2017-10-03

Quantile Factor Models

Chen, Liang

http://hdl.handle.net/10016/25299

UC3M Working papers
Economics
17-13
October, 2017
ISSN 2340-5031

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 916249875

Quantile Factor Models*

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

[1]*School of Economics, Shanghai University of Finance and Economics*
[2]*Department of Economics, European University Institute*
[3]*Department of Economics, Universidad Carlos III de Madrid*

October 3, 2017

Abstract

In contrast to Approximate Factor Models (AFM), our proposed Quantile Factor Models (QFM) allow for unobserved common factors shifting some parts of the distribution other than the means of observed variables in large panel datasets. When such extra factors exist, the standard estimation tools for AFM fail to extract them and their quantile factor loadings (QFL). Two alternative approaches are developed to estimate consistently the whole factor structure of QFM: (i) a two-step estimation procedure which is only valid when the same factors shift the means and the quantiles; and (ii) an iterative procedure which is able to extract (potentially) quantile-dependent factors and their QFL at a given quantile even when both sets of factors differ. Simulation results confirm that our QFM estimation approaches perform reasonably well in finite samples, while four empirical applications provide evidence that extra factors shifting quantiles could be relevant in practice.

Keywords: Factor models, quantile regression, generated regressors, incidental parameters.

JEL codes: C31, C33, C38.

* We are grateful to Dante Amengual, Manuel Arellano, Steve Bond, Guillaume Carlier, Valentina Corradi, Jean-Pierre Florens, Alfred Galichon, Peter R. Hansen, Jerry Hausman, Sophocles Mavroeidis, Bent Nielsen, Enrique Sentana, Guillermo Tellechea and participants at several seminars and workshops for many helpful discussions and insightful comments. Financial support from The Open Society Foundation, The Oxford Martin School, the Spanish Ministerio de Econom´ıa y Competitividad (grants ECO2016-78652 and Maria de Maeztu MDM 2014-0431), and MadEco-CM (grant S205/HUM-3444) is gratefully acknowledged.

1 Introduction

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

The fact that it is easy to construct theories involving common factors, at least in a narrative version, together with the availability of fairly straightforward estimation procedures for AFM— e.g., via Principal Components Analysis (PCA hereafter),[1] have led to their extensive use in many fields of economics. Early applications of AFM abound in Aggregation Theory, Consumer Theory, Business Cycle Analysis, Finance, Monetary Economics, and Monitoring and Forecasting, among others. More recently, a conventional characterization of cross-sectional dependence among error terms in Panel Data has relied on the use of a finite number of unobserved common factors. These originate from economy-wide shocks that affect all units with different intensities (loadings), in addition to idiosyncratic (individual-specific) disturbances. Interactive fixed-effects models can be easily estimated by PCA (see Bai 2009) or by common correlated effects (see Pesaran 2006). Likewise, the surge of Big Data technologies (machine learning, multiple scaling, social network analysis, etc.) has made factor models a key tool in predictive analytics, seeking to uncover patterns and capture relationships in very large datasets (see Diebold 2012 for a survey).

Our departure point in this paper is to notice that the standard regression interpretation of AFM as as a linear conditional mean model of X_{it} given F_t, that is, $\mathbb{E}(X_{it}|F_t) = \lambda_i' F_t$, entails two possibly restrictive features. On the one hand, PCA does not capture hidden factors that may shift characteristics (moments or quantiles) of the distribution of X_{it} other than its mean. On the other hand, the loadings λ_i are not allowed to vary across the distributional characteristics of each unit i in the panel. A simple way of assessing the limitations of the conventional formulation of AFM is to consider the factor structure in the following *location-scale shift* model: $X_{it} = \alpha_i f_t + g_t \epsilon_{it}$, with $g_t > 0$ and $\mathbb{E}(\epsilon_{it}) = 0$, where one factor shifts location, f_t, while a different factor shifts scale, g_t.[2] This model can be rewritten in quantile regression format as $X_{it} = \lambda_i'(\tau) F_t + \nu_{it}(\tau)$, with $0 < \tau < 1$, $\lambda_i(\tau) = [\alpha_i, \beta_i(\tau)]'$, where $\beta_i(\tau)$ represents the quantile function of ϵ_{it}, $F_t = [f_t, g_t]'$, $\nu_{it}(\tau) = g_t[\epsilon_{it} - \beta_i(\tau)]$, and the conditional quantile

[1] See, *inter alia*, Bai (2003), Bai and Ng (2008b), Stock and Watson (2011).

[2] This model corresponds to Example 4 in sub-section 2.2 further below, which includes a list of illustrative models that could be interpreted as potential Data Generating Processes (DGPs, hereafter) of X_{it}. Some of these models will be also briefly overviewed in the rest of this Section.

$Q_{\nu_{it}}[\tau|F_t] = 0.$[3] PCA will only extract the mean-shifting factor f_t in this model, while it will fail to capture (a rotation of) the scale-shifting factor g_t and the quantile-dependent loadings $\beta_i(\tau)$ in its quantile regression representation.[4]

That said, our goal in this paper is to develop a common factor methodology which is flexible enough to capture both the quantile-dependent loadings and those extra factors (e.g., $\beta_i(\tau)$ and g_t in the previous example) that standard AFM tools are unable to recover. We propose a new class of factor structures, labeled as *Quantile Factor Models* (QFMs hereafter), for which estimation and inference are analyzed. In a nutshell, QFM could be thought of as capturing the same type of flexible generalization that quantile regression techniques (QR henceforth) represent for conditional mean linear regression models.

To help understand how this new methodology works, we introduce the QFM sequentially in three different blocks. The first one deals with the case where the factors are not quantile dependent but their loadings for each variable are allowed to vary throughout the conditional distribution of X_{it}. These loadings are denoted as *Quantile Factor Loadings* (QFL hereafter) at quantile τ $(0 < \tau < 1)$, with $Q_{X_{it}}[\tau|F_t] = \lambda_i'(\tau)F_t$. They become analogues to factor loadings in standard AFM, but allowing factors to exhibit heterogeneous effects across different parts of the conditional distribution. In other words, the QFL at different quantiles $\lambda_i(\tau)$, labeled as *QFL processes*, are functions of τ. The previous location-scale shift model with the restriction that the *same* factor shifts location and scale ($f_t = g_t$): $X_{it} = \alpha_i f_t + f_t\epsilon_{it}$, with $f_t > 0$ and $\mathbb{E}(\epsilon_{it}) = 0$, provides an illustration of the type of model considered in this first block.[5] The QFM version of this example is $X_{it} = \lambda_i'(\tau)F_t + \nu_{it}(\tau)$, with $\lambda_i(\tau) = \alpha_i + \beta_i(\tau)$ where, as before, $\beta_i(\tau)$ represents the quantile function of ϵ_{it}, $F_t = f_t$, $\nu_{it}(\tau) = f_t[\epsilon_{it} - \beta_i(\tau)]$, and $Q_{\nu_{it}(\tau)}[\tau|F_t] = 0$. A straightforward *two-step* estimation procedure is proposed for this class of models which yields consistent estimates the common factor f_t and the QFL process $\lambda_i(\tau)$.

Inspired by the sequential approach adopted in the literature on factor-augmented regression models (see, e.g., Bai and Ng 2006), PCA is used in the first step to estimate the common factors F_t from X_{it}; in the second step, the QFL at various τs are estimated using QR for the time series of each of the N units, having replaced the unobserved factors by their PCA estimates. Although our setting differs from the one considered in factor-augmented models, we are able to establish uniform consistency, the specific rate of convergence, and weak convergence of the estimated QFL processes under rather general assumptions. The asymptotic distributions of

[3]Throughout the paper we use $Q_W[\tau|Z]$ to denote the conditional quantile of W given Z.

[4]As will be discussed further below in this section, for this specific DGP there is a very simple way of estimating g_t (once f_t has been estimated by PCA). Yet, it will be argued that such estimation procedure fails to retrieve the whole factor structure for slightly different DGPs than the one considered above, e.g., whenever g_t is a vector of factors instead of a single one.

[5]This restricted version of the location-scale shift model with two factors corresponds to Example 2 in subsection 2.2 below. Likewise, a conventional AFM (Example 1 in that subsection): $X_{it} = \alpha_i f_t + \epsilon_{it}$ can be expressed as the following QFM model: $X_{it} = \lambda_i'(\tau)F_t + \nu_{it}(\tau)$, with $\lambda_i(\tau) = [\alpha_i, \beta_i(\tau)]'$, with $F_t = [f_t, 1]'$, $\nu_{it}(\tau) = \epsilon_{it} - \beta_i(\tau)$.

the entire QFL process can then be used to test hypotheses in very general form. For instance, one could test the null that the loadings are equal to pre-specified values for a given τ, for all τs, or even more generally that they are equal, without pre-specifying their value.

Our two-step procedure provides an intuitive approach for estimation and inference in QFMs where the same factors shift the means and the quantiles. However, it fails to extract the extra factors when both sets of factors differ (see Examples 4, 5 and 7 in subsection 2.2), or when some factors themselves are subject to quantile dependence (see Examples 8 and 9). To address estimation in these more general class of models, a second block of the paper is devoted to propose a novel *fully iterative* estimation procedure. This estimation approach relies on the minimization of the standard *check* function in QR (instead of the standard quadratic loss function used in AFM) to estimate jointly the common factors and the QFL at a given quantile, $F_t(\tau)$ and $\lambda_i(\tau)$, respectively. The consistency of the such estimators is shown for N and T jointly going to infinity. The challenge of this proof comes from having to deal with $(N+T)*r$ incidental parameters (where $r \geq 1$ is the number of factors) and a non-smooth objective function.

Notice that location-scale shift models where $f_t \neq g_t$ are behind a current line of research in asset pricing which has been coined the "idiosyncratic volatility puzzle" by Ang et al. (2006). This approach focuses on the co-movements in the idiosyncratic volatilities of a panel of asset returns. Notice that the volatility co-movement does not arise from omitted factors in the AFM but from assuming a genuine factor structure in the idiosyncratic volatility processes (see Barigozzi and Hallin 2016, Herskovic et al. 2016 and Renault et al. 2017). This technique is valid if the model corresponds to a DGP akin to the ones presented in the examples above; yet, when there are co-movements in other moments or distributional characteristics of the returns, to the best of our knowledge, our QFM iterative approach becomes the first estimation procedure which can deal with these issues.

Finally, the third block of QFM is devoted to what we label the *combined estimation approach*. This estimation procedure is computationally simpler than the fully iterative approach and can be applied when the direct objects of interest are the quantile-dependent factors (i.e., g_t in the previous examples) and their QFL, rather than *all* factors. Going back to the location-scale shift model with different factors: $X_{it} = \alpha_i f_t + g_t \epsilon_{it}$, the idea is to first estimate (rotations of) α_i and f_t by PCA, and then apply the iterative procedure to the residuals $\hat{e}_{it} = X_{it} - \hat{\alpha}_i \hat{f}_t$ to extract g_t and its corresponding QFL process. As before, consistency of the estimators in the combined estimation approach is shown for N and T going to infinity.

Admittedly, for this specific model, a much simpler procedure would be to retrieve g_t from:

$$\frac{1}{N}\sum_{i=1}^{N} \hat{e}_{it}^2 = g_t^2 \cdot \frac{1}{N}\sum_{i=1}^{N} \epsilon_{it}^2 + \frac{1}{N}\sum_{i=1}^{N} (\alpha_i f_t - \hat{\alpha}_i \hat{f}_t)^2 + 2 g_t \cdot \frac{1}{N}\sum_{i=1}^{N} (\alpha_i f_t - \hat{\alpha}_i \hat{f}_t)\epsilon_{it},$$

where, under some assumptions, it holds that the last two terms are $o_P(1)$ and $N^{-1}\sum_{i=1}^{N}\epsilon_{it}^2 \to E(\epsilon_{it}^2)$. As a result, an estimator $\hat{g}_t$ can be derived (up to a scalar) from the square root of $N^{-1}\sum_{i=1}^{N}\hat{e}_{it}^2$. However, while the combined estimation approach is able to recover the whole QFM structure for more general DGPs than the previous model (see, e.g., Example 5 to 9 in sub-section 2.2), it will be shown that the estimation procedure above fails to do so.

Summing up, our paper relies on three different but related econometric literatures: quantile regressions, panel data and factor models. Our specific contributions to each of these research areas can be summarized as follows:

(i) First, as in standard factor augmented regressions, the true factors are replaced by the estimated factors when applying QR in the second step of our two-step procedure. However, the standard conditions on the relative asymptotics of N and T allowing for the estimated factors to be treated as known when the optimizing criteria are smooth object functions (e.g., the minimization of sums of squared residuals) do not hold in our QFM setup. In effect, while these conditions are $T^{1/2}/N \to 0$ for linear factor-augmented regressions (see Bai and Ng 2006) and $T^{5/8}/N \to 0$ for non-linear factor-augmented regressions (Bai and Ng 2008a), lack of smoothness in our check function criterion requires the stronger condition $T^{5/4}/N \to 0$.

(ii) Second, our QFM setup contributes to the growing literature on the intersection of QR and panel data models. The novelty here is that, while this literature deals with observable regressors, we consider unobservable ones (the common factors).[6] As a result, loading and factors become unknown incidental parameters in QFM, and our iterative approaches (fully or combined) provide a feasible way to estimate them consistently.

(iii) Lastly, the extra factors obtained by our iterative procedures can be used to improve the monitoring and forecasting performance in any augmented regression setup, as well as to help in the factor identification process, depending on the application at hand. For instance, in finance these "new" factors could be interpreted as variance, skewness or kurtosis common factors driving assets returns; with income data, they could represent common factors behind income inequality; and with climate data these factors could represent common features behind global extreme temperatures, etc.

Several empirical applications of the proposed estimation procedures are provided using large panels of financial, macroeconomic and climate change data. As for financial data, we use updated information on monthly US mutual fund returns, as well as the portfolio returns used by Fama and French (1993) in their classical paper on common risk factors. As regards macro data, we focus on the popular Stock and Watson (2016) dataset of quarterly economic indicators for the US. Finally, the climate change data is drawn from the Climate Research Unit (CRU) at University of East Anglia, which provides annual information from global temperatures across

[6]See, *inter alia*, Koenker (2004), Abrevaya and Dahl (2008), Graham et al. (2009), Lamarche (2010), Canay (2011), Rosen (2012), Kato et al. (2012) and Harding and Lamarche (2014).

different weather stations. Results based on the combined PCA-iterative estimation approach point out to the existence of extra factors in these datasets, which cannot be captured by standard PCA tools. These hidden factors shift the second moments (volatility) of the data and, in some instances, we even find some evidence of extra factors shifting higher moments (skewness and kurtosis) of the corresponding distributions. In particular, it is documented that several of our estimated quantile-shifting factors are similar to those obtained from applying PCA to the squared residuals (a procedure labeled PCA-SQ hereafter); as explained earlier, under the assumption that the squared idiosyncratic error terms exhibit a factor structure, this last procedure has been recently proposed in the empirical finance literature to capture the so-called *volatility* factors (see, e.g, Renault et al. 2017). Yet, as will be shown below, the PCA-SQ estimation procedure is more restrictive than our iterative approaches since it does not yield separate identification of quantile-shifting factors when there are more than one. In addition, it is shown that a generalization of this procedure to capture factor structures in higher powers (e.g., cubed or fourth powers) of the idiosyncratic error terms does not work well. By contrast, our fully iterative and combined estimation approaches achieve consistent estimators of the relevant extra factors and QFL in all these instances.

The rest of the paper is organized as follows. Section 2 defines QFM and provides a list of simple illustrative examples where the new factor methodology is the appropriate one. Section 3 presents the main asymptotic results for the two-step procedure focusing exclusively on the estimation and inference of QFL processes. Section 4 proposes a fully iterative procedure when the two-step procedure fails and proves the consistency of the loadings and factor estimators. Section 5 presents the combined estimation approach based on PCA in the first stage and the application of the iterative procedure to residuals in the second stage. Section 6 contains some simulation results to evaluate the performance in finite samples of our estimation procedures relative to other alternative approaches, such as PCA and PCA-SQ. Section 7 considers several empirical applications using four large panel datasets, where we document the relevance of factors shifting other moments of the distributions of the data rather than just their means. Finally, Section 8 concludes and suggests several avenues for further research. Proofs of the main results are collected in the Appendix.

2 Preliminaries, Notation and Examples

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

2.1 Quantile Factor Models

We consider a panel of observable random variables $\{X_{it}\}$ generated by:

$$X_{it} = \lambda_i'(U_{it})F_t, \text{ where } U_{it} \perp F_s, \text{ and } U_{it} \sim U[0,1] \tag{1}$$

for $i = 1, \dots, N$ and $s, t = 1, \dots, T$. The common factors F_t is a $r \times 1$ vector of unobservable random variables, with $F_t \in \mathcal{F} \subset \mathbb{R}^r$ for all t. Let $\mathcal{T}$ denote a closed subinterval of $(0,1)$, and suppose that $\lambda_i(\tau) \in \mathcal{A} \subset \mathbb{R}^r$ for all i and $\tau \in \mathcal{T}$. If we further assume that the mapping $\tau \mapsto \lambda_i'(\tau)f$ is strictly increasing for all i and any $f \in \mathcal{F}$, then $\lambda_i'(\tau)F_t$ is the τth quantile of X_{it} conditional on F_t since:

$$P\big[X_{it} \leq \lambda_i'(\tau)F_t | F_t\big] = P\big[\lambda_i'(U_{it})F_t \leq \lambda_i'(\tau)F_t | F_t\big] = P[U_{it} \leq \tau] = \tau.$$

In other words, model (1) implies:

$$Q_{X_{it}}[\tau | F_t] = \lambda_i'(\tau)F_t \text{ for all } \tau \in \mathcal{T}. \tag{2}$$

Therefore, conditional on F_t, the quantiles of X_{it} have a factor model structure. As a result, we label (1) as a *QFM*, while $\Lambda(\tau) = (\lambda_1(\tau), \dots, \lambda_N(\tau))'$ are denoted as *QFL at* τ.[7]

Similar representations for conditional quantiles can be found in Chernozhukov and Hansen (2006, 2008), Canay (2011), and many other papers referenced therein. QFM also has an interesting random coefficient interpretation (see Koenker 2005) as $\tilde{\lambda}_{it} = \lambda_i(U_{it})$ could be interpreted as random coefficients. Moreover, since the dependence among the elements of F_t is left unrestricted, the factors can include different transformations of the same variable, and thus model (1) can approximate nonlinear conditional quantile functions arbitrarily well by increasing the number of factors. In this sense, the linearity of the quantile factor model (1) is not as restrictive as it might look.

2.2 Examples

In this section we provide a few illustrative examples of QFMs derived from different specifications of location-scale shift models. The objective is to show, via some simple models, that there are cases where the standard AFM methodology is unable to obtain the full factor structure; hence, an alternative approach is required. To simplify the exposition, it is assumed in these examples that there is only one factor shifting location, f_t. As regards the scale, it is assumed

[7] Notice that, for convenience, the factors F_t in (1) are assumed not to depend on the quantiles. As shown in sub-section 2.2, this assumption is not restrictive for a wide set of illustrative models leading to representation (1). Yet, as discussed also in that sub-section, there could be some other cases (e.g., Examples 8 and 9) where factors may be quantile dependent. In such instances, F_t is a function of τ, i.e., $F_t(\tau)$, and (2) is generalized to $Q_{X_{it}}[\tau|F_t] = \lambda_i'(\tau)F_t(\tau)$. A detailed analysis of this broader class of DGPs is deferred to Section 4 below.

that either: (i) there is no factor structure (homoskedasticity), or (ii) the scale is affected by same factor that affects location, $f_t = g_t$, or by a different factor, $g_t \neq f_t$ (heteroskedasticity). We also consider some examples where g_t is a vector of factors, rather than a single one, as well as cases where there is more than one error term. Finally, we discuss an example where there is a factor structure in the higher moments of the idiosyncratic component and show that such factor structures lead to quantile-dependent factors, $F_t(\tau)$, whose estimation requires the use of our iterative estimation procedures.

Example 1. ***Location shift model.*** *$X_{it} = \alpha_i f_t + \epsilon_{it}$, where $\{\epsilon_{it}\}$ are zero-mean i.i.d errors independent of f_t with cumulative distribution function (CDF) F_ϵ. This is a standard AFM which can be equivalently written as $X_{it} = \alpha_i f_t + Q_\epsilon(U_{it})$, where $Q_\epsilon(\tau) = \mathsf{F}_\epsilon^{-1}(\tau) = \inf\{c : \mathsf{F}_\epsilon(c) \leq \tau\}$ is assumed to be uniquely defined for each $\tau \in (0,1)$, and $\{U_{it}\}$ are i.i.d and uniformly distributed over $[0,1]$. Thus, this simple model has a QFM representation (1) by defining $\lambda_i(U_{it}) = [Q_\epsilon(U_{it}), \alpha_i]'$ and $F_t = [1, f_t]'$, such that the loadings of the unit factor (intercepts) are the only quantile-dependent objects in the QFM representation.*

Example 2. ***Location-scale shift model (same sign-restricted factor).*** *$X_{it} = \alpha_i f_t + f_t\epsilon_{it}$, where $f_t > 0$ for all t and $\{\epsilon_{it}\}$ are defined as in Example 1. This model has a QFM representation (1) by defining $\lambda_i(U_{it}) = Q_\epsilon(U_{it}) + \alpha_i$ and $F_t = f_t$, such that the slopes of the factor f_t are quantile dependent objects.*

Example 3. ***Location-scale shift model (same sign-unrestricted factor).*** *$X_{it} = \alpha_i f_t + f_t\epsilon_{it}$, where $\{\epsilon_{it}\}$ are defined as in Example 1 and the sign of f_t is unrestricted. When $f_t \geq 0$, the conditional τth quantile of X_{it} given f_t is $(Q_\epsilon(\tau) + \alpha_i)f_t$ while, when $f_t < 0$, the conditional τth quantile of X_{it} given f_t is $(Q_\epsilon(1-\tau) + \alpha_i)f_t$. Therefore, for some uniformly distributed variable U_{it} this model has a QFM representation (1) with $\lambda_i(U_{it}) = [\alpha_i + Q_\epsilon(U_{it}), \alpha_i + Q_\epsilon(1 - U_{it})]'$, $F_t = [f_t^+, f_t^-]'$, $f_t^+ = f_t \cdot \mathbf{1}\{f_t \geq 0\}$ and $f_t^- = f_t \cdot \mathbf{1}\{f_t < 0\}$. Notice that, when the distribution of ϵ_{it} is symmetric, i.e., $Q_\epsilon(\tau) = -Q_\epsilon(1-\tau)$, this model is observationally equivalent to $X_{it} = \alpha_i f_t + |f_t|\,\epsilon_{it}$.*

Example 4. ***Location-scale shift model (different factors).*** *$X_{it} = \alpha_i f_t + g_t\epsilon_{it}$, where $\{\epsilon_{it}\}$ are defined as in Example 1 and $g_t > 0$. This model has a QFM representation (1) with $\lambda_i(U_{it}) = [\alpha_i, Q_\epsilon(U_{it})]'$ and $F_t = [f_t, g_t]'$.*

Example 5. ***Location-scale shift model with two scale factors.*** *$X_{it} = \alpha_i f_t + (\gamma_{1i}g_{1t} + \gamma_{2i}g_{2t})\epsilon_{it}$, where $\{\epsilon_{it}\}$ are defined as in Example 1 and $\gamma_{1i}g_{1t} + \gamma_{2i}g_{2t} > 0$. This model has a QFM representation (1) with $\lambda_i(U_{it}) = [\alpha_i, \gamma_{1i}Q_\epsilon(U_{it}), \gamma_{2i}Q_\epsilon(U_{it})]'$ and $F_t = [f_t, g_{1t}, g_{2t}]'$. This is an extension of Example 4 to the case of more than one scale-shifting factors; in general we could consider models of the form: $X_{it} = \alpha_i' f_t + (\gamma_i' g_t)\epsilon_{it}$ where f_t and g_t are vectors of dimension larger than unity.*

Example 6. ***Location-scale shift model with squared idiosyncratic errors.*** *$X_{it} = \alpha_i f_t + g_t\epsilon_{it}^2$, where $\{\epsilon_{it}\} \sim$ i.i.d $N(0,1)$ and $g_t > 0$. This model is observationally equivalent to $X_{it} =$*

$\alpha_i f_t + g_t v_{it}$ where v_{it} has a chi-square distribution with 1 degree of freedom. Thus, it has a QFM representation (1) with $\lambda_i(U_{it}) = [\alpha_i, Q_v(U_{it})]'$ and $F_t = [f_t, g_t]'$, where $Q_v(U_{it})$ is the quantile function of v_{it}.

Example 7. ***Location-scale shift model with two idiosyncratic errors.*** *$X_{it} = \alpha_i f_t + g_t\epsilon_{it} + h_t e_{it}$, where ϵ_{it} and e_{it} are two independent normal random variables with variances σ_ϵ^2 and σ_e^2. This model is observationally equivalent to $X_{it} = \alpha_i f_t + \sqrt{g_t^2\sigma_\epsilon^2 + h_t^2\sigma_e^2} \cdot v_{it}$ where v_{it} follows a standard normal distribution. Thus, it has a QFM representation (1) with $\lambda_i(U_{it}) = [\alpha_i, \Phi^{-1}(U_{it})]'$ and $F_t = [f_t, \sqrt{g_t^2\sigma_\epsilon^2 + h_t^2\sigma_e^2}]'$, where Φ^{-1} is the quantile function of the standard normal distribution.*

Example 8. ***Location-scale shift model with an idiosyncratic error and its cube.*** *$X_{it} = \alpha_i f_t + g_t\epsilon_{it} + c_i h_t\epsilon_{it}^3$, where ϵ_{it} is a standard normal random variable. Let g_t, h_t, c_i be positive, then X_{it} has an equivalent representation in form of (1) with $\lambda_i(U_{it}) = [\alpha_i, \Phi^{-1}(U_{it}), c_i\Phi^{-1}(U_{it})^3]'$ and $F_t = [f_t, g_t, h_t]'$. It follows that $Q_{X_{it}}[\tau|f_t, g_t, h_t] = \alpha_i f_t + g_t\Phi^{-1}(\tau) + c_i h_t\Phi^{-1}(\tau)^3$. In particular, if $c_i = 1$ for all i and noticing that the mapping $\tau \mapsto \Phi^{-1}(\tau)^3$ is strictly increasing, then we have $Q_{X_{it}}[\tau|f_t, g_t, h_t] = \alpha_i f_t + \Phi^{-1}(\tau) \cdot [g_t + h_t\Phi^{-1}(\tau)^2]$, so that there exists a QFM representation (1) with $\lambda_i(\tau) = [\alpha_i, \Phi^{-1}(\tau)]'$ and $F_t(\tau) = [f_t, g_t + h_t\Phi^{-1}(\tau)^2]'$. Notice that in this case, the second factor in $F_t(\tau)$, $g_t + h_t\Phi^{-1}(\tau)^2$, is quantile dependent.*

Example 9. ***Location-scale shift model with an idiosyncratic error and its odd powers.*** *$X_{it} = \alpha_i f_t + g_t\epsilon_{it} + \Sigma_{k\in\mathcal{O}} h_{kt}\epsilon_{it}^k$, where $\mathcal{O}$ is a sequence of odd numbers, such that $k = 3, 5, .., K$, provides a generalization of Example 8. In this case $\lambda_i(\tau) = [\alpha_i, \Phi^{-1}(\tau)]'$ and $F_t(\tau) = [f_t, g_t + \Sigma_{k\in\mathcal{O}} h_{kt}\Phi^{-1}(\tau)^{k-1}]'$ and, as in Example 8, the second factor is quantile dependent.*

Not surprisingly, the standard AFM methodology based on PCA only works in Example 1. In the remaining cases, PCA will yield consistent estimates of those factors shifting the means, but it will fail to capture those which shift the quantiles and their corresponding QFL. For instance, our two-step procedure works in Examples 1, 2 and 6 where all relevant factors shift the mean of X_{it}. By contrast, Examples 4, 5, 7, 8 and 9 contain hidden factors that do not affect the mean. In particular, Examples 8 and 9 lead to quantile-dependent factors. Thus, in all these cases, only our iterative procedures manage to capture the whole QFM structure.[8]

3 Two-step Estimation Procedure

Note that DGP (1) can be rewritten as:

$$X_{it} = \lambda_i'(\tau)F_t + [\lambda_i(U_{it}) - \lambda_i(\tau)]'F_t = \lambda_i'(\tau)F_t + v_{it}, \tag{3}$$

[8] Notice that in Example 7, where there is a identification issue related to g_t and h_t, our iterative procedures will only be able to capture f_t and weighted sums of squares of g_t and h_t. Nonetheless, in these cases, our iterative procedures still provide more information than PCA, which will only be able to extract f_t.

where $v_{it} = [\lambda_i(U_{it}) - \lambda_i(\tau)]' F_t$ and $P[v_{it} \leq 0|F_t] = \tau$. The main objects of interest are the common factors and the QFL at all $\tau \in \mathcal{T}$. If F_t were to be observed, using standard QR of X_{it} on F_t leads to consistent and asymptotically normally distributed estimators of $\lambda_i(\tau)$ for each i and $\tau \in \mathcal{T}$. However, since F_t are not observable, a feasible procedure is to estimate the factors first, and then run QR of X_{it} on the estimated factors, $\hat{F}_t$.

Define $\lambda_i = \mathbb{E}[\lambda_i(U_{it})]$, then model (1) can also be expressed as:

$$X_{it} = \lambda_i' F_t + [\lambda_i(U_{it}) - \lambda_i]' F_t = \lambda_i' F_t + e_{it}, \tag{4}$$

where $e_{it} = [\lambda_i(U_{it}) - \lambda_i]' F_t$, and hence $\mathbb{E}[e_{it}|F_t] = 0$. Thus, if λ_i, F_t and e_{it} satisfy some assumptions (see Assumption 1 below), (4) can be viewed as an AFM, and the factors can be consistently estimated by PCA as in Stock and Watson (2002) and Bai (2003).

Remark 1: Relative to a standard AFM (see Example 1 above), it is important to note that we impose stronger assumptions: U_{it} needs to be uniformly distributed, orthogonal to F_t and, more important, it is assumed to be i.i.d. across i and t. Thus, this is equivalent to assuming that ϵ_{it} in Example 1 is i.i.d. across i and t. Notice, however, that as in Bai and Ng (2002) we could have allowed for weak cross-sectional and serial correlations without much difficulty, as long as they are not too strong. In particular, weak cross-sectional correlations would not affect our asymptotic results for the two-step approach, although weak serial correlation may probably change the variance of the two-step estimators. Yet, because independence of the idiosyncratic error terms is essential for the proof of consistency in the iterative approach, we adopt this assumption in the derivation of the results to facilitate comparability of both estimation procedures. Define $X_t = [X_{1t}, \ldots, X_{Nt}]'$, $e_t = [e_{1t}, \ldots, e_{Nt}]'$ and $\Lambda = [\lambda_1, \ldots, \lambda_N]'$. Representation (4) then implies the following characterization of the covariance matrix of X_t: $\mathbb{E}(X_t X_t') = \Lambda \Sigma_F \Lambda' + \Sigma_e$, where $\Sigma_F = \mathbb{E}[F_t F_t']$ and $\Sigma_e = \mathbb{E}[e_t e_t']$ is a diagonal matrix. ■

The above representation leads us to the following two-step estimation approach (2SA, henceforth) for the common factors and the QFL at various τs:[9]

1. First, obtain the estimated factors $\hat{F}$. For example, following Bai (2003), one can use PCA where $\hat{F} = (\hat{F}_1, \ldots, \hat{F}_T)'$ are the r eigenvectors (multiplied by $\sqrt{T}$) of XX' associated with the r largest eigenvalues, where $X = \{X_{it}\}'$ is a $T \times N$ matrix collecting all the observed variables.

2. For $i = 1, \ldots, N$ and each $\tau \in \mathcal{T}$, the QR estimator $\hat{\lambda}_i(\tau)$ is then defined as:

$$\hat{\lambda}_i(\tau) = \arg\min_{\lambda \in \mathcal{A}} T^{-1} \sum_{t=1}^{T} \rho_\tau(X_{it} - \lambda' \hat{F}_t), \tag{5}$$

[9] As will be discussed in Section 4 below, this two-step procedure turns out to fail if the data are generated by the type of location and scale-shift models illustrated in Example 4 above.

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$ is the so-called *check* function which provides the basic optimizing criterion in QR.

Since 2SA can be easily implemented in standard econometric packages, it becomes a very convenient tool for practitioners. Furthermore, an observation of independent interest is that, whenever the errors e_{it} in model (4) have symmetric distributions around zero, our second step at $\tau = 0.5$ can be viewed as a median regression for estimating the factor loadings in an AFM, while the estimated factor loadings in Bai (2003) are obtained by OLS regressions of X_{it} on $\hat{F}_t$.

As is well known, a generic problem of factor analysis is the indeterminacy of the factors and factor loadings up to to a rotation, which also pertains to the QFMs defined above. In effect, for any invertible $r \times r$ matrix A, model (1) is observationally equivalent to $(\lambda_i'(U_{it})A^{-1})(AF_t)$. Therefore, identification of the factors and the QFL requires r^2 restrictions to pin down a unique rotation matrix. Although our main results in the next section are stated for a (possibly random) rotation of $\lambda_i(\tau)$, for illustrative purposes we use the normalization rule implicitly adopted in PCA estimation, labeled as PC1 by Bai and Ng (2013): $T^{-1}\sum_{t=1}^{T} F_t F_t' = I_r$ and $N^{-1}\sum_{i=1}^{N} \lambda_i \lambda_i'$ is diagonal. Other alternative normalizations proposed by these authors (PC2 or PC3) could have been used, and the derivation of the corresponding asymptotic results would be straightforward.[10]

3.1 Asymptotic Results

We next present asymptotic results regarding consistency and weak convergence of the QFL processes when 2SA is used to estimate the whole factor structure. In addition, we discuss how this procedure fails to retrieve the QFL process when the estimated factors in the first step are only consistent for a subspace of the factors in the QFM.

3.1.1 Consistency

To establish the uniform consistency of the estimated QFL, we impose the following assumptions for each $i = 1, \dots, N$:

Assumption 1. *Suppose that the observed data $\{X_{it}\}$ are generated by model (1) and*

(i) The sequence $\{F_t\}$ is strictly stationary and m-dependent with $\mathbb{E}\|F_t\|^4 < \infty$, and $\Sigma_F = \mathbb{E}(F_t F_t') > 0$.

(ii) The random variables $\{U_{it}\}$ are uniformly distributed over $[0, 1]$ and independent across i and t, and U_{it} is independent of F_s for all i, t, s.

[10] PC2 assumes that $T^{-1}\sum_{t=1}^{T} F_t F_t' = I_r$ and $[\lambda_1, \dots, \lambda_r]'$ is a lower triangular matrix, while PC3 assumes $[\lambda_1, \dots, \lambda_r]' = I_r$. All these sets of restrictions imply different rotation matrices but one has to resort to specific economic theories to determine which one becomes more appropriate.

(iii) *There is a compact set $\mathcal{A} \subset \mathbb{R}^r$ such that $\lambda_i(\tau) \in \mathcal{A}$ for all i and $\tau \in \mathcal{T}$, and there is a $\Sigma_\Lambda > 0$ such that $\|N^{-1}\sum_{i=1}^N \lambda_i\lambda_i' - \Sigma_\Lambda\| \to 0$ as $N \to \infty$. There is a constant $\bar{\lambda} < \infty$ such that $\|\lambda_i\| \leq \|\bar{\lambda}\|$ for all i. Define $u_{it} = \lambda_i(U_{it}) - \lambda_i$, then $\mathbb{E}[u_{it}^4] < \infty$ for all i, t.*

(iv) *The eigenvalues of $\Sigma_F\Sigma_\Lambda$ are distinct.*

(v) *The conditional density $\mathsf{f}_X(x|F_t = f)$ exists, and is bounded and uniformly continuous in x for all $f \in \mathcal{F}$; $J(\lambda_i(\tau)) = \mathbb{E}[\mathsf{f}_X(\lambda_i(\tau)'F_t|F_t)F_tF_t']$ is positive definite for all τ.*

Define $H_{NT} = (\Lambda'\Lambda/N)(F'\hat{F}/T)V_{NT}^{-1}$, where $\Lambda' = [\lambda_1, \ldots, \lambda_N]$, $F' = [F_1, \ldots, F_T]$, $\hat{F}' = [\hat{F}_1, \ldots, \hat{F}_T]$, and V_{NT} is a $r \times r$ diagonal matrix with the eigenvalues of $(NT)^{-1}XX'$ in decreasing order. Further, define $H_0 = \Sigma_\Lambda^{1/2}\Upsilon V^{-1/2}$, where V is a diagonal matrix with the eigenvalues of $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$ in decreasing order, and Υ is a matrix of corresponding eigenvectors. It can be shown that:

Theorem 1 (Uniform Consistency)**.** *Under Assumption 1, $\sup_{\tau\in\mathcal{T}} \|\hat{\lambda}_i(\tau) - H_{NT}^{-1}\lambda_i(\tau)\| = o_P(1)$ and $\sup_{\tau\in\mathcal{T}} \|\hat{\lambda}_i(\tau) - H_0^{-1}\lambda_i(\tau)\| = o_P(1)$ for all $i = 1, \ldots, N$.*

Remark 2: The proof of Theorem 1 involves two steps. First, it is shown that $T^{-1}\sum_{t=1}^T \rho_\tau(X_{it} - \lambda'\hat{F}_t)$ converges to $\mathbb{E}[\rho_\tau(X_{it} - \lambda' H_0' F_t)]$ uniformly in τ and λ. Next, given that $H_0^{-1}\lambda_i(\tau)$ is the unique minimizer of $\mathbb{E}[\rho_\tau(X_{it} - \lambda' H_0' F_t)]$ by Assumption 1(v) and that $\hat{\lambda}_i(\tau)$ is defined as the minimizer of $T^{-1}\sum_{t=1}^T \rho_\tau(X_{it} - \lambda'\hat{F}_t)$, the uniform consistency of $\hat{\lambda}_i(\tau)$ for $H_0^{-1}\lambda_i(\tau)$ follows from Lemma B.1 of Chernozhukov and Hansen (2006), which yields a generalization of the consistency of M-estimators to estimated processes. A key result to show the uniform convergence of $T^{-1}\sum_{t=1}^T \rho_\tau(X_{it} - \lambda'\hat{F}_t)$ to $\mathbb{E}[\rho_\tau(X_{it} - \lambda' H_0' F_t)]$, as well as to prove Theorem 2 below, is the following consistency result for the estimated factors: $T^{-1}\sum_{t=1}^T \|\hat{F}_t - H_{NT}'F_t\|^2 = o_P(1)$. This result emerges as a direct consequence of Theorem 1 in Bai and Ng (2002) if one shows that the factors, loadings λ_i and the error terms e_{it} in equation (4) all satisfy Assumptions A to D in their paper. Notice, however, that the error terms $e_{it} = (\lambda_i(U_{it}) - \lambda_i)'F_t$ in our setting violate Assumption C.5 of Bai and Ng (2002), which requires:

$$\mathbb{E}\left|N^{-1/2}\sum_{i=1}^N [e_{it}e_{is} - \mathbb{E}(e_{it}e_{is})]\right|^4 < \infty \text{ for all } t, s. \tag{6}$$

To see this, consider the simple case where $r = 1$. When $t = s$, we have:

$$\mathbb{E}\left|N^{-1/2}\sum_{i=1}^N [e_{it}e_{is} - \mathbb{E}(e_{it}e_{is})]\right|^2 = N^{-1}\sum_{i=1}^N\sum_{j=1}^N \Big(\mathbb{E}[e_{it}^2 e_{jt}^2] - \mathbb{E}[e_{it}^2]\mathbb{E}[e_{jt}^2]\Big).$$

Since in our setup, $\mathbb{E}[e_{it}^2 e_{jt}^2] - \mathbb{E}[e_{it}^2]\mathbb{E}[e_{jt}^2] = \mathbb{E}[u_{it}^2]\mathbb{E}[u_{jt}^2]\big(\mathbb{E}[F_t^4] - (\mathbb{E}[F_t^2])^2\big) \neq 0$ for any i, j, unless F_t^2 is a constant, the previous expression cannot be bounded, and thus Assumption C.5 of Bai and Ng (2002) is not satisfied. As shown in the Appendix, imposing the stronger condition

$\mathbb{E}\|F_t\|^4 < \infty$ allows us to prove that Theorem 1 of Bai and Ng (2002) still holds in our model.[11]
■

3.1.2 Weak Convergence

To establish the limiting distribution of the estimated QFL processes, we impose the following additional assumptions:

Assumption 2. *(i)* $\mathbb{E}\|F_t\|^8 < \infty$*; (ii)* $T^{5/4}/N \to 0$ *as* $N, T \to \infty$*; (iii) For each* $i \leq N$*, the eigenvalues of* $J_{H_0}(\lambda_i(\tau)) = H_0' J(\lambda_i(\tau)) H_0$ *are bounded below by a constant* $\rho^* > 0$ *uniformly in* τ*.*

Define $\varphi_\tau(u) = \mathbf{1}\{u < 0\} - \tau$, let $\mathcal{B}_r$ be a vector of r independent standard Brownian Bridges, and let $\ell^\infty(\mathcal{T})$ be the space of bounded functions on $\mathcal{T}$, then:

Theorem 2 (Weak Convergence)**.** *Under Assumptions 1 and 2, it holds that, for each* i*,*

$$J_{H_0}(\lambda_i(\cdot)) \cdot \sqrt{T}[\hat{\lambda}_i(\cdot) - H_{NT}^{-1}\lambda_i(\cdot)] = -\mathbb{V}_{iT}(\cdot) + o_P(1) \text{ in } \ell^\infty(\mathcal{T}),$$

where $\mathbb{V}_{iT}(\cdot) = T^{-1/2}\sum_{t=1}^{T} \varphi_\tau(X_t - \lambda_i(\cdot)' F_t) H_0' F_t$ *converges weakly to* $\mathcal{B}_r(\cdot)$ *in* $\ell^\infty(\mathcal{T})$*.*

Remark 3: Bai and Ng (2008a) show that, for extremum estimators with twice continuously differentiable object functions, the estimated factors can be treated as known regressors if (among other conditions) $T^{5/8}/N \to 0$. By contrast, the estimation-effects-free property of our estimators requires a much larger N compared to T, i.e., $T^{5/4}/N \to 0$, because the object function considered here is not smooth. Indeed, a necessary condition for the estimated factors to have no distributional effects is:

$$\sqrt{T} \cdot \max_{1 \leq t \leq T} \|\hat{F}_t - H_{NT}' F_t\| = o_P(1). \tag{7}$$

In contrast, in Bai and Ng (2008a), due to their smooth object function, it suffices to have:

$$\big(O_P(1) + O_P(\sqrt{T}/\sqrt{N})\big) \cdot \max_{1 \leq t \leq T} \|\hat{F}_t - H_{NT}' F_t\| = o_P(1).$$

We establish in the Appendix the following uniform convergence rate for the estimated factors:

$$\max_{1 \leq t \leq T} \|\hat{F}_t - H_{NT}' F_t\| = O_P(T^{-5/8}) + O_P(T^{1/8}/\sqrt{N}), \tag{8}$$

illustrating that the required condition $T^{5/4}/N \to 0$ is therefore a direct consequence of (7) and (8). ■

[11] Assumption A of Bai and Ng (2002) does require $\mathbb{E}\|F_t\|^4 < \infty$, which is only needed to prove Theorem 2 in their paper. To prove their Theorem 1, $\mathbb{E}\|F_t\|^2 < \infty$ is sufficient.

Remark 4: Suppose that $\mathcal{T} = [\epsilon, 1-\epsilon]$ for some $\epsilon > 0$. For small values of ϵ, Theorem 2 may not provide a good approximation for the finite sample distributions of the estimators. Usually, the Gaussian approximation performs well for $\epsilon > 30/T$ (e.g., when $T = 200$, $\epsilon > 0.15$) while for more extreme quantiles the small sample distributions are better approximated by the asymptotic distributions of extremal conditional quantiles (see Chernozhukov 2005). ■

As mentioned earlier, given that the asymptotic theory above involves a random rotation matrix of the original QFL, H_{NT}^{-1}, inference about the individual elements of the QFL process relies on the PC1 identification restrictions in Bai and Ng (2013), which we repeat here for convenience:

$$T^{-1}\sum_{t=1}^{T} F_t F_t' = I_r \text{ and } N^{-1}\sum_{i=1}^{N} \lambda_i \lambda_i' \text{ is diagonal.} \tag{9}$$

Hence, PC1 implies that the representation in Theorem 2 still holds if we replace H_{NT}^{-1} by I_r. Formally, we have:

Corollary 1. *Under Assumptions 1 and 2 , the following representation holds for each i if the restrictions in (9) are satisfied for large N and T:*

$$J_{H_0}(\lambda_i(\cdot)) \cdot \sqrt{T}[\hat{\lambda}_i(\cdot) - \lambda_i(\cdot)] = T^{-1/2}\sum_{t=1}^{T} \varphi_\tau(X_t - \lambda_i(\cdot)' F_t) H_0' F_t + o_P(1) \text{ in } \ell^\infty(\mathcal{T}). \tag{10}$$

The result above follows directly from Theorem 2 by noting that, as shown in Bai and Ng (2013), $H_{NT}^{-1} - I_r = O_P(\min[N,T]^{-1})$ under restrictions (9).

Theorem 2 also allows us to construct confidence intervals and make inference for the entire QFL process whenever uniform (in τ) consistent estimators of $J_{H_0}(\lambda_i(\tau))$ are available. Similar to Powell (1986), the following estimator is considered:

$$\hat{J}(\hat{\lambda}_i(\tau)) = \frac{1}{2h_T \cdot T}\sum_{t=1}^{T}\Big\{\mathbf{1}\{|X_{it} - \hat{\lambda}_i(\tau)'\hat{F}_t| \leq h_T\}\hat{F}_t\hat{F}_t'\Big\}, \tag{11}$$

and we need to make an additional assumption:

Assumption 3. *The bandwidth parameter h_T satisfies: $h_T \to 0$ and $h_T \cdot T^{1/2} \to \infty$ and $\|H_{NT} - H_0\|/h_T = o_P(1)$.*

Then, the following result shows that weak convergence still holds when $J_{H_0}(\lambda_i(\tau))$ is replaced by its estimate.

Theorem 3. *Under Assumptions 1 to 3, it holds that $\sup_{\tau\in\mathcal{T}} \big\|\hat{J}(\hat{\lambda}_i(\tau)) - J_{H_0}(\lambda_i(\tau))\big\| = o_P(1)$, and thus for each $i \leq N$, $\hat{J}(\hat{\lambda}_i(\cdot)) \cdot \sqrt{T}[\hat{\lambda}_i(\cdot) - H_{NT}^{-1}\lambda_i(\cdot)] \Rightarrow \mathcal{B}_r(\cdot)$ in $\ell^\infty(\mathcal{T})$.*

From Theorem 3, confidence bands can be derived for $H_{NT}^{-1}\lambda_i(\tau)$. For example, when $r = 1$, the α level confidence band is $\hat{\lambda}_i(\tau) \pm T^{-1/2}\hat{J}(\hat{\lambda}_i(\tau))^{-1}C_\alpha$, where C_α is the αth quantile of $\sup_{\tau\in\mathcal{T}} |\mathcal{B}(\tau)|$. Theorem 3 also implies that for each $i \leq N$ and each $\tau \in \mathcal{T}$,

$$[\tau(1-\tau)]^{-1/2} \cdot \hat{J}(\hat{\lambda}_i(\tau)) \cdot \sqrt{T}[\hat{\lambda}_i(\tau) - H_{NT}^{-1}\lambda_i(\tau)] \rightsquigarrow \mathcal{N}(0, I_r). \tag{12}$$

3.1.3 Hypothesis Testing

The asymptotic results provided in the previous subsection allow us to test various hypotheses of the form $R\lambda_i(\tau) = c$ about the QFL processes, once the identification restrictions in (9) have been imposed to pin down the rotation matrix.[12] First, it is straightforward to construct a Wald test for $H_0 : R\lambda_i(\tau) = c$ for a given τ based on (12) and Corollary 1. Second, based on Theorem 3 and Corollary 1, it is easy to construct a Kolmogorov-Smirnov (KS) test for the hypothesis:

$$H_0 : \lambda_{ij}(\lambda) = \lambda_{ij} \text{ for all } \tau \in \mathcal{T}$$

for $1 \leq j \leq r$, where λ_{ij} is known constant.

A more interesting class of hypotheses is that:

$$H_0 : \lambda_{ij}(\lambda) = \lambda_{ij} \text{ for all } \tau \in \mathcal{T}$$

for $1 \leq j \leq r$, where λ_{ij} is an *unknown* constant. For simplicity, we focus again on the case where the number of PCA factors is $r = 1$, though the following results can be easily generalized to models with $r > 1$. In line with our discussion in section 3.1.2, the aforementioned estimation results lead us to consider the following particular specification of a location-scale shift model:

$$X_{it} = \lambda_i F_t + (1 + \gamma_i F_t)\epsilon_{it}.$$

With $\mathbb{E}[\epsilon_{it}|F_t] = 0$, this is a conditional mean factor model with the same set of common factors shifting location and scale. In this model, the null hypothesis of constant factor loadings across quantiles is equivalent to:

$$H_0 : \gamma_i = 0.$$

Let $\hat{\lambda}_i(\cdot)$ be the estimated factor loading process for individual i, then it is natural to consider the process $\hat{\lambda}_i(\cdot) - \lambda_i$. However, since λ_i is unknown, we base our test on the process $\hat{\lambda}_i(\cdot) - \hat{\lambda}_i(0.5)$, where $\hat{\lambda}_i(0.5)$ can be replaced by any consistent estimator of λ_i under H_0.

Let $\hat{F}_t$ be the PCA estimator of F_t, and define:

$$\hat{\theta}_i(\tau) = [\hat{\alpha}_i(\tau), \hat{\lambda}_i(\tau)]' = \arg\min_{\theta\in\mathbb{R}^2} \sum_{t=1}^{T} \rho_\tau(X_{it} - \theta'(1, \hat{F}_t)) \text{ for all } \tau \in \mathcal{T}.$$

[12] An exception is the hypothesis $\lambda_i(\tau) = 0$ for a given τ, which is invariant to rotations.

Then, as in the proof of Theorem 2, we can show that, under H_0 and the previous assumption:

$$\mathsf{f}_\epsilon(Q_\epsilon(\cdot)) \cdot (1 - h_0^2(\mathbb{E}F_t)^2)^{1/2} \cdot \sqrt{T}(\hat{\lambda}_i(\cdot) - h^{-1}\lambda_i(\cdot)) \Rightarrow \mathcal{B}(\cdot) \text{ in } \ell^\infty(\mathcal{T}),$$

where $\lambda_i(\tau) = \lambda_i + \gamma_i Q_\epsilon(\tau)$, $h = (N^{-1}\sum_{i=1}^N \lambda_i^2)(T^{-1}\sum_{t=1}^T F_t\hat{F}_t)/v$, $h_0 = (\mathbb{E}[F_t^2])^{-1/2}$, and v is the largest eigenvalue of $(NT)^{-1}XX'$. Then, under H_0, it follows that,:

$$\begin{aligned}\hat{v}_T(\cdot) &= \mathsf{f}_\epsilon(Q_\epsilon(\cdot)) \cdot (1 - h_0^2(\mathbb{E}F_t)^2)^{1/2} \cdot \sqrt{T}(\hat{\lambda}_i(\cdot) - \hat{\lambda}_i(1/2)) \\ &= \mathsf{f}_\epsilon(Q_\epsilon(\cdot)) \cdot (1 - h_0^2(\mathbb{E}F_t)^2)^{1/2} \cdot \sqrt{T}(\hat{\lambda}_i(\cdot) - h^{-1}\lambda_i) + \mathsf{f}_\epsilon(Q_\epsilon(\cdot)) \cdot O_P(1), \quad (13)\end{aligned}$$

where the first term on the right converges weakly to a Brownian bridge, and the second term depends on the distribution of ϵ_{it}, which is usually unknown. Note that the second term, which makes the standard KS test $\sup_{\tau\in\mathcal{T}} |\hat{v}_T(\tau)|$ invalid, is due to the estimation of the unknown parameter λ_i, an issue which is known in the literature as Durbin's problem (see Durbin 1973 and Koenker and Xiao 2002). Following Koenker and Xiao (2002), one could easily make use of the Khamaladze transformation to purge the estimation effects and get a nuisance-parameter free test.

4 Fully Iterative Estimation Procedure

The 2SA estimation procedure presented above relies on the assumption that a QFM can be transformed into an AFM, from which all the factors can be extracted by PCA. One key restriction is Assumption 1(iii), which requires that the factors shifting the quantiles of X should also shift the means of X. To see this, consider Example 4 again: $X_{it} = \alpha_i f_t + g_t\epsilon_{it}$, where different factors affect location and scale. Recall that this model can be rewritten in QFM format as $X_{it} = \lambda_i'(U_{it})F_t$, with $\lambda_i(U_{it}) = [\alpha_i, Q_\epsilon(U_{it})]'$ and $F_t = [f_t, g_t]'$. Hence, $\lambda_i = \mathbb{E}[\lambda_i(U_{it})] = [\alpha_i, 0]$, implying the violation of Assumption 1(iii) since $N^{-1}\sum_{i=1}^N \lambda_i\lambda_i'$ fails to be asymptotically full rank. As a result, factor g_t cannot be recovered from the first-step PCA estimators. Therefore, in general, the first step in 2SA can only consistently estimate those factors that shift the means.

However, if we assume that the median of ϵ_{it} is equal to 0, then for each $\tau \in (0,1)$ and $\tau \neq 0.5$ we have:

$$Q_{X_{it}}[\tau|F_t] = \alpha_i f_t + Q_\epsilon(\tau)g_t,$$

where $F_t = [f_t, g_t]'$, and the loadings $\lambda_i(\tau) = [\alpha_i, Q_\epsilon(\tau)]'$ satisfy Assumption 1(iii) if α_i have enough cross-sectional variations. Even though the factor g_t plays no special role in the AFM form of the model (since it does not shift the means of X), the above expression implies that the quantiles of X across individual units at each τ are informative about the factor g_t. Moreover, in view of the discussion about Examples 8 and 9 above, where the second factor is a function of τ, let us now allow for the possibility that, not only the loadings, but also the factors could be quantile dependent. To address these issues we consider in this section the following more general setting for a given $\tau \in (0,1)$:

$$X_{it} = \lambda_i(\tau)' F_t(\tau) + v_{it}(\tau), \tag{14}$$

where the errors $v_{it}(\tau)$ satisfy:

$$P[v_{it}(\tau) \leq 0 | F_t(\tau)] = \tau,$$

and $\lambda_i(\tau)$ and $F_t(\tau)$ are $r(\tau) \times 1$ vector of factor loadings and factors at quantile τ. Thus, the main novelty as regards our first definition of QFM in Section 1 is that the factors and the number of factors are also allowed to differ across τ. Besides accounting for Examples 8 and 9, this new setup trivially nests all other examples listed above. For instance, even in the case of Example 4, there is quantile variation in the number of factors since, when $\tau \neq 0.5$, we have $r(\tau) = 2$ and $F_t(\tau) = [f_t, g_t]'$ while, when $\tau = 0.5$, we have $r(\tau) = 1$ and $F_t(\tau) = f_t$.

To simplify the notations, we suppress the dependence of $F_t(\tau)$, $\lambda_i(\tau)$, $r(\tau)$ and $v_{it}(\tau)$ on τ in the following discussion, and write $F = [F_1, \ldots, F_T]'$ and $\Lambda = [\lambda_1, \ldots, \lambda_N]'$. Also for simplicity, the number of quantile factors r is assumed to be known at τ, so F and Λ are $T \times r$ and $N \times r$ matrices respectively. To avoid confusions, let $F^0 = [F_1^0, \ldots, F_T^0]'$ and $\Lambda^0 = [\lambda_1^0, \ldots, \lambda_N^0]'$ denote the true values of the factor and loadings. Moreover, following the fixed-effects approach in dealing with incidental parameters, we treat F^0 and Λ^0 as fixed parameters to estimate. Alternatively, F^0 and Λ^0 can be viewed as random variables, but all the assumptions and results in the sequel should be understood as being conditional on F^0 and Λ^0.

As in standard QR, we consider the following check function to replace the least-square object function of the PCA estimators:

$$S_{NT}(F, \Lambda) = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau (X_{it} - \lambda_i' F_t).$$

Let $\mathcal{A}$ and $\mathcal{F}$ be two subsets of $\mathbb{R}^r$. The estimators are defined as:

$$[\hat{F}, \hat{\Lambda}] = \underset{F_t \in \mathcal{F}, \lambda_i \in \mathcal{A}}{\arg\min}\ S_{NT}(F, \Lambda), \tag{15}$$

where $\hat{F} = [\hat{F}_1, \ldots, \hat{F}_T]'$, $\hat{\Lambda} = [\hat{\Lambda}_1, \ldots, \hat{\Lambda}_N]'$. Starting with any $T \times r$ matrix $\hat{F}^{(1)}$, the factors in problem (15) can be estimated using the following fully iterative approach (FIA, hereafter):

1. Given $\hat{F}^{(m)} = [\hat{F}_1^{(m)}, \ldots, \hat{F}_T^{(m)}]$, using QR of $\{X_{it}\}_{t=1}^T$ on $\hat{F}^{(m)}$ to estimate $\hat{\Lambda}_i^{(m+1)}$ for $i = 1, \ldots, N$.
2. Given $\hat{\Lambda}^{(m+1)} = [\hat{\Lambda}_1^{(m+1)}, \ldots, \hat{\Lambda}_N^{(m+1)}]$, using QR of $\{X_{it}\}_{i=1}^N$ on $\hat{\Lambda}^{(m+1)}$ to estimate $\hat{F}_t^{(m+1)}$ for $t = 1, \ldots, T$.

3. Repeat Steps 1 and 2 until $\hat{F}^{(k)}$ and $\hat{F}^{(k+1)}$ become close enough.

There are two main difficulties in deriving the asymptotic properties of this iterative estimator. First, there is the presence of incidental parameters along both dimensions, and second there is the non-smoothness of the object function. There are some recent contributions in the literature on panel data models dealing with simplified versions of problem (15). For example, Fernández-Val and Weidner (2016) and Chen et al. (2014) consider bias-corrected fixed-effects estimators for nonlinear panel data models with both individual and time effects. Similar to our QFM (with $r = 1$), their models contain $N + T$ incidental parameters, but their object functions are assumed to be smooth and strictly concave. Likewise, Galvao and Kato (2016) analyze QR for panel data models where the check functions are replaced by some smooth object functions; yet, only N incidental parameters are considered in their model. In sum, Problem (15) requires specific analysis since it involves both a non-smooth object function and $(N + T) * r$ incidental parameters. Due to these difficulties, we focus in the sequel on the derivation of the consistency properties of our iterative estimation procedure which, to our knowledge is a novel result, while the derivation of the asymptotic distribution of the estimated factors and QFL under this approach is left for further research.

To derive the consistency properties of the above estimator, the following assumptions are needed:

Assumption 4. *(i) $\mathcal{A}$ and $\mathcal{F}$ are two compact subsets of $\mathbb{R}^r$. $\lambda_i^0 \in \mathcal{A}$ and $F_t^0 \in \mathcal{F}$ for all $i \leq N$ and $t \leq T$.*
(ii) There exists two positive definite matrices Σ_Λ and Σ_F such that $N^{-1}\Lambda^{0\prime}\Lambda^0 \to \Sigma_\Lambda$ as $N \to \infty$ and $T^{-1}F^{0\prime}F^0 \to \Sigma_F$ as $T \to \infty$.
(iii) $P[v_{it} \leq 0] = \tau$ for all i, t. $\{v_{it}\}$ are serially and cross-sectionally independent, their density functions $\{\mathsf{f}_{it}\}$ exist and are continuous. For any compact set C, there exists $\mathsf{f}_C > 0$ such that $\inf_{c \in C} \mathsf{f}_{it}(c) \geq \mathsf{f}_C$ for all i, t.
(iv) There exists a finite constant M such that $\mathbb{E}|v_{it} - Ev_{it}|^m \leq m!M^m$ for all i, t and all $m \geq 1$.

Assumption 4(i) is not restrictive when F and Λ are both treated as fixed parameters, and 4(ii) is a standard assumption in factor models; 4(iii) is the defining assumption for QFM. 4(iv) restricts the speed at which the central absolute moments of the error terms increase, but this is satisfied by Gaussian and many other standard distributions. Note that although we require the error terms to be independent, they are not assumed to be identically distributed. In fact, as we can see in Example 4, for given f_t and g_t, the errors $v_{it} = g_t\epsilon_{it}$ are identically distributed across i but not across t.

Let $P_A = A(A'A)^{-1}A'$ be the projection matrix of A. Then, in line with Bai (2003), it can be shown that the estimated factors and factor loadings at quantile τ are consistent for F^0 and Λ^0 in the following sense:

Theorem 4. *Suppose that Assumption 4 holds, then we have:*

$$\|P_{\hat{F}} - P_{F^0}\| = o_P(1) \text{ and } \|P_{\hat{\Lambda}} - P_{\Lambda^0}\| = o_P(1).$$

The above result ensures that the space of our estimated factors is close to the space spanned by the true factors shifting the quantiles of X. Going back to Example 4, this means that our estimated factors will be consistent for both f_t and g_t, because g_t shifts the quantiles (except the median) of X. On the contrary, the PCA estimators are only able to retrieve f_t, because g_t does not shift the mean of X.

Remark 5: Define $M_A = I - P_A$, then we have:

$$\|M_{F^0}\hat{F}/\sqrt{T}\| = \|(M_{F^0} - M_{\hat{F}})\hat{F}/\sqrt{T}\| = \|(P_{F^0} - P_{\hat{F}})\hat{F}/\sqrt{T}\| \leq \|P_{F^0} - P_{\hat{F}}\| \cdot \|\hat{F}/\sqrt{T}\|.$$

Under the normalization $\hat{F}'\hat{F}/T = I$, it follows from Theorem 4 that $\|M_{F^0}\hat{F}/\sqrt{T}\| = o_P(1)$, or equivalently,

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - \hat{H}F_t^0\|^2 = o_P(1),$$

where $\hat{H} = (\hat{F}'^0/T) \cdot (F^{0'}F^0/T)^{-1}$ is a random rotation matrix. The above result is similar to Theorem 1 of Bai and Ng (2002). ■

5 Combining PCA and the Iterative Procedure

The previous section shows that the space formed by the mean and quantile-shifting factors can be consistently estimated by FIA. However, in those cases where the quantile-shifting factors (e.g., the volatility factor in the empirical finance literature) and their QFL are the only objects of interest, it is feasible to estimate the spaces of the mean factors and quantile factors separately by means of a computationally simpler procedure. This can be achieved by combining PCA and FIA in a sequential way, leading to what we label the *combined* estimation approach (CA hereafter). For instance, in Example 4: $X_{it} = \alpha_i f_t + g_t\epsilon_{it}$, with $f_t \neq g_t$, the CA has the advantage that we can estimate the spaces of f_t and g_t separately, while FIA will estimate the space of $[f_t\ g_t]$ jointly, i.e., linear combinations of f_t and g_t. In effect, from PCA we could retrieve estimates of f_t and α_i, to obtain residuals $X_{it} - \hat{\alpha}_i\hat{f}_t$. Applying the iterative approach to these residuals would then yield an estimate of g_t and its QFL. Likewise, in Example 2: $X_{it} = \alpha_i f_t + f_t\epsilon_{it}$, PCA and the FIA applied to the residuals would yield two estimates of the same f_t, the first one with constant loadings while the second one with different loadings across τ. In this case, the space of $[f_t\ g_t] = [f_t\ f_t]$ does not have full rank, but this is not a problem for CA since it will extract the only quantile-shifting factor, f_t.

The above arguments can be extended to the case where there are several location and scale-shifting factors by considering the following more general model:

$$X_{it} = \lambda_i' F_t + e_{it},$$

where e_{it} satisfies $\mathbb{E}[e_{it}|F_t] = 0$, with $\lambda_i, F_t \in \mathbb{R}^p$. Suppose that:

$$e_{it} = \gamma_i' G_t + v_{it} \text{ for all } i, t,$$

for some $\gamma_i, G_t \in \mathbb{R}^k$, where $P[v_{it} \leq 0] = \tau$ for some $\tau \in (0,1)$, so that $Q_{e_{it}}[\tau] = \gamma_i' G_t$.

For simplicity, it is assumed that the mean-shifting factors F_t and the quantile-shifting factors G_t do not have elements in common, and that their dimensions (p and k, respectively) are known. As mentioned above, one can estimate (the space of) both sets of factors separately by implementing the following two steps. First, we obtain the estimated residuals: $\hat{e}_{it} = X_{it} - \hat{\lambda}_i' \hat{F}_t$, where $\hat{\lambda}_i'$ and $\hat{F}_t$ are simply the PCA estimators (with PC1 restrictions) as defined in Section 3. Next, the iterative approach is applied to $\hat{e}_{it}$ to estimate the quantile factors.

In particular, let $\mathcal{A}$ and $\mathcal{G}$ be subsets of $\mathbb{R}^k$, and define:

$$[\hat{G}, \hat{\Gamma}] = \underset{\gamma_i \in \mathcal{A}, G_t \in \mathcal{G}}{\arg\min} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(\hat{e}_{it} - \gamma_i' G_t).$$

Again, we treat $G = [G_1, \ldots, G_T]'$ and $\Gamma = [\gamma_1, \ldots, \gamma_N]'$ as fixed parameters, and make the following assumptions:

Assumption 5. *(i) There exists a constant M such that $\mathbb{E}\|F_t\|^4 \leq M$ and $\|\lambda_i\| \leq M$ for all i, t. There exists positive definite matrices Σ_F and Σ_Λ such that $\|T^{-1}\sum_{t=1}^{T} F_t F_t' - \Sigma_F\| = o_P(1)$ and $\|N^{-1}\sum_{i=1}^{N} \lambda_i \lambda_i' - \Sigma_\Lambda\| = o(1)$.*
(ii) $\{e_{it}\}$ are independent random variables with zero means, and there exists a finite constant M such that $\mathbb{E}|e_{it}|^m \leq m! M^m$ for all i, t and all $m \geq 1$.
(iii) $\mathcal{A}$ and $\mathcal{G}$ are two compact subsets of $\mathbb{R}^k$, and $\gamma_i \in \mathcal{A}$ and $G_t \in \mathcal{G}$ for all i, t.
(iv) There exists two positive definite matrices Σ_Γ and Σ_G such that $N^{-1}\Gamma'\Gamma \to \Sigma_\Gamma$ as $N \to \infty$ and $T^{-1}G'G \to \Sigma_G$ as $T \to \infty$.
(v) $P[v_{it} \leq 0] = \tau$ for all i, t. The density functions of v_{it} : $\{\mathsf{f}_{it}\}$ exist and are continuous. For any compact set C, there exists $\mathsf{f}_C > 0$ such that $\inf_{c \in C} \mathsf{f}_{it}(c) \geq \mathsf{f}_C$ for all i, t.

Then, it can be shown that :

Theorem 5. *Under Assumption 5, the estimated mean factors are consistent in the sense that: $T^{-1}\sum_{t=1}^{T} \|\hat{F}_t - H_{NT}' F_t\| = o_P(1)$, and the estimated quantile factors are consistent in the sense that: $\|P_{\hat{G}} - P_G\| = o_P(1)$, where the rotation matrix H_{NT} is defined as in Theorem 1 and the projection matrix P_A is defined as in Theorem 4.*

Remark 6: As mentioned in the Introduction, there have been recent attempts in the empirical finance literature to estimate some extra factors related to idiosyncratic volatility through the PCA-SQ approach (see, *inter alia*, Barigozzi and Hallin 2016, Herskovic et al. 2016 and Renault et al. 2017). The basic underlying model in this stream of the literature resembles Example 4: it allows for a single quantile-shifting factor (called the *volatility factor*) and possibly several mean-shifting factors. In particular, in line with our CA, Renault et al. (2017) first use PCA to estimate the mean factors and obtain the residuals $\hat{e}_{it}$, and then they propose to use the first principal component of $\hat{e}_t^2 = [\hat{e}_{1t}^2, \ldots, \hat{e}_{Nt}^2]'$ as an estimate of the volatility factor. This approach can be justified by the observation that the squared idiosyncratic errors in Example 4 have an AFM structure with g_t^2 as the common factor:

$$u_{it}^2 = g_t^2 \epsilon_{it}^2 = \sigma_i^2 g_t^2 + v_{it},$$

where $u_{it} = g_t \epsilon_{it}$ represents the idiosyncratic error, $\sigma_i^2 = \mathbb{E}[\epsilon_{it}^2]$, and $v_{it} = (\sigma_i^2 - \mathbb{E}[\epsilon_{it}^2])g_t^2$ with $\mathbb{E}[v_{it}|g_t] = 0$. However, when there are more than one quantile-shifting factors, like in Example 5, the PCA-SQ approach will not be able to estimate the space of all the quantile factors. In effect, the squared idiosyncratic errors of Example 5 can be written as:

$$(\gamma_{1i} g_{1t} + \gamma_{2i} g_{2t})^2 \epsilon_{it}^2 = \sigma_i^2 \gamma_{1i}^2 \cdot g_{1t}^2 + \sigma_i^2 \gamma_{2i}^2 \cdot g_{2t}^2 + \sigma_i^2 \gamma_{1t} \gamma_{2t} \cdot g_{1t} g_{2t} + v_{it},$$

where $\mathbb{E}[v_{it}|g_{1t}, g_{2t}] = 0$. Thus, the first principal component of the squared residuals in the PCA-SQ approach will converge to a linear combination of g_{1t}^2, g_{2t}^2 and $g_{1t}g_{2t}$, while our CA consistently estimates the space of $[g_{1t}, g_{2t}]$. ■

6 Simulations

6.1 Estimation of QFL Using the Two-Step Procedure

To evaluate the finite sample performance of our 2SA estimator, we consider Example 2 (with the same common factor shifting location and scale) as the DGP, namely:

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

where λ_i and ϵ_{it} are drawn independently from $\mathcal{N}(0, 1)$, and that $f_t \sim i.i.d$ lognormal$(0, 0.589)$. This DGP implies a linear QFM of form (1) with QFL defined as: $\lambda_i(\tau) = \lambda_i + \Phi^{-1}(\tau)$. The histograms of $[\tau(1-\tau)]^{-1/2} \cdot \hat{J}(\hat{\lambda}_1(\tau)) \cdot \sqrt{T}[\hat{\lambda}_1(\tau) - H_{NT}^{-1}\lambda_1(\tau)]$ from 5000 simulations are reported, together with the density function of a $\mathcal{N}(0, 1)$ random variable. The sample sizes are $T = 100, 200$ and $N = 100, 200, 500, 1000$, while the QFL estimates are displayed for a small subset of relevant quantiles, namely, $\tau = 0.25, 0.50, 0.75$.

Histograms and densities are displayed in Figures 1 to 3 for the above-mentioned three quantiles. In line with our weak convergence results in Section 3, it can be inspected that the histograms of the constructed statistics become closer to the benchmark density functions of a standard normal random variable as N and T grow. Yet, these approximations are slightly less accurate for the quantiles at the tails that at the middle of the distribution. These differences could be due to the choice of the bandwidth parameter h_T which, as in nonparametric density estimations, has a significant influence on the distributions of the test statistics. In our simulations we simply set $h_T = T^{-1/3}$, so there should be enough room for improvements in these approximations if h_T is allowed to be data dependent, an issue which is left for further research.

6.2 A Comparison of Different Estimation Approaches

To illustrate the advantages of using FIA relative to 2SA estimators when the latter fails to retrieve all relevant information, we generate data for $N = T = 20, 50, 100$, using this time Example 4 as the DGP: $X_{it} = \alpha_i f_t + g_t \epsilon_{it}$, where $f_t \sim i.i.d\ \mathcal{N}(0,1)$, $g_t \sim i.i.d$ lognormal$(0, 0.589)$ and $\alpha_i, \epsilon_{it} \sim i.i.d\ \mathcal{N}(0,1)$. Ideally, we would expect the FIA estimators to capture the two factors f_t and g_t at $\tau \neq 0.5$, while the 2SA estimators would only extract the mean-shifting factor f_t. To check whether this prediction holds, we use 2SA and FIA to estimate two factors in each case; then we compare how well they fit the generated factors f_t and g_t.

Columns 3 to 6 of Table 1 report the average $\bar{R}^2$ from regressing f_t and g_t, respectively: (i) on $\hat{F}_{PC}$ (columns 3 and 4), which denotes a vector of the two mean-shifting factors estimated by PCA (i.e., the first step of 2SA), and (ii) on $\hat{F}_{QR}$ (columns 5 and 6), which denotes a vector of the two estimated quantile-shifting factors using FIA at $\tau = 0.1, 0.25, 0.75, 0.9$.[13] It becomes evident from the high values of the reported average $\bar{R}^2$s in columns 3 and 5 (regressions of f on $\hat{F}_{PC}$ and $\hat{F}_{QR}$, respectively) that the mean-shifting factor f_t is well captured by both estimation methods as N, T get large. By contrast, the high average $\bar{R}^2 s$ reported in column 6 (regression of g on $\hat{F}_{QR}$) and the low average $\bar{R}^2$s shown in column 4 (regression of g on $\hat{F}_{PC}$) indicate that FIA is the only estimation procedure which is able to capture the quantile-shifting factor g_t as N, T get large.

Next, we focus directly on the quantile-shifting (volatility) factor g_t in the previous DGP and check how well it is estimated when using CA and PCA-SQ, namely, the square root of the first principal component of the squared residuals. Columns 7 and 8 in Table 1 report the average $\bar{R}^2$ from regressing g_t: (i) on $\hat{g}_{CA}$, which denotes the estimated quantile factor using the CA (column 7), and (ii) on $\hat{g}_{VF}$, which denotes the estimated volatility factor using PCA-SQ. From the reported average $\bar{R}^2$s it can be observed that, as the sample sizes increase, $\hat{g}_{VF}$ gets slightly closer to g_t than $\hat{g}_{CA}$ at $\tau = 0.25, 0.75$; yet, the two methods perform equally well at $\tau = 0.1, 0.9$.

[13] In this simulation and in the remaining ones reported in this sub-section, the average $\bar{R}^2$ is computed from 100 replications.

Hence, these findings support the idea that for this particular DGP, which is best suited for for the use of PCA-SQ, CA performs reasonably well in capturing the volatility factor. Yet, as we will see next, CA may outperform PCA-SQ for relevant deviations of Example 4, e.g., when either the number of scale-shifting factors exceeds one or when there may be quantile-dependent factors.

To illustrate the advantages of using CA in the first case, we proceed to compare the performances of $\hat{g}_{CA}$ and $\hat{g}_{VF}$ using Example 5 as the DGP. In particular, we generate the following datasets: $X_{it} = \alpha_i f_t + (\gamma_{1i} g_{1t} + \gamma_{2i} g_{2t})\epsilon_{it}$ for $N, T = 100, 200, 500$, where $f_t \sim i.i.d\ \mathcal{N}(0,1)$, $g_{jt} \sim i.i.d$ lognormal$(0, 0.589)$ for $j = 1, 2$, $\gamma_{1i}, \gamma_{2i} \sim i.i.d$ Uniform$(0,1)$, and $\alpha_i, \epsilon_{it} \sim i.i.d$ $\mathcal{N}(0,1)$. Columns 3 to 6 of Table 2 report the average $\bar{R}^2$ from regressing g_{1t} and g_{2t}, respectively: (i) on $\hat{g}_{CA}$ (columns 3 and 4), where we estimate two quantile-shifting factors from the residuals, and (ii) on $\hat{g}_{VF}$ (columns 5 and 6), which is defined as above. It is clear that, as N, T grow, CA is able to capture both quantile-shifting factors, while PCA-SQ fails to consistently estimate them.

Finally, to examine the performance of our CA in the presence of quantile-dependent factors, we consider Example 8 as the DGP: $X_{it} = \alpha_i f_t + g_t \epsilon_{it} + h_t \epsilon_{it}^3$, where $f_t \sim i.i.d\ \mathcal{N}(0,1)$, $g_t, \sqrt{2}h_t \sim i.i.d$ lognormal$(0, 0.589)$, and $\alpha_i, \epsilon_{it} \sim i.i.d\ \mathcal{N}(0,1)$. As discussed in Section 2, in this case the mean-shifting factor is f_t and the quantile-dependent factor is $s_t(\tau) = g_t + \Phi^{-1}(\tau)^2 h_t$. We consider three estimation methods of $s_t(\tau)$: PCA, CA and PCA-SQ, yielding $\hat{F}_{PC}$ (we estimate 2 mean-shifting factors), $\hat{g}_{CA}$, $\hat{g}_{VF}$ estimated factors, respectively. Columns 3 to 5 of Table 3 report the average $\bar{R}^2$ from regressing $s_t(\tau)$ on: (i) $\hat{F}_{PC}$ (column 3), (iv) $\hat{g}_{CA}$ (column 4) and, (iii) $\hat{g}_{VF}$ (column 5), for $\tau = 0.1, 0.25, 0.75, 0.9$ and $N, T = 100, 200, 500$. As can be inspected, CA behaves far better than PCA or PCA-SQ in capturing the quantile-dependent factor $s_t(\tau)$ at all τs, especially for large N and T.

7 Empirical Applications

In this section we consider empirical applications of our QFM estimation procedures using four datasets in macroeconomics, finance, and climate change:

1. The first dataset (SW for short) corresponds to an updated version of the popular panel of macroeconomic indicators which has been used by Stock and Watson to construct leading indicators for the US economy. This data can be downloaded from Mark Watson's website. SW consists of 167 quarterly macro variables from 1959 to 2014 ($N = 167, T = 221$). These macro variable are transformed into stationary series before estimating the factors (see Stock and Watson 2016 for the details of this dataset).
2. The second dataset (Climate for short) consists of the annual changes of temperature from 338 stations from 1916 to 2016 ($N = 338, T = 100$) drawn from the Climate Research Unit

(CRU) at the University of East Anglia, where information about global temperatures across different stations in the Northern and Southern Hemisphere is provided.

3. The third dataset (MF for short) contains the monthly returns of 2378 mutual funds from 2000 to 2014 ($N = 2378, T = 180$), obtained from the Center of Research for Security Prices (CRSP).
4. The last dataset (FF for short) contains the monthly excess returns of 100 portfolios from 1985 to 2012, constructed as in Fama and French (1993) and downloaded from Kenneth French's website ($N = 100, T = 324$).

Since the analysis of the estimated mean-shifting factors in these datasets is well documented in the literature, our focus in this section is exclusively on the estimation of the quantile-shifting factors. Both sets of factors are estimated by CA whose two stages are recalled for convenience: (i) first, we use PCA to estimate the mean-shifting factors, F_{PC}, and (ii) next apply the iterative procedure to the residuals to obtain the quantile factors at $\tau = 0.1, 0.25, 0.75, 0.9$, denoted as FQR_τ, with $FQR = [FQR_{10}, FQR_{25}, FQR_{75}, FQR_{90}]$. As a benchmark, we use PCA-SQ to estimate a volatility factor, denoted as VF_2, in each dataset.[14] Likewise, to capture potential factors affecting skewness and kurtosis, let VF_3 and VF_4 denote the cubic and quartic roots of the first principal component of the third and fourth powers of residuals, respectively. In the case of Example 4, where g_t is the only quantile-shifting factor, it can be easily seen that both VF_2 and VF_4 yield consistent estimators of g_t, while VF_3 will converge to 0 since the idiosyncratic error terms ϵ_{it} have zero means and are symmetrically distributed. At any rate, the idea behind computing VF_3 and VF_4 is to capture those common factors shifting higher moments of the distribution of the idiosyncratic errors were the data be generated by more general DGPs than the ones discussed above.

Let p and k denote the number of mean factors and the number of quantile factors, respectively. To choose p we apply the well-known selection procedures developed by Bai and Ng (2002) and Ahn and Horenstein (2013), as well as the evidence drawn on this issue from previous empirical studies using these datasets.[15] For example, in many available studies using SW dataset, the estimated number of mean factors is usually between 3 and 5, which coincides with the range of factors found when applying AFM to financial asset returns; see Fama and French (1993, 2017). As regards k, it is chosen to be 1 or 2. When $k = 1$, we estimate one quantile factor at each τ, so that vector FQR contains four estimated quantile factors; when

[14] Notice that the volatility factor was labeled $\hat{g}_{VF}$ in the simulation exercises provided in Section 6. However, for consistency with VF_3 and VF_4, it is VF_2 in this Section.

[15] We have simulation results (available upon request) showing that the estimators proposed by Bai and Ng (2002) and Ahn and Horenstein (2013) are inconsistent for p in the case of QFM. The reason is that the idiosyncratic errors are cross-sectionally uncorrelated but dependent in QFM, due to the common quantile factors, and that such dependences are not asymptotically negligible. As a consequence, we find that the method of Bai and Ng (2002) tends to overestimate the number of mean factors, while the method of Ahn and Horenstein (2013) tends to underestimate the number of mean factors. Thus, we use the estimated number of factors from these two methods as the upper and lower bounds for choosing p. In this fashion, we document how robust are our empirical results to the number of estimated PCA factors considered in each of the empirical applications.

$k = 2$, we estimate 2 quantile factors at each τ, and thus FQR contains 8 estimated quantile factors in total.

The first issue we need to examine is whether there is some overlapping between the sets the quantile-shifting factors and the mean-shifting factors (e.g., Examples 2 and 4 illustrate extreme cases where these sets of factors are identical and where there is no coincidence at all, respectively). To do this, we regress each element of FQR_τ on F_{PC} and compute the average $\overline{R}^2$ from these regressions across the chosen set of τs. Note that in Example 2, where both factors coincide (so that both F_{PC} and FQR_τ are consistent estimators of f_t), a regression of FQR_τ on F_{PC} will yield an average $\overline{R}^2$ close to 1. Conversely, Example 4 implies an average $\overline{R}^2$ close to 0 since the quantile-shifting factor differs from the mean-shifting factor. The low average $\overline{R}^2$s reported in Table 4 overwhelmingly indicate that the two types of factors differ. This would be in line (but perhaps not exactly identical) with a DGP close to Example 4 or some of the other related examples listed in Section 2. By contrast, it provides strong evidence against a DGP in line with Example 2.

Next, in Table 5, we report the $\overline{R}^2$s of regressing VF_2, VF_3, VF_4 on FQR, for different choices of p and k for each dataset. The idea is to check whether our CA estimated factors are able to reproduce the volatility factors obtained by Renault et al. (2017) in the case of VF_2, and the higher moment factors which result from extending their estimation procedure to higher powers of the residuals. There are three main takeaways from the results displayed in Table 5.

First, the $\overline{R}^2$s of regressing VF_2 on FQR are close to 1 for all applications and for all different choices of p and k (except in the FF dataset when $p = 3$), implying that our CA fares well relative to PCA-SQ in capturing those quantile factors that shift the second moment of the idiosyncratic residuals. To further check this finding, we plot FQR_τ together with VF_2 for all applications when $k = 1$ and some representative choices of p in Figure 4. It can be observed that, except perhaps for the Climate case, the estimated quantile factors and the estimated volatility factor are very close. Yet, as discussed earlier, the factor loadings obtained by these authors are always constant whereas the QFL obtained by our CA are quantile dependent, as documented in Figure 5, where we plot the estimated QFLs at $\tau = 0.1, 0.25, 0.75, 0.9$ when $k = 1$ for each dataset.[16]

Second, the $\overline{R}^2$s of regressing VF_3 on FQR are small in most cases. As already mentioned, these low values of the $\overline{R}^2$s could be explained by the definition of the idiosyncratic errors as zero-mean symmetrically distributed random variables, implying that their cubes will also be zero-meaned. Hence, to the extent that the first principal component of the cubed residuals are close to a linear combination of the cubed idiosyncratic errors, VF_3 would be close to a vector of zeros, leading to a low correlations with FQR. However, there is an exception for the Climate dataset for $p = 1$ and $k = 2$, where $\overline{R}^2 = 0.75$, pointing out to some evidence of a factor

[16] For readability we only plot the QFLs of the first 200 variables for MF.

structure in the skewness of the cubed idiosyncratic errors.

Finally, as discussed above, when there is only one quantile-shifting factor (e.g., like in Example 4), FQR_τ, VF_2, VF_4 are all consistent estimators of this factor. Hence, regressing VF_4 on FQR should lead to an $\overline{R}^2$ close to 1 when $k = 1$, which is is only the case in the MF dataset, where $\overline{R}^2 = 0.89$ for $p = 1$. Thus, although for the MF dataset there is some evidence of a factor structure in kurtosis driven by a single factor, the findings for the other three datasets do not support this hypothesis.

Overall, the results reported above provide support for the existence of more than one quantile-shifting factors in the examined datasets. Given the novelty of these findings, further analysis and interpretation of these factors is high in our research agenda.

8 Conclusions

Factor models have become a leading methodology for the joint modeling of large number of economic time series, with the big improvements in data collection and information technologies. They have proved very successful in finance and macroeconomic applications (monitoring and forecasting). This first generation of factor models are designed to reduced the dimensionality of big datasets by finding those common components which, by shifting the means of the observed variables with different intensities, are able to capture a large fraction of their co-movements. However, one could envisage the existence of other common factors (*extra factors*) that do not (or not only) shift the means but also affect other distributional characteristics (volatility, higher moments, extreme values, etc). This calls for a second generation of factor models. In this paper we present a novel methodology to address these new issues, under the label of *Quantile Factor Models* (QFM), that could be part of this second generation of factor models. These extra factors are useful for identification purposes, for instance mean factors versus volatility/skewness/kurtosis factors, as well as for forecasting purposes in augmented-regression setups. In effect, to the extent that the new factors represent hidden common characteristics of big datasets that are not captured by PCA, they could therefore help improving forecasts.

Using tools in the interface of quantile regressions (QR) and approximate factor (AFM) models, we have proposed different estimation approaches for factors and quantile-dependent loadings in QFM. These estimation procedures range from a simple two-stage approach (PCA in the first stage and QR in the second stage of each of the variables in the dataset on the estimated PCA factors in the first stage) to more computationally involved (fully and combined) iterative estimation procedures. In the latter, estimated factors and loadings at each quantile are obtained by minimizing check functions which depend on both sets of objects.

Not surprisingly, simplicity comes at a cost: the two-step approach is able to retrieve factors (and loadings) shifting the other parts of the distribution of the observed variables when these factors happen to be identical to those that shift the means (and trivially when only mean-shifting factors exist, as in standard AFM). Conversely, when both factors exist and are not the same, then only the iterative procedures yield consistent estimates of factors and loadings. Interestingly, if the extra factors constitute the main objects of interest, the use of the so-called combined estimation procedure (apply PCA first to obtain the mean-shifting factors, and then apply the iterative procedure to the residuals), proves to be less computationally burdensome than the fully iterative procedure (where all factors and loadings are estimated iteratively at each quantile) and still provides consistent estimators of those objects. Through simulations, we have shown that the combined approach performs reasonably for sufficiently large finite samples, and that it is able to overcome some shortcomings of recently available methods in empirical finance which propose to detect volatility factors by applying PCA to the squared residuals. In addition, our empirical applications to four panel datasets of financial, macro and climate data provide evidence that these extra factors may be relevant in practice.

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

- First, there is the derivation of the asymptotic distributions of quantile-dependent factors and loadings in the iterative procedure, where the interaction of a very large number of incidental parameters, unobserved variables (the factors) and non-smooth criterion functions poses big challenges for the tractability of this problem. Although we provide a novel proof of consistency, further research is clearly needed on how to deal with these issues in QFM. Here we have only tackled them in an admittedly informal way, using goodness-of-fit comparisons of our estimated factors in simulations and empirical applications.
- Second, there is the issue of developing new information criteria to determine the number of quantile-shifting factors. This is an issue that again we have dealt informally by checking how robust are our results in the empirical applications to the choice of a different number of factors. As pointed our earlier, we have found some simulation evidence that popular selection criteria for the number of mean-shifting factors in AFM (see, e.g., Bai and Ng 2002 and Ahn and Horenstein 2013) do not work well for QFM; thus further investigation on this issue is required.
- Third, there is the issue of checking the contributions of the extra factors in forecasting and monitoring (see Stock and Watson 2002 for this type of analysis in dynamic factor models), a topic of high interest for applied researchers, especially with the surge of Big Data technologies. Since in this paper we have mainly focused on introducing the new class of QFM and their basic properties, for the sake of brevity this topic has been left out for further research.
- Finally, given the evidence that these extra factors could be relevant in practice, another

interesting issue is how to interpret them in different economic and financial setups. Once the econometric techniques to detect and estimate extra factors have been established, attempts to provide new economic insights for these objects would help enrich the economic theory behind factor structures.

Figures and Tables

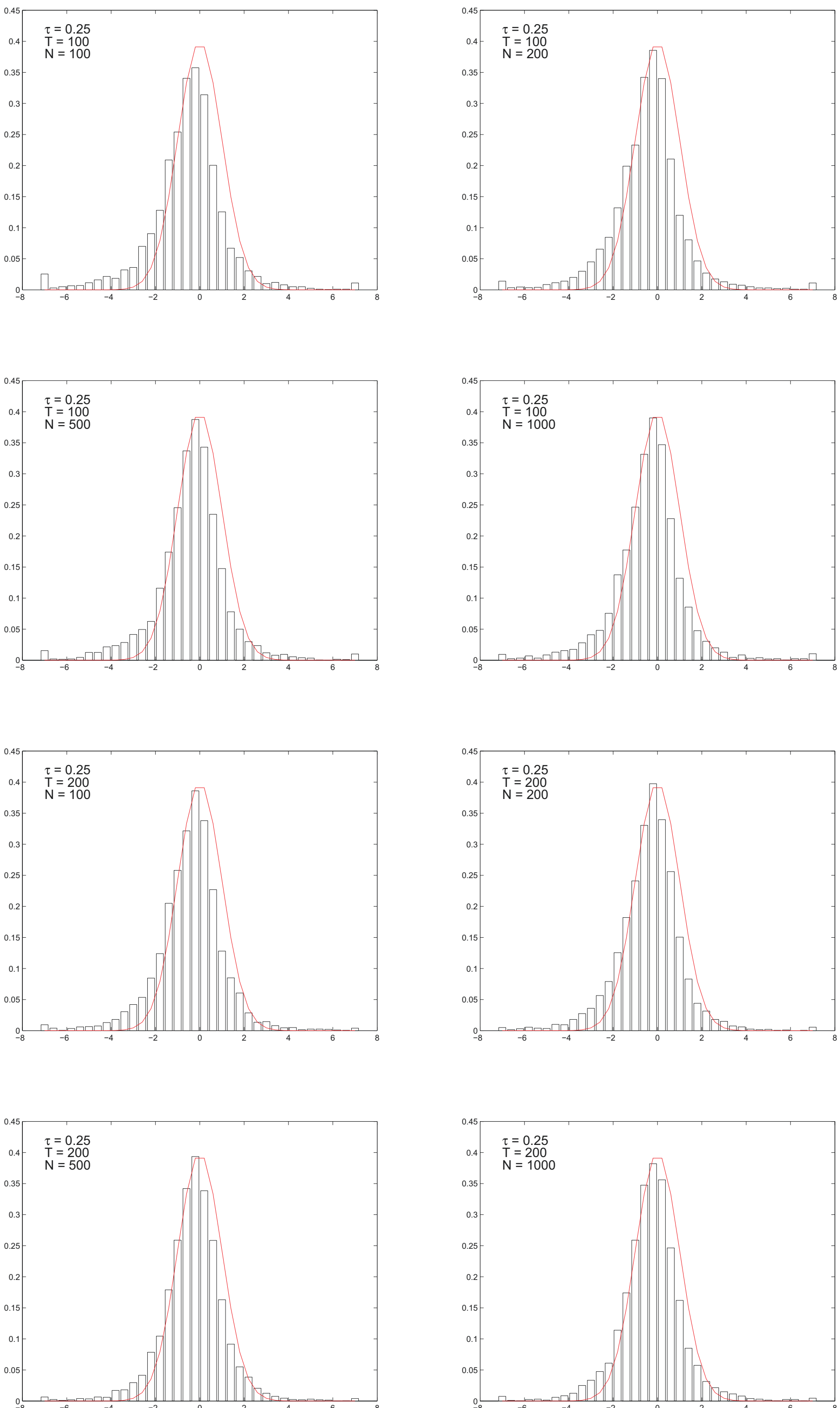

Figure 1: Histograms of $[\tau(1-\tau)]^{-1/2}\cdot\hat{J}(\hat{\lambda}_1(\tau))\cdot\sqrt{T}[\hat{\lambda}_1(\tau)-H_{NT}^{-1}\lambda_1(\tau)]$ and the density function of $\mathcal{N}(0,1)$ for $\tau=0.25$.

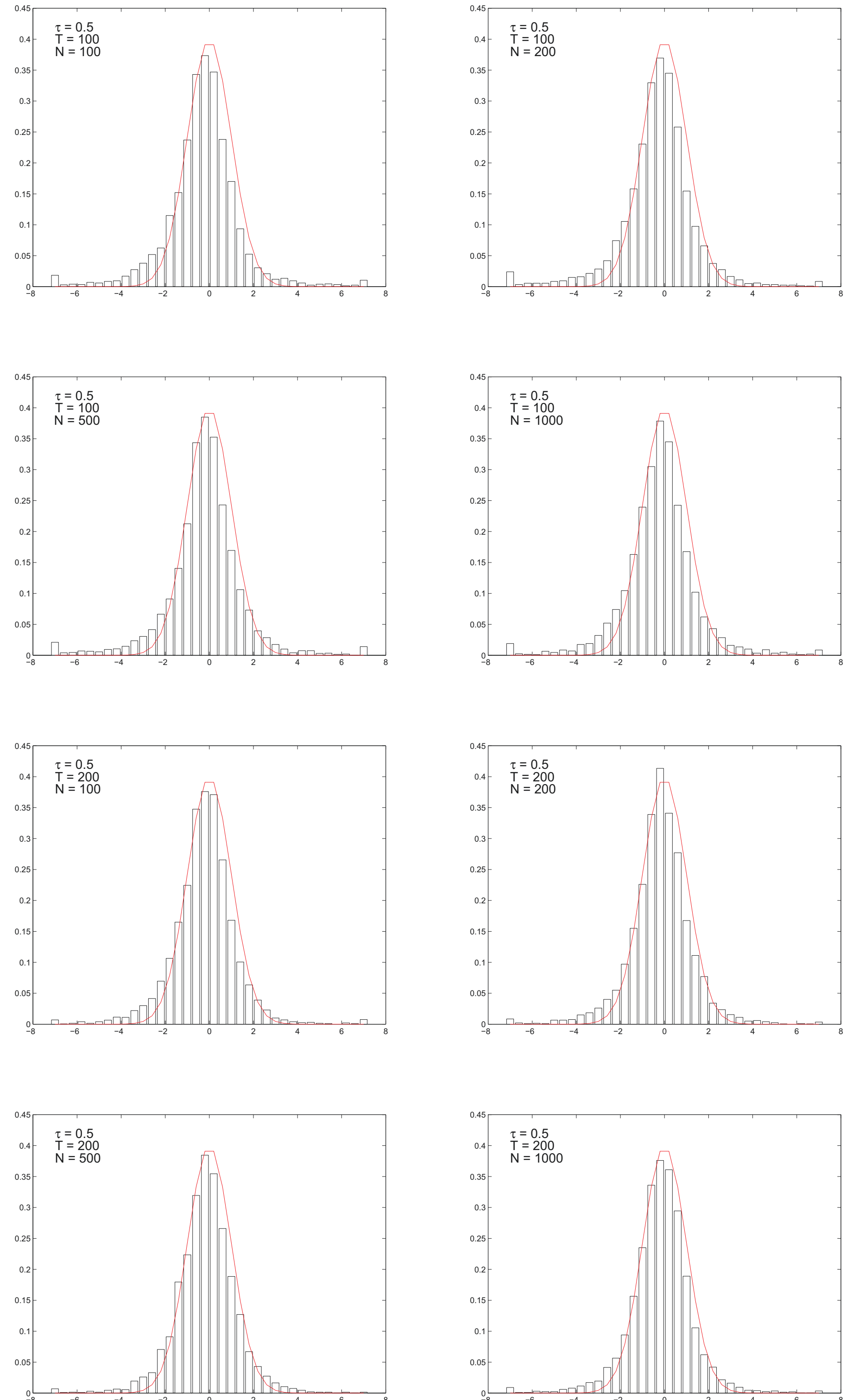

Figure 2: Histograms of $[\tau(1-\tau)]^{-1/2} \cdot \hat{J}(\hat{\lambda}_1(\tau)) \cdot \sqrt{T}[\hat{\lambda}_1(\tau) - H_{NT}^{-1}\lambda_1(\tau)]$ and the density function of $\mathcal{N}(0,1)$ for $\tau = 0.5$.

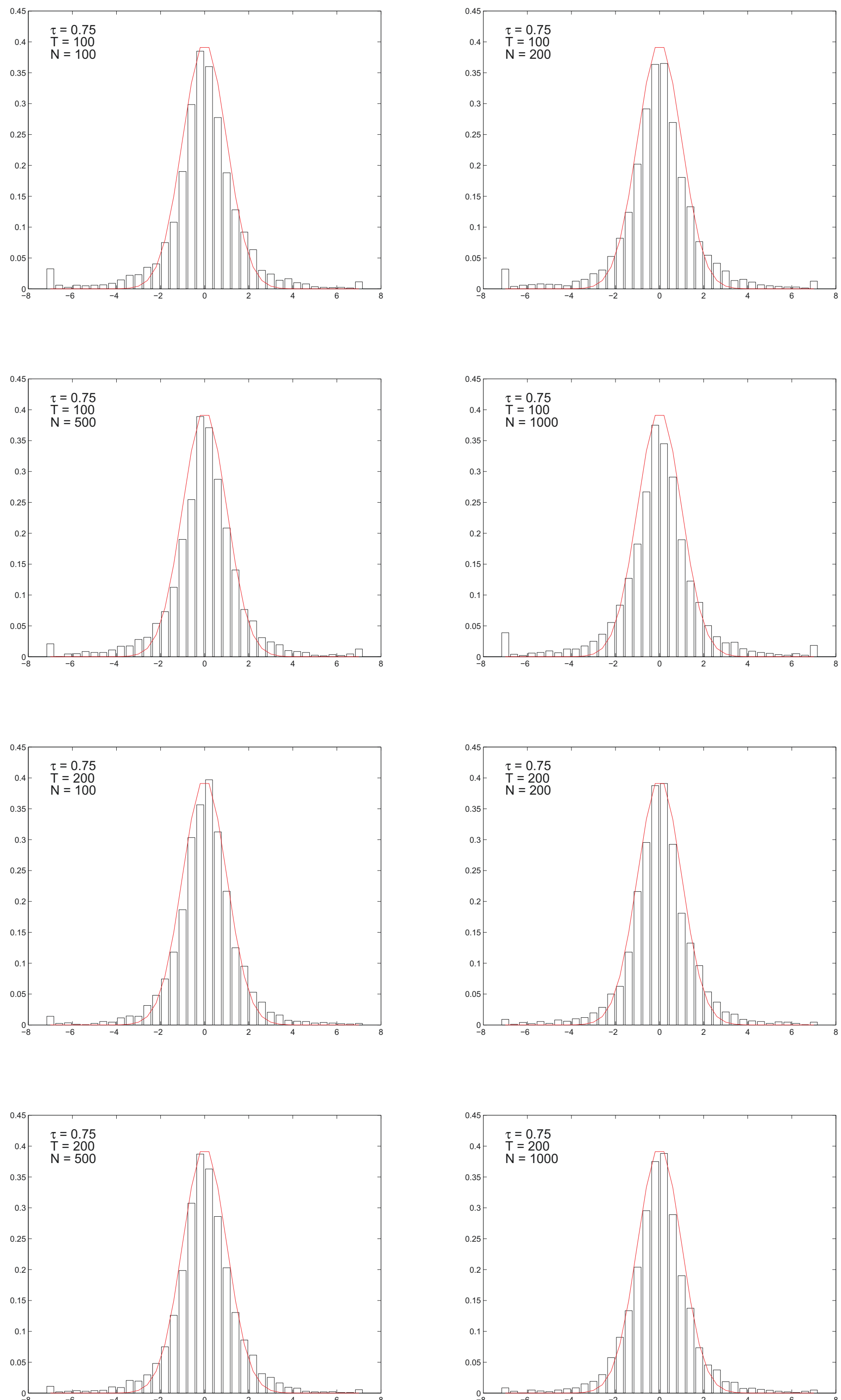

Figure 3: Histograms of $[\tau(1-\tau)]^{-1/2} \cdot \hat{J}(\hat{\lambda}_1(\tau)) \cdot \sqrt{T}[\hat{\lambda}_1(\tau) - H_{NT}^{-1}\lambda_1(\tau)]$ and the density function of $\mathcal{N}(0,1)$ for $\tau = 0.75$.

Figure 4: Comparison of estimated quantile factors (FQR; $\tau = 0.1, 0.25, 0.75$ and $= .9$) and volatility factors (VF_2) for SW, MF, FF ($k = 1, p = 5$) and Climate ($k = 1, p = 2$) datasets.

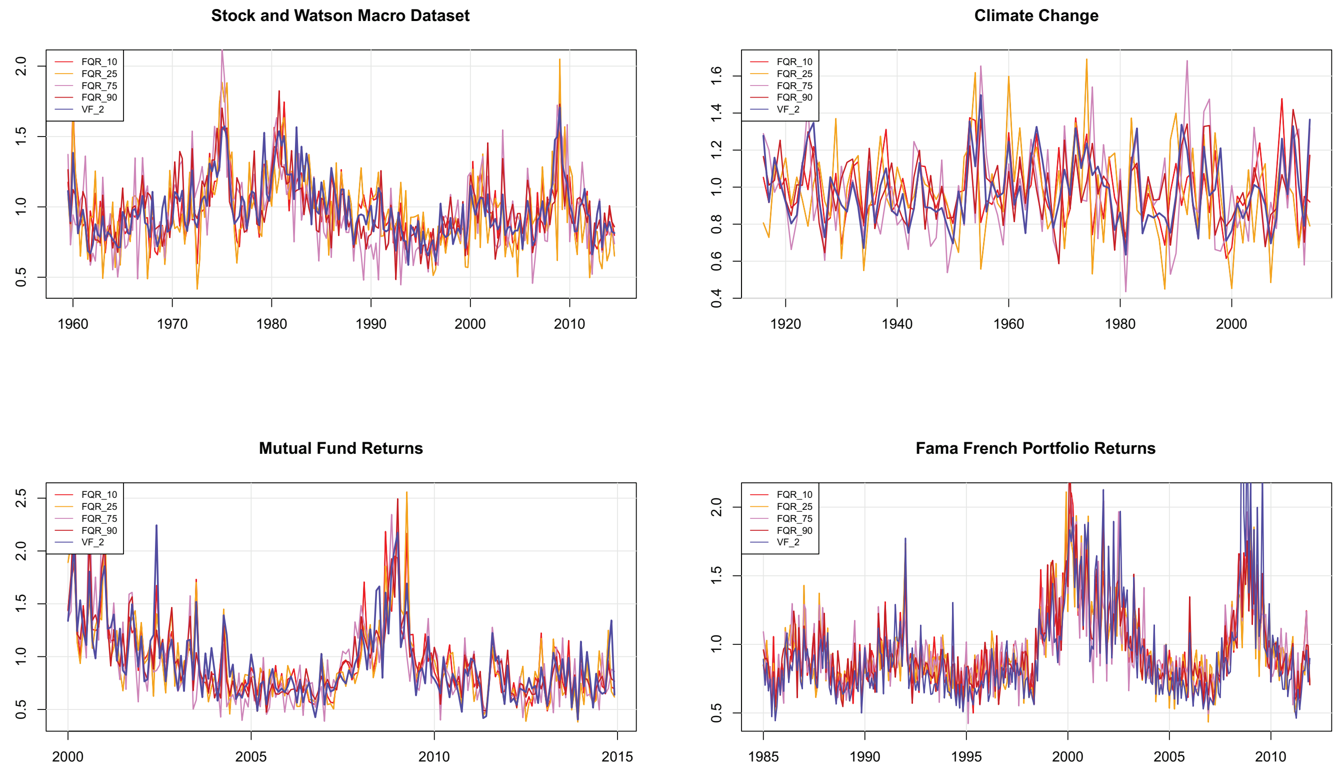

Figure 5: Estimated QFL ($\tau = 0.1, 0.25, 0.75$ and $= .9$) for SW, MF, FF ($k = 1, p = 5$) and Climate ($k = 1, p = 2$) datasets.

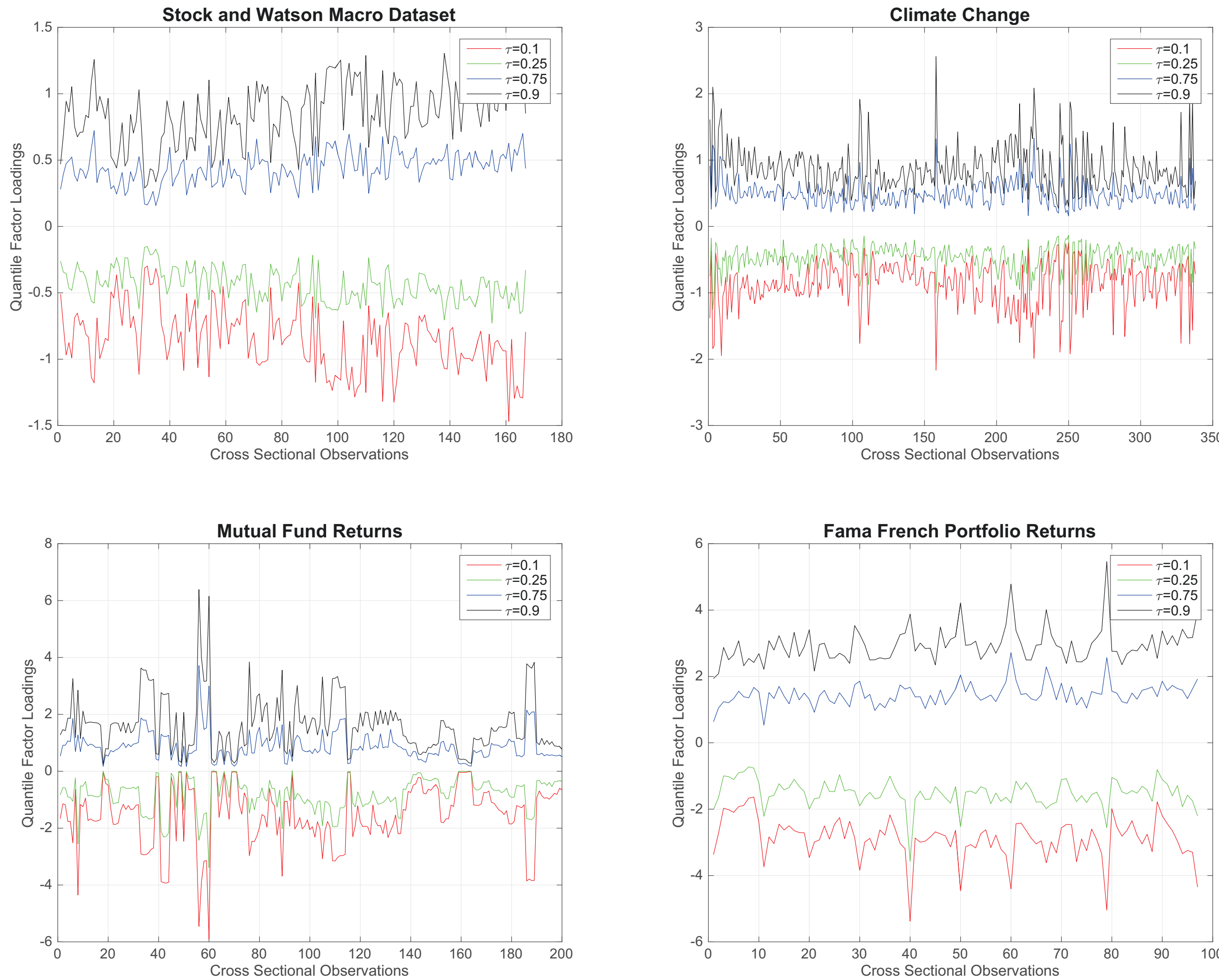

Table 1: Comparison of estimated factors: PCA, FIA, CA and PCA-SQ (Average $\bar{R}^2$s).

τ	$N = T$	f, $\hat{F}_{PC}$	g, $\hat{F}_{PC}$	f, $\hat{F}_{QR}$	g, $\hat{F}_{QR}$	g, $\hat{g}_{CA}$	g, $\hat{g}_{VF}$
0.10	20	.8519	.3294	.2581	.5171	.5701	.6766
	50	.9512	.2102	.6610	.8337	.8559	.8929
	100	.9788	.1872	.9184	.9377	.9222	.9343
0.25	20	.8743	.3399	.7148	.5479	.4395	.6890
	50	.9492	.2353	.9023	.7742	.7347	.8769
	100	.9783	.1548	.9602	.8812	.8698	.9456
0.75	20	.8531	.3277	.6943	.5389	.4169	.6690
	50	.9549	.2340	.9156	.7028	.7597	.8891
	100	.9779	.1492	.9606	.8765	.8749	.9452
0.90	20	.8569	.2912	.3396	.5164	.5498	.6868
	50	.9539	.1945	.6491	.8175	.8548	.8910
	100	.9790	.1607	.9221	.9404	.9369	.9473

* DGP: $X_{it} = \alpha_i f_t + g_t \epsilon_{it}$, where $f_t \sim i.i.d\ \mathcal{N}(0,1)$, $g_t = e^{h_t}$ with $h_t \sim i.i.d\ \mathcal{N}(0, 0.589)$ and $\alpha_i, \epsilon_{it} \sim i.i.d\ \mathcal{N}(0,1)$. Columns 3 to 6: average $\bar{R}^2$ (from 100 replications) of regressing f_t and g_t on $\hat{F}_{PC}$ and on $\hat{F}_{QR}$. Columns 7 and 8: average $\bar{R}^2$ of regressing g_t on $\hat{g}_{CA}$ and on $\hat{g}_{VF}$.

Table 2: Comparison of estimated quantile factors: CA and PCA-SQ (Average $\bar{R}^2$s).

τ	$N = T$	g_1, $\hat{g}_{CA}$	g_2, $\hat{g}_{CA}$	g_1, $\hat{g}_{VF}$	g_2, $\hat{g}_{VF}$
0.10	100	.3893	.3903	.3193	.3044
	200	.6114	.6135	.4725	.4684
	500	.9279	.9287	.4868	.4924
0.25	100	.4103	.4215	.4197	.4750
	200	.4762	.4562	.4865	.4488
	500	.7593	.7584	.4971	.4796
0.75	100	.4025	.4100	.4371	.4430
	200	.4937	.4642	.4884	.4653
	500	.7803	.7819	.4860	.4874
0.90	100	.4707	.4587	.4621	.4236
	200	.6319	.6313	.4796	.4551
	500	.9278	.9283	.4813	.4896

* DGP: $X_{it} = \alpha_i f_t + (\gamma_{1i} g_{1t} + \gamma_{2i} g_{2t})\epsilon_{it}$, where $f_t \sim i.i.d\ \mathcal{N}(0,1)$, $g_{jt} = e^{h_{jt}}$ with $h_{jt} \sim i.i.d\ \mathcal{N}(0, 0.589)$ for $j = 1, 2$, $\gamma_{1i}, \gamma_{2i} \sim i.i.d$ Uniform$(0,1)$, and $\alpha_i, \epsilon_{it} \sim i.i.d\ \mathcal{N}(0,1)$. Columns 3 to 6: average $\bar{R}^2$ (from 100 replications) of regressing g_{1t} and g_{2t} on $\hat{g}_{CA}$ and on $\hat{g}_{VF}$.

Table 3: Comparison of estimated quantile-dependent factors: Factor structure in cubic error terms (Average $\bar{R}^2$s).

τ	$N=T$	$s(\tau)$, $\hat{F}_{PC}$	$s(\tau)$, $\hat{g}_{CA}$	$s(\tau)$, $\hat{g}_{VF}$
0.10	100	.1102	.5623	.2365
	200	.0851	.7590	.2455
	500	.0420	.9218	.2833
0.25	100	.0330	.3177	.0807
	200	.0147	.6018	.0844
	500	.0121	.8647	.1057
0.75	100	.0411	.3429	.0891
	200	.0217	.5912	.0864
	500	.0111	.8601	.1005
0.90	100	.1049	.5594	.2288
	200	.0674	.7863	.2357
	500	.0414	.9241	.2929

* DGP: $X_{it} = \alpha_i f_t + g_t \epsilon_{it} + h_t \epsilon_{it}^3$, where $f_t \sim i.i.d\ \mathcal{N}(0,1)$, $g_t, \sqrt{2} h_t \sim i.i.d$ lognormal$(0, 0.589)$, and $\alpha_i, \epsilon_{it} \sim i.i.d\ \mathcal{N}(0,1)$. Columns 3 to 5: average $\bar{R}^2$ (from 100 replications) of regressing $s_t(\tau) = g_t + \Phi^{-1}(\tau)^2 h_t$ on $\hat{F}_{PC}$, on $\hat{g}_{CA}$ and on $\hat{g}_{VF}$.

Table 4: Comparison of mean and quantile factors in SW, Climate, MF and FF datasets (Average $\bar{R}^2$s).

SW	$k=1$	$k=2$	**Climate**	$k=1$	$k=2$
$p=3$	0.1359	0.0018	$p=1$	-0.0088	-0.0888
$p=4$	0.1086	0.0064	$p=2$	-0.0134	-0.0162
$p=5$	0.1682	0.0018	$p=3$	-0.0283	-0.0274
MF	$k=1$	$k=2$	**FF**	$k=1$	$k=2$
$p=1$	0.0263	0.0001	$p=3$	0.0163	-0.0007
$p=3$	0.0220	-0.0069	$p=4$	0.0241	-0.0012
$p=5$	0.0088	-0.0155	$p=5$	0.0242	-0.0046

* Average $\bar{R}^2$ of regressing each element of FQR_τ on $\hat{F}_{PC}$, for $\tau = 0.1, 0.25, 0.75, 0.9$, where FQR_τ are the estimated quantile factors, $\hat{F}_{PC}$ are the estimated mean factors, and p and k are the number of quantile factors and mean factors respectively.

Table 5: Regressions of volatility, skewness and kurtosis factors on quantile factors ($\bar{R}^2$s)

	VF_2 on FQR		VF_3 on FQR		VF_4 on FQR	
SW	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$
$p=3$	0.8480	0.8271	0.1386	0.3053	0.3200	0.3673
$p=4$	0.7949	0.8073	0.0160	0.0591	0.0094	0.0085
$p=5$	0.7694	0.8117	0.0080	0.0525	0.0093	-0.0064
Climate	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$
$p=1$	0.6824	0.6068	0.5001	0.7463	0.2141	0.2112
$p=2$	0.6516	0.5373	0.0109	0.2616	0.2861	0.1706
$p=3$	0.5738	0.5578	0.0450	0.6815	0.2716	0.2999
MF	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$
$p=1$	0.7590	0.7969	0.1991	0.5479	0.7377	0.8899
$p=3$	0.6648	0.7593	0.0013	0.1187	0.1089	0.2003
$p=5$	0.7837	0.7677	0.1920	0.2226	0.1749	0.2276
FF	$k=1$	$k=2$	$k=1$	$k=2$	$k=1$	$k=2$
$p=3$	0.2912	0.3133	0.0678	0.2153	0.2415	0.2647
$p=4$	0.6508	0.6445	0.0420	0.1900	0.3427	0.3314
$p=5$	0.7095	0.6894	0.0161	0.1310	0.2905	0.2917

* $\bar{R}^2$ of regressing VF_2, VF_3 and VF_4 on FQR, where VF_i $(i = 2, 3, 4)$ is the i-th root of the first principal component of the i-th power of the residuals. $FQR = [FQR_{0.1}, FQR_{0.25}, FQR_{0.75}, FQR_{0.9}]$, where FQR_τ are the estimated quantile factors at τ, and p and k are the number of quantile factors and mean factors respectively.

Appendix

In this Appendix we provide proofs of the main results in the paper (Theorems 1 to 5). To prove these Theorems, Lemmas 1 to 8 are required, whose proofs can be found in the online supplemental material.

A.1 Proof of Theorem 1

Lemma 1. *Define $C_{NT} = \min[N, T]$, the following results hold under Assumption 1:*
(i) $T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_{NT}'F_t\|^2 = O_P(C_{NT}^{-1})$.
(ii) $\|H_{NT} - H_0\| = o_P(1)$.

Proof of Theorem 1: Define:

$$\mathbb{Q}_\infty(\tau,\lambda) = \mathbb{E}[\rho_\tau(X_{it} - \lambda' H_0' F_t)],$$

$$\mathbb{Q}_T(\tau,\lambda) = T^{-1}\sum_{t=1}^{T}\rho_\tau(X_{it} - \lambda' H_0' F_t),$$

$$\hat{\mathbb{Q}}_T(\tau,\lambda) = T^{-1}\sum_{t=1}^{T}\rho_\tau(X_{it} - \lambda'\hat{F}_t).$$

First, it is easy to see that by Assumption 1 $H_0^{-1}\lambda_i(\tau)$ uniquely minimizes $\mathbb{Q}_\infty(\tau,\lambda)$ uniformly over $\tau \in \mathcal{T}$.

Second, notice that the function $(\tau,\lambda) \mapsto \rho_\tau(x - \lambda' H_0' f)$ is continuous for each $x \in \mathcal{X}$ and $f \in \mathcal{F}$, and $|\rho_\tau(X_{it} - \lambda' H_0' F_t)| \leq C \cdot \|\bar{\lambda}\| \cdot \|H_0\| \cdot \|F_t\|$ for some constant $C < \infty$ for all $(\tau,\lambda) \in \mathcal{T}\times\mathcal{A}$. Since $\mathbb{E}\|F_t\| < \infty$ and $\mathcal{A}$ is compact by Assumption 1, it follows that:

$$\sup_{(\tau,\lambda)\in\mathcal{T}\times\mathcal{A}} \|\mathbb{Q}_T(\tau,\lambda) - \mathbb{Q}_\infty(\tau,\lambda)\| = o_P(1) \tag{A.1}$$

by invoking Lemma 2.4 of Newey and McFadden (1994).

Third, by definition, $\hat{\lambda}_i(\tau)$ is the minimizer of $\hat{\mathbb{Q}}_T(\tau,\lambda)$ over $\mathcal{A}$ for each τ. Note that $\rho_\tau(u-v)-\rho_\tau(u) = v\varphi_\tau(u) + \int_0^v (\mathbf{1}\{u<s\} - \mathbf{1}\{u<0\})ds$, so:[17]

$$|\hat{\mathbb{Q}}_T(\tau,\lambda) - \mathbb{Q}_{T,H_0}(\tau,\lambda)| \leq C\cdot\|\lambda\|\cdot T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\| \leq C\cdot\|\lambda\|\cdot\sqrt{T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\|^2}$$

for some constant $C > 0$. By Lemma 1 we have $T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\|^2 \leq T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_{NT}'F_t\|^2 + \|H_{NT} - H_0\|^2\cdot T^{-1}\sum_{t=1}^{T}\|F_t\|^2 = o_P(1)$; it then follows that $\sup_{(\tau,\lambda)\in\mathcal{T}\times\mathcal{A}}|\hat{\mathbb{Q}}_T(\tau,\lambda) - \mathbb{Q}_T(\tau,\lambda)| = o_P(1)$.

[17] It then follows that $|\rho_\tau(u-v) - \rho_\tau(u)| \leq |v|\cdot|\mathbf{1}\{u<0\} - \tau| + |v|\cdot\mathbf{1}\{|u|<|v|\} \leq 3|v|$.

The latter result together with (A.1) imply: $\sup_{(\tau,\lambda)\in\mathcal{T}\times\mathcal{A}} |\hat{\mathbb{Q}}_T(\tau,\lambda) - \mathbb{Q}_\infty(\tau,\lambda)| = o_P(1)$. Since $\hat{\lambda}_i(\tau)$ is the minimizer of $\hat{\mathbb{Q}}_T(\tau,\lambda)$ by definition, and $H_0^{-1}\lambda_i(\tau)$ is the unique minimizer of $\mathbb{Q}_\infty(\tau,\lambda)$, it then follows from Lemma B.1 of Chernozhukov and Hansen (2006) that $\sup_{\tau\in\mathcal{T}} \|\hat{\lambda}_i(\tau) - H_0^{-1}\lambda_i(\tau)\| = o_P(1)$ for all i. Finally:

$$\sup_{\tau\in\mathcal{T}} \|\hat{\lambda}_i(\tau) - H_{NT}^{-1}\lambda_i(\tau)\| \leq \sup_{\tau\in\mathcal{T}} \|\hat{\lambda}_i(\tau) - H_0^{-1}\lambda_i(\tau)\| + \|H_{NT}^{-1} - H_0^{-1}\| \cdot \sup_{\tau\in\mathcal{T}} \|\lambda_i(\tau)\| = o_P(1).$$

A.2 Proof of Theorem 2

To simplify the notations, we suppress the subscription i and write $X_t, \lambda(\tau), \hat{\lambda}(\tau)$ instead of $X_{it}, \lambda_i(\tau), \hat{\lambda}_i(\tau)$. Define $\varphi_\tau(u) = \mathbf{1}\{u < 0\} - \tau$ and $\mathcal{D} = \{D \in \mathbb{R}^{r\times r} : D > 0 \text{ and } \|D\| < \infty\}$. Note that $H_0 \in \mathcal{D}$, and since $H_{NT} \xrightarrow{p} H_0$, we have that $H_{NT} \in \mathcal{D}$ with probability approaching 1. For any $D \in \mathcal{D}$, define:

$$\mathbb{S}_{\infty,D}(\tau,\lambda) = \mathbb{E}[\varphi_\tau(X_t - \lambda' D' F_t) D' F_t],$$

$$\mathbb{S}_{T,D}(\tau,\lambda) = T^{-1} \sum_{t=1}^{T} \varphi_\tau(X_t - \lambda' D' F_t) D' F_t,$$

$$\mathbb{G}_T(\tau,\lambda,D) = \sqrt{T}[\mathbb{S}_{T,D}(\tau,\lambda) - \mathbb{S}_{\infty,D}(\tau,\lambda)],$$

$$\hat{\mathbb{S}}_T(\tau,\lambda) = T^{-1} \sum_{t=1}^{T} \varphi_\tau(X_t - \lambda' \hat{F}_t) \hat{F}_t,$$

$$\hat{\mathbb{H}}_{T,D}(\tau,\lambda) = \sqrt{T}[\mathbb{S}_{T,D}(\tau,\lambda) - \hat{\mathbb{S}}_T(\tau,\lambda)].$$

The following lemmas hold under Assumptions 1 and 2:

Lemma 2. $\max_{1\leq t\leq T} \|\hat{F}_t - H_{NT}' F_t\| = O_P(T^{1/8}/\sqrt{N}) + O_P(T^{-5/8}) = o_P(T^{-1/2})$.

Lemma 3. $\sup_{\tau\in\mathcal{T}} \|\sqrt{T}\hat{\mathbb{S}}_T(\tau, \hat{\lambda}(\tau))\| = o_P(1)$.

Lemma 4. $\sup_{\tau\in\mathcal{T}} \|\mathbb{G}_T(\tau, \hat{\lambda}(\tau), H_{NT}) - \mathbb{G}_T(\tau, H_0^{-1}\lambda(\tau), H_0)\| = o_P(1)$.

Lemma 5. $\sup_{\tau\in\mathcal{T}} \left\|\hat{\mathbb{H}}_{T,H_{NT}}(\tau, \hat{\lambda}(\tau))\right\| = o_P(1)$.

Proof of Theorem 2:

First, we have the following expansion for each $\tau \in \mathcal{T}$:

$$\mathbb{S}_{\infty,H_{NT}}(\tau, \hat{\lambda}(\tau)) = \mathbb{S}_{\infty,H_{NT}}(\tau, H_{NT}^{-1}\lambda(\tau)) + H_{NT}'\mathbb{E}[\mathsf{f}_X(\lambda^*(\tau)' H_{NT}' F_t | F_t) F_t F_t'] H_{NT} \cdot [\hat{\lambda}(\tau) - H_{NT}^{-1}\lambda(\tau)],$$

where $\lambda^*(\tau)$ is on the line connecting $H_{NT}^{-1}\lambda(\tau)$ and $\hat{\lambda}(\tau)$ for each τ. Then, by uniform continuity of $\mathsf{f}_X(x|f)$ and uniform convergence of $\hat{\lambda}(\tau)$ for $H_{NT}^{-1}\lambda(\tau)$, we have that:

$$\mathbb{S}_{\infty,H_{NT}}(\tau, \hat{\lambda}(\tau)) = H_{NT}'[J(\lambda(\tau)) + o_P(1)] H_{NT} \cdot [\hat{\lambda}(\tau) - H_{NT}^{-1}\lambda(\tau)] \tag{A.2}$$

uniformly over $\mathcal{T}$ since $\mathbb{S}_{\infty,H_{NT}}(\tau, H_{NT}^{-1}\lambda(\tau)) = 0$.

Second, by definition we have:

$$\sqrt{T}\mathbb{S}_{\infty,H_{NT}}(\tau,\hat{\lambda}(\tau)) = -\mathbb{G}_T(\tau,\hat{\lambda}(\tau),H_{NT}) + \sqrt{T}\hat{\mathbb{S}}_T(\tau,\hat{\lambda}(\tau)) + \hat{\mathbb{H}}_{T,H_{NT}}(\tau,\hat{\lambda}(\tau)), \tag{A.3}$$

and combining Lemmas 1, 3, 4, 5, (A.2), and (A.3) yields:

$$[H_0'J(\lambda(\tau))H_0 + o_P(1)]\cdot\sqrt{T}[\hat{\lambda}(\tau) - H_{NT}^{-1}\lambda(\tau)] = -\mathbb{G}_T(\tau, H_0^{-1}\lambda(\tau), H_0) + o_P(1) \tag{A.4}$$

uniformly in $\tau \in \mathcal{T}$. It then follows from (A.4) and Assumption 2(iii) that:[18]

$$\sup_{\tau\in\mathcal{T}} \| -\mathbb{G}_T(\tau, H_0^{-1}\lambda(\tau), H_0) + o_P(1)\| \geq (\rho^* + o_P(1))\cdot \sup_{\tau\in\mathcal{T}}\sqrt{T}\|\hat{\lambda}(\tau) - H_{NT}^{-1}\lambda(\tau)\|. \tag{A.5}$$

Since the mapping $\tau \mapsto \lambda(\tau)$ is continuous due to implicit function theorem and Assumption 1(v) (see Angrist et al. 2006), the process $\mathbb{V}_T(\cdot) = \mathbb{G}_T(\cdot, H_0^{-1}\lambda(\cdot), H_0)$ is $\tilde{\rho}$-stochastic equicontinuous with:

$$\tilde{\rho}[\tau_1,\tau_2] = \rho[(\tau_1,\lambda(\tau_1)'H_0',H_0),(\tau_2,\lambda(\tau_2)'H_0',H_0)],$$

where ρ is defined in the proof of Lemma 4:

$$\rho[(\tau_1,\theta_1,D_1),(\tau_2,\theta_2,D_2)] = \sqrt{\max_{1\leq j\leq r}\mathbb{E}\Big[\varphi_{\tau_1}(X_t-\theta_1'F_t)D_1^{(j\cdot)}F_t - \varphi_{\tau_2}(X_t-\theta_2'F_t)D_2^{(j\cdot)}F_t\Big]^2},$$

where $D^{(\cdot j)}$ denotes the jth column of D. Then by stochastic equicontinuity and a standard multivariate central limit theorem, we have that

$$\mathbb{V}_T(\tau) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varphi_\tau(X_t - \lambda(\tau)'F_t)H_0'F_t$$

converges weakly to a zero mean Gaussian process $\mathbb{V}_\infty(\tau)$ defined by its covariance matrix:

$$\Sigma(\tau_1,\tau_2) = \mathbb{E}[\mathbb{V}_\infty(\tau_1)\mathbb{V}_\infty(\tau_2)] = [\min(\tau_1,\tau_2) - \tau_1\tau_2]H_0'\Sigma_F H_0.$$

It then follows from (A.5) that $\sup_{\tau\in\mathcal{T}}\sqrt{T}\|\hat{\lambda}(\tau) - H_{NT}^{-1}\lambda(\tau)\|$ is $O_P(1)$, and thus from (A.4) we can conclude that $[H_0'J(\lambda(\cdot)H_0]\cdot\sqrt{T}[\hat{\lambda}(\cdot) - H_{NT}^{-1}\lambda(\cdot)]$ converges weakly to $\mathbb{V}_\infty(\cdot)$ in $\ell^\infty(\mathcal{T})$. The desired result follows by noting that $H_0'\Sigma_F H_0 = I_r$. ■

A.3 Proof of Theorem 3

Proof of Theorem 3: Again, for simplicity, we suppress the subscript i. Recall that:

$$J_{H_0}(\lambda(\tau)) = \mathbb{E}\big[f_{X|F}(\lambda(\tau)'F_t|F_t)H_0'F_tF_t'H_0\big]$$

[18] For a symmetric positive definite matrix A and a non-zero vector a, $\|Aa\| = \sqrt{a'^2a} = \sqrt{(a/\|a\|)'^2(a/\|a\|)}\cdot\|a\| \geq \sqrt{\rho(A^2)}\|a\| = \rho(A)\|a\|$, where $\rho(\cdot)$ is the minimum eigenvalue.

and

$$\hat{J}(\hat{\lambda}(\tau)) = \frac{1}{2h_T \cdot T}\sum_{t=1}^{T}\Big\{\mathbf{1}\{|X_t - \hat{\lambda}(\tau)'\hat{F}_t| \leq h_T\}\hat{F}_t\hat{F}_t'\Big\}.$$

Define:

$$J(\hat{\lambda}(\tau)) = \frac{1}{2h_T \cdot T}\sum_{t=1}^{T}\Big\{\mathbf{1}\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T\}H_0'F_tF_t'H_0\Big\}.$$

Using Assumptions 1(i)(v), Lemma 1(ii) and the uniform consistency of $\hat{\lambda}(\tau)$ for $H_0^{-1}\lambda(\tau)$, it is easy to show that $\sup_{\tau\in\mathcal{T}}\|J_{H_0}(\lambda(\tau)) - J(\hat{\lambda}(\tau))\| = o_P(1)$[19]. Thus, the uniform consistency of $\hat{J}(\hat{\lambda}(\tau))$ follows from:

$$\sup_{\tau\in\mathcal{T}}\|\hat{J}(\hat{\lambda}(\tau)) - J(\hat{\lambda}(\tau))\| = o_P(1). \tag{A.6}$$

To prove (A.6), note that

$$\begin{aligned}
& 2h_T\big(\hat{J}(\hat{\lambda}(\tau)) - J(\hat{\lambda}(\tau))\big) \\
= & \underbrace{\frac{1}{T}\sum_{t=1}^{T}\Big\{\mathbf{1}\{|X_t - \hat{\lambda}(\tau)'\hat{F}_t| \leq h_T\}(\hat{F}_t\hat{F}_t' - H_0'F_tF_t'H_0)\Big\}}_{I} \\
& + \underbrace{\frac{1}{T}\sum_{t=1}^{T}\big[\mathbf{1}\{|X_t - \hat{\lambda}(\tau)'\hat{F}_t| \leq h_T\} - \mathbf{1}\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T\}\big]H_0'F_tF_t'H_0}_{II}.
\end{aligned}$$

First, we have that:

$$\begin{aligned}
\|I\| \leq\ & \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t\hat{F}_t' - H_0'F_tF_t'H_0\| \leq \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\|\|H_0'F_t\| + \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\|\|\hat{F}_t\| \\
\leq\ & 2\|H_0\|\cdot\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\|\|F_t\| + \frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_0'F_t\|^2 \\
\leq\ & 2\|H_0\|\cdot\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_{NT}'F_t\|\|F_t\| + 2\|H_0\|\cdot\|H_{NT} - H_0\|\cdot\frac{1}{T}\sum_{t=1}^{T}\|F_t\|^2 + 2\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_{NT}'F_t\|^2 \\
& +2\|H_{NT} - H_0\|^2\cdot\frac{1}{T}\sum_{t=1}^{T}\|F_t\|^2 \\
\leq\ & 2\|H_0\|\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_{NT}'F_t\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|F_t\|^2} + O_P(\|H_{NT} - H_0\|) + O_P(C_{NT}^{-1}) \\
=\ & \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H_{NT}'F_t\|^2}\cdot O_P(1) + \|H_{NT} - H_0\|\cdot O_P(1).
\end{aligned}$$

Then by Assumptions 2(ii), 3 and Lemma 1(i) we have $\|I\|/h_T = o_P(1)$ uniformly in τ.

Second, define $G_{ij,t}$ as the ith row and jth column of $H_0'F_tF_t'H_0$, and we first consider the case $i = j$

[19] The details of the proof are similar to that of equation (A.8) in Angrist et al. (2006) and is therefore omitted.

such that $G_{i,t} = G_{ii,t} \geq 0$. It is easy to see that:

$$II_{i,i} = \frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{1}\{|X_t - \hat{\lambda}(\tau)'\hat{F}_t| \leq h_T\} - \mathbf{1}\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T\}\right]G_{i,t}$$

is bounded below and above by:

$$\frac{1}{T}\sum_{t=1}^{T}\Big[\mathbf{1}\big\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T - \hat{k}_T(\tau)\big\} - \mathbf{1}\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T\}\Big]G_{i,t},$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\Big[\mathbf{1}\big\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T + \hat{k}_T(\tau)\big\} - \mathbf{1}\{|X_t - \hat{\lambda}(\tau)'H_{NT}'F_t| \leq h_T\}\Big]G_{i,t},$$

where $\hat{k}_T(\tau) = \max_{1\leq t\leq T}|\hat{\lambda}(\tau)'(\hat{F}_t - H_{NT}'F_t)|$. We now show that the upper bound is $o_P(T^{-1/2})$; the proof for the lower bound is similar. Consider the following empirical process:

$$\mathbb{C}_T(\theta, h) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big\{\mathbf{1}\{|X_t - \theta'F_t| \leq h\}\cdot G_{i,t} - \mathbb{E}[\mathbf{1}\{|X_t - \theta'F_t| \leq h\}\cdot G_{i,t}]\Big\}.$$

The upper bound can be written as:

$$\begin{aligned}
&\underbrace{T^{-1/2}\Big[\mathbb{C}_T(\theta,h)|_{h=h_T+\hat{k}_T(\tau),\theta=H_{NT}\hat{\lambda}(\tau)} - \mathbb{C}_T(\lambda,h)|_{h=h_T,\theta=H_{NT}\hat{\lambda}(\tau)}\Big]}_{III} \\
+\ &\underbrace{\mathbb{E}\big[\mathbf{1}\{|X_t - \theta'F_t| \leq h\}\cdot G_{i,t}\big]_{h=h_T+\hat{k}_T(\tau),\theta=H_{NT}\hat{\lambda}(\tau)} - \mathbb{E}\big[\mathbf{1}\{|X_t - \theta'F_t| \leq h\}\cdot G_{i,t}\big]_{h=h_T,\theta=H_{NT}\hat{\lambda}(\tau)}}_{IV}.
\end{aligned}$$

Since $\mathbb{C}_T(\theta, h)$ is stochastic equicontinous when $\mathbb{E}\|F_t\|^4 < \infty$ by Theorem 1 of Andrews (1994), it then follows that $\|III\|$ is $o_P(T^{-1/2})$ uniformly in τ given that:

$$\sup_{\tau\in\mathcal{T}}|\hat{k}_T(\tau)| \leq \sup_{\lambda\in\mathcal{A}}\|\lambda\|\cdot\max_{1\leq t\leq T}\|\hat{F}_t - H_{NT}'F_t\| = o_P(1). \tag{A.7}$$

Next, note that

$$\begin{aligned}
&\mathbb{E}\big[\mathbf{1}\{|X_t - \theta'F_t| \leq h_1\}\cdot G_{i,t}\big] - \mathbb{E}\big[\mathbf{1}\{|X_t - \theta'F_t| \leq h_2\}\cdot G_{i,t}\big] \\
=\ &\mathbb{E}\Big[\Big(f_{X|F}(\theta'F_t + h^*) - f_{X|F}(\theta'F_t + h^{**})\Big)(h_1 - h_2)G_{i,t}\Big],
\end{aligned}$$

where h^* and h^{**} are points on the lines connecting h_1 and h_2. We then have $\|IV\| \leq 2\bar{f}\cdot\mathbb{E}\|G_{i,t}\|\cdot|\hat{k}_T(\tau)| = o_P(T^{-1/2})$ uniformly in τ by (A.7). The proof for the case where $i \neq j$ is similar. Combining the above results and Assumption 3, (A.6) follows directly and thus the first statement in Theorem 3 is proved. The second statement follows trivially by Slutsky's Theorem and the fact that $\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t\hat{F}_t' - H_0'F_tF_t'H_0\| = o_P(1)$. ■

A.4 Proof of Theorem 4

We begin by introducing some notation and two useful lemmas. Let ψ be a nondecreasing, convex function with $\psi(0) = 0$ and Z a random variable. The Orlicz norm $\|X\|_\psi$ is defined as:

$$\|X\|_\psi = \inf\big\{C > 0 : E\psi\big(|X|/C\big) \leq 1\big\}.$$

Note that when $\psi(x) = x^p$ for $p \geq 1$, the Orlicz norm is simply the L_p-norm: $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. Define $\psi_p(x) = e^{x^p} - 1$, then we have the following inequality:

$$\|X\|_1 \leq \|X\|_{\psi_1}.$$

Lemma 6 (Lemma 2.2.10 of Van der Vaart and Wellner 1996)**.** *Let $Z_1, \dots, Z_m$ be arbitrary random variables that satisfy the tail bound:*

$$P[|Z|_i > z] \leq 2e^{-\frac{1}{2}\frac{z^2}{b+az}}$$

for all z and fixed $a, b > 0$. Then:

$$\Big\| \max_{1\leq i\leq m} Z_i \Big\|_{\psi_1} \leq K\Big(a\log(1+m) + \sqrt{b}\sqrt{\log(1+m)}\Big).$$

Lemma 7 (Lemma 2.2.11 of Van der Vaart and Wellner 1996)**.** *Let $Y_1, \dots, Y_n$ be independent random variables with zero mean such that $E|Y_i|^m \leq m!M^{m-2}v_i/2$, for every $m \geq 2$ and some constants M and v_i. Then:*

$$P[|Y_1 + \dots + Y_n| > x] \leq 2e^{-\frac{1}{2}\frac{x^2}{v+Mx}},$$

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

Lemma 8. *Let X be a random variable with mean μ, distribution function F_X, and continuous density function f_X. Then for any $x \in \mathbb{R}$ we have that:*

$$\sup_{c\in\mathbb{R}} |\rho_\tau(x-c) - \mathbb{E}\rho_\tau(X-c)| \leq |x-\mu| + \mathbb{E}|X-\mu|.$$

Proof of Theorem 4:

Step 1: The first step is to show that:

$$\sup_{F_t\in\mathcal{F},\lambda_i\in\mathcal{A}} \left|(NT)^{-1}\sum_{i=1,t=1}^{N,T}\big[\rho_\tau(X_{it}-\lambda_i'F_t) - \mathbb{E}\rho_\tau(X_{it}-\lambda_i'F_t)\big]\right| = o_P(1). \tag{A.8}$$

Define $\mathcal{D} = \{\lambda'F : \lambda \in \mathcal{A}, F \in \mathcal{F}\}$, then it is obvious that $\mathcal{D}$ is a compact subset of $\mathbb{R}$. Let $\tilde{B}_h(x)$ be a closed ball with center x and radius $h > 0$. Then for any $\gamma > 0$, $\mathcal{A}$ and $\mathcal{F}$ are covered by $\bigcup_{\lambda\in\mathcal{A}}\tilde{B}_\gamma(\lambda)$ and $\bigcup_{F\in\mathcal{F}}\tilde{B}_\gamma(F)$ respectively. It then follows from Assumption 4(i) that there exists finite constants K_1

and K_2, and $a_1, \dots a_{K_1} \in \mathcal{A}$, $f_1, \dots f_{K_2} \in \mathcal{F}$ such that:

$$\mathcal{A} \subset \bigcup_{1 \le k \le K_1} \tilde{B}_\gamma(a_k) \quad \text{and} \quad \mathcal{F} \subset \bigcup_{1 \le k \le K_2} \tilde{B}_\gamma(f_k).$$

These covering balls for $\mathcal{A}$ can be easily transformed into disjoint neighbourhoods $B_\gamma(a_1), \dots, B_\gamma(a_{K_1})$ whose union covers $\mathcal{A}$, and for each $a \in B_\gamma(a_k)$ we have $|a - a_k| \le \gamma$. The same procedure can be used for the covering balls of $\mathcal{F}$.

Thus, if $d \in \mathcal{D}$, then there exists λ, F and some $k_1 \le K_1$, $k_2 \le K_2$, such that $d = \lambda' F$, $\lambda \in B_\gamma(a_{k_1})$ and $F \in B_\gamma(f_{k_2})$. In other words, the set $\mathcal{D}$ is covered by:

$$\bigcup_{k=1}^{K_1} \bigcup_{s=1}^{K_2} H(a_k, f_s), \qquad \text{where } H(a_k, f_s) = \{\lambda' F : \lambda \in B_\gamma(a_k), F \in B_\gamma(f_s)\}.$$

Now suppose $d \in H(a_k, f_s)$, then by the above definition there exists $\lambda \in B_\gamma(a_k), F \in B_\gamma(f_s)$ such that $d = \lambda' F$. Define $M > 0$ as an upper bound for the norms of vectors in $\mathcal{F}$ and $\mathcal{A}$, and $\delta = \gamma^2 + 2\gamma M$. We have that:

$$|d - a_k' f_s| = |\lambda' F - a_k' f_s| \le |\lambda - a_k| \cdot |F - f_s| + |a_k| \cdot |F - f_s| + |\lambda - a_k| \cdot |f_s| \le \gamma^2 + 2\gamma M = \delta,$$

implying that $H(a_k, f_s) \subseteq \tilde{B}_\delta(a_k' f_s)$, and thus

$$\mathcal{D} \subset \bigcup_{k=1}^{K_1} \bigcup_{s=1}^{K_2} \tilde{B}_\delta(a_k' f_s).$$

Suppose there are K elements in the set $\{d : d = a_k' f_s, 1 \le k \le K_1, 1 \le s \le K_2\}$, then it is obvious that $K \le K_1 K_2$. Call these elements $d_1, d_2, \dots, d_K$, and define $D_K = \{d_1, \dots, d_K\}$. Then $\mathcal{D}$ is covered by the union of $\tilde{B}_\delta(d_1), \dots, \tilde{B}_\delta(d_K)$.

The above analysis implies that for any $\delta > 0$, there exists a finite constant K which depends on δ,[20] and a sequence $d_1, \dots, d_K \in \mathcal{D}$ such that $\mathcal{D}$ is covered by the union of K closed balls $\tilde{B}_\delta(d_1), \dots, \tilde{B}_\delta(d_K)$.

Returning to Equation (A.8). For given Λ and F, define $c_{it} = \lambda_i' F_t$ and:

$$d_{it} = a_k' f_s \in D_K \text{ if } \lambda_i \in B_\gamma(a_k), F_t \in B_\gamma(f_s). \tag{A.9}$$

Note that even though d_{it} is indexed by both i and t, it can only take K different values, and it is obvious

[20]This K depends on K_1, K_2 and therefore depends on γ, and finally depends on δ through the equality: $\delta = \gamma^2 + 2\gamma M$

that $|c_{it} - d_{it}| \le \delta$. First, note that:

$$
\begin{aligned}
& \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - \lambda_i' F_t) - \mathbb{E}\rho_\tau(X_{it} - \lambda_i' F_t) \right] \right| \\
\le \ & \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - c_{it}) - \rho_\tau(X_{it} - d_{it}) \right] \right| + \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\mathbb{E}\rho_\tau(X_{it} - c_{it}) - \mathbb{E}\rho_\tau(X_{it} - d_{it}) \right] \right| \\
& + \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - d_{it}) - \mathbb{E}\rho_\tau(X_{it} - d_{it}) \right] \right| \\
= \ & I + II + III.
\end{aligned}
$$

Next, define:

$$
s(x, \delta) = \sup_{c_1, c_2 \in \mathcal{D}, |c_1 - c_2| \le \delta} \left| \rho_\tau(x - c_1) - \rho_\tau(x - c_2) \right|.
$$

Since the mapping $c \mapsto \rho_\tau(x, c)$ is continuous for any x, it is uniform continuous over the compact set $\mathcal{D}$, implying that $s(X_{it}, \delta)$ converges to 0 almost surely as $\delta \to 0$. Moreover, it is easy to see that, based on Assumption 4, there exists a function $G(x) > 0$ such that $|\rho_\tau(x - c)| \le G(x)$ for all $c \in \mathcal{D}$, and $\mathbb{E}G(X_{it}) < \infty$ for all i, t. It follows from dominated convergence that $\mu_{it}(\delta) = \mathbb{E}[s(X_{it}, \delta)] \to 0$ as $\delta \to 0$ for all i, t. Therefore, we have that:

$$
\begin{aligned}
I \ & \le \ (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left| \rho_\tau(X_{it} - c_{it}) - \rho_\tau(X_{it} - d_{it}) \right| \\
& \le \ (NT)^{-1} \sum_{i=1,t=1}^{N,T} [s(X_{it}, \delta) - \mu_{it}(\delta)] + (NT)^{-1} \sum_{i=1,t=1}^{N,T} \mu_{it}(\delta) = I(i) + I(ii).
\end{aligned}
$$

It can be shown in a similar way that term II is bounded by $I(ii)$. Note that $I(i)$ and $I(ii)$ do not depend on λ_i, F_t, and that $I(i)$ is $o_P(1)$ by a law of large numbers, $I(ii)$ is positive and can be made arbitrarily small by choosing a small enough δ. So the term $I + II$ is uniformly $o_P(1)$, and it remains to show that III is uniformly $o_P(1)$.

For a given $\delta > 0$ and the corresponding K, define the following set:

$$
\Theta = \Big\{ \theta = [d_{11}, \ldots, d_{1T}, \ldots, d_{it}, \ldots, d_{N1}, \ldots, d_{NT}] : F_t \in \mathcal{F}, \lambda_i \in \mathcal{A}, 1 \le t \le T, 1 \le i \le N \Big\},
$$

where d_{it} is defined as in (A.9). This set contains all the possible values that θ can take when each λ_i and F_t are free to take values in $\mathcal{A}$ and $\mathcal{F}$. Recallr that, for any given δ, we can find a positive γ through the equation $\delta = \gamma^2 + 2\gamma M$, and therefore find finite constants K_1 and K_2 such that $\mathcal{A}$ and $\mathcal{F}$ are covered by K_1 and K_2 disjoint neighbourhoods respectively. Note that $d_{it} = a_k' f_s$ for all $\lambda_i \in B_\gamma(a_k)$ and all $F_t \in B_\gamma(f_s)$, so we have $S \le K_1^N K_2^T$ where S is the number of elements of the set Θ.[21] Denote these

[21] Without the restrictions in (A.9), each d_{it} can take K different values, and the number of elements of Θ will be K^{NT}.

elements as $\theta_1, \ldots, \theta_S$, where:

$$\theta_j = [d_{j,11}, \ldots, d_{j,1T}, \ldots, d_{j,it}, \ldots, d_{j,N1}, \ldots, d_{j,NT}],$$

and define:

$$S_{NT}(\theta_j) = \left| \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - d_{j,it}) - \mathbb{E}\rho_\tau(X_{it} - d_{j,it}) \right] \right|.$$

Then we have:

$$\sup_{F_t \in \mathcal{F}, \lambda_i \in \mathcal{A}} \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - d_{it}) - \mathbb{E}\rho_\tau(X_{it} - d_{it}) \right] \right| = (NT)^{-1} \max_{1 \leq j \leq S} S_{NT}(\theta_j).$$

Note that:

$$\begin{aligned} & \sup_{c \in \mathcal{D}} \left| \rho_\tau(X_{it} - c) - \mathbb{E}\rho_\tau(X_{it} - c) \right| \\ \leq \quad & \sup_{c \in \mathbb{R}} \left| \rho_\tau(X_{it} - c) - \mathbb{E}\rho_\tau(X_{it} - c) \right| \\ = \quad & \sup_{c \in \mathbb{R}} \left| \rho_\tau(v_{it} + \lambda_i^{0'} F_t^0 - c) - \mathbb{E}\rho_\tau(v_{it} + \lambda_i^{0'} F_t^0 - c) \right| \\ = \quad & \sup_{c \in \mathbb{R}} \left| \rho_\tau(v_{it} - c) - \mathbb{E}\rho_\tau(v_{it} - c) \right| \\ \leq \quad & |v_{it} - \mathbb{E}v_{it}| + \mathbb{E}|v_{it} - \mathbb{E}v_{it}|, \end{aligned}$$

where the last inequality follows from Lemma 8. Thus, for all i, t, it holds that:

$$\begin{aligned} & \mathbb{E}|\rho_\tau(X_{it} - d_{j,it}) - \mathbb{E}\rho_\tau(X_{it} - d_{j,it})|^m \\ \leq \quad & \mathbb{E}\Big[|v_{it} - \mathbb{E}v_{it}| + \mathbb{E}|v_{it} - \mathbb{E}v_{it}| \Big]^m \\ \leq \quad & 2^{m-1} \Big[\mathbb{E}|v_{it} - \mathbb{E}v_{it}|^m + (\mathbb{E}|v_{it} - \mathbb{E}v_{it}|)^m \Big]. \end{aligned}$$

It then follows from Assumption 4(iv) that there exists finite constants M_1 and M_2 such that:

$$\mathbb{E}|\rho_\tau(X_{it} - d_{j,it}) - \mathbb{E}\rho_\tau(X_{it} - d_{j,it})|^m \leq m! M_1^{m-2} M_2 / 2$$

for all i, t. It then follows from Lemma 7 that

$$P[S_{NT}(\theta_j) > x] \leq 2e^{-\frac{1}{2} \frac{x^2}{(N+T)M_2 + M_1 x}}$$

for all j. Finally, applying Lemma 6 we have that for some K^* that only depends on the function ψ_1,

$$\begin{aligned} \left\| \max_{1 \leq j \leq S} S_{NT}(\theta_j) \right\|_{\psi_1} \quad & \leq \quad K^* \Big(M_1 \log(1 + S) + \sqrt{(N+T)M_2} \sqrt{\log(1 + S)} \Big) \\ & \leq \quad M_3 \left(N \log K_1 + T \log K_2 + \sqrt{N+T} \sqrt{N \log K_1 + T \log K_2} \right) \end{aligned}$$

for some finite constant M_3. As a result of the above inequality, we have that:

$$\sup_{F_t \in \mathcal{F}, \lambda_i \in \mathcal{A}} \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - d_{it}) - \mathbb{E}\rho_\tau(X_{it} - d_{it}) \right] \right| = (NT)^{-1} \max_{1 \leq j \leq S} S_{NT}(\theta_j) = O_P(N^{-1} + T^{-1}) = o_P(1).$$

Thus, we have proved that Equation (A.8) holds by showing that I, II and III are all uniformly $o_P(1)$.

Step 2:

For any given Λ and F, we have that:

$$\begin{aligned}
& \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \rho_\tau(X_{it} - \lambda_i' F_t) - \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \rho_\tau(X_{it} - \lambda_i^{0'} F_t^0) \\
= \ & \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - \lambda_i' F_t) - \mathbb{E}\rho_\tau(X_{it} - \lambda_i' F_t) \right] + \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - \lambda_i^{0'} F_t^0) - \mathbb{E}\rho_\tau(X_{it} - \lambda_i^{0'} F_t^0) \right] \\
& + \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \left[\mathbb{E}\rho_\tau(X_{it} - \lambda_i' F_t) - \mathbb{E}\rho_\tau(X_{it} - \lambda_i^{0'} F_t^0) \right] \\
= \ & IV + V + VI.
\end{aligned}$$

The result of the first step implies that:

$$\sup_{F_t \in \mathcal{F}, \lambda_i \in \mathcal{A}} |IV + V| \leq 2 \sup_{F_t \in \mathcal{F}, \lambda_i \in \mathcal{A}} \left| (NT)^{-1} \sum_{i=1,t=1}^{N,T} \left[\rho_\tau(X_{it} - \lambda_i' F_t) - \mathbb{E}\rho_\tau(X_{it} - \lambda_i' F_t) \right] \right| = o_P(1).$$

Let $h_{it} = \lambda_i' F_t - \lambda_i^{0'} F_t^0$. By Taylor expansion we get:

$$\mathbb{E}\rho_\tau(X_{it} - \lambda_i' F_t) - \mathbb{E}\rho_\tau(X_{it} - \lambda_i^{0'} F_t^0) = \mathbb{E}\rho_\tau(v_{it} - h_{it}) - \mathbb{E}\rho_\tau(v_{it}) = \mathsf{f}_{it}(\tilde{v}_{it}) h_{it}^2 \geq \underline{\mathsf{f}} \cdot h_{it}^2,$$

for some $\underline{\mathsf{f}} > 0$ according to Assumption 4(iii), where $\tilde{v}_{it}$ is between 0 and h_{it}.

Next, by the above results and the definition of $\hat{\Lambda}$ and $\hat{F}$, we have

$$0 \geq \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \rho_\tau(X_{it} - \hat{\lambda}_i' \hat{F}_t) - \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \rho_\tau(X_{it} - \lambda_i^{0'} F_t^0) \geq o_P(1) + \underline{\mathsf{f}} \frac{1}{NT} \sum_{i=1,t=1}^{N,T} \hat{h}_{it}^2,$$

where the $o_P(1)$ is uniform over $\mathcal{A}$ and $\mathcal{F}$, and $\hat{h}_{it} = \hat{\lambda}_i' \hat{F}_t - \lambda_i^{0'} F_t^0$. Since $\underline{\mathsf{f}} > 0$ the above inequality implies that

$$\frac{1}{NT} \sum_{i=1,t=1}^{N,T} (\hat{\lambda}_i' \hat{F}_t - \lambda_i^{0'} F_t^0)^2 = \frac{1}{NT} \| \hat{F} \hat{\Lambda}' - F^0 \Lambda^{0'} \|^2 = o_P(1).$$

Finally, define $M_{\hat{F}} = I - \hat{F}(\hat{F}'\hat{F})^{-1}\hat{F}'$, and let $\|\cdot\|_s$ denote the spectral norm of a matrix. Then, it holds that:

$$\|M_{\hat{F}}(\hat{F}\hat{\Lambda}' - F^0\Lambda^{0'})\| \leq \sqrt{\text{rank}\left(M_{\hat{F}}(\hat{F}\hat{\Lambda}' - F^0\Lambda^{0'})\right)} \cdot \|M_{\hat{F}}\|_S \cdot \|\hat{F}\hat{\Lambda}' - F^0\Lambda^{0'}\|_S.$$

Since $\text{rank}(M_{\hat{F}}) = T - r$, $\text{rank}(\hat{F}\hat{\Lambda}' - F^0\Lambda^{0'}) \leq 2r$, $\|M_{\hat{F}}\|_S = 1$ and $\|\hat{F}\hat{\Lambda}' - F^0\Lambda^{0'}\|_S \leq \|\hat{F}\hat{\Lambda}' - F^0\Lambda^{0'}\|$, it follows that:

$$(NT)^{-1/2}\|M_{\hat{F}}F^0\Lambda^{0'}\| = \sqrt{\text{Tr}\left[\frac{F^{0'}M_{\hat{F}}F^0}{T} \cdot \frac{\Lambda^{0'}\Lambda^0}{N}\right]} = o_P(1).$$

Since $N^{-1}\Lambda^{0'}\Lambda^0$ converges to a full rank matrix by Assumption 4(ii), it then follows that:

$$\left\|\frac{F^{0'}M_{\hat{F}}F^0}{T}\right\| = o_P(1),$$

which implies:

$$F^{0'}F^0/T - (F^{0'}\hat{F}/T)(\hat{F}'^0/T) = o_P(1).$$

Consequently:

$$\|P_{\hat{F}} - P_{F^0}\|^2 = \text{Tr}[P_{\hat{F}}] + \text{Tr}[P_{F^0}] - 2\,\text{Tr}[P_{\hat{F}} \cdot P_{F^0}] = \text{Tr}\left[(F^{0'}F^0/T)^{-1}\left(F^{0'}F^0/T - (F^{0'}\hat{F}/T)(\hat{F}'^0/T)\right)\right],$$

which is equal to $o_P(1)$ since $F^{0'}F^0/T$ converges to a positive definite matrix by Assumption 4(ii). We can show that $\|P_{\hat{\Lambda}} - P_{\Lambda^0}\| = o_P(1)$ in a similar way. Then the proof of Theorem 4 is complete. ■

A.5 Proof of Theorem 5

Proof of Theorem 5: Assumption 5 (i) and (ii) ensures that Assumptions A to D of Bai and Ng (2002) are satisfied. It then follows directly from Theorem 1 of Bai and Ng (2002) that:

$$T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H'_{NT}F_t\|^2 = o_P(1)$$

and

$$N^{-1}\sum_{i=1}^{N}\|\hat{\lambda}_i - H_{NT}^{-1}\lambda_i\|^2 = o_P(1).$$

To prove the consistency of the estimated quantile-shifting factors, we need to show that:

$$\sup_{G_t\in\mathcal{G},\gamma_i\in\mathcal{A}}\left|(NT)^{-1}\sum_{i=1,t=1}^{N,T}\left[\rho_\tau(\hat{e}_{it} - \gamma_i'G_t) - \mathbb{E}\rho_\tau(e_{it} - \gamma_i'G_t)\right]\right| = o_P(1),$$

and then the rest of the proof follows exactly as the second step of the proof of Theorem 4. Noticing that:

$$
\begin{aligned}
&(NT)^{-1}\sum_{i=1,t=1}^{N,T}\left[\rho_\tau(\hat{e}_{it}-\gamma_i'G_t)-\mathbb{E}\rho_\tau(e_{it}-\gamma_i'G_t)\right]\\
&=(NT)^{-1}\sum_{i=1,t=1}^{N,T}\left[\rho_\tau(\hat{e}_{it}-\gamma_i'G_t)-\rho_\tau(e_{it}-\gamma_i'G_t)\right]+(NT)^{-1}\sum_{i=1,t=1}^{N,T}\left[\rho_\tau(e_{it}-\gamma_i'G_t)-\mathbb{E}\rho_\tau(e_{it}-\gamma_i'G_t)\right],
\end{aligned}
$$

it then follows from the proof of Theorem 4 that:

$$
\sup_{G_t\in\mathcal{G},\gamma_i\in\mathcal{A}}\left|(NT)^{-1}\sum_{i=1,t=1}^{N,T}\left[\rho_\tau(e_{it}-\gamma_i'G_t)-\mathbb{E}\rho_\tau(e_{it}-\gamma_i'G_t)\right]\right|=o_P(1).
$$

Therefore it remains to show that:

$$
\sup_{G_t\in\mathcal{G},\gamma_i\in\mathcal{A}}\left|(NT)^{-1}\sum_{i=1,t=1}^{N,T}\left[\rho_\tau(\hat{e}_{it}-\gamma_i'G_t)-\rho_\tau(e_{it}-\gamma_i'G_t)\right]\right|=o_P(1). \tag{A.10}
$$

It is easy to see that the left-hand side of the above equation is bounded (up to a constant) by:

$$
\begin{aligned}
\frac{1}{NT}\sum_{i=1,t=1}^{N,T}\left|\hat{e}_{it}-e_{it}\right|&=\frac{1}{NT}\sum_{i=1,t=1}^{N,T}\left|\hat{\lambda}_i'\hat{F}_t-\lambda_i'F_t\right|\\
&\leq\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i-\lambda_i\|\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t\|+\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i\|\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t-F_t\|.
\end{aligned}
$$

Finally, it follows from the consistency of PC estimators that:

$$
\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i-\lambda_i\|\leq\sqrt{\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i-\lambda_i\|^2}=o_P(1)\text{ and }\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t-F_t\|\leq\sqrt{\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t-F_t\|^2}=o_P(1),
$$

$T^{-1}\sum_{t=1}^{T}\|\hat{F}_t\|\leq\sqrt{T^{-1}\sum_{t=1}^{T}\|\hat{F}_t\|^2}=\sqrt{r}$, and $N^{-1}\sum_{i=1}^{N}\|\lambda_i\|$ is $O(1)$ by assumption, so (A.10) holds and the proof is complete. ■

References

Abrevaya, J. and C. M. Dahl (2008). The effects of birth inputs on birthweight: evidence from quantile estimation on panel data. *Journal of Business & Economic Statistics 26*(4), 379–397.

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

Andrews, D. W. (1994). Empirical process methods in econometrics. *Handbook of Econometrics 4*, 2247–2294.

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

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

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

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

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

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

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

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

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

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

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal 14*(3), 368–386.

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

Chen, M., I. Fernández-Val, and M. Weidner (2014). Nonlinear panel models with interactive effects. *Working paper, UCL.*

Chernozhukov, V. (2005). Extremal quantile regression. *Annals of Statistics*, 806–839.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics 132*(2), 491–525.

Chernozhukov, V. and C. Hansen (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics 142*(1), 379–398.

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

Diebold, F. X. (2012). A personal perspective on the origin(s) and development of 'big data': The phenomenon, the term, and the discipline. *Working paper, UPenn*.

Durbin, J. (1973). *Distribution theory for tests based on the sample distribution function*, Volume 9. Siam.

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 (2017). International tests of a five-factor asset pricing model. *Journal of Financial Economics 123*(3), 441–463.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large N, T. *Journal of Econometrics 192*(1), 291–312.

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

Graham, B. S., J. Hahn, and J. L. Powell (2009). The incidental parameter problem in a non-differentiable panel data model. *Economics Letters 105*(2), 181–182.

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

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

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.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis 91*(1), 74–89.

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

Koenker, R. and Z. Xiao (2002). Inference on the quantile regression process. *Econometrica 70*(4), 1583–1612.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics 157*(2), 396–408.

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

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

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics 32*(1), 143–155.

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

Rosen, A. M. (2012). Set identification via quantile restrictions in short panels. *Journal of Econometrics 166*(1), 127–137.

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

Stock, J. H. and M. Watson (2016). Factor models and structural vector autoregressions in macroeconomics. *Handbook of Marcroeconomics 2A*, 415 – 526.

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

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

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