Estimation of Characteristics-based Quantile Factor Models *

Liang Chen[1], Juan J. Dolado[2], Jesús Gonzalo[3], and Haozi Pan[4]

[1] *HSBC Business School, Peking University, chenliang@phbs.pku.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*
[4] *CEMFI, haozi.pan@cemfi.edu.es*

June 20, 2024

Abstract

This paper studies the estimation of characteristics-based quantile factor models where the factor loadings are unknown functions of observed individual characteristics, while the idiosyncratic error terms are subject to conditional quantile restrictions. We propose a three-stage estimation procedure that is easily implementable in practice and has nice properties. The convergence rates, the limiting distributions of the estimated factors and loading functions, plus a consistent selection criterion for the number of factors at each quantile are derived under general conditions. The proposed estimation methodology is shown to work satisfactorily when: (i) the idiosyncratic errors have heavy tails, (ii) the time dimension of the panel dataset is not large, and (iii) the number of factors exceeds the number of characteristics. Finite sample simulations and an empirical application aimed at estimating the loading functions of the daily returns of a large panel of S&P500 index securities help illustrate these properties.

Keywords: quantile factor models, nonparametric quantile regression, principal component analysis.

JEL codes: C14, C31, C33.

*We are grateful to Yuan Liao for extremely useful comments. We also thank participants at the 2nd Workshop on High Dimensional Data Analysis, BGE Summer Forum 2023, IAAE Annual Conference, XIIth Workshop in Time Series Econometrics, Bristol University and Hunan University. Financial support from MICIN/ AEI/10.13039/501100011033, grants PID2019-104960GB-I00, PID2020-118659RB-I00, TED2021-129784BI00, CEX2021-001181-M, and Comunidad de Madrid, grants EPUC3M11 (V PRICIT) and H2019/HUM-5891, is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

The generalization of the classical factor analysis has historically relied on two main approaches. On the one hand, there has been the development of *approximate factor models* (**AFM**) where factors are assumed to be unobserved and therefore need to be jointly estimated with their loadings.[1] As a result, AFM suffer from a generic identification problem as both sets of objects can only be identified up to a rotation matrix. On the other hand, a growing literature has emerged in finance trying to explain the cross-sectional co-movements of stock returns on the basis of observable factors. In this setup, the factors are usually approximated using the differences between the returns of portfolios sorted by some observed characteristics, e.g. market capitalization and book-to-market ratio. This very popular approach, pioneered by Fama and French (1993), has been extended to include some additional factors, such as *momentum, profitability* and *investment*, besides the well-known Fama-French three factors (see Fama and French (2015)).

Both approaches have pros and cons. In effect, while the *latent factors* approach involves easily implementable estimation methods, such as the *principal component analysis* (**PCA**), it is often criticized for the lack of interpretation of the estimated factors. Conversely, the Fama-French approach turns out to be unambiguous about this interpretation; yet, their method of constructing the factor proxies quickly becomes unreliable for typical sample sizes as the number of factors grows (see Connor and Linton (2007)).

A setup that tries to exploit the advantages of both approaches while avoiding their shortcomings is the so-called *characteristics-based factor models* (**CFM**) introduced by Rosenberg (1974) and later extended by Connor and Linton (2007) and Connor, Hagmann, and Linton (2012). According to this procedure, the factor loadings are assumed to be smooth nonlinear functions of some observed characteristics of the different units, while the factors remain unobserved, as in AFM. Thereby, the latent factors in CFM can be easily estimated even when the number of factors is not small, whereas their interpretation hinges on the choice of the observed characteristics. Subsequently, Fan, Liao, and Wang (2016) have generalized this framework by allowing both the number of factors to differ from the number of observed characteristics and the factor loadings to be not fully explained by those characteristics. For the estimation of this kind of models, these authors introduced a new methodology called *projected principal component analysis* (**PPCA**), showing that such estimators exhibit faster rates of convergence than the conventional PCA estimators for AFM.

The main goal in this paper is to extend the analysis of Connor et al. (2012) and Fan

[1] AFM were first proposed by Chamberlain and Rothschild (1983) to characterize the co-movement of a large set of financial asset returns. The estimation and inference theory of these models and their subsequent extensions have been developed, *inter alia*, by Stock and Watson (2002), Bai and Ng (2002), and Bai (2003); Fan, Li, and Liao (2021) provide a recent overview of this line of research.

et al. (2016) to a new class of factor models labeled *characteristics-based quantile factor models* (**CQFM**). Note that, in relation to CFM, the main difference is that the idiosyncratic errors in CQFM are subject to quantile restrictions instead of mean restrictions. Moreover, in CQFM, the latent factors, the loading functions, and the number of factors are all allowed to vary across quantiles. Thus, this approach provides a more complete picture of how the joint distributions of many asset returns are driven by a few common risk factors.

In particular, our main contributions can be summarized as follows. First, we develop a new three-stage estimation method for CQFM, labeled *quantile-projected principal components analysis* (**QPPCA**), which is computationally simpler than the other available estimation procedures for this kind of factor models (see below). This procedure works as follows. First, at each time period, the outcomes (e.g. stock returns) are projected onto the space of the observed characteristics by means of sieve quantile regressions. Second, using the quantile fitted values from the first step, their factors and loadings are estimated by means of PCA. Third, the whole set of loading functions are retrieved by projecting the estimated loadings onto the basis of the sieve space. Finally, we propose a novel estimator for the number of factors at each quantile which is shown to perform satisfactorily for reasonable sample sizes.

The rates of convergence and the limiting distributions of the estimated factors and loading functions in QPPCA are derived under very general conditions, whereas the estimator for the number of factors at each quantile is also shown to be consistent. More specifically, a relevant contribution of QPPCA to this literature is that all its asymptotic properties are obtained without assuming any moment restrictions on the idiosyncratic errors. Thus, it becomes a nice tool to analyze data from financial markets, income inequality and climate change where the error distributions are known to exhibit heavy tails. Moreover, we only require the number of cross section observations (denoted as n) to diverge in our asymptotic analysis, while the number of time series observations (denoted as T) can be taken to be either fixed or diverging.

It is noteworthy that the QPPCA estimation approach shares several similarities with the PPCA approach of Fan et al. (2016). Yet, the main difference is that, given our less restrictive assumption on the error terms, sieve quantile regressions are implemented in the first step of QPPCA to project the observed outcomes, whereas sieve least square regressions are used in PPCA. Accordingly, QPPCA estimates turn out to be more robust to heavy tails and outliers, while the consistency of the PPCA estimators requires much stronger moment restrictions on the error terms. Foremost, a clear advantage of focusing on factor structures at the quantiles (QPPCA) rather than at the mean (PPCA) is that the former method will be able to estimate factors that the latter will not be able to obtain, like e.g. factors in the variance or in higher moments. In exchange for these advantages, our estimators present a trade-off in relation to PPCA: in principle, we have to assume that the factor loadings are fully explained by the observed characteristics. However, as explained below, this assumption can be later relaxed

using an alternative method that allows to test for omitted characteristics.

Another closely related paper to ours is Ma, Linton, and Gao (2021), which also addresses the estimation and inference for CQFM by considering a *semiparametric quantile factor analysis* (**SQFA**) approach, where observed characteristics are potentially allowed to affect stock returns in a nonlinear fashion. Thus, as in QPPCA, SQFA extends Connor et al. (2012) to the quantile restriction case. Yet, several important differences exist between the two approaches. First, as in Connor and Linton (2007) and Connor et al. (2012), SQFA assumes that the number of factors is known and is equal to the number of observed characteristics. In contrast, QPPCA does not only allow for the number of factors (which can vary across quantiles) to be different from the number of characteristics, but also implies that the number of factors at each quantile can be consistently estimated from the data. Second, while SQFA's initial estimator of the quantile loading functions also relies on sieve quantile regressions, its subsequent steps are based on an iterative minimization algorithm to jointly estimate the factors and loadings. As acknowledged by these authors, this algorithm can be computationally costly. However, this potential problem can be easily solved with the QPPCA methodology since its second and third steps are based on PCA, which are much easier to compute. Third, the asymptotic results of Ma et al. (2021) are obtained as $n, T \to \infty$, while all our results hold either T is fixed or $T \to \infty$ as $n \to \infty$. Finally, there are some differences in the assumptions imposed in these two approaches, which will be further discussed in the next sections once the main theoretical results are presented.

Next, it is important to highlight that the CQFM can be viewed as also being closely related to the quantile factor models (**QFM**) and its associated quantile factor analysis (**QFA**) estimators, proposed by Chen, Dolado, and Gonzalo (2021) to generalize AFM to quantile regressions. In effect, while no restrictions on the factor loadings are imposed in QFM (except for a standard rank condition), the loadings in CQFM become unknown functions of some observed characteristics, as in SQFA, entailing the risk of mis-specifying the choice of the relevant characteristics. Furthermore, in relation to the QFA estimators, an additional advantage of the CQFM setup is that the model can be consistently estimated even when T is fixed, while the QFA estimators turn out to be consistent only when both n and T go to infinity.

In traditional characteristics-based factor models, each factor loading is fully explained by an observed characteristic, which is also the assumption imposed in Ma et al. (2021). As mentioned above, an important extension of Fan et al. (2016) is that they allow a more general setup where the cross-sectional variations of the factor loadings may not be fully explained by the observed characteristics. In this paper, we also consider this general setup in the context of quantile factor models, and propose another estimation approach, labeled **QFA-Sieve**, relying on a combination of QFA and sieve regressions. This new approach allows us to heuristically test for the existence of other characteristics that explain the variations of the factor loadings when T is large (see Subsection 3.4 and the simulations in Subsection 4.3) and that are not considered

in our CQFM.

Lastly, we provide an empirical application of the proposed estimators to analyze the behavior of the risk factors and their loadings in a panel dataset of excess stock returns that has been used in other studies. Our main finding is that QPPCA allows to uncover substantial variations of the estimated loading functions across different quantiles, a feature which is ignored by using PPCA.

The outline of the rest of the paper is as follows. Section 2 introduces the model and the estimators. Section 3 derives their asymptotic properties, proposes a novel consistent estimator of the number of factors at each quantile, and presents a solution for the case where not all loadings are functions of observed characteristics. Section 4 provides several Monte Carlo simulation results for finite samples. Section 5 reports an empirical application of the proposed estimators. Finally, Section 6 concludes. An online appendix gathers detailed proofs of the theorems.

Notations: For any matrix $\boldsymbol{C}$, $\|\boldsymbol{C}\|$ and $\|\boldsymbol{C}\|_S$ denote the Frobenius norm and the spectral norm of $\boldsymbol{C}$, respectively; $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of $\boldsymbol{C}$, respectively, when the all eigenvalues are real; and $\boldsymbol{C} > 0$ signifies that $\boldsymbol{C}$ is a positive definite matrix. For two sequences of positive constants $\{a_1, \ldots, a_n, \ldots\}$ and $\{b_1, \ldots, b_n, \ldots\}$, $a_n \asymp b_n$ means that a_n/b_n is bounded below and above for all large n. The symbol $\lesssim$ means that the left side is bounded by a positive constant times the right side. Finally, for a random vector (Y, X), $\mathsf{Q}_\tau[Y|X = x]$ denotes the τ-quantile of Y given $X = x$.

Acronyms: Given the large number of acronyms used throughout the paper, we repeat them here broken down into two categories to facilitate the reading of the paper:

(i) Models: **AFM** (approximate factor models), **CFM** (characteristics-based factor models), **QFM** (quantile factor models), **CQFM** (characteristics-based quantile factor models).

(ii) Estimation methods: **PPCA** (projected principal component analysis), **QPPCA** (quantile-projected principal component analysis), **QFA** (quantile factor analysis), **SQFA** (semiparametric quantile factor analysis), **QFA-Sieve** (QFA combined with sieve estimation).

2 Model and Estimators

2.1 Model

For a panel of observed data $\{y_{it}\}_{1\leq i\leq n, 1\leq t\leq T}$, Chen et al. (2021) consider the following quantile factor model (QFM):

$$y_{it} = \boldsymbol{\lambda}_i'(\tau)\boldsymbol{f}_t(\tau) + u_{it}(\tau), \quad \tau \in (0,1), \tag{1}$$

where $\boldsymbol{\lambda}_i(\tau), \boldsymbol{f}_t(\tau) \in \mathbb{R}^R$ are quantile-dependent *unobserved* factor loadings and factors, respectively, R is the number of factors at quantile τ, and $u_{it}(\tau)$ is the idiosyncratic error satisfying $\mathsf{Q}_\tau[u_{it}(\tau)|\boldsymbol{\lambda}_i(\tau), \boldsymbol{f}_t(\tau)] = 0$. Note that the dependence of R on τ is suppressed to ease the notations.

Our focus in this paper is on the CQFM model of Ma et al. (2021), which nests the special case of the QFM in (1), where $\boldsymbol{\lambda}_i(\tau)$ are unrestricted. For simplicity, we deal with the case where all loadings are functions of the considered set of characteristics, and relegate the case of other types of loadings to Subsection 3.4 below. Specifically we assume the existence of a vector of *observed* characteristics $\boldsymbol{x}_i = (x_{i1}, x_{i2}, .., x_{iD}) \in \mathbb{R}^D$ for unit i such that

$$\boldsymbol{\lambda}_i(\tau) = \boldsymbol{g}_\tau(\boldsymbol{x}_i), \tag{2}$$

where $\boldsymbol{g}_\tau(\cdot) : \mathbb{R}^D \mapsto \mathbb{R}^R$ is a vector of unknown functions for each τ. As in Fan et al. (2016), we suppose that the rth element of $\boldsymbol{g}_\tau(\boldsymbol{x}_i)$ is given by the following additive function

$$g_{\tau,r}(\boldsymbol{x}_i) = \sum_{d=1}^{D} g_{\tau,rd}(x_{id}),$$

where $g_{\tau,r1}, \ldots, g_{\tau,rD}$ are unknown functions. Following the related literature, it is assumed that $g_{\tau,r}$ is time-invariant so that the loadings capture the cross-sectional heterogeneity only. As Fan et al. (2016) argue, such a specification is not stringent since in many factor-model applications to stationary time series, the analysis is carried out within each fixed time window with either a fixed or slowly-growing T. Yet, even if there are individual time-varying characteristics, like e.g. firm size or firm age, following these authors we expect the conclusions in the current paper to remain valid if some smoothness assumptions are added for the time-varying components of those covariates. For example, in our empirical application in Section 5 below, the dependent variable is daily stock returns, while the characteristics are only available every quarter. Therefore, to study the factor structure of the stock returns in a given quarter, as in Fan et al. (2016), the characteristics should be treated as time invariant.[2].

Let $\boldsymbol{Y}$ be the $n \times T$ matrix of y_{it}, $\boldsymbol{F}_\tau$ be the $T \times R$ matrix of $\boldsymbol{f}_t(\tau)$, $\boldsymbol{X}$ be the $n \times D$ matrix of $\boldsymbol{x}_i$, $\boldsymbol{G}_\tau(\boldsymbol{X})$ be the $n \times R$ matrix of $\boldsymbol{g}_\tau(\boldsymbol{x}_i)$, $\boldsymbol{U}_\tau$ be the $n \times T$ matrix of $u_{it}(\tau)$. Then, models (1) and (2) can be rewritten in compact matrix form as:

$$\boldsymbol{Y} = \boldsymbol{G}_\tau(\boldsymbol{X})\boldsymbol{F}_\tau' + \boldsymbol{U}_\tau. \tag{3}$$

As already highlighted, the above setup is more general than those considered in the models of Connor et al. (2012) and Ma et al. (2021). In the latter, the dimension of the vector of

[2]This reasoning echoes the one used for standard AFM models where loadings only have cross-sectional variation while factors are time varying

characteristics is required to be equal to the number of factors ($D = R$), while each of the loading functions is assumed to be linked to only one of the observed characteristics, i.e., $g_{\tau,r}(\boldsymbol{x}_i) = g_{\tau,r}(x_{ir})$ for $r = 1, \ldots, R$. Note that these assumptions facilitate the interpretation of the estimated factors, e.g. the first estimated factor would be the value factor, the second one would be the momentum factor, and so on. However, both conditions also could be restrictive in other setups. For example, if y_{it} represents the profit flow of firm i at time t and there are two factors capturing, say, a monetary shock and a fiscal shock, then it seems more reasonable to allow for dependence of the response of the firm's profit to the macro shocks on a wide range of firm characteristics — such as size, leverage, growth, etc. — that exceeds the number of factors. Moreover, a drawback of the two above-mentioned approaches is that they are more difficult to estimate and therefore require algorithms involving multiple iterations, particularly when the number of characteristics is large (say there are tens of characteristics, then there will be tens of factors). In contrast, in our setup, it is potentially easier to generalize CQFM to allow for high dimensional characteristics since the number of factors can be much smaller than the number of characteristics. Lastly, it should be noted that, in the previous estimation methods, one needs to assume that the number of factors R is a priori known, while in this paper we propose a method that consistently estimates R from the data at each quantile (see Section 3.3 below).

Relative to the semiparametric factor models considered by Fan et al. (2016), the most salient difference is that the idiosyncratic errors in CQFM are subject to conditional quantile restrictions, rather than to conditional mean restrictions. From this perspective, as pointed out in Chen et al. (2021), the QFM framework allows to recover different factor structures (including the factors, the loadings, the number of factors) across different quantiles, even when the distribution of the idiosyncratic errors exhibits heavy tails. Hence, CQFM becomes a useful tool to analyze the co-movement of the financial market variables, where the correlation of the tail risks between different assets becomes the main object of interest.

2.2 Estimators

To simplify the notations even further, in the rest of the paper we suppress the τ-subscripts in the model and use $\boldsymbol{g}(), \boldsymbol{G}(\cdot), \boldsymbol{F}, \boldsymbol{U}$ instead of $\boldsymbol{g}_\tau(\cdot), \boldsymbol{G}_\tau(\cdot), \boldsymbol{F}_\tau, \boldsymbol{U}_\tau$.

Write $\theta_{0t}(\boldsymbol{x}_i) = \boldsymbol{g}(\boldsymbol{x}_i)'\boldsymbol{f}_t = \sum_{r=1}^{R} g_r(\boldsymbol{x}_i) f_{tr} = \sum_{r=1}^{R}(\sum_{d=1}^{D} g_{rd}(x_{id})) f_{tr}$. Let Θ be a space of continuous functions such that $\theta_{0t} \in \Theta$ for all $t = 1, \ldots, T$, while $\{\Theta_n\}$ is a sequence of sieve spaces approximating Θ. In particular, let us consider the following finite dimensional linear spaces:

$$\Theta_n = \left\{ h : \mathcal{X} \mapsto \mathbb{R}, \quad h(\boldsymbol{x}) = \sum_{d=1}^{D} \sum_{j=1}^{k_n} a_{jd} \phi_j(x_d) : (a_{11}, \ldots, a_{jd}, \ldots, a_{k_n D}) \in \mathbb{R}^{D k_n} \right\},$$

where $\mathcal{X} \subset \mathbb{R}^D$ is the support of $\boldsymbol{x}_i$, and $\phi_1, \ldots, \phi_{k_n}$ is a set of continuous basis functions. Write

$$\underbrace{\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)}_{Dk_n \times 1} = [\phi_1(x_{i1}), \ldots, \phi_{k_n}(x_{i1}), \ldots \phi_1(x_{id}), \ldots, \phi_{k_n}(x_{id}), \ldots, \phi_1(x_{iD}), \ldots, \phi_{k_n}(x_{iD})]'.$$

Suppose that for $r = 1, \ldots, R$, there exist $\boldsymbol{b}_{01}, \ldots, \boldsymbol{b}_{0R} \in \mathbb{R}^{Dk_n}$ such that for some constant $\alpha > 0$,

$$\max_{1 \le r \le R} \sup_{\boldsymbol{x} \in \mathcal{X}} \left| g_r(\boldsymbol{x}) - \boldsymbol{b}_{0r}' \boldsymbol{\phi}_{k_n}(\boldsymbol{x}) \right| = O(k_n^{-\alpha}). \tag{4}$$

Then, for $\boldsymbol{B}_0 = (\boldsymbol{b}_{01}, \ldots, \boldsymbol{b}_{0R}) \in \mathbb{R}^{Dk_n \times R}$, $\boldsymbol{a}_{0t} = \boldsymbol{B}_0 \boldsymbol{f}_t$ and $\pi_n \theta_{0t}(\cdot) = \boldsymbol{a}_{0t}' \boldsymbol{\phi}_{k_n}(\cdot)$, we have $\pi_n \theta_{0t} \in \Theta_n$ for all t and

$$\max_{1 \le t \le T} \sup_{\boldsymbol{x} \in \mathcal{X}} |\pi_n \theta_{0t}(\boldsymbol{x}) - \theta_{0t}(\boldsymbol{x})| = O(k_n^{-\alpha}). \tag{5}$$

Once the definitions above have been established, the next stage is to introduce our QPPCA estimation method which consists of the following three steps.

Step 1: Obtain the sieve estimator of θ_{0t}. Let $\rho_\tau(u) = (\tau - \mathbf{1}\{u \le 0\})u$ be the check function, and define $l(\theta, y_{it}, \boldsymbol{x}_i) = \rho_\tau(y_{it} - \theta(\boldsymbol{x}_i)) - \rho_\tau(y_{it} - \theta_{0t}(\boldsymbol{x}_i))$, $L_n(\theta) = n^{-1} \sum_{i=1}^n l(\theta, y_{it}, \boldsymbol{x}_i)$. Then the sieve estimator $\hat{\theta}_{nt}$ is defined by

$$L_n(\hat{\theta}_{nt}) \le \inf_{\theta \in \Theta_n} L_n(\theta).$$

In practice, $\hat{\theta}_{nt}$ can be obtained by means of a simple parametric quantile regression as follows:

$$\hat{\boldsymbol{a}}_t = \arg\min_{\boldsymbol{a} \in \mathbb{R}^{Dk_n}} \sum_{i=1}^n \rho_\tau \left(y_{it} - \boldsymbol{a}' \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i) \right) \quad \text{and} \quad \hat{\theta}_{nt}(\cdot) = \hat{\boldsymbol{a}}_t' \boldsymbol{\phi}_{k_n}(\cdot).$$

Step 2: Write $\hat{y}_{it} = \hat{\theta}_{nt}(\boldsymbol{x}_i) = \hat{\boldsymbol{a}}_t' \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ and let $\hat{\boldsymbol{Y}}$ be the $n \times T$ matrix of $\hat{y}_{it}$. Then, the estimator of $\boldsymbol{F}$, denoted as $\hat{\boldsymbol{F}}$, is the matrix of eigenvectors (multiplied by $\sqrt{T}$) associated with R largest eigenvalues of the $T \times T$ matrix $\hat{\boldsymbol{Y}}' \hat{\boldsymbol{Y}}$. Moreover, the estimator of the characteristics-based loading matrix $\boldsymbol{G}(\boldsymbol{X})$ is given by $\hat{\boldsymbol{G}}(\boldsymbol{X}) = \hat{\boldsymbol{Y}} \hat{\boldsymbol{F}}/T$. It is well known that these estimators are the ones that minimize the objective function: $L_{nT}(\boldsymbol{G}(\boldsymbol{X}), \boldsymbol{F}) = \|\hat{\boldsymbol{Y}} - \boldsymbol{G}(\boldsymbol{X})' \boldsymbol{F}\|^2$, subject to the standard normalization, namely, $\boldsymbol{F}'\boldsymbol{F}/T = \boldsymbol{I}_R$ and $\boldsymbol{G}(\boldsymbol{X})'\boldsymbol{G}(\boldsymbol{X})/n$ is diagonal (see Stock and Watson (2002)).[3]

Step 3: Estimate the factor loading functions: $g_r(\cdot)$ for $r = 1, \ldots, R$. Define $\hat{\boldsymbol{A}} = (\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_T)$.

[3]Note, however, that the estimator is invariant to the rotation transformations of the sieve bases.

Then $\boldsymbol{B}_0$ can be estimated as

$$\hat{\boldsymbol{B}} = \hat{\boldsymbol{A}}\hat{\boldsymbol{F}}/T. \tag{6}$$

Lastly, the estimator of $\boldsymbol{g}(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathcal{X}$ is given by $\hat{\boldsymbol{g}}(\boldsymbol{x})' = \boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\hat{\boldsymbol{B}}$.

The insight for the 3-step estimation procedure is as follows. First, note that by (4),

$$\boldsymbol{Y} = \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}' + \boldsymbol{U} \approx \boldsymbol{\Phi}\boldsymbol{B}_0\boldsymbol{F}' + \boldsymbol{U} = \boldsymbol{\Phi}\boldsymbol{A}_0 + \boldsymbol{U}$$

with $\boldsymbol{A}_0 = (\boldsymbol{a}_{01}, \ldots, \boldsymbol{a}_{0T}) = \boldsymbol{B}_0\boldsymbol{F}'$. The purpose of the first-step quantile projection is to purge the idiosyncratic errors $\boldsymbol{U}$, so that the matrix of fitted value $\hat{\boldsymbol{Y}} = \boldsymbol{\Phi}\hat{\boldsymbol{A}} \approx \boldsymbol{\Phi}\boldsymbol{A}_0 = \boldsymbol{\Phi}\boldsymbol{B}_0\boldsymbol{F}' \approx \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'$ has a low-rank structure. Second, it is easy to see that this low-rank structure is essentially a factor model with vanishing error terms (i.e., the estimation errors from the first step), which motivates the second-step PCA estimation of $\boldsymbol{G}(\boldsymbol{X})$ and $\boldsymbol{F}$ using $\hat{\boldsymbol{Y}}$. Third, since $\hat{\boldsymbol{A}} \approx \boldsymbol{A}_0 = \boldsymbol{B}_0\boldsymbol{F}' \approx \boldsymbol{B}_0\hat{\boldsymbol{F}}'$, postmultiplying both sides by $\hat{\boldsymbol{F}}$ yields the estimator of $\boldsymbol{B}_0$ under the normalization $\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}}/T = \boldsymbol{I}_R$. Finally, given $\hat{\boldsymbol{B}}$ and (4), the estimator of $\boldsymbol{g}(\boldsymbol{x})$ follows naturally because it is approximated by $\boldsymbol{\phi}_{k_n}(\boldsymbol{x})'\boldsymbol{B}_0$ as k_n diverges.

Remark 1. *The main difference between this three-stage estimation approach and the PPCA of Fan et al. (2016) is how we project $\boldsymbol{y}_t$ onto the space of $\boldsymbol{X}$ in the first step, namely, how $\boldsymbol{a}_{01}, \ldots, \boldsymbol{a}_{0T}$ are estimated. The use of sieve quantile regressions instead of the least squares projections is a natural choice given that the idiosyncratic errors in CQFM are subject to conditional quantile restrictions. When the distributions of the errors are symmetric around 0, the QPPCA estimators at $\tau = 0.5$ can be viewed as a robust version of the PPCA estimators since the consistency of the QPPCA estimators does not rely on moment restrictions of the errors (see Theorem 1 below).*

Remark 2. *The SQFA estimation method advocated by Ma et al. (2021) chooses $\boldsymbol{B}$ and $\boldsymbol{F}$ in an iterative fashion to minimize the following objective function:*

$$L_{nT}(\boldsymbol{B}, \boldsymbol{F}) = \sum_{i=1}^{n}\sum_{t=1}^{T} \rho_\tau(y_{it} - \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)'\boldsymbol{B}\boldsymbol{f}_t),$$

*while the quantile factor analysis (**QFA**) proposed by Chen et al. (2021) for QFM relies on a similar approach which estimates the factor loadings $\boldsymbol{G}(\boldsymbol{X})$ and $\boldsymbol{F}$ jointly. Accordingly, both SQFA and QFA require n and T go to infinity in order to establish the consistency of the estimators. By contrast, as will be shown in the next section, the consistency of the QPPCA estimators can be established either when T is fixed or T goes to infinity along with n.*

3 Asymptotic Properties of the Estimators

In this section, we derive the rates of convergence and the asymptotic distributions of the QPPCA estimators. To simplify the discussion, the number of factors is taken to be known in the first two subsections while this assumption is relaxed in the last subsection where a consistent estimator of R is introduced.

As in Chen et al. (2021) and Ma et al. (2021), the quantile factors are treated as non-random constants in the asymptotic analysis. Hence, the conditional quantile restrictions on the idiosyncratic errors imply that

$$P[u_{it} \leq 0 | \boldsymbol{x}_i = \boldsymbol{x}] = \tau \text{ for any } \boldsymbol{x} \in \mathcal{X}. \tag{7}$$

Alternatively, all the assumptions and results to be presented below could be understood as being conditional on the realizations of the factors.

Lastly, note all the results to be presented below hold either when: (i) T is fixed and $n \to \infty$, or (ii) $n, T \to \infty$. The first case is also called the *high-dimension-low-sample-size* setup in the statistics literature (see Shen, Shen, and Marron (2013) and Jung and Marron (2009)). One of the main insights of Fan et al. (2016) is that dimensionality is a blessing rather than a curse in the context of CFM, implying that their PPCA estimators are consistent even when T is fixed. Our results below extend the finite-T-consistency results of Fan et al. (2016) to CQFM.

3.1 Rates of convergence

Suppose that the observed data $\{y_{it}\}$ are generated by (3) and that $\{u_{it}\}$ satisfy (7). Let

$$\varepsilon_n = \sqrt{k_n/n} \vee k_n^{-\alpha} \quad \text{and} \quad \varepsilon_{nT} = \sqrt{\ln T} \cdot \varepsilon_n.$$

For any $\theta_1, \theta_2 \in \Theta$, define the pseudo-metric $d(\theta_1, \theta_2) \equiv \sqrt{\mathbb{E}\left(\theta_1(\boldsymbol{x}_i) - \theta_2(\boldsymbol{x}_i)\right)^2}$. The following set of conditions are required to establish the uniform rate of convergence of $\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nT}$, which is a crucial result to prove the theorems in the sequel.

Assumption 1. *Let M be a generic bounded constant.*
(i) Define $\boldsymbol{z}_i = (u_{i1}, \ldots, u_{iT}, \boldsymbol{x}_i)$. Then, $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ are i.i.d. Moreover, the distributions of $(u_{i1}, \boldsymbol{x}_i), \ldots, (u_{iT}, \boldsymbol{x}_i)$ are identical for each i.
(ii) Equation (4) holds for some $\alpha \geq 1$.
(iii) $\mathcal{X} \subset \mathbb{R}^D$ is bounded, and $\sup_{\theta \in \Theta} \sup_{\boldsymbol{x} \in \mathcal{X}} |\theta(\boldsymbol{x})| < M$. $\|\boldsymbol{f}_t\| < M$ for all $t = 1, \ldots, T$.
(iv) The conditional density of u_{it} given $\boldsymbol{x}_i = \boldsymbol{x}$, denoted as $\mathsf{f}(\cdot|\boldsymbol{x})$, satisfies: $0 < \inf_{\mathcal{X}} \mathsf{f}(0|\boldsymbol{x}) \leq \sup_{\mathcal{X}} \mathsf{f}(0|\boldsymbol{x}) < \infty$ and $\sup_{\mathcal{X}} |\mathsf{f}(z|\boldsymbol{x}) - \mathsf{f}(0|\boldsymbol{x})| \to 0$ as $|z| \to 0$.
(v) As $n \to \infty$, $k_n \to \infty$ and $\varepsilon_{nT} \to 0$.

Assumption 1(i) can be relaxed to allow for weak cross-sectional dependence though it excludes e.g. nonstationary error terms — see Remark 3 below for the details. Assumption 1(ii) is a general condition on the sieve approximations that can be easily verified using more primitive conditions. For instance, it holds if Θ is an $\alpha-$smooth Hölder space (see Chen (2007) for further examples). Assumption 1(iii) and Assumption 1(iv) are also standard in sieve quantile regressions, noting that the latter imposes very mild restrictions on the size of T when it goes to infinity jointly with n.

Proposition 1. *Under Assumption 1, it holds that* $\max_{1\leq t\leq T} d(\hat{\theta}_{nt}, \theta_{0t}) = O_P(\varepsilon_{nT})$ *when either* T *is fixed or* $T \to \infty$ *as* $n \to \infty$.

Remark 3. *The proof of Proposition 1 is based on Corollary 1 of Chen and Shen (1998). In particular, we show that*

$$P\left[\max_t d(\hat{\theta}_{nt}, \theta_{0t}) \geq C\varepsilon_{nT}\right] \leq \sum_{t=1}^{T} P\left[d(\hat{\theta}_{nt}, \theta_{0t}) \geq C\varepsilon_{nT}\right] \leq c_1 \exp\left\{C^2 \ln T(1 - c_2 n\varepsilon_n^2)\right\}$$

for any $C \geq 1$ *and some constants* c_1, c_2. *Moreover, as shown in Chen and Shen (1998), the above inequality holds when the observations are generated from a stationary uniform (*ϕ*-) mixing sequence with* $\phi(j) \lesssim j^{-\zeta}$ *for some* $\zeta > 1$. *Thus, in line with Connor and Korajczyk (1993), Lee and Robinson (2016) and Ma et al. (2021), one can assume the existence of a reordering of the cross-sectional units such that their dependence can be characterized by the uniform mixing condition mentioned above, and the conclusion of Proposition 1 will still hold.*

To establish the convergence rates of the estimated factors and loading functions, some further assumptions are required.

Assumption 2. *Let* M *be a generic bounded constant.*
(i) Let $\boldsymbol{\Sigma}_\phi = \mathbb{E}[\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']$. *Then, there exist constants* c_1, c_2 *such that* $0 < c_1 \leq \lambda_{\min}(\boldsymbol{\Sigma}_\phi) \leq \lambda_{\max}(\boldsymbol{\Sigma}_\phi) \leq c_2 < \infty$ *for all* n.
(ii) $k_n^2/n \to 0$ *as* $n \to \infty$.
(iii) There exist a constant $c > 0$ *such that* $\lambda_{\min}(\boldsymbol{F}'\boldsymbol{F}/T) > c$ *for all* T.
(iv) $\hat{\boldsymbol{\Sigma}}_g \equiv n^{-1}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{x}_i)\boldsymbol{g}(\boldsymbol{x}_i)' \xrightarrow{P} \boldsymbol{\Sigma}_g > 0$ *as* $n \to \infty$.
(v) The eigenvalues of $\boldsymbol{\Sigma}_g \cdot \boldsymbol{F}'\boldsymbol{F}/T$ *are distinct.*

The conditions in Assumption 2 are all standard in the literature on factor models and sieve estimation. In particular, Assumption 2(ii) strengthens Assumption 1(v), and Assumption 2(iii) implicitly requires that $T \geq R$. In comparison, Assumption A0 of Ma et al. (2021) imposes that $\liminf_{T\to\infty} |T^{-1}\sum_{t=1}^T f_{tr}| > 0$ for all $r = 1, \ldots, R$, which excludes the possibility that the underlying time series generating $\boldsymbol{F}$ has zero mean. The following theorem gives the rates of convergence of the estimated factors and loading functions.

Theorem 1. *Let $\hat{\boldsymbol{\Omega}}$ be the diagonal matrix whose elements are the eigenvalues of $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}/(nT)$, and define $\hat{\boldsymbol{H}} = \hat{\boldsymbol{\Sigma}}_g(\boldsymbol{F}'\hat{\boldsymbol{F}}/T)\hat{\boldsymbol{\Omega}}^{-1}$. Then, under Assumptions 1 and 2, the following results hold either when T is fixed or $T \to \infty$ as $n \to \infty$,:*
(i) $\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T} = O_P(\varepsilon_{nT})$.
(ii) $\|\hat{\boldsymbol{G}}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X})(\hat{\boldsymbol{H}}')^{-1}\|/\sqrt{n} = O_P(\varepsilon_{nT})$.
(iii) $\sup_{\boldsymbol{x}\in\mathcal{X}} \|\hat{\boldsymbol{g}}(\boldsymbol{x}) - \hat{\boldsymbol{H}}^{-1}\boldsymbol{g}(\boldsymbol{x})\| = O_P(\sqrt{k_n}\varepsilon_{nT})$.

A few remarks on this result are relevant. First, it is worth highlighting that Theorem 1 (and Theorem 2 below) holds without requiring any restrictions on the time-series dependence of the idiosyncratic errors, while both Fan et al. (2016) and Ma et al. (2021) impose some kind of weak-time-series-dependence conditions. Second, while our setup does not require any moment restrictions on u_{it}, Assumption 3.4(iv) of Fan et al. (2016) needs the error terms to have exponential tails. This is one of the main advantages of using quantile projections. Third, we allow the error terms $\{u_{it}, t = 1, \ldots, T\}$ and the characteristics $\boldsymbol{x}_i$ to be dependent while Assumption 3.4(i) of Fan et al. (2016) requires them to be independent. We believe that, as in the case of omitted covariates in standard regression models, the assumption of dependence is more realistic than the one of independence. As a consequence, the convergence rates given in Theorem 1 are generally slower than those of Fan et al. (2016). In Theorem 2 below, we show that the convergence rates of estimated factors using QPPCA and PPCA are identical if $\boldsymbol{x}_i$ is assumed to be independent of $(u_{i1}, \ldots, u_{iT})$.

Assumption 3. *Let L be a generic bounded constant and let $\mathsf{f}(\cdot)$ denote the p.d.f. of u_{it}.*
(i) For each i, $\boldsymbol{x}_i$ is independent of $(u_{i1}, \ldots, u_{iT})$.
(ii) $|\mathsf{f}(c) - \mathsf{f}(0)| \leq L|c|$ for any c in a neighborhood of 0.
(iii) Equation (4) holds for some $\alpha \geq 3$.

Assumption 3(i) essentially requires that the observed characteristics only affect the location but not the scale of the distributions of y_{it}. In such a case, the leading term in the Bahadur representation of $\hat{\boldsymbol{a}}_t$ has a similar structure to the least square estimators (see Lemma 2 in the online appendix). Thus, this assumption implies an improved convergence rate of $\hat{\boldsymbol{F}}$ which happens to be as fast as the rate of the PPCA estimators (see Theorem 4.1 of Fan et al. (2016)).

Theorem 2. *Let $\eta_{nT} = \sqrt{\ln(k_n^{-1/4}\varepsilon_{nT}^{-1/2})} \cdot k_n^{5/4}\varepsilon_{nT}^{1/2}n^{-1/2}$. Under Assumptions 1 to 3, we have*

$$\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T} = O_P\left(n^{-1/2} \vee k_n^{-\alpha} \vee \eta_{nT} \vee \varepsilon_{nT}^2\right).$$

Moreover, if $T \asymp n^{\gamma_1}$ and $k_n \asymp n^{1/(6+\gamma_2)}$ for some $\gamma_1 \geq 0$ and $\gamma_2 > 0$, then

$$\|\hat{\boldsymbol{F}} - \boldsymbol{F}\hat{\boldsymbol{H}}\|/\sqrt{T} = O_P\left(n^{-1/2} \vee k_n^{-\alpha}\right).$$

Remark 4. *The term η_{nT} in Theorem 2 represents the higher-order terms in the Bahadur representation of $\hat{\boldsymbol{a}}_t$. When α is large, η_{nT} is approximately equal to $k_n^{3/2}n^{-3/4}$. Note that this slightly unusual expression of η_{nT} is mainly due to the non-smoothness of the check function. Indeed, similar terms can be found in Theorem 2 of Horowitz and Lee (2005), Theorem 3.2 of Kato et al. (2012) and Theorem 2 of Ma et al. (2021).*

3.2 Asymptotic distribution

Define $\boldsymbol{\Sigma}_{\mathsf{f}\phi} = \mathbb{E}[\mathsf{f}(0|\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']$ and $\sigma_{k_n}^2 = \boldsymbol{\phi}'_{k_n}(\boldsymbol{x})\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Sigma}_{\phi}\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x})$.

Assumption 4. *Let L be a generic bounded constant.*
(i) $u_{i1}, \ldots, u_{iT}$ are independent conditional on $\boldsymbol{x}_i$.
(ii) $|\mathsf{f}(c|\boldsymbol{x}) - \mathsf{f}(0|\boldsymbol{x})| \leq L|c|$ for any c in a neighborhood of 0 and any $\boldsymbol{x} \in \mathcal{X}$.
(iii) There exist constants c_1, c_2 such that $0 < c_1 \leq \lambda_{\min}(\boldsymbol{\Sigma}_{\mathsf{f}\phi}) \leq \lambda_{\max}(\boldsymbol{\Sigma}_{\mathsf{f}\phi}) \leq c_2 < \infty$ for all k_n.
(iv) $(nT)^{1/2}k_n^{1/2-\alpha}\sigma_{k_n}^{-1} = o(1)$ and $(nT)^{1/2}k_n^{1/2}\eta_{nT}\sigma_{k_n}^{-1} = o(1)$.

Assumption 4(i) is adopted for simplicity; yet it could be replaced by $\beta-$mixing conditions at the cost of getting more complex asymptotic covariance matrices. When $\sigma_{k_n} \asymp k_n^{1/2}$ and T is fixed, Assumption 4(iv) essentially requires that $n^{1/2}k_n^{-\alpha} = o(1)$ and $n^{1/2}\eta_{nT} = o(1)$, or $k_n^6 \ll n \ll k_n^{2\alpha}$. As a result, we need (4) to hold with $\alpha > 3$. The remaining conditions in Assumption 4 are standard (see, e.g. Assumptions 3 and 5 of Horowitz and Lee (2005)). We are now in the position of establishing the asymptotic distribution of the estimated loading functions.

Theorem 3. *Under Assumptions 1, 2, and 4, it holds that for any $\boldsymbol{x} \in \mathcal{X}$*

$$\boldsymbol{\Sigma}_{T,\tau}^{-1/2}(\hat{\boldsymbol{H}}')^{-1} \cdot \frac{\sqrt{nT}}{\sigma_{k_n}} \left(\hat{\boldsymbol{g}}(\boldsymbol{x}) - (\boldsymbol{F}'\hat{\boldsymbol{F}}/T)'\boldsymbol{g}(\boldsymbol{x})\right) \xrightarrow{d} N(0, \boldsymbol{I}_R),$$

where $\boldsymbol{\Sigma}_{T,\tau} = \tau(1-\tau)(\boldsymbol{F}'\boldsymbol{F}/T)$.

The asymptotic distribution of $\hat{\boldsymbol{F}}$ is more difficult to derive, especially when n and T go to infinity simultaneously. For this reason, instead of focusing on $\hat{\boldsymbol{F}}$, let us consider the following updated estimator for the factors:

$$\tilde{\boldsymbol{F}} = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{G}}(\boldsymbol{X}) \cdot (\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X}))^{-1}.$$

In addition, let $\tilde{\boldsymbol{H}} = (\boldsymbol{G}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)(\hat{\boldsymbol{G}}(\boldsymbol{X})'\hat{\boldsymbol{G}}(\boldsymbol{X})/n)^{-1}$ and

$$\boldsymbol{\Xi}_\tau = \tau(1-\tau) \cdot \boldsymbol{\Sigma}_g^{-1}\mathbb{E}[\boldsymbol{g}(\boldsymbol{x}_i)\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)']\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\boldsymbol{\Sigma}_{\phi}\boldsymbol{\Sigma}_{\mathsf{f}\phi}^{-1}\mathbb{E}[\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{g}(\boldsymbol{x}_i)']\boldsymbol{\Sigma}_g^{-1}.$$

Then, the following additional assumption is needed to derive the asymptotic distribution of $\tilde{\boldsymbol{f}}_t$.

Assumption 5. *Conditions (i) to (iii) of Assumption 4 hold and* $n^{1/2}k_n^{-\alpha}/\|\boldsymbol{\Xi}_\tau\|^{1/2} = o(1)$, $n^{1/2}\eta_{nT}/\|\boldsymbol{\Xi}_\tau\|^{1/2} = o(1)$, $\varepsilon_{nT}\sqrt{k_n} = o(1)$.

Theorem 4. *Under Assumptions 1, 2, and 5, it holds for all* $t = 1, \ldots, T$,

$$\boldsymbol{\Xi}_\tau^{-1/2}(\hat{\boldsymbol{H}}')^{-1}\sqrt{n}(\tilde{\boldsymbol{f}}_t - \tilde{\boldsymbol{H}}'\boldsymbol{f}_t) \xrightarrow{d} N(0, \boldsymbol{I}_R).$$

When $\|\boldsymbol{\Xi}_\tau\| \asymp k_n$, the convergence rate of $\tilde{\boldsymbol{f}}_t$ is $O_P(\sqrt{n/k_n})$, and Assumption 5 requires that $n^{1/2}k_n^{-\alpha-1/2} = o(1)$ and $n^{1/2}\eta_{nT}k_n^{-1/2} = o(1)$, or $k_n^4 \ll n \ll k_n^{2\alpha+1}$. As a result, we need (4) to hold with $\alpha \geq 2$. Alternatively, if Assumption 3(i) holds, it can be shown that

$$\|\boldsymbol{\Xi}_\tau - \tau(1-\tau)\cdot\boldsymbol{\Sigma}_g^{-1}\cdot \mathsf{f}^{-2}(0)\| = O(k_n^{-\alpha}).$$

In this case, the convergence rate of $\tilde{\boldsymbol{f}}_t$ is $\sqrt{n}$ for each t.

Remark 5. *As in Proposition 1 of Bai (2003), it can be shown that* $\hat{\boldsymbol{H}}$, $\tilde{\boldsymbol{H}}$ *and* $\boldsymbol{F}'\hat{\boldsymbol{F}}/T$ *all converge in probability to some positive definite matrices as* $n, T \to \infty$. *In particular, if* $\boldsymbol{F}'\boldsymbol{F}/T = \boldsymbol{I}_R$ *and* $\hat{\boldsymbol{\Sigma}}_g$ *is diagonal, the probability limits of* $\hat{\boldsymbol{H}}$, $\tilde{\boldsymbol{H}}$ *and* $\boldsymbol{F}'\hat{\boldsymbol{F}}/T$ *are all equal to* $\boldsymbol{I}_R$.

Remark 6. *Note that both Theorems 3 and 4 are fulfilled when* T *is fixed as well as when* $T \to \infty$ *as* $n \to \infty$. *In the latter case, if* $n \asymp T$, *Chen et al. (2021) show that the estimators of the quantile factors are* $\sqrt{n}$*-consistent and asymptotically normal under more general conditions. Thus, whenever* T *is as large as* n *and the quantile factors are the main objects of interest, the estimators of Chen et al. (2021) should be preferable. However, if* T *is small and* n *is large, Theorem 4 above shows that the QPPCA estimators proposed here remain consistent and asymptotically normal.*

3.3 Estimating the number of factors

Given that $\hat{\boldsymbol{Y}} = \boldsymbol{\Phi}(\boldsymbol{X})\hat{\boldsymbol{A}} \approx \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{A}_0 = \boldsymbol{\Phi}(\boldsymbol{X})\boldsymbol{B}_0\boldsymbol{F}' \approx \boldsymbol{G}(\boldsymbol{X})\boldsymbol{F}'$, the rank of $\hat{\boldsymbol{Y}}$ is asymptotically equal to R. Let $\hat{\rho}_1, \ldots, \hat{\rho}_{\bar{R}}$ be the $\bar{R}$ largest eigenvalues of $\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'/(nT)$ in descending order. Then, the estimator of R is given by the number of non-vanishing eigenvalues of $\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'/(nT)$, i.e.

$$\hat{R} = \sum_{j=1}^{\bar{R}} \mathbf{1}\{\hat{\rho}_j > p_n\}, \tag{8}$$

where $\{p_n\}$ is a sequence of non-increasing positive constants. The following theorem provides conditions on the threshold p_n to establish the consistency of $\hat{R}$ which, following Chen et al. (2021), is denoted as the rank minimization estimator of the number of factors.

Theorem 5. *Suppose that* $\bar{R} \geq R$ *and* $p_n \to 0$, $p_n\varepsilon_{nT}^{-1} \to \infty$ *as* $n \to \infty$, *then under Assumptions 1 and 2, we have* $P[\hat{R} = R] \to 1$ *as* $n \to \infty$.

To prove Theorem 5, we show that the largest R eigenvalues of $\hat{\boldsymbol{Y}}\hat{\boldsymbol{Y}}'/(nT)$ converge in probability to some positive constants, while the remaining eigenvalues are all $O_P(\varepsilon_{nT})$. Then, the decreasing sequence $\{p_n\}$ is chosen to dominate the vanishing eigenvalues in the limit. Again, this result also holds even when T is fixed.

In theory, the choice of p_n is determined by α, which depends on the smoothness of the unknown quantile loading functions. Thus, a conservative choice of p_n can rely on assuming that $\alpha = 1$. In this case, $\varepsilon_{nT} = (k_n^{1/2} n^{-1/2} \vee k_n^{-1}) \ln T$, and the optimal choice of k_n is $k_n^* \asymp n^{1/3}$. Hence, to satisfy the condition of Theorem 5, we need $p_n \gg n^{-1/3} \ln T$. The following choice is recommended in practice:

$$p_n = d \cdot \hat{\rho}_1^{1/2} \cdot n^{-1/4} \ln T, \tag{9}$$

where d is a positive constant and $\hat{\rho}_1^{1/2}$ plays the role of a normalization factor.

Remark 7. *Alternatively, to avoid the choice of the threshold sequence $\{p_n\}$, use could be made of the approach proposed by Ahn and Horenstein (2013) to estimate the number of factors by maximizing the ratios of consecutive eigenvalues, i.e.*

$$\tilde{R} = \underset{j=1,\ldots,\bar{R}}{\arg\max} \frac{\hat{\rho}_j}{\hat{\rho}_{j+1}}.$$

This is the estimator considered by Fan et al. (2016) in the context of AFM where the error terms are required to be sub-Gaussian. For this reason, a formal proof of the consistency of this estimator in the context of QFM is technically challenging, being left for further research.

3.4 Unobserved Characteristics

As anticipated above, an interesting extension of Fan et al. (2016) with respect to Connor et al. (2012) is that some of the factor loadings are allowed not to depend on the chosen set of observed characteristics. Here, we also consider the case where some of the loadings are functions of other unobserved random variables. Given their conditional mean restrictions, obtaining estimates of those extra loading components is quite straightforward: they correspond to the residuals of projecting the PPCA loading matrix on the set of sieve functions. However, allowing for this more general setup in the context of QFM would be very challenging. To see this, in line with Fan et al. (2016), assume that

$$\boldsymbol{\lambda}_i(\tau) = \boldsymbol{g}_\tau(\boldsymbol{x}_i) + \boldsymbol{\gamma}_i, \tag{10}$$

where $\boldsymbol{\gamma}_i$ is unobserved and independent of $\boldsymbol{x}_i$. Then, representation (1) can be replaced by the more general model

$$y_{it} = \boldsymbol{g}_\tau(\boldsymbol{x}_i)' \boldsymbol{f}_t(\tau) + \tilde{u}_{it}(\tau) \quad \text{where} \quad \tilde{u}_{it}(\tau) = u_{it}(\tau) + \boldsymbol{\gamma}_i' \boldsymbol{f}_t(\tau).$$

This extended model can be viewed as a CQFM with *measurement errors*, where the new error terms $\tilde{u}_{it}(\tau)$ no longer satisfy the conditional quantile restrictions, even when $\boldsymbol{x}_i$ and $\boldsymbol{\gamma}_i$ are independent. The insight is that imposing quantile conditional restrictions for the two elements of $\tilde{u}_{it}(\tau)$ does not imply that this conditional restriction should hold for their sum.[4] As a result, Step 1 of the QPPCA method will not produce consistent estimators of $\{\theta_{0t}\}$.

To overcome this problem, we consider an alternative approach labeled **QFA-Sieve** which works as follows: first, the quantile factor loadings $\{\boldsymbol{\lambda}_i(\tau)\}_{i=1}^n$ are estimated using the QFA approach of Chen et al. (2021); second, the estimated loadings are projected on the observed characteristics using least square sieves to obtain estimators of $\boldsymbol{g}(\cdot)$. The insight for this procedure relies on the following reasoning: since the first-step QFA estimators are consistent (see Chen et al. (2021)) for the factor loadings, the least-squares projection in the second step will also yield consistent estimators of the loading functions under very general conditions by treating $\boldsymbol{\lambda}_i$ as the regression errors; however, unlike the QPPCA estimator that is consistent even when T is fixed, QFA-Sieve needs both n and T to be large, as required by the QFA method. The performance of this method in finite samples is evaluated through Monte Carlo simulations in Subsection 4.3 below. Given that it fares satisfactorily, we conclude that the difference between our QPPCA and QFA-Sieve estimators can be used to heuristically check if the CQFM misses some characteristics in practice.[5]

4 Simulations

In this section, we report the results of a few Monte Carlo simulations aimed at studying the behavior in finite samples of the QPPCA estimators regarding the estimation of the number of factors, the factors themselves and their loading functions. In most cases, unless otherwise explicitly said, we take the number of characteristics to be $D = 5$ and assume that all of them, $\{x_{id}, d = 1, \ldots, 5\}$, are drawn independently from the uniform distribution: $U[-1, 1]$.

4.1 Estimating the number of factors

Consider the following DGP:

$$y_{it} = \sum_{r=1}^{3} \lambda_{ir} f_{tr} + \left(x_{i1}^2 + x_{i2}^2 + x_{i3}^2\right) u_{it},$$

[4] In fact, dealing with measurement errors is far from being a trivial issue even in standard quantile regressions (see e.g. Hausman, Liu, Luo, and Palmer (2021)).

[5] The idea is that, when T is large and there is no missing characteristic, i.e., $\boldsymbol{\gamma}_i = 0$, both QPPCA and QFA-Sieve will produce consistent estimators of $\boldsymbol{g}_\tau(\cdot)$, thus these two estimators will be close to each other. On the other hand, if there exist missing characteristics, QFA-Sieve is consistent while QPPCA is not, implying these two estimators will be different.

where $f_{t1} = 1$, $f_{t2}, f_{t3} \sim i.i.d\ N(0,1)$. Note that the chosen DGP is a location-scale shift model where the scale is driven by a subset of the five characteristics. This type of heteroskedasticity implies that the quantile loading functions exhibit variations across quantiles, unlike a pure location-shift model where the loading functions would be the same (up to a constant) at different quantiles.

Let $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ and $g_3(x) = cos(\pi x)$, such that

$$\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id}), \quad \lambda_{i2} = \sum_{d=1,2} g_2(x_{id}), \quad \lambda_{i3} = \sum_{d=3,4} g_3(x_{id}).$$

As for the idiosyncratic component, u_{it} are i.i.d. draws from three alternative distributions: (i) the standard normal distribution, $N(0,1)$, (ii) the Student's t distribution with 3 degrees of freedom, $t(3)$, and (iii) the standard Cauchy distribution, Cauchy(0,1). We set $k_n = n^{1/3}$ in the quantile sieve estimation, and make use of the *Chebyshev polynomials of the first kind* as the basis functions. Moreover, in order to implement the rank minimization estimator for the number of factors in (8), the threshold p_n is chosen as in (9), with $d = 1/4$.

First, Table 1 displays the results of the number of factors estimated with the rank minimization criterion for $\tau \in \{0.25, 0.5, 0.75\}$, $T \in \{5, 10\}$ and $n \in \{50, 100, 200, 1000\}$ from 1000 simulation replications. For each combination of τ, n and T, the reported results represent: [frequency of $\hat{R} < R$; frequency of $\hat{R} = R$; frequency of $\hat{R} > R$]. Next, for comparison, Table 2 reports the corresponding results when the number of factors is estimated using the Ahn and Horenstein (2013)'s eigen-ratio estimator discussed in Remark 7.

[Tables 1 and 2 about here]

There are three main takeaways from these simulation results. First, both selection criteria accurately estimate the number of factors when T is small and n is large, supporting our previous claim about their consistency even when T is fixed. Second, when n is large (=1000), both estimators perform well, even when the errors follow the standard Cauchy distribution. Hence, this result also provides support for the claim that our estimator is consistent in the absence of moment restrictions on the error terms. Third, although both estimators yield similar results when $n = 1000$, the rank minimization estimator outperforms the eigen-ratio estimator when n is not sufficiently large.

4.2 Estimating the factors

4.2.1 Comparison of QPPCA with PCA and PPCA

Following Chen et al. (2021), we consider the following DGP:

$$y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it},$$

where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. As before, let $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ but now $g_3(x) = |cos(\pi x)|$. The factor loading functions and the error terms are also generated as in Subsection 4.1. Note that, in this DGP, there are two location shift factors, f_{t1} and f_{t2}, that affect the mean of y_{it} and only one scale shift factor f_{t3} that affects the variance of y_{it}.

As discussed Chen et al. (2021), the methods based on PCA are not able to capture the scale factor f_{t3}. Thus, we focus on the estimation of the two location factors: f_{t1} and f_{t2}. Three competing estimation methods are considered: (i) the proposed method with $\tau = 0.5$ (QPPCA); (ii) the projection estimator proposed by Fan et al. (2016) (PPCA); and (iii) the standard estimator of Bai and Ng (2002) for AFM (PCA). For the first method, the choices of k_n and the basis functions are again the same as in Subsection 4.1. Regarding the choices of n and T, we fix $T = 10, 50$ and let n increase from 50 to 500. For each estimation method, the number of location factors ($R = 2$) is assumed to be known, and we report the average Frobenius error as a measure of fit: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\boldsymbol{F}\|/\sqrt{T}$ from 1000 replications.

The results are plotted in Figure 1. As can be inspected from the three graphs on the left panel, for small T ($T = 10$), the PCA estimators perform worse than the PPCA and QPPCA estimators when u_{it} is either drawn from the $N(0,1)$ or $t(3)$ distributions. Moreover, when the distribution is a standard Cauchy, the QPPCA estimator performs much better than any of its competitors. These findings agree again with our previous theoretical results showing that this estimator is consistent even when T is fixed or the moments of u_{it} do not exist. On the other hand, it can be seen from the three graphs on the right panel that when T is not small ($T = 50$), the three methods perform similarly when the errors are normal, but again QPPCA method stands out as the best procedure when the distribution of the errors become heavy-tailed.

[Figure 1 about here]

In sum, the results in this subsection show that, when estimating the mean factors, (i) the QPPCA estimator is consistent even when T is fixed, and (ii) compared with the PCA and PPCA estimators, the QPPCA estimator is more robust to outliers.

4.2.2 Comparison of QPPCA with QFA and SQFA when $D > R$

The DGP is the same as in the previous subsection, with $D = 5$ and $R = 3$. Now, we focus on the estimation of the two location factors: f_{t1} and f_{t2} plus the scale factor f_{t3}. Three competing estimation methods are considered: (i) QPPCA; (ii) the SQFA estimator proposed by Ma et al. (2021); (iii) the quantile factor analysis estimator (QFA) of Chen et al. (2021). For the first two methods, the choices of k_n and the basis functions are again the same as in Subsections 4.1 and 4.2.1. All the methods are applied for $\tau = 0.25, 0.75$. Again, we fix $T = 10, 50$ and let n increase from 50 to 500. For each estimation method, the number of factors ($R = 3$) is assumed to be known, and we report the average Frobenius error as a measure of fit: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\boldsymbol{F}\|/\sqrt{T}$ from 1000 replications.

The results for $\tau = 0.25$ and $\tau = 0.75$ are plotted in Figure 2 and Figure 3 respectively. There are two main takeaways: first, for small T ($T = 10$), the QFA estimators perform much worse than the SQFA and QPPCA estimators in all three cases, which is as expected since the consistency of QFA estimators requires both n and T to diverge; and second, when T becomes large ($T = 50$), the three methods perform similarly — they all successfully span the space of the location and scale factors even when the errors are drawn from Cauchy distributions, where SQFA performs slightly better than QPPCA for $n < 200$. However, it should be noted that for the QFA and QPPCA methods, the number of estimated factors is 3, while for the SQFA estimator the number of estimated factors is 5, because this method assumes that $R = D$.[6] In the next subsection, we show the main advantage of the QPPCA estimator compared with the SQFA estimator when D is smaller than R.

[Figures 2 and 3 about here]

4.2.3 Comparison of QPPCA with SQFA when $D < R$

Consider the following location-scale model as the DGP:

$$y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it},$$

where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. Now, the number of characteristics is 2 and, as in the previous simulations, all characteristics x_{id} ($i = 1, ..., n$ and $d = 1, 2$) are independently drawn from the uniform distribution: $U[-1, 1]$. Let $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. Moreover, let $\lambda_{i1} = \sum_{d=1,2} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ *and* $\lambda_{i3} = \sum_{d=1,2} g_3(x_{id})$. Again, u_{it} are generated from three different distributions.

For each estimator, we consider $\tau \in \{0.25, 0.75\}$, $T = 50$, and n increases from 50 to 500.

[6] In fact, we have also run simulations with $R = D$ (the case for which SQFA is designed) finding that SQFA and QPPCA yield very similar results

For each τ, $R = 3$ factors are estimated using QPPCA. Note that the SQFA method chooses the number of factors as the number of characteristics by default, implying that only $D = 2$ factors will be estimated. Moreover, the choices of the basis functions and k_n are the same as in Subsection 4.1. As before, we report the average Frobenius error as a measure of fit: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\boldsymbol{F}\|/\sqrt{T}$ from 1000 replications. The results are plotted in Figure 4 for $\tau = 0.25$ (left panel) and $\tau = 0.75$ (right panel). It is not surprising to check that the QPPCA estimator outperforms the SQFA estimator since the latter is restricted to estimating only $D = 2$ factors. Further analysis shows that the scale factor f_{t3} can not be consistently estimated by the SQFA estimator.

[Figure 4 about here]

Thus, the main overall findings of this and the previous subsection are that the QPPCA and SQFA estimators perform similarly whenever the number of characteristics exceeds the number of factors while QPPCA beats SFQA by a large margin when the opposite happens. The insight for this result is similar to adding irrelevant covariates in a regression, where consistency remains, against the case where relevant controls are omitted, leading to biased estimators.

4.3 Estimating the loading functions

We next focus on the estimation of the loading functions. Motivated by the empirical application in the next section, we consider the specification: $n = 355, T = 62, R = 1$.

First, consider the following DGP:

$$y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it},$$

where x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3, 4$) are independently drawn from the uniform distribution, and these four characteristics are all assumed to be observed. Let $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$, and u_{it} are i.i.d draws from the $t(3)$ distribution. Moreover, $f_t = |g_t|$ and $\{g_t\}$ are i.i.d draws from $N(0, 1)$. Except the QPPCA estimator, we also consider the QFA-Sieve estimator proposed in Section 3.4 for $g_j(\cdot)$, $j = 1, \ldots, 4$ at $\tau = 0.25, 0.75$. Specifically, the QFA estimator is applied to y_{it} first to obtain the estimated factor loadings, which are then projected onto the space of $x_{id}, d = 1, \ldots, 4$ using sieve least squares to estimate the loading functions. The choices of the basis functions and k_n are the same as in Section 4.1.

The estimated results are reported in Figure 5 and Figure 6 for $\tau = 0.25$ and $\tau = 0.75$ respectively. The black lines are the true loading functions, while the red lines give the 95% and 5% point-wise empirical quantiles of the estimated functions using QPPCA from 1000 replications, and the green lines provide the results for the QFA-Sieve estimator. It can be seen that the estimated functions using both methods trace the true functions very well, and

confidence intervals between the 5% and 95% quantiles are very narrow. The fact that the QFA-Sieve estimator and the QPPCA estimator for the loading functions are almost identical is a strong indication that there are no missing characteristics.

[Figures 5 and 6 about here]

Next, consider the following DGP:

$$y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3}) + 1.5cos(\pi x_{i5})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it},$$

where the only difference with the previous DGP is that now we assume that only the first four characteristics are observed while x_{i5} is drawn independently from $U[-1, 1]$ and it is assumed to be unobserved. Again, we consider the QPPCA estimator and the QFA-Sieve estimator for $g_j(\cdot)$, $j = 1, \ldots, 4$ at $\tau = 0.25, 0.75$, treating x_{i5} as the unobserved characteristic.

The estimated results are reported in Figure 7 and Figure 8 for $\tau = 0.25$ and $\tau = 0.75$ respectively. It can be seen that, as in the previous subsection, all the loading functions can be consistently estimated by the QFA-Sieve method, despite the assumption that the fifth characteristic remains unobserved. On the other hand, the QPPCA estimator of $g_4(\cdot)$ is obviously inconsistent. These results support our claim that the QFA-Sieve method is robust against unobserved or missing variables in the factor loading functions. Moreover, the difference between the two estimators of $g_4(\cdot)$ can be used as the evidence for the existence of missing characteristics.

However, there are two observations worth mentioning. First, due to the presence of unobserved characteristics, the confidence intervals of the QFA-Sieve estimators are much wider than in the case of no missing characteristics. Second, as discussed in Section 3.4, a sufficiently large T is necessary for the QFA-Sieve method to show its full strength in dealing with unobserved characteristics. The simulation results of Chen et al. (2021) show that the QFA estimator performs well for $T \geq 50$ and longer samples, implying that the first-step estimation error of the QFA-Sieve method is very small for DGPs ($T = 62$) considered in this subsection. Conversely, the additional simulations shown in the online appendix (Figures A.1 to A.6) illustrate that QFA-Sieve does not fare well when T is small. In particular, when $T = 10$ in the DGP considered there, the confidence intervals of the QPPCA estimators contain the true loading functions, and their widths shrink as n increases, in contrast to those of the QFA-Sieve estimators which are much wider and do not shrink as n increases.

[Figures 7 and 8 about here]

5 Empirical Application

In this section, the QPPCA estimation approach is used to investigate the factor structure of security returns. As Fan et al. (2016), we use a dataset that includes information on the daily

returns of S&P500 index securities with complete daily closing price records from 2005 to 2013.[7] The sample consists of 355 stocks, whose book value and market capitalization are drawn from Compustat. Moreover, as in most of this literature, the 1-month US treasury bond rate is used as the risk-free rate to compute the daily excess return of each stock.

Following Connor et al. (2012), Fan et al. (2016), and Ma et al. (2021), four characteristics are considered: *size, value, momentum* and *volatility*, which are standardized to have zero means and unit standard deviations. Similar to Fan et al. (2016), we analyze the data corresponding to the first quarter of 2006, which includes $T = 62$ observations. As before, the Chebyshev polynomials of the first kind are used as the basis functions in the sieve regressions with $k_n = 4$.

First, Table 3 shows the estimated number of mean factors using the eigen-ratio estimator proposed by Fan et al. (2016) and the estimated numbers of quantile factors using the QPPCA rank minimization estimator for $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. The five largest eigenvalues of $\hat{Y}\hat{Y}'$ and the threshold p_n are also displayed in this table. In addition, the estimated numbers of quantile factors using the QFA rank minimization estimator are reported in the last column. Overall, the results provide strong evidence in favor of the existence of a single location factor and one factor at each quantile.

[Table 3 about here]

Second, Table 4 shows the correlation coefficients between the estimated location factor by PPCA and the estimated quantile factors by QPPCA for the aforementioned values of τ. The sample means of each estimated factor are also reported in the last column. As can be inspected, the PPCA factor is highly correlated with the QPPCA factors. Combined with the results of Table 3, the high correlations between the mean factor and the quantile factor at different quantile levels support the existence of a single factor that shifts the mean and quantiles of the return distributions.

[Table 4 about here]

Third, Figure 9 to Figure 12 show the estimation quantile loading functions (red lines) of the four characteristics using QPPCA at $\tau = 0.1, 0.25, 0.75, 0.9$, along with their point-wise 95% confidence intervals (shaded areas) constructed using the asymptotic distribution of Theorem 3. We also plot the estimated quantile loading functions using QFA-Sieve (green lines), which allow us to determine the existence of missing characteristics in light of the simulation results in Section 4.3. These plots suggest that all characteristics are observed. Moreover, the estimated loading functions using PPCA (blue lines) are also included for comparison.

[Figures 9 to 12 about here]

Overall, a few salient findings emerge from this exercise. First, for each characteristic, the quantile variation of their loading functions is larger at the more extreme quantiles (i.e.

[7] This dataset is downloaded from CRSP (Center for Research in Security Prices).

$\tau = 0.1, 0.9$). For example, for the "size" characteristic (Figure 9), the loading functions at $\tau = 0.1, 0.25, 0.75$ are very similar and almost flat, being clearly different from the decreasing function at $\tau = 0.9$, and implying that the impact of "size" on the factor loadings mainly affects the upper quantiles of the returns. Second, in most cases, the estimated loading functions using QFA-Sieve lie within the confidence intervals of the QPPCA estimators, indicating that there is no missing characteristic. Moreover, the loading functions of "size" and "volatility" (Figures 9 and 12) behave monotonically at all quantiles while, for "value" and "momentum" (Figures 10 and 11), they exhibit a clear non-linear pattern, mostly looking U-shaped and inverted U-shaped, respectively. Interestingly, the shapes of the loading function resemble those reported by Ma et al. (2021) except for "value" (i.e. a value stock refers to shares of a company that appears to trade at a lower price relative to its fundamentals, such as dividends, earnings, or sales) which these authors find to have an inverted U-shape.

In sum, this empirical evidence points out that the estimated loading functions vary substantially across different quantiles, a feature that cannot be uncovered using the PPCA method. Yet, this is a useful finding since deviating from the efficient market hypothesis, factors contributing to alpha generation can have different relevance depending on the distribution of excess returns. Thus, the standard asset valuation techniques based on CAPM and the Fama-French factors should take these features into consideration to create a portfolio delivering excess returns over time which would beat the market.

Finally, it is worth highlighting that QPPCA allows estimating the conditional quantile of excess returns $\mathsf{Q}_\tau[y_{it}|\boldsymbol{x}_i] = \boldsymbol{g}_\tau(\boldsymbol{x}_i)'\boldsymbol{f}_t$, yielding $\hat{y}_{it}(\tau) = \hat{\boldsymbol{a}}_t'\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ as its estimator, where $\hat{\boldsymbol{a}}_t$ is obtained from the cross-sectional quantile regressions in step 1 of the three-step procedure introduced in Section 2. One could interpret $\hat{y}_{it}(\tau)$ as the "quantile return" which is interesting from the perspective of empirical applications since it is idiosyncratic free, that is, much less noisy than the realized return y_{it}. Just as the literature on asset pricing has increasingly appreciated the concept of "expected returns" because it is noiseless (see e.g. Elton (1999)),"quantile returns" are also interesting on their own and could perhaps help provide a better explanation of the distribution of returns, an issue which remains high in our research agenda.

6 Conclusions

This paper proposes a three-stage estimation method for characteristics-based quantile factor models (CQFM). The convergence rates of the proposed estimators, labeled QPPCA, are established, and the asymptotic distributions of the estimated factors and loading functions are derived under very general conditions. Compared with the existing estimation methods of CQFM, not only QPPCA estimators are easier to implement in practice, but also they are consistent for fixed T as long as n goes to infinity, as well as being robust to heavy tails and outliers in the

distribution of the idiosyncratic errors. Moreover, the number of quantile factors are allowed to be different from the number of the characteristics, and this number can be consistently estimated using a new rank-minimization estimator proposed in this paper. Finally, we propose a method labeled as QFA-Sieve that can consistently estimate the loading functions of the observed characteristics if T is large and can be employed to identify the presence of unobserved characteristics in our CQFM formulation.

Simulation results show that the proposed estimators perform satisfactorily in finite samples, especially when the number of cross-section observations is large. An application of QPPCA to a dataset consisting of individual stock returns reveals that the quantile factor loadings are nonlinear functions of some observed characteristics and that these functions exhibit considerable variations across quantiles. Further, we do provide evidence that no omitted characteristics are relevant. We conjecture that these results lead to the concept of *quantile returns* which generalizes the standard concept of *expected returns*, typically proxied by averages of realized returns. The methodology associated with QPPCA is useful to derive the convergence rates and asymptotic properties of the (average) quantile returns which remains high in our ongoing research agenda.

A Figures and Tables

Table 1: Estimating the number of factors: rank minimization estimator

	T	n	$N(0,1)$			$t(3)$			Cauchy(0,1)		
$\tau = 0.25$	5	50	[0.13	0.65	0.23]	[0.03	0.41	0.56]	[0.01	0.10	0.89]
	5	100	[0.10	0.72	0.19]	[0.02	0.44	0.54]	[0.00	0.03	0.97]
	5	200	[0.23	0.77	0.00]	[0.12	0.82	0.06]	[0.00	0.17	0.83]
	5	1000	[0.17	0.83	0.00]	[0.16	0.84	0.00]	[0.06	0.81	0.13]
	10	50	[0.17	0.76	0.07]	[0.03	0.50	0.47]	[0.02	0.06	0.92]
	10	100	[0.08	0.89	0.03]	[0.03	0.65	0.46]	[0.00	0.03	0.97]
	10	200	[0.07	0.93	0.00]	[0.05	0.95	0.00]	[0.00	0.24	0.76]
	10	1000	[0.03	0.97	0.00]	[0.02	0.98	0.00]	[0.01	0.98	0.01]
$\tau = 0.5$	5	50	[0.19	0.71	0.10]	[0.09	0.56	0.35]	[0.00	0.15	0.85]
	5	100	[0.17	0.76	0.08]	[0.07	0.59	0.34]	[0.00	0.20	0.80]
	5	200	[0.23	0.77	0.00]	[0.19	0.80	0.01]	[0.06	0.75	0.19]
	5	1000	[0.18	0.82	0.00]	[0.15	0.85	0.00]	[0.13	0.87	0.00]
	10	50	[0.20	0.78	0.03]	[0.08	0.76	0.15]	[0.00	0.13	0.87]
	10	100	[0.12	0.87	0.01]	[0.05	0.87	0.08]	[0.00	0.24	0.76]
	10	200	[0.05	0.95	0.00]	[0.05	0.95	0.00]	[0.03	0.94	0.03]
	10	1000	[0.01	0.99	0.00]	[0.02	0.98	0.00]	[0.02	0.99	0.00]
$\tau = 0.75$	5	50	[0.11	0.68	0.21]	[0.04	0.41	0.56]	[0.01	0.09	0.90]
	5	100	[0.10	0.71	0.19]	[0.02	0.42	0.56]	[0.00	0.04	0.96]
	5	200	[0.22	0.78	0.00]	[0.14	0.81	0.05]	[0.00	0.15	0.85]
	5	1000	[0.18	0.82	0.00]	[0.17	0.83	0.00]	[0.04	0.82	0.15]
	10	50	[0.15	0.78	0.08]	[0.04	0.50	0.46]	[0.01	0.05	0.94]
	10	100	[0.11	0.86	0.04]	[0.03	0.65	0.32]	[0.00	0.03	0.97]
	10	200	[0.06	0.94	0.00]	[0.05	0.94	0.01]	[0.01	0.27	0.73]
	10	1000	[0.02	0.98	0.00]	[0.02	0.98	0.00]	[0.02	0.97	0.01]

Note: The DGP is $y_{it} = \sum_{r=1}^{3} \lambda_{ir} f_{tr} + \left(x_{i1}^2 + x_{i2}^2 + x_{i3}^2\right) u_{it}$, where $f_{t1} = 1$, $f_{t2}, f_{t3} \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are drawn independently from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ and $g_3(x) = cos(\pi x)$, and $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$, $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. u_{it} are i.i.d variables drawn from three different distributions. In the first step quantile sieve estimation, $k_n = n^{1/3}$ and we use the *Chebyshev polynomials of the first kind* as the basis functions. For the estimator of the number of factors, the threshold p_n is chosen as in (9) with $d = 1/4$. The reported results are [frequency of $\hat{R} < R$; frequency of $\hat{R} = R$; frequency of $\hat{R} > R$] from 1000 replications.

Table 2: Estimating the number of factors: eigen-ratio estimator

	T	n	$N(0,1)$			$t(3)$			Cauchy(0,1)		
$\tau = 0.25$	5	50	[0.57	0.25	0.19]	[0.54	0.22	0.25]	[0.54	0.17	0.29]
	5	100	[0.58	0.33	0.09]	[0.58	0.27	0.15]	[0.59	0.15	0.26]
	5	200	[0.44	0.54	0.01]	[0.54	0.43	0.04]	[0.62	0.24	0.14]
	5	1000	[0.23	0.77	0.00]	[0.31	0.69	0.00]	[0.56	0.42	0.02]
	10	50	[0.46	0.37	0.17]	[0.45	0.18	0.37]	[0.47	0.07	0.46]
	10	100	[0.37	0.59	0.04]	[0.46	0.42	0.11]	[0.60	0.09	0.31]
	10	200	[0.09	0.91	0.00]	[0.19	0.80	0.01]	[0.59	0.31	0.11]
	10	1000	[0.01	0.99	0.00]	[0.03	0.97	0.00]	[0.17	0.83	0.00]
$\tau = 0.5$	5	50	[0.58	0.28	0.14]	[0.57	0.22	0.20]	[0.50	0.20	0.30]
	5	100	[0.58	0.33	0.09]	[0.57	0.28	0.15]	[0.56	0.21	0.22]
	5	200	[0.42	0.57	0.01]	[0.46	0.51	0.03]	[0.54	0.41	0.06]
	5	1000	[0.21	0.79	0.00]	[0.23	0.77	0.00]	[0.28	0.72	0.00]
	10	50	[0.41	0.46	0.13]	[0.46	0.33	0.21]	[0.42	0.10	0.48]
	10	100	[0.30	0.66	0.04]	[0.36	0.57	0.07]	[0.51	0.24	0.26]
	10	200	[0.06	0.94	0.00]	[0.11	0.89	0.00]	[0.22	0.76	0.02]
	10	1000	[0.01	0.99	0.00]	[0.02	0.98	0.00]	[0.03	0.97	0.00]
$\tau = 0.75$	5	50	[0.58	0.25	0.17]	[0.54	0.22	0.24]	[0.55	0.17	0.28]
	5	100	[0.57	0.32	0.10]	[0.59	0.24	0.17]	[0.56	0.20	0.24]
	5	200	[0.43	0.55	0.02]	[0.52	0.43	0.04]	[0.65	0.21	0.14]
	5	1000	[0.24	0.76	0.00]	[0.33	0.67	0.00]	[0.55	0.44	0.01]
	10	50	[0.46	0.36	0.18]	[0.44	0.20	0.37]	[0.47	0.05	0.48]
	10	100	[0.36	0.59	0.06]	[0.46	0.40	0.14]	[0.63	0.09	0.28]
	10	200	[0.11	0.89	0.00]	[0.19	0.80	0.01]	[0.58	0.31	0.11]
	10	1000	[0.01	0.99	0.00]	[0.03	0.97	0.00]	[0.16	0.83	0.01]

Note: The DGP is $y_{it} = \sum_{r=1}^{3} \lambda_{ir} f_{tr} + \left(x_{i1}^2 + x_{i2}^2 + x_{i3}^2\right) u_{it}$, where $f_{t1} = 1$, $f_{t2}, f_{t3} \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} are drawn independently from the uniform distribution: $U[-1,1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ and $g_3(x) = cos(\pi x)$, and $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$, $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. u_{it} are i.i.d variables drawn from three different distributions. In the first step quantile sieve estimation, $k_n = n^{1/3}$ and we use the *Chebyshev polynomials of the first kind* as the basis functions. The estimator for the number of factors is the integer that maximizes the eigen-ratios. The reported results are [frequency of $\hat{R} < R$; frequency of $\hat{R} = R$; frequency of $\hat{R} > R$] from 1000 replications.

Table 3: Estimated numbers of factors

		Five largest eigenvalues of $\hat{Y}\hat{Y}'$					p_n	$\hat{r}$	$\hat{r}_{QFA}$
mean(PPCA)		0.929	0.090	0.081	0.066	0.043		1	
quantile	τ=0.1	4.913	0.145	0.127	0.115	0.086	0.527	1	1
	τ=0.25	1.713	0.110	0.098	0.059	0.047	0.311	1	1
	τ=0.5	0.887	0.094	0.084	0.053	0.043	0.224	1	1
	τ=0.75	2.706	0.115	0.087	0.074	0.067	0.391	1	1
	τ=0.9	8.000	0.297	0.209	0.151	0.122	0.672	1	1

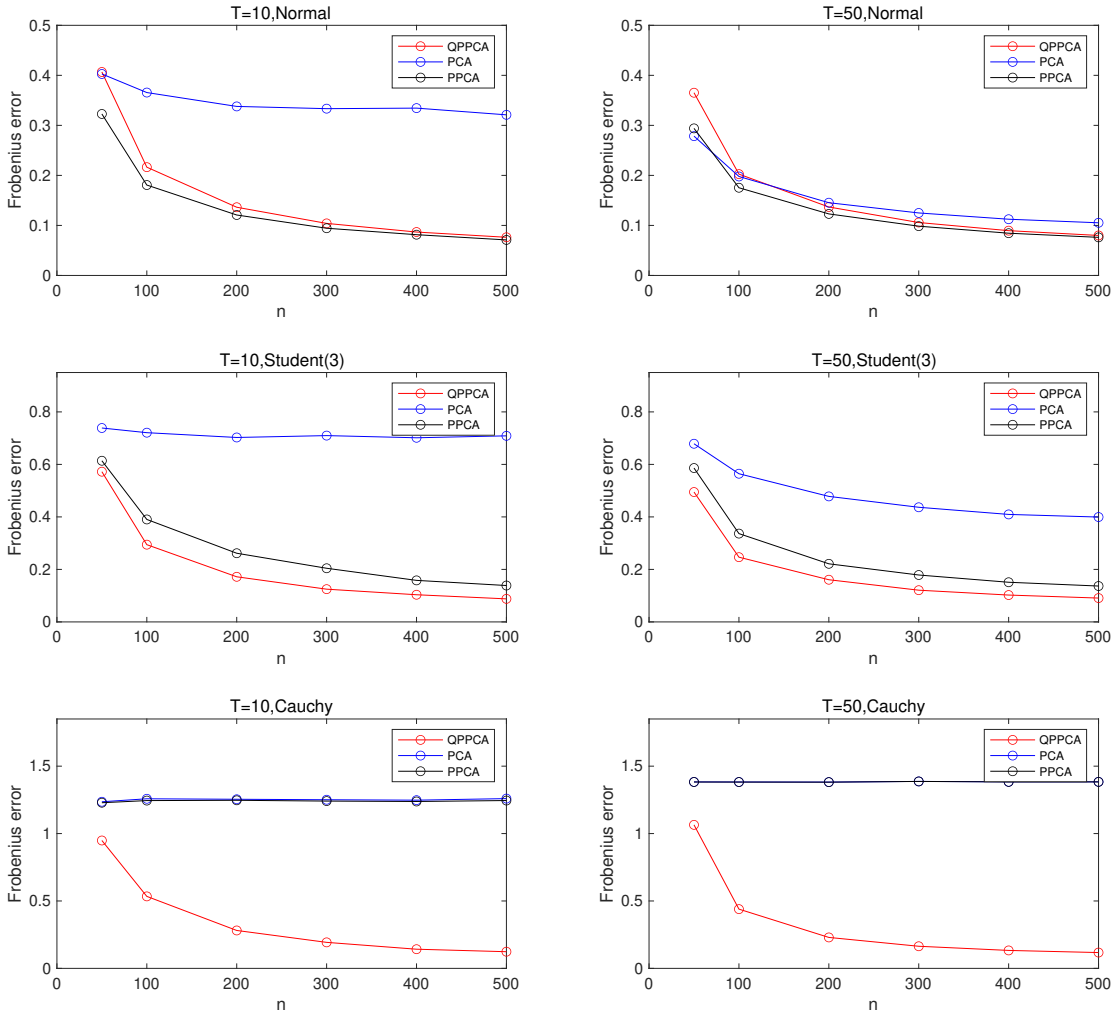
Note: This table shows the estimated numbers of factors using the eigen-ratio estimator proposed by Fan et al. (2016), the proposed estimator in this paper, and the rank-minimization estimator proposed by Chen et al. (2021) for different τs. Column 3 to Column 7 give the 5 largest eigenvalues of $\hat{Y}\hat{Y}'$, where $\hat{Y}$ is the matrix of fitted values in the first-step sieve regressions, and p_n is the threshold value defined in (9).

Table 4: Correlations and means of estimated factors

	τ=0.1	τ=0.25	τ=0.5	τ=0.75	τ=0.9	PPCA	Mean
τ=0.1	1	0.966	0.910	0.838	0.763	0.915	0.906
τ=0.25		1	0.975	0.924	0.856	0.973	0.738
τ=0.5			1	0.971	0.914	0.990	-0.121
τ=0.75				1	0.967	0.979	-0.784
τ=0.9					1	0.943	-0.917
PPCA						1	-0.231

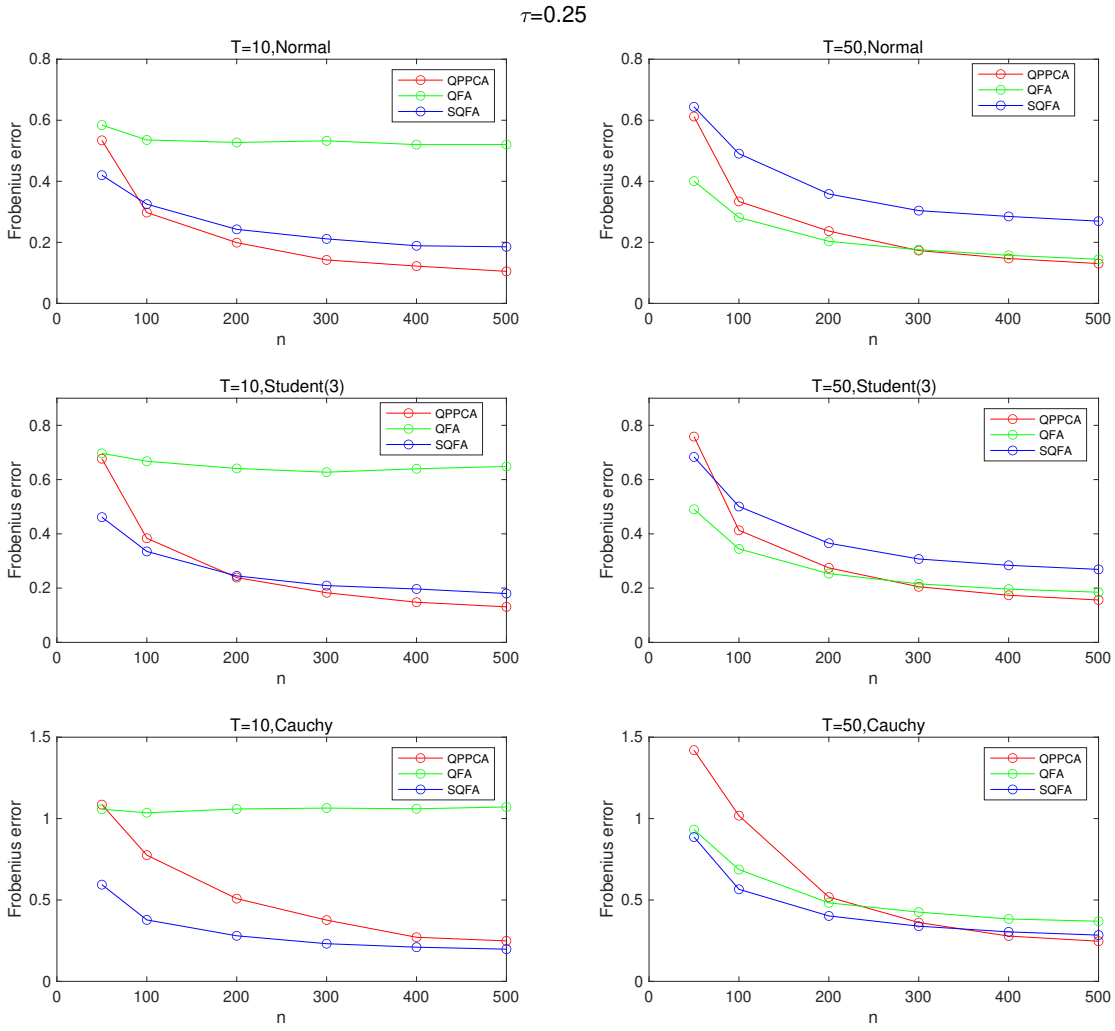
Note: This Table shows the correlations and sample means of the estimated mean factor using PPCA and the estimated quantile factors at different τs using QPPCA.

Figure 1: Estimation of factors: QPPCA, PCA and PPCA.



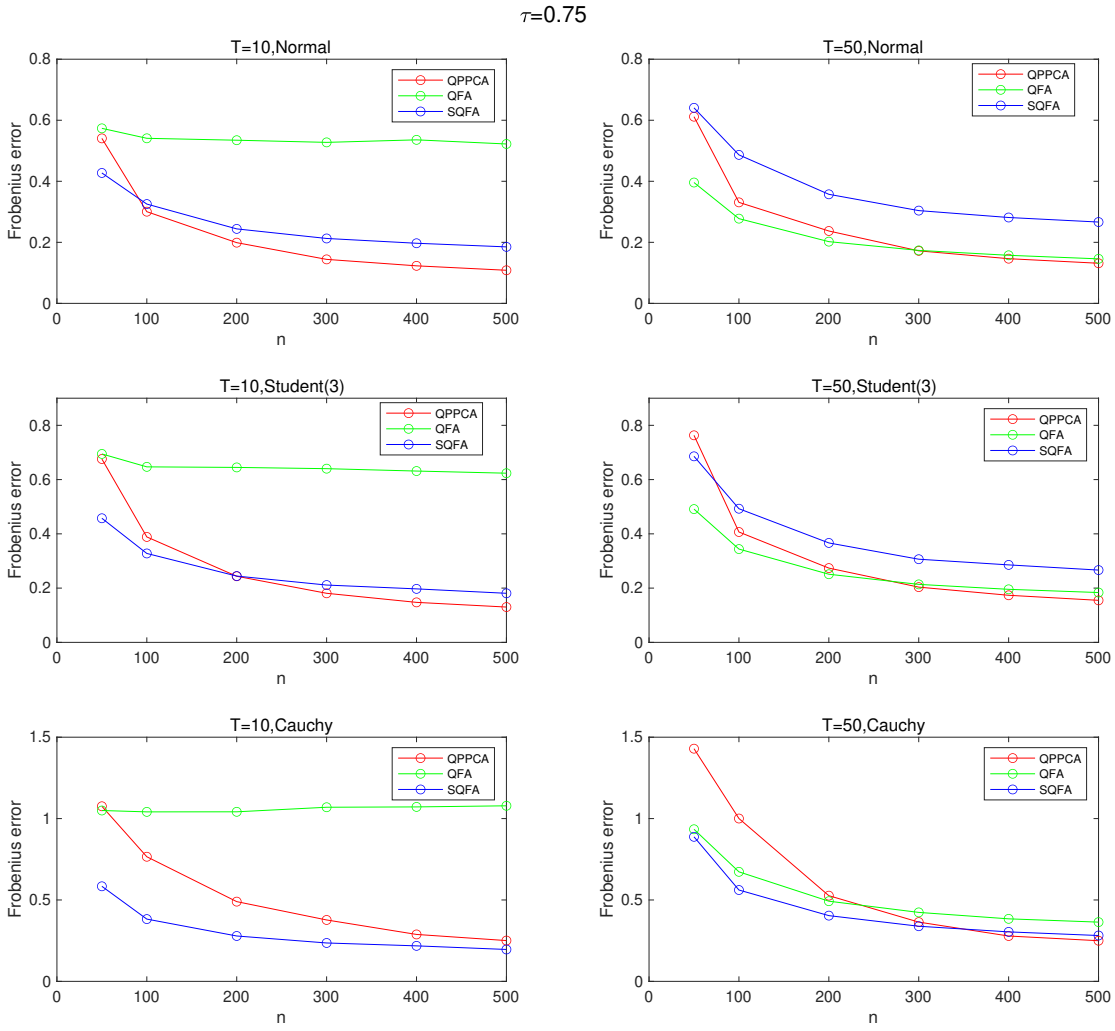
Note: The DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2, 3, 4, 5)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The mean factors (f_{t1} and f_{t2}) are estimated by 3 methods: PCA, PPCA, and QPPCA at $\tau = 0.5$. The reported results are the average Frobenius errors: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}\|/\sqrt{T}$ from 1000 repetitions.

Figure 2: Estimation of factors: QPPCA, QFA and SQFA, with $D > R$.



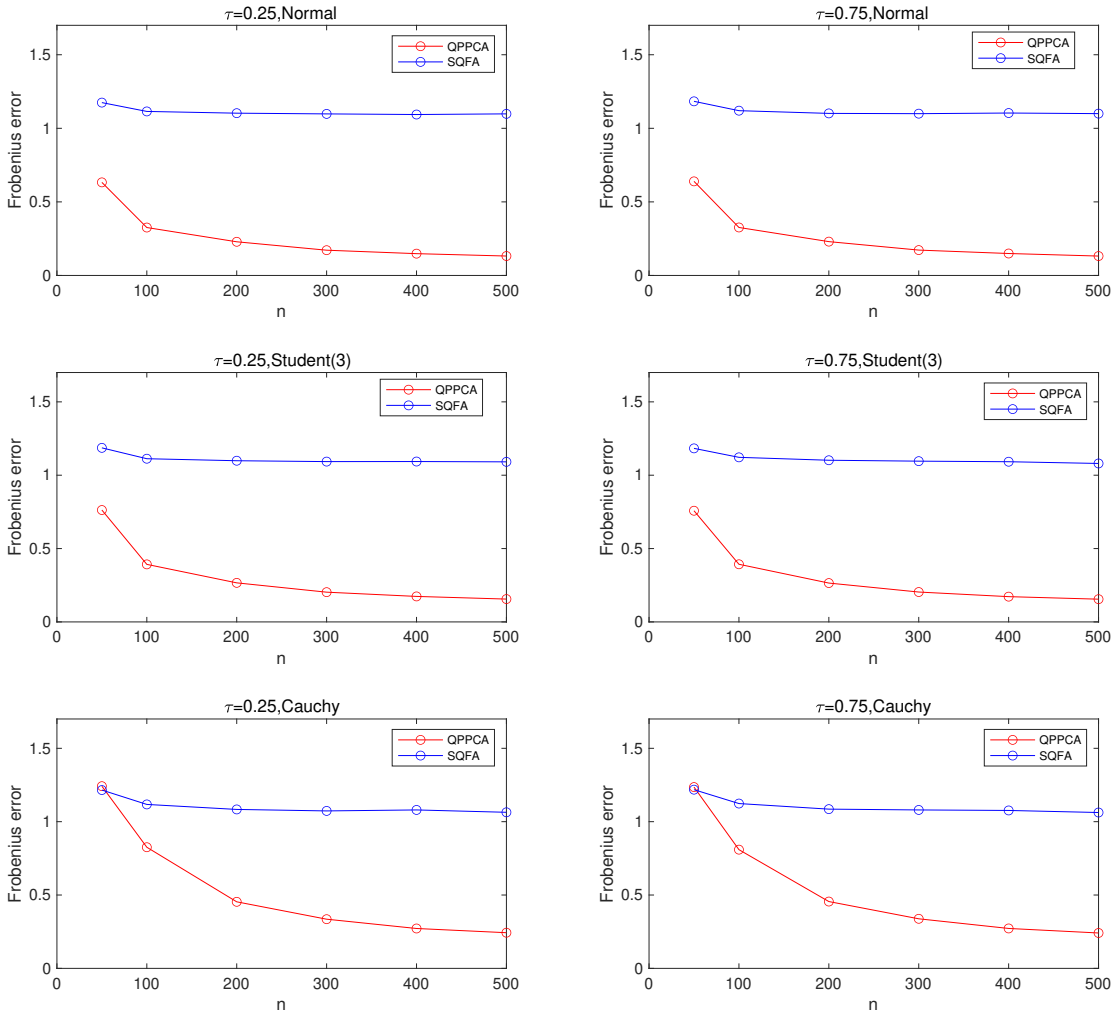
Note: The DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2, 3, 4, 5)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The three factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.25$. The reported results are the average Frobenius errors: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}\|/\sqrt{T}$ from 1000 repetitions.

Figure 3: Estimation of factors: QPPCA, QFA and SQFA, with $D > R$.



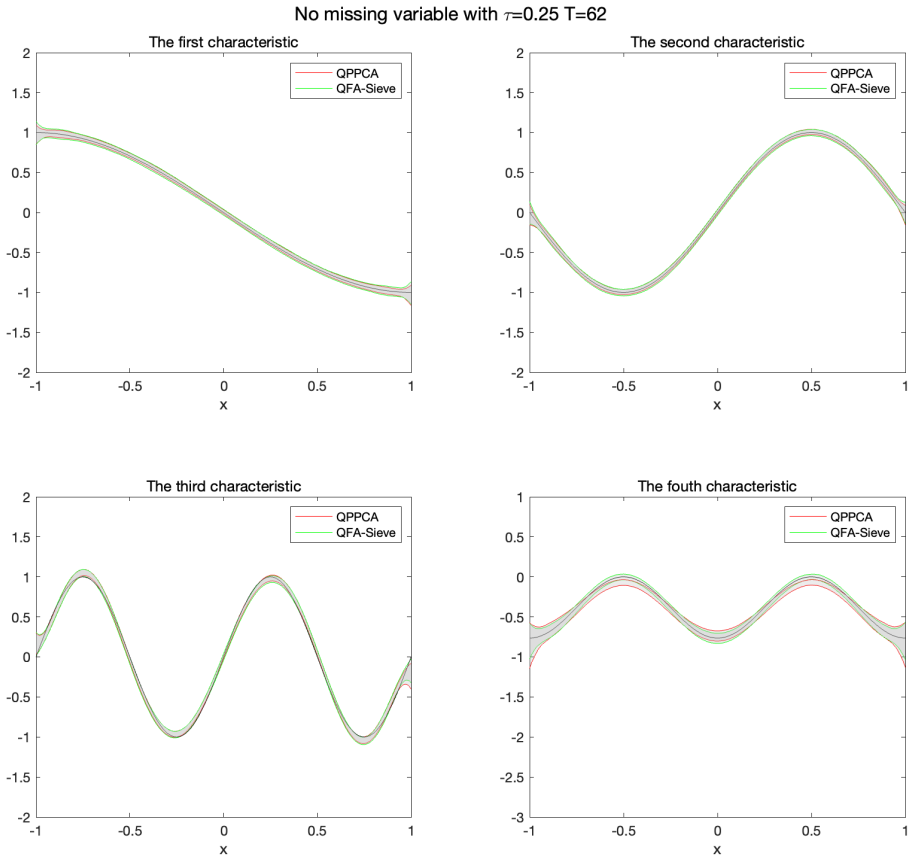
Note: The DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 5 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2, 3, 4, 5)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,3,5} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=3,4} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The three factors are estimated by 3 methods: QPPCA, QFA and SQFA at $\tau = 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}\|/\sqrt{T}$ from 1000 repetitions.

Figure 4: Estimation of factors: QPPCA and SQFA, with $T = 50$ and $D < R$.



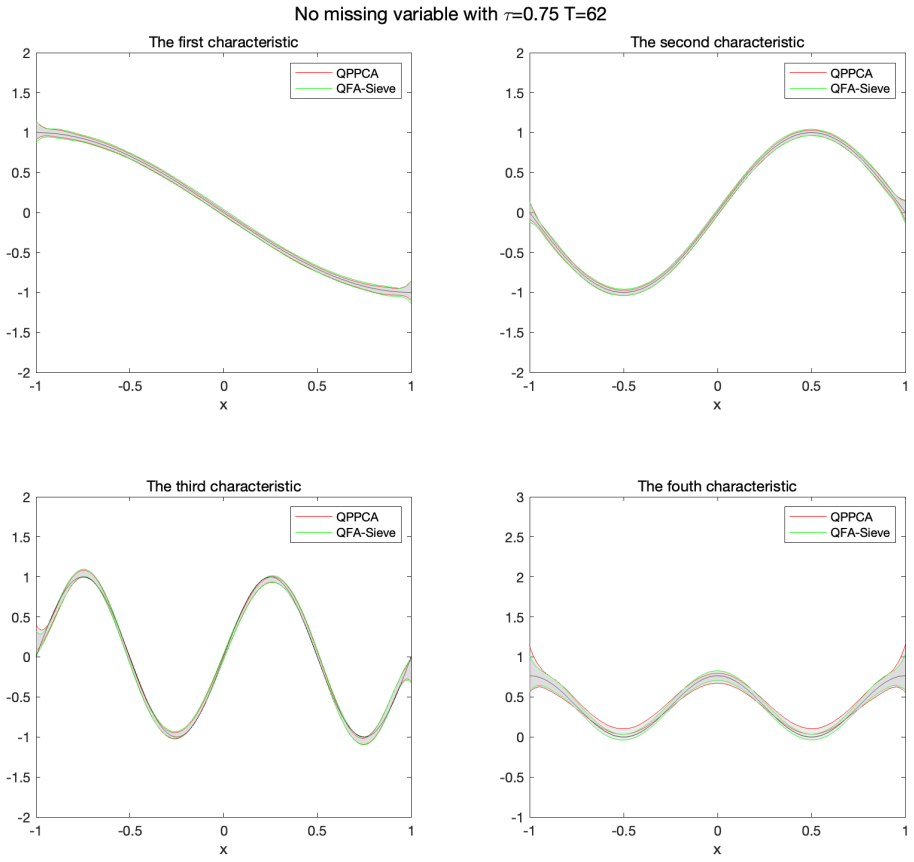
Note: The DGP is $Y_{it} = \lambda_{i1} f_{t1} + \lambda_{i2} f_{t2} + (\lambda_{i3} f_{t3}) u_{it}$, where $f_{t3} = |h_t|, f_{t1}, f_{t2}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = sin(2\pi x)$, $g_2(x) = sin(\pi x)$ *and* $g_3(x) = |cos(\pi x)|$. The factor loading functions are generated as $\lambda_{i1} = \sum_{d=1,2} g_1(x_{id})$, $\lambda_{i2} = \sum_{d=1,2} g_2(x_{id})$ and $\lambda_{i3} = \sum_{d=1,2} g_3(x_{id})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The three factors are estimated by 2 methods: QPPCA and SQFA at $\tau = 0.25, 0.75$. The reported results are the average Frobenius errors: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}\|/\sqrt{T}$ from 1000 repetitions.

Figure 5: Estimated Loading functions at $\tau = 0.25$, no missing variable: QPPCA and QFA-Sieve



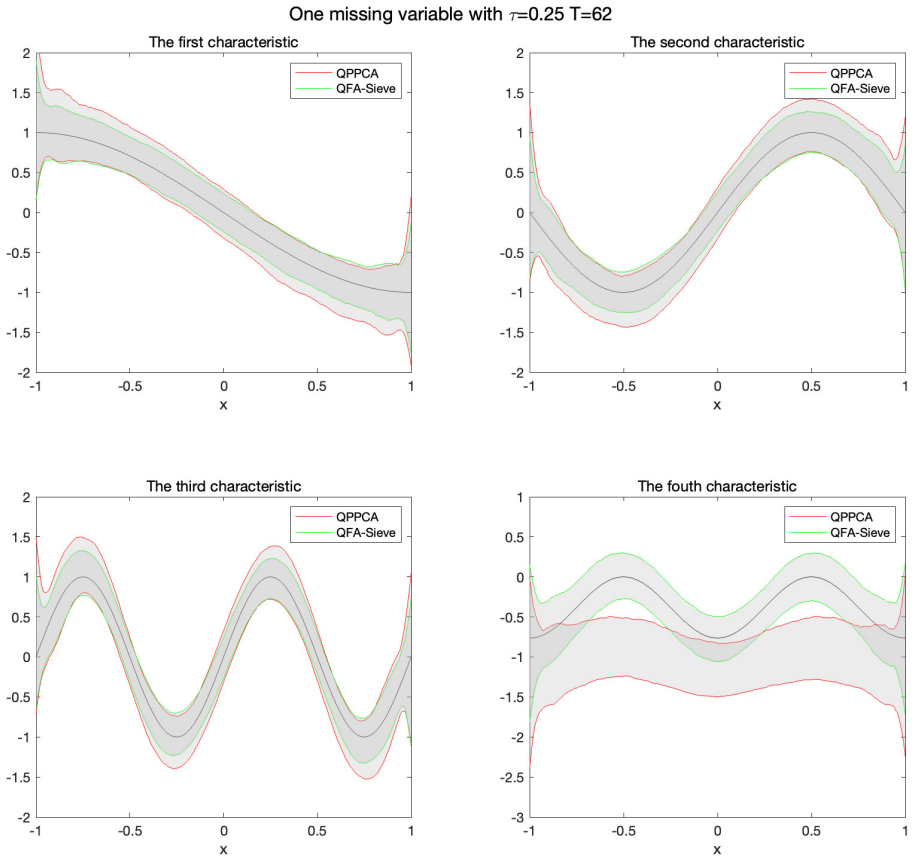
Note: $n = 355, T = 62, R = 1$ and $D = 4$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0,1)$, x_{id} $(i = 1, ..., n$ and $d = 1, 2, 3, 4)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black line) at $\tau = 0.25$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure 6: Estimated Loading functions at $\tau = 0.75$, no missing variable: QPPCA and QFA-Sieve



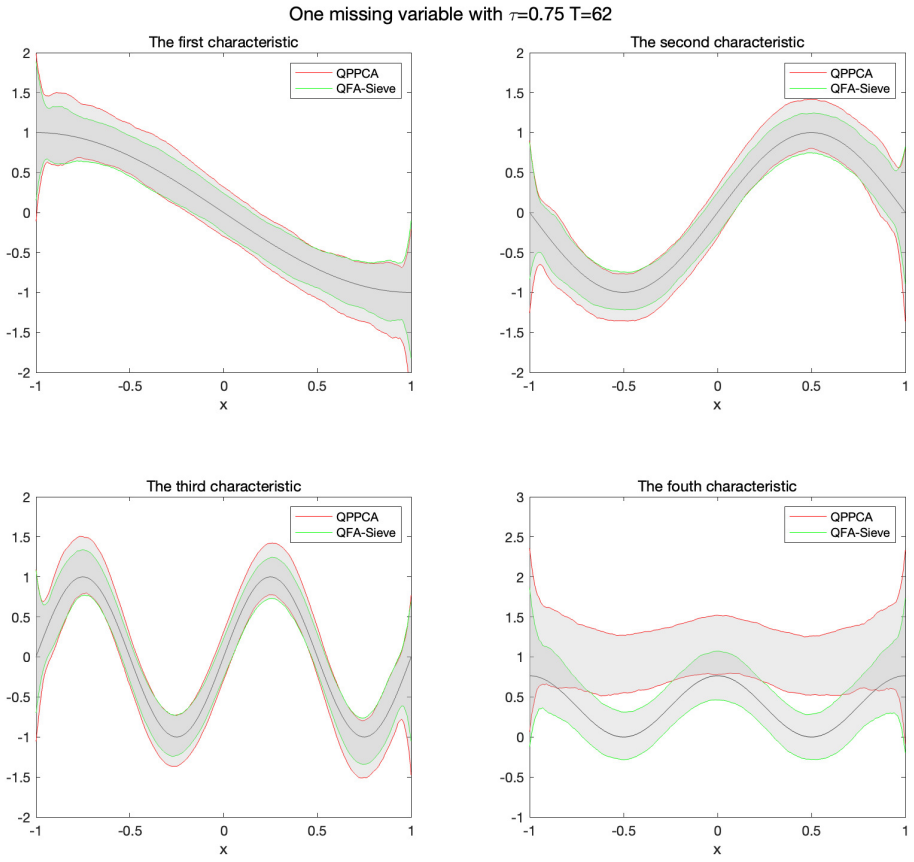
Note: $n = 355, T = 62, R = 1$ and $D = 4$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0,1)$, x_{id} $(i = 1, ..., n$ and $d = 1, 2, 3, 4)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black line) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure 7: Estimated Loading functions at $\tau = 0.25$, 1 missing variable: QPPCA and QFA-Sieve



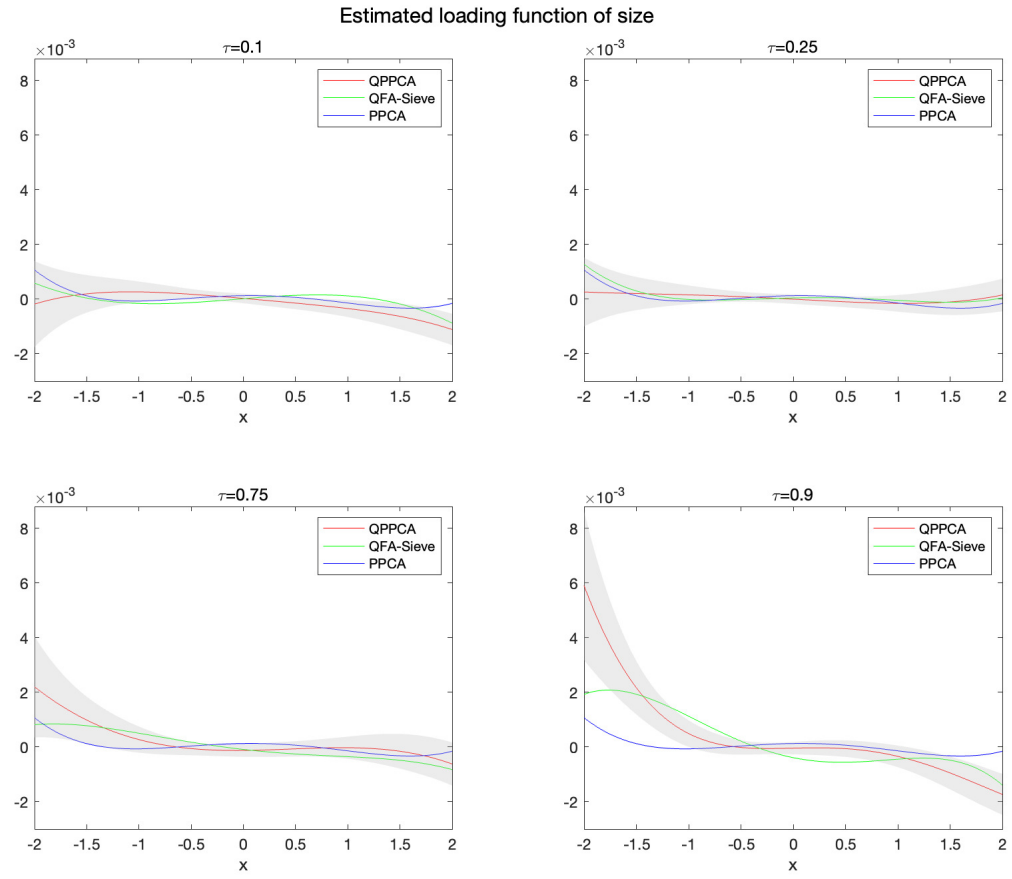
Note: $n = 355, T = 62, R = 1$ and $D = 5$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3}) + g_5(x_{i5})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0,1)$, x_{id} $(i = 1, ..., n$ and $d = 1, \ldots, 5)$ are independently drawn from the uniform distribution: $U[-1, 1]$. Assume that only the first 4 characteristics are observed while x_{i5} is unobserved. $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$ and $g_5(x) = 1.5cos(\pi x)$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black line) at $\tau = 0.25$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure 8: Estimated Loading functions at $\tau = 0.75$, 1 missing variable: QPPCA and QFA-Sieve



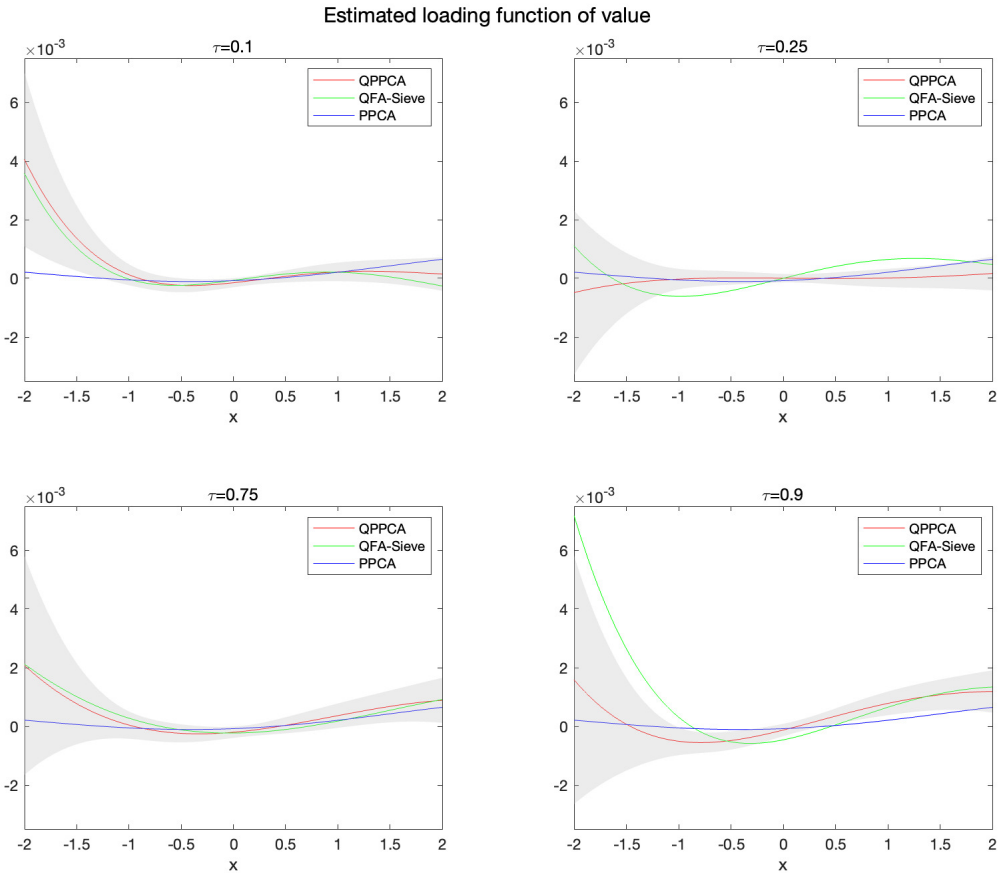
Note: $n = 355, T = 62, R = 1$ and $D = 5$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3}) + g_5(x_{i5})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0,1)$, x_{id} $(i = 1, ..., n$ and $d = 1, \ldots, 5)$ are independently drawn from the uniform distribution: $U[-1, 1]$. Assume that only the first 4 characteristics are observed while x_{i5} is unobserved. $g_1(x) = -sin(0.5\pi x)$, $g_2(x) = sin(\pi x)$, $g_3(x) = sin(2\pi x)$, $g_4(x) = cos^2(\pi x)$ and $g_5(x) = 1.5cos(\pi x)$ and $\{u_{it}\}$ are i.i.d draws from the $t(3)$ distribution. The graphs show the true loading functions (the black line) at $\tau = 0.75$, and the empirical point-wise 5% and 95% quantiles of the estimated loading functions using QPPCA (the red lines) and QFA-Sieve (the green lines) from 1000 repetitions.

Figure 9: Estimated loading functions using PPCA, QPPCA and QFA-Sieve



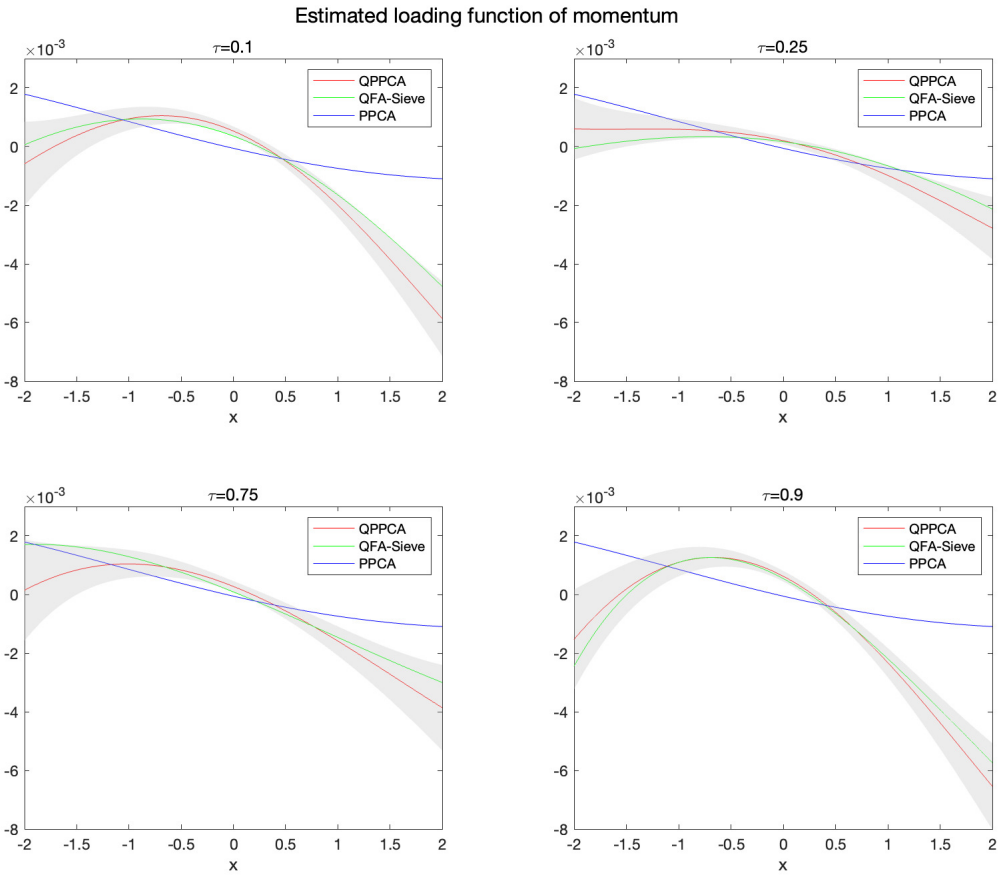
Note: This Figure plots the estimated quantile factor loading functions of **size** using QPPCA and QFA-Sieve at $\tau = 0.1, 0.25, 0.75, 0.9$. and the estimated loading function of **size** using PPCA. The shaded areas show the 95% point-wise confidence intervals constructed using the asymptotic distribution of the QPPCA estimator.

Figure 10: Estimated loading functions using PPCA, QPPCA and QFA-Sieve



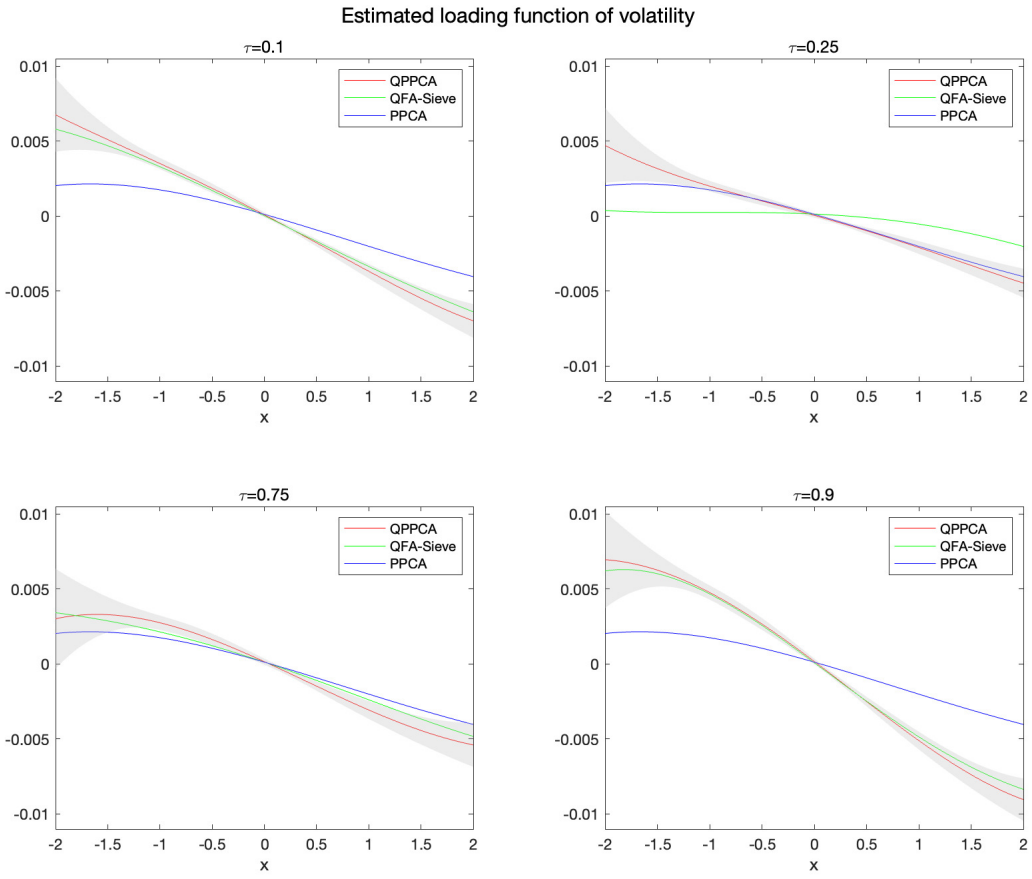
Note: This Figure plots the estimated quantile factor loading functions of **value** using QPPCA and QFA-Sieve at $\tau = 0.1, 0.25, 0.75, 0.9$. and the estimated loading function of **value** using PPCA. The shaded areas show the 95% point-wise confidence intervals constructed using the asymptotic distribution of the QPPCA estimator.

Figure 11: Estimated loading functions using PPCA, QPPCA and QFA-Sieve



Note: This Figure plots the estimated quantile factor loading functions of **momentum** using QPPCA and QFA-Sieve at $\tau = 0.1, 0.25, 0.75, 0.9$. and the estimated loading function of **momentum** using PPCA. The shaded areas show the 95% point-wise confidence intervals constructed using the asymptotic distribution of the QPPCA estimator.

Figure 12: Estimated loading functions using PPCA, QPPCA and QFA-Sieve



Note: This Figure plots the estimated quantile factor loading functions of **volatility** using QPPCA and QFA-Sieve at $\tau = 0.1, 0.25, 0.75, 0.9$. and the estimated loading function of **volatility** using PPCA. The shaded areas show the 95% point-wise confidence intervals constructed using the asymptotic distribution of the QPPCA estimator.

References

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

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

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

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1281–1304.

Chen, L., J. J. Dolado, and J. Gonzalo (2021). Quantile factor models. *Econometrica 89*(2), 875–910.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics 6*, 5549–5632.

Chen, X. and X. Shen (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 289–314.

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the Fama–French model and extensions. *Econometrica 80*(2), 713–754.

Connor, G. and R. A. Korajczyk (1993). A test for the number of factors in an approximate factor model. *the Journal of Finance 48*(4), 1263–1291.

Connor, G. and O. Linton (2007). Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance 14*(5), 694–717.

Elton, E. J. (1999). Expected return, realized return, and asset pricing tests. *The Journal of Finance 54*(4), 1199–1220.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fan, J., K. Li, and Y. Liao (2021). Recent developments in factor models and applications in econometric learning. *Annual Review of Financial Economics 13*, 401–430.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *Annals of Statistics 44*(1), 219.

Hausman, J., H. Liu, Y. Luo, and C. Palmer (2021). Errors in the dependent variable of quantile regression models. *Econometrica 89*(2), 849–873.

Horowitz, J. L. and S. Lee (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association 100*(472), 1238–1249.

Jung, S. and J. S. Marron (2009). Pca consistency in high dimension, low sample size context. *The Annals of Statistics 37*(6B), 4104–4130.

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

Lee, J. and P. M. Robinson (2016). Series estimation under cross-sectional dependence. *Journal of Econometrics 190*(1), 1–17.

Ma, S., O. Linton, and J. Gao (2021). Estimation and inference in semiparametric quantile factor models. *Journal of Econometrics 222*(1), 295–323.

Rosenberg, B. (1974). Extra-market components of covariance in security returns. *Journal of Financial and Quantitative Analysis 9*(2), 263–274.

Shen, D., H. Shen, and J. S. Marron (2013). Consistency of sparse pca in high dimension, low sample size contexts. *Journal of Multivariate Analysis 115*, 317–333.

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.