Quantile Factor Models[*]

Liang Chen[1], Juan J. Dolado[2], and Jesús Gonzalo[3]

[1]*School of Economics, Shanghai University of Finance and Economics, chen.liang@mail.shufe.edu.cn*
[2]*Department of Economics, Universidad Carlos III de Madrid, dolado@eco.uc3m.es*
[3]*Department of Economics, Universidad Carlos III de Madrid, jgonzalo@est-eco.uc3m.es*

March 23, 2020

Abstract

Quantile Factor Models (QFM) represent a new class of factor models for high-dimensional panel data. Unlike Approximate Factor Models (AFM), where only location-shifting factors can be extracted, QFM also capture unobserved factors shifting other relevant parts of the distributions of observables. We propose a quantile regression approach, labeled Quantile Factor Analysis (QFA), to consistently estimate all the quantile-dependent factors and loadings. Their asymptotic distribution is derived using a kernel-smoothed version of the QFA estimators. Two consistent model-selection criteria, based on information criteria and rank minimization, are developed to determine the number of factors at each quantile. Moreover, in contrast to the conditions required by Principal Components Analysis in AFM, QFA estimation remains valid even when the idiosyncratic errors have heavy-tailed distributions. Three empirical applications (regarding climate, macroeconomic and finance panel data) illustrate that extra factors shifting quantiles other than the means could be relevant for causality analysis, prediction and economic interpretation of common factors.

Keywords: Factor models, quantile regression, incidental parameters.
JEL codes: C31, C33, C38.

1 Introduction

Following the key contributions by Ross (1976), Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986) to the theory of approximate factor models (AFM henceforth) in the context of asset pricing, the analysis and applications of this class of models have proliferated thereafter. As it is well known, AFM imply that a panel $\{X_{it}\}$ of N variables (units), each with T observations, has the representation $X_{it} = \lambda_i' f_t + \epsilon_{it}$, where $\lambda_i = [\lambda_{i1}, .., \lambda_{ir}]'$ and $f_t = [f_{t1}, .., f_{tr}]'$ are $r \times 1$ vectors of factor loadings and common factors, respectively, with $r \ll N$, and $\{\epsilon_{it}\}$ are zero-mean weakly dependent idiosyncratic disturbances which are uncorrelated with the factors.

The fact that it is easy to construct theories involving common factors, at least in a narrative version, together with the availability of fairly straightforward estimation procedures for AFM—e.g. via Principal Components Analysis (PCA), has led to their extensive use in many fields of economics.[1]

Inspired by the generalization of linear regression to quantile regression (QR) models, our starting point in this paper is to claim that the standard regression interpretation of static AFM as linear conditional mean models of X_{it} given f_t (i.e. $\mathbb{E}(X_{it}|f_t) = \lambda_i' f_t$), entails two possibly restrictive features. On the one hand, PCA does not capture hidden factors that may shift characteristics (moments or quantiles) of the distribution of X_{it} other than its mean. On the other hand, neither the loadings λ_i nor the factors f_t are allowed to vary across the distributional characteristics of each unit in the panel.

Highlighting these limitations, a growing literature in empirical finance has been documenting a much more pronounced co-movement of financial asset returns at the lower part than at the rest of their distributions. In particular, Amengual and Sentana (2018) reject the null hypothesis of a Gaussian copula when analyzing the cross-sectional dependence among monthly returns on individual US stocks, for which they find nonlinear tail dependence, co-skeweness and co-kurtosis. Likewise, Ando and Bai (2020) (AB 2020, hereafter) report that the common factor structures explaining the asset return distributions in global financial markets since the subprime crisis are different in the lower and the upper tails. In international finance, Maravalle and Rawdanowicz (2018) find that, while global and regional factors were key in explaining fluctuations of European government bond yields between the Great Moderation and the onset of the Great Recession (lower tails of the distributions of bond yields), their role decreased afterwards, giving way to

---

[1]Early applications of AFM abound in Aggregation Theory, Consumer Theory, Business Cycle Analysis, Finance, Monetary Economics, and Monitoring and Forecasting; see, *inter alia*, Bai (2003), Bai and Ng (2008b), Stock and Watson (2011). More recently, the characterization of cross-sectional dependence among error terms in Panel Data has relied on the use of a finite number of unobserved common factors which originate from economy-wide shocks affecting units with different intensities (loadings). Interactive fixed-effects models can be easily estimated by PCA (see Bai 2009) or by common correlated effects (see Pesaran 2006), and there are even generalizations of these techniques for nonlinear panel single-index models (see Chen et al. 2018). Lastly, the surge of Big Data and Machine Learning technologies has made factor models a key tool for dimension reduction and predictive analytics when using very large datasets (see Athey and Imbens 2019 for a survey).

country-specific factors as the main driving forces.[2] On the macro side, Adrian et al. (2019) find that, while the estimated lower conditional quantiles of the distribution of future GDP growth in the US exhibit strong dependence on current financial conditions, the upper quantiles are stable over time. Lastly, in micro theory, de Castro and Galvao (2019) have recently extended the expected utility model of rational behavior to quantile utility preferences, where e.g. factor structures determining hedonic pricing of consumption goods or financial stocks may exhibit large differences across quantiles.

A simple way of illustrating the above limitations of the conventional formulation of AFM is to consider the factor structure in a *location-scale shift* model with the following Data Generating Process (DGP): $X_{it} = \alpha_i f_{1t} + \eta_i f_{2t}\epsilon_{it}$, with $f_{1t} \neq f_{2t}$ (both are scalars), $\eta_i, f_{2t} > 0$ and $\mathbb{E}(\epsilon_{it}) = 0$. The first factor ($f_{1t}$) shifts location, whereas the second factor (f_{2t}) shifts the scale and therefore governs the volatility of shocks to X_{it}. This model has been proposed by Herskovic et al. (2016) to empirically document a strong co-movement of the volatilities in the idiosyncratic component of individual stock returns and firm-level cash flows.[3] Such a DGP can be rewritten in QR format as $X_{it} = \lambda_i'(\tau)f_t + u_{it}(\tau)$, with $0 < \tau < 1$, $\lambda_i(\tau) = [\alpha_i, \eta_i \mathsf{Q}_\epsilon(\tau)]'$, where $\mathsf{Q}_\epsilon(\tau)$ represents the quantile function of ϵ_{it}, $f_t = [f_{1t}, f_{2t}]'$, $u_{it}(\tau) = \eta_i f_{2t}[\epsilon_{it} - \mathsf{Q}_\epsilon(\tau)]$, and the conditional quantile $\mathsf{Q}_{u_{it}(\tau)}[\tau|f_t] = 0$.[4] PCA will only extract the location-shifting factor f_{1t} in this model, but it will fail to capture the scale-shifting factor f_{2t} and the quantile-dependent loadings $\lambda_i(\tau)$ in its QR representation. As will be explained below, our estimation procedure allows to estimate the space spanned by f_{1t} and f_{2t}.[5] Also notice that, when the distribution of ϵ_{it} is symmetric, then f_t can be considered as being quantile dependent, i.e. $f_t(\tau)$, since $f_t(\tau) = f_{1t}$ for $\tau = 0.5$, and $f_t(\tau) = [f_{1t}, f_{2t}]'$ for $\tau \neq 0.5$. Together with the remaining examples discussed below, this means that the general class of models to be considered in the sequel would be one where the loadings, factors and the number of factors are all allowed to be quantile-dependent objects, namely, $\lambda_i(\tau)$, $f_t(\tau)$ and $r(\tau)$ for $\tau \in (0,1)$. In what follows, we coin this class of models *Quantile Factor Models* (QFM, hereafter), whose detailed definition is provided in Section 2 below.

That said, our goal in this paper is to develop a common factor methodology for QFM which is flexible enough to capture the quantile-dependent objects that standard AFM tools fail to recover. To do so, we analyze their estimation and inference, including selection criteria for the number of factors at each quantile τ. Put succinctly, QFM could be thought of as capturing the same type of flexible generalization that QR techniques represent for linear regression models.

To help understand how this new methodology works, we start by proposing an estimation

---

[2] The diminishing role of the former factors could result from quantitative easing and the increase in yields (upper tail of the distribution), while the emergence of the latter factors could be due to the financial fragmentation in a number of vulnerable euro-area countries during sovereign debt crisis.

[3] This DGP is further discussed in subsection 2.2 below, where we present a larger set of illustrative models as examples of potential DGPs for X_{it}. Notice that the simplifying assumption of a known number of factors in this specific example is later relaxed.

[4] Throughout the paper we use $\mathsf{Q}_W[\tau|Z]$ to denote the conditional quantile of W given Z.

[5] Given that f_{1t} can be consistently estimated by PCA, it is also feasible to separate f_{2t} from their joint space.

approach for the quantile-dependent objects in QFM, labeled *Quantile Factor Analysis* (QFA, henceforth). Our QFA estimation procedure relies on the minimization of the standard *check* function in QR (instead of the conventional quadratic loss function used in AFM) to estimate jointly the common factors $f_t(\tau)$ and the loadings $\lambda_i(\tau)$ at a given quantile τ, once the number of factors has been selected. However, since the objective function for QFM is not convex in the relevant parameters, we introduce an iterative QR algorithm which yields estimators of the quantile-dependent objects. We then derive their average rates of convergence, and propose two consistent selection criteria (one based on information criteria and another on rank minimization) for the number of factors at each τ. In addition, we establish asymptotic normality for QFA estimators based on smoothed QR (see e.g., Horowitz 1998 and Galvao and Kato 2016). Moreover, given that QFA estimation captures all quantile-shifting factors (including those affecting the means of observed variables), our asymptotic results and the proposed selection criteria provide a natural way to differentiate AFM from QFM.

In sum, the key contributions of this paper to the literature on factor models can be summarized as follows:

1. We provide a complete asymptotic analysis for a new class of factor models: QFM. In particular, we show that the average convergence rates of the QFA estimators are the same as the PCA estimators of Bai and Ng (2002) (BN 2002, hereafter), which is a crucial result for proving the consistency of the two selection criteria of the number of factors at each τ. In addition, similar to Bai (2003), our QFA estimators based on smoothed QR are shown to converge at the parametric rates ($\sqrt{N}$ and $\sqrt{T}$) to normal distributions.

2. We argue that the problems of incidental parameters and non-smooth object functions require the use of some novel techniques in our proofs, which are borrowed from the theory of empirical processes. Moreover, our proof strategy can be easily extended to some other nonlinear factor models (e.g., probit and logit factor models considered by Chen et al. 2018) with smooth object functions. Finally, as a byproduct of our approach (and in exchange for some restrictions on the dependence of the idiosyncratic errors in an AFM; see Assumption 1 below), it is shown that the QFA estimators inherit from QR certain robustness properties to the presence of outliers and heavy-tailed distributions in those error terms, which would render PCA invalid.

3. We show in the empirical section how QFA could provide a useful tool for quantile causal analysis, density forecasting, and economic interpretation of factors by applying the proposed methodology to three different datasets related to climate, macroeconomic aggregates and stock returns.

Related literature

There is a recent literature that attempts to make the AFM setup more flexible. For example,

Su and Wang (2017) allow for the factor loadings to be time-varying and Pelger and Xiong (2018) admit these loadings to be state dependent. Chen et al. (2009) provide a theory for nonlinear PCA, where they favor sieve estimation to retrieve nonlinear factors. Finally, Gorodnichenko and Ng (2017) propose an algorithm to estimate level and volatility factors simultaneously. Different from these studies, our approach to modelling nonlinearities in factor models is through the conditional quantiles of the observed data.

On top of this, there is an emerging literature on heterogeneous panel quantile models with factor structures, especially in financial economics. The main idea is that a few unobservable factors explain co-movements of asset return distributions in a large range of asset returns observed at high frequencies, as in stock markets. In parallel and independent research, there have been two papers related to ours. First, Ma et al. (2019) propose estimation and inference procedures in semiparametric quantile factor models. In these models, factor loadings/betas are smooth functions of a small number of observables under the assumption that the included factors all have non zero mean. Then, sieve techniques are used to obtain preliminary estimation of these functions for each time period. Finally, the factor structure is imposed in a sequential fashion to estimate the factor returns by GLS under weak conditions on cross-sectional and temporal dependence. We depart from these authors in that we do not need to assume the loadings to depend on observables and, foremost, in that not only loadings but also factors are quantile-dependent objects in our setup. Second, in a closely related paper, AB (2020) use a similar setup to ours, where the unobservable factor structure is also allowed to be quantile dependent. These authors use Bayesian MCMC and frequentist estimation approaches, the latter building upon our proposed iterative procedure, as duly acknowledged in their paper. However, we differ from AB (2020) in several respects which make our QFA approach valuable: (i) our assumptions are less restrictive, since we rely on properties of the density, as in QR, while AB (2020) needs all the moments of the idiosyncratic errors to exist, (ii) our proofs of the main results are different from theirs, and (iii) our rank-minimization selection criterion to estimate the number of factors is novel, behaves well in finite samples and is computationally more efficient than the information criteria-based method.

Finally, it is noteworthy that the illustrative location-scale shift model above, where $f_{1t} \neq f_{2t}$, is behind a current line of research in asset pricing which has been coined the "idiosyncratic volatility puzzle" by Ang et al. (2006). This approach focuses on the co-movements in the idiosyncratic volatilities of a panel of asset returns, and consists of applying PCA to (or taking cross-sectional averages of) the squared residuals, once the mean (PCA) factors have been removed from the original variables (a procedure labeled PCA-SQ hereafter).[6] For example, this technique would be valid for our illustrative example above. Yet, while the QFA approach is

---

[6]See, e.g., Barigozzi and Hallin (2016), Herskovic et al. (2016) and Renault et al. (2017). Notice that the volatility co-movement does not arise from omitted factors in the AFM but from assuming a genuine factor structure in the idiosyncratic volatility processes.

able to recover the whole QFM structure for more general DGPs than the previous model (see subsection 2.2), PCA-SQ fails to do so. It will also fail when the idiosyncratic errors do not have bounded eighth moments. Hence, to the best of our knowledge, our QFA approach becomes the first estimation procedure capable of dealing with these issues.

Structure of the Paper

The rest of the paper is organized as follows. Section 2 defines QFM and provides a list of simple illustrative examples where the new QFM methodology applies. In Section 3, we present the QFA estimator and its computational algorithm, establish the average rates of convergence of the quantile-dependent factors and factor loadings, and propose two consistent selection criteria to choose the number of factors at each quantile. Section 4 introduces a kernel-smoothed version of the QFA estimators to derive their asymptotic distributions. Section 5 contains some Monte Carlo simulation results to evaluate the performance in finite samples of our estimation procedures relative to other alternative approaches with different assumptions about the idiosyncratic error terms. Section 6 considers three empirical applications using three large panel datasets, where we document the relevance of extra factors in causal analysis, forecasting and economic interpretation of common factors. Finally, Section 7 concludes and suggests several avenues for further research. Proofs of the main results are collected in the Online Appendix.

Notations

The Frobenius norm is denoted as $\| \cdot \|$. For a matrix A with real eigenvalues, $\rho_j(A)$ denotes the jth largest eigenvalue. Following van der Vaart and Wellner (1996), the symbol $\lesssim$ means "left side bounded by a positive constant times the right side" (the symbol $\gtrsim$ is defined similarly), and $D(\cdot, g, \mathcal{G})$ denotes the packing number of space $\mathcal{G}$ endowed with semimetric g.

2   The Model and Some Illustrative Examples

This section starts by introducing the main definitions to be used throughout the paper. Next, we show how to derive the QFM representation of several illustrative DGPs exhibiting different factor structures.

2.1   Quantile Factor Models

Suppose that the observed variable X_{it}, with $i = 1, 2, .., N$ and $t = 1, 2, ..., T$, has the following QFM structure at some $\tau \in (0, 1)$:

$$\mathsf{Q}_{X_{it}}[\tau | f_t(\tau)] = \lambda_i'(\tau) f_t(\tau),$$

where the common factors $f_t(\tau)$ is a $r(\tau) \times 1$ vector of unobservable random variables, $\lambda_i(\tau)$ is a $r(\tau) \times 1$ vector of non-random factor loadings with $r(\tau) \ll N$. Note that in the QFM defined above, the factors, the loadings, and the number of factors are all allowed to be quantile-dependent.

Alternatively, the above equation implies that

$$X_{it} = \lambda_i'(\tau)f_t(\tau) + u_{it}(\tau), \tag{1}$$

where the quantile-dependent idiosyncratic error $u_{it}(\tau)$ is assumed to satisfy the following quantile restrictions:

$$P[u_{it}(\tau) \leq 0 | f_t(\tau)] = \tau.$$

2.2 Examples

In this section we provide a few illustrative examples of how QFMs can be derived from different specifications of location-scale shift models and related ones. The goal of these simple illustrations is to show instances where the standard AFM methodology may fail to capture the full factor structure, therefore requiring the use of the alternative QFM approach.

Example 1. *Location-shift model.* $X_{it} = \alpha_i f_{1t} + \epsilon_{it}$, *where* $\{\epsilon_{it}\}$ *are zero-mean i.i.d errors independent of* $\{f_{1t}\}$ *with cumulative distribution function (CDF)* F_ϵ. *Let* $\mathsf{Q}_\epsilon(\tau) = \mathsf{F}_\epsilon^{-1}(\tau) = \inf\{c : \mathsf{F}_\epsilon(c) \leq \tau\}$ *be the quantile function of* ϵ_{it}. *Moreover, assume that the median of* ϵ_{it} *is 0, i.e.,* $\mathsf{Q}_\epsilon(0.5) = 0$, *then this simple model has a QFM representation* (1) *by defining* $\lambda_i(\tau) = [\mathsf{Q}_\epsilon(\tau), \alpha_i]'$, $f_t(\tau) = [1, f_{1t}]'$ *for* $\tau \neq 0.5$, *and* $\lambda_i(\tau) = \alpha_i$, $f_t(\tau) = f_{1t}$ *for* $\tau = 0.5$. *However, note that the standard estimation method (PCA) for this AFM may not be consistent if the distribution of* ϵ_{it} *has heavy tails. For example, Assumption C of BN (2002) requires* $\mathbb{E}[\epsilon_{it}^8] < \infty$, *which is not satisfied if, e.g.* ϵ_{it} *follows the standard Cauchy or some Pareto distributions.*

Example 2. *Location-scale shift model (same sign-restricted factor).* $X_{it} = \alpha_i f_{1t} + \eta_i f_{1t} \epsilon_{it}$, *where* $\eta_i f_{1t} > 0$ *for all* i, t *and* $\{\epsilon_{it}\}$ *are defined as in Example 1. This model has a QFM representation* (1) *by defining* $\lambda_i(\tau) = \eta_i \mathsf{Q}_\epsilon(\tau) + \alpha_i$ *and* $f_t(\tau) = f_{1t}$ *for all* τ, *such that the loadings of the factor* f_{1t} *are the only quantile-dependent objects.*

Example 3. *Location-scale shift model (different factors).* $X_{it} = \alpha_i' f_{1t} + (\eta_i' f_{2t})\epsilon_{it}$, *where* $\{\epsilon_{it}\}$ *are defined as in Example 1,* $\alpha_i, f_{1t} \in \mathbb{R}^{r_1}$, $\eta_i, f_{2t} \in \mathbb{R}^{r_2}$, *and* $\eta_i' f_{2t} > 0$. *When* f_{1t} *and* f_{2t} *do not share common elements, this model has a QFM representation* (1) *with* $\lambda_i(\tau) = [\alpha_i', \eta_i' \mathsf{Q}_\epsilon(\tau)]'$, $f_t(\tau) = [f_{1t}', f_{2t}']$ *for* $\tau \neq 0.5$, *and* $\lambda_i(\tau) = \alpha_i$, $f_t(\tau) = f_{1t}$ *for* $\tau = 0.5$.

Example 4. *Location-scale shift model with an idiosyncratic error and its cube.* $X_{it} = \alpha_i f_{1t} + f_{2t}\epsilon_{it} + c_i f_{3t}\epsilon_{it}^3$, *where* ϵ_{it} *is a standard normal random variable whose CDF is denoted as* $\Phi(\cdot)$. *Let* f_{2t}, f_{3t}, c_i *be positive, then* X_{it} *has an equivalent representation in form*

of (1) with $\lambda_i(\tau) = [\alpha_i, \Phi^{-1}(\tau), c_i\Phi^{-1}(\tau)^3]'$, $f_t(\tau) = (f_{1t}, f_{2t}, f_{3t})'$ for $\tau \neq 0.5$, and $\lambda_i(\tau) = \alpha_i$, $f_t(\tau) = f_{1t}$ for $\tau = 0.5$. In particular, if $c_i = 1$ for all i and noticing that the mapping $\tau \mapsto \Phi^{-1}(\tau)^3$ is strictly increasing, then we have for $\tau \neq 0.5$, $Q_{X_{it}}[\tau|f_t(\tau)] = \alpha_i f_{1t} + \Phi^{-1}(\tau) \cdot [f_{2t} + f_{3t}\Phi^{-1}(\tau)^2]$, so that there exists a QFM representation (1) with $\lambda_i(\tau) = [\alpha_i, \Phi^{-1}(\tau)]'$ and $f_t(\tau) = [f_{1t}, f_{2t} + f_{3t}\Phi^{-1}(\tau)^2]'$ for $\tau \neq 0.5$. Notice that in this case, the second factor in $f_t(\tau)$, $f_{2t} + f_{3t}\Phi^{-1}(\tau)^2$, is quantile dependent even for $\tau \neq 0.5$.

Not surprisingly, the standard AFM methodology based on PCA only works in Example 1, insofar as the idiosyncratic errors satisfy certain moment conditions. In the remaining examples, PCA will only yield consistent estimates of those factors shifting the locations; however (except in Example 2), it will fail to capture those extra factors which shift quantiles other than the means, or their corresponding quantile-varying loadings. In the sequel, QFA is therefore proposed as a new estimation procedure to estimate both sets of quantile-dependent objects in QFM.

3 Estimators and their Asymptotic Properties

To simplify the notations, we suppress hereafter the dependence of $f_t(\tau), \lambda_i(\tau), r(\tau)$ and $u_{it}(\tau)$ on τ, so that the QFM in (1) is rewritten as:

$$X_{it} = \lambda_i' f_t + u_{it}, \quad P[u_{it} \leq 0|f_t] = \tau, \tag{2}$$

where $\lambda_i, f_t \in \mathbb{R}^r$. Suppose that we have a sample of observations $\{X_{it}\}$ generated by (2) for $i = 1, \ldots, N$, and $t = 1, \ldots, T$, where the realized values of $\{f_t\}$ are $\{f_{0t}\}$ and the true values of $\{\lambda_i\}$ are $\{\lambda_{0i}\}$. We take a fixed-effects approach by treating $\{\lambda_{0i}\}$ and $\{f_{0t}\}$ as parameters to be estimated, and our asymptotic analysis is conditional on $\{f_{0t}\}$. In Section 3.1, we consider the estimation of $\{\lambda_{0i}\}$ and $\{f_{0t}\}$ while r is assumed to be known. Finally, Section 3.2 deals with the estimation of r for each quantile.

3.1 Estimating Factors and Loadings

It is well known in the literature on factor models that $\{\lambda_{0i}\}$ and $\{f_{0t}\}$ cannot be separately identified without imposing normalizations (see BN 2002). Without loss of generality, we choose the following normalizations:

$$\frac{1}{T}\sum_{t=1}^{T} f_t f_t' = \mathbb{I}_r, \quad \frac{1}{N}\sum_{i=1}^{N} \lambda_i \lambda_i' \text{ is diagonal with non-increasing diagonal elements.} \tag{3}$$

Let $M = (N + T)r$, $\theta = (\lambda_1', \ldots, \lambda_N', f_1', \ldots, f_T')'$, and $\theta_0 = (\lambda_{01}', \ldots, \lambda_{0N}', f_{01}', \ldots, f_{0T}')'$ denotes the vector of true parameters, where we also suppress the dependence of θ and θ_0 on M

to save notation. Let $\mathcal{A}, \mathcal{F} \subset \mathbb{R}^r$ and define:

$$\Theta^r = \left\{ \theta \in \mathbb{R}^M : \lambda_i \in \mathcal{A}, f_t \in \mathcal{F} \text{ for all } i, t, \{\lambda_i\} \text{ and } \{f_t\} \text{ satisfy the normalizations in (3)} \right\}.$$

Further, define:

$$\mathbb{M}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda_i' f_t),$$

where $\rho_\tau(u) = (\tau - \mathbf{1}\{u \leq 0\})u$ is the check function. The QFA estimator of θ_0 is defined as:

$$\hat{\theta} = (\hat{\lambda}_1', \dots, \hat{\lambda}_N', \hat{f}_1', \dots, \hat{f}_T')' = \arg\min_{\theta \in \Theta^r} \mathbb{M}_{NT}(\theta).$$

It is obvious that the way in which our estimator is related to the PCA estimator studied by BN (2002) and Bai (2003) is analogous to how QR is related to standard least-squares regressions. However, unlike Bai (2003)'s PCA estimator, our estimator $\hat{\theta}$ does not yield an analytical closed form. This makes it difficult not only to find a computational algorithm that would yield the estimator, but also the analysis of its asymptotic properties. In the sequel, we introduce a computational algorithm called *iterative quantile regression* (IQR, hereafter) that can effectively find the stationary points of the object function. In parallel, Theorem 1 shows that $\hat{\theta}$ achieves the same convergence rate as the PCA estimators for AFM.

To describe the algorithm, let $\Lambda = (\lambda_1, \dots, \lambda_N)'$, $F = (f_1, \dots, f_T)'$, and define the following averages:

$$\mathbb{M}_{i,T}(\lambda, F) = \frac{1}{T} \sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda' f_t) \quad \text{and} \quad \mathbb{M}_{t,N}(\Lambda, f) = \frac{1}{N} \sum_{i=1}^{N} \rho_\tau(X_{it} - \lambda_i' f).$$

Note that we have $\mathbb{M}_{NT}(\theta) = N^{-1} \sum_{i=1}^{N} \mathbb{M}_{i,T}(\lambda_i, F) = T^{-1} \sum_{t=1}^{T} \mathbb{M}_{t,N}(\Lambda, f_t)$. The main difficulty in finding the global minimum of $\mathbb{M}_{NT}$ is that this object function is not convex in θ. However, for given F, $\mathbb{M}_{i,T}(\lambda, F)$ happens to be convex in λ for each i and likewise, for given Λ, $\mathbb{M}_{t,N}(\Lambda, f)$ is convex in f for each t. Thus, both optimization problems can be efficiently solved by various linear programming methods (see Chapter 6 of Koenker 2005). Based on this observation, we propose the following iterative procedure:

Iterative quantile regression (IQR):
Step 1: Choose random starting parameters: $F^{(0)}$.
Step 2: Given $F^{(l-1)}$, solve $\lambda_i^{(l-1)} = \arg\min_\lambda \mathbb{M}_{i,T}(\lambda, F^{(l-1)})$ for $i = 1, \dots, N$; given $\Lambda^{(l-1)}$, solve $f_t^{(l)} = \arg\min_f \mathbb{M}_{t,N}(\Lambda^{(l-1)}, f)$ for $t = 1, \dots, T$.
Step 3: For $l = 1, \dots, L$, iterate the second step until $\mathbb{M}_{NT}(\theta^{(L)})$ is close to $\mathbb{M}_{NT}(\theta^{(L-1)})$, where $\theta^{(l)} = (\text{vech}(\Lambda^{(l)})', \text{vech}(F^{(l)})')'$.
Step 4: Normalize $\Lambda^{(L)}$ and $F^{(L)}$ so that they satisfy the normalizations in (3).

To see the connection between the IQR algorithm and the PCA estimator of Bai (2003), suppose that $r = 1$, and replace the check function in the IQR algorithm by the least-squares loss function. Then, it is easy to show that the second step of the algorithm above yields $\Lambda^{(l-1)} = (X'F^{(l-1)})/\|F^{(l-1)}\|^2$ and $F^{(l)} = (X\Lambda^{(l-1)})/\|\Lambda^{(l-1)}\|^2 = XX'F^{(l-1)}/C_{l-1}$, where X is the $T \times N$ matrix with elements $\{X_{it}\}$, and $C_l = \|F^{(l)}\|^2 \cdot \|\Lambda^{(l)}\|^2$. Thus, with proper normalizations at each step, the iterative procedure is equivalent to the well-known *power method* of Hotelling (1933), and the sequence $F^{(0)}, F^{(1)}, \ldots$ will converge to the eigenvector associated with the largest eigenvalue of XX'. In the more general case $r > 1$, if we replace the check function in the IQR algorithm by the least-squares loss function and normalize $F^{(l-1)}, \Lambda^{(l-1)}$ to satisfy (3) at step 2, it can be shown that the above iterative procedure is similar to the method of *orthogonal iteration* (see Section 7.3.2 of Golub and Van Loan 2013) for calculating the eigenvectors associated with the r largest eigenvalues of XX', which is the PCA estimator of Bai (2003). Therefore, the IQR algorithm and its corresponding QFA estimator can be viewed as an extension of PCA to QFM.

Similar algorithms have been proposed in the machine learning literature to reduce the dimensions for binary data, where the check function is replaced by some smooth nonlinear link functions, e.g. Collins et al. (2001). However, unlike PCA, whether such methods guarantee finding the global minimum remains an important open question. Nonetheless, in all of our Monte Carlo simulations we found that the QFA estimators of the factors using the IQR algorithm always converge to the space of the true factors, which is somewhat reassuring in this respect.

To prove the consistency of the QFA estimator $\hat{\theta}$, we make the following assumptions:

Assumption 1. *(i)* $\mathcal{A}$ *and* $\mathcal{F}$ *are compact sets and* $\theta_0 \in \Theta^r$. *In particular,* $N^{-1} \sum_{i=1}^{N} \lambda_{0i} \lambda'_{0i} = diag(\sigma_{N1}, \ldots, \sigma_{Nr})$ *with* $\sigma_{N1} \geq \sigma_{N2} \cdots \geq \sigma_{Nr}$, *and* $\sigma_{Nj} \to \sigma_j$ *as* $N \to \infty$ *for* $j = 1, \ldots, r$ *with* $\infty > \sigma_1 > \sigma_2 \cdots > \sigma_r > 0$.
(ii) The conditional density function of u_{it} *given* $\{f_{0t}\}$, *denoted as* f_{it}, *is continuous, and satisfies that: for any compact set* $C \subset \mathbb{R}$ *and any* $u \in C$, *there exists a positive constant* $\underline{\mathsf{f}} > 0$ *(depending on* C*) such that* $\mathsf{f}_{it}(u) \geq \underline{\mathsf{f}}$ *for all* i, t.
(iii) Given $\{f_{0t}, 1 \leq t \leq T\}$, $\{u_{it}, 1 \leq i \leq N, 1 \leq t \leq T\}$ *is independent across* i *and* t.

Assumptions 1 (i) is essentially the *strong factors* assumption that is standard in the literature (see Assumption B of Bai 2003). The requirement that $\sigma_1, \ldots, \sigma_r$ are distinct is similar to Assumption G of Bai (2003), which is a convenient assumption to order the factors. Assumptions 1 (ii) and (iii) are similar to (C1) and (C2) in AB (2020), except that we do not require moments of u_{it} to exist. Also notice that Assumption (iii), which allows for both cross-sectional and time series heteroskedasticity, requires the idiosyncratic errors to be mutually independent. This stems from the use of Hoeffding's inequality in the proofs of some results, which provides a sub-Gaussian tail bound for the sum of bounded independent random variables. There have

been attempts to relax this assumption (see Remark 1.4 below) but it is difficult to characterize the minimal set of conditions that the error terms should satisfy to achieve the sub-Gaussian inequality required in our proofs. Notice, however, that in exchange for the independence assumption, we can dispense with the bounded moment conditions in the idiosyncratic terms, whose violation would render PCA invalid. At any rate, in sub-section 5.2 we run some Monte Carlo simulation on the performance of our QFA estimation when error terms are allowed to exhibit mild cross-sectional and serial dependence in order to check the robustness of our results to these features.

Write $\hat{\Lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_N)'$, $\Lambda_0 = (\lambda_{01}, \ldots, \lambda_{0N})'$, $\hat{F} = (\hat{f}_1, \ldots, \hat{f}_T)'$, $F_0 = (f_{01}, \ldots, f_{0T})'$, and let $L_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. The following theorem provides the average rate of convergence of $\hat{\Lambda}$ and $\hat{F}$.

Theorem 1. *Under Assumption 1, there exists a diagonal matrix* $\mathsf{S} \in \mathbb{R}^{r \times r}$ *whose diagonal elements are either* 1 *or* -1, *such that as* $N, T \to \infty$,

$$\|\hat{\Lambda} - \Lambda_0 \mathsf{S}\|/\sqrt{N} = O_P(1/L_{NT}) \quad \text{and} \quad \|\hat{F} - F_0 \mathsf{S}\|/\sqrt{T} = O_P(1/L_{NT}).$$

The sign matrix S appears in the above result due the intrinsic sign indeterminacy of factors and loadings – that is, the factor structure remains unchanged if a factor and its loading are both multiplied by -1 (e.g., see Theorem 1.b of Stock and Watson 2002 for a similar result).

Remark 1.1: Since our proof strategy is substantially different from that of BN (2002), we briefly sketch the main ideas underlying our proof here. To facilitate the discussion, for any $\theta_a, \theta_b \in \Theta^r$ define the semimetric d by:

$$d(\theta_a, \theta_b) = \sqrt{\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\lambda'_{ai} f_{at} - \lambda'_{bi} f_{bt})^2} = \frac{1}{\sqrt{NT}} \left\| \Lambda_a F'_a - \Lambda_b F'_b \right\|,$$

and let

$$\bar{\mathbb{M}}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[\rho_\tau(X_{it} - \lambda'_i f_t)].$$

The semimetric d plays an important role in our asymptotic analysis. We first show that $d(\hat{\theta}, \theta_0) = o_P(1)$. Next, it can be shown that:

$$\bar{\mathbb{M}}_{NT}(\hat{\theta}) - \bar{\mathbb{M}}_{NT}(\theta_0) \gtrsim d^2(\hat{\theta}, \theta_0), \tag{4}$$

and that for sufficiently small $\delta > 0$,

$$\mathbb{E}\left[\sup_{\theta \in \Theta^r(\delta)} \left| \mathbb{M}_{NT}(\theta) - \bar{\mathbb{M}}_{NT}(\theta) - \mathbb{M}_{NT}(\theta_0) + \bar{\mathbb{M}}_{NT}(\theta_0) \right| \right] \lesssim \frac{\delta}{L_{NT}}, \tag{5}$$

where $\Theta^r(\delta) = \{\theta \in \Theta^r : d(\theta, \theta_0) \leq \delta\}$. Intuitively, the above two inequalities and $d(\hat{\theta}, \theta_0) = o_P(1)$ imply that $d^2(\hat{\theta}, \theta_0) \lesssim d(\hat{\theta}, \theta_0)/L_{NT}$, or $d(\hat{\theta}, \theta_0) \lesssim L_{NT}^{-1}$. Then, the desired results follow from the fact that $\|\hat{\Lambda} - \Lambda_0 \mathsf{S}\|/\sqrt{N} + \|\hat{F} - F_0 \mathsf{S}\|/\sqrt{T} \lesssim d(\hat{\theta}, \theta_0)$.

Inequality (4) follows easily from a Taylor expansion of $\bar{\mathbb{M}}_{NT}(\hat{\theta})$ around θ_0, together with Assumption 1(ii). It is worth stressing that the proof of (5) requires the chaining argument which is commonly used in the theory of empirical processes. In particular, using Hoeffding's inequality and the fact that $|\rho_\tau(u) - \rho_\tau(v)| \leq 2|u - v|$, it can be shown that, for any given $\theta_a, \theta_b \in \Theta^r$,

$$P\left[\sqrt{NT} \left|\mathbb{M}_{NT}(\theta_a) - \bar{\mathbb{M}}_{NT}(\theta_a) - \mathbb{M}_{NT}(\theta_b) + \bar{\mathbb{M}}_{NT}(\theta_b)\right| \geq c\right] \leq e^{-\frac{c^2}{Kd^2(\theta_a, \theta_b)}} \tag{6}$$

for some constant K. Then, along the lines of Theorem 2.2.4 of van der Vaart and Wellner (1996), it follows that the left-hand side of (5) is bounded (up to a positive constant) by $\int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^r(\delta))} d\epsilon/\sqrt{NT}$. Finally, we can prove that $\int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^r(\delta))} d\epsilon \lesssim \delta\sqrt{M}$, from which inequality (5) follows.

Remark 1.2: Compared to BN (2002), recall that, in exchange for Assumption 1(iii) we do not require any moment of u_{it} to be finite. Thus, for the canonical AFM (e.g., Example 1) where the idiosyncratic errors have median equal to zero and satisfy Assumption 1(iii), our estimator for the case $\tau = 0.5$ can be interpreted as a least absolute deviation (LAD) estimator which is robust to heavy tails and outliers. In relation to this issue, it is important to point out that the LAD estimator is related to robust PCA in the machine learning literature that aims to recover a low rank matrix from a large panel of observables. For example, the *Principal Components Pursuit* method proposed by Candès et al. (2011) features a combination of the L_1 norm (as in LAD) and a nuclear norm on the low rank matrix (see Chapter 3 of Vidal et al. 2016 and Bai and Ng 2019 for other robust PCA methods). In section 5 below, we will illustrate the robustness of the LAD estimator relative to the PCA estimator by Monte Carlo simulations.

Remark 1.3: If the true parameters do not satisfy the normalizations (3), they can still be in the space Θ^r after some normalizations. Let H_{NT} be a $r \times r$ invertible matrix and define $\bar{f}_{0t} = H'_{NT} f_{0t}$, $\bar{\lambda}_{0i} = (H_{NT})^{-1} \lambda_{0i}$. Note that $\lambda'_{0i} f_{0t} = \bar{\lambda}'_{0i} \bar{f}_{0t}$. For $\{\bar{f}_{0t}\}$ and $\{\bar{\lambda}_{0i}\}$ to satisfy the normalizations (3), we require:

$$\frac{1}{T} \sum_{t=1}^T \bar{f}_{0t} \bar{f}'_{0t} = H'_{NT} \Sigma_{T,F} H_{NT} = \mathbb{I}_r \quad \text{and} \quad \frac{1}{N} \sum_{i=1}^N \bar{\lambda}_{0i} \bar{\lambda}'_{0i} = (H_{NT})^{-1} \Sigma_{N,\Lambda} (H'_{NT})^{-1} = \mathbb{D}_N,$$

where $\Sigma_{T,F} = T^{-1} \sum_{t=1}^T f_{0t} f'_{0t}$, $\Sigma_{N,\Lambda} = \frac{1}{N} \sum_{i=1}^N \lambda_{0i} \lambda'_{0i}$, and $\mathbb{D}_N$ is a diagonal matrix with non-

increasing diagonal elements. The above equalities imply that:

$$\Sigma_{T,F}^{1/2}\Sigma_{N,\Lambda}\Sigma_{T,F}^{1/2} \cdot \Sigma_{T,F}^{1/2}H_{NT} = \Sigma_{T,F}^{1/2}H_{NT} \cdot \mathbb{D}_N.$$

Thus, the rotation matrix H_{NT} can be chosen as $\Sigma_{T,F}^{-1/2}\Gamma_{NT}$, where Γ_{NT} is the matrix of eigenvectors of $\Sigma_{T,F}^{1/2}\Sigma_{N,\Lambda}\Sigma_{T,F}^{1/2}$. Note that when the eigenvalues of $\Sigma_{T,F}^{1/2}\Sigma_{N,\Lambda}\Sigma_{T,F}^{1/2}$ are distinct, its eigenvectors are unique up to signs, i.e., each eigenvector can be replaced by the negative of itself. As a result, Theorem 1 can be stated as follows:

$$\|\hat{\Lambda} - \Lambda_0(H'_{NT})^{-1}\mathsf{S}\|/\sqrt{N} = O_P(1/L_{NT}) \quad \text{and} \quad \|\hat{F} - F_0 H_{NT}\mathsf{S}\|/\sqrt{T} = O_P(1/L_{NT}),$$

where S is the diagonal matrix defined above. Notice that the rotation matrix H_{NT} is slightly different from the rotation matrix of Bai (2003). Moreover, because both λ_{0i} and f_{0t} are τ-dependent, H_{NT} also varies across quantiles, although we did not make it explicitly quantile dependent in the previous discussion to simplify notation.

Remark 1.4: Compared to BN (2002), our Assumption 1(iii) is admittedly strong. However, note that this assumption is made conditional on $\{f_{0t}\}$, so cross-sectional and temporal dependence of u_{it} due to the common factors are still allowed for. Moreover, the independence assumption is only used to establish the sub-Gaussian inequality (6). Thus, Assumption 1(iii) can be relaxed as long as the sub-Gaussian inequality holds.[7]

3.2 Selecting the Number of Factors

In the previous section, we assumed the number of quantile-dependent factors $r(\tau)$ to be known at each τ. In this subsection we propose two different procedures to select the correct number of factors at each quantile with probability approaching one. The first one selects the number of factors by rank minimization while the second one uses information criteria (IC). As before, the dependence of the quantile-dependent objects on τ, including $r(\tau)$, is suppressed for notational simplicity.

3.2.1 Model Selection by Rank Minimization

Let k be a positive integer larger than r, and $\mathcal{A}^k$ and $\mathcal{F}^k$ be compact subsets of $\mathbb{R}^k$. In particular, let us assume that $[\lambda'_{0i} \quad \mathbf{0}_{1\times(k-r)}]' \in \mathcal{A}^k$ for all i.

Let $\lambda_i^k, f_t^k \in \mathbb{R}^k$ for all i,t and write $\theta^k = (\lambda_1^{k'},\ldots,\lambda_N^{k'},f_1^{k'},\ldots,f_T^{k'})'$, $\Lambda^k = (\lambda_1^k,\ldots,\lambda_N^k)'$,

---

[7]See van de Geer (2002) for the properties of Hoeffding inequalities for martingales.

$F^k = (f_1^k, \ldots, f_T^k)'$. Consider the following normalizations:

$$\frac{1}{T}\sum_{t=1}^{T} f_t^k f_t^{k'} = \mathbb{I}_k, \quad \frac{1}{N}\sum_{i=1}^{N} \lambda_i^k \lambda_i^{k'} \text{ is diagonal with non-increasing diagonal elements.} \quad (7)$$

Define $\Theta^k = \{\theta^k : \lambda_i^k \in \mathcal{A}^k, f_t^k \in \mathcal{F}^k, \text{ and } \lambda_i^k, f_t^k \text{ satisfy (7)}\}$, and

$$\hat{\theta}^k = (\hat{\lambda}_1^{k'}, \ldots, \hat{\lambda}_N^{k'}, \hat{f}_1^{k'}, \ldots, \hat{f}_T^{k'})' = \arg\min_{\theta^k \in \Theta^k} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda_i^{k'} f_t^k).$$

Moreover, define $\hat{\Lambda}^k = (\hat{\lambda}_1^k, \ldots, \hat{\lambda}_N^k)'$ and write

$$(\hat{\Lambda}^k)'\hat{\Lambda}^k/N = \text{diag}\left(\hat{\sigma}_{N,1}^k, \ldots, \hat{\sigma}_{N,k}^k\right).$$

The first estimator of the number of factors r is defined as:

$$\hat{r}_{\text{rank}} = \sum_{j=1}^{k} \mathbf{1}\{\hat{\sigma}_{N,j}^k > P_{NT}\},$$

where P_{NT} is a sequence that goes to 0 as $N, T \to \infty$. In other words, $\hat{r}_{\text{rank}}$ is equal to the number of diagonal elements of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$ that are larger than the threshold P_{NT}. We call $\hat{r}_{\text{rank}}$ the *rank-minimization estimator* because, as discussed below in Remark 2.1, it can be interpreted as a rank estimator of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$.

It can then be shown that:

Theorem 2. *Under Assumption 1, $P[\hat{r}_{rank} = r] \to 1$ as $N, T \to \infty$ if $k > r$, $P_{NT} \to 0$ and $P_{NT}L_{NT}^2 \to \infty$.*

Remark 2.1: In the proof of Theorem 2, we show that for $k > r$, it holds that (up to sign)

$$\left\|\hat{F}^{k,r} - F_0\right\|/\sqrt{T} = O_P(1/L_{NT}) \quad \text{and} \quad \left\|\hat{\Lambda}^k - \Lambda_0^*\right\|/\sqrt{N} = O_P(1/L_{NT}),$$

where $\hat{F}^{k,r}$ is the first r columns of $\hat{F}^k$ and $\Lambda_0^* = [\Lambda_0, \mathbf{0}_{N\times(k-r)}]$. It then follows from Assumption 1 that $\hat{\sigma}_{N,j}^k \xrightarrow{p} \sigma_j > 0$ for $j = 1, \ldots, r$ and $\hat{\sigma}_{N,j}^k = N^{-1}\sum_{i=1}^{N}\left(\hat{\lambda}_{i,j}^k\right)^2 = O_P(1/L_{NT}^2)$ for $j = r+1, \ldots, k$. Thus, the first r diagonal components of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$ converge in probability to positive constants while the remaining diagonal components are all $O_P(1/L_{NT}^2)$. In other words, $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$ converges to a matrix with rank r, and P_{NT} can be viewed as a cutoff value to choose the asymptotic rank of $(\hat{\Lambda}^k)'\hat{\Lambda}^k/N$.

3.2.2 Model Selection by Information Criteria

The second estimator of r is similar to the IC-based estimator of BN (2002). Let l denote a positive integer smaller or equal to k, and $\mathcal{A}^l$ and $\mathcal{F}^l$ be compact subsets of $\mathbb{R}^l$. In particular, for $l > r$, assume that $[\lambda'_{0i} \quad \mathbf{0}_{1\times(l-r)}]' \in \mathcal{A}^l$ for all i. Moreover, we can define $\Theta^l, \hat{\theta}^l, \hat{f}^l_t, \hat{\lambda}^l_i, \hat{F}^l$ and $\hat{\Lambda}^l$ in a similar fashion.

Define the IC-based estimator of r as follows:

$$\hat{r}_{\mathrm{IC}} = \underset{1 \leq l \leq k}{\arg\min} \left[\mathbb{M}_{NT}(\hat{\theta}^l) + l \cdot P_{NT} \right].$$

We can show that:

Theorem 3. *Suppose Assumption 1 holds, and assume that for any compact set $C \subset \mathbb{R}$ and any $u \in C$, there exists $\bar{\mathsf{f}} > 0$ (depending on C) such that $\mathsf{f}_{it}(u) \leq \bar{\mathsf{f}}$ for all i, t. Then $P[\hat{r}_{IC} = r] \to 1$ as $N, T \to \infty$ if $k > r$, $P_{NT} \to 0$ and $P_{NT} L^2_{NT} \to \infty$.*

Remark 3.1: AB (2020) obtain a similar result, but the difference with ours is that we only need the density function of the idiosyncratic errors to be uniformly bounded above and below, while AB (2020) requires all the moments of the errors to be bounded. The reason why we can obtain the same result here with less restrictions is that our proof is based on the innovative argument discussed in Remark 1.1 and on the average convergence rate of the estimators, while the proof of AB (2020) depends on the uniform convergence rate of the estimators.

Remark 3.2: Let X denote the $T \times N$ matrix of observed variables, and let $\check{F}^l, \check{\Lambda}^l$ denote the matrices of PCA estimators of BN (2002) when the number of factors is specified as l. Then BN (2002)'s estimator of r can be written as:

$$\hat{r} = \underset{1 \leq l \leq k}{\arg\min} \hat{S}(l) \quad \text{where} \quad \hat{S}(l) = (NT)^{-1} \left\| X - \check{F}^l \check{\Lambda}^{l'} \right\|^2 + l \cdot P_{NT},$$

$k > r$, and P_{NT} is defined as in Theorem 2 above. It can be shown that IC-based estimator $\hat{r}$ is equivalent to the number of diagonal elements in $\check{\Lambda}^{k'} \check{\Lambda}^k / N$ that are larger than P_{NT}. Thus, the two seemly different estimators of the number of factors are equivalent in AFM. However, due to the differences of the object functions, such equivalence does not exist in QFM.

Remark 3.3: The choice of P_{NT} for $\hat{r}_{\mathrm{rank}}$ and $\hat{r}_{\mathrm{IC}}$ can be different in practice. In particular, it can differ from those penalties used by BN (2002). AB (2020) choose

$$P_{NT} = \log \left(\frac{NT}{N+T} \right) \cdot \frac{N+T}{NT}$$

for $\hat{r}_{\text{IC}}$, similar to IC_{p1} of BN (2002). However, as shown in AB's (2020) simulation results, this choice does not perform very well even for N, T as large as 300.

Remark 3.4: Even though $\hat{r}_{\text{rank}}$ and $\hat{r}_{\text{IC}}$ are both consistent estimators of r, the computational cost of $\hat{r}_{\text{rank}}$ is much lower than that of $\hat{r}_{\text{IC}}$, because for $\hat{r}_{\text{rank}}$ we only estimate the model once, while for $\hat{r}_{\text{IC}}$ we need to estimate the model k times. Thus, in the simulations and empirical applications we will focus on $\hat{r}_{\text{rank}}$, and we refer to AB (2020) for the corresponding simulation results of $\hat{r}_{\text{IC}}$. In particular, we find that the choice

$$P_{NT} = \hat{\sigma}_{N,1}^k \cdot \left(\frac{1}{L_{NT}^2} \right)^{1/3}$$

for $\hat{r}_{\text{rank}}$ works fairly well as long as $\min\{N, T\}$ is 100. This is also the value used in all of our simulations and applications.

4 Estimators Based on Smoothed Quantile Regressions

The derivation of the asymptotic distribution of the QFA estimator $\hat{\theta}$ becomes a difficult task due to the non-smoothness of the check function and the problem of incidental parameters. As in the asymptotic analysis of conventional QR, one can expand the expected score function (which is smooth and continuously differentiable) and obtain a stochastic expansion for $\hat{\lambda}_i - \mathsf{S}\lambda_{0i}$; yet the following term appears in the expansion:

$$\frac{1}{T} \sum_{t=1}^{T} \left\{ \left(\mathbf{1}\{X_{it} \le \hat{\lambda}_i' \hat{f}_t\} - \mathbb{E}[\mathbf{1}\{X_{it} \le \hat{\lambda}_i' \hat{f}_t\}] \right) \hat{f}_t - \left(\mathbf{1}\{X_{it} \le \lambda_{0i}' f_{0t}\} - \tau \right) f_{0t} \right\}. \tag{8}$$

The next step would be to show that the above expression is a higher-order term (i.e. $o_P(T^{-0.5})$) thus it does not affect the asymptotic distribution of $\hat{\lambda}_i$. However, due to the presence of the indicator functions in (8), this is not an easy task. To see this, let's consider a similar problem for the PCA estimators of AFM. Let $\check{\lambda}_i$ and $\check{f}_t$ be the PCA estimators. In the stochastic expansion of $\check{\lambda}_i - \lambda_{0i}$, the analogous term to (8) happens to be:

$$\frac{1}{T} \sum_{t=1}^{T} \epsilon_{it} (\check{f}_t - f_{0t}),$$

15

where ϵ_{it} is the idiosyncratic error in the AFM. Note that, based on the result $T^{-1} \sum_{t=1}^{T} \|\check{f}_t - f_{0t}\| = O_P(L_{NT}^{-1})$, one can only show that:

$$\left\| \frac{1}{T} \sum_{t=1}^{T} \epsilon_{it}(\check{f}_t - f_{0t}) \right\| \leq \sqrt{\frac{1}{T} \sum_{t=1}^{T} \epsilon_{it}^2} \cdot \sqrt{\frac{1}{T} \sum_{t=1}^{T} \|\check{f}_t - f_{0t}\|^2} = O_P(L_{NT}^{-1}).$$

Instead, one has to use the stochastic expansion of $\check{f}_t - f_{0t}$ to show that $T^{-1} \sum_{t=1}^{T} \epsilon_{it}(\check{f}_t - f_{0t}) = O_P(L_{NT}^{-2})$ (see the proof of Lemma B.1 of Bai 2003). Likewise, to show that (8) is $o_P(T^{-0.5})$, establishing the convergence rate of $\hat{f}_t - \mathsf{S}f_{0t}$ is not enough, and the stochastic expansion of $\hat{f}_t - \mathsf{S}f_{0t}$ is required. However, due the non-smoothness of the indicator functions, it is not clear how to explore this stochastic expansion in (8).

To overcome this problem, we proceed to define a new estimator of θ_0, denoted as $\tilde{\theta}$, relying on the following smoothed quantile regressions (SQR):

$$\tilde{\theta} = (\tilde{\lambda}_1', \ldots, \tilde{\lambda}_N', \tilde{f}_1', \ldots, \tilde{f}_T')' = \arg\min_{\theta \in \Theta^r} \mathbb{S}_{NT}(\theta),$$

where

$$\mathbb{S}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tau - K\left(\frac{X_{it} - \lambda_i' f_t}{h} \right) \right] (X_{it} - \lambda_i' f_t),$$

$K(z) = 1 - \int_{-1}^{z} k(z)dz$, $k(z)$ is a continuous function with support $[-1, 1]$, and h is a bandwidth parameter that goes to 0 as N, T diverge.

Define

$$\Phi_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathsf{f}_{it}(0) f_{0t} f_{0t}' \quad \text{and} \quad \Psi_t = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \mathsf{f}_{it}(0) \lambda_{0i} \lambda_{0i}'$$

for all i, t. We impose the following assumptions:

Assumption 2. *Let $m \geq 8$ be a positive integer,*
(i) $\Phi_i > 0$ and $\Psi_t > 0$ for all i, t.
(ii) λ_{0i} is an interior point of $\mathcal{A}$ and f_{0t} is an interior point of $\mathcal{F}$ for all i, t.
(iii) $k(z)$ is symmetric around 0 and twice continuously differentiable. $\int_{-1}^{1} k(z)dz = 1$, $\int_{-1}^{1} z^j k(z)dz = 0$ for $j = 1, \ldots, m-1$ and $\int_{-1}^{1} z^m k(z)dz \neq 0$.
(iv) f_{it} is $m+2$ times continuously differentiable. Let $\mathsf{f}_{it}^{(j)}(u) = (\partial/\partial u)^j \mathsf{f}_{it}(u)$ for $j = 1, \ldots, m+2$. For any compact set $C \subset \mathbb{R}$ and any $u \in C$, there exists $-\infty < \underline{l} < \bar{l} < \infty$ such that $\underline{l} \leq \mathsf{f}_{it}^{(j)}(u) \leq \bar{l}$ and $\underline{\mathsf{f}} \leq \mathsf{f}_{it}(u) \leq \bar{l}$ for $j = 1, \ldots, m+2$ and for all i, t.
(v) As $N, T \to \infty$, $N \propto T$, $h \propto T^{-c}$ and $m^{-1} < c < 1/6$.

The above conditions are standard in SQR, with the exception of (v). Note that, like Galvao and Kato (2016), we need $k(z)$ to be a higher-order kernel function to control the higher-order terms in the stochastic expansions of the estimators. However, Galvao and Kato (2016) assume

that $m^{-1} < c < 1/3$ (or $m \geq 4$), while we need $m^{-1} < c < 1/6$ (or $m \geq 8$). The difference is due to the fact that the incidental parameters (λ_i and f_t) in QFM enter the model interactively, while in the panel quantile models considered by these authors there are no interactive fixed-effects.

Then, we can show that:

Theorem 4. *Under Assumptions 1 and 2, there exists a diagonal matrix* $\mathsf{S} \in \mathbb{R}^{r \times r}$ *whose diagonal elements are either* 1 *or* -1, *such that*

$$\sqrt{T}(\tilde{\lambda}_i - \mathsf{S}\lambda_{0i}) \overset{d}{\to} \mathcal{N}(0, \tau(1-\tau)\Phi_i^{-2}) \quad and \quad \sqrt{N}(\tilde{f}_t - \mathsf{S}f_{0t}) \overset{d}{\to} \mathcal{N}(0, \tau(1-\tau)\Psi_t^{-1}\Sigma_\Lambda\Psi_t^{-1})$$

for each i *and* t, *where* $\Sigma_\Lambda = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$.

Remark 4.1: Similar to the proof of Theorem 1, we can show that

$$\|\tilde{\Lambda} - \Lambda_0\mathsf{S}\|/\sqrt{N} = O_P(1/L_{NT}) + O_P(h^{m/2}) \quad \text{and} \quad \|\tilde{F} - F_0\mathsf{S}\|/\sqrt{T} = O_P(1/L_{NT}) + O_P(h^{m/2}),$$

where the extra $O_P(h^{m/2})$ term is due the approximation bias of the smoothed check function. However, Assumption 2(v) implies that $1/L_{NT} \gg h^{m/2}$, and then it follows that average convergence rates of $\tilde{\Lambda}$ and $\tilde{F}$ are both L_{NT}.

Remark 4.2: Similar to Theorems 1 and 2 of Bai (2003), we show that the new estimator is free of incidental-parameter biases. That is, the asymptotic distribution of $\tilde{\lambda}_i$ is the same as if we would observe $\{f_{0t}\}$, and likewise the asymptotic distribution of $\tilde{f}_t$ is the same as if $\{\lambda_{0i}\}$ were observed. The proof of this result is not trivial. To see why this is the case, first define $\varrho(u) = [\tau - K(u/h)]u$ and $\mathbb{S}_{i,T}(\lambda, F) = T^{-1}\sum_{t=1}^T \varrho(X_{it} - \lambda'f_t)$, then we can write $\tilde{\lambda}_i = \arg\min_{\lambda \in \mathcal{A}} \mathbb{S}_{i,T}(\lambda, \tilde{F})$. Expanding $\partial\mathbb{S}_{i,T}(\tilde{\lambda}_i, \tilde{F})/\partial\lambda$ around (λ_{0i}, F_0) yields

$$\left(\frac{1}{T}\sum_{t=1}^T \varrho^{(2)}(u_{it})f_{0t}f_{0t}'\right)(\tilde{\lambda}_i - \mathsf{S}\lambda_{0i}) \approx \frac{1}{T}\sum_{t=1}^T \varrho^{(1)}(u_{it})\mathsf{S}f_{0t} + \frac{1}{T}\sum_{t=1}^T \varrho^{(1)}(u_{it})(\tilde{f}_t - \mathsf{S}f_{0t})$$

$$-\frac{1}{T}\sum_{t=1}^T \varrho^{(2)}(u_{it})f_{0t}\lambda_{0i}'(\tilde{f}_t - \mathsf{S}f_{0t}), \quad (9)$$

where $\varrho^{(j)}(u) = (\partial/\partial u)^j \varrho(u)$. The key step is to show that the last two terms on the right-hand side of the above equation are both $o_P(1/\sqrt{T})$. This is relatively easier for the PCA estimator of Bai (2003), since $(\tilde{f}_t - \mathsf{S}f_{0t})$ has an analytical form (like e.g. in equation A.1 of Bai 2003). In our case, we would also need a stochastic expansion for $(\tilde{f}_t - \mathsf{S}f_{0t})$, which in turn depends on the stochastic expansion of $(\tilde{\lambda}_i - \mathsf{S}\lambda_{0i})$ due to the nature of factor models. As in Chen et al. (2018), this problem can be partly solved by showing that the expected Hessian matrix is asymptotically block-diagonal (see Lemma 11 in the Online Appendix). However, the proof of Chen et al. (2018) is only applicable to a special infeasible normalization, namely $\sum_{i=1}^N \lambda_{0i}\lambda_i = \sum_{t=1}^T f_{0t}f_t'$, while

17

our proof of Lemma 11 allows for normalization (3) and can be generalized to any of the other normalizations considered by Bai and Ng (2013) that uniquely pin down the rotation matrix.

Remark 4.3: As discussed in Remark 1.3, if the true parameters do not satisfy the normalizations (3), the results of Theorem 3 can be stated as

$$\sqrt{T}\left(\tilde{\lambda}_i - \mathsf{S}H_{NT}^{-1}\lambda_{0i}\right) \overset{d}{\to} \mathcal{N}\left(0, \tau(1-\tau)H^{-1}\Phi_i^{-1}\Sigma_F\Phi_i^{-1}(H^{-1})'\right),$$

$$\sqrt{N}\left(\tilde{f}_t - \mathsf{S}H_{NT}'f_{0t}\right) \overset{d}{\to} \mathcal{N}\left(0, \tau(1-\tau)H'\Psi_t^{-1}\Sigma_\Lambda\Psi_t^{-1}H\right),$$

where S and H_{NT} are defined in Remark 1.3, $\Sigma_F = \lim_{T\to\infty}\Sigma_{T,F}$, $\Sigma_\Lambda = \lim_{T\to\infty}\Sigma_{N,\Lambda}$, $H = \Sigma_F^{-1/2}\Gamma$, and Γ is the matrix of eigenvectors of $\Sigma_F^{1/2}\Sigma_\Lambda\Sigma_F^{1/2}$.

Remark 4.4: Let $l(z)$ be a continuous kernel function with support $[-1,1]$ where $l^{(j)}(z) = \partial^j l(z)/\partial z^j$ exists and $|l^{(j)}(z)|$ is bounded for $j = 1, 2$. Let b a bandwidth. Estimators for the asymptotic variance matrices of $\tilde{\lambda}_i$ and $\tilde{f}_t$ can be simply constructed as

$$\tilde{V}_{\lambda_i} = \tau(1-\tau)\tilde{\Phi}_i^{-2} \text{ where } \tilde{\Phi}_i = \frac{1}{Tb}\sum_{t=1}^{T} l(\tilde{u}_{it}/b) \cdot \tilde{f}_t\tilde{f}_t',$$

and

$$\tilde{V}_{f_t} = \tau(1-\tau)\tilde{\Psi}_t^{-1}\tilde{\Sigma}_\Lambda\tilde{\Psi}_t^{-1} \text{ where } \tilde{\Psi}_t = \frac{1}{Nb}\sum_{i=1}^{N} l(\tilde{u}_{it}/b) \cdot \tilde{\lambda}_i\tilde{\lambda}_i', \quad \tilde{\Sigma}_\Lambda = \tilde{\Lambda}'\tilde{\Lambda}/N,$$

with $\tilde{u}_{it} = X_{it} - \tilde{\lambda}_i'\tilde{f}_t$. In Section A.5 of the Online Appendix we show that under Assumptions 1 and 2, the above estimators of the asymptotic covariance matrices are consistent if $b \to 0$ and $Nb^3 \to \infty$. Note that this is different from the usual condition $Nb^2 \to \infty$ in standard quantile regressions (see e.g. Powell 1984 and Angrist et al. 2006). Moreover, the above estimators are also consistent for the asymptotic covariance matrices discussed in Remark 4.3.

Remark 4.5: A restrictive DGP within class (1) would be a QFM where the PCA factors coincide with the quantile factors and only the factor loadings are quantile dependent. The representation for such restricted subset of QFM is as follows:

$$X_{it} = \lambda_i'(\tau)f_t + u_{it}(\tau), \text{ for } \tau \in (0,1). \tag{10}$$

As a result, the main objects of interest are the common factors and the quantile-varying loadings. Notice that, if the factors f_t were to be observed, using standard QR of X_{it} on f_t would lead to consistent and asymptotically normally distributed estimators of $\lambda_i(\tau)$ for each i and $\tau \in (0,1)$. However, since f_t are not observable, a feasible two-stage approach is to first estimate the factors by PCA, denoted as $\hat{f}_{PCA,t}$, and next run QR of X_{it} on $\hat{f}_{PCA,t}$ to obtain

estimates of $\lambda_i(\tau)$ as follows:

$$\hat{\lambda}_i(\tau) = \arg\min_{\lambda} T^{-1} \sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda' \hat{f}_{PCA,t}). \tag{11}$$

As explained in Chen et al. (2017), unlike the QFA estimators (see Remark 1.2), this two-stage procedure requires moments of the idiosyncratic term u_{it} to be bounded in order to apply PCA in the first stage. However, an interesting result (see Chen et al. 2017, Theorem 2) is that the standard conditions on the relative asymptotics of N and T allowing for the estimated factors to be treated as known do not hold when applying this two-stage estimation approach. In effect, while these conditions are $T^{1/2}/N \to 0$ for linear factor-augmented regressions (see Bai and Ng 2006) and $T^{5/8}/N \to 0$ for nonlinear factor-augmented regressions (Bai and Ng 2008a), lack of smoothness in the object (check) function at the second stage requires the stronger condition $T^{5/4}/N \to 0$. Moreover, Theorem 3 in Chen et al. (2017) shows how to run inference on the quantile-varying loadings (e.g., testing the null that they are constant across all quantiles or a subset of them).

5   Finite Sample Simulations

In this section we report the results from several Monte Carlo simulations regarding the performance of our proposed QFM methodology in finite samples. In particular, we focus on four relevant issues: (i) how well does our preferred estimator of the number of factors perform relative to other methods when the distribution of the idiosyncratic errors in an AFM exhibits heavy tails or outliers, (ii) how well do PCA and QFA estimate the true factors under the previous circumstances, (iii) how robust is the QFA estimation procedure when the errors terms are serially and cross-sectionally correlated, instead of being independent, and (iv) how good are the normal approximations given in Theorem 4 for the QFA estimators based on SQR.

5.1   Estimation of AFM with Outliers

As pointed out in Remark 1.2, since the consistency of our QFA estimator does not require the moments of the idiosyncratic errors to exist, at $\tau = 0.5$ it can be viewed as a robust QR alternative to the PCA estimators commonly used in practice. For the same token, our estimator of the number of factors should also be more robust to outliers and heavy tails than the IC-based method of BN (2002). In this subsection we confirm the above claims by means of Monte Carlo simulations.

19

We consider the following DGP:

$$X_{it} = \sum_{j=1}^{3} \lambda_{ji} f_{jt} + u_{it},$$

where $f_{1t} = 0.8 f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5 f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = 0.2 f_{3,t-1} + \epsilon_{3t}$, $\lambda_{ji}, \epsilon_{jt}$ are all independent draws from $\mathcal{N}(0,1)$, and $u_{it} \sim i.i.d \ B_{it} \cdot \mathcal{N}(0,1) + (1 - B_{it}) \cdot \text{Cauchy}(0,1)$, where B_{it} are i.i.d Bernoulli random variables with means equal to 0.98 and $\text{Cauchy}(0,1)$ denotes the standard Cauchy distribution. In this way, approximately 2% of the idiosyncratic errors are generated as outliers.

We consider four estimators of the number of factors r: two estimators based on PC_{p1}, IC_{p1} of BN (2002), the Eigenvalue Ratio (ER) estimator proposed by Ahn and Horenstein (2013) and our rank-minimization estimator discussed in subsection 3.2, having chosen

$$P_{NT} = \hat{\sigma}_{N,1}^{k} \cdot \left(\frac{1}{L_{NT}^2} \right)^{1/3}.$$

We set $k = 8$ for all four estimators, and consider $N, T \in \{50, 100, 200, 500\}$.

Table 1 reports the following fractions:

$$[\text{proportion of } \hat{r} < 3 \ , \ \text{proportion of } \hat{r} = 3 \ , \ \text{proportion of } \hat{r} > 3 \ ]$$

for each estimator having run 1000 replications.

It becomes evident from the results in Table 1 that PC_{p1} and IC_{p1} almost always overestimate the number factors, while the ER estimator tends to underestimate them, though to a lesser extent than what PC_{p1} and IC_{p1} overestimate them. By contrast, our rank-minimization estimator chooses accurately the right number of factors as long as $\min\{N, T\} \geq 100$.

Next, to compare the PCA and QFA estimators of the common factors in the previous DGP, we assume that $r = 3$ is known. We first get the PCA estimator (denoted as $\hat{F}_{PCA}$), and then obtain the QFA estimator at $\tau = 0.5$ (denoted $\hat{F}_{QFA}^{0.5}$) using the IQR algorithm. Next, we regress each of the true factors on $\hat{F}_{PCA}$ and $\hat{F}_{QFA}^{0.5}$ separately, and report the average R^2 from 1000 replications in Table 2 as an indicator of how well the space of the true factors is spanned by the estimated factors.[8] As shown in the first three columns of Table 2, while the PCA estimators are not very successful in capturing the true common factors, the QFA estimators approximate them very well, even when N, T are not too large.

As discussed earlier, the overall findings reported in Tables 1 and 2 are in line with our theoretical results. In effect, while the standard PCA estimators of BN (2002) fail to capture

---

[8]All the R^2 we use in this section and the next section are adjusted R^2.

the true factors because they require the eighth moments of all the idiosyncratic errors to be bounded (unlike the DGP above), our QFA estimators succeed to do so since they only need the density function to exist and be continuously differentiable, like in the previous DGP. Thus, this simulation exercise provides strong evidence about the substantial gains that can be achieved by using QFA rather than PCA in those cases where the idiosyncratic error terms in AFM exhibit heavy tails and outliers.

5.2   Estimation of QFM: Heavy-tailed and Dependent Idiosyncratic Errors

In this subsection we consider the following DGP:

$$X_{it} = \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + (\lambda_{3i}f_{3t}) \cdot e_{it},$$

where $f_{1t} = 0.8f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = |g_t|$, $\lambda_{1i}, \lambda_{2i}, \epsilon_{1t}, \epsilon_{2t}, g_t$ are all independent draws from $\mathcal{N}(0,1)$, and λ_{3i} are independent draws from $U[1,2]$. Following BN (2002), the following specification for e_{it} is used:

$$e_{it} = \beta\, e_{i,t-1} + v_{it} + \rho \cdot \sum_{j=i-J, j\neq i}^{i+J} v_{jt},$$

where v_{it} are independent draws from $\mathcal{N}(0,1)$ except in the second case below. The autoregressive coefficient β captures the serial correlation of e_{it}, while the parameters ρ and J capture the cross-sectional correlations of e_{it}. We consider four cases:

Case 1: Independent errors: $\beta = 0$ and $\rho = 0$.
Case 2: Independent errors with heavy tails: $\beta = \rho = 0$, and $v_{it} \sim i.i.d$ Student(3).
Case 3: Serially correlated errors: $\beta = 0.2$ and $\rho = 0$.
Case 4: Serially and cross-sectionally correlated errors: $\beta = 0.2$ and $\rho = 0.2$, and $J = 3$.

For each of the previous cases and each $\tau \in \{0.25, 0.5, 0.75\}$, we first estimate $\hat{r}$ using our rank-minimization estimator, having set k and P_{NT} as described in the previous subsection. Second, we estimate $\hat{r}$ factors by means of the QFA estimation approach, which we denote $\hat{F}^\tau_{QFA}$. Finally, we regress each of the true factors on $\hat{F}^\tau_{QFA}$ and calculate the R^2s. This procedure is repeated 1000 times where, for each τ, we report the averages of $\hat{r}$ and the R^2s in these 1000 replications.

The results for Case 1 and Case 2 (where the heavy tails are captured this time by a Student(3) rather than by a Cauchy distribution) are reported in Table 3 and Table 4, respectively, for $N, T \in \{50, 100, 200\}$. Notice that for $\tau = 0.25, 0.75$, we have $r(\tau) = 3$ while, for $\tau = 0.5$, we get $r(\tau) = 2$, since the factor f_{3t} does not affect the median of X_{it}. It can be observed that both our selection criterion and the QFA estimators perform very well in choosing the number of

21

QFA factors and in estimating them. It should be noticed that at $\tau = 0.25, 0.75$ the estimation of the scale factor f_{3t} is not as good as the mean factors f_{1t}, f_{2t} for small N and T. However, such differences vanish as N and T increase.

The results for Case 3 and Case 4 are in turn reported in Table 5 and Table 6, respectively. It can be inspected that the QFA estimators still perform well, even though the independence assumption is violated in these DGPs. Thus, despite adopting independence in Assumption 1 (iii) for tractability in the proofs (see Remark 1.4), it seems that QFA estimation still works properly when the error terms are allowed to exhibit mild serial and cross-sectional correlations.

5.3   Normal Approximations of the Estimators Based on SQR

To evaluate the normal approximations of Theorem 4 for the estimators based on SQR, we consider the following DGP:

$$X_{it} = \lambda_i f_t + f_t \epsilon_{it},$$

where $f_t \sim i.i.d\, \mathcal{U}(1,2)$ and they are normalized such that $F'F/T = 1$, $\lambda_i \sim i.i.d\, \mathcal{N}(0,1)$ and $\epsilon_{it} \sim i.i.d\, \mathcal{N}(0,1)$. Note that since our Theorem 4 is conditional on the factors and the loadings, f_t and λ_i are fixed in the simulations. To smooth the indicator function, we use the following eighth-order kernel function (see Muller, 1984):

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

while the Epanechnikov kernel $l(z) = 0.75(1 - z^2) \cdot \mathbf{1}\{|z| \leq 1\}$ is applied to estimate the variance.

Figure 1 and Figure 2 plot the histograms of the standardized estimators of the factors: $\hat{V}_{f_t}^{-1}\sqrt{N}(\tilde{f}_t - f_{0t})$ at $\tau = 0.25$, $t = T/2$ from 1000 replications[9], where $\hat{V}_{f_t}$ is estimated using the formula in Remark 4.4. To check how the bandwidths affect the finite distributions of the estimators, we display results for different choices of h and b.

From the reported histograms of the standardized estimators and the superimposed density function of the standard normal distribution, it becomes clear that the asymptotic distributions given in Theorem 4 provide reasonably good approximations for the finite sample distributions of the estimators based on SQR, even for $N = T = 50$, and that such approximations are not very sensitive to the choice of bandwidths.

---

[9]We choose the signs of $\tilde{f}_t$ such that $\mathsf{S} = 1$.

6 Empirical Applications

In this section we argue that QFA could provide a useful tool for causal analysis, predictive exercises and the economic interpretation of factors. In particular, we focus on applying our proposed methodology to three different datasets related to climate, macro aggregates and stock returns.

6.1 Climate Change and CO_2 Emissions

In our first empirical application we investigate how CO_2 emissions affect temperatures, which is a long-standing issue in climate change science and economics (see e.g. Hansen et al. 1981, and Hsiang and Kopp 2018). The dataset (coined Climate for short) we use consists of the annual changes of temperature from 441 stations from 1917 to 2018 ($N = 441, T = 102$), drawn from the Climate Research Unit at the University of East Anglia, where information about global temperatures across different stations in the Northern and Southern Hemisphere is collected. The annual global CO_2 emissions data is downloaded from The Global Change Data Lab.

Table 7 (column labeled Climate) reports the estimated number of factors using PC_{p1} of BN (2002), the ER estimator and the rank-minimization estimator for a grid of quantiles ranging from 0.01 to 0.99.[10] The maximum number of factors k is set to 8 for all estimators. PC_{p1} selects the maximum number of factors (8), while the ER estimator selects only one. Thus, the tendencies to overestimate (resp. underestimate) the number of PCA factors by the former (resp. latter) criteria mirror our simulation results in Section 5.1. By contrast, it can be observed that the numbers of factors estimated by the rank-minimization vary across quantiles. In particular, the number of QFA factors decreases as we move away from the median.

To compare the QFA factors (denoted as $\hat{F}_{QFA}^{\tau}$) and the PCA factors (denoted as $\hat{F}_{PCA}$), we regress each element of $\hat{F}_{QFA}^{\tau}$ on the 8 PCA factors selected by PC_{p1} and compute the R^2 in these regressions.[11] The results are shown in the upper panel of Table 8, where it becomes clear that the median factors ($\hat{F}_{QFA}^{0.5}$) are highly correlated with the PCA factors, with all the R^2s above 0.95. By contrast, the QFA factors at the upper and lower quantiles ($\tau = 0.01, 0.05, 0.95, 0.99$) exhibit much lower correlations with the PCA factors, with R^2s around 0.60. Thus, there seems to be room for using QFA in this application.

Next, to analyze the impact of CO_2 emissions on climate change, bivariate Granger non-causality tests are implemented. We regress the QFA factors at each relevant quantile on their own lags and the lagged growth rates of CO_2 emissions, labeled $\Delta \log(CO_2)$, where the lag length is chosen according to BIC. Table 9 reports the p-values of these tests. The results of this novel

---

[10]In all the applications, before estimating the factors and the number of factors, each variable is standardized to have zero mean and variance equal to one.

[11]We choose the number of PCA factors estimated by PC_{p1} in these regressions to play conservative.

approach to analyze quantile causality indicate that the growth rate of CO_2 emissions strongly Granger causes the QFA factors at the lower quantiles ($\tau = 0.01, 0.05$), with p-values below 0.01, as well as some of the median factors, albeit to a lessser extent (p-values below 0.04). Moreover, the null of Granger non-causality is not rejected for QFA factors at the upper quantiles. Given that CO_2 emissions lead to global warming, the results for the lower quantiles of temperatures are in line with the evidence reported by Gadea and Gonzalo (2020). Using a similar climate dataset but different quantile techniques to ours, these authors find that global warming over the last century seems to be mainly due to a different behaviour in the lower tail than in the central and upper tails of the distribution of global temperatures.

6.2 Macroeconomic Forecasting in a Data-Rich Environment

In the second application, we extend the diffusion-index forecasting exercise popularized by Stock and Watson (2002) to explore the predictive power of the QFA factors. The main goal is to extract a few common factors (by both PCA and QFA) from a large panel of macroeconomic variables, and then use these factors to forecast e.g. real GDP growth and the inflation rate.

The FRED-QD dataset (coined Macro here) is used to estimate PCA and QFA factors. This is a quarterly panel consists of 211 US macroeconomic variables from 1960Q1 to 2019Q2 ($N = 211, T = 238$). It emulates the popular dataset used by Stock and Watson (2002), but also contains several additional time series. The variables in this dataset are updated in a timely manner and can be downloaded for free.[12] Before estimation, each series is transformed to be stationary using Matlab codes that are also available on the FRED-QD data website.

As with the climate data, while PC_{p1} selects 8 PCA factors, ER selects only one. The estimated numbers of QFA factors are reported in Table 7 (column labeled Macro). As can be seen, the number of QFA factors varies significantly across different quantiles, pointing to the existence of a nonstandard factor structure for this dataset. Moreover, the middle panel of Table 8 reports the R^2s of regressing each of the QFA factors on the 8 PCA factors. It becomes clear that the QFA factors at τ close to 0.5 are all well explained by the PCA factors. However, the first QFA factor at $\tau = 0.9$ (denoted $\hat{F}_{QFA}^{0.9}$) and those at $\tau = 0.95, 0.99$ (denoted as $\hat{F}_{QFA}^{0.95}$ and $\hat{F}_{QFA}^{0.99}$) contain some extra information that could be potentially helpful for forecasting macroeconomic variables. Since $\hat{F}_{QFA}^{0.95}$ exhibits a very high correlation with $\hat{F}_{QFA}^{0.9}$ and $\hat{F}_{QFA}^{0.99}$, we exclusively focus on the predictive power of $\hat{F}_{QFA}^{0.9}$ and $\hat{F}_{QFA}^{0.99}$ in the subsequent analysis.

Let y_{t+1} denote the realized value of real GDP growth/inflation at period $t + 1$. The fore-

---

[12]Link to the dataset: http://research.stlouisfed.org/econ/mccracken/. We refer to McCracken and Ng (2016) for the details of a very similar dataset that contains monthly macroeconomic variables.

casting equation we consider is as follows:

$$y_{t+1} = \alpha + \sum_{j=0}^{p_{max}} \beta_j y_{t-j} + \gamma' F_t + \epsilon_{t+1},$$

where F_t is vector containing several unobserved common factors extracted from the large macroeconomic dataset. The predicted value of y_{t+1}, based on a vector of estimated factors $\hat{F}_t$, is simply constructed as $\hat{y}_{t+1} = \hat{\alpha} + \sum_{j=0}^{\hat{p}} \hat{\beta}_j y_{t-j} + \hat{\gamma}' \hat{F}_t$, where $\hat{\alpha}, \hat{\beta}_j, \hat{\gamma}$ are OLS estimates of the coefficients and $\hat{p}$ is the optimal lag length according to BIC. We compare five different specifications for F_t: (i) $F_t = 0$, which is the benchmark AR model, (ii) AR plus $\hat{F}_t$ only including $\hat{F}_{PCA}$, (iii) AR plus $\hat{F}_t$ including $\hat{F}_{PCA}$ and $\hat{F}_{QFA}^{0.9}$, (iv) AR plus $\hat{F}_t$ including $\hat{F}_{PCA}$ and $\hat{F}_{QFA}^{0.99}$, and (v) AR plus $\hat{F}_t$ including $\hat{F}_{PCA}$, $\hat{F}_{QFA}^{0.9}$ and $\hat{F}_{QFA}^{0.99}$. Following Chudik et al. (2018), the initial estimation period is 1960Q1 to 1989Q4 (120 periods), and the forecast evaluation period is split into pre-crisis (1990Q1 to 2007Q2) and crisis and recovery (2007Q3 to 2019Q2) sub-periods. A rolling window of 120 periods is used both to estimate the coefficients and generate the rolling forecasts. In particular, following Chudik et al. (2018), the number of mean factors is estimated using PC_{p1} at each rolling window, where the maximum number of factors is set equal to 5.

The mean squared error (MSE) of these procedures, and their relative MSE (R-MSE) to the benchmark AR model are reported in Table 10 for the whole evaluation period and each relevant sub-sample. As can be observed, in regards to real GDP growth, adding the upper tail QFA factors ranks better in terms of R-MSE than the AR and AR+$\hat{F}_{PCA}$ models for the three considered periods. The gains are not sizeable but yet they are relevant. As for the inflation rate, the results are weaker, though there are some gains for the crisis and recovery period.

A well-known shortcoming of point forecasts is that their uncertainty is generally unknown, hence it is difficult to quantify their precision at any given period of time. To address this problem, it has became customary among central banks to report density forecasts for important macroeconomic variables. In this respect, Adrian et al. (2019) argue that a simple way of producing such densities is via QR. Following these authors, we next evaluate the predictive power of the QFA factors for forecasting the densities of real GDP growth and inflation. In particular, we first predict the conditional quantiles of the target variable y_{t+h} by $\hat{q}_{\tau,t+h} = \hat{\alpha}_\tau + \sum_{j=0}^{p} \hat{\beta}_{\tau,j} y_{t-j} + \hat{\gamma}_\tau' \hat{F}_{\tau,t}$ for $\tau \in \{0.05, 0.25, 0.75, 0.95\}$, where $\hat{\alpha}_\tau, \hat{\beta}_{\tau,j}, \hat{\gamma}_\tau$ are estimated coefficients by running QR of y_{t+h} on $[1, y_t, \ldots, y_{t-p}, \hat{F}_{\tau,t}]$, and $\hat{F}_{\tau,t}$ is a vector of estimated quantile factors using the IQR algorithm.[13] Next, given the predicted quantiles: $[\hat{q}_{0.05,t+h}, \hat{q}_{0.25,t+h}, \hat{q}_{0.75,t+h}, \hat{q}_{0.95,t+h}]$, the predicted density of y_{t+h} is constructed as the density of a skewed t-distribution by matching the predicted quantiles.[14] Finally, the accuracy of the density forecast is measured by the predictive score, which is the predicted density evaluated at the realized value of y_{t+h}. Higher predictive scores

---

[13]Only 1 QFA factor is estimated at $\tau = 0.05, 0.95$, whereas 5 QFA factors are estimated at $\tau = 0.25, 0.75$.

[14]We refer to Adrian et al. (2019) for the details and to Azzalini and Capitanio (2003) for the definition and properties of the skewed t-distribution.

indicate more accurate predictions. The out-of-sample density forecasts are constructed using rolling windows with the most recent 120 observations, and the evaluation period is 1990Q1 to 2019Q2. Moreover, we set $p = 3$, and the benchmark model is the one where $\hat{F}_{\tau,t} = 0$, i.e. the quantiles of y_{t+h} are predicted only using its own lags. Figure 3 displays the predictive scores of the one-quarter-ahead ($h = 1$) and one-year-ahead ($h = 4$) density forecasts for both variables. It can be seen that in both instances the predictive scores of the "AR + Quantile Factors" procedure is frequently above that of the "AR benchmark" model, sometimes by a large margin, indicating that the QFA factors could indeed be very informative for density forecasting of highly relevant macroeconomic variables.

6.3 Interpretation of Financial Factors

Our last application concerns the interpretation of the quantile factors extracted from financial asset returns. The dataset (Finance in short) contains the monthly returns of 429 stocks from 1980M01 to 2014M12 ($N = 429, T = 420$), obtained from The Center of Research for Security Prices (CRSP).[15]

Except at $\tau = 0.5$, the estimated number of QFA factors reported in Table 7 (column labeled Finance) are all equal to 1, which agrees with the choice of PCA factors by the ER estimator but again is less than the 4 factors selected by PC_{p1}.[16]

The lower panel of Table 8 reports the R^2s of regressing each of the QFA factors on the 8 PCA factors. As can be inspected, most of these factors are well explained by the PCA factors, with the exception of those at $\tau = 0.01, 0.99$, where the R^2s are below 60%. Interestingly, as discussed in the Introduction this evidence is seemingly consistent with the findings of the financial literature on the existence of tail factors in the distribution of asset returns, as reported e.g. by Andersen et al. (2018). Thus, it is interesting to check whether the extra quantile factors at the lower tail and upper tail of the returns distribution could yield some confirmation of that hypothesis.

First, as shown in the upper panel of Figure 4, the QFA factors at $\tau = 0.01$ and 0.99 (both with variance = 0.12) are much less volatile than those at $\tau = 0.5$ (both with variance = 1), meaning that the tails of the distributions of returns are more stable than the median. Second, we find that the interquantile range (defined as the difference between the quantile factors at $\tau = 0.99$ and 0.01) provides a good measure of uncertainty for financial markets.[17] Finally, as shown in the lower panel of Figure 4, the interquantile range is highly correlated with the volatility factor (with a correlation of 0.87) constructed by applying PCA-SQ to the squared

---

[15]The panel is balanced by only keeping stocks that have no missing observations during this time period.

[16]PC_{p3} and IC_{p3} of BN (2002) chose 8 factors while all the other 6 information criteria choose 4 factors.

[17]The results with the interpercentile range and interquartile range turn out to be similar.

residuals of an AFM.[18] On the contrary, the correlations between the two median factors and the volatility factor only reach 0.08 and -0.05, respectively. Thus, this evidence seems supportive of the the presence of extra common factors affecting the tails and the volatility of the asset returns, with our results providing a link between them.

7   Conclusions

Approximate Factor Models (AFM) have become a leading methodology for the joint modelling of large number of economic time series with the big improvements in data collection and information technologies. This first generation of AFM was designed to reduce the dimensionality of big datasets through finding those common components (mean factors) which, by shifting the means of the observed variables with different intensities, are able to capture a large fraction of their co-movements. However, one could envisage the existence of other common factors that do not (or not only) shift the means but also affect other distributional characteristics (volatility, higher moments, extreme values, etc.). This calls for a second generation of factor models.

Inspired by the generalization of linear regressions to quantile regressions (QR), this paper proposes Quantile Factor Models (QFM) as a new class of factor models. In QFM, both factors and loadings are allowed to be quantile-dependent objects. These extra factors could be useful for identification purposes, for instance mean factors vs. volatility/skewness/kurtosis factors, as well as for forecasting purposes in factor-augmented regressions and FAVAR setups.

Using tools in the interface of QR, Principal Component Analysis (PCA) and the theory of empirical processes, we propose an estimation procedure of the quantile-dependent objects in QFM, labelled Quantile Factor Analysis (QFA), which yields consistent and asymptotically normal estimators of factors and loadings at each quantile. An important advantage of QFA is that it is able to extract simultaneously all mean and extra (non-mean) factors determining the factor structure of QFM, in contrast to PCA which can only extract mean factors. In addition, we propose novel selection criteria to estimate consistently the number of factors at each quantile. Finally, another relevant result is that QFA estimators remain valid (under some restrictive assumption on the idiosyncratic error terms – see Assumption1 (iii) – which are adopted to simplify the proofs) when the idiosyncratic error terms in AFM exhibit heavy tails and outliers, a case where PCA is rendered invalid.

The previous theoretical findings receive support in finite samples from a range of Monte Carlo simulations. Furthermore, it is shown in these simulations that QFA estimation performs well when we depart from some of simplifying assumptions used in the theory section for tractability, like, e.g., independence of the idiosyncratic errors. Lastly, our empirical applica-

---

[18]Following Renault et al. (2017), we first project out the 8 PCA factors from the returns, and the volatility factor is obtained as the cross-sectional average of the squared residuals.

tions to three large panel datasets of financial, macro and climate variables provide evidence that some of these extra factors may be highly relevant in practice for causality analysis, forecasting, and economic interpretation purposes.

Any time a novel methodology is proposed, new research issues emerge for future investigation. Among the ones which have been left out of this paper (some are part of our current research agenda), four topics stand out as important:

- Factor augmented regressions and FAVAR: In relation to this topic, it would also be interesting to check in great detail the contributions of the extra factors in forecasting and monitoring multivariate systems. This is an issue of high interest for applied researchers, especially with the surge of Big Data technologies. For example, one could analyze the role of the extra factors in the estimation and shock identification in FAVAR. Recent developments in quantile VAR estimation, as in White et al. (2015), provide useful tools in addressing these issues.

- Relaxing the independence assumptions: In view of the simulation results in Tables 5 and 6, we conjecture that the main theoretical results of our paper continue to hold when the error terms in QFM are allowed to have weak cross-sectional and serial dependence. Providing a formal justification for this conjecture remains high in our research agenda. As discussed in Remark 1.4, the goal here is to provide more general conditions on u_{it} under which the sub-Gaussian type inequalities still hold.

- Dynamic QFM: Although our methodology admits factors to have dependence, provided Assumption 2(i) holds, there is still the pending issue of how to extend our results for static QFM to dynamic QFM, where the set of quantile-dependent variables include lagged factors (see Forni et al. 2000 and Stock and Watson 2011). Since our main aim in this paper has been to introduce the new class of QFM and their basic properties, for the sake of brevity, we have focused on static QFM, leaving this topic for further research.

- Economic interpretation of QFA factors in empirical applications: Given the evidence that extra factors could be relevant in practice, another interesting issue is how to interpret them in different economic and financial setups. As illustrated in subsection 6.3, once the econometric techniques to detect and estimate extra factors in QFM have been established, attempts to provide new economic insights for these objects would help enrich the economic theory underlying this type of factor structures.

A   Tables and Figures

Table 1: AFM with Outliers in the Idiosyncratic Errors: Estimating the Number of Factors

N	T	PC_{p1} of BN			IC_{p1} of BN			Eigenvalue Ratio			Rank Estimator		
50	50	[0.00	0.04	0.96]	[0.00	0.14	0.86]	[0.26	0.30	0.44]	[0.47	0.53	0.00]
50	100	[0.00	0.02	0.98]	[0.00	0.05	0.95]	[0.33	0.19	0.48]	[0.40	0.60	0.00]
50	200	[0.00	0.00	1.00]	[0.00	0.01	0.99]	[0.41	0.12	0.47]	[0.33	0.67	0.00]
50	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.56	0.07	0.37]	[0.29	0.71	0.00]
100	50	[0.00	0.02	0.98]	[0.00	0.05	0.95]	[0.34	0.18	0.48]	[0.39	0.61	0.00]
100	100	[0.00	0.00	1.00]	[0.00	0.01	0.99]	[0.41	0.13	0.46]	[0.10	0.90	0.00]
100	200	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.48	0.07	0.45]	[0.06	0.94	0.00]
100	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.65	0.05	0.30]	[0.02	0.98	0.00]
200	50	[0.00	0.00	1.00]	[0.00	0.01	0.99]	[0.45	0.10	0.45]	[0.37	0.63	0.00]
200	100	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.48	0.08	0.44]	[0.10	0.90	0.00]
200	200	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.63	0.06	0.31]	[0.00	1.00	0.00]
200	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.76	0.08	0.16]	[0.00	1.00	0.00]
500	50	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.57	0.08	0.35]	[0.36	0.64	0.00]
500	100	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.68	0.06	0.26]	[0.05	0.95	0.00]
500	200	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.76	0.08	0.16]	[0.00	1.00	0.00]
500	500	[0.00	0.00	1.00]	[0.00	0.00	1.00]	[0.80	0.10	0.10]	[0.00	1.00	0.00]

Note: The DGP considered in this Table is: $X_{it} = \sum_{j=1}^{3} \lambda_{ji} f_{jt} + u_{it}$, where $f_{1t} = 0.8 f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5 f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = 0.2 f_{3,t-1} + \epsilon_{3t}$, $\lambda_{ji}, \epsilon_{jt} \sim i.i.d \, \mathcal{N}(0,1)$, $u_{it} \sim i.i.d \, B_{it} \cdot \mathcal{N}(0,1) + (1 - B_{it}) \cdot \text{Cauchy}(0,1)$ where $B_{it} \sim i.i.d \, \text{Bernoulli}(0.98)$. For each estimation method, the [proportion of $\hat{r} < 3$, proportion of $\hat{r} = 3$, proportion of $\hat{r} > 3$] is reported from 1000 replications.

Table 2: AFM with Outliers in the Idiosyncratic Errors: Estimation of the Factors

		Regress F on $\hat{F}_{PCA}$			Regress F on $\hat{F}_{QFA}^{0.5}$		
N	T	f_1	f_2	f_3	f_1	f_2	f_3
50	50	0.939	0.810	0.686	0.987	0.975	0.968
50	100	0.931	0.718	0.578	0.987	0.975	0.968
50	200	0.890	0.589	0.412	0.987	0.975	0.968
50	500	0.807	0.405	0.252	0.988	0.975	0.968
100	50	0.928	0.738	0.595	0.993	0.986	0.984
100	100	0.921	0.630	0.441	0.994	0.988	0.984
100	200	0.857	0.479	0.285	0.994	0.988	0.985
100	500	0.713	0.294	0.138	0.994	0.988	0.984
200	50	0.890	0.657	0.513	0.997	0.994	0.992
200	100	0.858	0.514	0.333	0.997	0.994	0.993
200	200	0.779	0.358	0.178	0.997	0.994	0.992
200	500	0.530	0.131	0.051	0.997	0.994	0.992
500	50	0.819	0.501	0.371	0.998	0.997	0.996
500	100	0.725	0.327	0.196	0.999	0.998	0.997
500	200	0.546	0.165	0.062	0.999	0.998	0.997
500	500	0.273	0.036	0.018	0.999	0.998	0.997

Note: The DGP considered in this Table is: $X_{it} = \sum_{j=1}^{3} \lambda_{ji} f_{jt} + u_{it}$, where $f_{1t} = 0.8 f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5 f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = 0.2 f_{3,t-1} + \epsilon_{3t}$, $\lambda_{ji}, \epsilon_{jt} \sim i.i.d \ \mathcal{N}(0,1)$, $u_{it} \sim i.i.d \ B_{it} \cdot \mathcal{N}(0,1) + (1 - B_{it}) \cdot \text{Cauchy}(0,1)$ where $B_{it} \sim i.i.d \ \text{Bernoulli}(0.98)$. For each estimation method, we report the average R^2 in the regression of (each of) the true factors on the estimated factors by PCA and QFA (assuming the number of factors to be known).

Table 3: Estimation of QFM: Independent Idiosyncratic Errors

N	T	$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
		$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}
50	50	2.21	0.866	0.721	0.339	1.91	0.956	0.808	0.013	2.23	0.926	0.738	0.334
50	100	2.42	0.943	0.758	0.483	1.88	0.968	0.839	0.003	2.38	0.946	0.708	0.463
50	200	2.43	0.933	0.703	0.485	1.88	0.971	0.842	0.001	2.40	0.951	0.698	0.445
100	50	2.14	0.944	0.681	0.337	1.80	0.980	0.786	0.014	2.13	0.948	0.694	0.357
100	100	2.71	0.977	0.898	0.688	1.98	0.985	0.954	0.001	2.72	0.968	0.890	0.707
100	200	2.82	0.983	0.904	0.757	1.99	0.987	0.966	0.003	2.86	0.982	0.908	0.793
200	50	2.35	0.970	0.826	0.490	1.87	0.989	0.867	0.008	2.29	0.973	0.745	0.489
200	100	2.80	0.990	0.934	0.782	2.00	0.993	0.987	0.001	2.81	0.990	0.977	0.772
200	200	2.99	0.992	0.986	0.940	2.00	0.994	0.988	0.000	2.99	0.992	0.986	0.935

Note: The DGP considered in this Table is: $X_{it} = \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + (\lambda_{3i}f_{3t}) \cdot e_{it}$, $f_{1t} = 0.8f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = |g_t|$, $\lambda_{1i}, \lambda_{2i}, \epsilon_{1t}, \epsilon_{2t}, g_t \sim i.i.d\ \mathcal{N}(0,1)$, and $\lambda_{3i} \sim i.i.d\ U[1,2]$. $e_{it} = \beta e_{i,t-1} + v_{it} + \rho \cdot \sum_{j=i-J, j\neq i}^{i+J} v_{jt}$, $v_{it} \sim i.i.d\ \mathcal{N}(0,1)$, $\beta = \rho = 0$. For each τ, the first column reports the averages of the rank estimator $\hat{r}$ from 1000 replications, while the second to the fourth columns report the average R^2 in the regression of (each of) the true factors on the QFA factors $\hat{F}_{QFA}^{\tau}$, obtained from the IQR algorithm.

Table 4: Estimation of QFM: Independent Idiosyncratic Errors with Heavy Tails

N	T	$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
		$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}
50	50	2.81	0.911	0.727	0.585	2.38	0.954	0.827	0.031	2.95	0.925	0.711	0.617
50	100	2.79	0.934	0.782	0.621	2.03	0.963	0.885	0.005	2.79	0.933	0.783	0.658
50	200	2.82	0.942	0.811	0.680	1.91	0.966	0.855	0.000	2.76	0.943	0.790	0.648
100	50	3.20	0.962	0.851	0.737	2.67	0.977	0.907	0.076	3.07	0.942	0.828	0.682
100	100	3.06	0.972	0.897	0.840	2.21	0.983	0.939	0.018	3.06	0.974	0.931	0.801
100	200	3.00	0.974	0.944	0.867	1.99	0.983	0.958	0.000	2.98	0.974	0.943	0.860
200	50	3.24	0.971	0.839	0.753	2.82	0.984	0.903	0.106	3.31	0.970	0.858	0.773
200	100	3.10	0.985	0.937	0.897	2.31	0.991	0.975	0.018	3.09	0.987	0.949	0.883
200	200	3.02	0.989	0.977	0.932	2.07	0.992	0.985	0.005	3.02	0.988	0.978	0.933

Note: The DGP considered in this Table is: $X_{it} = \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + (\lambda_{3i}f_{3t}) \cdot e_{it}$, $f_{1t} = 0.8f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = |g_t|$, $\lambda_{1i}, \lambda_{2i}, \epsilon_{1t}, \epsilon_{2t}, g_t \sim i.i.d\ \mathcal{N}(0,1)$, and $\lambda_{3i} \sim i.i.d\ U[1,2]$. $e_{it} = \beta e_{i,t-1} + v_{it} + \rho \cdot \sum_{j=i-J, j\neq i}^{i+J} v_{jt}$, $v_{it} \sim i.i.d$ Student(3), $\beta = \rho = 0$. For each τ, the first column reports the averages of the rank estimator $\hat{r}$ from 1000 replications, while the second to the fourth columns report the averages of R^2 in the regression of (each of) the true factors on the QFA factors $\hat{F}_{QFA}^{\tau}$, obtained from the IQR algorithm.

Table 5: Estimation of QFM: Serially Correlated Idiosyncratic Errors

N	T	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}
		$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
50	50	2.31	0.900	0.698	0.400	1.97	0.961	0.805	0.023	2.32	0.924	0.705	0.416
50	100	2.40	0.927	0.722	0.475	1.91	0.968	0.863	0.005	2.38	0.940	0.709	0.453
50	200	2.66	0.956	0.841	0.586	1.95	0.970	0.904	0.000	2.70	0.948	0.824	0.628
100	50	2.33	0.945	0.736	0.479	1.91	0.980	0.857	0.005	2.32	0.942	0.737	0.478
100	100	2.72	0.978	0.863	0.704	1.98	0.985	0.957	0.000	2.72	0.978	0.895	0.690
100	200	2.87	0.983	0.924	0.801	1.98	0.987	0.955	0.000	2.88	0.965	0.948	0.805
200	50	2.35	0.974	0.724	0.540	1.92	0.989	0.859	0.021	2.40	0.963	0.758	0.531
200	100	2.75	0.987	0.929	0.734	1.98	0.993	0.960	0.000	2.76	0.990	0.912	0.760
200	200	2.98	0.993	0.984	0.927	2.00	0.994	0.987	0.000	2.99	0.992	0.975	0.942

Note: The DGP considered in this Table is: $X_{it} = \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + (\lambda_{3i}f_{3t}) \cdot e_{it}$, $f_{1t} = 0.8f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = |g_t|$, $\lambda_{1i}, \lambda_{2i}, \epsilon_{1t}, \epsilon_{2t}, g_t \sim i.i.d \; \mathcal{N}(0,1)$, and $\lambda_{3i} \sim i.i.d \; U[1,2]$. $e_{it} = \beta * e_{i,t-1} + v_{it} + \rho \cdot \sum_{j=i-J,j\neq i}^{i+J} v_{jt}$, $v_{it} \sim i.i.d \; \mathcal{N}(0,1)$, $\beta = 0.2$, $\rho = 0$. For each τ, the first column reports the average rank estimator $\hat{r}$ from 1000 replications, while the second to the fourth columns report the average R^2 in the regression of (each of) the true factors on the QFA factors $\hat{F}_{QFA}^\tau$, obtained from the IQR algorithm.

Table 6: Estimation of QFM: Serially and Cross-Sectionally Correlated Idiosyncratic Errors

N	T	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}	$\hat{r}$	f_{1t}	f_{2t}	f_{3t}
		$\tau = 0.25$				$\tau = 0.5$				$\tau = 0.75$			
50	50	2.54	0.926	0.705	0.409	2.16	0.952	0.808	0.029	2.53	0.921	0.700	0.423
50	100	2.49	0.941	0.703	0.397	1.95	0.959	0.845	0.001	2.50	0.934	0.723	0.423
50	200	2.66	0.945	0.803	0.460	1.97	0.963	0.881	0.000	2.64	0.939	0.756	0.471
100	50	2.52	0.942	0.780	0.495	2.02	0.977	0.820	0.021	2.41	0.946	0.744	0.472
100	100	2.91	0.976	0.896	0.697	2.06	0.981	0.945	0.006	2.87	0.977	0.893	0.686
100	200	2.90	0.979	0.924	0.702	2.01	0.983	0.966	0.000	2.92	0.980	0.933	0.713
200	50	2.47	0.967	0.732	0.569	2.05	0.987	0.870	0.032	2.52	0.969	0.785	0.576
200	100	2.88	0.989	0.913	0.802	2.00	0.991	0.982	0.000	2.89	0.989	0.938	0.788
200	200	3.00	0.990	0.982	0.866	2.00	0.992	0.983	0.000	3.00	0.990	0.981	0.866

Note: The DGP considered in this Table is: $X_{it} = \lambda_{1i}f_{1t} + \lambda_{2i}f_{2t} + (\lambda_{3i}f_{3t}) \cdot e_{it}$, $f_{1t} = 0.8f_{1,t-1} + \epsilon_{1t}$, $f_{2t} = 0.5f_{2,t-1} + \epsilon_{2t}$, $f_{3t} = |g_t|$, $\lambda_{1i}, \lambda_{2i}, \epsilon_{1t}, \epsilon_{2t}, g_t \sim i.i.d \; \mathcal{N}(0,1)$, and $\lambda_{3i} \sim i.i.d \; U[1,2]$. $e_{it} = \beta e_{i,t-1} + v_{it} + \rho \cdot \sum_{j=i-J,j\neq i}^{i+J} v_{jt}$, $v_{it} \sim i.i.d \; \mathcal{N}(0,1)$, $\beta = \rho = 0.2$ and $J = 3$. For each τ, the first column reports the average rank estimator $\hat{r}$ from 1000 replications, while the second to fourth columns report the average R^2 in the regression of (each of) the true factors on the QFA factors $\hat{F}_{QFA}^\tau$, obtained from the IQR algorithm.

Table 7: All Empirical Applications: Number of Factors

	Climate	Macro	Finance
(N, T)	(441,102)	(211,238)	(429,420)
PC_{p1}	8	8	4
ER	2	1	1
$\hat{r}_{\text{rank}}\ \tau = 0.01$	1	1	1
$\hat{r}_{\text{rank}}\ \tau = 0.05$	1	1	1
$\hat{r}_{\text{rank}}\ \tau = 0.10$	2	2	1
$\hat{r}_{\text{rank}}\ \tau = 0.25$	4	4	1
$\hat{r}_{\text{rank}}\ \tau = 0.50$	6	5	2
$\hat{r}_{\text{rank}}\ \tau = 0.75$	4	5	1
$\hat{r}_{\text{rank}}\ \tau = 0.90$	2	2	1
$\hat{r}_{\text{rank}}\ \tau = 0.95$	1	1	1
$\hat{r}_{\text{rank}}\ \tau = 0.99$	1	1	1

Note: This Table provides the estimated numbers of mean factors using PC_{p1} of BN (2002), the ER estimator of Ahn and Horenstein (2013), and the estimated numbers of quantile factors at $\tau \in \{0.01, 0.05, 0.1, 0.25, 0.75, 0.9, 0.95, 0.99\}$ using the rank-minimization estimator proposed in subsection 3.2.

Table 8: All Empirical Applications: Comparison of $\hat{F}_{QFA}$ and $\hat{F}_{PCA}$

Dataset	τ	Elements of $\hat{F}_{QFA}^{\tau}$					
		1	2	3	4	5	6
Climate	0.01	0.599					
	0.05	0.623					
	0.10	0.759	0.848				
	0.25	0.939	0.961	0.965	0.941		
	0.50	0.995	0.995	0.992	0.988	0.980	0.970
	0.75	0.950	0.961	0.966	0.933		
	0.90	0.755	0.905				
	0.95	0.629					
	0.99	0.567					
Macro	0.01	0.657					
	0.05	0.733					
	0.10	0.796	0.871				
	0.25	0.952	0.932	0.939	0.890		
	0.50	0.993	0.976	0.964	0.945	0.923	
	0.75	0.906	0.945	0.943	0.903	0.882	
	0.90	0.316	0.911				
	0.95	0.261					
	0.99	0.266					
Finance	0.01	0.560					
	0.05	0.731					
	0.10	0.803					
	0.25	0.921					
	0.50	0.993	0.977				
	0.75	0.945					
	0.90	0.783					
	0.95	0.660					
	0.99	0.492					

Note: This Table reports the R^2 of regressing each element of $\hat{F}_{QFA}$ on $\hat{F}_{PCA}$. For $\hat{F}_{QFA}$, the numbers of estimated factors is obtained from Table 7 while, for $\hat{F}_{PCA}$, the numbers of estimated factors are 8 for all datasets.

Table 9: Climate: P-values of Granger Non-Causality Tests

	Elements of $\hat{F}^{\tau}_{QFA}$					
τ	1	2	3	4	5	6
0.01	**0.004**					
0.05	**0.010**					
0.10	0.225	0.336				
0.25	0.231	0.371	0.093	0.834		
0.50	0.462	0.457	**0.027**	0.362	0.808	**0.037**
0.75	0.381	0.404	0.229	0.423		
0.90	0.340	0.769				
0.95	0.621					
0.99	0.958					

Note: This Table reports the p-values of Granger non-causality tests, where each of the QFA factors is regressed on their own lags and the lags of $\Delta \log(CO_2)$, and the lag lengths are chosen according to BIC.

Table 10: Macro Forecasting: MSE of Different Methods

	Pre-Crisis		Crisis- Post-Crisis		Full	
	MSE	R. MSE	MSE	R. MSE	MSE	R. MSE
	Real GDP Growth					
AR Benchmark	4.526	1.000	5.456	1.000	4.904	1.000
AR + $\hat{F}_{PCA}$	4.282	0.946	5.373	0.985	4.725	0.964
AR + $\hat{F}_{PCA}$ + $\hat{F}^{90}_{QFA}$	4.155	**0.918**	5.331	0.977	4.634	**0.945**
AR + $\hat{F}_{PCA}$ + $\hat{F}^{99}_{QFA}$	4.354	0.962	5.270	**0.966**	4.728	0.964
AR + $\hat{F}_{PCA}$ + $\hat{F}^{90}_{QFA}$ + $\hat{F}^{99}_{QFA}$	4.191	0.926	5.456	1.000	4.688	0.956
	Inflation					
AR Benchmark	0.266	1.000	0.790	1.000	0.479	1.000
AR + $\hat{F}_{PCA}$	0.246	0.926	0.732	0.926	0.444	0.926
AR + $\hat{F}_{PCA}$ + $\hat{F}^{90}_{QFA}$	0.246	0.926	0.732	0.926	0.444	0.927
AR + $\hat{F}_{PCA}$ + $\hat{F}^{99}_{QFA}$	0.247	0.927	0.739	0.935	0.447	0.932
AR + $\hat{F}_{PCA}$ + $\hat{F}^{90}_{QFA}$ + $\hat{F}^{99}_{QFA}$	0.245	**0.922**	0.726	**0.919**	0.441	**0.920**

Note: This Table reports the MSE of five alternative 1-quarter-ahead forecasting methods for real GDP growth and inflation, and their relative MSE (R. MSE) compared with the AR benchmark model (the lowest R. MSE are shown in bold characters). The out-of-sample forecasting is implemented using rolling windows with 120 observations. The full forecasting evaluation period is from 1990Q1 to 2019Q2, the pre-crisis period is from 1990Q1 to 2007Q2, and the crisis plus post-crisis period is from 2007Q3 to 2019Q2.

Figure 1: Normal Approximations of the Estimated Factors using SQR for $N = T = 50$.
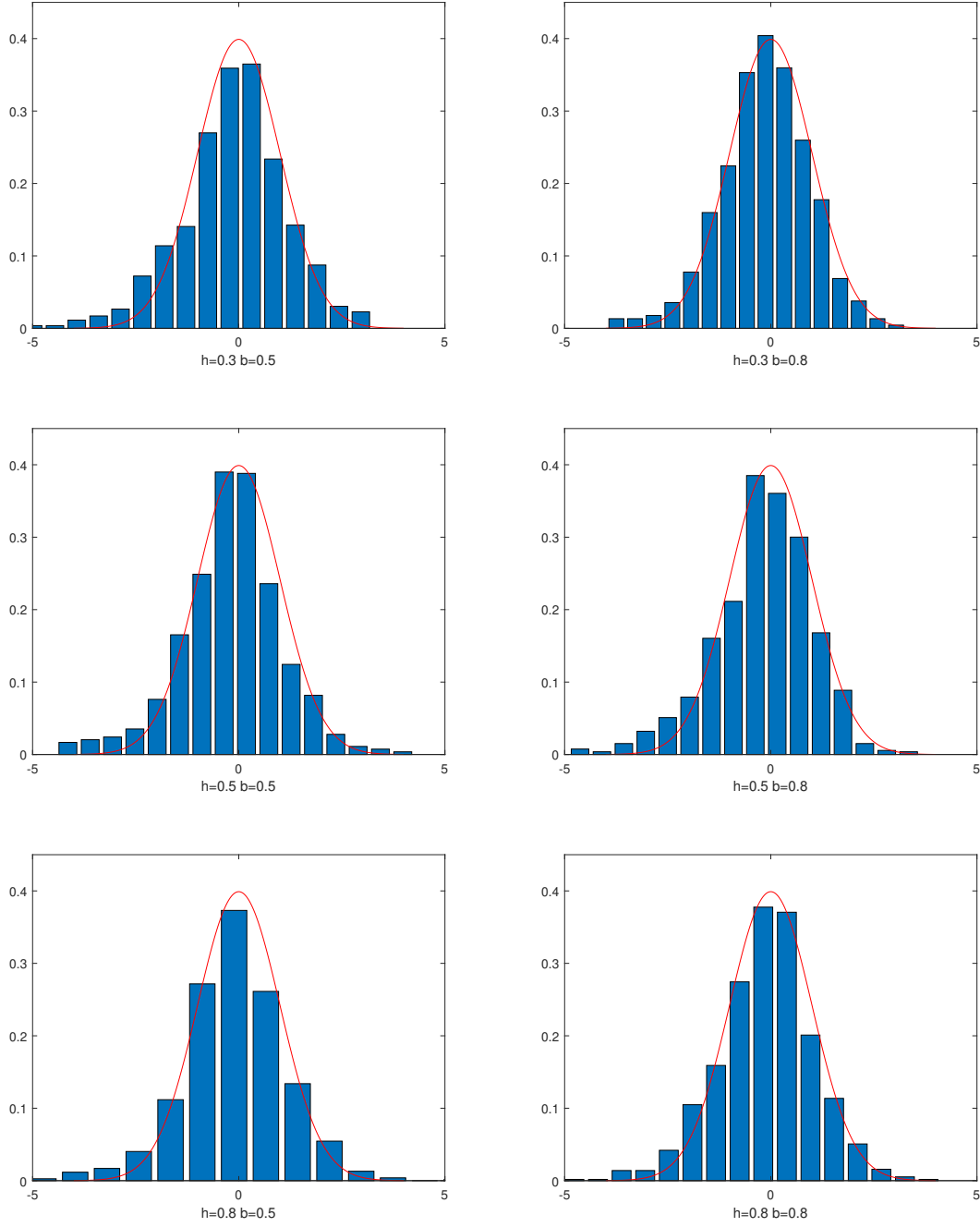
DGP: $X_{it} = \lambda_i f_t + f_t \epsilon_{it}$, where $f_t \sim i.i.d \, \mathcal{U}(1, 2)$ and they are normalized such that $F'F/T = 1$, $\lambda_i \sim i.i.d \, \mathcal{N}(0, 1)$ and $\epsilon_{it} \sim i.i.d \, \mathcal{N}(0, 1)$. The figure plots the histograms of the standardized estimators of the factors using SQR: $\hat{V}_{f_t}^{-1} \sqrt{N}(\tilde{f}_t - f_{0t})$ at $\tau = 0.25, t = T/2$ from 1000 replications, where $\hat{V}_{f_t}$ is estimated using the formula in Remark 4.4, h is the bandwidth parameter in the smoothed check function, and b is the bandwidth parameter used in $\hat{V}_{f_t}$.

Figure 2: Normal Approximations of the Estimated Factors using SQR for $N = T = 200$.
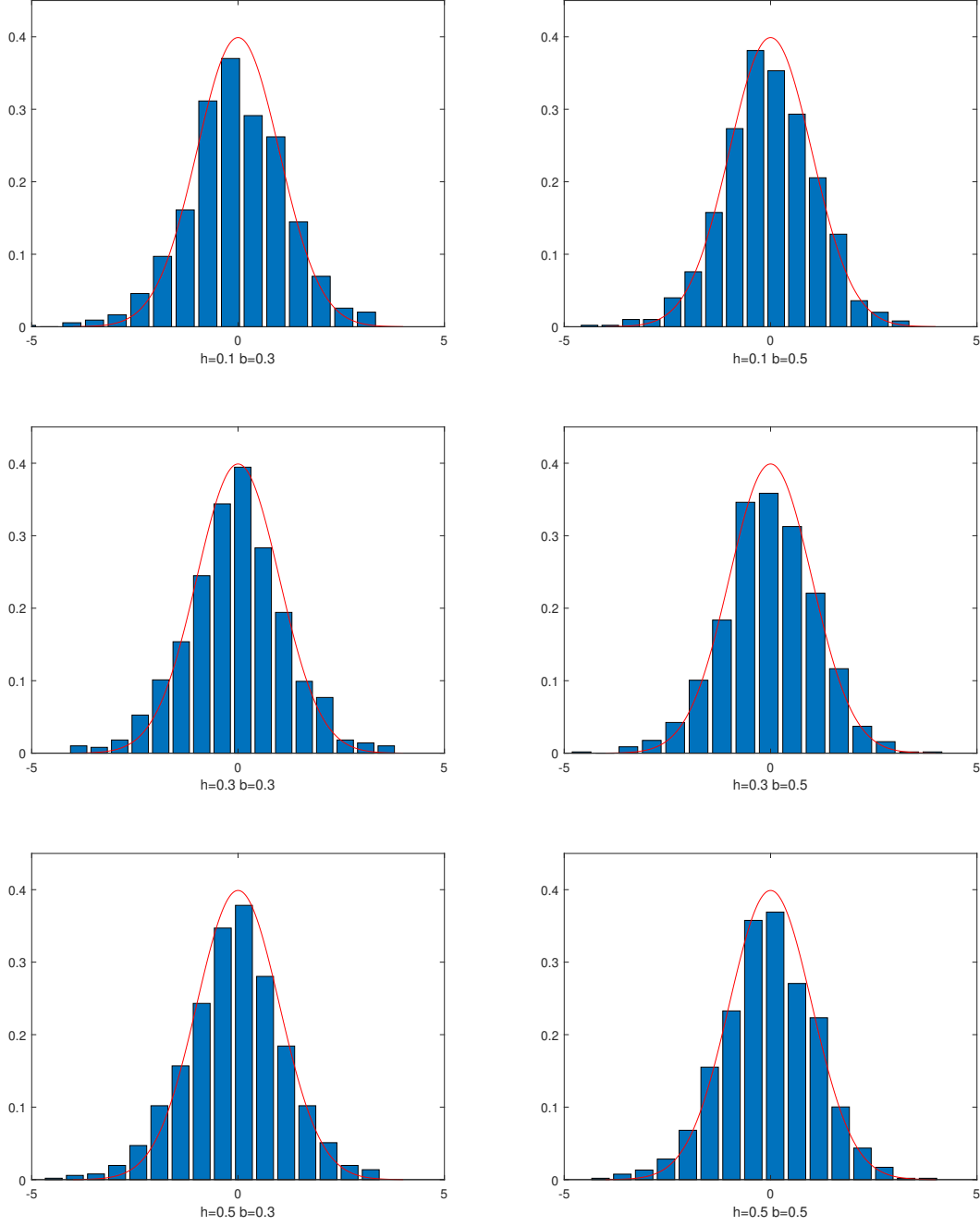
DGP: $X_{it} = \lambda_i f_t + f_t \epsilon_{it}$, where $f_t \sim i.i.d\ \mathcal{U}(1, 2)$ and they are normalized such that $F'F/T = 1$, $\lambda_i \sim i.i.d\ \mathcal{N}(0, 1)$ and $\epsilon_{it} \sim i.i.d\ \mathcal{N}(0, 1)$. The figure plots the histograms of the standardized estimators of the factors using SQR: $\hat{V}_{f_t}^{-1}\sqrt{N}(\tilde{f}_t - f_{0t})$ at $\tau = 0.25, t = T/2$ from 1000 replications, where $\hat{V}_{f_t}$ is estimated using the formula in Remark 4.4, h is the bandwidth parameter in the smoothed check function, and b is the bandwidth parameter used in $\hat{V}_{f_t}$.

Figure 3: Macro Forecasting: Predictive Scores of Density Forecasts for GDP Growth and Inflation
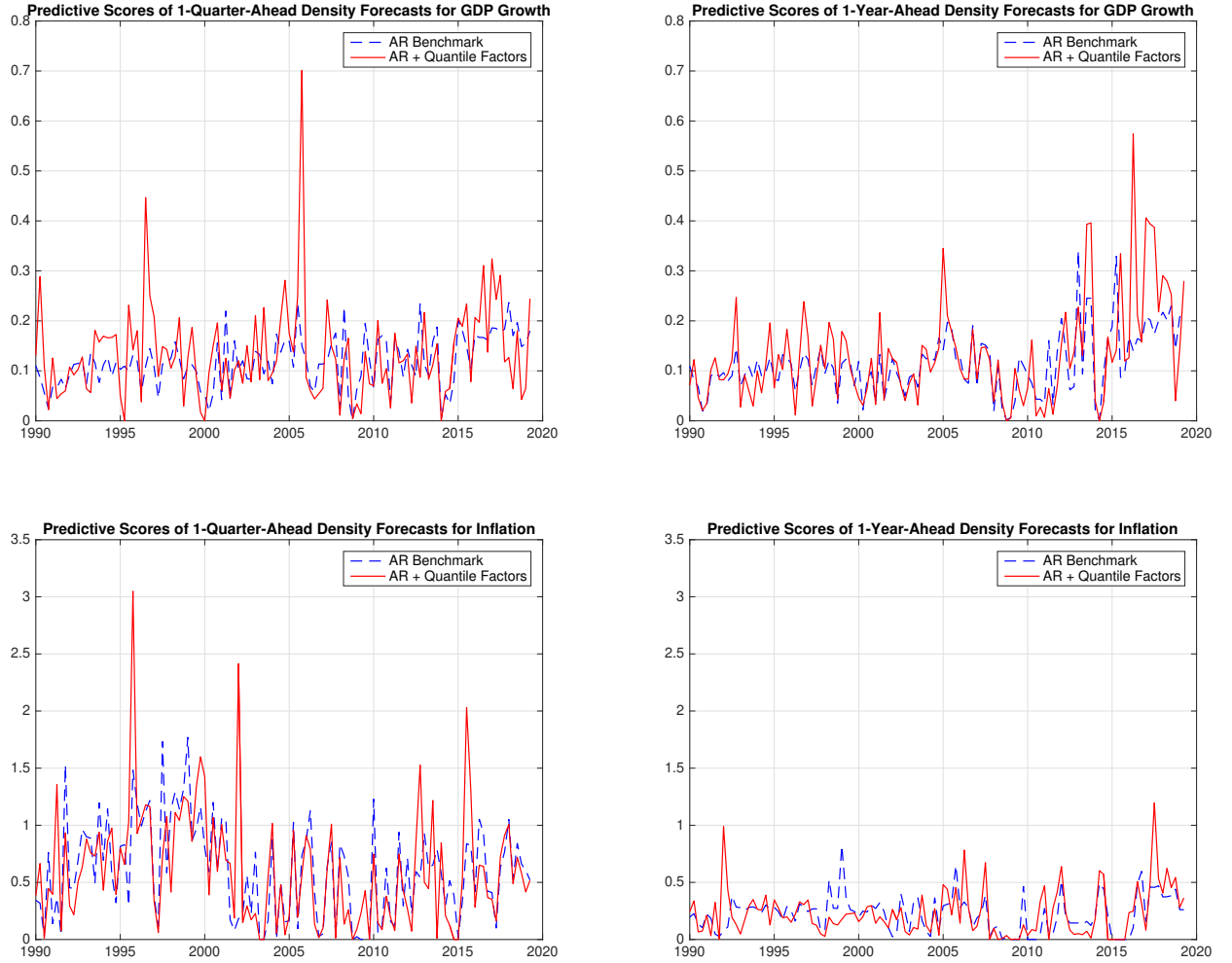
Note: The graphs plot the predictive scores of 1-quarter-ahead and 1-year-ahead density forecasts for real GDP growth and inflation. The evaluation period is from 1990Q1 to 2019Q2, and the out-of-sample forecasting is implemented using rolling windows with 120 observations. The predicted τ-quantiles are constructed using quantile regressions of the target variable on its owns lags and the estimated quantile factors at τ (denoted as $\hat{F}_{QFA}^{\tau}$). The predicted densities are constructed as the density functions of skewed t-distributions by matching the predicted quantiles of the target variable at $\tau \in \{0.05, 0.25, 0.75, 0.95\}$. The predictive scores are the predicted densities evaluated at the realized values of the target variable. Higher scores indicate more accurate forecasts. The dotted blue line is the predictive scores of the benchmark AR model where only the lags of the target variable are used to predict the τ-quantiles, the red line is the predictive scores of the model where $\hat{F}_{QFA}^{\tau}$ is also used to predict the τ-quantiles.

Figure 4: Finance: Quantile Factors, Interquantile Range and Volatility Factor
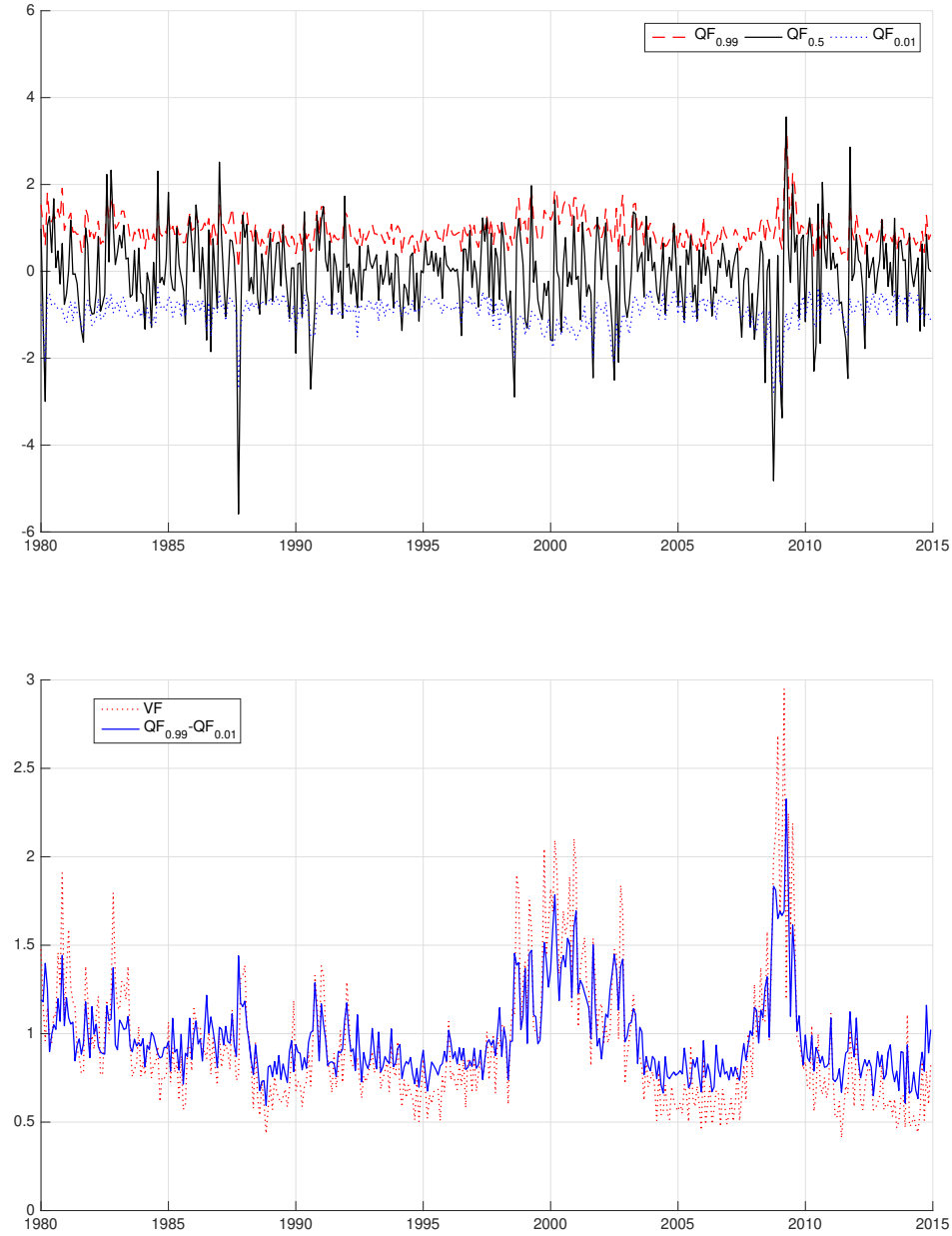
Note: The upper panel presents the QFA factors at $\tau = 0.01, 0.09$ and the first QFA factor at $\tau = 0.5$ from the financial datasets. The lower panel shows the interquantile range (blue line), defined as $\hat{F}_{QFA}^{0.99} - \hat{F}_{QFA}^{0.01}$, alongside the volatility factor (dotted red line) constructed by PCA-SQ. Both series are normalized to have unit length, and their correlation is 0.87.

References

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

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

Amengual, D. and E. Sentana (2018). Is a normal copula the right copula? *Journal of Business & Economic Statistics 0*(0), 1–17.

Andersen, T., N. Fusari, and V. Todorov (2018). The pricing of tail risk and the equity premium: Evidence from international option markets. CREATES research papers, Department of Economics and Business Economics, Aarhus University.

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

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *The Journal of Finance 61*(1), 259–299.

Angrist, J., V. Chernozhukov, and I. Fernández-Val (2006). Quantile regression under misspecification, with an application to the US wage structure. *Econometrica 74*(2), 539–563.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*(1), 685–725.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 367–389.

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

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

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

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2008a). Extremum estimation when the predictors are estimated from large panels. *Annals of Economics and Finance 9*(2), 201–222.

Bai, J. and S. Ng (2008b). *Large dimensional factor analysis.* Now Publishers Inc.

Bai, J. and S. Ng (2013). Principal components estimation and identification of static factors. *Journal of Econometrics 176*(1), 18–29.

Bai, J. and S. Ng (2019). Rank regularized estimation of approximate factor models. *Journal of Econometrics 212*(1), 78–96.

Barigozzi, M. and M. Hallin (2016). Generalized dynamic factor models and volatilities: recovering the market volatility shocks. *The Econometrics Journal 19*(1), 33–63.

Candès, E. J., X. Li, Y. Ma, and J. Wright (2011). Robust principal component analysis? *Journal of the ACM 58*(3), 1–37.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1281–304.

Chen, L., J. J. Dolado, and J. Gonzalo (2017). Quantile factor models. *Working paper, Universidad Carlos III de Madrid*.

Chen, M., I. Fernández-Val, and M. Weidner (2018). Nonlinear factor models for network and panel data. *Working paper, UCL*.

Chen, X., L. P. Hansen, J. Scheinkman, et al. (2009). Nonlinear principal components and long-run implications of multivariate diffusions. *The Annals of Statistics 37*(6B), 4279–4312.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica 86*(4), 1479–1512.

Collins, M., S. Dasgupta, and R. E. Schapire (2001). A generalization of principal component analysis to the exponential family. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS?01, Cambridge, MA, USA, pp. 617?624. MIT Press.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

de Castro, L. and A. F. Galvao (2019). Dynamic quantile models of rational behavior. *Econometrica 87*(6), 1893–1939.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics 82*(4), 540–554.

Gadea, M. D. and J. Gonzalo (2020). Trends in distributional characteristics: Existence of global warming. *Journal of Econometrics 214*(1), 153–174.

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

Golub, G. H. and C. F. Van Loan (2013). *Matrix Computations*, Volume 3. JHU Press.

Gorodnichenko, Y. and S. Ng (2017). Level and volatility factors in macroeconomic data. *Journal of Monetary Economics 91*, 52–68.

Hansen, J., D. Johnson, A. Lacis, S. Lebedeff, P. Lee, D. Rind, and G. Russell (1981). Climate impact of increasing atmospheric carbon dioxide. *Science 213*(4511), 957–966.

Herskovic, B., B. Kelly, H. Lustig, and S. Van Nieuwerburgh (2016). The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *Journal of Financial Economics 119*(2), 249–283.

Horowitz, J. (1998). Bootstrap methods for median regression models. *Econometrica 66*(6), 1327–1352.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology 24*(6), 417.

Hsiang, S. and R. E. Kopp (2018). An economist's guide to climate change science. *Journal of Economic Perspectives 32*(4), 3–32.

Koenker, R. (2005). *Quantile Regression*. Number 38. Cambridge University Press.

Ma, S., O. Linton, and J. Gao (2019). Estimation and inference in semiparametric quantile factor models. Technical report, Faculty of Economics, University of Cambridge.

Maravalle, A. and Ł. Rawdanowicz (2018). Changes in economic and financial synchronisation. *OECD ECO/WKP 65*.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

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

Pelger, M. and R. Xiong (2018). Interpretable proximate factors for large dimensions. *arXiv preprint arXiv:1805.03373*.

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

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics 25*(3), 303–325.

Renault, E., T. Van Der Heijden, and B. Werker (2017). Arbitrage pricing theory for idiosyncratic variance factors. *Working paper, Brown University*.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Stock, J. H. and M. W. Watson (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting 1*, 35–59.

Su, L. and X. Wang (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics 198*(1), 84–101.

van de Geer, S. A. (2002). On Hoeffding's inequality for dependent random variables. In *Empirical process techniques for dependent data*, pp. 161–169. Springer.

van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes*. Springer, New York.

Vidal, R., Y. Ma, and S. Sastry (2016). *Generalized Principal Component Analysis*, Volume 40. Springer.

White, H., T.-H. Kim, and S. Manganelli (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics 187*(1), 169–188.

Online Appendix to "Quantile Factor Models"

Liang Chen[1], Juan J. Dolado[2], and Jesús Gonzalo[3]

[1] *School of Economics, Shanghai University of Finance and Economics, chen.liang@mail.shufe.edu.cn*
[2] *Department of Economics, Universidad Carlos III de Madrid, dolado@eco.uc3m.es*
[3] *Department of Economics, Universidad Carlos III de Madrid, jgonzalo@est-eco.uc3m.es*

March 24, 2020

A Proofs of the Main Results

Definitions and Notations: Throughout the Appendix, $K_1, K_2, \ldots$ denote some positive constants that do not depend on N, T. For any random variable Y, define the Orlicz norm $\|Y\|_\psi$ as:

$$\|Y\|_\psi = \inf \left\{ C > 0 : \mathbb{E}\psi\left(|Y|/C\right) \leq 1 \right\},$$

where ψ is a non-decreasing, convex function with $\psi(0) = 0$. In particular, when $\psi(x) = e^{x^2} - 1$, the norm is written as $\|Y\|_{\psi_2}$. We use $\|\cdot\|$, $\|\cdot\|_S$ and $\|\cdot\|_{\max}$ to denote the Frobenius norm, the spectral norm, and the max norm for matrices, respectively. Notice that, when considering vectors, $\|\cdot\|$ is the Euclidean norm. For a matrix A with real eigenvalues, let $\rho_j(A)$ denote the jth largest eigenvalue. Following van der Vaart and Wellner (1996), the symbol $\lesssim$ means "left side bounded by a positive constant times the right side" (the symbol $\gtrsim$ is defined similarly), and $D(\cdot, g, \mathcal{G})$ and $C(\cdot, g, \mathcal{G})$ denote the packing and covering numbers of space $\mathcal{G}$ endowed with metric g.

Note that all the proofs in this appendix remain valid if we replace λ_{0i}, f_{0t} by $\mathsf{S}\lambda_{0i}, \mathsf{S}f_{0t}$, where S is the sign matrix defined in Theorem 1 and Theorem 4. Thus, without loss of generality, except for the proof of Lemma 2, we assume that $\mathsf{S} = \mathbb{I}_r$ to simplify the notations.

A.1 Proof of Theorem 1

Define

$$\tilde{\mathbb{M}}_{NT}(\theta) = \mathbb{M}_{NT}(\theta) - \bar{\mathbb{M}}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \rho_\tau(X_{it} - \lambda_i' f_t) - \mathbb{E}[\rho_\tau(X_{it} - \lambda_i' f_t)] \right\},$$

$$\mathbb{W}_{NT}(\theta) = \tilde{\mathbb{M}}_{NT}(\theta) - \tilde{\mathbb{M}}_{NT}(\theta_0).$$

Then, the following results hold.

Lemma 1. *Under Assumption 1, we have $d(\hat{\theta}, \theta_0) = o_P(1)$ as $N, T \to \infty$.*

Proof. First, for any $\lambda_i \in \mathcal{A}$ and $f_t \in \mathcal{F}$, expanding $\mathbb{E}[\rho_\tau(X_{it} - \lambda_i' f_t)]$ around $c_{0,it} = \lambda_{0i} f_{0t}$, we have

$$\mathbb{E}[\rho_\tau(X_{it} - \lambda_i' f_t)] = \mathbb{E}[\rho_\tau(X_{it} - \lambda_{0i} f_{0t})] + 0.5 \cdot \mathsf{f}_{it}(c_{it}^*) \cdot (\lambda_i' f_t - \lambda_{0i} f_{0t})^2,$$

where c_{it}^* is between $\lambda_i' f_t$ and $c_{0,it}$. It then follows from Assumption 1(ii) that for all $\lambda_i \in \mathcal{A}$ and $f_t \in \mathcal{F}$,

$$\mathbb{E}[\rho_\tau(X_{it} - \lambda_i' f_t)] - \mathbb{E}[\rho_\tau(X_{it} - \lambda_{0i} f_{0t})] \gtrsim (\lambda_i' f_t - \lambda_{0i}' f_{0t})^2$$

since $|\lambda_i' f_t|$ and $|\lambda_{0i}' f_{0t}|$ are both bounded. Therefore, for any $\theta \in \Theta^r$, we have:

$$\bar{\mathbb{M}}_{NT}(\theta) - \bar{\mathbb{M}}_{NT}(\theta_0) \gtrsim d^2(\theta, \theta_0). \tag{A.1}$$

Second, by the definition of $\hat{\theta}$, we have $\mathbb{M}_{NT}(\hat{\theta}) - \mathbb{M}_{NT}(\theta_0) \leq 0$, or equivalently $\mathbb{W}_{NT}(\hat{\theta}) + \bar{\mathbb{M}}_{NT}(\hat{\theta}) - \bar{\mathbb{M}}_{NT}(\theta_0) \leq 0$. It then follows from (A.1) that

$$0 \leq d^2(\hat{\theta}, \theta_0) \lesssim \bar{\mathbb{M}}_{NT}(\hat{\theta}) - \bar{\mathbb{M}}_{NT}(\theta_0) \leq \sup_{\theta \in \Theta^r} |\mathbb{W}_{NT}(\theta)|.$$

Thus, it remains to be shown that

$$\sup_{\theta \in \Theta^r} |\mathbb{W}_{NT}(\theta)| = o_P(1). \tag{A.2}$$

Choose K_1 large enough such that $\|\lambda_{0i}\|, \|f_{0t}\|, \|\lambda_i\|, \|f_t\| \leq K_1$ for all i, t for any $\theta \in \Theta^r$. Let $B_r(K_1)$ be a Euclidean ball in $\mathbb{R}^r$ with radius K_1. For any $\epsilon > 0$, let $\lambda_{(1)}, \ldots, \lambda_{(J)}$ be a maximal set of points in $B_r(K_1)$ such that $\|\lambda_{(j)} - \lambda_{(h)}\| > \epsilon/K_1$ for any $j \neq h$ where "maximal" signifies that no point can be added without violating the validity of the inequality. Similarly, let $f_{(1)}, \ldots, f_{(J)}$ be a maximal set of points in $B_r(K_1)$ such that $\|f_{(j)} - f_{(h)}\| > \epsilon/K_1$ for any $j \neq h$. It is well known that J, the packing number of $B_r(K_1)$, is equal to $K_2(K_1/\epsilon)^r$.

For any $\theta \in \Theta^r$, define $\theta^* = (\lambda_1^{*'}, \ldots, \lambda_N^{*'}, f_1^{*'}, \ldots, f_T^{*'})'$, where $\lambda_i^* = \{\lambda_{(j)} : j \leq J, \|\lambda_{(j)} - \lambda_i\| \leq \epsilon/K_1\}$ and $f_t^* = \{f_{(j)} : j \leq J, \|f_{(j)} - f_t\| \leq \epsilon/K_1\}$. Thus, we can write

$$\mathbb{W}_{NT}(\theta) = \mathbb{W}_{NT}(\theta^*) + \mathbb{W}_{NT}(\theta) - \mathbb{W}_{NT}(\theta^*).$$

Note that $|\rho_\tau(X_{it} - \lambda_i' f_t) - \rho_\tau(X_{it} - \lambda_i^{*'} f_t^*)| \leq 2|\lambda_i' f_t - \lambda_i^{*'} f_t^*| \leq 2\|\lambda_i\|\|f_t - f_t^*\| + 2\|f_t^*\|\|\lambda_i - \lambda_i^*\| \leq 4\epsilon$. Thus,

$$\sup_{\theta \in \Theta^r} |\mathbb{W}_{NT}(\theta) - \mathbb{W}_{NT}(\theta^*)| \lesssim \epsilon. \tag{A.3}$$

Also, note that $|\rho_\tau(X_{it} - \lambda_{0i}' f_{0t}) - \rho_\tau(X_{it} - \lambda_i^{*'} f_t^*)| \leq 2|\lambda_{0i}' f_{0t} - \lambda_i^{*'} f_t^*|$. Then, by Hoeffding's inequality we have

$$P[|\sqrt{NT}\mathbb{W}_{NT}(\theta^*)| > c] \leq 2e^{-\frac{c^2}{2 \cdot d^2(\theta^*, \theta_0)}},$$

and by Lemma 2.2.1 of van der Vaart and Wellner (1996) it follows that $\|\mathbb{W}_{NT}(\theta^*)\|_{\psi_2} \lesssim d(\theta^*, \theta_0)/\sqrt{NT}$. Since θ^* can take at most $J^{N+T} \lesssim (K_1/\epsilon)^{r(N+T)}$ different values, and $d(\theta^*, \theta_0) \leq 2K_1$, it follows from

Lemma 2.2.2 of van der Vaart and Wellner (1996) that

$$
\mathbb{E}\left[\sup_{\theta\in\Theta^r}|\mathbb{W}_{NT}(\theta^*)|\right] \leq \left\|\sup_{\theta\in\Theta^r}|\mathbb{W}_{NT}(\theta^*)|\right\|_{\psi_2} \lesssim \sqrt{\log(K_1/\epsilon)}\sqrt{r(N+T)}/\sqrt{NT} \lesssim \sqrt{\log(K_1/\epsilon)}/L_{NT}.
\tag{A.4}
$$

Finally, by Markov's inequality and (A.3), for any $\delta > 0$,

$$
\begin{aligned}
P\left[\sup_{\theta\in\Theta^r}|\mathbb{W}_{NT}(\theta)| > \delta\right] &\leq P\left[\sup_{\theta\in\Theta^r}|\mathbb{W}_{NT}(\theta^*)| > \delta/2\right] + P\left[\sup_{\theta\in\Theta^r}|\mathbb{W}_{NT}(\theta) - \mathbb{W}_{NT}(\theta^*)| > \delta/2\right] \\
&\leq 2\cdot\mathbb{E}\left[\sup_{\theta\in\Theta^r}|\mathbb{W}_{NT}(\theta^*)|\right]/\delta + P\left[K_2\epsilon > \delta/2\right].
\end{aligned}
$$

Thus, (A.2) follows from (A.4) since ϵ is arbitrary. This concludes the proof. $\qquad\square$

Next, define $\Theta^r(\delta) = \{\theta \in \Theta^r : d(\theta,\theta_0) \leq \delta\}$.

Lemma 2. *Unnder Assumption 1 and for sufficiently small $\delta > 0$, there exists a diagonal matrix $\mathsf{S} \in \mathbb{R}^{r\times r}$ whose diagonal elements are either 1 or -1, such that*

$$
\|\Lambda - \Lambda_0\mathsf{S}\|/\sqrt{N} + \|F - F_0\mathsf{S}\|/\sqrt{T} \leq K_3\delta
$$

for any $\theta \in \Theta^r(\delta)$.

Proof. First, let $\mathsf{U} \in \mathbb{R}^{r\times r}$ be a diagonal matrix whose diagonal elements are either 1 or -1. Since $F'F/T = F_0'F_0/T = \mathbb{I}_r$ and $\|\Lambda_0\|/\sqrt{N} \leq K_4$ by Assumption 1(i), we have

$$
\begin{aligned}
\|\Lambda - \Lambda_0\mathsf{U}\|/\sqrt{N} &= \|(\Lambda - \Lambda_0\mathsf{U})F'\|/\sqrt{NT} = \|\Lambda F' - \Lambda_0 F_0' + \Lambda_0 F_0' - \Lambda_0\mathsf{U}F'\|/\sqrt{NT} \\
&\leq \|\Lambda F' - \Lambda_0 F_0'\|/\sqrt{NT} + \|\Lambda_0\|/\sqrt{N}\cdot\|F - F_0\mathsf{U}\|/\sqrt{T} \\
&\leq d(\theta,\theta_0) + K_4\|F - F_0\mathsf{U}\|/\sqrt{T}.
\end{aligned}
$$

Thus, for $\theta \in \Theta^r(\delta)$,

$$
\|\Lambda - \Lambda_0\mathsf{U}\|/\sqrt{N} + \|F - F_0\mathsf{U}\|/\sqrt{T} \leq \delta + (1 + K_4)\|F - F_0\mathsf{U}\|/\sqrt{T}.
\tag{A.5}
$$

Second,

$$
\begin{aligned}
\|F - F_0\mathsf{U}\|/\sqrt{T} &= \|F_0\mathsf{U} - F(F'F_0\mathsf{U}/T) + F(F'F_0\mathsf{U}/T) - F\|/\sqrt{T} \\
&\leq \|F_0\mathsf{U} - F(F'F_0\mathsf{U}/T)\|/\sqrt{T} + \|F(F'F_0\mathsf{U}/T) - F\|/\sqrt{T} = \|M_F F_0\|/\sqrt{T} + \|F'F_0/T - \mathsf{U}\|,
\end{aligned}
\tag{A.6}
$$

where $P_A = A(A'A)^{-1}A'$ and $M_A = \mathbb{I} - P_A$.

Third, we have

$$
\begin{aligned}
\frac{1}{\sqrt{NT}}\|(\Lambda F' - \Lambda_0 F_0')M_F\| &\leq \sqrt{\text{rank}[(\Lambda F' - \Lambda_0 F_0')M_F]}\cdot\|M_F\|_S\cdot\|\Lambda F' - \Lambda_0 F_0'\|_S/\sqrt{NT} \\
&\lesssim \|\Lambda F' - \Lambda_0 F_0'\|/\sqrt{NT} = d(\theta,\theta_0),
\end{aligned}
\tag{A.7}
$$

3

and since

$$\|(\Lambda F' - \Lambda_0 F_0')M_F\|/\sqrt{NT} = \|\Lambda_0 F_0' M_F\|/\sqrt{NT} = \sqrt{\text{Tr}\left[(\Lambda_0'\Lambda_0/N)\cdot(F_0'M_F F_0/T)\right]}$$
$$\geq \sqrt{\sigma_{Nr}}\sqrt{\text{Tr}\left(F_0'M_F F_0/T\right)} = \sqrt{\sigma_{Nr}}\|M_F F_0\|/\sqrt{T}, \quad \text{(A.8)}$$

it follows from (A.7) and (A.8) that

$$\|M_F F_0\|/\sqrt{T} \lesssim \sqrt{\frac{1}{\sigma_{Nr}}}d(\theta,\theta_0). \tag{A.9}$$

Similarly, it can be shown that

$$\|M_{F_0}F\|/\sqrt{T} \lesssim \sqrt{\frac{1}{\rho_{\min}(\Lambda'\Lambda/N)}}d(\theta,\theta_0), \tag{A.10}$$

where $\rho_{\min}$ denotes the minimum eigenvalue.

Fourth, we have

$$\frac{1}{\sqrt{NT}}\|(\Lambda F' - \Lambda_0 F_0')P_F\| \leq \frac{1}{\sqrt{NT}}\|\Lambda F' - \Lambda_0 F_0'\|\cdot\|P_F\| = \sqrt{r}d(\theta,\theta_0),$$

so

$$\frac{1}{\sqrt{NT}}\|(\Lambda F' - \Lambda_0 F_0')P_F\| = \frac{1}{\sqrt{NT}}\|\Lambda F' - \Lambda_0(F_0'F/T)F'\| = \frac{1}{\sqrt{N}}\|\Lambda - \Lambda_0(F_0'F/T)\| \leq \sqrt{r}d(\theta,\theta_0). \tag{A.11}$$

Likewise, it can be shown that

$$\frac{1}{\sqrt{N}}\|\Lambda_0 - \Lambda(F'F_0/T)\| \leq \sqrt{r}d(\theta,\theta_0). \tag{A.12}$$

Fifth, define $R_T = F'F_0/T$. Note that $FR_T = FF'F_0/T = P_F F_0$, thus

$$\mathbb{I}_r = F_0'F_0/T = R_T'(F'F/T)R_T + F_0'F_0/T - R_T'(F'F/T)R_T$$
$$= R_T'R_T + F_0'F_0/T - F_0'FR_T/T + F_0'FR_T/T - R_T'(F'F/T)R_T$$
$$= R_T'R_T + F_0'(F_0 - FR_T)/T = R_T'R_T + F_0'M_F F_0/T \tag{A.13}$$

because

$$F_0'FR_T/T - R_T'(F'F/T)R_T = (F_0 - FR_T)'FR_T/T = F_0'M_F FR_T/T = 0.$$

Moreover,

$$\Lambda_0'\Lambda_0/N = R_T'(\Lambda'\Lambda/N)R_T + (\Lambda_0'\Lambda_0/N - R_T'(\Lambda'\Lambda/N)R_T)$$
$$= R_T'(\Lambda'\Lambda/N)R_T + \Lambda_0'(\Lambda_0 - \Lambda R_T)/N + (\Lambda_0 - \Lambda R_T)'\Lambda R_T/N. \tag{A.14}$$

4

Similarly to the proof of (A.13), we have

$$\mathbb{I}_r = R_T R_T' + F'(F - F_0 R_T')/T = R_T R_T' + F' M_{F_0} F/T. \tag{A.15}$$

From (A.14) we have

$$\Lambda_0' \Lambda_0/N = R_T'(\Lambda'\Lambda/N)(R_T')^{-1} R_T' R_T + \Lambda_0'(\Lambda_0 - \Lambda R_T)/N + (\Lambda_0 - \Lambda R_T)'\Lambda R_T/N$$
$$= R_T'(\Lambda'\Lambda/N)(R_T')^{-1} + R_T'(\Lambda'\Lambda/N)(R_T')^{-1}(R_T' R_T - \mathbb{I}_r) + \Lambda_0'(\Lambda_0 - \Lambda R_T)/N + (\Lambda_0 - \Lambda R_T)'\Lambda R_T/N,$$

and it follows from the above equation and (A.13) that

$$\left(\Lambda_0' \Lambda_0/N + D_{NT}\right) R_T' = R_T'(\Lambda'\Lambda/N), \tag{A.16}$$

where

$$D_{NT} = R_T'(\Lambda'\Lambda/N)(R_T')^{-1} F_0' M_F F_0/T - \Lambda_0'(\Lambda_0 - \Lambda R_T)/N - (\Lambda_0 - \Lambda R_T)'\Lambda R_T/N.$$

From (A.9) (A.11) and (A.12) we have that $\|D_{NT}\| \lesssim d(\theta, \theta_0)$. Hence, by the Bauer-Fike Theorem (see Theorem 7.2.2 of Golub and Van Loan 2013), it holds that

$$|\rho_{\min}[\Lambda'\Lambda/N] - \rho_{\min}[\Lambda_0'\Lambda_0/N]| \leq \|D_{NT}\|_S \leq \|D_{NT}\| \lesssim d(\theta, \theta_0). \tag{A.17}$$

Moreover, by Assumption 1(i) and the perturbation theory for eigenvectors (see Section 6.12 of Franklin 2012), there exists a diagonal matrix $\mathsf{S} \in \mathbb{R}^{r \times r}$ whose diagonal elements are either 1 or -1 such that

$$\|R_T' V_T \mathsf{S} - \mathbb{I}_r\| = \|R_T' V_T - \mathsf{S}\| \lesssim d(\theta, \theta_0), \tag{A.18}$$

where $V_T = \text{diag}\left((R_{T,1} R_{T,1}')^{-1/2}, \ldots, (R_{T,r} R_{T,r}')^{-1/2}\right)$, and $R_{T,j}'$ is the jth column of R_T'.

In addition, (A.10) and (A.17) imply that $\rho_{\min}[\Lambda'\Lambda/N]$ is bounded below by a positive constant, and that

$$\|M_{F_0} F\|/\sqrt{T} \lesssim d(\theta, \theta_0). \tag{A.19}$$

Note that the triangular inequality implies that

$$\|R_T' - \mathsf{S}\| \leq \|R_T' V_T - \mathsf{S}\| + \|R_T' V_T - R_T\| \leq \|R_T' V_T - \mathsf{S}\| + \|R_T\| \cdot \|V_T - \mathbb{I}_r\|. \tag{A.20}$$

From (A.15) and (A.19)

$$\|V_T - \mathbb{I}_r\| \leq \|R_T R_T' - \mathbb{I}_r\| = \|M_{F_0} F\|^2/T \lesssim d^2(\theta, \theta_0). \tag{A.21}$$

For small enough $d(\theta, \theta_0)$, it then follows from (A.18) (A.20) and (A.21) that

$$\|F' F_0/T - \mathsf{S}\| = \|R_T - \mathsf{S}\| \lesssim d(\theta, \theta_0). \tag{A.22}$$

Finally, setting $\mathsf{U} = \mathsf{S}$, it follows from (A.6) (A.9) and (A.22) that for sufficiently small $d(\theta, \theta_0)$

$$\|F - F_0 \mathsf{S}\|/\sqrt{T} \lesssim d(\theta, \theta_0). \tag{A.23}$$

5

Then the desired result follows from (A.5) and (A.23). □

Lemma 3. *Under Assumption 1, for sufficiently small δ, we have*

$$\mathbb{E}\left[\sup_{\theta \in \Theta^r(\delta)} |\mathbb{W}_{NT}(\theta)|\right] \lesssim \frac{\delta}{L_{NT}}.$$

Proof. In the proof of Lemma 1 we have shown that

$$\left\|\sqrt{NT}|\mathbb{W}_{NT}(\theta_a) - \mathbb{W}_{NT}(\theta_b)|\right\|_{\psi_2} \lesssim d(\theta_a, \theta_b), \tag{A.24}$$

and therefore

$$\left\|\sqrt{NT}\mathbb{W}_{NT}(\theta)\right\|_{\psi_2} \lesssim d(\theta, \theta_0).$$

Construct nested sets $\Theta_1^r(\delta) \subset \Theta_2^r(\delta) \dots \subset \Theta^r(\delta)$ such that each $\Theta_j^r(\delta)$ is a maximal set of points with $d(\theta_a, \theta_b) > \delta/2^j$ for every $\theta_a \neq \theta_b$ in $\Theta_j^r(\delta)$. In particular, let $\Theta_0^r(\delta) = \{\theta_0\}$.

For each point θ in $\Theta_{j+1}^r(\delta)$, let θ_* be a point in $\Theta_j^r(\delta)$ such that $d(\theta, \theta_*) \leq \delta/2^j$. It then follows by the triangle inequality that

$$\max_{\Theta_{j+1}^r(\delta)} |\mathbb{W}_{NT}(\theta)| \leq \max_{\Theta_{j+1}^r(\delta)} |\mathbb{W}_{NT}(\theta_*)| + \max_{\Theta_{j+1}^r(\delta)} |\mathbb{W}_{NT}(\theta) - \mathbb{W}_{NT}(\theta_*)|. \tag{A.25}$$

Note that

$$\max_{\Theta_{j+1}^r(\delta)} |\mathbb{W}_{NT}(\theta_*)| \leq \max_{\Theta_j^r(\delta)} |\mathbb{W}_{NT}(\theta)|,$$

and taking $\|\|_{\psi_2}$ norm on both sides of (A.25) gives

$$\left\|\max_{\Theta_{j+1}^r(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta)|\right\|_{\psi_2} \leq \left\|\max_{\Theta_j^r(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta)|\right\|_{\psi_2} + \left\|\max_{\Theta_{j+1}^r(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta) - \mathbb{W}_{NT}(\theta_*)|\right\|_{\psi_2}.$$

Let $m_j = \#\Theta_j^r$, the number of points in Θ_j^r. The second term on the RHS of the last inequality, according to Lemma 2.2.2 of van der Vaart and Wellner (1996), is bounded by

$$K_5\sqrt{\log(1 + m_{j+1})} \cdot \max_{\Theta_{j+1}^r(\delta)} \left\|\sqrt{NT} |\mathbb{W}_{NT}(\theta) - \mathbb{W}_{NT}(\theta_*)|\right\|_{\psi_2},$$

which according to (A.24) is bounded by $\delta/2^j \cdot \sqrt{\log(1 + m_{j+1})}$ multiplied by a positive constant. Thus, we have

$$\left\|\max_{\Theta_{j+1}^r(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta)|\right\|_{\psi_2} \leq \left\|\max_{\Theta_j^r(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta)|\right\|_{\psi_2} + K_6\delta/2^j \cdot \sqrt{\log(1 + m_{j+1})},$$

which implies that for $J > 1$,

$$\left\|\max_{\Theta_{J+1}^r(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta)|\right\|_{\psi_2} \leq K_6 \sum_{j=1}^{J} \delta/2^j \cdot \sqrt{\log(1 + m_{j+1})}.$$

6

Letting $J \to \infty$, the above inequality yields [1]

$$\left\| \sup_{\Theta^r(\delta)} \sqrt{NT} \, |\mathbb{W}_{NT}(\theta)| \right\|_{\psi_2} \lesssim \sum_{j=1}^{\infty} \delta/2^j \cdot \sqrt{\log(m_{j+1})}. \tag{A.26}$$

Note that $m_{j+1} \leq D(\delta/2^{j+1}, d, \Theta^r(\delta))$, which is the packing number of $\Theta^r(\delta)$. Further note that

$$\delta/2^j \cdot \sqrt{\log\left(D(\delta/2^{j+1}, d, \Theta^r(\delta))\right)} \lesssim \int_{\delta/2^{j+2}}^{\delta/2^{j+1}} \sqrt{\log\left(D(\epsilon, d, \Theta^r(\delta))\right)} d\epsilon.$$

It follows from the above inequality that the RHS of (A.26) is bounded by $\int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^r(\delta))}$, and we finally have

$$\mathbb{E}\left[\sup_{\Theta^r(\delta)} \sqrt{NT} \, |\mathbb{W}_{NT}(\theta)| \right] \lesssim \left\| \sup_{\Theta^r(\delta)} \sqrt{NT} \, |\mathbb{W}_{NT}(\theta)| \right\|_{\psi_2} \lesssim \int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^r(\delta))} d\epsilon.$$

Thus, it remains to be shown that

$$\int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^r(\delta))} d\epsilon = O(\sqrt{N+T}\delta). \tag{A.27}$$

To prove (A.27), first note that for any $\theta_a, \theta_b \in \Theta^r$,

$$
\begin{aligned}
d(\theta_a, \theta_b) = \frac{1}{\sqrt{NT}}\|\Lambda_a F_a' - \Lambda_b F_b'\| &= \frac{1}{\sqrt{NT}}\|\Lambda_a F_a' - \Lambda_b F_a' + \Lambda_b F_a' - \Lambda_b F_b'\| \\
&\leq \frac{1}{\sqrt{N}}\|\Lambda_a - \Lambda_b\| + \frac{\|\Lambda_b\|}{\sqrt{N}} \cdot \frac{\|F_a - F_b\|}{\sqrt{T}} \leq K_7 \left(\frac{\|\Lambda_a - \Lambda_b\|}{\sqrt{N}} + \frac{\|F_a - F_b\|}{\sqrt{T}} \right),
\end{aligned}
$$

where $K_7 \geq 1$. Now define

$$d^*(\theta_a, \theta_b) = 2K_7 \sqrt{\frac{\|\Lambda_a - \Lambda_b\|^2}{N} + \frac{\|F_a - F_b\|^2}{T}}.$$

Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \leq 2\sqrt{a+b}$, we have $d(\theta_a, \theta_b) \leq d^*(\theta_a, \theta_b)$, and thus $D(\epsilon, d, \Theta^r(\delta)) \leq D(\epsilon, d^*, \Theta^r(\delta))$. Moreover, by Lemma 2, $\theta \in \Theta^r(\delta)$ implies $d^*(\theta, \theta_0) \leq K_8 d(\theta, \theta_0) \leq K_8 \delta$. Therefore, we have $\Theta^r(\delta) \subset \Theta^{r*}(\delta)$, where $\Theta^{r*}(\delta) = \{\theta \in \Theta^r : d^*(\theta, \theta_0) \leq K_8 \delta\}$. It then follows that

$$D(\epsilon, d, \Theta^r(\delta)) \leq D(\epsilon, d^*, \Theta^r(\delta)) \leq D(\epsilon, d^*, \Theta^{r*}(\delta)) \leq C(\epsilon/2, d^*, \Theta^{r*}(\delta)). \tag{A.28}$$

Next, we calculate an upper bound for $C(\epsilon/2, d^*, \Theta^{r*}(\delta))$. Let $\eta = \epsilon/2$, and $\theta_1^*, \ldots, \theta_J^*$ be a maximal set in $\Theta^{r*}(\delta)$ such that $d^*(\theta_j^*, \theta_l^*) > \eta$ for any $j \neq l$. Define $B(\theta, c) = \{\gamma \in \Theta^r : d^*(\gamma, \theta) \leq c\}$. Then, the balls $B(\theta_1^*, \eta), \ldots, B(\theta_J^*, \eta)$ cover $\Theta^{r*}(\delta)$, and thus $C(\epsilon/2, d^*, \Theta^{r*}(\delta)) \leq J$. Moreover, the balls $B(\theta_1^*, \eta/4), \ldots, B(\theta_J^*, \eta/4)$ are disjoint and

$$\cup_{j=1}^J B(\theta_j^*, \eta/4) \subset \Theta^{r*}(\delta + \eta/4).$$

---

[1] Note that the process $\mathbb{W}_{NT}(\theta)$ is separable.

Note that the volume of a ball define by the metric d^* with radius c is the volume of an ellipsoid, which is equal to $h_M \cdot c^M$, where h_M is a constant that depends on N, T and r, but not on c. Therefore, we have

$$J \cdot h_M \cdot (\eta/4)^M \leq h_M \cdot (K_8(\delta + \eta/4))^M,$$

which implies

$$J \leq \left(\frac{K_8(4\delta + \eta)}{\eta}\right)^M = \left(\frac{K_8(8\delta + \epsilon)}{\epsilon}\right)^M \leq \left(\frac{K_9\delta}{\epsilon}\right)^M \tag{A.29}$$

for $\epsilon \leq \delta$. It then follows from (A.28) and (A.29) that

$$\int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^r(\delta))}d\epsilon \quad \leq \quad \int_0^\delta \sqrt{\log C(\epsilon/2, d^*, \Theta^{r*}(\delta))}d\epsilon \quad \leq \quad \sqrt{(N+T)r}\int_0^\delta \sqrt{\log(K_9\delta/\epsilon)}d\epsilon.$$

It is easy to show that $\int_0^\delta \sqrt{\log(K_9\delta/\epsilon)}d\epsilon = O(\delta)$ and thus (A.27) follows. This concludes the proof of Lemma 3. $\square$

Proof of Theorem 1:

Proof. The parameter space Θ^r can be partitioned into shells $S_j = \{\theta \in \Theta^r : 2^{j-1} < L_{NT} \cdot d(\theta, \theta_0) \leq 2^j\}$. If $L_{NT} \cdot d(\hat{\theta}, \theta_0)$ is larger than 2^V for a given integer V, then $\hat{\theta}$ is in one of the shells S_j with $j > V$. In such a case the infimum of the mapping $\theta \mapsto \mathbb{M}_{NT}(\theta) - \mathbb{M}_{NT}(\theta_0)$ over this shell is nonpositive by the definition of $\hat{\theta}$. From this we conclude that, for every $\eta > 0$,

$$P\left[L_{NT} \cdot d(\hat{\theta}, \theta_0) > 2^V\right]$$

$$\leq \sum_{j>V, 2^{j-1}\leq \eta L_{NT}} P\left[\inf_{\theta \in S_j}(\mathbb{M}_{NT}(\theta) - \mathbb{M}_{NT}(\theta_0)) \leq 0\right] + P[d(\hat{\theta}, \theta_0) > \eta].$$

For arbitrarily small $\eta > 0$, the second probability on the RHS of the above equation converges to 0 as $N, T \to \infty$ by Lemma 1.

Next, note that by (A.1), for each θ in S_j we have

$$-[\bar{\mathbb{M}}_{NT}(\theta) - \bar{\mathbb{M}}_{NT}(\theta_0)] \lesssim -d_{NT}^2(\theta, \theta_0) \leq -\frac{2^{2j-2}}{L_{NT}^2}.$$

Thus, $\inf_{\theta \in S_j}(\mathbb{M}_{NT}(\theta) - \mathbb{M}_{NT}(\theta_0)) \leq 0$ implies that

$$\inf_{\theta \in S_j} \mathbb{W}_{NT}(\theta) \leq -\frac{2^{2j-2}}{L_{NT}^2},$$

8

and therefore

$$\sum_{j>V,2^{j-1}\leq\eta L_{NT}} P\left[\inf_{\theta\in S_j}\left(\mathbb{M}_{NT}(\theta)-\mathbb{M}_{NT}(\theta_0)\right)\leq 0\right]$$

$$\leq \sum_{j>V,2^{j-1}\leq\eta L_{NT}} P\left[\sup_{\theta\in S_j}|\mathbb{W}_{NT}(\theta)|\geq\frac{2^{2j-2}}{L_{NT}^2}\right].$$

By Lemma 3 and Markov's inequality, we have

$$P\left[\sup_{\theta\in S_j}|\mathbb{W}_{NT}(\theta)|\geq\frac{2^{2j-2}}{L_{NT}^2}\right]\lesssim\frac{L_{NT}^2}{2^{2j}}\cdot\mathbb{E}\left[\sup_{\theta\in S_j}|\mathbb{W}_{NT}(\theta)|\right]\lesssim\frac{L_{NT}^2}{2^{2j}}\cdot\frac{2^j}{L_{NT}^2}=2^{-j},$$

which implies that

$$\sum_{j>V,2^{j-1}\leq\eta L_{NT}} P\left[\inf_{\theta\in S_j}\left(\mathbb{M}_{NT}(\theta)-\mathbb{M}_{NT}(\theta_0)\right)\leq 0\right]\lesssim\sum_{j>V}2^{-j}.$$

The RHS of the previous expression convergences to 0 as $V\to\infty$, implying that $L_{NT}\cdot d(\hat{\theta},\theta_0)=O_P(1)$, or $d(\hat{\theta},\theta_0)=O_P(1/L_{NT})$. The desired result then follows from Lemma 2. $\qquad\square$

A.2 Proof of Theorem 2

With a little abuse of notation, for $\theta_a\in\Theta^k$ and $\theta_b\in\Theta^r$, let

$$d(\theta_a,\theta_b)=\|\Lambda_a F_a'-\Lambda_b F_b'\|/\sqrt{NT}.$$

In this case, d is not a metric because θ_a and θ_b belong to different spaces. However, the following proofs will not be affected if we replace $d(\theta_a,\theta_b)$ by a different notation (say $\bar{d}(\theta_a,\theta_b)$).

For sufficiently small δ, define $\Theta^k(\delta)=\{\theta^k\in\Theta^k:d(\theta^k,\theta_0)\leq\delta\}$. Let $F^{k,r}$ denote the first r columns of F^k, and let $F^{k,-r}$ denote the remaining $k-r$ columns of F^k. $\Lambda^{k,r}$ and $\Lambda^{k,-r}$ are defined similarly.

Lemma 4. *Suppose that Assumption 1 holds and $r<k<\infty$. Then for any $\theta^k\in\Theta^k(\delta)$ and sufficiently small δ, we have*

$$\|F^{k,r}-F_0\|/\sqrt{T}\lesssim\delta,\quad\|\Lambda^{k,r}-\Lambda_0\|/\sqrt{N}\gtrsim\delta,\quad\|\Lambda^{k,-r}\|/\sqrt{N}\lesssim\delta.$$

Proof. First, similar to (A.9) and (A.10) we can show that for any $\theta^k\in\Theta^k(\delta)$,

$$\|M_{F^k}F_0\|/\sqrt{T}=\|F_0-F^k(F^{k'}F_0/T)\|/\sqrt{T}\lesssim\sqrt{\frac{1}{\sigma_{Nr}}}\cdot d(\theta^k,\theta_0),\qquad(\text{A.30})$$

$$\|M_{F_0}F^{k,r}\|/\sqrt{T}=\|F^{k,r}-F_0(F_0'F^{k,r}/T)\|/\sqrt{T}\lesssim\sqrt{\frac{1}{\rho_r(\Lambda^{k'}\Lambda^k/N)}}\cdot d(\theta^k,\theta_0).\qquad(\text{A.31})$$

9

Similar to (A.11) and (A.12) we can show that

$$\frac{1}{\sqrt{N}}\|\Lambda^k - \Lambda_0(F_0'F^k/T)\| \leq \sqrt{k}d(\theta^k,\theta_0), \quad \frac{1}{\sqrt{N}}\|\Lambda_0 - \Lambda^k(F^{k'}F_0/T)\| \leq \sqrt{r}d(\theta^k,\theta_0). \quad (A.32)$$

With a little abuse of notation again, define $R_T = F^{k'}F_0/T$. From the above inequality, we have

$$\|(\Lambda^k)'\Lambda^k/N - R_T(\Lambda_0'\Lambda_0/N)R_T'\| \leq \left(\frac{\|\Lambda_0 R_T'\|}{\sqrt{N}} + \frac{\|\Lambda^k\|}{\sqrt{N}}\right) \cdot \frac{\|\Lambda^k - \Lambda_0 R_T'\|}{\sqrt{N}} \lesssim d(\theta^k,\theta_0). \quad (A.33)$$

Note that the matrix $R_T(\Lambda_0'\Lambda_0/N)R_T'$ has rank less or equal to r. Thus, according to the Bauer-Fike Theorem, we have

$$\rho_{r+j}\left((\Lambda^k)'\Lambda^k/N\right) - \rho_{r+j}\left(R_T(\Lambda_0'\Lambda_0/N)R_T'\right) = \sigma_{N,r+j}^k \lesssim d(\theta^k,\theta_0) \text{ for } 1 \leq j \leq k-r, \quad (A.34)$$

where $(\Lambda^k)'\Lambda^k/N = \text{diag}\left(\sigma_{N,1}^k,\ldots,\sigma_{N,k}^k\right)$.

Define $R_T^r = F^{k,r'}F_0/T$ and $R_T^{-r} = (F^{k,-r})'F_0/T$, then $R_T = (R_T^{r'}, R_T^{-r'})'$. It then follows from (A.33) and (A.34) that

$$\|R_T^{-r}(\Lambda_0'\Lambda_0/N)R_T^{-r'}\| \leq \|R_T^{-r}(\Lambda_0'\Lambda_0/N)R_T^{-r'} - \text{diag}\left(\sigma_{N,j+1}^k,\ldots,\sigma_{N,k}^k\right)\| + \|\text{diag}\left(\sigma_{N,j+1}^k,\ldots,\sigma_{N,k}^k\right)\| \lesssim d(\theta,\theta_0)$$

which in turn implies that

$$\|R_T^{-r}\|^2 \lesssim \|R_T^{-r}\|_{\max}^2 \lesssim \sqrt{1/\sigma_{N,r}} \cdot d(\theta^k,\theta_0). \quad (A.35)$$

Next, we can write

$$\mathbb{I}_r = F_0'F_0/T = R_T'R_T + F_0'F_0/T - R_T'(F^{k'}F^k/T)R_T = R_T^{r'}R_T^r + R_T^{-r'}R_T^{-r} + \frac{F_0'}{\sqrt{T}} \cdot \frac{F_0 - F^k R_T}{\sqrt{T}}.$$

$$\Lambda_0'\Lambda_0/N = R_T'(\Lambda^{k'}\Lambda^k/N)R_T + \Lambda_0'\Lambda_0/N - R_T'(\Lambda^{k'}\Lambda^k/N)R_T = R_T^{r'}\left(\text{diag}(\sigma_{N,1}^k,\ldots,\sigma_{N,r}^k)\right)R_T^r$$
$$+ R_T^{-r'}\left(\text{diag}(\sigma_{N,r+1}^k,\ldots,\sigma_{N,k}^k)\right)R_T^{-r} + \Lambda_0'(\Lambda_0 - \Lambda^k R_T)/N + (\Lambda_0 - \Lambda^k R_T)'\Lambda R_T/N.$$

Then, similar to (A.16), we can write

$$(\Lambda_0'\Lambda_0/N + D_{NT})R_T^{r'} = R_T^{r'}\left(\text{diag}(\sigma_{N,1}^k,\ldots,\sigma_{N,r}^k)\right), \quad (A.36)$$

where

$$D_{NT} = R_T^{-r'}R_T^{-r} + \frac{F_0'}{\sqrt{T}} \cdot \frac{F_0 - F^k R_T}{\sqrt{T}} - R_T^{-r'}\left(\text{diag}(\sigma_{N,r+1}^k,\ldots,\sigma_{N,k}^k)\right)R_T^{-r}$$
$$- \Lambda_0'(\Lambda_0 - \Lambda^k R_T)/N - (\Lambda_0 - \Lambda^k R_T)'\Lambda R_T/N,$$

and it follows from (A.30) to (A.35) that $\|D_{NT}\| \lesssim d(\theta^k,\theta_0)$. Therefore, similar to the proof of Lemma

10

2, we have

$$\|R_T^r - \mathbb{I}_r\| \lesssim d(\theta^k, \theta_0) \quad \text{and} \quad \|\sigma_{N,j}^k - \sigma_{N,j}\| \lesssim d(\theta^k, \theta_0) \text{ for } j = 1, \ldots, r, \tag{A.37}$$

and we can show that

$$\|F^{k,r} - F_0\|/\sqrt{T} \lesssim d(\theta^k, \theta_0), \quad \|\Lambda^{k,r} - \Lambda_0\|/\sqrt{N} \lesssim d(\theta^k, \theta_0). \tag{A.38}$$

From (A.38), we have

$$\|R_T^{-r}\| = \|F^{k,-r'} F_0/T\| = \|F^{k,-r'}(F_0 - F^{k,r})/T\| \leq \frac{\|F^{k,-r'}\|}{\sqrt{T}} \cdot \frac{\|F_0 - F^{k,r}\|}{\sqrt{T}} \lesssim d(\theta^k, \theta_0). \tag{A.39}$$

It then follows from (A.32) and (A.39) that

$$\|\Lambda^{k,-r}\|/\sqrt{N} \leq \|\Lambda^{k,-r} - \Lambda_0 R_T^{-r'}\|/\sqrt{N} + \|\Lambda_0 R_T^{-r'}\|/\sqrt{N} \lesssim d(\theta^k, \theta_0). \tag{A.40}$$

Then the desired results follow from (A.38) and (A.40). □

Write

$$\mathbb{M}_{NT}(\theta^k) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda_i^{k'} f_t^k), \quad \bar{\mathbb{M}}_{NT}(\theta^k) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[\rho_\tau(X_{it} - \lambda_i^{k'} f_t^k)],$$

$$\mathbb{W}_{NT}(\theta^k) = \mathbb{M}_{NT}(\theta^k) - \bar{\mathbb{M}}_{NT}(\theta^k) - \left(\mathbb{M}_{NT}(\theta_0) - \bar{\mathbb{M}}_{NT}(\theta_0)\right).$$

Lemma 5. *Suppose that Assumption 1 holds and $r < k < \infty$. For sufficiently small δ, we have:*

$$\mathbb{E}\left[\sup_{\theta^k \in \Theta^k(\delta)} |\mathbb{W}_{NT}(\theta^k)|\right] \lesssim \frac{\delta}{L_{NT}}.$$

Proof. Similar to the proof of Lemma 3, we can show that

$$\mathbb{E}\left[\sup_{\Theta^k(\delta)} \sqrt{NT} |\mathbb{W}_{NT}(\theta^k)|\right] \lesssim \int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^k(\delta))} d\epsilon. \tag{A.41}$$

Thus, it remains to be shown that

$$\int_0^\delta \sqrt{\log D(\epsilon, d, \Theta^k(\delta))} d\epsilon = O((N+T)\delta). \tag{A.42}$$

Note that according to the previous lemma, $d(\theta^k, \theta_0) \leq \delta$ implies that $\|F^{k,r} - F_0\|/\sqrt{T} + \|\Lambda^{k,r} - \Lambda_0\|/\sqrt{N} + \|\Lambda^{k,-r}\|/\sqrt{N} \leq K_{10}\delta$. Moreover, we have for some $K_{11} > 0$,

$$\|F^{k,r} - F_0\|/\sqrt{T} + \|\Lambda^{k,r} - \Lambda_0\|/\sqrt{N} + \|\Lambda^{k,-r}\|/\sqrt{N} \geq K_{11}\sqrt{\frac{\|F^{k,r} - F_0\|^2}{T} + \frac{\|\Lambda^{k,r} - \Lambda_0\|^2}{N} + \frac{\|\Lambda^{k,-r}\|^2}{N}}.$$

11

Thus, the set $\Theta^k(\delta)$ is contained in $\Theta^{k**}(\delta)$ where

$$\Theta^{k**}(\delta) = \left\{ \theta^k \in \Theta^k : \sqrt{\frac{\|F^{k,r} - F_0\|^2}{T} + \frac{\|\Lambda^{k,r} - \Lambda_0\|^2}{N} + \frac{\|\Lambda^{k,-r}\|^2}{N}} \leq K_{12}\delta, \frac{\|F^{k,-r}\|}{\sqrt{T}} \leq K_{13} \right\}.$$

In addition, as in the proof of Lemma 3, we can show that for $\theta_a^k, \theta_b^k \in \Theta^k(\delta)$,

$$d(\theta_a^k, \theta_b^k) = \frac{1}{\sqrt{NT}} \|\Lambda_a^k F_a^{k'} - \Lambda_b^k F_b^{k'}\|$$

$$= \frac{1}{\sqrt{NT}} \|\Lambda_a^k F_a^{k'} - \Lambda_b^k F_a^{k'} + \Lambda_b^k F_a^{k'} - \Lambda_b^k F_b^{k'}\| \leq \frac{1}{\sqrt{N}} \|\Lambda_a^k - \Lambda_b^k\| + \frac{1}{\sqrt{NT}} \|\Lambda_b^k (F_a^{k'} - F_b^{k'})\|$$

$$\leq \frac{1}{\sqrt{N}} \|\Lambda_a^k - \Lambda_b^k\| + \frac{1}{\sqrt{N}} \|\Lambda_b^{k,r}\| \cdot \frac{1}{\sqrt{T}} \|F_a^{k,r} - F_b^{k,r}\| + \frac{1}{\sqrt{N}} \|\Lambda_b^{k,-r}\| \cdot \frac{1}{\sqrt{T}} \|F_a^{k,-r} - F_b^{k,-r}\| \lesssim d^{**}(\theta_a^k, \theta_b^k),$$

where

$$d^{**}(\theta_a^k, \theta_b^k) = K_{14} \sqrt{\frac{\|F_a^{k,r} - F_b^{k,r}\|^2}{T} + \frac{\|\Lambda_a^k - \Lambda_b^k\|^2}{N}} + K_{15}\delta \cdot \frac{\|F_a^{k,-r} - F_b^{k,-r}\|}{\sqrt{T}}$$

because $\|\Lambda_b^{k,-r}\|/\sqrt{N} \lesssim \delta$ by the previous lemma. Thus, we have $D(\epsilon, d, \Theta^k(\delta)) \leq D(\epsilon, d^{**}, \Theta^{k**}(\delta))$.

Next, we calculate $D(\epsilon, d^{**}, \Theta^{k**}(\delta))$. Let $(F_1^{k,r}, \Lambda_1^k), \ldots, (F_{m(\epsilon)}^{k,r}, \Lambda_{m(\epsilon)}^k)$ be a maximal set of points in $\Theta^{k**}(\delta)$ such that

$$\sqrt{\frac{\|F_p^{k,r} - F_q^{k,r}\|^2}{T} + \frac{\|\Lambda_p^k - \Lambda_q^k\|^2}{N}} > \epsilon/(2K_{14})$$

for any $p, q \leq m(\epsilon)$ and $p \neq q$. Similar to the proof of Lemma 3, we can show that $m(\epsilon) = (K_{16}\delta/\epsilon)^{Tr+Nk}$. Let $F_1^{k,-r}, \ldots, F_{n(\epsilon)}^{k,-r}$ be a maximal set of points in $\Theta^{k**}(\delta)$ such that

$$\frac{\|F_p^{k,-r} - F_q^{k,-r}\|}{\sqrt{T}} > \epsilon/(2K_{15}\delta)$$

for any $p, q \leq n(\epsilon)$ and $p \neq q$. Then similarly we can show that $n(\epsilon) = (K_{17}\delta/\epsilon)^{T(k-r)}$. Therefore, for any $\theta^k \in \Theta^{k**}(\delta)$, we can find $p^* \leq m(\epsilon)$ and $q^* \leq n(\epsilon)$ such that

$$\sqrt{\frac{\|F^{k,r} - F_{p^*}^{k,r}\|^2}{T} + \frac{\|\Lambda^k - \Lambda_{p^*}^k\|^2}{N}} \leq \epsilon/(2K_{14}), \frac{\|F^{k,-r} - F_{q^*}^{k,-r}\|}{\sqrt{T}} \leq \epsilon/(2K_{15}\delta).$$

Let θ^{k*} consist of $F_{p^*}^{k,r}, F_{q^*}^{k,-r}$ and $\Lambda_{p^*}^k$. Then, it follows by the definition of d^{**} that $d^{**}(\theta^k, \theta^{k*}) \leq \epsilon$. As θ^k varies in $\Theta^{k**}(\delta)$, the number of possible choices for θ^{k*} is bounded by $m(\epsilon) \cdot n(\epsilon) = (K_{18}\delta/\epsilon)^{(T+N)k}$. This means that

$$D(\epsilon, d, \Theta^k(\delta)) \leq D(\epsilon, d^{**}, \Theta^{k**}(\delta)) \leq (K_{18}\delta/\epsilon)^{(T+N)k}. \tag{A.43}$$

Finally, (A.42) follows easily from (A.43) and the desired result follows. $\square$

Proof of Theorem 2:

Proof. First, similar to the proof of Lemma 1, we can show that $d(\hat{\theta}^k, \theta_0) = o_P(1)$. Second, similar to

12

the proof of Theorem 1, it follows from the previous lemma that

$$d(\hat{\theta}^k, \theta_0) = O_P(L_{NT}^{-1}). \tag{A.44}$$

Next, it follows from (A.37), (A.40) and Assumption 1 that

$$|\hat{\sigma}_{N,j}^k - \sigma_j| = o_P(1) \text{ for } j = 1, \ldots, r, \tag{A.45}$$

$$\sum_{j=r+1}^{k} \hat{\sigma}_{N,j}^k = \|\hat{\Lambda}^{k,-r}\|^2 / N \le d(\hat{\theta}^k, \theta_0)^2 = O_P(L_{NT}^{-2}). \tag{A.46}$$

Thus, by (A.45) and (A.46), we have

$$P[\hat{r}_{\mathrm{rank}} \ne r] = P[\hat{r}_{\mathrm{rank}} < r] + P[\hat{r}_{\mathrm{rank}} > r] \le P[\hat{\sigma}_{N,r}^k \le P_{NT}] + P[\hat{\sigma}_{N,r+1}^k > P_{NT}] = o(1). \tag{A.47}$$

It then follows that $P[\hat{r}_{\mathrm{rank}} = r] \to 1$. $\square$

A.3 Proof of Theorem 3

Proof. Following the proof of Bai and Ng (2002), it suffices to show that for some $C > 0$,

$$\mathbb{M}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\hat{\theta}^r) > C + o_P(1) \text{ for } l < r, \tag{A.48}$$

and

$$\mathbb{M}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\hat{\theta}^r) = O_P(1/L_{NT}^2) \text{ for } l > r. \tag{A.49}$$

Case 1: Consider $l < r$. Adding and subtracting terms, we have

$$\begin{aligned}
& \mathbb{M}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\hat{\theta}^r) \\
= \ & \mathbb{M}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\theta_0) + \mathbb{M}_{NT}(\theta_0) - \mathbb{M}_{NT}(\hat{\theta}^r) \\
= \ & \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\theta_0) + \Big(\mathbb{M}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) + \bar{\mathbb{M}}_{NT}(\theta_0) - \mathbb{M}_{NT}(\theta_0) \Big) + \Big(\mathbb{M}_{NT}(\theta_0) - \mathbb{M}_{NT}(\hat{\theta}^r) \Big).
\end{aligned}$$

As in the proof of Lemma 1, it can be shown that

$$\left| \mathbb{M}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) + \bar{\mathbb{M}}_{NT}(\theta_0) - \mathbb{M}_{NT}(\theta_0) \right| = o_P(1).$$

Next, by Lemma 1

$$\left| \mathbb{M}_{NT}(\theta_0) - \mathbb{M}_{NT}(\hat{\theta}^r) \right| \lesssim d(\theta_0, \hat{\theta}^r) = o_P(1).$$

Thus, it remains to show that

$$\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\theta_0) \ge C.$$

Note that by Taylor expansion and Assumption 1(ii), it can be shown that $\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\theta_0) \gtrsim d^2(\hat{\theta}^l, \theta_0)$. Next, similar to (A.9) we can show that $\|M_{\hat{F}^l} F_0\|/\sqrt{T} \lesssim d(\hat{\theta}^l, \theta_0)$. It then follows that

$$\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\theta_0) \gtrsim \|M_{\hat{F}^l} F_0\|^2 / T. \tag{A.50}$$

13

Note that
$$\|M_{\hat{F}^l} F_0\|^2/T = \text{Trace}\left[\mathbb{I}_r - F_0'\hat{F}^l \hat{F}^{l'} F_0/T^2\right] \geq \rho_{max}\left[\mathbb{I}_r - F_0'\hat{F}^l \hat{F}^{l'} F_0/T^2\right]. \tag{A.51}$$

By Lemma A.5 of Ahn and Horenstein (2013), we have
$$\rho_{max}\left[\mathbb{I}_r - F_0'\hat{F}^l \hat{F}^{l'} F_0/T^2\right] + \rho_{min}\left[F_0'\hat{F}^l \hat{F}^{l'} F_0/T^2\right] \geq \rho_{min}\left[\mathbb{I}_r\right]. \tag{A.52}$$

Since $F_0'\hat{F}^l \hat{F}^{l'} F_0$ is a $r \times r$ symmetric matrix with rank less or equal to l, we have $\rho_{min}\left[F_0'\hat{F}^l \hat{F}^{l'} F_0/T^2\right] = 0$, and the above inequality implies that
$$\rho_{max}\left[\mathbb{I}_r - F_0'\hat{F}^l \hat{F}^{l'} F_0/T^2\right] \geq 1. \tag{A.53}$$

Thus, (A.48) follows from (A.50) to (A.53).

Case 2: Now consider $l > r$. Adding and subtracting terms we can write
$$\mathbb{M}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\hat{\theta}^r) = \left(\mathbb{M}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\theta_0) + \bar{\mathbb{M}}_{NT}(\theta_0)\right)$$
$$- \left(\mathbb{M}_{NT}(\hat{\theta}^r) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^r) - \mathbb{M}_{NT}(\theta_0) + \bar{\mathbb{M}}_{NT}(\theta_0)\right) + \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^r). \tag{A.54}$$

First, similar to the proof of Theorem 2, we can show that for sufficiently small δ,
$$\mathbb{E}\left[\sup_{d(\theta^l,\theta_0)\leq\delta} \left|\mathbb{M}_{NT}(\theta^l) - \bar{\mathbb{M}}_{NT}(\theta^l) - \mathbb{M}_{NT}(\theta_0) + \bar{\mathbb{M}}_{NT}(\theta_0)\right|\right] \lesssim \frac{\delta}{L_{NT}},$$

and $d(\hat{\theta}^l, \theta_0) = O_P(1/L_{NT})$. It then follows that
$$\mathbb{M}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \mathbb{M}_{NT}(\theta_0) + \bar{\mathbb{M}}_{NT}(\theta_0) = O_P(1/L_{NT}^2). \tag{A.55}$$

Second, similar to the proof of Lemma 3 and Theorem 1 we can show that
$$\mathbb{M}_{NT}(\hat{\theta}^r) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^r) - \mathbb{M}_{NT}(\theta_0) + \bar{\mathbb{M}}_{NT}(\theta_0) = O_P(1/L_{NT}^2). \tag{A.56}$$

Finally, consider $\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^r)$. We can write
$$\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^r) = \bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\theta_0) - \left(\bar{\mathbb{M}}_{NT}(\hat{\theta}^r) - \bar{\mathbb{M}}_{NT}(\theta_0)\right).$$

Using Taylor expansion and the assumption that $|\mathsf{f}_{it}(\cdot)|$ is uniformly (in i and t) bounded above, it can be shown that
$$\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\theta_0) \lesssim d^2(\hat{\theta}^l, \theta_0) \quad \text{and} \quad \bar{\mathbb{M}}_{NT}(\hat{\theta}^r) - \bar{\mathbb{M}}_{NT}(\theta_0) \lesssim d^2(\hat{\theta}^r, \theta_0).$$

It then follows from $d(\hat{\theta}^l, \theta_0) = O_P(1/L_{NT})$ and $d(\hat{\theta}^r, \theta_0) = O_P(1/L_{NT})$ that
$$\bar{\mathbb{M}}_{NT}(\hat{\theta}^l) - \bar{\mathbb{M}}_{NT}(\hat{\theta}^r) = O_P(1/L_{NT}^2). \tag{A.57}$$

14

Thus, (A.49) follows from (A.54), (A.55), (A.56) and (A.57). The this concludes the proof. □

A.4 Proof of Theorem 4

We only prove the asymptotic distribution of $\tilde{\lambda}_i$ since the proof for $\tilde{f}_t$ is symmetric. Define $\varrho(u) = [\tau - K(u/h)]u$, then we can write

$$\mathbb{S}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varrho(X_{it} - \lambda_i' f_t).$$

Let $\varrho^{(j)}(u) = (\partial/\partial u)^j \varrho(u)$ for $j = 1, 2, 3$. For fixed λ_i, f_t, define

$$\bar{\varrho}(X_{it} - \lambda_i' f_t) = \mathbb{E}\left[\varrho(X_{it} - \lambda_i' f_t)\right], \quad \bar{\varrho}^{(j)}(X_{it} - \lambda_i' f_t) = \mathbb{E}\left[\varrho^{(j)}(X_{it} - \lambda_i' f_t)\right] \text{ for } j = 1, 2, 3.$$

When the functions defined above are evaluated at the true parameters, we suppress their arguments to further simplify the notations. For example, $\varrho_{it} = \varrho(X_{it} - \lambda_{0i}' f_{0t})$, $\bar{\varrho}_{it} = \bar{\varrho}(X_{it} - \lambda_{0i}' f_{0t})$. Moreover, define

$$\bar{\mathbb{S}}_{NT}(\theta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \bar{\varrho}(X_{it} - \lambda_i' f_t),$$

$$\mathbb{U}_{NT}(\theta) = \mathbb{S}_{NT}(\theta) - \bar{\mathbb{S}}_{NT}(\theta) - (\mathbb{S}_{NT}(\theta_0) - \bar{\mathbb{S}}_{NT}(\theta_0)).$$

Using $\bar{O}(1)$ to denote a sequence that is uniformly (over i and t) bounded. Then we can show that

Lemma 6. *Under Assumptions 1 and 2,*
(i) There exists a constant $\bar{C} > 0$ such that $h^{j-1}|\varrho^{(j)}(u)| \leq \bar{C}$ for $j = 1, 2, 3$.
(ii) $\bar{\varrho}_{it}^{(1)} = \bar{O}(h^m)$, $\bar{\varrho}^{(2)}(X_{it} - \lambda_i' f_t) = \mathsf{f}_{it}(\lambda_i' f_t - \lambda_{0i}' f_{0t}) + \bar{O}(h^m)$, and $\bar{\varrho}^{(3)}(X_{it} - \lambda_i' f_t) = \mathsf{f}_{it}^{(1)}(\lambda_i' f_t - \lambda_{0i}' f_{0t}) + \bar{O}(h^m)$.
(iii) $\mathbb{E}(\varrho_{it}^{(1)})^2 = \tau(1 - \tau) + \bar{O}(h)$, and $h \cdot \mathbb{E}\left[(\varrho_{it}^{(2)})^2\right] = \bar{O}(1)$.

Proof. The proof is similar to the standard calculations of the means of kernel density estimators, and it is omitted here to save space. Similar results can be found in Horowitz (1998) and Galvao and Kato (2016). □

Lemma 7. *Under Assumptions 1 and 2, $d(\tilde{\theta}, \theta_0) = o_P(1)$ as $N, T \to \infty$ and $h \to 0$.*

Proof. By definition we have $\mathbb{S}_{NT}(\tilde{\theta}) \leq \mathbb{S}_{NT}(\theta_0)$. Adding and subtracting terms and using (A.1) we have

$$d^2(\tilde{\theta}, \theta_0) \lesssim \bar{\mathbb{M}}_{NT}(\tilde{\theta}) - \bar{\mathbb{M}}_{NT}(\theta_0) \leq \mathbb{M}_{NT}(\tilde{\theta}) - \mathbb{S}_{NT}(\tilde{\theta}) + \mathbb{S}_{NT}(\theta_0) - \mathbb{M}_{NT}(\theta_0) +$$
$$\bar{\mathbb{M}}_{NT}(\tilde{\theta}) - \mathbb{M}_{NT}(\tilde{\theta}) + \mathbb{M}_{NT}(\theta_0) - \bar{\mathbb{M}}_{NT}(\theta_0).$$

It follows that

$$d^2(\tilde{\theta}, \theta_0) \lesssim \sup_{\theta \in \Theta^r} |\mathbb{M}_{NT}(\theta) - \mathbb{S}_{NT}(\theta)| + \sup_{\theta \in \Theta^r} |\mathbb{W}_{NT}(\theta)|.$$

15

It is easy to see that the first term on the RHS of the above inequality is $O_p(h)$, and the second term is $o_P(1)$ as proved in Lemma 1. Then, the desired result follows. $\qquad\square$

Lemma 8. *Under Assumptions 1 and 2, $d(\tilde{\theta}, \theta_0) = O_P(1/L_{NT})$ as $N, T \to \infty$.*

Proof. First, since $\varrho^{(1)}(u)$ is uniformly bounded, we have $|\varrho(X_{it} - c_1) - \varrho(X_{it} - c_2)| \lesssim |c_1 - c_2|$. Then similar to the proof of Lemma 3 we can show that

$$\mathbb{E}\left[\sup_{\theta \in \Theta^r(\delta)} |\mathbb{U}_{NT}(\theta)|\right] \lesssim \frac{\delta}{L_{NT}}. \tag{A.58}$$

Similar to the proof Theorem 1, the parameter space Θ^r can be partitioned into shells $S_j = \{\theta \in \Theta^r : 2^{j-1} < L_{NT} \cdot d(\theta, \theta_0) \le 2^j\}$. Conclude that, for a given integer V and for every $\eta > 0$,

$$P\left[L_{NT} \cdot d(\tilde{\theta}, \theta_0) > 2^V\right] \le \sum_{j > V, 2^{j-1} \le \eta L_{NT}} P\left[\inf_{\theta \in S_j}(\mathbb{S}_{NT}(\theta) - \mathbb{S}_{NT}(\theta_0)) \le 0\right] + P[d(\tilde{\theta}, \theta_0) \ge \eta].$$

For arbitrarily small $\eta > 0$, the second probability on the RHS of the above equation converges to 0 as $N, T \to \infty$ by Lemma 7.

Next, expanding $\mathbb{S}_{NT}(\theta)$ around θ_0 and taking expectations yield

$$\bar{\mathbb{S}}_{NT}(\theta) - \bar{\mathbb{S}}_{NT}(\theta_0) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \bar{\varrho}_{it}^{(1)} \cdot (\lambda_i' f_t - \lambda_{0i}' f_{0t}) + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T 0.5\bar{\varrho}^{(2)}(c_{it}^*) \cdot (\lambda_i' f_t - \lambda_{0i}' f_{0t})^2,$$

where c_{it}^* lies between $\lambda_i' f_t$ and $\lambda_{0i}' f_{0t}$. It then follows from Lemma 6 and Assumption 2 that

$$\bar{\mathbb{S}}_{NT}(\theta) - \bar{\mathbb{S}}_{NT}(\theta_0) \ge O(h^m) + 0.5\underline{\mathfrak{f}} \cdot d^2(\theta, \theta_0).$$

Thus, for each θ in S_j we have

$$-[\bar{\mathbb{S}}_{NT}(\theta) - \bar{\mathbb{S}}_{NT}(\theta_0)] \le -\underline{\mathfrak{f}} \cdot d_{NT}^2(\theta, \theta_0) + O(h^m) \le -\underline{\mathfrak{f}} \cdot \frac{2^{2j-1}}{L_{NT}^2} + O(h^m).$$

Therefore, $\inf_{\theta \in S_j}(\mathbb{S}_{NT}(\theta) - \mathbb{S}_{NT}(\theta_0)) \le 0$ implies that

$$\inf_{\theta \in S_j} \mathbb{U}_{NT}(\theta) \le -\underline{\mathfrak{f}} \cdot \frac{2^{2j-1}}{L_{NT}^2} + O(h^m),$$

and it follows that

$$\sum_{j > V, 2^{j-1} \le \eta L_{NT}} P\left[\inf_{\theta \in S_j}(\mathbb{S}_{NT}(\theta) - \mathbb{S}_{NT}(\theta_0)) \le 0\right]$$

$$\le \sum_{j > V, 2^{j-1} \le \eta L_{NT}} P\left[\sup_{\theta \in S_j} |\mathbb{U}_{NT}(\theta)| \ge \underline{\mathfrak{f}} \cdot \frac{2^{2j-1}}{L_{NT}^2} + O(h^m)\right].$$

16

By ([A.58](#)) and Markov's inequality, we have

$$P\left[\sup_{\theta \in S_j} |\mathbb{U}_{NT}(\theta)| \geq \underline{\mathtt{f}} \cdot \frac{2^{2j-1}}{L_{NT}^2} + O(h^m)\right] \lesssim \frac{2^j}{2^{2j} + O(L_{NT}^2 \cdot h^m)}.$$

By Assumption 2(v), $O(L_{NT}^2 \cdot h^m) = o(1)$. Thus, the above inequality implies that

$$\sum_{j>V, 2^{j-1} \leq \eta L_{NT}} P\left[\inf_{\theta \in S_j} (\mathbb{M}_{NT}(\theta) - \mathbb{M}_{NT}(\theta_0)) \leq 0\right] \lesssim \sum_{j>V} 2^{-j}.$$

The RHS of the previous expression convergences to 0 as $V \to \infty$, implying that $L_{NT} \cdot d(\tilde{\theta}, \theta_0) = O_P(1)$, or $d(\tilde{\theta}, \theta_0) = O_P(1/L_{NT})$. $\qquad\square$

Define:

$$\mathbb{S}_{i,T}(\lambda, F) = \frac{1}{T}\sum_{t=1}^T \varrho(X_{it} - \lambda' f_t), \quad \bar{\mathbb{S}}_{i,T}(\lambda, F) = \frac{1}{T}\sum_{t=1}^T \bar{\varrho}(X_{it} - \lambda' f_t),$$

$$\mathbb{M}_{i,T}(\lambda, F) = \frac{1}{T}\sum_{t=1}^T \rho_\tau(X_{it} - \lambda' f_t), \quad \bar{\mathbb{M}}_{i,T}(\lambda, F) = \frac{1}{T}\sum_{t=1}^T \mathbb{E}\rho_\tau(X_{it} - \lambda' f_t).$$

Lemma 9. *Under Assumptions 1 and 2, we have $\|\tilde{\lambda}_i - \lambda_{0i}\| = o_P(1)$ for each i.*

Proof. Note that

$$\tilde{\lambda}_i = \operatorname*{arg\,min}_{\lambda \in \mathcal{A}} \mathbb{S}_{i,T}(\lambda, \tilde{F}).$$

First, we show that

$$\sup_{\lambda \in \mathcal{A}} |\mathbb{S}_{i,T}(\lambda, \tilde{F}) - \bar{\mathbb{M}}_{i,T}(\lambda, F_0)| = o_P(1). \tag{A.59}$$

Adding and subtracting terms we have

$$\mathbb{S}_{i,T}(\lambda, \tilde{F}) - \bar{\mathbb{M}}_{i,T}(\lambda, F_0) = \mathbb{S}_{i,T}(\lambda, \tilde{F}) - \mathbb{M}_{i,T}(\lambda, \tilde{F}) + \mathbb{M}_{i,T}(\lambda, \tilde{F}) - \mathbb{M}_{i,T}(\lambda, F_0) + \mathbb{M}_{i,T}(\lambda, F_0) - \bar{\mathbb{M}}_{i,T}(\lambda, F_0),$$

thus,

$$\sup_{\lambda \in \mathcal{A}} |\mathbb{S}_{i,T}(\lambda, \tilde{F}) - \bar{\mathbb{M}}_{i,T}(\lambda, F_0)| \leq \sup_{\lambda \in \mathcal{A}} |\mathbb{S}_{i,T}(\lambda, \tilde{F}) - \mathbb{M}_{i,T}(\lambda, \tilde{F})| +$$
$$\sup_{\lambda \in \mathcal{A}} |\mathbb{M}_{i,T}(\lambda, \tilde{F}) - \mathbb{M}_{i,T}(\lambda, F_0)| + \sup_{\lambda \in \mathcal{A}} |\mathbb{M}_{i,T}(\lambda, F_0) - \bar{\mathbb{M}}_{i,T}(\lambda, F_0)|.$$

It is easy to show that

$$\sup_{\lambda \in \mathcal{A}} |\mathbb{S}_{i,T}(\lambda, \tilde{F}) - \mathbb{M}_{i,T}(\lambda, \tilde{F})| \lesssim h,$$

$$\sup_{\lambda \in \mathcal{A}} |\mathbb{M}_{i,T}(\lambda, \tilde{F}) - \mathbb{M}_{i,T}(\lambda, F_0)| \lesssim \sup_{\lambda \in \mathcal{A}} \|\lambda\| \cdot \frac{1}{T}\sum_{t=1}^T \|\tilde{f}_t - f_{0t}\| \lesssim \|\tilde{F} - F_0\|/\sqrt{T} = O_P(1/L_{NT}),$$

$$\sup_{\lambda \in \mathcal{A}} |\mathbb{M}_{i,T}(\lambda, F_0) - \bar{\mathbb{M}}_{i,T}(\lambda, F_0)| = \sup_{\lambda \in \mathcal{A}} \left|\frac{1}{T}\sum_{t=1}^T [\rho_\tau(X_{it} - \lambda' f_{0t}) - \mathbb{E}\rho_\tau(X_{it} - \lambda' f_{0t})]\right| = o_P(1).$$

17

Then (A.59) follows as $h \to 0$.

Second, we can show that for any $\epsilon > 0$, and $B_i(\epsilon) = \{\lambda \in \mathcal{A} : \|\lambda - \lambda_{0i}\| \leq \epsilon\}$,

$$\inf_{\lambda \in B_i^C(\epsilon)} \bar{\mathbb{M}}_{i,T}(\lambda, F_0) - \bar{\mathbb{M}}_{i,T}(\lambda_{0i}, F_0) > 0, \tag{A.60}$$

e.g. the proof of Proposition 3.1 of Galvao and Kato (2016).

Finally, given (A.59) and (A.60), the consistency of $\tilde{\lambda}_i$ follows from a standard proof for the consistency of M-estimators (see Theorem 2.1 of Newey and McFadden 1994), so that the proof concludes. $\qquad \square$

Lemma 10. *Under Assumptions 1 and 2, we have $\|\tilde{\lambda}_i - \lambda_{0i}\| = O_P\left(T^{-1/2}h^{-1}\right)$ for each i.*

Proof. For any fixed $\lambda_i \in \mathcal{A}$ and $f_t \in \mathcal{F}$, expanding $\varrho^{(1)}(X_{it} - \lambda_i' f_t) f_t$ gives

$$
\begin{aligned}
&\varrho^{(1)}(X_{it} - \lambda_i' f_t) f_t \\
=&\varrho^{(1)}(X_{it} - \lambda_{0i}' f_t) f_t - \varrho^{(2)}(X_{it} - \lambda_{0i}' f_t) f_t f_t' \cdot (\lambda_i - \lambda_{0i}) + 0.5 \varrho^{(3)}(X_{it} - \lambda_i^{*'} f_t) f_t [(\lambda_i - \lambda_{0i})' f_t]^2 \\
=&\varrho_{it}^{(1)} f_{0t} + \varrho^{(1)}(X_{it} - \lambda_{0i}' f_t^*)(f_t - f_{0t}) - \varrho^{(2)}(X_{it} - \lambda_{0i}' f_t^*) f_t^* \lambda_{0i}'(f_t - f_{0t}) - \varrho_{it}^{(2)} f_t f_t' \cdot (\lambda_i - \lambda_{0i}) \\
&+ \varrho^{(3)}(X_{it} - \lambda_{0i}' f_t^*) f_t f_t' \cdot (\lambda_i - \lambda_{0i}) \lambda_{0i}'(f_t - f_{0t}) + 0.5 \varrho^{(3)}(X_{it} - \lambda_i^{*'} f_t) f_t [(\lambda_i - \lambda_{0i})' f_t]^2,
\end{aligned}
$$

where λ_i^* lies between λ_i and λ_{0i} and f_t^* lies between f_t and f_{0t}. Taking expectations of both sides of the above equation, and setting $\lambda_i = \tilde{\lambda}_i, f_t = \tilde{f}_t$, it follows from Lemma 4 that:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} \bar{\varrho}^{(1)}(X_{it} - \tilde{\lambda}_i' \tilde{f}_t)\tilde{f}_t =\ & \frac{1}{T}\sum_{t=1}^{T} \bar{\varrho}_{it}^{(1)} f_{0t} - \left(\frac{1}{T}\sum_{t=1}^{T} \bar{\varrho}_{it}^{(2)} \tilde{f}_t \tilde{f}_t'\right)(\tilde{\lambda}_i - \lambda_{0i}) \\
& + O_P\left(T^{-1/2}\|\tilde{F} - F_0\|\right) + O_P(\|\tilde{\lambda}_i - \lambda_{0i}\|) \cdot O_P\left(T^{-1/2}\|\tilde{F} - F_0\|\right) + O_P(\|\tilde{\lambda}_i - \lambda_{0i}\|^2).
\end{aligned}
$$

Lemma 6, Lemma 8 and Assumption 2 imply that:

$$\frac{1}{T}\sum_{t=1}^{T} \bar{\varrho}_{it}^{(2)} \tilde{f}_t \tilde{f}_t' = \frac{1}{T}\sum_{t=1}^{T} \bar{\varrho}_{it}^{(2)} f_{0t} f_{0t}' + o_P(1) = \Phi_i + o_P(1).$$

Then, from Lemma 6, Lemma 8 and Lemma 9 we get

$$\Phi_i(\tilde{\lambda}_i - \lambda_{0i}) + o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|) = O(h^m) + O_P(1/L_{NT}) - \frac{1}{T}\sum_{t=1}^{T} \bar{\varrho}^{(1)}(X_{it} - \tilde{\lambda}_i' \tilde{f}_t)\tilde{f}_t. \tag{A.61}$$

Note that we can write

$$\frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t$$

$$= \quad -\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}f_{0t} - \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t - \tilde{\varrho}_{it}^{(1)}f_{0t}\right]$$

$$= \quad -\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}f_{0t} - \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t - \tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'f_{0t})f_{0t}\right] - \frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'f_{0t}) - \tilde{\varrho}_{it}^{(1)}\right]f_{0t}.$$

The first term on the RHS of the above equation is $O_P(T^{-1/2})$ by Lemma 6 and Lyapunov's CLT. As for the second term on the RHS of the above equation, we have that

$$\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t - \tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'f_{0t})f_{0t}\right] =$$

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'f_t^*)(\tilde{f}_t - f_{0t}) - \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}^{(2)}(X_{it}-\tilde{\lambda}_i'f_t^*)f_t^*\tilde{\lambda}_i'(\tilde{f}_t - f_{0t}), \quad \text{(A.62)}$$

where f_t^* lies between $\tilde{f}_t$ and f_{0t}. The first term on the right of (A.62) is $O_P(1/L_{NT})$ because $\varrho^{(1)}$ is uniformly bounded and $T^{-1}\sum_{t=1}^{T}\|\tilde{f}_t - f_{0t}\| = O_P(1/L_{NT})$ by Lemma 8. Similarly, the second term on the RHS of (A.62) is $O_P(1/(L_{NT}h))$ because $h\varrho^{(2)}(u)$ is uniformly bounded. Finally, we can show that (see, e.g., Lemma B.2 of Galvao and Kato 2016)

$$\frac{1}{T}\sum_{t=1}^{T}\left[\tilde{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'f_{0t}) - \tilde{\varrho}_{it}^{(1)}\right]f_{0t} = O_P(\|\tilde{\lambda}_i - \lambda_{0i}\|) \cdot O_P(1/\sqrt{Th}) = o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|).$$

Combining the above results yields

$$\frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t = O_P\left(\frac{1}{L_{NT}h}\right) + o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|), \quad \text{(A.63)}$$

and the desired result follows from (A.61), (A.63) and Assumption 2. $\qquad\square$

To derive the asymptotic distribution of $\tilde{\lambda}_i$, it is essential to obtain the stochastic expansion of $\tilde{f}_t$.

To do so, define

$$\mathbb{P}_{NT}(\theta) = b\left[\frac{1}{2N}\sum_{p=1}^{r}\sum_{q>p}^{r}\left(\sum_{i=1}^{N}\lambda_{ip}\lambda_{iq}\right)^2 + \frac{1}{2T}\sum_{p=1}^{r}\sum_{q>p}^{r}\left(\sum_{t=1}^{T}f_{tp}f_{tq}\right)^2 + \frac{1}{8T}\sum_{k=1}^{r}\left(\sum_{t=1}^{T}f_{tk}^2 - T\right)^2\right]$$

for some $b > 0$. Define

$$\mathcal{S}^*(\theta) = \Big[\underbrace{\ldots, -\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\bar{\varrho}^{(1)}(X_{it}-\lambda_i'f_t)f_t', \ldots}_{1\times Nr}, \underbrace{\ldots, -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\bar{\varrho}^{(1)}(X_{it}-\lambda_i'f_t)\lambda_i', \ldots}_{1\times Tr}\Big]',$$

19

$$S(\theta) = S^*(\theta) + \partial \mathbb{P}_{NT}(\theta)/\partial\theta, \quad \mathcal{H}(\theta) = \partial S^*(\theta)/\partial\theta' + \partial^2 \mathbb{P}_{NT}(\theta)/\partial\theta\partial\theta',$$

and let $\mathcal{H} = \mathcal{H}(\theta_0)$.

Expanding $S(\tilde{\theta})$ around $S(\theta_0)$ gives:

$$S(\tilde{\theta}) = S(\theta_0) + \mathcal{H} \cdot (\tilde{\theta} - \theta_0) + 0.5\mathcal{R}(\tilde{\theta}), \tag{A.64}$$

where

$$\mathcal{R}(\tilde{\theta}) = \left(\sum_{j=1}^{M} \partial \mathcal{H}(\theta^*)/\partial\theta_j \cdot (\tilde{\theta}_j - \theta_{0j}) \right) (\tilde{\theta} - \theta_0),$$

and θ^* lies between $\tilde{\theta}$ and θ_0.

Further, define

$$\mathcal{H}_d = \begin{pmatrix} \mathcal{H}_d^\Lambda & 0 \\ 0 & \mathcal{H}_d^F \end{pmatrix}, \quad \mathcal{H}_d^\Lambda = \frac{\sqrt{T}}{\sqrt{N}} \,\mathrm{diag}\,[\Phi_{T,1}, \dots, \Phi_{T,i}, \dots, \Phi_{T,N}], \quad \mathcal{H}_d^F = \frac{\sqrt{N}}{\sqrt{T}} \,\mathrm{diag}\,[\Psi_{N,1}, \dots, \Psi_{N,t}, \dots, \Psi_{N,T}],$$

where

$$\Phi_{T,i} = \frac{1}{T} \sum_{t=1}^{T} \bar{\varrho}_{it}^{(2)} f_{0t} f_{0t}', \quad \Psi_{N,t} = \frac{1}{N} \sum_{i=1}^{N} \bar{\varrho}_{it}^{(2)} \lambda_{0i} \lambda_{0i}'.$$

The following lemma is important for the stochastic expansion of $\tilde{f}_t$.

Lemma 11. *The matrix $\mathcal{H}$ is invertible and $\|\mathcal{H}^{-1} - \mathcal{H}_d^{-1}\|_{\max} = O(1/T)$.*

Proof. To simplify the notations, we consider the case $r = 2$, but the proof can be easily generalized to the case $r > 2$. Note that $\lambda_{0i} = (\lambda_{0i,1}, \lambda_{0i,2})'$ and $f_{0t} = (f_{0t,1}, f_{0t,2})'$, and $\mathbb{P}_{NT}(\theta)$ simplifies to

$$\mathbb{P}_{NT}(\theta) = b \left[\frac{1}{2N} \left(\sum_{i=1}^{N} \lambda_{i1}\lambda_{i2} \right)^2 + \frac{1}{2T} \left(\sum_{t=1}^{T} f_{t1}f_{t2} \right)^2 + \frac{1}{8T} \left(\sum_{t=1}^{T} f_{t1}^2 - T \right)^2 + \frac{1}{8T} \left(\sum_{t=1}^{T} f_{t2}^2 - T \right)^2 \right].$$

First, define

$$\begin{aligned}
\gamma_1' &= [\mathbf{0}_{1\times 2N}, (f_{01,1}, 0), \dots, (f_{0t,1}, 0), \dots, (f_{0T,1}, 0)]/\sqrt{T}, \\
\gamma_2' &= [\mathbf{0}_{1\times 2N}, (0, f_{01,2}), \dots, (0, f_{0t,2}), \dots, (0, f_{0T,2})]/\sqrt{T}, \\
\gamma_3' &= [\mathbf{0}_{1\times 2N}, (f_{01,2}, f_{01,1}), \dots, (f_{0t,2}, f_{0t,1}), \dots, (f_{0T,2}, f_{0T,1})]/\sqrt{T}, \\
\gamma_4' &= [(\lambda_{01,2}, \lambda_{01,1}), \dots, (\lambda_{0i,2}, \lambda_{0i,1}), \dots, (\lambda_{0N,2}, \lambda_{0N,1}), \mathbf{0}_{1\times 2T}]/\sqrt{N},
\end{aligned}$$

and note that $\partial^2 \mathbb{P}_{NT}(\theta_0)/\partial\theta\partial\theta' = b\left(\sum_{k=1}^{4} \gamma_k \gamma_k' \right)$.

Second, define

$$\omega_1' = \Big[\underbrace{(\lambda_{01,1},0)/\sqrt{N},\ldots,(\lambda_{0N,1},0)/\sqrt{N}}_{\omega_{1\Lambda}'},\underbrace{(-f_{01,1},0)/\sqrt{T},\ldots,(-f_{0T,1},0)/\sqrt{T}}_{\omega_{1F}'}\Big],$$

$$\omega_2' = \Big[\underbrace{(0,\lambda_{01,2})/\sqrt{N},\ldots,(0,\lambda_{0N,2})/\sqrt{N}}_{\omega_{2\Lambda}'},\underbrace{(0,-f_{01,2})/\sqrt{T},\ldots,(0,-f_{0T,2})/\sqrt{T}}_{\omega_{2F}'}\Big],$$

$$\omega_3' = \Big[\underbrace{(\lambda_{01,2},0)/\sqrt{N},\ldots,(\lambda_{0N,2},0)/\sqrt{N}}_{\omega_{3\Lambda}'},\underbrace{(0,-f_{01,1})/\sqrt{T},\ldots,(0,-f_{0T,1})/\sqrt{T}}_{\omega_{3F}'}\Big],$$

$$\omega_4' = \Big[\underbrace{(0,\lambda_{01,1})/\sqrt{N},\ldots,(0,\lambda_{0N,1})/\sqrt{N}}_{\omega_{4\Lambda}'},\underbrace{(-f_{01,2},0)/\sqrt{T},\ldots,(-f_{0T,2},0)/\sqrt{T}}_{\omega_{4F}'}\Big],$$

and $\omega = [\omega_1,\omega_2,\omega_3,\omega_4]$. It is easy to check that $\omega_p'\omega_q = 0$ for $p \neq q$. Moreover, we have

$$\omega\omega' = \sum_{k=1}^{4}\omega_k\omega_k' = \begin{pmatrix} \sum_{k=1}^{4}\omega_{k\Lambda}\omega_{k\Lambda}' & -(NT)^{-1/2}\{f_{0t}\lambda_{0i}'\}_{i\leq N,t\leq T} \\ -(NT)^{-1/2}\{\lambda_{0i}f_{0t}'\}_{t\leq T,i\leq N}, & \sum_{k=1}^{4}\omega_{kF}\omega_{kF}' \end{pmatrix}, \tag{A.65}$$

where $\{f_{0t}\lambda_{0i}'\}_{i\leq N,t\leq T}$ denotes a $2N \times 2T$ matrix whose $\{i,t\}$th block is $f_{0t}\lambda_{0i}'$. Further, it is easy to see that under our normalizations,

$$\omega'\omega = \begin{pmatrix} \sigma_{N1}+1 & 0 & 0 & 0 \\ 0 & \sigma_{N2}+1 & 0 & 0 \\ 0 & 0 & \sigma_{N2}+1 & 0 \\ 0 & 0 & 0 & \sigma_{N1}+1 \end{pmatrix}.$$

Next, we project γ_k onto ω, and write $\gamma_k = \omega\beta_k + \zeta_k$ for $k = 1,\ldots,4$, where $\beta_k = (\omega'\omega)^{-1}\omega'\gamma_k$. In particular,

$$\beta_1 = \begin{pmatrix} -\frac{1}{\sigma_{N1}+1} \\ 0 \\ 0 \\ 0 \end{pmatrix},\quad \beta_2 = \begin{pmatrix} 0 \\ -\frac{1}{\sigma_{N2}+1} \\ 0 \\ 0 \end{pmatrix},\quad \beta_3 = \begin{pmatrix} 0 \\ 0 \\ -\frac{1}{\sigma_{N2}+1} \\ -\frac{1}{\sigma_{N1}+1} \end{pmatrix},\quad \beta_4 = \begin{pmatrix} 0 \\ 0 \\ \frac{\sigma_{N2}}{\sigma_{N2}+1} \\ \frac{\sigma_{N1}}{\sigma_{N1}+1} \end{pmatrix}.$$

Define $B_N = \sum_{k=1}^{4}\beta_k\beta_k'$. It is easy to show that there exists $\underline{\rho} > 0$ such that $\rho_{\min}(B_N) > \underline{\rho}$ for all large N as long as $\sigma_{N1} - \sigma_{N2}$ is bounded below by a positive constant for all large N, which is true under our assumption that $\sigma_{N1} \to \sigma_1$, $\sigma_{N2} \to \sigma_2$, and $\sigma_1 > \sigma_2$. It then follows that

$$\partial^2\mathbb{P}_{NT}(\theta_0)/\partial\theta\partial\theta' = b\left(\sum_{k=1}^{4}\gamma_k\gamma_k'\right) = b\cdot\omega\left(\sum_{k=1}^{4}\beta_k\beta_k'\right)\omega' + b\left(\sum_{k=1}^{4}\zeta_k\zeta_k'\right)$$

$$= b\underline{\rho}\cdot\omega\omega' + b\cdot\omega(B_N - \underline{\rho}\cdot\mathbb{I}_4)\omega' + b\left(\sum_{k=1}^{4}\zeta_k\zeta_k'\right). \tag{A.66}$$

21

Now let $\underline{b} = \min\{\underline{f}, b\rho\}$. Then it follows from (A.66) that:

$$
\begin{aligned}
\mathcal{H} &= \partial \mathcal{S}^*(\theta_0)/\partial\theta' + \partial^2 \mathbb{P}_{NT}(\theta_0)/\partial\theta\partial\theta' \\
&= \partial \mathcal{S}^*(\theta_0)/\partial\theta' + \underline{b} \cdot \omega\omega' + \underbrace{(b\rho - \underline{b}) \cdot \omega\omega'}_{\geq 0} + \underbrace{b \cdot \omega(B_N - \rho I_4)\omega'}_{\geq 0} + b \underbrace{\left(\sum_{k=1}^{4} \zeta_k \zeta_k' \right)}_{\geq 0} \\
&\geq \partial \mathcal{S}^*(\theta_0)/\partial\theta' + \underline{b} \cdot \omega\omega'.
\end{aligned}
$$

Moreover, we can write

$$
\begin{aligned}
&\partial \mathcal{S}^*(\theta_0)/\partial\theta' \\
&= \begin{pmatrix} (NT)^{-1/2} \operatorname{diag}\left[\left\{\sum_{t=1}^{T} \bar{\varrho}_{it}^{(2)} f_{0t} f_{0t}'\right\}_{i \leq N}\right] & (NT)^{-1/2} \left\{\bar{\varrho}_{it}^{(2)} f_{0t} \lambda_{0i}'\right\}_{i \leq N, t \leq T} \\ (NT)^{-1/2} \left\{\bar{\varrho}_{it}^{(2)} \lambda_{0i} f_{0t}'\right\}_{t \leq T, i \leq N} & (NT)^{-1/2} \operatorname{diag}\left[\left\{\sum_{i=1}^{N} \bar{\varrho}_{it}^{(2)} \lambda_{0i} \lambda_{0i}'\right\}_{t \leq N}\right] \end{pmatrix} \\
&= \underbrace{\underline{b} \begin{pmatrix} \operatorname{diag}\left[\left\{(NT)^{-1/2} \sum_{t=1}^{T} f_{0t} f_{0t}'\right\}_{i \leq N}\right] & \mathbf{0}_{2N \times 2T} \\ \mathbf{0}_{2T \times 2N} & \operatorname{diag}\left[\left\{(NT)^{-1/2} \sum_{i=1}^{N} \lambda_{0i} \lambda_{0i}'\right\}_{t \leq N}\right] \end{pmatrix}}_{I} \\
&+ \underbrace{\underline{b} \begin{pmatrix} \mathbf{0}_{2N \times 2N} & (NT)^{-1/2} \left\{f_{0t} \lambda_{0i}'\right\}_{i \leq N, t \leq T} \\ (NT)^{-1/2} \left\{\lambda_{0i} f_{0t}'\right\}_{t \leq T, i \leq N} & \mathbf{0}_{2T \times 2T} \end{pmatrix}}_{II} \\
&+ \underbrace{\begin{pmatrix} (NT)^{-1/2} \operatorname{diag}\left[\left\{\sum_{t=1}^{T}(\bar{\varrho}_{it}^{(2)} - \underline{b}) f_{0t} f_{0t}'\right\}_{i \leq N}\right] & (NT)^{-1/2} \left\{(\bar{\varrho}_{it}^{(2)} - \underline{b}) f_{0t} \lambda_{0i}'\right\}_{i \leq N, t \leq T} \\ (NT)^{-1/2} \left\{(\bar{\varrho}_{it}^{(2)} - \underline{b}) \lambda_{0i} f_{0t}'\right\}_{t \leq T, i \leq N} & (NT)^{-1/2} \operatorname{diag}\left[\left\{\sum_{i=1}^{N}(\bar{\varrho}_{it}^{(2)} - \underline{b}) \lambda_{0i} \lambda_{0i}'\right\}_{t \leq N}\right] \end{pmatrix}}_{III}.
\end{aligned}
$$

Note that by our assumptions there exists a constant $\underline{c} > 0$ such that:

$$
I = \underline{b} \begin{pmatrix} \sqrt{T/N} \cdot \mathbb{I}_{2N} & \mathbf{0}_{2N \times 2T} \\ \mathbf{0}_{2T \times 2N} & \sqrt{N/T} \cdot \mathbb{I}_T \otimes \operatorname{diag}(\sigma_{N1}, \sigma_{N2}) \end{pmatrix} \geq \underline{c} \cdot \mathbb{I}_{2(N+T)}. \tag{A.67}
$$

From (A.65) we have

$$
II + \underline{b} \cdot \omega\omega' = \underline{b} \cdot \begin{pmatrix} \sum_{k=1}^{4} \omega_{k\Lambda} \omega_{k\Lambda}' & \mathbf{0}_{2N \times 2T} \\ \mathbf{0}_{2T \times 2N} & \sum_{k=1}^{4} \omega_{kF} \omega_{kF}' \end{pmatrix} \geq 0. \tag{A.68}
$$

For the last term we have for N, T large enough,

$$
III = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\bar{\varrho}_{it}^{(2)} - \underline{b}) \mu_{it} \mu_{it}' \geq 0, \tag{A.69}
$$

where $\mu_{it} = [\underbrace{\mathbf{0}_{1\times2}, \ldots, f_{0t}', \ldots, \mathbf{0}_{1\times2}}_{1 \times 2N}, \underbrace{\mathbf{0}_{1\times2}, \ldots, \lambda_{0i}', \ldots, \mathbf{0}_{1\times2}}_{1 \times 2T}]'$, because Assumption 1 and Lemma 6 imply

that $\bar{\varrho}_{it}^{(2)} \geq \underline{f}$ for all i, t. It then follows from (A.67), (A.68) and (A.69) that

$$\mathcal{H} \geq \partial \mathcal{S}^*(\theta_0)/\partial \theta' + \underline{b} \cdot \omega \omega' = I + II + III + \underline{b} \cdot \omega \omega' \geq \underline{c} \cdot \mathbb{I}_{2(N+T)},$$

and thus

$$\mathcal{H}^{-1} \leq \underline{c}^{-1} \cdot \mathbb{I}_{2(N+T)}. \tag{A.70}$$

Finally, write $\mathcal{H} = \mathcal{H}_d + \mathcal{C}$, where

$$\mathcal{C} = \begin{pmatrix} \mathbf{0}_{2N \times 2N} & (NT)^{-1/2} \left\{ \bar{\varrho}_{it}^{(2)} f_{0t} \lambda'_{0i} \right\}_{i \leq N, t \leq T} \\ (NT)^{-1/2} \left\{ \bar{\varrho}_{it}^{(2)} \lambda_{0i} f'_{0t} \right\}_{t \leq T, i \leq N} & \mathbf{0}_{2T \times 2T} \end{pmatrix} + b \left(\sum_{k=1}^{4} \gamma_k \gamma'_k \right).$$

Note that $\mathcal{H}^{-1} - \mathcal{H}_d^{-1} = -\mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}_d^{-1} + \mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}^{-1} \mathcal{C} \mathcal{H}_d^{-1}$, and thus $\left\| \mathcal{H}^{-1} - \mathcal{H}_d^{-1} \right\|_{\max} \leq \left\| \mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}_d^{-1} \right\|_{\max} + \left\| \mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}^{-1} \mathcal{C} \mathcal{H}_d^{-1} \right\|_{\max}$. Inequality (A.70) implies that $\mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}^{-1} \mathcal{C} \mathcal{H}_d^{-1} \leq \underline{c}^{-1} \mathcal{H}_d^{-1} \mathcal{C}^2 \mathcal{H}_d^{-1}$, and thus the jth diagonal element of $\mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}^{-1} \mathcal{C} \mathcal{H}_d^{-1}$ is smaller than the jth diagonal element of $\underline{c}^{-1} \mathcal{H}_d^{-1} \mathcal{C}^2 \mathcal{H}_d^{-1}$. It then follows that $\left\| \mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}^{-1} \mathcal{C} \mathcal{H}_d^{-1} \right\|_{\max} \leq \underline{c}^{-1} \left\| \mathcal{H}_d^{-1} \mathcal{C}^2 \mathcal{H}_d^{-1} \right\|_{\max}$ and therefore

$$\left\| \mathcal{H}^{-1} - \mathcal{H}_d^{-1} \right\|_{\max} \leq \left\| \mathcal{H}_d^{-1} \mathcal{C} \mathcal{H}_d^{-1} \right\|_{\max} + \underline{c}^{-1} \left\| \mathcal{H}_d^{-1} \mathcal{C}^2 \mathcal{H}_d^{-1} \right\|_{\max}$$

because the entry with the largest absolute value of a positive semidefinite matrix is always on the diagonal. Since $\mathcal{H}_d^{-1}$ is a block-diagonal matrix whose elements are all $O(1)$ by Assumption 2, and both $\left\| \mathcal{C} \right\|_{\max}$ and $\left\| \mathcal{C}^2 \right\|_{\max}$ can be shown to be $O(1/T)$, then the desired result follows. $\qquad \square$

Since $\partial \mathbb{P}_{NT}(\tilde{\theta})/\partial \theta = \partial \mathbb{P}_{NT}(\theta_0)/\partial \theta = 0$, (A.64) imply that

$$\tilde{\theta} - \theta_0 = \mathcal{H}^{-1} \mathcal{S}^*(\tilde{\theta}) - \mathcal{H}^{-1} \mathcal{S}^*(\theta_0) - 0.5 \mathcal{H}^{-1} \mathcal{R}(\tilde{\theta}). \tag{A.71}$$

Define

$$\mathcal{S}_{NT}^*(\theta) = \left[\underbrace{\ldots, -\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \varrho^{(1)}(X_{it} - \lambda'_i f_t) f'_t, \ldots,}_{1 \times Nr} \underbrace{\ldots, -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varrho^{(1)}(X_{it} - \lambda'_i f_t) \lambda'_i, \ldots}_{1 \times Tr} \right]',$$

$\tilde{\mathcal{S}}^*(\theta) = \mathcal{S}_{NT}^*(\theta) - \mathcal{S}^*(\theta)$ and $\mathcal{D} = \mathcal{H}^{-1} - \mathcal{H}_d^{-1}$. Note that by the first order conditions, $\mathcal{S}_{NT}^*(\tilde{\theta}) = 0$. As a result, we can write

$$\mathcal{H}^{-1} \mathcal{S}^*(\tilde{\theta}) = \mathcal{H}_d^{-1} \mathcal{S}^*(\tilde{\theta}) + \mathcal{D} \mathcal{S}^*(\tilde{\theta}) = -\mathcal{H}_d^{-1} \tilde{\mathcal{S}}^*(\tilde{\theta}) + \mathcal{D} \mathcal{S}^*(\tilde{\theta})$$

$$= -\mathcal{H}_d^{-1} \tilde{\mathcal{S}}^*(\theta_0) - \mathcal{H}_d^{-1} \left(\tilde{\mathcal{S}}^*(\tilde{\theta}) - \tilde{\mathcal{S}}^*(\theta_0) \right) + \mathcal{D} \mathcal{S}^*(\tilde{\theta}) \tag{A.72}$$

$$= -\mathcal{H}_d^{-1} \tilde{\mathcal{S}}^*(\theta_0) - \mathcal{H}_d^{-1} \left(\tilde{\mathcal{S}}^*(\tilde{\theta}) - \tilde{\mathcal{S}}^*(\theta_0) \right) - \mathcal{D} \tilde{\mathcal{S}}^*(\theta_0) - \mathcal{D} \left(\tilde{\mathcal{S}}^*(\tilde{\theta}) - \tilde{\mathcal{S}}^*(\theta_0) \right). \tag{A.73}$$

Next, let $\mathcal{R}(\tilde{\theta})_j$ denote the vector containing the $(j-1)r + 1$th to the jrth elements of $\mathcal{R}(\tilde{\theta})$ for $j = 1, \ldots, N+T$, and let $\bar{O}_P()$ denote a stochastic order that is uniformly in i and t.[2] Then by the result

---

[2] For example, $Z_{it} = \bar{O}_P(1)$ means that $\max_{i \leq N, t \leq T} \|Z_{it}\| = O_P(1)$

of Lemma 8, it can be shown that:

$$\mathcal{R}(\tilde{\theta})_i = \bar{O}_P(1)\|\tilde{\lambda}_i - \lambda_{0i}\|^2 + \bar{O}_P(1/\sqrt{T})\|\tilde{\lambda}_i - \lambda_{0i}\| + \bar{O}_P(1/T) \tag{A.74}$$

for $i = 1, \ldots, N$, and

$$\mathcal{R}(\tilde{\theta})_{N+t} = \bar{O}_P(1)\|\tilde{f}_t - f_{0t}\|^2 + \bar{O}_P(1/\sqrt{T})\|\tilde{f}_t - f_{0t}\| + \bar{O}_P(1/T) \tag{A.75}$$

for $t = 1, \ldots, T$. Write $\mathcal{D}_{j,s}$ as the $r \times r$ matrix containing the $(j-1)r+1$ to jr rows and $(s-1)r+1$ to sr columns of $\mathcal{D}$. Note that Lemma 6 and Lemma 11 imply that $\|\mathcal{H}^{-1}\mathcal{S}^*(\theta_0)\|_{\max} = \bar{O}(h^m)$. Then, from (A.71) to (A.75) we can write

$$\tilde{f}_t - f_{0t} = (\Psi_{N,t})^{-1}\frac{1}{N}\sum_{j=1}^{N} \tilde{\varrho}_{jt}^{(1)}\lambda_{0j} + \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\mathcal{D}_{N+t,j} \cdot \tilde{\varrho}_{js}^{(1)} \cdot f_{0s} + \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\mathcal{D}_{N+t,N+s} \cdot \tilde{\varrho}_{js}^{(1)} \cdot \lambda_{0j}$$

$$+ (\Psi_{N,t})^{-1}\frac{1}{N}\sum_{j=1}^{N}\left\{\tilde{\varrho}^{(1)}(X_{jt} - \tilde{\lambda}_j'\tilde{f}_t)\tilde{\lambda}_j - \tilde{\varrho}_{jt}^{(1)}\lambda_{0j}\right\} + \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\mathcal{D}_{N+t,j}\left\{\tilde{\varrho}^{(1)}(X_{js} - \tilde{\lambda}_j'\tilde{f}_s)\tilde{f}_s - \tilde{\varrho}_{js}^{(1)}f_{0s}\right\}$$

$$+ \frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\mathcal{D}_{N+t,N+s}\left\{\tilde{\varrho}^{(1)}(X_{js} - \tilde{\lambda}_j'\tilde{f}_s)\tilde{\lambda}_j - \tilde{\varrho}_{js}^{(1)}\lambda_{0j}\right\}$$

$$- 0.5(\Psi_{N,t})^{-1}\mathcal{R}(\tilde{\theta})_{N+t} - 0.5\sum_{j=1}^{N}\mathcal{D}_{N+t,j}\mathcal{R}(\tilde{\theta})_j - 0.5\sum_{s=1}^{T}\mathcal{D}_{N+t,N+s}\mathcal{R}(\tilde{\theta})_{N+s} + \bar{O}(h^m). \tag{A.76}$$

Lemma 12. *Let $c_1, \ldots, c_T$ be a sequence of uniformly bounded constants. Then under Assumptions 1 and 2 we have:*

$$\frac{1}{T}\sum_{t=1}^{T}c_t(\tilde{f}_t - f_{0t}) = O_P\left(\frac{1}{Th}\right).$$

Proof. Define $d_j = \sqrt{NT}\cdot T^{-1}\sum_{t=1}^{T}c_t\mathcal{D}_{N+t,j}$ for $j = 1, \ldots, N+T$. Lemma 11 implies that $\max_{1\leq j\leq N+T}\|d_j\|$ is bounded. From (A.76), we have

$$\frac{1}{T}\sum_{t=1}^{T}c_t(\tilde{f}_t - f_{0t}) = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}c_t(\Psi_{N,t})^{-1}\tilde{\varrho}_{jt}^{(1)}\lambda_{0j} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}d_j\tilde{\varrho}_{js}^{(1)}f_{0s} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}d_{N+s}\tilde{\varrho}_{js}^{(1)}\lambda_{0j}$$

$$+ \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}c_t(\Psi_{N,t})^{-1}\left\{\tilde{\varrho}^{(1)}(X_{jt} - \tilde{\lambda}_j'\tilde{f}_t)\tilde{\lambda}_j - \tilde{\varrho}_{jt}^{(1)}\lambda_{0j}\right\} + \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}d_j\left\{\tilde{\varrho}^{(1)}(X_{js} - \tilde{\lambda}_j'\tilde{f}_s)\tilde{f}_s - \tilde{\varrho}_{js}^{(1)}f_{0s}\right\}$$

$$+ \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}d_{N+s}\left\{\tilde{\varrho}^{(1)}(X_{js} - \tilde{\lambda}_j'\tilde{f}_s)\tilde{\lambda}_j - \tilde{\varrho}_{js}^{(1)}\lambda_{0j}\right\} - 0.5\frac{1}{T}\sum_{t=1}^{T}c_t(\Psi_{N,t})^{-1}\mathcal{R}(\tilde{\theta})_{N+t}$$

$$- 0.5\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}d_j\mathcal{R}(\tilde{\theta})_j - 0.5\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}d_{N+s}\mathcal{R}(\tilde{\theta})_{N+s} + \bar{O}(h^m). \tag{A.77}$$

First, by Lyapunov's CLT, it is easy to see that the first three terms on the right of (A.77) are all $O_P(1/\sqrt{NT})$.

24

Next, it follows from Lemma 8, (A.74), (A.75) and Assumption 2(v) that the last four terms on the right of (A.77) are all $O_P(1/L_{NT}^2)$. Finally, we will show that the remaining three terms on the right of (A.77) are all $O_P(1/(Th))$, from which the desired result follows.

Define

$$\mathbb{V}_{NT}(\theta) = \frac{1}{NT} \sum_{j=1}^{N} \sum_{s=1}^{T} d_j \left\{ \tilde{\varrho}^{(1)}(X_{js} - \lambda_j' f_s) f_s - \tilde{\varrho}_{js}^{(1)} f_{0s} \right\},$$

and $\Delta_{NT}(\theta_a, \theta_b) = \sqrt{NT} h \left[\mathbb{V}_{NT}(\theta_a) - \mathbb{V}_{NT}(\theta_b) \right]$. Note that

$$\Delta_{NT}(\theta_a, \theta_b) = \underbrace{\frac{h}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{s=1}^{T} d_j \cdot \tilde{\varrho}^{(1)}(X_{js} - \lambda_{aj}' f_{as}) \cdot (f_{as} - f_{bs})}_{\Delta_{1,NT}(\theta_a, \theta_b)}$$

$$+ \underbrace{\frac{h}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{s=1}^{T} d_j \cdot \left[\tilde{\varrho}^{(1)}(X_{js} - \lambda_{aj}' f_{as}) - \tilde{\varrho}^{(1)}(X_{js} - \lambda_{bj}' f_{bs}) \right] \cdot f_{bs}}_{\Delta_{2,NT}(\theta_a, \theta_b)}.$$

Lemma 6 implies that

$$h \left| d_j \cdot \tilde{\varrho}^{(1)}(X_{js} - \lambda_{aj}' f_{as}) \cdot (f_{as} - f_{bs}) \right| \lesssim \left| \lambda_{aj}' f_{as} - \lambda_{bj}' f_{bs} \right|,$$

$$h \left| \varrho^{(1)}(X_{js} - \lambda_{aj}' f_{as}) - \varrho^{(1)}(X_{js} - \lambda_{bj}' f_{bs}) \right| \lesssim \left| \lambda_{aj}' f_{as} - \lambda_{bj}' f_{bs} \right|,$$

By Hoeffding's inequality and Lemma 2.2.1 of van der Vaart and Wellner (1996), we can show that for $d(\theta_a, \theta_b)$ sufficiently small,

$$\|\Delta_{1,NT}(\theta_a, \theta_b)\|_{\psi_2} \lesssim d(\theta_a, \theta_b), \quad \text{and} \quad \|\Delta_{2,NT}(\theta_a, \theta_b)\|_{\psi_2} \lesssim d(\theta_a, \theta_b).$$

Thus,

$$\|\Delta_{NT}(\theta_a, \theta_b)\|_{\psi_2} \leq \|\Delta_{1,NT}(\theta_a, \theta_b)\|_{\psi_2} + \|\Delta_{2,NT}(\theta_a, \theta_b)\|_{\psi_2} \lesssim d(\theta_a, \theta_b).$$

Therefore, similar to the proof of Lemma 3, we can show that for sufficiently small $\delta > 0$,

$$\mathbb{E} \left[\sup_{\theta \in \Theta^r(\delta)} |\mathbb{V}_{NT}(\theta)| \right] \lesssim \frac{\delta}{L_{NT} h}. \tag{A.78}$$

It then follows from (A.78) and Lemma 8 that $\mathbb{V}_{NT}(\tilde{\theta}) = O_P(1/(L_{NT}^2 h)) = O_P(1/(Th))$, e.g., the fifth term on the right of (A.77) is $O_P(1/(Th))$. Similar results can be obtained for the fourth and sixth terms on the right of (A.77), and the thus the desired result follows. $\qquad\square$

Lemma 13. *Under Assumptions 1 and 2, for each i we have*

$$\frac{1}{T} \sum_{t=1}^{T} \tilde{\varrho}_{it}^{(1)} (\tilde{f}_t - f_{0t}) = O_P\left(\frac{1}{Th}\right) \quad and \quad \frac{1}{T} \sum_{t=1}^{T} \tilde{\varrho}_{it}^{(2)} f_{0t}(\tilde{f}_t - f_{0t})' = O_P\left(\frac{1}{Th^2}\right).$$

25

Proof. To save space we only prove the second result, and proof of the first result is similar. Using (A.72) and (A.76) we can write

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}(\hat{f}_{t}-f_{0t})'$$

$$=\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\tilde{\varrho}_{it}^{(2)}\tilde{\varrho}_{jt}^{(1)}f_{0t}\lambda_{0j}'(\Psi_{N,t})^{-1}+\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\tilde{\varrho}_{it}^{(2)}f_{0t}\cdot\left\{\bar{\varrho}^{(1)}(X_{jt}-\tilde{\lambda}_{j}'\tilde{f}_{t})\tilde{\lambda}_{j}'-\tilde{\varrho}_{jt}^{(1)}\lambda_{0j}'\right\}(\Psi_{N,t})^{-1}$$

$$-\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}\tilde{f}_{s}'\mathcal{D}_{N+t,j}'\right)\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_{j}'\tilde{f}_{s})$$

$$-\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}\tilde{\lambda}_{j}'\mathcal{D}_{N+t,N+s}'\right)\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_{j}'\tilde{f}_{s})$$

$$+\frac{1}{2T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}\mathcal{R}(\tilde{\theta})_{N+t}'(\Psi_{N,t})^{-1}+\frac{1}{2T}\sum_{j=1}^{N}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}\mathcal{R}(\tilde{\theta})_{j}'\mathcal{D}_{N+t,j}'$$

$$+\frac{1}{2T}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}\mathcal{R}(\tilde{\theta})_{N+s}'\mathcal{D}_{N+t,N+s}'+O(h^{m-1}).\quad\text{(A.79)}$$

First, we can write

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\tilde{\varrho}_{it}^{(2)}\tilde{\varrho}_{jt}^{(1)}f_{0t}\lambda_{0j}'(\Psi_{N,t})^{-1}=\frac{1}{NT}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}\tilde{\varrho}_{it}^{(1)}f_{0t}\lambda_{0i}'(\Psi_{N,t})^{-1}+\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1,j\neq i}^{N}\tilde{\varrho}_{it}^{(2)}\tilde{\varrho}_{jt}^{(1)}f_{0t}\lambda_{0j}'(\Psi_{N,t})^{-1}.$$

Since $h\varrho_{it}^{(2)}(\cdot)$ is uniformly bounded by Lemma 6, $\max_{t\leq T}\|(\Psi_{N,t})^{-1}\|=O(1)$ for large N by Assumption 2, the first term on the RHS of the above equation is $O_P((Nh)^{-1})$. Using Lyapunov's CLT and Lemma 6, the second term on the RHS of the above equation can be shown to be $O_P((NTh)^{-1/2})$. Thus, the first term on the RHS of (A.79) is $O_P((Th)^{-1})$.

Second, consider the second term on the RHS of (A.79), which can be written as

$$O_P\left(\frac{1}{Nh}\right)+\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1,j\neq i}^{N}\tilde{\varrho}_{it}^{(2)}f_{0t}\cdot\left\{\bar{\varrho}^{(1)}(X_{jt}-\tilde{\lambda}_{j}'\tilde{f}_{t})\tilde{\lambda}_{j}'-\tilde{\varrho}_{jt}^{(1)}\lambda_{0j}'\right\}(\Psi_{N,t})^{-1}.$$

Similar to the proof of Lemma 12, the second term of the above expression can be shown to be $O_P(1/(Th^2))$. So the second term on the RHS of (A.79) is $O_P(1/(Th^2))$.

Next, for the third term on the RHS of (A.79), its p,qth element is given by

$$\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\chi_{i,j}\cdot\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_{j}'\tilde{f}_{s})\tilde{f}_{s}$$

where $\chi_{i,j} = T^{-1}\sum_{t=1}^{T}(\sqrt{NT}\mathcal{D}_{N+t,j,q})\cdot f_{0t,p}\tilde{\varrho}_{it}^{(2)}$, and $\mathcal{D}_{N+t,j,q}$ is the qth row of $\mathcal{D}_{N+t,j}$. Therefore,

$$\left\|\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\chi_{i,j}\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_j'\tilde{f}_s)\tilde{f}_s\right\| \leq \sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\chi_{i,j}\|^2}\cdot\sqrt{\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left[\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_j'\tilde{f}_s)\right]^2\|\tilde{f}_s\|^2}.$$

Since $|\sqrt{NT}f_{0t,p}\mathcal{D}_{N+t,j,q}|$ is uniformly is bounded by Lemma 11, it can be shown that $\mathbb{E}\|\chi_{i,j}\|^2 = O((Th)^{-1})$. Moreover

$$\left[\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_j'\tilde{f}_s)\right]^2 = \left[\bar{\varrho}_{it}^{(1)}+\bar{\varrho}^{(2)}(X_{js}-\lambda_j^{*'}f_s^*)\cdot(\lambda_{0i}'f_{0t}-\tilde{\lambda}_j'\tilde{f}_s)\right]^2 \lesssim O(m^2)+(\lambda_{0i}'f_{0t}-\tilde{\lambda}_j'\tilde{f}_s)^2,$$

thus

$$\sqrt{\frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\left[\bar{\varrho}^{(1)}(X_{js}-\tilde{\lambda}_j'\tilde{f}_s)\right]^2\|\tilde{f}_s\|^2} \lesssim d(\tilde{\theta},\theta_0) = O_P(1/L_{NT})$$

by Lemma 8. So, the third term on the RHS of (A.79) is $O_P(T^{-1}h^{-1/2})$. the fourth term on the RHS of (A.79) can be shown to be $O_P(T^{-1}h^{-1/2})$ in the same way.

Finally, it follows from Lemma 8 and (A.75) that the fifth term on the RHS of (A.79) is $O_P((L_{NT})^{-2}h^{-1}) = O_P((Th)^{-1})$. The p,qth element of the sixth term on the RHS of (A.79) can be written as $(2\sqrt{NT})^{-1}\sum_{j=1}^{N}\chi_{i,j}\mathcal{R}(\tilde{\theta})_j$, which is bounded by

$$\frac{\sqrt{N}}{2\sqrt{T}}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\chi_{i,j}\|^2}\sqrt{\frac{1}{N}\sum_{j=1}^{N}\|\mathcal{R}(\tilde{\theta})_j\|^2} = O_P((Th)^{-1/2})O_P(L_{NT}^{-2}) = O_P(T^{-3/2}h^{-1/2}).$$

The same bound for the seventh term on the RHS of (A.79) can be obtained using the same argument. Thus, combining the above results, we get

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}(\tilde{f}_t-f_{0t})' = O_P\left(\frac{1}{Th^2}\right).$$

$\square$

Proof of Theorem 4:

Proof. From the expansion in the proof of Lemma 10, we have

$$\Phi_{T,i}(\tilde{\lambda}_i-\lambda_{0i}) = -\frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}^{(1)}(X_{it}-\tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t + \frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}_{it}^{(1)}f_{0t} + \frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}_{it}^{(1)}(\tilde{f}_t-f_{0t})$$

$$-\frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}_{it}^{(2)}f_{0t}(\tilde{f}_t-f_{0t})'\lambda_{0i} + O_P\left(T^{-1}\|\tilde{F}-F_0\|^2\right) + o_P(\|\tilde{\lambda}_i-\lambda_{0i}\|).$$

It then follows from Lemma 6, Lemma 8 and Lemma 12 that

$$\Phi_{T,i}(\tilde{\lambda}_i - \lambda_{0i}) = -\frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}^{(1)}(X_{it} - \tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t + O(h^m) + O_P\left(\frac{1}{Th}\right) + o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|).$$

Note that

$$-\frac{1}{T}\sum_{t=1}^{T}\bar{\varrho}^{(1)}(X_{it} - \tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}^{(1)}(X_{it} - \tilde{\lambda}_i'\tilde{f}_t)\tilde{f}_t$$

$$= \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}\tilde{f}_t - \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}\cdot(\tilde{\lambda}_i'\tilde{f}_t - \lambda_{0i}'f_{0t})\tilde{f}_t + 0.5\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(3)}(*)(\tilde{\lambda}_i'\tilde{f}_t - \lambda_{0i}'f_{0t})^2\tilde{f}_t,$$

where $\tilde{\varrho}_{it}^{(3)}(*) = \tilde{\varrho}_{it}^{(3)}(c_{it}^*)$ and c_{it}^* is between $\lambda_{0i}'f_{0t}$ and $\tilde{\lambda}_i'\tilde{f}_t$.

First, by Lemma 13 we have

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}\tilde{f}_t = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}f_{0t} + \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}(\tilde{f}_t - f_{0t}) = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(1)}f_{0t} + O_P\left(\frac{1}{Th}\right).$$

Second,

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}\cdot(\tilde{\lambda}_i'\tilde{f}_t - \lambda_{0i}'f_{0t})\tilde{f}_t = \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}\tilde{f}_t\cdot(\tilde{f}_t - f_{0t})'\tilde{\lambda}_i + \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}\tilde{f}_t f_{0t}'\cdot(\tilde{\lambda}_i - \lambda_{0i})$$

$$= \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}\cdot(\tilde{f}_t - f_{0t})'\tilde{\lambda}_i + \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}(\tilde{f}_t - f_{0t})\cdot(\tilde{f}_t - f_{0t})'\tilde{\lambda}_i + \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}f_{0t}f_{0t}'\cdot(\tilde{\lambda}_i - \lambda_{0i})$$

$$+ \frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}(\tilde{f}_t - f_{0t})f_{0t}'\cdot(\tilde{\lambda}_i - \lambda_{0i}). \quad \text{(A.80)}$$

It then follows from Lemma 8 that the second term on the RHS of (A.80) is $O_P((Th)^{-1})$, and Lemma 13 implies that the first term is $O_P(T^{-1}h^{-2})$. It is easy to show that the last two terms on the right of (A.80) are both $o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|)$. Thus, we have

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(2)}\cdot(\tilde{\lambda}_i'\tilde{f}_t - \lambda_{0i}'f_{0t})\tilde{f}_t = O_P(T^{-1}h^{-2}) + o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|). \quad \text{(A.81)}$$

Next, it is also easy to show that

$$\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(3)}(*)(\tilde{\lambda}_i'\tilde{f}_t - \lambda_{0i}'f_{0t})^2\tilde{f}_t\right\| \le K_{19}\|\tilde{\lambda}_i - \lambda_{0i}\|^2\frac{1}{T}\sum_{t=1}^{T}|\tilde{\varrho}_{it}^{(3)}(*)| + \frac{K_{20}}{T}\sum_{t=1}^{T}|\tilde{\varrho}_{it}^{(3)}(*)|\cdot\|\tilde{f}_t - f_{0t}\|^2,$$

Therefore, from Lemma 6, Lemma 8 and Lemma 10 we have

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\varrho}_{it}^{(3)}(*)(\tilde{\lambda}_i'\tilde{f}_t - \lambda_{0i}'f_{0t})^2\tilde{f}_t = O_P(\|\tilde{\lambda}_i - \lambda_{0i}\|)\cdot O_P(T^{-1/2}h^{-3}) + O_P(T^{-1}h^{-2}),$$

which is $o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|) + O_P(T^{-1}h^{-2})$ under the assumption that $\sqrt{T}h^3 \to \infty$.

Finally, combining all the results above we get

$$\Phi_{T,i}(\tilde{\lambda}_i - \lambda_{0i}) = \frac{1}{T}\sum_{t=1}^T \tilde{\varrho}_{it}^{(1)} f_{0t} + o_P(\|\tilde{\lambda}_i - \lambda_{0i}\|) + O_P\left(\frac{1}{Th^2}\right) + O(h^m), \qquad \text{(A.82)}$$

and from Lemma 6 it is easy to show that:

$$\frac{1}{T}\sum_{t=1}^T \bar{\varrho}_{it}^{(2)} f_{0t} f_{0t}' \to \Phi_i > 0 \quad \text{and} \quad \frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{\varrho}_{it}^{(1)} f_{0t} \xrightarrow{d} \mathcal{N}(0, \tau(1-\tau)\mathbb{I}_r). \qquad \text{(A.83)}$$

Since our assumption implies that $\sqrt{T}h^2 \to \infty$ and $\sqrt{T}h^m \to 0$, the desired results follow from (A.82) and (A.83). $\qquad\square$

A.5   Estimating the Covariance Matrices

In this subsection we show that the estimators of the covariance matrices proposed in Remark 4.4 are consistent. Recall that

$$V_{f_t} = \tau(1-\tau)\Psi_t^{-1}\Sigma_\Lambda\Psi_t^{-1}, \quad \tilde{V}_{f_t} = \tau(1-\tau)\tilde{\Psi}_t^{-1}\tilde{\Sigma}_\Lambda\tilde{\Psi}_t^{-1},$$

where $\tilde{\Psi}_t = (Nb)^{-1}\sum_{i=1}^N l(\tilde{u}_{it}/b) \cdot \tilde{\lambda}_i \tilde{\lambda}_i'$ and $\tilde{\Sigma}_\Lambda = \tilde{\Lambda}'\tilde{\Lambda}/N$. It suffices to show that $\|\tilde{\Sigma}_\Lambda - \Sigma_\Lambda\| = o_P(1)$ and $\|\tilde{\Psi}_t - \Psi_t\| = o_P(1)$.

First, from Lemma 2 and Lemma 7,

$$
\begin{aligned}
\|\tilde{\Sigma}_\Lambda - \Sigma_\Lambda\| &\leq \left\|\tilde{\Lambda}'\tilde{\Lambda}/N - \Lambda_0'\Lambda_0/N + \Lambda_0'\Lambda_0/N - \Sigma_\Lambda\right\| \\
&\leq \left\|\tilde{\Lambda}'\tilde{\Lambda}/N - \Lambda_0'\Lambda_0/N\right\| + o(1) \\
&\leq 2 \cdot \frac{\|\Lambda_0\|}{\sqrt{N}} \cdot \frac{\|\tilde{\Lambda} - \Lambda_0\|}{\sqrt{N}} + \left(\frac{\|\tilde{\Lambda} - \Lambda_0\|}{\sqrt{N}}\right)^2 + o(1) = o_P(1).
\end{aligned}
$$

Second,

$$\tilde{\Psi}_t = \frac{1}{Nb}\sum_{i=1}^N l(u_{it}/b)\cdot\lambda_{0i}\lambda_{0i}' + \frac{1}{Nb}\sum_{i=1}^N [l(\tilde{u}_{it}/b) - l(u_{it}/b)]\cdot\lambda_{0i}\lambda_{0i}' + \frac{1}{Nb}\sum_{i=1}^N l(\tilde{u}_{it}/b)\cdot(\tilde{\lambda}_i\tilde{\lambda}_i' - \lambda_{0i}\lambda_{0i}'). \quad \text{(A.84)}$$

For the first term on the RHS of (A.84), following the standard proof of kernel density estimator, we have

$$\frac{1}{Nb}\sum_{i=1}^N l(u_{it}/b)\cdot\lambda_{0i}\lambda_{0i}' = \frac{1}{N}\sum_{i=1}^N \mathsf{f}_{it}(0)\lambda_{0i}\lambda_{0i}' + O_P(b^2) + O_P(1/\sqrt{Nb}).$$

For the third term on the RHS of (A.84), by Lemma 8 one can show that

$$\left\|\frac{1}{Nb}\sum_{i=1}^N l(\tilde{u}_{it}/b)\cdot(\tilde{\lambda}_i\tilde{\lambda}_i' - \lambda_{0i}\lambda_{0i}')\right\| \lesssim \frac{1}{Nb}\sum_{i=1}^N \left\|\tilde{\lambda}_i\tilde{\lambda}_i' - \lambda_{0i}\lambda_{0i}'\right\| = O_P(1/\sqrt{Nb^2}).$$

29

As regards the second term on the RHS of (A.84), we have

$$\frac{1}{Nb}\sum_{i=1}^{N}[l(\tilde{u}_{it}/b) - l(u_{it}/b)] \cdot \lambda_{0i}\lambda_{0i}'$$

$$= \frac{1}{Nb^2}\sum_{i=1}^{N} l^{(1)}(u_{it}/b) \cdot \lambda_{0i}\lambda_{0i}' \cdot (\tilde{u}_{it} - u_{it}) + \frac{1}{Nb^3}\sum_{i=1}^{N} l^{(2)}(u_{it}^*/b) \cdot \lambda_{0i}\lambda_{0i}' \cdot (\tilde{u}_{it} - u_{it})^2$$

$$= \frac{1}{Nb^2}\sum_{i=1}^{N} l^{(1)}(u_{it}/b) \cdot \lambda_{0i}\lambda_{0i}'\lambda_{0i} \cdot (\tilde{f}_t - f_{0t}) + \frac{1}{Nb^2}\sum_{i=1}^{N} l^{(1)}(u_{it}/b) \cdot \lambda_{0i}\lambda_{0i}'(\tilde{\lambda}_i - \lambda_{0i}) \cdot \tilde{f}_t$$

$$+ \frac{1}{Nb^3}\sum_{i=1}^{N} l^{(2)}(u_{it}^*/b) \cdot \lambda_{0i}\lambda_{0i}' \cdot (\tilde{u}_{it} - u_{it})^2.$$

It is easy to show that

$$\frac{1}{Nb^2}\sum_{i=1}^{N} l^{(1)}(u_{it}/b) \cdot \lambda_{0i}\lambda_{0i}'\lambda_{0i} = O_P(1),$$

$$\left\|\frac{1}{Nb^2}\sum_{i=1}^{N} l^{(1)}(u_{it}/b) \cdot \lambda_{0i}\lambda_{0i}'(\tilde{\lambda}_i - \lambda_{0i}) \cdot \tilde{f}_t\right\| \lesssim \frac{1}{b^{3/2}}\sqrt{\frac{1}{Nb}\sum_{i=1}^{N}\left[l^{(1)}(u_{it}/b)\right]^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i - \lambda_{0i}\|^2}$$

$$= O_P(b^{-3/2})O_P(1)O_P(L_{NT}^{-1}) = O_P(1/\sqrt{Nb^3}).$$

and

$$\left\|\frac{1}{Nb^3}\sum_{i=1}^{N} l^{(2)}(u_{it}^*/b) \cdot \lambda_{0i}\lambda_{0i}' \cdot (\tilde{u}_{it} - u_{it})^2\right\| \lesssim \frac{1}{b^3}\cdot\frac{1}{N}\sum_{i=1}^{N}(\tilde{u}_{it} - u_{it})^2 = O_P(1/Nb^3).$$

Thus, it follows that

$$\left\|\frac{1}{Nb}\sum_{i=1}^{N}[l(\tilde{u}_{it}/b) - l(u_{it}/b)] \cdot \lambda_{0i}\lambda_{0i}'\right\| = O_P(N^{-1}) + O_P(1/\sqrt{Nb^3}) + O_P(1/Nb^3).$$

Finally, it follows from the definition of Ψ_t and the condition $b \to 0, Nb^3 \to \infty$ that $\tilde{\Psi}_t \xrightarrow{p} \Psi_t$ and thus $\tilde{V}_{f_t} \xrightarrow{p} V_{f_t}$. The proof for $\tilde{\Phi}_i$ is similar and therefore is omitted.

References

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

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

Franklin, J. N. (2012). *Matrix theory*. Courier Corporation.

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

Golub, G. H. and C. F. Van Loan (2013). *Matrix Computations*, Volume 3. JHU Press.

Horowitz, J. (1998). Bootstrap methods for median regression models. *Econometrica 66*(6), 1327–1352.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes.* Springer, New York.