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*

May 13, 2025

Abstract

This paper studies the estimation of characteristics-based quantile factor models where the factor loadings are unknown functions of observed individual characteristics, and the idiosyncratic error terms are subject to conditional quantile restrictions. We propose a three-stage estimation procedure that is easily implementable 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. Our proposed estimators are 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 differs from the number of characteristics. Monte Carlo 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 a Co-editor, two anonymous referees and 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 MICIU/AEI /10.13039/501100011033 and grants TED2021-129784BI00, CEX2021-001181-M, 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, *approximate factor models* (**AFM**) treat factors and loadings as latent objects which can be only identified up to a rotation matrix.[1] On the other hand, there is a vast literature in finance, pioneered by Fama and French (1993), that instead explains the cross-sectional co-movements of stock returns on the basis of observable factors. These factors were initially approximated by differences among the returns of portfolios sorted by some observed characteristics, like *size risk*, *value risk* and *market risk*, to which other factors were later added, such as *momentum, profitability* and *investment* (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 lack of interpretation of the estimated factors. Conversely, while the Fama-French approach turns out to be unambiguous about this interpretation, 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 exploits the advantages of both approaches while avoiding their shortcomings is the so-called *characteristics-based factor models* (**CFM**) introduced by Rosenberg (1974), empirically applied by Daniel and Titman (1997), and later extended by Connor and Linton (2007) and Connor, Hagmann, and Linton (2012), which forms the bedrock of this paper. This procedure assumes the factor loadings are smooth functions of some observed characteristics of the different units, while factors are treated as unknown objects, like in AFM. Thereby, the latent factors in CFM can be easily estimated even when the number of factors is large, whereas their interpretation hinges on the choice of the characteristics variables. Subsequently, Fan, Liao, and Wang (2016) have generalized this framework by allowing both the number of factors to differ from the number of characteristics and the factor loadings to be not fully explained by those characteristics. To estimate this class of models, these authors propose a new methodology called *projected principal component analysis* (**PPCA**) which, by de-noising the data, yields estimators with faster rates of convergence than the conventional PCA estimators for AFM.

In line with the extension of conditional mean AFM to quantile factor models (**QFM**) (see Chen, Dolado, and Gonzalo (2021)), our main goal in this paper is to extend the above-mentioned CFM and PPCA to a new class of factor models labeled *characteristics-based quantile factor models* (**CQFM**). In this setup latent factors are allowed to shift specific features (moments or quantiles) of the joint distribution of the data and not only their means. Thus, the idiosyncratic

[1] AFM were first proposed by Chamberlain and Rothschild (1983) to characterize the co-movement of large sets 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.

errors in CQFM are subject to quantile restrictions while they are only subject mean restrictions in CFM. Moreover, the latent factors, the loading functions, and the number of factors in CQFM are all allowed to vary across quantiles. In this fashion they provide a more complete picture of how, for example, the joint distributions of many asset returns are driven by a few common risk factors. Inspired by PPCA, we develop a new three-stage estimation method for CQFM, labeled *quantile-projected principal components analysis* (**QPPCA**), which is computationally simpler than other available estimation procedures for this class 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. By doing that, we eliminate noise at each quantile. Second, using the quantile fitted values from the first step, PCA is applied at each quantile to estimate their factors and loadings. Third, the whole set of loading functions is retrieved by projecting the estimated loadings onto the basis of the sieve space. Additionally, we propose a novel estimator for the number of factors at each quantile which is shown to perform satisfactorily for reasonable sample sizes. Finally, as in Fan, Liao, and Wang (2016), 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 as either fixed or diverging.

Regarding the literature on CQFM, a clear precedent of our work is Ma, Linton, and Gao (2021) which, like us, also extend the ideas developed by Connor et al. (2012) to a quantile-restriction framework. In particular, they address the estimation and inference of CQFM by considering a *semiparametric quantile factor analysis* (**SQFA**) approach, where the observed characteristics are potentially allowed to affect stock returns in a nonlinear fashion. Yet, we differ from SQFA in several important dimensions. First, following Connor and Linton (2007) and Connor et al. (2012), SQFA relies on the restrictive assumption that the number of latent factors is known and 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 this number can be consistently estimated at each quantile. Second, while the initial estimator of the quantile loading functions in SQFA also relies on sieve quantile regressions, its subsequent steps to jointly estimate the factors and loadings are based on an alternate optimization algorithm which, as these authors acknowledge, can be computationally cumbersome. This is not the case with QPPCA since its second and third steps are based on PCA, whose computation is much simpler. 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, some differences in the assumptions imposed in these two approaches will be further discussed once the main theoretical results are presented below.

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 concretely, an advantage of QPPCA over PPCA that all its asymptotic properties are obtained without assuming any of the moment restrictions on the idiosyncratic errors that PPCA requires for proving its consistency; therefore it becomes a useful tool to analyze data from financial markets, income inequality and climate change where the error distributions are known to exhibit heavy tails. 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. In other words, not only QPPCA estimates turn out to be more robust to heavy tails and outliers but it will also be able to estimate factors that PPCA will miss, like e.g. factors in the variance or in higher moments.

Next, as already underscored, the CQFM models and the SQFA approach are also closely related to the quantile factor analysis (**QFA**) proposed by Chen, Dolado, and Gonzalo (2021) to estimate QFM. In effect, while no restrictions on the factor loadings are imposed in QFA (except for a standard rank condition), the loadings in CQFM become unknown smooth functions of some observed characteristics of the units. This implies that the model can be consistently estimated even when T is fixed, while consistency of the QFA estimators requires both n and T go to infinity. However, the choice of which characteristics to include in those functions entails the risk of mis-specification and, precisely for this reason, one of the most important contributions in Fan et al. (2016)' s PPCA is to deal with a more general setup where the cross-sectional variations of the factor loadings may not be fully explained by the observed characteristics. We also consider this general setup in the context of QFM. To do so, we propose an alternative estimation approach to QPPCA, labeled **QFA-Sieve**, which relies on a combination of QFA and sieve quantile regressions which allows for the existence of other characteristics, besides those initially chosen, that may explain the variations of the factor loadings. Yet, QFA-Sieve only works when T is large (see Section 4 and the simulations in Subsection 5.3). Hence, if the model is correctly specified and T is not large, QPPCA should be the preferred method.

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 widely used in other studies. We find that, unlike PPCA, QPPCA and QFA-Sieve allow to uncover substantial variations of the estimated loading functions across different quantiles.

The rest of the paper is organized as follows. Section 2 introduces the model and the proposed estimators. Section 3 derives their asymptotic properties and introduces a novel consistent estimator of the number of factors at each quantile. Section 4 presents a solution for the case where not all loadings are functions of observed characteristics. Section 5 provides several Monte Carlo simulation results for finite samples. Section 6 reports an empirical application. Finally, Section 7 concludes. An online appendix gathers detailed proofs of the theorems and some supplementary material on simulation results discussed in the main text.

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 $\{y_{it}\}_{1\le i\le n, 1\le t\le T}$, 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 $\mathsf{Q}_\tau[u_{it}(\tau)|\boldsymbol{\lambda}_i(\tau), \boldsymbol{f}_t(\tau)] = 0$. The dependence of R on τ has been suppressed to ease the notations.

Our focus in this paper is on the CQFM model, where the loadings $\boldsymbol{\lambda}_i(\tau)$ in (1) are assumed to be functions of some observed characteristics $\boldsymbol{x}_i = (x_{i1}, x_{i2}, .., x_{iD}) \in \mathbb{R}^D$. Specifically, for unit i, it is assumed that:

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

where $\boldsymbol{g}_\tau(\cdot) : \mathbb{R}^D \mapsto \mathbb{R}^R$ is a vector of unknown functions for each τ, and $\boldsymbol{\gamma}_i(\tau) \in \mathbb{R}^R$ represents the remaining part of $\boldsymbol{\lambda}_i(\tau)$ that cannot be explained by $\boldsymbol{x}_i$. To avoid the curse of dimensionality, we follow Fan et al. (2016) in assuming that the rth element of $\boldsymbol{g}_\tau(\boldsymbol{x}_i)$ is given by the following additive function, which is plausible when no strong interactions between the different characteristics in the loading functions are present:

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

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 daily returns in a given quarter, as in Fan et al. (2016), the characteristics could 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 $\boldsymbol{g}_\tau(\boldsymbol{x}_i)$, $\boldsymbol{\Gamma}_\tau$ be the $n \times R$ matrix of $\boldsymbol{\gamma}_i(\tau)$, $\boldsymbol{U}_\tau$ be the $n \times T$ matrix of $u_{it}(\tau)$. Then, models (1) and (2) can be rewritten in compact matrix form as:

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

As already highlighted, the above setup is more general than those considered in Connor et al. (2012) and Ma et al. (2021) in terms of how the unobserved factor loadings are related to the observed characteristics. In the latter, the dimension of the vector of 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 rather 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 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, our setup provides an

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

easier tool to estimate CQFM in such a case since the number of factors can be quite smaller than the number of characteristics. Lastly, both papers mentioned above assume that the factor loadings are fully explained by the observed characteristics, i.e., $\boldsymbol{\gamma}_i(\tau) = 0$ for all i, which can also be too restrictive in many applications.

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), our 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, for example, our approach 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.

In the rest of this section, two estimation methods of the CQFM will be introduced. The first method (QPPCA) deals with the case where the factor loadings are fully explained by the observed characteristics, while the second method (QFA-Sieve) applies to the more general case where $\boldsymbol{\gamma}_i \neq 0$ in (2). To simplify the notations even further, we suppress the τ-subscripts in the sequel and use $\boldsymbol{g}(), \boldsymbol{\lambda}_i, \boldsymbol{f}_t, \boldsymbol{\gamma}_i, \boldsymbol{G}(\cdot), \boldsymbol{F}, \boldsymbol{U}, \boldsymbol{\Gamma}$ instead of $\boldsymbol{g}_\tau(\cdot), \boldsymbol{\lambda}_i(\tau), \boldsymbol{f}_t(\tau), \boldsymbol{\gamma}_i(\tau), \boldsymbol{G}_\tau(\cdot), \boldsymbol{F}_\tau, \boldsymbol{U}_\tau, \boldsymbol{\Gamma}_\tau$.

2.2 The QPPCA Estimators

In this subsection, we focus on the case where the factor loadings are fully explained by the observed characteristics, i.e. $\boldsymbol{\gamma}_i = 0$ for all i. Write $\theta_t(\boldsymbol{x}_i) = \boldsymbol{g}(\boldsymbol{x}_i)'\boldsymbol{f}_t = \sum_{r=1}^R g_r(\boldsymbol{x}_i) f_{tr} = \sum_{r=1}^R (\sum_{d=1}^D g_{rd}(x_{id})) f_{tr}$. Let Θ be a space of continuous functions such that $\theta_t \in \Theta$ 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:

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

where $\mathcal{X} \subset \mathbb{R}^D$ is the support of $\boldsymbol{x}_i$, $\phi_1, \ldots, \phi_{k_n}$ is a set of continuous basis functions, and $a_{11}, \ldots, a_{k_n D}$ are coefficients, with k_n denoting the number of basis functions used to approximate $\boldsymbol{g}(\boldsymbol{x})$, namely, the dimension of the sieve space. Then 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 $r = 1, \dots, R$, there exist a set of coefficients $\boldsymbol{b}_{01}, \dots, \boldsymbol{b}_{0R} \in \mathbb{R}^{Dk_n}$ 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 $\boldsymbol{B}_0 = (\boldsymbol{b}_{01}, \dots, \boldsymbol{b}_{0R}) \in \mathbb{R}^{Dk_n \times R}$ and $\boldsymbol{a}_{0t} = \boldsymbol{B}_0 \boldsymbol{f}_t$, we have that $\boldsymbol{a}_{0t}' \boldsymbol{\phi}_{k_n}(\cdot) \in \Theta_n$ for all t and

$$\max_{1 \le t \le T} \sup_{\boldsymbol{x} \in \mathcal{X}} \left| \boldsymbol{a}_{0t}' \boldsymbol{\phi}_{k_n}(\cdot) - \theta_t(\boldsymbol{x}) \right| = O(k_n^{-\alpha}). \tag{5}$$

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

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

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

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

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

Step 2: Write $\hat{y}_{it} = \hat{\theta}_t(\boldsymbol{x}_i) = \hat{\boldsymbol{a}}_t' \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)$ 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 in PCA, 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 $r = 1, \dots, R$. Define $\hat{\boldsymbol{A}} = (\hat{\boldsymbol{a}}_1, \dots, \hat{\boldsymbol{a}}_T)$. 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 $\boldsymbol{G}(\boldsymbol{X}) \approx \boldsymbol{\Phi} \boldsymbol{B}_0$ by (4),

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

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

with $\boldsymbol{A}_0 = (\boldsymbol{a}_{01}, \ldots, \boldsymbol{a}_{0T}) = \boldsymbol{B}_0\boldsymbol{F}' \in \mathbb{R}^{Dk_n \times T}$ and $\boldsymbol{\Phi}$ be the $n \times Dk_n$ matrix of $\boldsymbol{\phi}_{k_n}(\boldsymbol{x})$. The first-step quantile projection purges 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 leads to 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 $\boldsymbol{F}$ at each quantile 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 proposed 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 Chen et al. (2021)' s quantile factor analysis (QFA) 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.

2.3 The QFA-Sieve Estimators

As already mentioned, an interesting extension of Fan et al. (2016) with respect to Connor et al. (2012) for QFM is to allow for some of the factor loadings not to depend on the chosen set of observed characteristics. In this subsection, we consider the same generalization with respect to Ma et al. (2021) for CQFM, namely, $\boldsymbol{\gamma}_i \neq 0$ in (2).

Note that model (1) and (2) can be written as:

$$y_{it} = \boldsymbol{g}(\boldsymbol{x}_i)'\boldsymbol{f}_t + \tilde{u}_{it} \quad \text{where} \quad \tilde{u}_{it} = u_{it} + \boldsymbol{\gamma}_i'\boldsymbol{f}_t. \tag{7}$$

As shown by Fan et al. (2016) for CFM, when u_{it} has zero mean, the mean of the new error term $\tilde{u}_{it}$ is also equal to 0 as long as $\boldsymbol{\gamma}_i$ is properly normalized. Therefore, the presence of extra components in the factor loadings does not invalidate the mean restrictions in CFM, implying that their PPCA method works well irrespective of whether $\boldsymbol{\gamma}_i = 0$ or not. However, allowing for this more general setup in the context of QFM would be very challenging. The insight is that $\tilde{u}_{it}$ no longer satisfy the conditional quantile restrictions imposed on u_{it}, even when $\boldsymbol{x}_i$ and $\boldsymbol{\gamma}_i$ are independent, since imposing quantile conditional restrictions for the two elements of $\tilde{u}_{it}(\tau)$ does not imply that this restriction should hold for their sum.[4] Consequently, Step 1 of the QPPCA method will not produce consistent estimators of $\boldsymbol{g}(\boldsymbol{x}_i)'\boldsymbol{f}_t$.

To address this problem, we consider an alternative approach labeled **QFA-Sieve** which works as follows: first, the quantile factor loadings $\{\boldsymbol{\lambda}_i\}_{i=1}^n$ at $\tau \in (0,1)$ are estimated using the QFA approach of Chen et al. (2021):

$$(\check{\boldsymbol{\lambda}}_1, \ldots, \check{\boldsymbol{\lambda}}_N; \check{\boldsymbol{f}}_1, \ldots, \check{\boldsymbol{f}}_T) = \arg\min_{\boldsymbol{\Lambda}, \boldsymbol{F}} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - \boldsymbol{\lambda}_i' \boldsymbol{f}_t),$$

such that $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}/N$ is diagonal and $\boldsymbol{F}'\boldsymbol{F}/T = \boldsymbol{I}_R$. Second, the estimated loadings are projected on the observed characteristics using least square sieves to obtain estimators of $\boldsymbol{g}(\cdot)$. Specifically, let $\boldsymbol{\Phi} \in \mathbb{R}^{n \times Dk_n}$ be the matrix of the basis functions $\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i), i = 1, \ldots, n$ defined in Section 2.2, and $\check{\boldsymbol{\Lambda}}$ be the $n \times R$ matrix of $\check{\boldsymbol{\lambda}}_i$, then

$$\check{\boldsymbol{G}}(\boldsymbol{X}) = \boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\check{\boldsymbol{\Lambda}} \quad \text{and} \quad \check{\boldsymbol{g}}(\boldsymbol{x}) = \check{\boldsymbol{\Lambda}}'\boldsymbol{\Phi}(\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\phi}_{k_n}(\boldsymbol{x}).$$

The intuition for this procedure is as follows: since the first-step QFA estimators are consistent (see Chen et al. (2021)) for all the factor loadings embedded in the true factor structure, the least-squares projection in the second step will also yield consistent estimators of the loading functions under very general conditions by treating $\boldsymbol{\gamma}_i$ as the regression errors. Hence, a comparison of QFA-Sieve with QPPCA provides a heuristic test of the correct specification of the chosen set of characteristics in the latter procedure,

3 Asymptotic Properties of the QPPCA 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.

[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)).

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

$$P[u_{it} \leq 0 | \boldsymbol{x}_i = \boldsymbol{x}] = \tau \text{ for any } \boldsymbol{x} \in \mathcal{X}. \tag{8}$$

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 this finite-T-consistency result to CQFM.

3.1 Rates of convergence

Suppose that the observed data $\{y_{it}\}$ are generated by (3) and that $\{u_{it}\}$ satisfy (8). Let

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

The following set of conditions are required to establish the uniform rate of convergence of $\hat{\theta}_1, \ldots, \hat{\theta}_T$ (see Proposition 1 of the online appendix), 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) $k_n \to \infty$, $k_n^2/n \to 0$ and $\varepsilon_{nT} \to 0$ as $n \to \infty$.

Assumption 1(i) can be relaxed to allow for weak cross-sectional dependence though e.g. nonstationary error terms are excluded. 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, and the results of this section will still hold. 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). Assumptions 1(iii) and 1(iv) are also standard in sieve quantile regressions, though the latter imposes very mild restrictions on the size of T when it goes to infinity jointly with n.

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) There exists a constant $c > 0$ such that $\lambda_{\min}(\boldsymbol{F}'\boldsymbol{F}/T) > c$ for all T.
(iii) $\hat{\boldsymbol{\Sigma}}_g \equiv n^{-1}\sum_{i=1}^n \boldsymbol{g}(\boldsymbol{x}_i)\boldsymbol{g}(\boldsymbol{x}_i)' \xrightarrow{P} \boldsymbol{\Sigma}_g > 0$ as $n \to \infty$.
(iv) 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) 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 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 that assumption in 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). However, 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 $(u_{i1}, \ldots, u_{iT})$.

In particular, if the following set of assumptions are added to Assumptions 1 and 2, then it is possible to show that QPPCA achieves faster convergence rates, which are comparable to those of PPCA.

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) $|\mathsf{f}(c) - \mathsf{f}(0)| \le L|c|$ for any c in a neighborhood of 0.
(iii) Equation (4) holds for some $\alpha \ge 3$.

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 $\gamma_1 \ge 0$ 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 3. *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) $|\mathsf{f}(c|\boldsymbol{x}) - \mathsf{f}(0|\boldsymbol{x})| \le L|c|$ 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 $0 < c_1 \le \lambda_{\min}(\boldsymbol{\Sigma}_{\mathsf{f}\phi}) \le \lambda_{\max}(\boldsymbol{\Sigma}_{\mathsf{f}\phi}) \le c_2 < \infty$ 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 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 4. *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 5. *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.

$$\hat{R} = \sum_{j=1}^{\bar{R}} \mathbf{1}\{\hat{\rho}_j > p_n\}, \tag{9}$$

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:

$$p_n = d \cdot \hat{\rho}_1^{1/2} \cdot n^{-1/4} \ln T, \tag{10}$$

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

Remark 6. *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.*

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

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.

4 Asymptotic Properties of the QFA-Sieve Estimators

Since the QFA method of Chen et al. (2021) produces consistent estimates of R and $\boldsymbol{F}$, in this section we focus on the asymptotic properties of the estimated factor loading functions using the QFA-Sieve method: $\check{\boldsymbol{G}}(\boldsymbol{X})$ and $\check{\boldsymbol{g}}(\cdot)$.

Assumption 6. *Let $\{l_{nT}\}$ be a sequence of positive constants that goes to 0 as n, T diverge, and let $\boldsymbol{H} \in \mathbb{R}^{R\times R}$ be a positive-definite matrix with $\|\boldsymbol{H}\| < \infty$.*
(i) Let $\check{\boldsymbol{\Lambda}}$ be the estimated quantile factor loadings satisfying: $\|\check{\boldsymbol{\Lambda}} - \boldsymbol{\Lambda}\boldsymbol{H}\|/\sqrt{n} = O_P(l_{nT})$;
(ii) The basis functions $\{\boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i), i = 1, \ldots, n\}$ satisfy Assumption 2(i);
(iii) $\|n^{-1/2}\sum_{i=1}^{n} \boldsymbol{\phi}_{k_n}(\boldsymbol{x}_i)\boldsymbol{\gamma}_i'\| = O_P(\sqrt{k_n})$.

Despite only considering the QFA estimator of the loadings in the first step of the algorithm, Assumption 6(i) is a general assumption on the estimated quantile factor loadings, implying that our results below also apply to other consistent estimators of $\boldsymbol{\Lambda}$ as those momentarily discussed. For the QFA estimator, Chen et al. (2021) show that this condition is satisfied with $l_{nT} = n^{-1/2} \vee T^{-1/2}$ and $\boldsymbol{H} = \boldsymbol{\Upsilon}'\boldsymbol{\Sigma}_F^{-1/2}$, where $\boldsymbol{\Sigma}_F = \boldsymbol{F}'\boldsymbol{F}/T$, $\boldsymbol{\Sigma}_\Lambda = \boldsymbol{\Lambda}'\boldsymbol{\Lambda}/n$, and $\boldsymbol{\Upsilon}$ is the matrix of eigenvectors of $\boldsymbol{\Sigma}_F^{1/2}\boldsymbol{\Sigma}_\Lambda\boldsymbol{\Sigma}_F^{1/2}$.[5] Alternatively, one can use the methods proposed by Feng (2024) or Chen and Feng (2025) to obtain the initial estimator of $\boldsymbol{\Lambda}$. The nuclear-norm-penalized estimator of Feng (2024) features an objective function that is convex in the factors and their loadings, but its convergence rate is slower than the QFA estimator of Chen et al. (2021). On the other hand, the estimator of Chen and Feng (2025) requires stronger assumptions on the quantile factors and its computation is more complicated but its convergence rate is very close to that of the QFA estimator, i.e. $l_{nT} = n^{-1/2} \vee (\log T/T)^{1/2}$. However, it should be noted that, while all the above-mentioned methods require both n and T going to infinity to prove the consistency of $\check{\boldsymbol{\Lambda}}$, the QPPCA method only needs $n \to \infty$. Finally, Assumption 6(ii) is standard in sieve regressions, while Assumption 6(iii) holds under quite general conditions, e.g. when $\mathbb{E}[\boldsymbol{\gamma}_i|\boldsymbol{x}_i] = 0$ and the correlations of $(\boldsymbol{x}_i, \boldsymbol{\gamma}_i)$ across i are sufficiently weak.

Theorem 6. *Suppose Assumption 6 holds and $\epsilon_n = \sqrt{k_n/n} \vee k_n^{-\alpha}$. Then we have:*

[5] Note that $\boldsymbol{H} = \boldsymbol{I}$ if $\boldsymbol{\Sigma}_\Lambda$ is diagonal and $\boldsymbol{\Sigma}_F = \boldsymbol{I}$.

(i) $\|\check{\boldsymbol{G}}(\boldsymbol{X}) - \boldsymbol{G}(\boldsymbol{X})\boldsymbol{H}\|/\sqrt{n} = O_P(\epsilon_n) + O_P(l_{nT})$;
(ii) $\sup_{\boldsymbol{x}\in\mathcal{X}} \|\check{\boldsymbol{g}}(\boldsymbol{x}) - \boldsymbol{H}'\boldsymbol{g}(\boldsymbol{x})\| = O_P(\sqrt{k_n}\varepsilon_n) + O_P(\sqrt{k_n}l_{nT})$.

5 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 and QFA-Sieve procedures regarding the estimation of the number of factors, the factors themselves, and their loading functions. In most cases, unless otherwise explicitly stated, the characteristics $\{x_{id}, d = 1, \ldots, D\}$, are drawn independently from the uniform distribution: $U[-1, 1]$.

5.1 Estimating the number of factors

Consider the following DGP with $R = 3$ and $D = 5$:

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

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

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

The idiosyncratic components, u_{it}, are i.i.d. draws from three alternative distributions: (i) the standard normal, $N(0, 1)$, (ii) Student's t distribution with 3 degrees of freedom, $t(3)$, and (iii) the standard Cauchy distribution, Cauchy(0,1). To implement the rank minimization estimator for the number of factors in (9), the threshold p_n is chosen as in (10), with $d = 1/4$. Finally, 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 for this and remaining simulations reported in this section.

Table 1 displays 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 $\hat{R} > R$]. For comparison, Table B1 in the online appendix 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.

[Table 1 about here]

The main takeaway from these simulation exercises is that both selection criteria accurately estimate the number of factors when T is small and n is large, in line with their consistency property for fixed T. Nonetheless, when $n < 10000$, the rank minimization estimator outperforms the eigen-ratio estimator.

5.2 Estimating the factors

Following Chen et al. (2021), we next consider the following DGP with $D = R = 2$ in the whole subsection:

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

where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. Now we let $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$, such that

$$\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2}), \quad \lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2}),$$

with the error terms u_{it} being generated as in Subsection 5.1. Note that this DGP includes one mean factor, f_{t1}, affecting the mean of y_{it}, and one scale-shift factor f_{t2} that affects its variance.

5.2.1 Comparison of QPPCA and QFA-Sieve with PCA and PPCA

As shown in Chen et al. (2021), PCA and QPPCA are unable to capture the scale factor f_{t2}. Thus, our focus here will be on the estimation of the location factor, f_{t1}. In this subsection, four competing estimation procedures are considered: (i) the proposed QPPCA method with $\tau = 0.5$; (ii) the proposed QFA-Sieve method with $\tau = 0.5$;; (iii) the projection estimator proposed by Fan et al. (2016) (PPCA); (iv) the standard estimator of Bai and Ng (2002) for AFM (PCA). Regarding the sample sizes, we fix $T = 10, 50$ and let n increase from 50 to 500. For each method, the number of location factors ($R = 1$) is assumed to be known, and the average Frobenius error is reported as a measure of fit: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}\|/\sqrt{T}$ from 1000 replications.

The results are plotted in Figure 1. On the one hand, for small T ($T = 10$), the three graphs on the left panel show that PCA and QFA-Sieve perform worse than PPCA and QPPCA when u_{it} is either drawn from the $N(0, 1)$ or $t(3)$ distributions. Moreover, when the distribution is a standard Cauchy, QPPCA beats all of its competitors in line with its consistency even when T is fixed or the moments of u_{it} do not exist. On the other hand, when T is larger ($T = 50$), the three graphs on the right panel show that all methods perform similarly when the errors are $N(0, 1)$, while QPPCA and QFA-Sieve behave much better as the distribution of the errors becomes heavy-tailed. The estimation results for the loading matrix (see Figure B1 in the online appendix) are very similar.

[Figure 1 about here]

In summary, the previous set of results show that, when estimating the mean factors: (i) QPPCA is consistent even for small T, and (ii) QPPCA yields more robust estimators to outliers than PCA and PPCA, and (iii) similar findings hold for QFA-Sieve when T is sufficiently large.

5.2.2 Comparison of QPPCA and QFA-Sieve with SQFA

To compare QPPCA and QFA-Sieve with its closest competitor SQFA, we next consider the same DGP as before with the three estimation procedures being applied for $\tau = 0.25, 0.75$. Again, we fix $T = 10, 50$ and let n increase from 50 to 500. For each procedure, the number of factors ($R = 2$) is assumed to be known, while the average Frobenius errors from 1000 repetitions are reported as a measure of fit: $\|\boldsymbol{F} - \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}'\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}'\boldsymbol{F}\|/\sqrt{T}$ and $\|\boldsymbol{G} - \hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-1}\hat{\boldsymbol{G}}'\boldsymbol{G}\|/\sqrt{n}$.

Figure 2 displays the estimation errors for the factors at $\tau = 0.25$, while those at $\tau = 0.75$ are presented in Figure B2 in the online appendix since they are qualitatively similar. Likewise, the corresponding errors for the loading matrix are shown in Figures B3 and B4 in the online appendix to save space. Two main findings stand out. First, for small T ($T = 10$), QFA-Sieve performs much worse than QPPCA in all instances because its consistency relies on both n and T diverging. Second, QPPCA outperforms SQFA in nearly all cases. This result is neither surprising since the two loadings depend on both characteristics, therefore violating the basic assumption of SQFA requiring each loading to depend on a single characteristic.

Finally, we have run additional simulations when $D > R$ ($D = 3$ and $R = 2$) and $D < R$ ($D = 1$ and $R = 2$), which can also be found in the online appendix B (Figures B5 to B12). All in all, we find that as long as some of the loadings are allowed to depend on multiple characteristics (which must be the case whenever $D > R$), QPPCA beats SFQA by a large margin.

[Figure 2 about here]

5.3 Estimating the loading functions

We next focus on the estimation of the loading functions. Motivated by the available sample sizes in the empirical application in the next section, we set $n = 355, T = 62, R = 1$ in 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 $N(0,1)$. We restrict attention here to how QPPCA and QFA-Sieve perform in estimating $g_j(\cdot)$, $j = 1, \ldots, 4$ at $\tau = 0.25, 0.75$.

The results are presented in Figure 3 for $\tau = 0.25$ while, as before, those for $\tau = 0.75$ are displayed in the Figure B13 of the online appendix. The black lines are the true loading functions, the red solid lines give the 95% and 5% point-wise empirical quantiles of the estimated functions using QPPCA from 1000 replications, and the green dashed lines provide the results for QFA-Sieve. It can be inspected the estimated functions by both methods trace the true functions very accurately, and that confidence intervals between the 5% and 95% quantiles are very narrow. The similarity of the loading functions estimated by these two procedures indicates that no missing characteristics are relevant.

[**Figure 3 about here**]

Finally, to compare our two proposed methods to SQFA in the case of omitted characteristics, the following DGP is considered:

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

where the only difference with the previous DGP is that that only the first four characteristics are now assumed to be observed while x_{i5} is treated as unobserved, all being drawn independently from $U[-1, 1]$.

The results are reported in Figure 4 for $\tau = 0.25$ and in Figure B14 of the online appendix for $\tau = 0.75$. It can be seen that, in contrast to QPPCA, which fails to consistently estimate the $g_4(\cdot)$ function when x_{i5} is omitted, QFA-Sieve yields much better estimates of this function, supporting our claim that this method is robust against unobserved or missing characteristics. Moreover, in contrast to Figure 3, the difference between the two estimators of $g_4(\cdot)$ can be interpreted this time as being supportive to the existence of missing characteristics.

Two additional results are noteworthy. First, due to the presence of unobserved characteristics, the confidence intervals of QFA-Sieve are much wider than when they are not present. Second, as discussed in Section 4, QFA-Sieve requires a sufficiently large T to correctly deal with this case. The simulation results of Chen et al. (2021) show that the QFA estimator performs well for $T \geq 50$, implying that the first-step estimation error of QFA-Sieve is very small for the time sample size ($T = 62$) considered in this subsection. Conversely, the additional simulations shown in the online appendix (Figures B15 and B16) illustrate that QFA-Sieve does not fare well when T is small. In particular, for $T = 10$, the confidence intervals of QPPCA contain the true loading functions, with their widths shrinking as n increases. By contrast, the corresponding intervals of the QFA-Sieve estimators much wider and do not shrink as n increases.

[**Figure 4 about here**]

6 Empirical Application

In this section, QPPCA is applied to investigate the factor structure of security returns. As in 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.[6] The sample consists of 355 stocks, whose book value and market capitalization are drawn from Compustat. Moreover, as is standard in this literature, we use the 1-month US treasury bond rate 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, Chebyshev polynomials of the first kind are used as the basis functions in the sieve regressions with $k_n = 4$.

First, Table 2 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 as well as only one factor at each quantile.

[**Table 2 about here**]

Second, Table 3 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 and the QPPCA factors are highly correlated. Combined with the results in Table 2, this evidence points to the existence of a single factor that shifts the means and quantiles of the joint distribution of the returns. Not surprisingly, as shown in Table 4, this factor is closely related with the market factor, with correlations ranging from 0.89 at $\tau = 0.1$ to 0.98 at $\tau = 0.5$.

[**Tables 3 and 4 about here**]

Third, Figure 5 to Figure 8 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, in light of the simulation results in Section 5.3, allow us to detect missing characteristics. The

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

similarity of these plots suggest that all characteristics are observed. Moreover, the estimated loading functions using PPCA (blue lines) are also included for comparison.

[Figures 5 to 8 about here]

A few salient findings emerge from this empirical exercise. First, for each characteristic, the quantile variation of their loading functions is larger at the more extreme quantiles (i.e. $\tau = 0.1, 0.9$). For example, the loading function at $\tau = 0.9$ is decreasing in the "size" characteristic (Figure 13), while it is flat elsewhere, indicating therefore that the impact of this variable on the factor loadings mainly affects the upper quantiles of the returns. Second, in most instances, the estimated loading functions using QFA-Sieve lie within the confidence intervals of the QPPCA estimators, pointing to no missing characteristics being relevant. Moreover, while the loading functions of "size" and "volatility" (Figures 13 and 16) behave monotonically at all quantiles, those of "value" and "momentum" (Figures 14 and 15) look U-shaped and inverted U-shaped, respectively. Interestingly, the shapes of our loading function resemble those reported by Ma et al. (2021) at $\tau = 0.5$ except for "value" and "volatility" which these authors find to be increasing and decreasing, respectively, instead of U-shaped and inverted U-shaped.

In sum, this empirical evidence points out that the estimated loading functions vary substantially across different quantiles, a feature that PPCA cannot uncover.

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 QPPCA. One could interpret $\hat{y}_{it}(\tau)$ as the "quantile return" which is interesting from the perspective of financial 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.

7 Conclusions

This paper proposes a three-stage estimation method for characteristics-based quantile factor models (CQFM). The convergence rates of the proposed estimator based on quantile-projected principal component analysis (QPPCA), are established, and the asymptotic distributions of the estimated factors and loading functions are derived under very general conditions. Compared to the existing estimation methods of CQFM, QPPCA estimators are not only easier to compute, but also they are consistent for fixed T (as long as n goes to infinity) and robust to heavy tails and outliers in the distribution of the idiosyncratic errors. Moreover, the number of quantile factors are allowed to differ from the number of the characteristics, and this number can be consistently estimated using a novel rank-minimization estimator that it is also proposed. Finally, we develop an alternative procedure, labeled as QFA-Sieve, that yields consistent estimators of the loading functions when T is large and provides a useful diagnostic tool to detect the presence of unobserved characteristics in the factor structure of 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 in financial markets and that, in some cases, these functions exhibit considerable variations across quantiles. Furthermore, we do not find evidence of other omitted relevant characteristics besides the ones used in this application. 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 (10) 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: 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 (10).

Table 3: 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

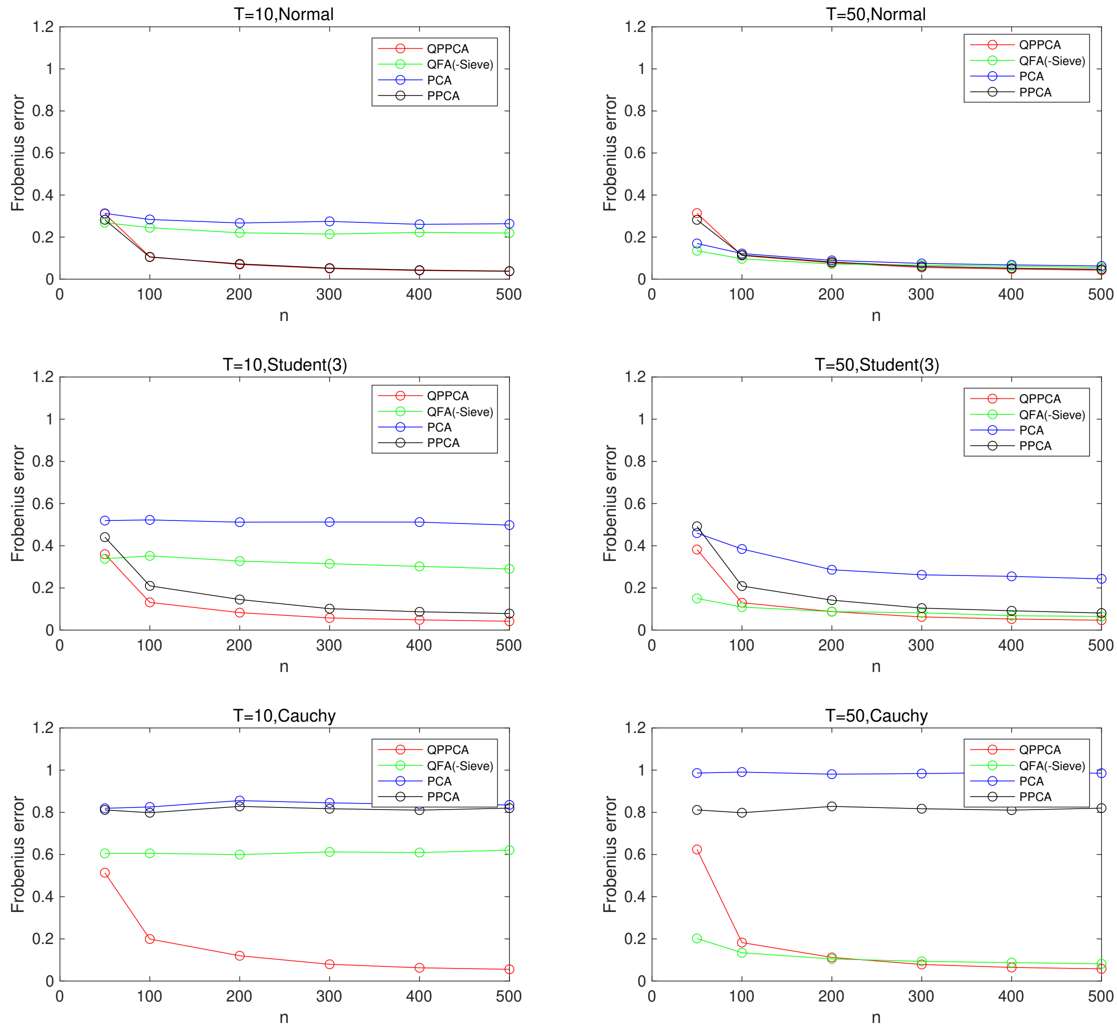
Note: This Table shows the correlations and sample means of the estimated mean factor using PPCA and the estimated quantile factors at different τs using QPPCA.

Table 4: Correlations of estimated factors and market factor

	τ =0.1	τ =0.25	τ =0.5	τ =0.75	τ =0.9	PPCA
Mean return	0.907	0.973	0.992	0.980	0.937	0.997
Market factor	0.893	0.957	0.982	0.958	0.905	0.974

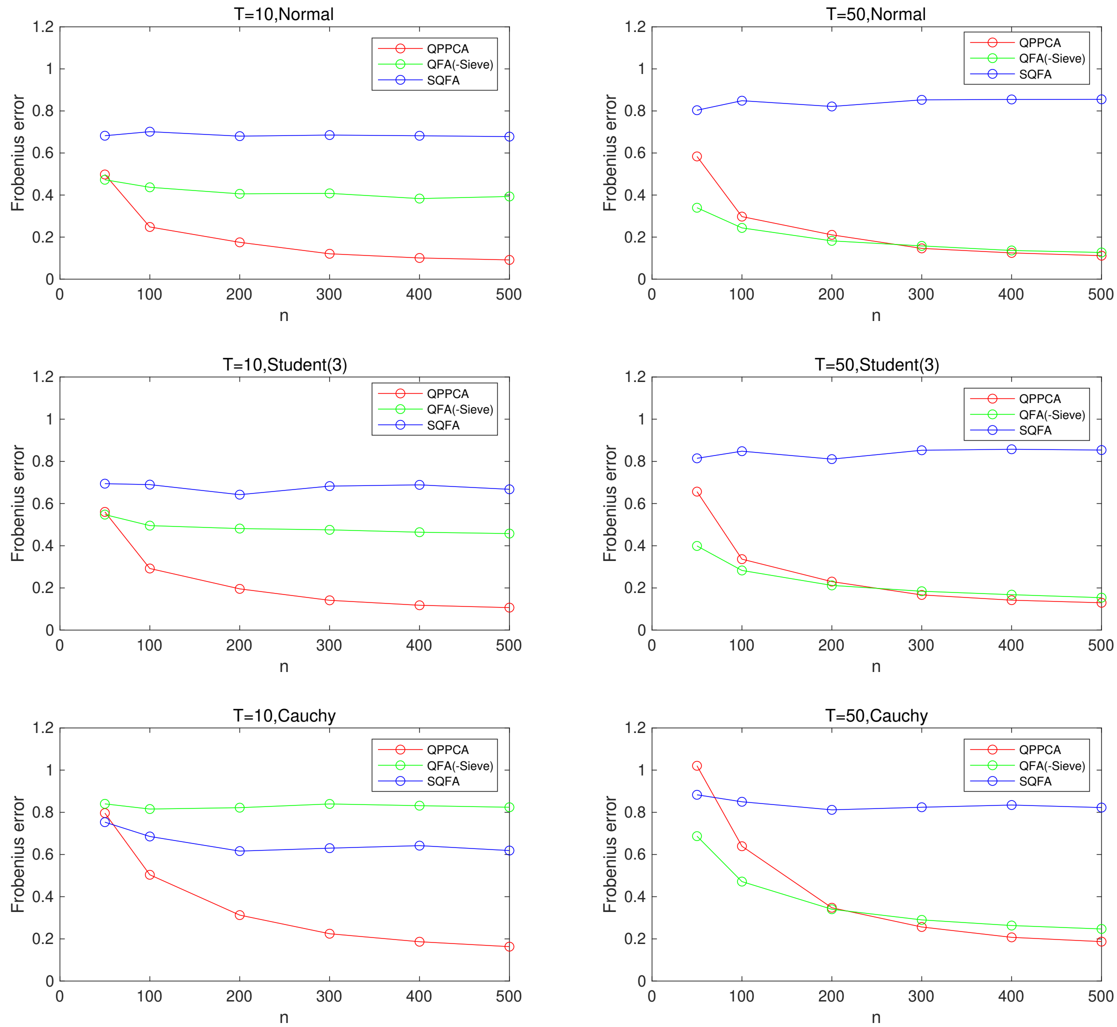
Note: This Table shows the correlations between the market factor and the estimated mean factor using PPCA and the estimated quantile factors at different τs using QPPCA.

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



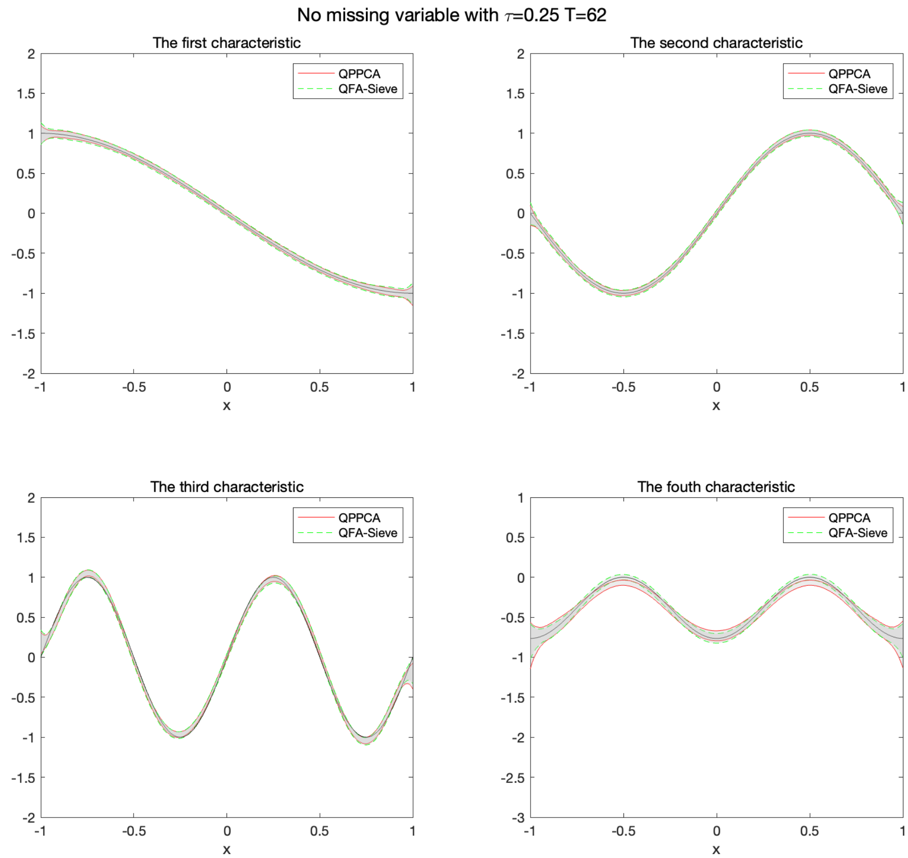
Note: The DGP is $Y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0, 1)$. The number of characteristics is 2 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2)$ are independently drawn from the uniform distribution: $U[-1, 1]$. $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$. The factor loading functions are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two factors are estimated by 3 methods: PCA, PPCA, QPPCA and QFA-Sieve 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-Sieve and SQFA, at $\tau = 0.25$, with $D = R$.



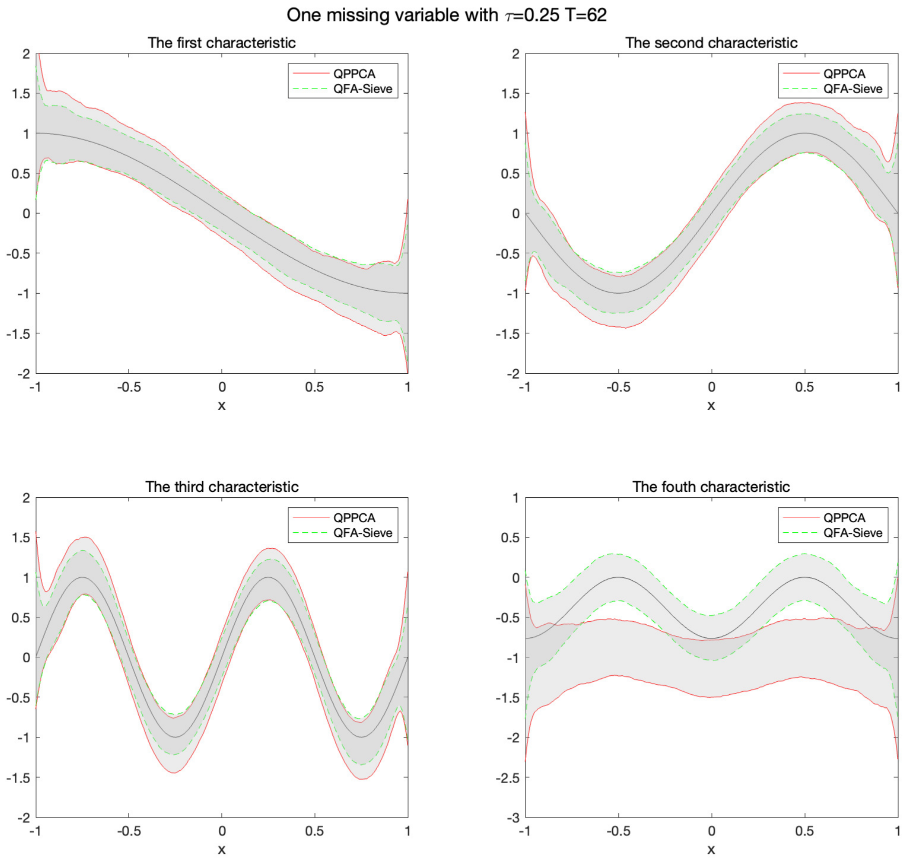
Note: The DGP is $Y_{it} = \lambda_{i1} f_{t1} + (\lambda_{i2} f_{t2}) u_{it}$, where $f_{t2} = |h_t|, f_{t1}, h_t \sim i.i.d\ N(0,1)$. The number of characteristics is 2 and all characteristics x_{id} $(i = 1, ..., n$ and $d = 1, 2)$ are independently drawn from the uniform distribution: $U[-1,1]$. $g_{11}(x) = g_{12}(x) = sin(2\pi x)$ and $g_{21}(x) = g_{22}(x) = cos^2(\pi x)$. The factor loading functions are generated as $\lambda_{i1} = g_{11}(x_{i1}) + g_{12}(x_{i2})$, $\lambda_{i2} = g_{21}(x_{i1}) + g_{22}(x_{i2})$. $\{u_{it}\}$ are i.i.d draws from three different distributions. The two 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: Estimated Loading functions at $\tau = 0.25$, no missing variable: QPPCA and QFA-Sieve



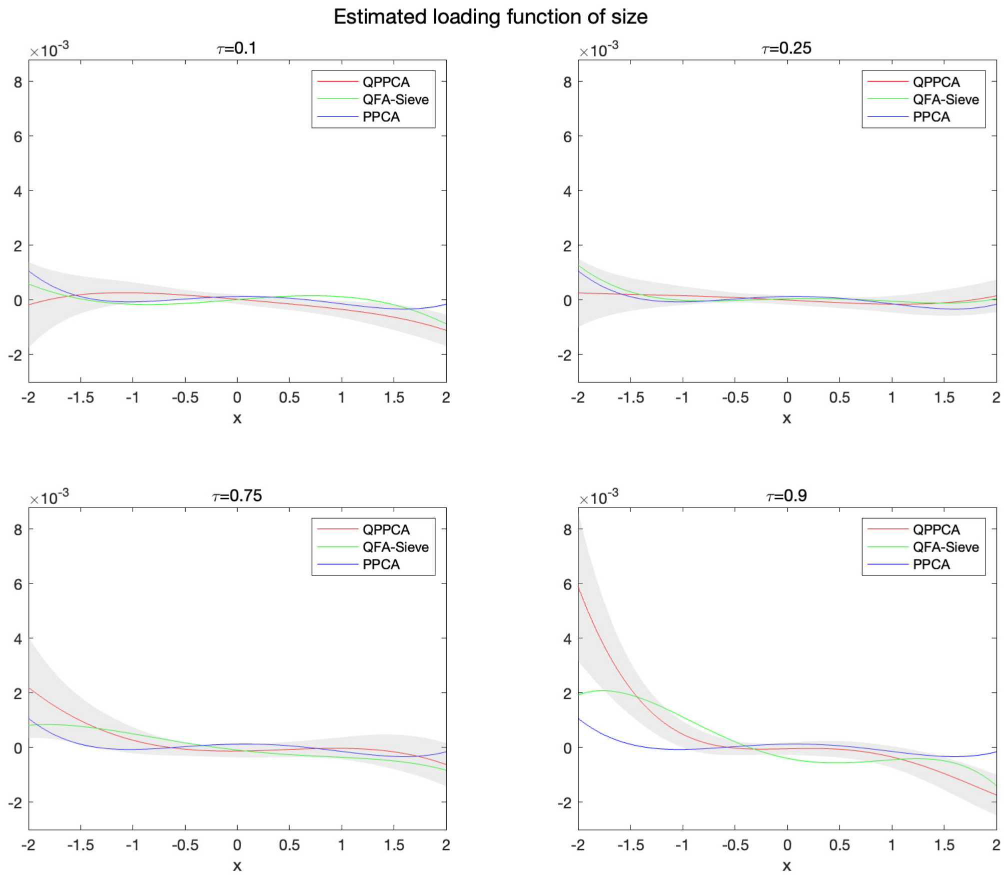
Note: $n = 355, T = 62, R = 1$ and $D = 4$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0,1)$, x_{id} ($i = 1, ..., n$ and $d = 1, 2, 3, 4$) are independently drawn from the 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 4: Estimated Loading functions at $\tau = 0.25$, 1 missing variable: QPPCA and QFA-Sieve



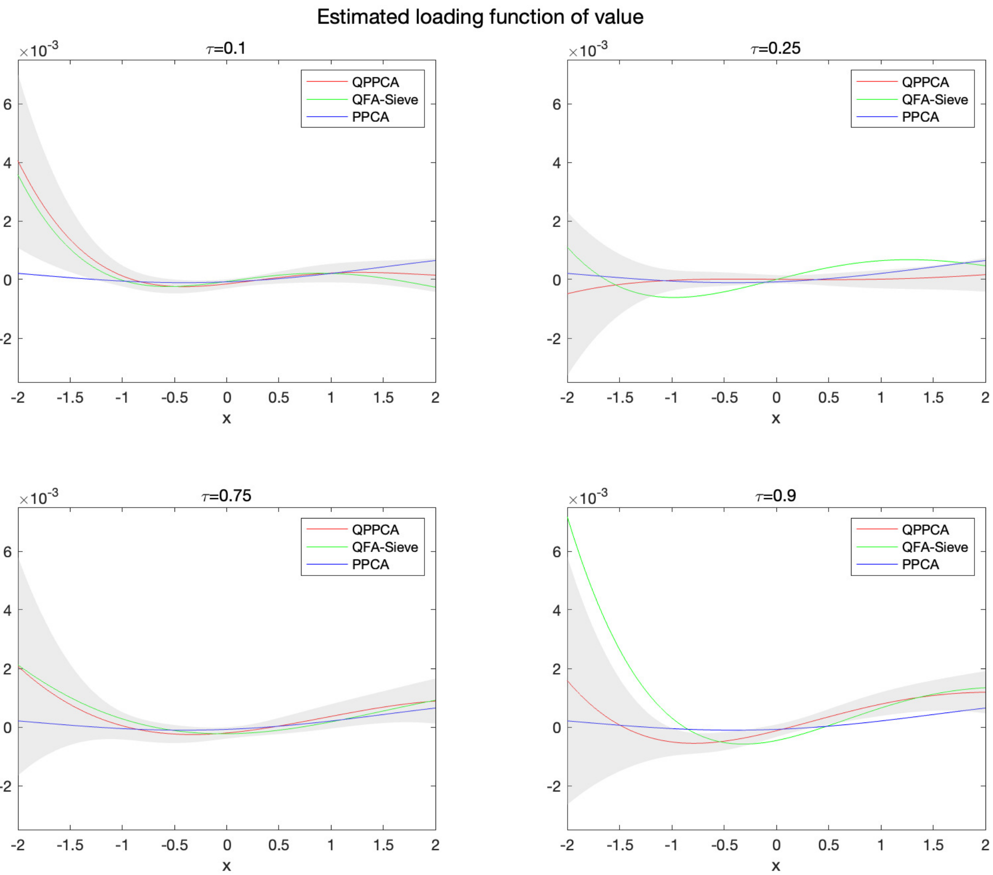
Note: $n = 355, T = 62, R = 1$ and $D = 5$. The DGP is: $y_{it} = (g_1(x_{i1}) + g_2(x_{i2}) + g_3(x_{i3}) + g_5(x_{i5})) \cdot f_t + g_4(x_{i4}) \cdot f_t \cdot u_{it}$, where $f_t = |h_t|$ and h_t are independently drawn from $N(0, 1)$, x_{id} ($i = 1, ..., n$ and $d = 1, \ldots, 5$) are independently 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 5: Estimated loading functions using PPCA, QPPCA and QFA-Sieve



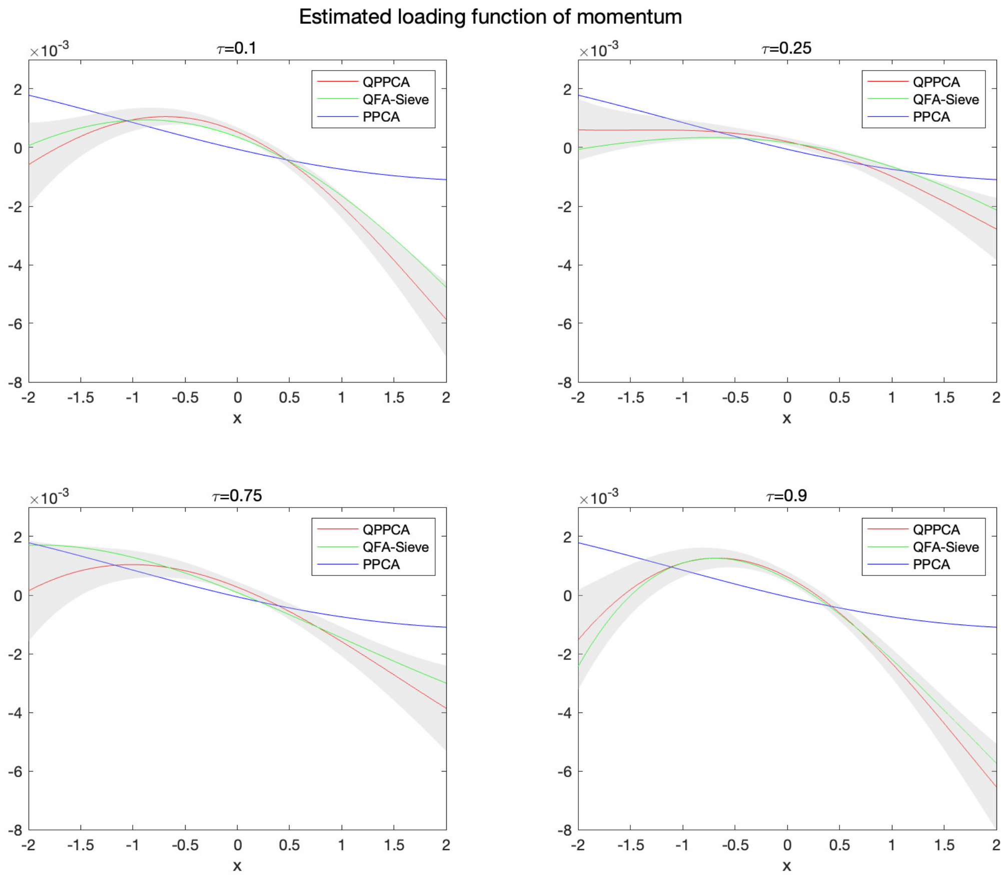
Note: This Figure plots the estimated quantile factor loading functions of **size** using QPPCA and QFA-Sieve at $\tau = 0.1, 0.25, 0.75, 0.9$. and the estimated loading function of **size** using PPCA. The shaded areas show the 95% point-wise confidence intervals constructed using the asymptotic distribution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

Figure 6: Estimated loading functions using PPCA, QPPCA and QFA-Sieve



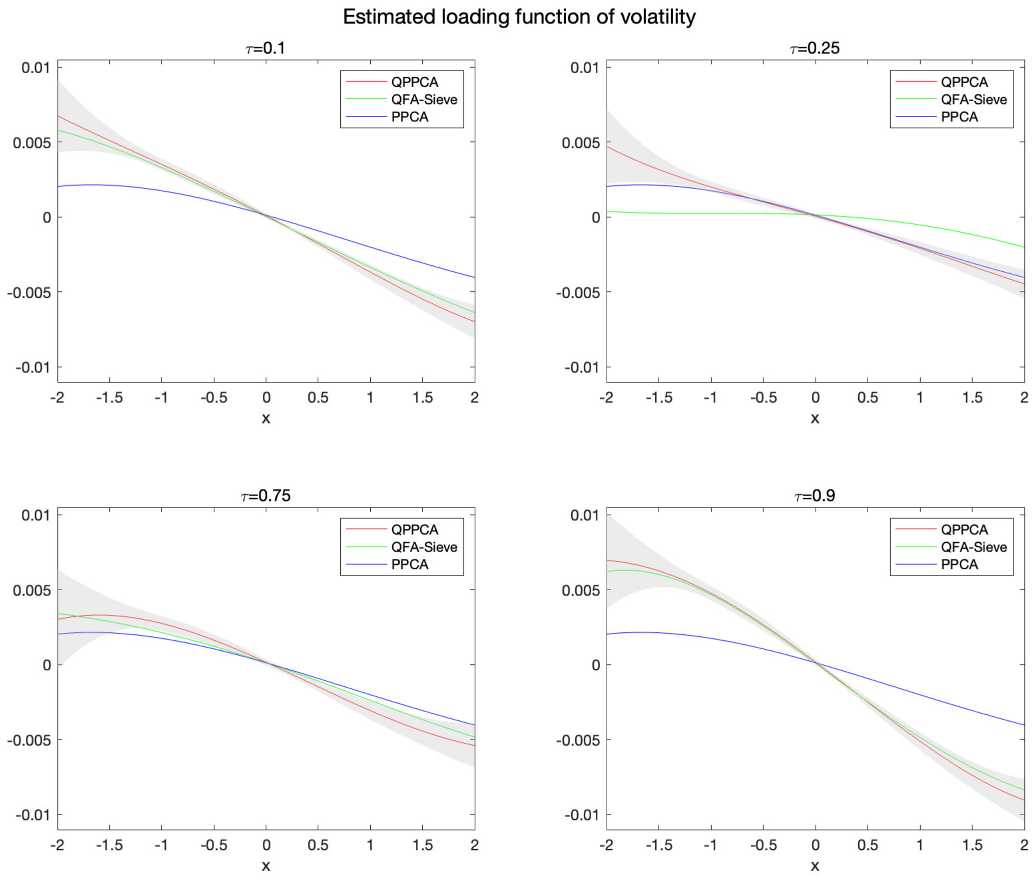
Note: This Figure plots the estimated quantile factor loading functions of **value** using QPPCA and QFA-Sieve at $\tau = 0.1, 0.25, 0.75, 0.9.$ and the estimated loading function of **value** using PPCA. The shaded areas show the 95% point-wise confidence intervals constructed using the asymptotic distribution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

Figure 7: 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 8: 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, S. and J. Feng (2025). Robust quantile factor analysis.

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

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.

Feng, J. (2024). Nuclear norm regularized quantile regression with interactive fixed effects. *Econometric Theory 40*(6), 1391–1421.

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.