(1).
\]

Combined with \( \frac{\vartheta}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{i,t} \hat{b}_i + f'_t \hat{\lambda}_i - f'_{t,0} \lambda_{i,0})^2 \geq 0 \) for any \( \hat{b}_i, \hat{\lambda}_i \), and \( f_t \), it must be true that

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

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

\[
\tilde{U}_r(\hat{B}, \hat{F}, \hat{\Lambda}) = \frac{1}{NT} \sum_{i=1}^{N} \left\| M_F \left( X_i \hat{b}_i - F_0 \lambda_{i,0} \right) \right\|^2
\] (5)

then

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

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

Now, by (5)

\[
\tilde{U}_r(\hat{B}, \hat{F}, \hat{\Lambda}) = \frac{1}{NT} \sum_{i=1}^{N} \hat{b}'_i X_i' M_F X_i \hat{b}_i - \frac{2}{NT} \sum_{i=1}^{N} \hat{b}'_i X_i' M_F F_0 \lambda_{i,0} + \frac{1}{NT} \sum_{i=1}^{N} X_i' F_0' M_F F_0 \lambda_{i,0}
\]

\[
= \frac{1}{N} \sum_{i=1}^{N} \hat{b}'_i A_i \hat{b}_i - \frac{2}{N} \sum_{i=1}^{N} \hat{b}'_i C_i \eta + \frac{1}{N} \sum_{i=1}^{N} \eta' B_i \eta,
\]

where

\[
A_i = \frac{1}{T} X_i' M_F X_i, \quad B_i = (\lambda_{i,0} X_i' \lambda_{i,0}) \otimes I_T, \quad C_i' = \frac{1}{\sqrt{T}} X_i' \otimes (X_i' M_F), \quad \eta = \frac{1}{\sqrt{T}} \text{vec}(M_F F_0).
\]

Completing the square,

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

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

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

\[
\frac{1}{N} \sum_{i=1}^{N} \left( \hat{b}_i + A_i^{-1} C_i \eta \right)' A_i \left( \hat{b}_i + A_i^{-1} C_i \eta \right) = o_p(1).
\] (7)
From Assumption D, the matrix $N^{-1} \sum_{i=1}^{N} E_i$ is positive definite, and thus equation (6) implies that $\|\eta\|^2 = o_p(1)$. This result implies that

$$\|M_F - M_{F_0}\| = \|P_F - P_{F_0}\| = o_p(1). \quad (8)$$

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 $\|\eta\|^2 = o_p(1)$, equation (7) implies that

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

where $0 < \rho_A$ is the lower bound of the eigenvalues of $A_{i,0} = \frac{1}{T} X_i^T M_{F_0} X_i$ $i = 1, ..., N$.

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 $\hat{b}_i$ follows from $\frac{1}{N} \sum_{i=1}^{N} \hat{b}_i^T \hat{b}_i = o_p(1)$ (recall we normalize $b_{i,0} = 0$). The average consistency of $\hat{f}_i$ and the average consistency of $\hat{b}_i$ further imply the average consistency of $\hat{\lambda}_i$ (see Ando and Bai (2015)). That is, $N^{-1} \sum_{i=1}^{N} \|\hat{\gamma}_i - \gamma_{i,0}\|^2 = o_p(1)$. 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

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

**B.1 Proof of Theorem 1**

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

$$L_{NT,i}(\gamma_i, F) \equiv \frac{1}{T} \sum_{t=1}^{T} \rho_r (y_{it} - \mathbf{x}_i' \hat{b}_i - \mathbf{f}_i' \hat{\lambda}_i),$$

and its centered version

$$\bar{L}_{NT,i}(\gamma_i, F) \equiv L_{NT,i}(\gamma_i, F) - L_{NT,i}(\gamma_{i,0}, F_0)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \rho_r (y_{it} - \mathbf{x}_i' \hat{b}_i - \mathbf{f}_i \hat{\lambda}_i) - \frac{1}{T} \sum_{t=1}^{T} \rho_r (y_{it} - \mathbf{x}_i' b_{i,0} - \mathbf{f}_i' \lambda_{i,0}).$$

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

$$B_{i,T}(\delta) \equiv \{ \gamma_i : \|\gamma_i - \gamma_{i,0}\| \leq \delta \}$$

42
and

\[ B_T(e) \equiv \{ F : \| F - F_0 \|/T^{1/2} \leq e \}, \]

where \( e > 0 \) is a 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 \( \hat{b}_i \) and \( \hat{\lambda}_i \) in the sense of

\[
\max_{1 \leq i \leq N} \| \hat{b}_i - b_{i,0} \| = o_p(1) \text{ and } \max_{1 \leq i \leq N} \| \hat{\lambda}_i - \lambda_{i,0} \| = o_p(1). 
\]

Fix any \( \delta > 0 \). We can assume that the common factor \( F \) satisfies \( \| F - F_0 \|/T^{1/2} < e \). For each \( \gamma_i \notin B_{i,T}(\delta) \), define

\[ \hat{\gamma}_i = s_i \gamma_i + (1 - s_i) \gamma_{i,0}, \]

where \( \gamma_{i,0} = (b_{i,0}', \lambda_{i,0}')' \) and \( s_i = \delta / \{ \| \gamma_i - \gamma_{i,0} \| \} \in (0, 1) \). The convexity of the objective function, given \( F \), leads

\begin{align*}
& s_i \{ L_{NT,i}(\gamma_i, F) - L_{NT,i}(\gamma_{i,0}, F_0) \} \\
& \geq L_{NT,i}(\hat{\gamma}_i, F) - L_{NT,i}(\gamma_{i,0}, F_0) + L_{NT,i}(\gamma_{i,0}, F) - L_{NT,i}(\gamma_{i,0}, F) \\
& > E \left[ L_{NT,i}(\hat{\gamma}_i, F) \right] + \left\{ L_{NT,i}(\hat{\gamma}_i, F) - E \left[ \tilde{L}_{NT,i}(\hat{\gamma}_i, F) \right] \right\} - C \times e, \quad (9)
\end{align*}

where \( E[\tilde{L}_{NT,i}(\gamma_i, F)] = E[L_{NT,i}(\gamma_i, F)] - E[L_{NT,i}(\gamma_{i,0}, F_0)] \) and the expectation \( E[\cdot] \) is taken with respect to the true conditional distribution of \( \{ y_{it} : i = 1, \ldots, N, \ t = 1, \ldots, T \} \) conditional on \( X, F_0, \Lambda_0 \) and \( B_0 \). Here, we used

\[ |L_{NT,i}(\gamma_{i,0}, F_0) - L_{NT,i}(\gamma_{i,0}, F)| \leq C \times T^{-1} \sum_{t=1}^{T} \| f_t - f_{t,0} \| \leq C \times \| F - F_0 \|/T^{1/2} < C \times e \]

with \( C \) being a positive constant. Note that \( F \in B_T(e) \). By choosing small \( e \), this term is dominated by the term \( E[\tilde{L}_{NT,i}(\hat{\gamma}_i, F)] + \{ \tilde{L}_{NT,i}(\hat{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\hat{\gamma}_i, F)] \} \). Thus, this term is negligible in our analysis below.

Let \( \omega_{it} = (x_{it}, f_{t,0}', \lambda_{i,0}) \). Under \( \| \gamma_i - \gamma_{i,0} \| \leq \delta \), the identity of Knight (1998) leads to

\begin{align*}
& \frac{1}{T} \sum_{t=1}^{T} E \left[ \rho_\tau (y_{it} - x_{it}' b_i - f_{t}' \lambda_i) - \rho_\tau (y_{it} - x_{it}' b_{i,0} - f_{t,0}' \lambda_{i,0}) \right] \\
& = \frac{1}{T} \sum_{t=1}^{T} \int_{0}^{1} \left( x_{it}' (b_t - b_{i,0}) + f_{t,0}' (\lambda_t - \lambda_{i,0}) + (f_t - f_{t,0})' \lambda_i \right) \{ G_{it}(s | \omega_{it}) - \tau \} \ ds \\
& = \frac{1}{T} \sum_{t=1}^{T} g_{it}(\tilde{h}_{it} | \omega_{it}) \left\{ x_{it}' (b_t - b_{i,0}) + f_{t,0}' (\lambda_t - \lambda_{i,0}) + (f_t - f_{t,0})' \lambda_i \right\}^2 \\
& = \frac{1}{T} \sum_{t=1}^{T} \sum_{l=1}^{L} g_{il}(\tilde{h}_{il} | \omega_{it}) \left\{ x_{it}' (b_t - b_{i,0}) + f_{t,0}' (\lambda_t - \lambda_{i,0}) \right\}^2 + \frac{1}{T} \sum_{t=1}^{T} g_{it}(\tilde{h}_{it} | \omega_{it}) \left\{ (f_t - f_{t,0})' \lambda_i \right\}^2 \\
& \quad + \frac{2}{T} \sum_{t=1}^{T} g_{it}(\tilde{h}_{it} | \omega_{it}) \left\{ x_{it}' (b_t - b_{i,0}) + f_{t,0}' (\lambda_t - \lambda_{i,0}) \right\} \left\{ (f_t - f_{t,0})' \lambda_i \right\} \right) (10)
\end{align*}
where \( \tilde{h}_{it} \) is between 0 and \( x'_i(b_i - b_{i,0}) + f_{t,0}'(\lambda_i - \lambda_{i,0}) + (f_i - f_{t,0})'\lambda_i \), \( G_{it}(\cdot|\omega_{it}) \) is the conditional cumulative distribution of \( \varepsilon_{it} = y_{it} - x'_i b_{i,0} - f_{t,0}'\lambda_{i,0} \) conditional on \( \omega_{it} \).

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 \( 0 < \tilde{g} \leq g_{it} (\tilde{h}_{it}|\omega_{it}) \). Together with Assumption (D2), there exists a positive constant \( \epsilon_\delta \) such that for \( 0 < \epsilon_\delta \leq \delta \),

\[
\frac{1}{T} \sum_{t=1}^{T} \tilde{g} \left\{ x'_i(b_i - b_{i,0}) + f_{t,0}'(\lambda_i - \lambda_{i}) \right\}^2 > \epsilon_\delta
\]

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

\[
E[\tilde{L}_{NT,i}(\tilde{\gamma}_i, F)] \geq \frac{\epsilon_\delta}{2}.
\]

From (9), we obtain the following inclusion relation

\[
\left\{ \max_{i} \|\tilde{\gamma}_i - \gamma_{i,0}\| > \delta, \ F \in B_T(e) \right\}
= \left\{ \|\tilde{\gamma}_i - \gamma_{i,0}\| > \delta, \ \exists i, \ F \in B_T(e) \right\}
\leq \left\{ L_{NT,i}(\gamma_i, F) \leq L_{NT,i}(\gamma_{i,0}, F_0), \ 1 \leq \exists i \leq N, \ \exists \gamma_i \notin B_{i,T}(\delta), \ F \in B_T(e) \right\}
\leq \left\{ \max_{1 \leq i \leq N} \sup_{\gamma_i \in B_{i,T}(\delta), \ F \in B_T(e)} \left| \tilde{L}_{NT,i}(\gamma_i, F) - E[\tilde{L}_{NT,i}(\gamma_i, F)] \right| \geq \frac{\epsilon_\delta}{2} \right\},
\]

where the following argument is used for the second inclusion. From (9), we have

\[
l_{i} \{ L_{NT,i}(\gamma_i, F) - L_{NT,i}(\gamma_{i,0}, F_0) \} \geq E[\tilde{L}_{NT,i}(\tilde{\gamma}_i, F)] + \{ L_{NT,i}(\tilde{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\gamma}_i, F)] \}.
\]

If \( L_{NT,i}(\gamma_i, F) \leq L_{NT,i}(\gamma_{i,0}, F_0), \ 1 \leq \exists i \leq N, \ \exists \gamma_i \notin B_{i,T}(\delta) \), this leads to

\[
0 \geq E[\tilde{L}_{NT,i}(\tilde{\gamma}_i, F)] + \{ L_{NT,i}(\tilde{\gamma}_i, F) - E[\tilde{L}_{NT,i}(\tilde{\gamma}_i, F)] \}.
\]

From (12), \( E[\tilde{L}_{NT,i}(\tilde{\gamma}_i, F)] \geq \epsilon_\delta/2 \), note that \( \tilde{\gamma}_i \in B_{i,T}(\delta) \). Thus, we should have

\[
\max_{1 \leq i \leq N} \sup_{\gamma_i \in B_{i,T}(\delta), \ F \in B_T(e)} \left| \tilde{L}_{NT,i}(\gamma_i, F) - E[\tilde{L}_{NT,i}(\gamma_i, F)] \right| \geq \epsilon_\delta/2.
\]

The second inclusion is obtained. We show this event has small probability.

It suffices to show that for every \( \epsilon > 0 \),

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

It further suffices to prove that

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

Let \( h(\gamma_i, f_i) \equiv \rho_r(y_{it} - x'_i b_{i,t} - f_{t,0}'\lambda_i) - \rho_r(y_{it} - x'_i b_{i,0} - f_{t,0}'\lambda_{i,0}) \). Observe that

\[
|h(\gamma_i, f_i) - h(\gamma_i, f_{t,0})| \leq C(\|x_{it}\| + \|f_{t,0}\|)\|\gamma_i - \gamma_i\| + C(\|f_i - f_{t,0}\|)
\]

and
\[ |h(\gamma_i; f_{t,0}) - h(\gamma_i; f_{t,0})| \leq C(||x_{it}|| + ||f_{t,0}||)||\gamma_i - \gamma_i| \] for some universal constant \( C > 0 \).

Put \( W_{it} \equiv C(||x_{it}|| + ||f_{t,0}||) \) and \( \kappa \equiv \sup_{i,t} W_{it} \). Since \( B_{i,T}(\delta) \) is compact subset, there exist \( C_T \) balls with centers \( \gamma_k = (\tilde{b}_k', \tilde{\lambda}_k') \), \( k = 1, ..., C_T \) and radius \( \varepsilon/(8\kappa) \) such that the collection of these balls covers \( B_{i,T}(\delta) \). Note that \( C_T \) can be chosen such that \( C_T(\varepsilon) = O(1/\varepsilon^{p+r}) \) and \( \varepsilon \to 0 \). For each \( \gamma_k \in B_{i,T}(\delta) \), there is \( \gamma_k \) such that \( |h(\gamma_i; f_t) - h(\gamma_k; f_{t,0})| \leq W_{it}\varepsilon/(8\kappa) \). These investigations lead to

\[
\begin{align*}
\tilde{L}_{NT,i}(\gamma_i; F) - E[\tilde{L}_{NT,i}(\gamma_i; F)] & \leq |\tilde{L}_{NT,i}(\gamma_i, F_0) - E[\tilde{L}_{NT,i}(\gamma_i, F_0)]| + C \frac{T}{T} \sum_{t=1}^{T} ||f_t - f_{t,0}|| \\
& \leq |\tilde{L}_{NT,i}(\gamma_k, F_0) - E[\tilde{L}_{NT,i}(\gamma_k, F_0)]| + \frac{\varepsilon}{8\kappa} \times \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \{W_{it} + E[W_{it}]\} + C \frac{T}{T} \sum_{t=1}^{T} ||f_t - f_{t,0}||,
\end{align*}
\]

where the last term can be made arbitrarily small under large \( N \) and \( T \) because

\[
T^{-1} \sum_{t=1}^{T} ||f_t - f_{t,0}||^2 \leq \varepsilon, \text{ which implies} \n\]

\[
T^{-1} \sum_{t=1}^{T} ||f_t - f_{t,0}|| \leq \sqrt{\sum_{t=1}^{T} ||f_t - f_{t,0}||^2} \leq \sqrt{\varepsilon}. \n\]

Therefore, we have

\[
P \left\{ \sup_{\gamma_i \in B_{i,T}(\delta), F \in B_{T}(\varepsilon)} \left| \tilde{L}_{NT,i}(\gamma_i; F_0) - E[\tilde{L}_{NT,i}(\gamma_i, F_0)] \right| \geq \varepsilon \right\}
\]

\[
\leq \sum_{k=1}^{C_T(\varepsilon)} P \left\{ \left| \tilde{L}_{NT,i}(\gamma_k, F_0) - E[\tilde{L}_{NT,i}(\gamma_k, F_0)] \right| \geq \frac{\varepsilon}{3} \right\}
\]

\[
+ P \left\{ \left| \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \{W_{it} + E[W_{it}]\} \right| \geq \frac{8\kappa}{3} \right\} + P \left\{ \left| \frac{1}{T} \sum_{t=1}^{T} ||f_t - f_{t,0}|| \right| \geq \frac{\varepsilon}{3\kappa} \right\}
\]

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 \( 2\kappa \).

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 \( x_{it} \) and \( f_t \), we have \( |h(\gamma_i; f_t) - E[h(\gamma_i; f_t)]| \leq M \) where \( M \) is bounded constant. From the independence property of \( \varepsilon_{it} \), Lemma 3 leads

\[
P \left\{ \left| \tilde{L}_{NT,i}(\gamma_k, F_0) - E[\tilde{L}_{NT,i}(\gamma_k, F_0)] \right| \geq \frac{\varepsilon}{3} \right\}
\]

\[
= \exp \left( \frac{-T^2\varepsilon^2}{18\sum_{t=1}^{T} M^2} \right),
\]

which leads

\[
\sum_{k=1}^{C_T(\varepsilon)} P \left\{ \left| \tilde{L}_{NT,i}(\gamma_k, F_0) - E[\tilde{L}_{NT,i}(\gamma_k, F_0)] \right| \geq \frac{\varepsilon}{3} \right\}
\]

45
Next, we prove that the estimated common factor is $\hat{\gamma}$ taken with respect to the true conditional distribution of $\{y_{it} : i = 1, ..., N, t = 1, ..., T\}$, which completes the proof of the uniform consistency of $\hat{\gamma}$ where

$$E \left[ \gamma \right] = \gamma_0$$

and its centered version $\hat{\gamma} - \gamma_0$. From (11), we obtain the inclusion relation

$$\left\{ \max_{1 \leq t \leq T} \| \hat{f}_t - f_{t,0} \| > \delta \right\} \subset \left\{ \| \hat{f}_t - f_{t,0} \| > \delta, \; \exists t \right\} \subset \left\{ S_{NT,t}(\hat{\gamma}, f_t) \leq S_{NT,t}(\gamma, f_{t,0}), \; 1 \leq \exists t \leq T, \exists f_t, \; \text{s.t.} \; \| f_t - f_{t,0} \| > \delta \right\} \subset \left\{ \max_{1 \leq t \leq T} \sup_{\| f_t - f_{t,0} \| \leq \delta} \left| E[S_{NT,t}(\hat{\gamma}, f_t) - E[S_{NT,t}(\gamma, f_{t,0})]] \right| \geq \frac{\epsilon_d}{2} \right\},$$

where $E[S_{NT,t}(\gamma, f_t)] = E[S_{NT,t}(\gamma, f_t)] - E[S_{NT,t}(\gamma, f_{t,0})]$ 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 $\Lambda_0$. The second inclusion is obtained as follows. Because of the convexity of the objective function, we have

$$s_t \{ S_{N,T,t}(\hat{\gamma}, f_t) - S_{N,T,t}(\hat{\gamma}, f_{t,0}) \}$$

$$\geq S_{N,T,t}(\hat{\gamma}, \tilde{f}_t) - S_{N,T,t}(\hat{\gamma}, f_{t,0})$$

$$= \tilde{S}_{N,T,t}(\hat{\gamma}, \tilde{f}_t) - \tilde{S}_{N,T,t}(\hat{\gamma}, f_{t,0})$$

$$= \left[ \tilde{S}_{N,T,t}(\hat{\gamma}, \tilde{f}_t) - E \left[ \tilde{S}_{N,T,t}(\gamma_0, \tilde{f}_t) \right] \right] + E \left[ \tilde{S}_{N,T,t}(\hat{\gamma}, f_{t,0}) \right]$$

$$- \left\{ \tilde{S}_{N,T,t}(\hat{\gamma}, f_{t,0}) - E \left[ \tilde{S}_{N,T,t}(\hat{\gamma}, f_{t,0}) \right] \right\} + E \left[ \tilde{S}_{N,T,t}(\gamma_0, \tilde{f}_t) \right].$$

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

$$s_t \{ S_{N,T,t}(\hat{\gamma}, f_t) - S_{N,T,t}(\hat{\gamma}, f_{t,0}) \}$$

$$\geq \left[ \tilde{S}_{N,T,t}(\hat{\gamma}, \tilde{f}_t) - E \left[ \tilde{S}_{N,T,t}(\gamma_0, \tilde{f}_t) \right] \right] + e_\delta + o_p(1).$$

If $S_{N,T,t}(\hat{\gamma}, f_t) \leq S_{N,T,t}(\gamma_0, f_t)$, we should have $\max_{1 \leq t \leq T, \| f_t - f_{t,0} \| \leq \delta} \| \tilde{S}_{N,T,t}(\hat{\gamma}, f_t) - E[\tilde{S}_{N,T,t}(\gamma_0, f_t)] \| \geq e_\delta/2$. Note that $\| f_t - f_{t,0} \| = \delta$. Therefore, the second inclusion is obtained. Therefore, it suffices to show that for every $\varepsilon > 0$,

$$\lim_{N,T \to \infty} P \left\{ \max_{1 \leq t \leq T} \sup_{\| f_t - f_{t,0} \| \leq \delta} \left| \tilde{S}_{N,T,t}(\hat{\gamma}, f_t) - E[\tilde{S}_{N,T,t}(\gamma_0, f_t)] \right| \geq \varepsilon \right\} = 0.$$

It suffices to prove that

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

and

$$\lim_{N,T \to \infty} P \left\{ \max_{1 \leq t \leq T} \sup_{\| f_t - f_{t,0} \| \leq \delta} \left| E[\tilde{S}_{N,T,t}(\hat{\gamma}, f_t)] - E[\tilde{S}_{N,T,t}(\gamma_0, f_t)] \right| \geq \varepsilon \right\} = 0.$$

Since

$$\left| \tilde{S}_{N,T,t}(\hat{\gamma}, f_t) - \tilde{S}_{N,T,t}(\gamma_0, f_t) \right| \leq C \times \left\{ \frac{1}{N} \sum_{i=1}^{N} \| x_{it} \| \| \tilde{b}_i - b_{i,0} \| + \frac{1}{N} \sum_{i=1}^{N} \| f_t \| \| \tilde{\lambda}_i - \lambda_{i,0} \| \right\},$$

and $\sup_{it} \| x_{it} \| < \infty$ and $\sup_{i} \| f_t \| < \infty$, consistency of $\hat{\gamma}$ implies (17).
Finally, we prove (16). Because we already proved the consistency of \( \hat{\gamma}_i \) \((i = 1, \ldots, N)\), it is enough to show

\[
\max_{1 \leq t \leq T} P \left\{ \sup_{\gamma_i \in B_t(\delta), \|f_t - f_{t,0}\| \leq \delta} \left| \bar{S}_{NT, t}(\gamma, f_t) - E[\bar{S}_{NT, t}(\gamma, f_t)] \right| \geq \varepsilon \right\} = o(T^{-1}).
\] (18)

For each \( \delta > 0 \), we define

\[
B_{t, N}(\delta) \equiv \left\{ f_t : \|f_t - f_{t,0}\| \leq \delta \right\}.
\]

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

\[
\begin{align*}
&\left| \bar{S}_{NT, t}(\gamma, f_t) - E[\bar{S}_{NT, t}(\gamma, f_t)] \right| \\
\leq &\left| \bar{S}_{NT, t}(\gamma_0, f_t) - E[\bar{S}_{NT, t}(\gamma_0, f_t)] \right| + \left| \bar{S}_{NT, t}(\gamma, f_t) - \bar{S}_{NT, t}(\gamma_0, f_t) \right| \\
&+ \left| E[\bar{S}_{NT, t}(\gamma, f_t)] - E[\bar{S}_{NT, t}(\gamma_0, f_t)] \right| \\
\leq &\left| \bar{S}_{NT, t}(\gamma_0, f_k) - E[\bar{S}_{NT, t}(\gamma_0, f_k)] \right| + C \frac{1}{N} \sum_{i=1}^{N} \|\gamma - \gamma_0\| \\
\leq &\left| \bar{S}_{NT, t}(\gamma_0, \bar{f}_k) - E[\bar{S}_{NT, t}(\gamma_0, \bar{f}_k)] \right| + \varepsilon \frac{1}{8\beta} \frac{1}{N} \sum_{i=1}^{N} \{K(\lambda_i) + E[K(\lambda_i)]\} + C \frac{N}{N} \sum_{i=1}^{N} \|\gamma - \gamma_0\|,
\end{align*}
\]

where \( C \) is universal constant. Therefore, we have

\[
P \left\{ \sup_{\gamma_i \in B_t(\delta), F \in B_T(\varepsilon)} \left| \bar{S}_{NT, t}(\gamma, f_t) - E[\bar{S}_{NT, t}(\gamma, f_t)] \right| \geq \varepsilon \right\} \\
\leq \sum_{k=1}^{C_f(\varepsilon)} P \left\{ \left| \bar{S}_{NT, t}(\gamma_k, F_0) - E[\bar{S}_{NT, t}(\gamma_k, F_0)] \right| \geq \frac{\varepsilon}{3} \right\} \\
+ P \left\{ \frac{1}{N} \sum_{i=1}^{N} \{K_i + E[K_i]\} \geq \frac{8\beta}{3} \right\} + P \left\{ \frac{1}{N} \sum_{i=1}^{N} \|\lambda_i - \lambda_{i,0}\| \geq \frac{\varepsilon}{3C} \right\},
\]

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 \( x_{ik} \) and \( f_t \), we have \( |h(\gamma_{i,0}, f_t) - E[h(\gamma_{i,0}, f_t)]| \leq M \) where \( M \) is bounded constant. The first term is

\[
\sum_{k=1}^{C_f(\varepsilon)} P \left\{ \left| \bar{S}_{NT, t}(\gamma_k, \bar{f}_k) - E[\bar{S}_{NT, t}(\gamma_k, \bar{f}_k)] \right| \geq \frac{\varepsilon}{3} \right\}.
\]
\[
\leq C_f(\varepsilon) \times \exp\left(\frac{-N^2\varepsilon^2}{18 \sum_{i=1}^{N} M_i^2}\right) \\
= O\left(\varepsilon^{-r} \times \exp(-N)\right) \\
= O\left(\exp\left[-N \left\{1 + r \frac{\log(\varepsilon)}{N}\right\}\right]\right) \\
= o(T^{-2})
\] (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.

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

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

We note that these results are the preliminary convergence rates of \(\hat{\gamma}_i\) and \(\hat{f}_t\). 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 \(0 \leq a \leq V/(sd)\),

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

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 \(\sigma\)-field. We denote \(\mathcal{H}\) being a class of measurable functions on the measurable space \((\mathcal{S}, \mathcal{Q})\). For a process \(Z(h)\) defined on \((\mathcal{S}, \mathcal{Q})\), we define \(\|Z(h)\|_{\mathcal{H}} = \sup_{h \in \mathcal{H}} |Z(h)|\). 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)
We define \( z_{it} = (x_{it}', f_{i,0}')', z_{it,0} = (x_{it}', f_{i,0}')', \hat{z}_{it} = (x_{it}', \hat{f}_{i}')', \omega_{it} = \{x_{it}, f_{i,0}, \lambda_{i,0}\} \) and

\[
Q_{NT,i}(\gamma_i, F) \equiv \frac{1}{T} \sum_{t=1}^{T} \left( \tau - I(\varepsilon_{it} \leq x_{it}'(b_i - b_{i,0}) + f_i', \lambda_i - f_{i,0}' \lambda_{i,0}) \right) z_{it},
\]

\[
\overline{Q}_{NT,i}(\gamma_i, F) \equiv E[Q_{NT,i}(\gamma_i, F)] = \frac{1}{T} \sum_{t=1}^{T} E \left[ \left( \tau - G_{it}(x_{it}'(b_i - b_{i,0}) + f_i', \lambda_i - f_{i,0}' \lambda_{i,0}| \omega_{it}) \right) z_{it} \right],
\]

where the expectation of \( E[Q_{NT,i}(\gamma_i, F)] \) is taken with respect to the true conditional distribution of \( \{y_{it} : i = 1, ..., N, \; t = 1, ..., T\} \) conditional on \( X, F_0 \) and \( \Lambda_0 \), \( G_{it}(|x_{it}) \) is the conditional cumulative distribution function of \( \varepsilon_{it} \). Because of the computational property of the quantile regression estimator (Gutenbrunner and Jureckova (1992)), it is known that \( \max_{1 \leq i \leq N} |Q_{NT,i}(\gamma_i, \hat{F})| \) is bounded by

\[
O_p(T^{-1} \max_{1 \leq i \leq N, \; 1 \leq t \leq T} \|x_{it}\|) + O_p(T^{-1} \max_{1 \leq t \leq T} \|\hat{f}_t\|) = O_p(T^{-1}).
\]

Here, we used \( \max_{1 \leq t \leq T} \|\hat{f}_t\| = O_p(1) \). We thus have

\[
O_p(T^{-1}) = Q_{NT,i}(\gamma_i, \hat{F}) = Q_{NT,i}(\gamma_{i,0}, F_0) + \overline{Q}_{NT,i}(\gamma_i, \hat{F}) + \left\{ Q_{NT,i}(\gamma_i, \hat{F}) - \overline{Q}_{NT,i}(\gamma_i, \hat{F}) - Q_{NT,i}(\gamma_{i,0}, F_0) \right\}.
\]

Expanding \( \overline{Q}_{NT,i}(\gamma_i, \hat{F}) \) at \( (\gamma_{i,0}, F_0) \), we obtain

\[
\overline{Q}_{NT,i}(\gamma_i, \hat{F}) = -\frac{1}{T} \sum_{t=1}^{T} E[g_{it}(0|\omega_{it})z_{it,0}z_{it,0}'](\gamma_i - \gamma_{i,0}) - \frac{1}{T} \sum_{t=1}^{T} E[g_{it}(0|\omega_{it})z_{it,0}z_{it,0}'](\hat{f}_t - f_{i,0}) + o_p(\|\gamma_i - \gamma_{i,0}\|) + o_p\left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{f}_t - f_{i,0}\| \right),
\]

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

\[
\Gamma_i(\gamma_i - \gamma_{i,0}) + o_p(\|\gamma_i - \gamma_{i,0}\|) = Q_{NT,i}(\gamma_{i,0}, F_0) + \left\{ Q_{NT,i}(\lambda_i, \hat{F}) - \overline{Q}_{NT,i}(\lambda_i, \hat{F}) - Q_{NT,i}(\gamma_{i,0}, F_0) \right\}
\]

\[
- \left[ \frac{1}{T} \sum_{t=1}^{T} E[g_{it}(0|\omega_{it})z_{it,0}z_{it,0}'](\hat{f}_t - f_{i,0}) \right] + O_p\left( \frac{1}{T} \sum_{t=1}^{T} \|\hat{f}_t - f_{i,0}\| \right),
\]

C.1 Proof of Lemma 4

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

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

for all \( s > 0 \).
Putting (26) into (24), we have

\[ V_{NT,t}(\gamma, f_t) = \frac{1}{N} \sum_{i=1}^{N} \left( \tau - I(\varepsilon_{it} \leq x_{it}'(b_i - b_{i,0}) + f'_i \lambda_i - f'_{t,0} \lambda_{i,0}) \right) \lambda_i, \]

and

\[ \nabla_{NT,t}(\gamma, f_t) \equiv E \left[ \frac{1}{N} \sum_{i=1}^{N} \left( \tau - G_{it}(x_{it}'(b_i - b_{i,0}) + f'_i \lambda_i - f'_{t,0} \lambda_{i,0} | \omega_{it}) \right) \lambda_i \right], \]

where \( G_{it}(\cdot | \omega_{it}) \) is the true cumulative distribution function of \( \varepsilon_{it} \) conditioned on \( \omega_{it} \).

Noting that \( \nabla_{NT,t}(\gamma_0, f_{t,0}) = 0 \), the expansion of \( \nabla_{NT,t}(\hat{\gamma}_t, \hat{f}_t) \) at \( (\gamma_0, f_{t,0}) \) leads to

\[
\begin{align*}
\nabla_{NT,t}(\hat{\gamma}_t, \hat{f}_t) &= -\frac{1}{N} \sum_{i=1}^{N} E[g_{it}(0 | \omega_{it}) \lambda_{i,0} x_{i,t}'(f_t - f_{t,0})] - \frac{1}{N} \sum_{i=1}^{N} E[g_{it}(0 | \omega_{it}) \lambda_{i,0} x_{i,t}'(\hat{f}_t - f_{t,0})] + o_p\left(\frac{1}{N} \sum_{i=1}^{N} \|\gamma_i - \gamma_{i,0}\|\right).
\end{align*}
\]

(25)

Viewing the factor loadings \( \lambda_i \) as regressors and \( f_t \) as regression coefficient, the similar argument that derived the equation (24) leads to

\[
\begin{align*}
\hat{f}_t - f_{t,0} + o_p(\|f_t - f_{t,0}\|)
\end{align*}
\]

\[
\begin{align*}
= \Psi_t^{-1} V_{NT,t}(\gamma_0, f_{t,0}) - \Psi_t^{-1} \left\{ V_{NT,t}(\hat{\gamma}_t, \hat{f}_t) - \nabla_{NT,t}(\hat{\gamma}_t, \hat{f}_t) - V_{NT,t}(\gamma_0, f_{t,0}) \right\} \\
+ \Psi_t^{-1} \left( \frac{1}{N} \sum_{i=1}^{N} E[g_{it}(0 | \omega_{it}) \lambda_{i,0} x_{i,t}'(\hat{f}_t - f_{t,0})] \right) + o_p\left(\frac{1}{N} \sum_{i=1}^{N} \|\gamma_i - \gamma_{i,0}\|\right) + O_p\left(\frac{1}{N}\right),
\end{align*}
\]

(26)

where \( \Psi_t = N^{-1} \sum_{i=1}^{N} E[g_{it}(0 | \omega_{it}) \lambda_{i,0} x_{i,t}'] \).

Putting (26) into (24), we have

\[
\begin{align*}
(\hat{\gamma}_i - \gamma_{i,0})
\end{align*}
\]

\[
\begin{align*}
= \Gamma_i^{-1} Q_{NT,i}(\gamma_{i,0}, F_0) - \Gamma_i^{-1} \left\{ Q_{NT,i}(\hat{\gamma}_i, \hat{F}_i) - Q_{NT,i}(\gamma_{i,0}, F_0) \right\} \\
- \frac{1}{T} \sum_{t=1}^{T} \Gamma_i^{-1} E[g_{it}(0 | \omega_{it}) x_{i,t}'(\hat{f}_t - f_{t,0})] \Psi_t^{-1} \left( \frac{1}{N} \sum_{j=1}^{N} E[g_{jt}(0 | \omega_{jt}) \lambda_{j,0} x_{j,t}'(\hat{f}_t - f_{t,0})] \right) \\
+ \frac{1}{T} \sum_{t=1}^{T} \Gamma_i^{-1} E[g_{it}(0 | \omega_{it}) x_{i,t}'(\hat{f}_t - f_{t,0})] \Psi_t^{-1} V_{NT,t}(\gamma_0, f_{t,0}) \\
+ \frac{1}{T} \sum_{t=1}^{T} \Gamma_i^{-1} E[g_{it}(0 | \omega_{it}) x_{i,t}'(\hat{f}_t - f_{t,0})] \Psi_t^{-1} \left\{ V_{NT,t}(\hat{\gamma}_t, \hat{f}_t) - \nabla_{NT,t}(\hat{\gamma}_t, \hat{f}_t) - V_{NT,t}(\gamma_0, f_{t,0}) \right\} \\
+ O_p\left(\frac{1}{N}\right) + O_p\left(\frac{1}{T}\right) + o_p(\|\gamma_i - \gamma_{i,0}\|),
\end{align*}
\]

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

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

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

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

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 \( \Gamma_i^{-1} \{ Q_{NT,i}(\hat{\gamma}_i, \hat{F}) - \overline{Q}_{NT,i}(\hat{\gamma}_i, \hat{F}) - Q_{NT,i}(\gamma_{i,0}, F_0) \} \) in (27), we apply Lemma 6. Define

\( h(u|x, \gamma_i, f_i) = I(u \leq x' (b_i - b_{i,0}) + f_{t,0} \lambda_{i,0} - f_i \lambda_i) - I(u \leq 0) \). From the result of Theorem 1, we define \( \mathcal{H} = \{ h(u|x, \gamma_i, f_i) \} \) \( ||\hat{\gamma}_i - \gamma_{i,0}|| < \kappa, \ ||\hat{f}_i - f_{t,0}|| < \kappa \) with \( \kappa \to 0 \). It is obvious that \( E[h(u|x, \gamma_i, f_i)] = 0 \) and that each element in \( \mathcal{H} \) is bounded by \( 2 \). Also, \( E[\{h(u|x, \gamma_i, f_i) - E[h(u|x, \gamma_i, f_i)]\}^2] \leq C \times \kappa^2 \) for \( h \in \mathcal{H} \). Put \( Z_i = \| \sum_{t=1}^{T} \{ \xi_{it}|x, \gamma_i, f_i \} - E[\xi_{it}|x, \gamma_i, f_i] \|_{\mathcal{H}} \) 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

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

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 \( P(T^{-1} Z_i > T^{-1/2}) \leq N^{-2} \). 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

$$\sqrt{T} \left[ \frac{1}{T} \sum_{t=1}^{T} \Gamma_i^{-1} J_{it} \Psi_t^{-1} V_{NT,t}(\gamma_{0,t}, f_{t,0}) \right] = \frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Gamma_i^{-1} J_{it} \Psi_t^{-1} (\tau - I(\varepsilon_{it} \leq 0)) \lambda_i = o_p(1), \tag{30}$$

where \( J_{it} = E[g_{it}(0|\omega_{it})z_{it,0}\lambda_{0,t}] \). Because \( \| \Gamma_i^{-1} J_{it} \Psi_t^{-1} \| < \infty \), it is enough to show that

$$\frac{1}{N \sqrt{T}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tau - I(\varepsilon_{it} \leq 0)) \lambda_i = o_p(1). \tag{31}$$

The expected value of its second moment is

$$\frac{1}{N^2 T} E \left[ \sum_{i=1}^{N} \sum_{t,s=1}^{T} (\tau - I(\varepsilon_{it} \leq 0)) \lambda_i \lambda_j' (\tau - I(\varepsilon_{js} \leq 0)) \right] = \frac{1}{N^2 T} E \left[ \sum_{i=1}^{N} \sum_{t,s=1}^{T} (\tau - I(\varepsilon_{it} \leq 0)) \lambda_i \lambda_j' (\tau - I(\varepsilon_{is} \leq 0)) \right] = \frac{1}{N^2 T} E \left[ \sum_{i=1}^{N} \sum_{t=1}^{T} (\tau - I(\varepsilon_{it} \leq 0))^2 \lambda_i \lambda_j' \right],$$

52
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 $\{\hat{\gamma}_t, \hat{F}_t\}$ and the uniform boundedness of $\|\Gamma_i^{-1}E[g_{it}(0|\omega_{it})z_{it,0}]\|$, $\lambda_i$ and $\lambda_{i,0}$, it suffices to show that

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

converges to zero. Here $\zeta_{it} = \varepsilon_{it}$, $h(u|x, \gamma_{it}, f_{it}) = I(u \leq \alpha'(b_{it} - b_{i,0}) + f_{i,0}\lambda_{i,0} - f_i\lambda_i) - I(u \leq 0)$, $B_\gamma(\kappa) \equiv \{ \gamma; \|\gamma_i - \gamma_{i,0}\| \leq \kappa, i = 1, \ldots, N \}$, and $B_F(\kappa) \equiv \{ F; \|f_t - f_{t,0}\| \leq \kappa, t = 1, \ldots, T \}$ with $\kappa \to 0$.

For some positive constant $C_t$, put $W_i \equiv C_t \times (\|\lambda_i\| + \|\lambda_{i,0}\|)$, $\alpha \equiv \sup_i E[W_i]$ and $\varepsilon = \frac{1}{T^{1/4}}$.

There exist $C_{\gamma_i}$ balls with centers $\gamma_{ik}$, $k = 1, \ldots, C_{\gamma_i}$ and radius $\varepsilon/(8\alpha)$ such that the collection of these balls covers $B_{\gamma_i}(\kappa) \equiv \{ \gamma_i; \|\gamma_i - \gamma_{i,0}\| \leq \kappa \}$. Similarly, because $B_{f_{it}}(e)$ is compact subset, there exist $C_{f_{it}}$ balls with centers $\tilde{f}_{ij}$, $j = 1, \ldots, C_{f_{it}}$ and radius $\varepsilon/(8\alpha)$ such that the collection of these balls covers $B_{f_{it}}(e)$. We note that $C_{\gamma_i}$ and $C_{f_{it}}$ can be chosen such that $C_{\gamma_i}(\varepsilon) = O(1/\varepsilon^p+r)$ and $C_{f_{it}}(\varepsilon) = O(1/\varepsilon^r)$ with $\varepsilon \to 0$.

Because

$$\left| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \{ h(\zeta_{it}|x_{it}, \gamma_{it}, f_{t}) - E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_{t})] \} \right|$$

$$\leq \left| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \{ h(\zeta_{it}|x_{it}, \gamma_{ik}, f_{it}) - E[h(\zeta_{it}|x_{it}, \gamma_{ik}, f_{it})] \} \right| + \left| \frac{\varepsilon}{8\alpha} \cdot \frac{1}{N} \sum_{i=1}^N \left\{ W_i + E[W_i] \right\} \right|,$$

we have

$$P\left\{ \sup_{\gamma \in B_\gamma(\delta), F \in B_F(\varepsilon)} \left| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \{ h(\zeta_{it}|x_{it}, \gamma_{it}, f_{t}) - E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_{t})] \} \right| \geq \frac{1}{T^{1/4}} \right\} \quad \leq \sum_{k_1=1}^{C_{\gamma_1}(\varepsilon)} \cdots \sum_{k_N=1}^{C_{\gamma_N}(\varepsilon)} \sum_{j_1=1}^{C_{f_1}(\varepsilon)} \cdots \sum_{j_T=1}^{C_{f_T}(\varepsilon)} \left\{ \left| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \{ h(\zeta_{it}|x_{it}, \gamma_{ik}, f_{it}) - E[h(\zeta_{it}|x_{it}, \gamma_{ik}, f_{it})] \} \right| \geq \frac{1}{2T^{1/4}} \right\}$$

$$+ P\left\{ \left| \varepsilon \cdot \frac{1}{N} \sum_{i=1}^N \left\{ W_i + E[W_i] \right\} \right| \geq \frac{1}{2T^{1/4}} \right\},$$

where the second term is zero because $\left| \frac{1}{N} \sum_{i=1}^N \left\{ W_i + E[W_i] \right\} \right| < 2\alpha$ 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,

$$E \left\{ \left\{ \frac{1}{T} \sum_{t=1}^T h(\zeta_{it}|x_{it}, \gamma_{it}, f_{t}) - \frac{1}{T} \sum_{t=1}^T E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_{t})] \right\}^2 \right\}$$

53
\[
\sum_{i=1}^{N} E \left\{ T^{-1} \sum_{t=1}^{T} h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t}) - T^{-1} \sum_{t=1}^{T} E[h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t})] \right\}^2 \leq C \times N \times o(T^{-1})
\]

where \( C \) is some positive constant. Take

\[ Z_i = T^{-1} \sum_{t=1}^{T} h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t}) - T^{-1} \sum_{t=1}^{T} E[h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t})] \]

in Lemma 5, we then have

\[
P \left\{ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{t=1}^{T} h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t}) - E[h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t})] \right\} \geq \frac{1}{2T^{1/4}} \}
\]

which leads

\[
\sum_{k_1=1}^{C_{\gamma_1}(\varepsilon)} \cdots \sum_{k_N=1}^{C_{\gamma_N}(\varepsilon)} \sum_{j_1=1}^{C_{\gamma_1}(\varepsilon)} \cdots \sum_{j_T=1}^{C_{\gamma_T}(\varepsilon)} P \left\{ \left| \frac{1}{N^2} \sum_{i=1}^{N} \sum_{t=1}^{T} h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t}) - E[h(\zeta_{it} | x_{it}, \gamma_{ik}, \bar{f}_{t})] \right| \geq \frac{1}{2T^{1/4}} \right\}
\]

\[
\leq C_{\gamma_1}(\varepsilon) \times \cdots \times C_{\gamma_N}(\varepsilon) \times C_{\gamma_1}(\varepsilon) \times \cdots \times C_{\gamma_T}(\varepsilon) \times \exp \left( -C \times N^{T^{1/2}} \right)
\]

\[
= O \left( \varepsilon^{-N(p+r)} \times \varepsilon^{-Tr} \times \exp \left[ -NT^{1/2} \right] \right)
\]

\[
= O \left( \exp \left[ -NT^{1/2} \left( 1 - \frac{N(p+r) \log(\varepsilon)}{NT^{1/2}} - \frac{Tr \log(\varepsilon)}{NT^{1/2}} \right) \right] \right)
\]

\[
= o(1),
\]

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

\[
(\hat{\gamma}_i - \gamma_{i,0}) = -\frac{1}{T} \sum_{t=1}^{T} \Gamma^{-1} E[g_{it}(0|\omega_{it})z_{it,0} \lambda_{i,0}] \Psi^{-1} \left( \frac{1}{N} \sum_{j=1}^{N} E[g_{jt}(0|\omega_{jt}) \lambda_{j,0} z_{jt,0}^j] (\hat{\gamma}_j - \gamma_{j,0}) \right) + O_p \left( \frac{1}{T^{1/4}} \right),
\]

which leads the following expression

\[
(\hat{\gamma} - \gamma_0) = \frac{1}{N} K(\hat{\gamma} - \gamma_0) + O_p(1/T^{1/4}),
\]

54
where \( \hat{\gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_N)', \( \gamma_0 = (\gamma_{1,0}, ..., \gamma_{N,0})', \) and

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

where \( K_{ij} = T^{-1} \sum_{t=1}^{T} \Gamma_{it}^{-1} J_{it} \Psi_{it}^{-1} J_{jt}^{T}, \) \( J_{it} = E[g_{it}(0)\omega_{it}]z_{it,0}\lambda_{i,0}' \). Because \( (I - \frac{1}{n} K) \) is positive definite matrix, we obtain \( \max_{1 \leq i \leq N} \| \hat{\gamma}_i - \gamma_{i,0} \| = O_p(1/T^{1/4}) \).

From \( \max_{1 \leq i \leq N} \| \hat{\gamma}_i - \gamma_{i,0} \| = O_p(1/T^{1/4}) \), we further can improve the convergence rate of the fifth term in (27). We show that

\[
P\left\{ \sup_{\gamma \in B_{\gamma}(T^{-1/4}), F \in B_{F}(N^{-1/4})} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ h_{it}(\zeta_{it}|x_{it},\gamma_{i},f_{i}) - E[h_{it}(\zeta_{it}|x_{it},\gamma_{i},f_{i})] \} \right\| \geq \frac{1}{T^{3/8}} \right\}
\]

converges to zero. Here \( \zeta_{it} = \epsilon_{it}, \gamma_{i} = (b_{i}', \lambda_{i}')', \)

\[
h(u|x, \gamma_{i}, f_{i}) = I(u \leq x'(b_{i} - b_{i,0}) + f_{i,0}' \lambda_{i,0} - f_{i}' \lambda_{i}) - I(u \leq 0),
\]

\( B_{\gamma}(T^{-1/4}) \equiv \{ \gamma; \| \gamma_{i} - \gamma_{i,0} \| \leq T^{-1/4}, \ i = 1, ..., N \} \), and

\( B_{F}(N^{-1/4}) \equiv \{ F; \| f_{t} - f_{t,0} \| \leq N^{-1/4}, \ t = 1, ..., T \} \).

For some positive constant \( C_{l} \), put \( M_{i} \equiv C_{l} \times (\| \lambda_{i} \| + \| \lambda_{i,0} \|), \alpha \equiv \sup_{i} M_{i} \) and \( \epsilon = T^{-3/8} \).

There exist a positive constant \( C \) and \( D_{\gamma_{i}} = [C \times (T^{2/3})^{p+r}] \) balls with centers \( \gamma_{i,k} \), \( k = 1, ..., D_{\gamma_{i}} \) and radius \( T^{-3/8}/(3\alpha) \) such that the collection of these balls covers \( B_{\gamma_{i}}(\kappa) \equiv \{ \gamma_{i}; \| \gamma_{i} - \gamma_{i,0} \| \leq T^{-1/4} \} \). 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 \( \bar{f}_{t,j} \) \( (j = 1, ..., D_{F_{t}}) \) and radius \( T^{-3/8}/(8\alpha) \) such that the collection of these balls covers \( B_{F_{t}}(N^{-1/4}) \).

Because

\[
\left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ h(\zeta_{it}|x_{it},\gamma_{i},f_{i}) - E[h(\zeta_{it}|x_{it},\gamma_{i},f_{i})] \} \right| 
\leq \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ h(\zeta_{it}|x_{it},\bar{\gamma}_{i,k},\bar{f}_{t,j}) - E[h(\zeta_{it}|x_{it},\bar{\gamma}_{i,k},\bar{f}_{t,j})] \} \right| + \left| \frac{T^{-3/8}}{8\alpha} \cdot \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ M_{i} + E[M_{i}] \} \right|
\]

we have

\[
P\left\{ \sup_{\gamma \in B_{\gamma}(\delta), F \in B_{F}(\epsilon)} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ h(\zeta_{it}|x_{it},\gamma_{i},f_{i}) - E[h(\zeta_{it}|x_{it},\gamma_{i},f_{i})] \} \right| \geq \frac{1}{T^{3/8}} \right\}
\]

\[
\leq \sum_{k_{1}=1}^{D_{\gamma_{1}}(\epsilon)} \cdots \sum_{k_{N}=1}^{D_{\gamma_{N}}(\epsilon)} \sum_{j_{1}=1}^{D_{F_{1}}(\epsilon)} \cdots \sum_{j_{T}=1}^{D_{F_{T}}(\epsilon)} P\left\{ \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ h(\zeta_{it}|x_{it},\bar{\gamma}_{i,k},\bar{f}_{t,j}) - E[h(\zeta_{it}|x_{it},\bar{\gamma}_{i,k},\bar{f}_{t,j})] \} \right| \geq \frac{1}{2T^{3/8}} \right\}
\]

\[
+ P\left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \{ M_{i} + E[M_{i}] \} \right| \geq 4\alpha \right\},
\]

55
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,

\[
E \left[ \frac{1}{T} \sum_{t=1}^{T} h(\zeta_{it}\mid x_{it}, \gamma_i, f_t) - \frac{1}{T} \sum_{t=1}^{T} E[h(\zeta_{it}\mid x_{it}, \gamma_i, f_t)] \right]^2 = \frac{1}{T^2} \sum_{t=1}^{T} E \left[ \left( h(\zeta_{it}\mid x_{it}, \gamma_i, f_t) - E[h(\zeta_{it}\mid x_{it}, \gamma_i, f_t)] \right)^2 \right] = O \left( \frac{1}{T} \times \frac{T^{1/2} + N^{1/2}}{N^{1/2}T^{1/2}} \right),
\]

where we used the result of Lemma 4 such that

\( E[\{ h(\zeta_{it}\mid x_{it}, \gamma_i, f_t) - E[h(\zeta_{it}\mid x_{it}, \gamma_i, f_t)] \}^2] = (T^{1/2} + N^{1/2})/(N^{1/2}T^{1/2}) \). Therefore,

\[
\sum_{i=1}^{N} E \left[ T^{-1} \sum_{t=1}^{T} h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j}) - T^{-1} \sum_{t=1}^{T} E[h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j})] \right]^2 \leq C \times \left( \frac{N^{1/2}(T^{1/2} + N^{1/2})}{T^{3/2}} \right),
\]

where \( C \) is some positive constant. Take \( Z_i = T^{-1} \sum_{t=1}^{T} h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j}) - T^{-1} \sum_{t=1}^{T} E[h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j})] \) in Lemma 5, we then have

\[
P \left\{ \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j}) - E[h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j})] \right| \geq \frac{1}{2T^{3/8}} \right\} \\
= P \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j}) - \frac{1}{T} \sum_{t=1}^{T} E[h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j})] \right) \right| \geq \frac{N}{2T^{3/8}} \right\} \\
\leq \exp \left( -C \times \frac{N^2}{T^{3/4}} \times \frac{T^{3/2}}{N^{1/2}(T^{1/2} + N^{1/2})} \right),
\]

where \( C \) is some positive constant. This leads

\[
\sum_{k_1=1}^{D_{\gamma_1}(\varepsilon)} \cdots \sum_{k_N=1}^{D_{\gamma_N}(\varepsilon)} \sum_{j_1=1}^{D_{f_1}(\varepsilon)} \cdots \sum_{j_T=1}^{D_{f_T}(\varepsilon)} P \left\{ \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j}) - E[h(\zeta_{it}\mid x_{it}, \gamma_{i_k}, f_{i_j})] \right| \geq \frac{1}{2T^{3/8}} \right\} \\
\leq D_{\gamma_1}(\varepsilon) \times \cdots \times D_{\gamma_N}(\varepsilon) \times D_{f_1}(\varepsilon) \times \cdots \times D_{f_T}(\varepsilon) \times \exp \left( -C \times \frac{N^3T^{3/4}}{N^{1/2} + T^{1/2}} \right)
\]
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

$$
\max_{1 \leq i \leq N} \| \hat{\gamma} - \gamma_0 \| = O_p(1/T^{1/4}),
$$

we obtain the claim $\max_{1 \leq i \leq N} \| \hat{\gamma}_i - \gamma_{i,0} \| = O_p(T^{-3/8})$.

By using $\max_{1 \leq i \leq N} \| \hat{\gamma}_i - \gamma_{i,0} \| = O_p(T^{-3/8})$, we can further strengthen this result to $\max_{1 \leq i \leq N} \| \hat{\gamma}_i - \gamma_{i,0} \| = O_p(T^{-7/16})$.

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

$$
P\left\{ \sup_{\gamma \in B_n(T^{-7/16}), F \in B_F(N^{-7/16})} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ h(u|\bar{x}_{it}, \bar{\gamma}_i, \bar{f}_t) - E[h(u|\bar{x}_{it}, \bar{\gamma}_i, \bar{f}_t)] \right\} \geq \frac{\eta}{T^{1/2}} \right\}
$$

converges to zero. Here $\zeta_{it}$ and $h(u|x_{it}, \gamma_{it}, f_t)$ are defined before, $B_n(T^{-7/16})$ and $B_F(N^{-7/16})$ are defined in the proof of Lemma 4. By using the same argument, we have

$$
P\left\{ \sum_{k_1=1}^{E_{T/16}(e)} \cdots \sum_{k_N=1}^{E_{T/16}(e)} \sum_{j_1=1}^{E_{T/16}(e)} \cdots \sum_{j_T=1}^{E_{T/16}(e)} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ h(u|\bar{x}_{it}, \bar{\gamma}_i, \bar{f}_t) - E[h(u|\bar{x}_{it}, \bar{\gamma}_i, \bar{f}_t)] \right\} \geq \frac{\eta}{T^{1/2}} \right\}
$$

and

$$
E \left\{ \left( \frac{1}{T} \sum_{t=1}^{T} h(\zeta_{it}|x_{it}, \gamma_{it}, f_t) - \frac{1}{T} \sum_{t=1}^{T} E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_t)] \right)^2 \right\} = O \left( \frac{1}{T} \times \frac{T^{7/8} + N^{7/8}}{N^{1/8}(T^{7/8} + N^{7/8})} \right).
$$

Therefore,

$$
\sum_{i=1}^{N} E \left\{ T^{-1} \sum_{t=1}^{T} h(\zeta_{it}|x_{it}, \gamma_{it}, f_t) - T^{-1} \sum_{t=1}^{T} E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_t)] \right\}^2 \leq C \left( \frac{N^{1/8}(T^{7/8} + N^{7/8})}{T^{15/8}} \right),
$$

where $C$ is some positive constant. Again, we take $Z_i$ as

$$
Z_i = T^{-1} \sum_{t=1}^{T} h(\zeta_{it}|x_{it}, \gamma_{it}, f_t) - T^{-1} \sum_{t=1}^{T} E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_t)]
$$

in Lemma 5, we then have

$$
P\left\{ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} h(\zeta_{it}|x_{it}, \gamma_{it}, f_t) - E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_t)] \right\} \geq \frac{\eta}{2T^{1/2}} \right\}
$$

and

$$
\sum_{k_1=1}^{E_{T/16}(e)} \cdots \sum_{k_N=1}^{E_{T/16}(e)} \sum_{j_1=1}^{E_{T/16}(e)} \cdots \sum_{j_T=1}^{E_{T/16}(e)} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ h(\zeta_{it}|x_{it}, \gamma_{it}, f_t) - E[h(\zeta_{it}|x_{it}, \gamma_{it}, f_t)] \right\} \geq \frac{1}{2T^{1/2}} \right\}
$$

57
Lemma 4, we can obtain the following expression. where the last line is obtained by using $\sqrt{T}/N^{1-\gamma} \to 0$ for small value of $\gamma$ $(1/16 < \gamma)$. 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)

$$
\hat f_t - f_{t,0}
\leq \frac{D_{\gamma_1}(\varepsilon) \times \cdots \times D_{\gamma_N}(\varepsilon) \times D_{f_1}(\varepsilon) \times \cdots \times D_{f_T}(\varepsilon) \times \exp \left( -C_n \times \frac{N^{15/8}T^{7/8}}{N^{7/8} + T^{7/8}} \right)
= O \left( \left( \frac{T^{1/16}}{N^{7/16}} \right)^{N(p+r)} \times \left( \frac{T^{1/2}}{N^{7/16}} \right)^{T^r} \times \exp \left[ \frac{-N^{15/8}T^{7/8}}{N^{7/8} + T^{7/8}} \right] \right)
= O \left( \exp \left[ \left( \frac{N^{15/8}T^{7/8}}{N^{7/8} + T^{7/8}} \right) \left\{ 1 - \frac{\{N(p+r)\}(N^{7/8} + T^{7/8}) \log(T^{1/16})}{N^{15/8}T^{7/8}} \right\} \right) \right)
= O \left( \exp \left[ \left( \frac{N^{15/8}T^{7/8}}{N^{7/8} + T^{7/8}} \right) \left\{ 1 - \frac{\log(T^{1/16})}{T^{7/8}} - \frac{\log(T^{1/16})}{N^{7/8}} - \frac{T^{1/2}/N^{7/16}}{N} \right\} \right) \right)
= o(1),
$$

where the last line is obtained by using $\sqrt{T}/N^{1-\gamma} \to 0$ for small value of $\gamma$ $(1/16 < \gamma)$. 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)

$$
\hat f_t - f_{t,0}
= \psi_t^{-1} V_{NT,t}(\gamma_0, f_{t,0}) - \psi^{-1} \left\{ V_{NT,t}(\gamma_0, \hat f_t) - V_{NT,t}(\gamma_0, f_{t,0}) \right\}
+ \frac{1}{N} \sum_{i=1}^{N} \psi_t^{-1} E[g_{it}(0|\omega_{it})\lambda_{i,0}z'_{i,t,0}] \Gamma_i^{-1} \left\{ \frac{1}{T} \sum_{t=1}^{T} E[g_{it}(0|\omega_{it})z_{i,t,0}\lambda_{i,0}] (\hat f_t - f_{t,0}) \right\}
+ \frac{1}{N} \sum_{i=1}^{N} \psi_t^{-1} E[g_{it}(0|\omega_{it})\lambda_{i,0}z'_{i,t,0}] \Gamma_i^{-1} Q_{NT,i}(\gamma_{i,0}, F_0)
+ \frac{1}{N} \sum_{i=1}^{N} \psi_t^{-1} E[g_{it}(0|\omega_{it})\lambda_{i,0}z'_{i,t,0}] \Gamma_i^{-1} \left\{ Q_{NT,i}(\hat \gamma_i, \hat \lambda_i) - Q_{NT,i}(\gamma_{i,0}, F_0) \right\}
+ O_p \left( \frac{1}{N} \right) + O_p \left( \frac{1}{T} \right) + o_p(\|\hat f_t - f_{t,0}\|).
$$

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

## D  Proof of Theorem 2

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

$$
\sqrt{T}(\hat \gamma_t - \gamma_{t,0}) = \Gamma_i^{-1} Q_{NT,i}(\gamma_{i,0}, F_0) - \frac{1}{T} \sum_{i=1}^{T} \Gamma_i^{-1} J_{\alpha} \psi_t^{-1} \left( \frac{1}{N} \sum_{j=1}^{N} J_{jt}(\hat \gamma_j - \gamma_{j,0}) \right) + o_p(1).
$$
This part of analysis is similar to Song (2013) and Ando and Bai (2015). We have

\[
\sqrt{T}(\hat{\gamma} - \gamma_0) = \frac{1}{N} \sqrt{T}K(\hat{\gamma} - \gamma_0) + \sqrt{T}\eta + o_p(1),
\]

where \( \hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_N)', \gamma_0 = (\gamma_{1,0}, \ldots, \gamma_{N,0})', \eta = (\eta'_1, \ldots, \eta'_N)' \) with \( \eta_i = \Gamma_i^{-1}Q_{NT,i}(\gamma_{i,0}, F_0) \), and \( K \) is defined in (33). Solving (37) in terms of \( \sqrt{T}(\hat{\gamma} - \gamma_0) \), we have

\[
\sqrt{T}(\hat{\gamma} - \gamma_0) = \left( I - \frac{1}{N} K \right)^{-1} \left( \sqrt{T}\eta + o_p(1) \right) = \left( I + \frac{1}{N} K + \frac{1}{N^2} K^2 + \frac{1}{N^3} K^3 + \cdots \right) \left( \sqrt{T}\eta + o_p(1) \right).
\]

Note that we can ignore the higher order terms related to \( \sqrt{T}N^{-\alpha}K\eta \) 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

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

We thus see that the asymptotic distribution of \( T^{1/2}(\hat{\gamma}_i - \gamma_{i,0}) \) is normal with mean zero and variance-covariance matrix \( \Sigma_i \).

It is obvious that we can again employ the similar argument that employed to derive the asymptotic distribution of \( T^{1/2}(\hat{\gamma}_i - \gamma_{i,0}) \). We then have

\[
N^{1/2}(\hat{f}_t - f_{t,0}) = \Psi_t^{-1} \left( N^{1/2}V_{NT,t}(\gamma_{0,0}, f_{t,0}) \right) + o_p(1),
\]

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

\section{Proof of Theorem 3}

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

\textbf{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 \( Q(\tau | x_{it}, f_{t,0}, \lambda_{i,0}) \) with the true dimension of the interactive effects \( r_0 \) is given as

\[
Q(\tau | x_{it}, f_{t,0}, \lambda_{i,0}) = x_{it}'b_{t,0} + f_{t,0}'\lambda_{i,0}.
\]
where the dependency of \( \tau \) is dropped from \( b_{i,t,\tau}, f_{i,t,\tau} \) and \( \lambda_{i,0,\tau} \). The true parameters \( \{B_0, \Lambda_0, F_0\} \) are minimizer of the loss function:

\[
\ell_{NT}(B, \Lambda, F) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_{\tau}(y_{it} - x'_{it}b_{i} - f'_{i}\lambda_{i}) - \rho_{\tau}(y_{it} - x'_{it}b_{i,0} - f'_{i,0}\lambda_{i,0})
\]

\[
= \mathbb{E} \left[ \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \int_{0}^{1} \left( x'_{it}(B_{i}-b_{i,0}) + (f'_{i}\lambda_{i}-f'_{i,0}\lambda_{i,0}) \right) \left\{ G_{it}(s|x_{it}, f_{i,0}, \lambda_{i,0}) - \tau \right\} ds \right],
\]

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 \( \Lambda_0 \).

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

\[
Q(\tau|x_{it}, f_{i,0}, \lambda_{i,0}) = x'_{it}b_{i,0} + (H_r f_{i,0})'(H_r^+ \lambda_{i,0}).
\]

It is clear that these transformed true parameters \( F_0(r) = (f_{i,0}(r), ..., f_{T,0}(r))' \) and \( \Lambda_0(r) = (\lambda_{1,0}(r), ..., \lambda_{N,0}(r))' \) with \( f_{i,0}(r) \equiv H_r f_{i,0} \) and \( \lambda_{i,0}(r) \equiv H_r^+ \lambda_{i,0} \) together with \( B_0 \), will let the loss \( \ell_{NT}(B, \Lambda, F) \) 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 \( \gamma_i(r) = (\dot{b}_{i}(r), \dot{\lambda}_{i}(r)) \) and \( \dot{f}_i(r) \) 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

\[
\max_{1 \leq i \leq N} \|\gamma_i(r) - \gamma_{i,0}(r)\| = O_p(\log(N)/\sqrt{T}) \quad \text{and} \quad \max_{1 \leq i \leq T} \|\dot{f}_i(r) - f_{i,0}(r)\| = O_p(\log(T)/\sqrt{N}).
\]

Here \( \gamma_{i,0}(r) = (b_{i,0}', \lambda_{i,0}(r)')' \). Note that \( b_{i,0} \) is the true parameter under the true model with the number of factors \( r_0 \).

Using Knight’s identity \( \rho_r(u - \nu) - \rho_r(u) = -\nu \psi_r(u) + \int_{0}^{u} (I(u \leq s) - I(u \leq 0))ds \) with \( \psi_r(u) = \tau - I(u \leq 0) \), we express

\[
V(r)
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_{\tau}(y_{it} - x'_{it}\dot{b}_{i}(r) - \dot{f}_i(r)\dot{\lambda}_{i}(r))
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_{\tau}(y_{it} - x'_{it}b_{i,0} - f'_{i,0}\lambda_{i,0}(r)) - \rho_{\tau}(y_{it} - x'_{it}b_{i,0} - f'_{i,0}\lambda_{i,0}(r))
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_{\tau}(\varepsilon_{it}) + \sum_{t=1}^{T} \sum_{i=1}^{N} (x'_{it}(b_{i}(r) - b_{i,0}) + \dot{f}_i(r)\dot{\lambda}_{i}(r) - f'_{i,0}(r)\lambda_{i,0}(r)) \psi_{\varepsilon_{t,i}}
+ \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \int_{0}^{1} (x'_{it}(b_{i}(r) - b_{i,0}) + \dot{f}_i(r)\dot{\lambda}_{i}(r) - f'_{i,0}(r)\lambda_{i,0}(r)) \left(I(\varepsilon_{t,i} \leq s) - I(\varepsilon_{t,i} \leq 0)\right)ds
= \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_{\tau}(\varepsilon_{it}) + I_1(r) + I_2(r).
\]
First, we investigate $I_2(r)$. For a notational simplicity, we denote $\min\{N, T\}$ as $C_{NT}$.

\[
(NT) \times I_2(r) \\
= \sum_{t=1}^{T} \sum_{i=1}^{N} \int_{0}^{s} \left( x_{it} - \hat{b}_i - \hat{f}_t \right) + \left( \hat{f}_t - f_{t,0} \right) \hat{\lambda}_i \left( r \right) + f_{t,0} \hat{\lambda}_i \left( r \right) - \lambda_i \left( 0 \right) \left( I_{\varepsilon_{t,i}} \geq s \right) - I_{\varepsilon_{t,i}} \left( 0 \right) \right) ds \\
= \frac{\log N}{C_{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[ \int_{0}^{s} \left( x_{it} - \hat{b}_i - \hat{f}_t \right) + \left( \hat{f}_t - f_{t,0} \right) \hat{\lambda}_i \left( r \right) + f_{t,0} \hat{\lambda}_i \left( r \right) - \lambda_i \left( 0 \right) \right] ds \\
= \log N \sum_{t=1}^{T} \sum_{i=1}^{N} \left[ \int_{0}^{s} \left( x_{it} - \hat{b}_i - \hat{f}_t \right) + \left( \hat{f}_t - f_{t,0} \right) \hat{\lambda}_i \left( r \right) + f_{t,0} \hat{\lambda}_i \left( r \right) - \lambda_i \left( 0 \right) \right] ds + \omega_p \left( N \right) \\
= \log N \sum_{t=1}^{T} \sum_{i=1}^{N} \left[ \int_{0}^{s} \left( x_{it} - \hat{b}_i - \hat{f}_t \right) + \left( \hat{f}_t - f_{t,0} \right) \hat{\lambda}_i \left( r \right) + f_{t,0} \hat{\lambda}_i \left( r \right) - \lambda_i \left( 0 \right) \right] ds + \omega_p \left( N \right) \\
= C \times \sum_{t=1}^{T} \sum_{i=1}^{N} \left( x_{it} - \hat{b}_i - \hat{f}_t \right) + \left( \hat{f}_t - f_{t,0} \right) \hat{\lambda}_i \left( r \right) + f_{t,0} \hat{\lambda}_i \left( r \right) - \lambda_i \left( 0 \right) \right] ds \\
= \omega_p \left( T \right) \\
\]

where we used $N^{-1} \sum_{i=1}^{N} \| \hat{x}_{it} - \hat{x}_i \| \leq \omega_p \left( T^{-1} \right)$,

$N^{-1} \sum_{i=1}^{N} \| \hat{x}_i - x_i \| \leq \omega_p \left( T^{-1} \right)$,

$N^{-1} \sum_{i=1}^{N} \| \hat{\lambda}_i - \lambda_i \| \leq \omega_p \left( T^{-1} \right)$,

and $0 < \omega_p \left( N \right) < \infty$ for $t = 1, \ldots, N$.

Next, we evaluate the term $I_1(r)$. Noting that $\sum_{i=1}^{N} \| \hat{x}_{it} - \hat{x}_i \| \leq \omega_p \left( T^{-1} \right)$,

$\sum_{i=1}^{N} \| \hat{\lambda}_i - \lambda_i \| \leq \omega_p \left( T^{-1} \right)$,

and $\sum_{t=1}^{T} \| \hat{\lambda}_i - \lambda_i \| \leq \omega_p \left( T^{-1} \right)$,

we have

\[
(NT) \times I_1(r) \leq C \times \omega_p \left( T^{-1} \right) \\
= \omega_p \left( \sqrt{NT/C_{NT}^2} \right) \\
= \omega_p \left( \sqrt{NT/C_{NT}^2} \right) \\
\]

Thus, we obtain

\[
V(r) - V(r_0) = \omega_p \left( C_{NT}^{-1} \right). \\
\]

Using the same argument of the proof of Corollary 1 in Bai and Ng (2002), for $r > r_0$, this implies that $V(r)/V(r_0) = 1 + \omega_p \left( 1/C_{NT} \right)$. Thus, $\log(V(r)/V(r_0)) = \omega_p \left( C_{NT}^{-1} \right)$. Because
(r - r_0)q(N, T) \geq q(N, T), which converges to zero slower rate than O_p\left(1/C_{NT}\right), it follows that

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

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:

\[
\bar{\ell}_{NT}(B(r), \Lambda(r), F(r)) = \frac{1}{NT} E \left[ \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_r \left( y_{it} - \bar{x}_{it}^\prime \hat{b}_i(r) - \hat{f}_t(r)^\prime \hat{\lambda}_i(r) \right) - \rho_r \left( y_{it} - \bar{x}_{it}^\prime \hat{b}_i(0) - \hat{f}_t(0)^\prime \hat{\lambda}_i(0) \right) \right]
\]

\[
= \frac{1}{NT} E \left[ \sum_{t=1}^{T} \sum_{i=1}^{N} \int_0^{\infty} x_{it}^\prime (b_i(r) - b_i(0)) + (f_t(r) - f_t(0)) \right. \left\{ g_{it}(s|x_{it}, f_t(0), \lambda_i(0) - \tau) \right\} ds \right].
\]

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

\[
\lim_{N,T \rightarrow \infty} \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[ \rho_r \left( y_{it} - \bar{x}_{it}^\prime \hat{b}_i(r) - \hat{f}_t(r)^\prime \hat{\lambda}_i(r) \right) - \rho_r \left( y_{it} - \bar{x}_{it}^\prime \hat{b}_i(0) - \hat{f}_t(0)^\prime \hat{\lambda}_i(0) \right) \right] > C,
\]

where \( \hat{b}_i(r) \), \( \hat{f}_t(r) \) and \( \hat{\lambda}_i(r) \) 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 \( q(N, T) \rightarrow 0 \), 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

\[
P\left( IC(r) - IC(r_0) < 0 \right) \rightarrow 0.
\]

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 \( \{ (\hat{b}_{i,\tau}', \hat{\lambda}_{i,\tau}') ; \ i = 1, ..., N \} \) 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 \( \{ (\hat{b}_{i,0.05}', \hat{\lambda}_{i,0.05}') ; \ i = 1, ..., N \} \), 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 \( \{ (\hat{b}_{i,0.95}', \hat{\lambda}_{i,0.95}') ; \ i = 1, ..., N \} \). 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 \( \{ (\hat{b}_{i,0.05}', \hat{\lambda}_{i,0.05}') ; \ i = 1, ..., N \} \), 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 \( \{ (\hat{b}_{i,0.95}', \hat{\lambda}_{i,0.95}') ; \ i = 1, ..., N \} \). 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 $x_{i,t}$ are correlated with the unobservable factor structures $f'_i \lambda_{i,t}$.

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 $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Then, we generate the data from the following structure:

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

where $x_{it} = (x_{it,1}, \ldots, x_{it,T})'$ 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 $\tau$ is

$$ Q_{y_{it}} (\tau | x_{it}, f(\tau), \lambda(\tau)) = G^{-1}(\tau) + x'_{it} b_{i}(\tau) + f(\tau)' \lambda_{i}(\tau). $$

The quantile restriction $P(y_{it} - Q_{y_{it}} (\tau | \omega_{it}) \leq 0) = \tau$ is satisfied.

We generate $T \times 5$ common factor matrix $F = (f_{it})$ 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

$$ f_{i,u_{it}} = \begin{cases} (f_{t1}, f_{t2}, f_{t3})' & \text{if } u_{it} \leq 0.2 \\ (f_{t1}, f_{t2}, f_{i3}, f_{t4})' & \text{if } 0.2 < u_{it} \leq 0.8 \\ (f_{t1}, f_{t2}, f_{i3}, f_{t4}, f_{t5})' & \text{if } 0.8 < u_{it} \end{cases}. $$

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

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

where $\zeta_{ik}$ 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 $i = 1, \ldots, N$, we generate the set of regressors as follows:

\[
\begin{align*}
    x_{it,1} &= v_{it,1} + 0.02f_{1t}^2 + 0.02\zeta_{i1}^2, \\
    x_{it,3} &= v_{it,3} - 0.01f_{2t}^2 + 0.02\zeta_{i2}^2, \\
    x_{it,5} &= v_{it,5} - 0.01f_{3t}^2 + 0.03\zeta_{i3}^2, \\
    x_{it,k} &= v_{it,k} \quad (k \neq 1, 3, 5),
\end{align*}
\]

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

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

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 ~ Table 5. Tables show the mean squared error (MSE) between the true structure and the estimates.

\[
\begin{align*}
    \text{MSE}_1 &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2, \\
    \text{MSE}_2 &= \frac{1}{Np} \sum_{i=1}^{N} \|b_{i,0,\tau} - \hat{b}_{i,\tau}\|^2,
\end{align*}
\]

where $\hat{b}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $b_{i,\tau,0}$ and the true $\tau$-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

\[
\ell_\tau(Y | X, B_\tau) = \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - x_{it}'b_{i,\tau}).
\]

Table 4 ~ 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 $\tau$.

Figure 6 shows the boxplots of the average mean squared error between the true parameter vectors $b_{i,\tau,0}$ and its estimates $\hat{b}_{i,\tau}$ over $i$; $\text{MSE} = \frac{1}{Np} \sum_{i=1}^{N} \|b_{i,0,\tau} - \hat{b}_{i,\tau}\|^2$. 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 frequentist 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 frequentist 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 frequentist 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


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).

<table>
<thead>
<tr>
<th>Financial Firm</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amman</td>
<td>68</td>
<td>68</td>
<td>67</td>
</tr>
<tr>
<td>Australian</td>
<td>99</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Bangkok</td>
<td>86</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>Berlin</td>
<td>140</td>
<td>144</td>
<td>146</td>
</tr>
<tr>
<td>Borsa Istanbul</td>
<td>46</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>BSE Ltd</td>
<td>285</td>
<td>286</td>
<td>277</td>
</tr>
<tr>
<td>Copenhagen Stock Exchange</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>Dhaka</td>
<td>69</td>
<td>68</td>
<td>69</td>
</tr>
<tr>
<td>Euronext.liffe Paris</td>
<td>83</td>
<td>83</td>
<td>81</td>
</tr>
<tr>
<td>Frankfurt</td>
<td>531</td>
<td>536</td>
<td>529</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>203</td>
<td>202</td>
<td>199</td>
</tr>
<tr>
<td>Indonesia</td>
<td>76</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>Johannesburg</td>
<td>40</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>Karachi</td>
<td>56</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>Korea Stock Exchange</td>
<td>53</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>Kuala Lumpur</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>Kuwait City</td>
<td>74</td>
<td>72</td>
<td>74</td>
</tr>
<tr>
<td>London</td>
<td>144</td>
<td>146</td>
<td>149</td>
</tr>
<tr>
<td>Milan</td>
<td>44</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>NASDAQ</td>
<td>374</td>
<td>375</td>
<td>370</td>
</tr>
<tr>
<td>National India</td>
<td>103</td>
<td>101</td>
<td>103</td>
</tr>
<tr>
<td>New York Stock Exchange (NYSE)</td>
<td>186</td>
<td>187</td>
<td>183</td>
</tr>
<tr>
<td>Non NASDAQ OTC</td>
<td>1370</td>
<td>1364</td>
<td>1289</td>
</tr>
<tr>
<td>OTC Bulletin Board</td>
<td>49</td>
<td>51</td>
<td>52</td>
</tr>
<tr>
<td>Philippine Stock Exchange</td>
<td>82</td>
<td>82</td>
<td>81</td>
</tr>
<tr>
<td>Santiago</td>
<td>42</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>Shanghai</td>
<td>73</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Shenzen</td>
<td>55</td>
<td>55</td>
<td>54</td>
</tr>
<tr>
<td>Singapore</td>
<td>49</td>
<td>49</td>
<td>47</td>
</tr>
<tr>
<td>SIX Swiss</td>
<td>68</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>Stockholm</td>
<td>45</td>
<td>44</td>
<td>44</td>
</tr>
<tr>
<td>Stuttgart</td>
<td>63</td>
<td>61</td>
<td>63</td>
</tr>
<tr>
<td>Taiwan</td>
<td>61</td>
<td>61</td>
<td>60</td>
</tr>
<tr>
<td>Tel Aviv</td>
<td>107</td>
<td>105</td>
<td>106</td>
</tr>
<tr>
<td>Thailand</td>
<td>58</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>Tokyo Stock Exchange</td>
<td>177</td>
<td>175</td>
<td>177</td>
</tr>
<tr>
<td>Toronto</td>
<td>81</td>
<td>80</td>
<td>77</td>
</tr>
<tr>
<td>TSX Ventures</td>
<td>64</td>
<td>60</td>
<td>56</td>
</tr>
<tr>
<td>XETRA</td>
<td>83</td>
<td>83</td>
<td>82</td>
</tr>
<tr>
<td>Others</td>
<td>781</td>
<td>785</td>
<td>780</td>
</tr>
</tbody>
</table>
Table 4: First data-generating process. Comparison of the proposed estimator with factor structure $f'_{t,\tau} \lambda_{i,\tau}$, and the standard estimator without factor structure $f'_{t,\tau}$. The mean squared errors are defined as $\text{MSE}_1 = \frac{1}{N_T} \sum_{i=1}^{N} \sum_{t=1}^{T} \{Q_{it}(\tau) - \hat{Q}_{it}(\tau)\}^2$, $\text{MSE}_2 = \frac{1}{N_p} \sum_{i=1}^{N} \| b_{i,0,\tau} - \hat{b}_{i,\tau} \|^2$, where $\hat{b}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $b_{i,\tau,0}$ and the true $\tau$-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.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$T$</th>
<th>$N$</th>
<th>With factor structure</th>
<th>Without factor structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\text{MSE}_1$</td>
<td>$\text{MSE}_2$</td>
</tr>
<tr>
<td>0.05</td>
<td>300</td>
<td>300</td>
<td>0.7140</td>
<td>0.2397</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0524</td>
<td>0.0137</td>
</tr>
<tr>
<td>0.05</td>
<td>300</td>
<td>900</td>
<td>0.7004</td>
<td>0.2389</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0451</td>
<td>0.0088</td>
</tr>
<tr>
<td>0.05</td>
<td>900</td>
<td>300</td>
<td>0.5649</td>
<td>0.0822</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0374</td>
<td>0.0039</td>
</tr>
<tr>
<td>0.05</td>
<td>900</td>
<td>900</td>
<td>0.5063</td>
<td>0.0820</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0130</td>
<td>0.0024</td>
</tr>
<tr>
<td>0.5</td>
<td>300</td>
<td>300</td>
<td>0.3382</td>
<td>0.1279</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0116</td>
<td>0.0053</td>
</tr>
<tr>
<td>0.5</td>
<td>300</td>
<td>900</td>
<td>0.3178</td>
<td>0.1305</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0117</td>
<td>0.0056</td>
</tr>
<tr>
<td>0.5</td>
<td>900</td>
<td>300</td>
<td>0.3054</td>
<td>0.0745</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0090</td>
<td>0.0049</td>
</tr>
<tr>
<td>0.5</td>
<td>900</td>
<td>900</td>
<td>0.2697</td>
<td>0.0766</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0049</td>
<td>0.0022</td>
</tr>
<tr>
<td>0.95</td>
<td>300</td>
<td>300</td>
<td>0.7438</td>
<td>0.2779</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0240</td>
<td>0.0140</td>
</tr>
<tr>
<td>0.95</td>
<td>300</td>
<td>900</td>
<td>0.6892</td>
<td>0.2651</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0220</td>
<td>0.0075</td>
</tr>
<tr>
<td>0.95</td>
<td>900</td>
<td>300</td>
<td>0.6171</td>
<td>0.0904</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0227</td>
<td>0.0041</td>
</tr>
<tr>
<td>0.95</td>
<td>900</td>
<td>900</td>
<td>0.5506</td>
<td>0.0890</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0169</td>
<td>0.0032</td>
</tr>
</tbody>
</table>
Table 5: Second data-generating process. Comparison of the proposed estimator with factor structure $f'_{t,\tau} \lambda_{i,\tau}$, and the standard estimator without factor structure $f'_{t,\tau}$. The mean squared errors are defined as $\text{MSE}_1 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ Q_{it}(\tau) - \hat{Q}_{it}(\tau) \}^2$, $\text{MSE}_2 = \frac{1}{Np} \sum_{i=1}^{N} \| b_{i,0,\tau} - \hat{b}_{i,\tau} \|^2$, where $\hat{b}_{i,\tau}$ and $\hat{Q}_{it}(\tau)$ are the estimates of the true parameter vector $b_{i,\tau,0}$ and the true $\tau$-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.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$T$</th>
<th>$N$</th>
<th>With factor structure</th>
<th>Without factor structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\text{MSE}_1$</td>
<td>$\text{MSE}_2$</td>
</tr>
<tr>
<td>0.05</td>
<td>300</td>
<td>300</td>
<td>0.9283</td>
<td>0.3532</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0400</td>
<td>0.0140</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>900</td>
<td>0.8691</td>
<td>0.3517</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0408</td>
<td>0.0121</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>300</td>
<td>0.6723</td>
<td>0.1209</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0330</td>
<td>0.0042</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>900</td>
<td>0.6012</td>
<td>0.1217</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0139</td>
<td>0.0028</td>
</tr>
<tr>
<td>0.5</td>
<td>300</td>
<td>300</td>
<td>0.3429</td>
<td>0.1338</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0110</td>
<td>0.0087</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>900</td>
<td>0.3283</td>
<td>0.1375</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0091</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>300</td>
<td>0.3062</td>
<td>0.0735</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0146</td>
<td>0.0041</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>900</td>
<td>0.2703</td>
<td>0.0761</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0041</td>
<td>0.0038</td>
</tr>
<tr>
<td>0.95</td>
<td>300</td>
<td>300</td>
<td>0.9736</td>
<td>0.4000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0403</td>
<td>0.0172</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>900</td>
<td>0.8970</td>
<td>0.3878</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0237</td>
<td>0.0081</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>300</td>
<td>0.7750</td>
<td>0.1309</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0283</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>900</td>
<td>0.6833</td>
<td>0.1300</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0217</td>
<td>0.0041</td>
</tr>
</tbody>
</table>
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 $\tau = 95\%$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>$T$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.05 \ (r_{0,\tau} = 3)$</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>900</td>
</tr>
</tbody>
</table>

| $0.95 \ (r_{0,\tau} = 5)$ | 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

<table>
<thead>
<tr>
<th>$\hat{r}$</th>
<th>$T$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.05 \ (r_{0,\tau} = 3)$</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>900</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>900</td>
</tr>
</tbody>
</table>

| $0.95 \ (r_{0,\tau} = 5)$ | 300 | 300 | 0 | 0 | 0 | 16 | 29 | 55 | 0 | 0 | 0 |
| | 300 | 900 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| | 900 | 300 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
| | 900 | 900 | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 |
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).
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).
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 frequentist 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.
Figure 6: Boxplot of the average mean squared errors: \( \text{MSE} = \frac{1}{N_p} \sum_{i=1}^{N} \left\| b_{t, \tau, 0} - \hat{b}_{t, \tau} \right\|^2 \), between the true parameter vector \( b_{t, \tau, 0} \) and its estimate \( \hat{b}_{t, \tau} \). (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.