Journal of Econometrics 244 (2024) 105765

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Functional quantile autoregression ☆

Chaohua Dong [a], Rong Chen [b], Zhijie Xiao [c], Weiyi Liu [d,*]

[a] Zhongnan University of Economics and Law, China
[b] Rutgers University, USA
[c] Boston College, USA
[d] Capital University of Economics and Business, China

ARTICLE INFO

JEL classification:
C14
C22

Keywords:
Distributional dynamics
Functional dependence
GARCH
Quantile autoregression
Sieve estimation

ABSTRACT

This paper proposes a new class of time series models, the functional quantile autoregression (FQAR) models, in which the conditional distribution of the observation at the current time point is affected by its past distributional information, and is expressed as a functional of the past conditional quantile functions. Different from the conventional functional time series models which are based on functionally observed data, the proposed FQAR method studies functional dynamics in traditional time series data. We propose a sieve estimator for the model. Asymptotic properties of the estimators are derived. Numerical investigations are conducted to highlight the proposed method.

1. Introduction

Distributional information plays an important role in many time series applications. In the field of economics and finance, the distribution of many economic and financial variables may heavily depend on the past distributional information of itself or other variables. Such distributional information may include quantiles, volatility, skewness, kurtosis, left and right tail information, Gini index and many others. Accumulated empirical evidence indicates that tail information of asset returns in the past can affect future distributions of many financial and economic variables; the income distribution in the past may have important influence on future consumption in the economy; the distributional information of stock returns contains important information for risk management and investment. See, *inter alia*, Engle and Manganelli (2004), Sim and Zhou (2015), Adrian et al. (2019), Millick et al. (2019), Patton et al. (2019) and Chang and Schorfheide (2022). For example, Sim and Zhou (2015) studied the empirical relationship between the distributions of oil price shocks and stock returns and found distributional dependence between these time series. Adrian et al. (2019) studied the dynamics of the conditional distribution of real GDP growth and found strong evidence that both the left tail and the center quantiles of the future conditional distribution are affected by the past left tail distributional information.

Any attempt at studying, predicting, or testing such time series requires appropriate models to capture such kind distributional dependence. Most existing approaches focus on modeling selected distributional characteristics, such as the conditional variance or a

☆ We are most grateful to the Associate Editor for carefully going through earlier versions of the paper and providing many pertinent comments and constructive suggestions that markedly improved the paper. We also thank the Editor, three referees, Marco Del Negro, Jean-Jacques Forneron, Jihyung Lee, Zhongjun Qu, Neil Shephard, Songhua Tan, and participants of the 2022 NBER/NSF conference on time series for helpful comments and suggestions on previous versions of this paper. Dong thanks the financial support from National Natural Science Foundation of China (Grant No. 72073143). Xiao thanks Boston College for research support. Liu thanks the support by the 111 Project (Grant No. D23001), National Natural Science Foundation of China (Grant No. 72071134), Humanities and Social Science Fund of Ministry of Education of China (Grant No. 21YJC790078), and the Project of Cultivation for Young Top-Notch Talents of Beijing Municipal Institutions (Grant No. BPHR202203176).

* Corresponding author.
E-mail address: liuweiyi@cueb.edu.cn (W. Liu).

https://doi.org/10.1016/j.jeconom.2024.105765
Received 15 December 2022; Received in revised form 3 April 2024; Accepted 26 April 2024
Available online 14 May 2024
0304-4076/

specific quantile of the conditional distributions. For example, the ARCH or GARCH type of models emphasize the role of conditional volatility and model the dynamics of the conditional variance of returns. The (G)ARCH-in-Mean model studies how returns are affected by past distributional information represented by their conditional mean and conditional variance. The autoregressive conditional Fréchet model (Zhao et al., 2019) studies how the maxima of financial time series evolves through the parameters of a conditional Fréchet Distribution. The Conditional Autoregressive Value at Risk (CaViaR) model (Engle and Manganelli, 2004) argues that the conditional quantile of the future return depends on a conditional quantile of the previous return distribution.

There is no doubt that these selected distributional characteristics contain important information about the underlying distribution. However, focusing exclusively on a limited subset of distributional characteristics may also potentially miss interesting relationships among other characteristics of the distributions. As argued in the finance and economic literature, information of various quantiles of the whole past distribution can have important influence on the future behavior of many economic variables. For example, Adrian et al. (2019) found that the entire conditional distribution of GDP growth evolves over time. Yahya et al. (2020) and Pham (2021) observed cross-quantile dependence in financial markets. Researchers also found that past income distributional information plays an important role on the future economic behavior. Specifically, when the information of the entire past income distribution is considered, the empirical results may differ from those that only use some specific quantile information of the past distribution (see, e.g., Coibion et al., 2017; Chang and Schorfheide, 2022).

We believe that functional time series models are able to provide a useful approach in addressing some of these issues in modeling dependence structure over distributions. The quantile-based analysis can provide a flexible yet tractable alternative to the existing literature in modeling dependence (see, e.g., Koenker and Xiao, 2006). In this paper, we propose a functional quantile autoregression (FQAR) model where the future conditional quantile function is dependent on the past conditional quantile functions. In particular, we seek to relax the restriction of focusing only on a few distributional characteristics of the past distribution, and consider functional quantile autoregression models where information from the entire previous distribution can potentially affect the current distribution of the variable of interest.

The FQAR model is similar to the GARCH models in its structure. In GARCH models, the conditional variance of the current distribution is assumed to be related to the past conditional variance (autoregressive) and past observations or other covariates. In the FQAR model, we do not restrict the dynamic relationship to the second moment but allow the current conditional quantile function $q_t(\tau)$ to be related to the past conditional quantile functions $q_{t-j}(\tau)$ in a functional autoregressive form, and the past observations or other covariates.

The proposed FQAR model is also similar to the general functional time series models that have attracted a large amount of research attention in recent years (see, e.g. Bosq, 2000; Hörmann and Kokoszka, 2012; Park and Qian, 2012; Liu et al., 2016; Chang et al., 2021, etc.). There is also a literature on functional quantile regression models (see, e.g. Cardot et al., 2005; Ferraty and Vieu, 2005; Chen and Müller, 2012; Müller and Stadtmüller, 2005; Kato, 2012, etc.). However, the existing functional time series or functional quantile regression models consider the relationship of functionally observed data (curve variables). These functional models require a richer data environment to estimate, say, the cross-sectional distributions of repeated panels, or, the intra-period distributions of a time series observed at high frequency. Comparing to the existing literature, instead of looking at curve time series observations, we consider traditional univariate time series yet allow for a flexible functional dependence structure on the underlying dynamic mechanism. In proposing the FQAR model, we complement the literature by capturing systematic influences of the entire past distributional information on the future distribution of the response.

Statistical properties of the FQAR model are studied. Particularly, under stationarity conditions, the FQAR process admits a convergent infinity summation representation, resulting in that our estimation approach does not depend on the initial quantile function asymptotically. We propose a sieve method to estimate the FQAR model. Consistency and the limiting distributions of the estimators are investigated. A Monte Carlo experiment is conducted to evaluate the finite sample performance of the proposed estimator. Finally, we apply the proposed FQAR model to study the dynamics in the S&P 500 index time series, and illustrate how the current conditional quantile (distribution) is affected by the previous conditional distributions in a functional manner. We show that the conditional distribution of $\{y_t\}$ are influenced by the whole distribution of $\{y_{t-l}\}$.

The rest of the paper is organized as follows. The model is proposed in Section 2, along with some discussions of its properties. Section 3 discusses the sieve estimation procedure. Asymptotic properties of the estimators are given in Section 4. Results of a Monte Carlo investigation are reported in Section 5, and a financial application is presented in Section 6.

Throughout the paper, we use the following notation. For a function h in a Hilbert space, $\|h\|$ denotes the induced norm $\|h\| = \sqrt{\langle h, h\rangle}$, where the inner product is specified in the context. For a square matrix A, $\|A\|_F$ is the Frobenius norm. $\mathbb{E}_T$ is the expectation with respect to the empirical measure, and $\mathbb{G}_T$ denotes $\sqrt{T}(\mathbb{E}_T - E)$. We also use $a_T \lesssim b_T$ to imply that the inequality $a_T \le C b_T$ holds for all T with a constant C independent of T.

2. The FQAR model

Let y_t be a stationary time series, and $\mathcal{F}_t$ be the information set at time t, which includes $\{y_s, s \le t\}$ and maybe some other potential covariates observable at time t. We denote the conditional distribution of y_t given past information $\mathcal{F}_{t-1}$ by $F_t(\cdot)$. In addition, we denote the corresponding conditional quantile function by

$$q_t(\tau) = Q_{y_t}(\tau|\mathcal{F}_{t-1}) = F_t^{-1}(\tau) = \inf\{y : F_t(y) \ge \tau\},\ \tau \in [0, 1].$$

Being the inverse of the conditional distribution function, the conditional quantile function $q_t(\tau)$ fully captures the relationship between y_t and the past information.

To study the distributional dynamics in the time series $\{y_t\}$, we propose in this paper a functional quantile autoregression (FQAR) model in which the (conditional) distribution of y_t is affected by the previous (conditional) distribution of y_{t-d} $(d = 1, \ldots)$, and potentially a covariate x_t. A general first order functional quantile autoregression model can be represented as:

$$q_t(\tau) = \mathcal{H} q_{t-1}(\tau) + g(\tau, x_t), \tag{1}$$

where $\mathcal{H}$ is an operator $L^2[0,1] \mapsto L^2[0,1]$ in the functional space $L^2[0,1] = \{\varphi(u) : \int_0^1 \varphi^2(u)du < \infty\}$ associated with inner product $\langle \varphi_1, \varphi_2 \rangle = \int_0^1 \varphi_1(u)\varphi_2(u)du$, and x_t is a vector of lagged y_t and possibly other covariates available at time t. The function $g(\tau, x_t)$ brings additional information in x_t, similar to the GARCH term in a GARCH model. Due to the nature of quantile functions, we consider the space $L^2[0,1]$ as the domain of the operator $\mathcal{H}$. It is known that $L^2[0,1]$ is a Hilbert space with the induced norm $\|\varphi\| = \sqrt{\langle \varphi, \varphi \rangle}$. Notice that, for any variable Z with $E[Z^2|X] < \infty$ almost surely, X being any variable, its conditional quantile function $q_z(\tau|X) \in L^2[0,1]$. In fact, $\int_0^1 q_z^2(\tau|X)d\tau = \int u^2 dF_{Z|X}(u) = E[Z^2|X] < \infty$ provided that the conditional distribution function $F_{Z|X}(u)$ is strictly increasing. Additional discussions on the conditional quantile functions can be found in de Castro et al. (2023).

In this paper we consider the first order FQAR in the following form:

$$\begin{aligned} y_t &\sim q_t(\cdot) \\ q_t(\tau) &= \int_0^1 H(\tau, u) q_{t-1}(u)du + g(\tau, x_t), \quad \tau \in [0,1], \end{aligned} \tag{2}$$

where $H(\tau, u)$ and $g(\tau, x)$ are unknown functions. Corresponding to the general form of FQAR model (1), model (2) assumes that the operator $\mathcal{H}$ is an integral transformation

$$\mathcal{H}\varphi(\cdot) = \int_0^1 H(\cdot, u)\varphi(u)du, \tag{3}$$

with a kernel function $H(\cdot,\cdot) \in L^2(D) = \{h(u,v) : \iint_D h^2(u,v)dudv < \infty\}$, $D = [0,1]^2$, associated with inner product $\langle h_1, h_2 \rangle = \iint_D h_1(u,v)h_2(u,v)dudv$. We say that $\mathcal{H}$ is generated by the kernel $H(\cdot,\cdot)$. The coefficient function $H(\tau, u)$ acts as a weight function that relates current quantile with the entire past distribution; for a fixed τ, the current conditional quantile $q_t(\tau)$ is a weighted average (the integral) of the entire previous conditional quantile function $q_{t-1}(\cdot)$, with additional contribution from $g(\tau, x_t)$.

The proposed FQAR model constitutes a significant extension of traditional time series models. Many existing time series models, such as the ARMA processes, the QAR processes, and the linear GARCH processes, can be represented as special cases of the FQAR model. Here we demonstrate with a few examples.

Example 1. The linear GARCH model takes the form of

$$y_t = \sigma_t \varepsilon_t, \quad \sigma_t = \gamma_0 + \gamma_1 \sigma_{t-1} + \delta_1 |y_{t-1}|,$$

where $\{\varepsilon_t\}$ is a sequence of iid innovations with quantile function $q_\varepsilon(\tau)$. It follows that

$$q_t(\tau) = \gamma_1 q_{t-1}(\tau) + (\gamma_0 + \delta_1 |y_{t-1}|) q_\varepsilon(\tau).$$

It is a FQAR model (1) with $\mathcal{H} q_{t-1}(\tau) = \gamma_1 q_{t-1}(\tau)$ and $g(\tau, x_t) = (\gamma_0 + \delta_1 |y_{t-1}|) q_\varepsilon(\tau)$.

This is a simple model that allows the previous quantile $q_{t-1}(\tau)$ affecting $q_t(\tau)$ (but only at the same quantile). It can be rephrased as model (2), with $H(\tau, u)$ being specified as $\gamma_1 \delta(\tau - u)$ where $\delta(\cdot)$ is the Dirac delta function.

Example 2. The linear location-scale functional autoregressive model with conditional heteroskedasticity is in the form of $y_t = \mu_t + \sigma_t \varepsilon_t$ where

$$\begin{aligned} \mu_t &= c_1 + \int_0^1 \rho_1(u) q_{t-1}(u) du, \\ \sigma_t &= c_2 |y_{t-1}| + \int_0^1 \rho_2(u) q_{t-1}(u) du. \end{aligned}$$

Accordingly, it is a FQAR model with

$$H(u, \tau) = \rho_1(u) + \rho_2(u) q_\varepsilon(\tau) \text{ and } g(\tau, x_t) = c_1 + c_2 |y_{t-1}| q_\varepsilon(\tau).$$

Example 3. The FQAR model can also be viewed as an extension of a multi-quantile CaViaR model to the functional space. The CaViaR model of Engle and Manganelli (2004) takes the form of

$$y_t = q_t + \varepsilon_t \tag{4}$$

$$q_t = \beta_0 + \beta_1 q_{t-1} + \beta_2 g(x_t), \tag{5}$$

where ε_t follows a given distribution with its τth quantile being zero, hence q_t is the τth conditional quantile of y_t. The model has important applications in finance, but it focuses on only one particular τ. White et al. (2010) extended the original CaViaR model to accommodate multiple quantile levels and cross-quantile dependence. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$ be the selected quantile levels and $\boldsymbol{q}_t = (q_t(\tau_1), \ldots, q_t(\tau_m))$ where $q_t(\cdot)$ is the conditional quantile function of y_t. The vector version of (5) then becomes

$$\boldsymbol{q}_t = \boldsymbol{\beta}_0 + \boldsymbol{H} \boldsymbol{q}_{t-1} + \boldsymbol{g}(x_t), \tag{6}$$

where $\boldsymbol{H}$ is a $m \times m$ matrix and $\boldsymbol{g}(x_t)$ is a m-dimensional linear function corresponding to each component of $\boldsymbol{\tau}$. Naturally, our FQAR model (2) may be viewed as a further extension from multiple quantiles to the entire quantile function, wherein the coefficient matrix $\boldsymbol{H}$ in (6) corresponds to the kernel function $H(\tau, u)$ of the integration operator in (2), and the m-dimensional function $\boldsymbol{g}(x_t)$ corresponds to $g(\tau, x_t)$ in (2). The FQAR model (2) is fully non-parametric.

Remark 1. Various extensions with more complicated structures of FQAR processes can be proposed and studied. For simplicity and without loss of generality, we focus our discussion on model (2) in this paper.

Remark 2. The FQAR processes consider an autoregressive-type of relationship between the conditional quantile functions. The conditional quantile function q_t in (2) is completely specified given the observations $\{x_s, s = 0, \ldots, t\}$ and the initial quantile function $q_0(\cdot)$; and the likelihood function $f(y_1, \ldots, y_T \mid H, g, q_0(\cdot), x_1, \ldots, x_T)$ is completely specified. As we do not specify a parametric form of the unknown functions H and g, a nonparametric approach is needed for estimation.

2.1. Stationary conditions

The operator $\mathcal{H}$ in (3) is linear, that is, for any $\varphi, \psi \in L^2[0,1]$, $\mathcal{H}(\varphi + \psi) = \mathcal{H}\varphi + \mathcal{H}\psi$ and $\mathcal{H}(c\varphi) = c\mathcal{H}\varphi$ for all c. Notice that for any $\varphi \in L^2[0,1]$, $\|\mathcal{H}\varphi\| \leq \lceil H \rceil \|\varphi\|$, implying that $\|\mathcal{H}\| \leq \lceil H \rceil$, where $\|\mathcal{H}\|$ and $\lceil H \rceil$ are the norms of the mapping $\mathcal{H}$ and the function $H(u,v) \in L^2(D)$, respectively. Define the power of the operator for any integer $i \geq 1$ by $\mathcal{H}^i = \mathcal{H}(\mathcal{H}^{i-1})$, with $\mathcal{H}^0 = \mathcal{I}$ being the identical operator, i.e. $\mathcal{I}\varphi = \varphi$ for any $\varphi \in L^2[0,1]$. Moreover, if $\|\mathcal{H}\| < 1$, then $\mathcal{I} - \mathcal{H}$ is invertible and $(\mathcal{I} - \mathcal{H})^{-1} = \sum_{i=0}^{\infty} \mathcal{H}^i$. See Theorem 1.4 in Taylor and Lay (1980, p. 192). Denote $\|g(\tau, x)\|_\tau$ as the norm of $g(\cdot, x)$ in L^2 space for a fixed x.

We have the following theorem.

Theorem 1. *Assume that, in model (2), $\|\mathcal{H}\| < 1$ and $\sup_{j \geq 0} \|g(\tau, x_{t-j})\|_\tau < \infty$ in probability. In addition, assume $\sup_t \|q_t(\tau)\|_{L^2[0,1]} < \infty$ (or equivalently $\sup_t var(y_t \mid \mathcal{F}_{t-1}) < \infty$). We have*

(i) $q_t(\tau) = \sum_{j=0}^{t-1} \mathcal{H}^j g(\tau, x_{t-j}) + O_P(\|\mathcal{H}\|^t)$ *for large t;*

(ii) $q_t(\tau) = \sum_{j=0}^{\infty} \mathcal{H}^j g(\tau, x_{t-j})$ *in function norm sense.*

(iii) *Moreover, suppose $\lceil H \rceil < 1$, $\{x_t\}$ is stationary and $E|g(\tau, x_t)|^2 < \infty$, $\forall \tau \in [0,1]$. Then for each $\tau \in [0,1]$, $\{q_t(\tau)\}$, as a time series, is (weakly) stationary, with representation $q_t(\tau) = \sum_{j=0}^{\infty} \mathcal{H}^j g(\tau, x_{t-j})$ in which $\sum_{j=0}^{\infty} |\mathcal{H}^j g(\tau, x_{t-j})| < \infty$ with probability one.*

The proof is relegated to Appendix A.

Remark 3. The theorem shows several useful representations of the quantile functions under model (2). Here, the condition $\|\mathcal{H}\| < 1$ is necessary. It performs similarly to the autoregressive coefficient in an AR(1) model; it is also crucial for our estimation approach as it ensures that the estimation does not depend on the initial quantile function $q_0(\tau)$ asymptotically. The FQAR process shares some similar properties to the conventional AR process. Result (i) provides an approximation of $q_t(\tau)$ in terms of the underlying process and the integral operator. Result (ii) is a property parallel to causality for traditional time series, see Proposition 3.1.2 in Brockwell and Davis (1991, p88). Result (iii) shows pointwise evolution of the process $q_t(\tau)$ at each fixed τ, and how the past information is accumulated and contributed to the current distribution in an AR(∞) representation. The slightly stronger condition $\lceil H \rceil < 1$ in (iii) is needed for the convergence of $\sum_{j=0}^{\infty} |\mathcal{H}^j g(\tau, x_{t-j})| < \infty$ with probability one.

2.2. Monotonicity of the quantile function

By definition, the quantile functions should be monotone in τ. Although it is possible to impose some restrictions on various components of the model to achieve monotonicity, such sufficient conditions may not be practical. In practice, one popular way to achieve monotonicity is to transfer a potentially nonmonotone function into a monotone one, hence changing model (2) to

$$q_t(\tau) = \mathcal{L}\left(\int H(\tau, u) q_{t-1}(u) du + g(\tau, x_t) \right),$$

where $\mathcal{L}(\cdot)$ is a monotonizing operator.

There are many such operators, including the monotone rearrangement procedures of Dette et al. (2006) and Chernozhukov et al. (2009, 2010) and projection procedures of Daouia and Park (2013) and Westling et al. (2020). The readers are also referred to Barlow et al. (1972) and Kato (2012) for related discussions and references. In this paper, we adapt the approach to first obtain the unrestricted estimators of H and g, then obtain their induced estimated quantile functions, and finally apply the monotone transformation when needed to obtain the final estimate of the quantile functions, for prediction and other purposes. Specifically, we consider the following transformation similar to that in Chernozhukov et al. (2009) and Kato (2012). Let $w(u)$ be a possibly non-monotone function for $u \in [0,1]$. Let $Y = w(U)$ where $U \sim Unif[0,1]$. The cdf of Y is

$$F(y) = \int_0^1 1\,(w(u) \leq y)\, du,$$

which is naturally monotone in y. Its corresponding quantile function is

$$w^*(u) = F^{-1}(u) = \inf\{y : F(y) \geq u\},$$

which is again naturally monotone in u. The monotonizing operator $\mathcal{L}(\cdot)$ is defined as $w(u) \to w^*(u)$. For more discussions on the monotone rearrangement, see, e.g., Chernozhukov et al. (2009) and Kato (2012).

2.3. An AR representation of FQAR

Let U_t be an i.i.d sequence, following $Unif[0,1]$. Assume $x_t = y_{t-1}$ in (2). Under appropriate assumptions, the data generating process of the time series y_t in (2) can be written as

$$y_t = q_t(U_t) = \int_0^1 H(U_t, u) q_{t-1}(u) du + g(U_t, y_{t-1}). \tag{7}$$

Let $g_1(U_t, y_{t-1}) = g(U_t, y_{t-1})$ and $g_j(U_t, y_{t-1}) = \int_0^1 H(U_t, u) g_{j-1}(u, y_{t-j}) du,\ j \geq 2$. Then, y_t can be represented by $y_t = \sum_{j=1}^{\infty} g_j(U_t, y_{t-j})$. Denote $g_j^*(y_{t-j}) = E(g_j(U_t, y_{t-j}) \mid \mathcal{F}_{t-1})$, then

$$y_t = \sum_{j=1}^{\infty} g_j^*(y_{t-j}) + v_t, \tag{8}$$

where $v_t = \sum_{j=1}^{\infty}(g_j(U_t, y_{t-j}) - g_j^*(y_{t-j}))$. It is clear that $E(v_t) = 0$ and $E(v_t v_{t-j}) = 0$ since U_t and U_{t-j} are independent. Hence $\{v_t\}$ is a sequence of mean zero uncorrelated random variables, though they are not independent and have a complicated dependence structure. The model structure in (8) is often referred to as nonlinear additive AR model (Chen and Tsay, 1993; Huang and Yang, 2004) but (8) is of infinity order and with a dependent noise process.

The structure can be seen more clearly if the function $g(\cdot)$ is linear in y_{t-1}, $g(\tau, y_{t-1}) = \gamma_0(\tau) + \gamma_1(\tau) y_{t-1}$, then

$$y_t = \alpha^* + \sum_{j=1}^{\infty} \gamma_j^* y_{t-j} + v_t,$$

where the constants α^* and γ_j^* depend on the functions $\gamma_0(\cdot)$ and $\gamma_1(\cdot)$. Thus, in this example, y_t can be re-written as a linear AR process with the coefficients α^* and γ_j^*, and a martingale difference sequence innovation term v_t. The complicated and valuable information on distributional dynamics in the FQAR process are thrown into the uncorrelated but dependent residual term v_t.

It is interesting to note that the autocovariance function of this stationary FQAR process is simply that of a stationary linear AR (infinity) process with fixed AR parameters. Thus, from the point of view of traditional time series analysis, these two processes are very similar. In particular, classical time series analysis based on autocorrelations, partial autocorrelations and others will tend to identify data generated from this FQAR process as a linear AR model with fixed coefficients.

3. A sieve estimator for the FQAR model

Given observations $\{y_t, 1 \leq t \leq T\}$, we need to estimate the unknown functions $H(\tau, u)$ and $g(\tau, x)$ in (2) and the conditional quantile functions $q_t(\cdot)$ at each t. Toward this, we propose an orthogonal series method.

3.1. The estimation procedure

Let $\{\varphi_i(\cdot), i \geq 0\}$ be a complete orthonormal sequence in $L^2[0,1]$, viz, $\int_0^1 \varphi_i(u)\varphi_j(u) du = \delta_{ij}$ the Kronecker delta, in particular $\varphi_0(\cdot) \equiv 1$. Accordingly, a basis of $L^2(D)$ can be formed by the tensor product of $\{\varphi_i(\cdot), i \geq 0\}$ with itself, where $D = [0,1]^2$. For $H(\tau, u) \in L^2(D)$, we then have

$$H(\tau, u) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{ij}\varphi_i(\tau)\varphi_j(u), \qquad c_{ij} = \iint_D H(\tau, u)\varphi_i(\tau)\varphi_j(u) d\tau du. \tag{9}$$

As will become more clear in our later discussions, the convergence rate of our estimator of $g(\tau, x)$ may differ in the cases of bounded and unbounded supports of x_t. For this reason, we consider two leading cases for the function space to which $g(\tau, x)$ belongs. Specifically, we assume that $g(\tau, x) \in L^2([0,1] \times [a,b], \pi(x))$ where the support of x_t is given by a bounded interval $[a,b]$ on which $\pi(x)$ is a density, or $g(\tau, x) \in L^2([0,1] \times \mathbb{R})$, implying that x_t possesses support $\mathbb{R}$. The orthogonal expansion of $g(\tau, x)$ given below in these two cases have different convergence rate. See Dong et al. (2021). In later discussions, if not necessary to differentiate them, we simply use $\mathcal{X}$ to denote the support of x_t.

Suppose that $\{p_j(\cdot), j \geq 0\}$ is an orthonormal complete system associated with Hilbert space $L^2([a,b], \pi(x))$ or $L^2(\mathbb{R})$, $\langle p_i, p_j \rangle = \delta_{ij}$ where $\langle p_i, p_j \rangle = \int_a^b p_i(x)p_j(x)\pi(x)dx$ or $\int_{\mathbb{R}} p_i(x)p_j(x)dx$. As a result, the tensor product $\{\varphi_i(\cdot)\} \otimes \{p_j(\cdot)\}$ becomes an orthonormal basis in $L^2([0,1] \times [a,b], \pi(x))$ or $L^2([0,1] \times \mathbb{R})$. Thus, we have the orthogonal expansion

$$g(\tau, x) = \sum_{j=0}^{\infty} b_j(\tau) p_j(x) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} b_{ij}\varphi_i(\tau) p_j(x), \tag{10}$$

where $b_j(\tau) = \langle g(\tau, \cdot), p_j(\cdot) \rangle$ and $b_{ij} = \int_0^1 \varphi_i(\tau) b_j(\tau) d\tau,\ i, j \geq 0$.

We may write these in matrix form for conciseness. Denoting $\Phi(u) = (\varphi_0(u), \varphi_1(u), \ldots)'$ and $\Psi(x) = (p_0(x), p_1(x), p_2(x), \ldots)'$, then $H(\tau, u)$ and $g(\tau, x)$ can be expressed as

$$H(\tau, u) = \Phi(\tau)' C \Phi(u); \tag{11}$$

$$g(u, x) = \Phi(u)' B \Psi(x), \tag{12}$$

where $C = (C_{ij})$ and $B = (B_{ij})$ are infinite dimensional matrices with $C_{ij} = c_{i-1,j-1}$ and $B_{ij} = b_{i-1,j-1}$. As a result, the estimation of the unknown functions H and g is converted to the estimation of the matrices B and C.

In terms of the orthogonal series expansions (11) and (12), Eq. (2) can be rephrased as

$$q_t(\tau) = \Phi(\tau)' C \int_0^1 \Phi(u) q_{t-1}(u) du + \Phi(\tau)' B \Psi(x_t), \tag{13}$$

for all $t \geq 1$. By iteration and using the property of orthogonality, we have

$$\boldsymbol{\alpha}_{t-1} := C\boldsymbol{\alpha}_{t-2} + B\Psi(x_{t-1}), \quad \text{where } \boldsymbol{\alpha}_s = \int_0^1 \Phi(u) q_s(u) du,$$

for $t = 1, \ldots, T$. Hence, from (13) we have

$$q_t(\tau) = \Phi(\tau)' C^t \boldsymbol{\alpha}_0 + \Phi(\tau)' C^{t-1} B \Psi(x_1) + \cdots + \Phi(\tau)' C B \Psi(x_{t-1}) + \Phi(\tau)' B \Psi(x_t). \tag{14}$$

Following the sieve estimation procedure, we truncate the matrices C and B, along with Ψ and Φ, to k dimensions as an approximation. This is reasonable because the elements in the infinite dimensional matrices C, B and vector $\boldsymbol{\alpha}_0$ attenuate to zero as their indices tend to infinity, due to the Parseval equality in Hilbert space, i.e., $\|C\| = \|H\|_{L^2}$, $\|B\|^2 = \|g\|^2_{L^2}$, $\|\boldsymbol{\alpha}_0\| = \|q_0\|_{L^2}$. In addition, under the assumption $\|C\| < 1$, $C^i \approx 0$ for large i. Thus, we may ignore the higher order terms of C^i in (14) for all $i > t_0$ where t_0 is pre-determined and large. Then for $t > t_0$, the term $\Phi(\tau)' C^t \boldsymbol{\alpha}_0$ in (14) disappears, and the initial quantile $q_0(\cdot)$ contained in $\boldsymbol{\alpha}_0$ is eradicated. Hence, only using the observations $y_{t_0+1}, \ldots, y_T$, our estimation methodology is free of the initial quantile function.

Thus, given the truncation parameter k, $q_t(\tau)$ in (14) is approximated by

$$q_t^{(k)}(\tau) = \Phi_k(\tau)' C_k^{t_0} B_k \Psi_k(x_{t-t_0}) + \cdots + \Phi_k(\tau)' C_k B_k \Psi_k(x_{t-1}) + \Phi_k(\tau)' B_k \Psi_k(x_t), \tag{15}$$

for $t = t_0 + 1, \ldots, T$, where C_k and B_k are the left-top $k \times k$ sub-matrices of C and B, respectively, and $\Phi_k(\cdot)$ and $\Psi_k(\cdot)$ are the top k-dimensional sub-vectors of $\Phi(\cdot)$ and $\Psi(\cdot)$, respectively.

Since $q_t(\tau)$ minimizes $E[\rho_\tau(y_t - a)|\mathcal{F}_{t-1}]$ where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, we use the integrated loss function as our objective function for estimation,

$$(\hat{C}_k, \hat{B}_k) = \underset{C_k, B_k}{\operatorname{argmin}} \sum_{t=t_0+1}^{T} \int_0^1 \rho_\tau(y_t - q_t^{(k)}(\tau)) d\tau.$$

This can be done through a nonlinear optimization on C_k, B_k for a fixed k. Alternatively, if we ignore the algebraic relationship for the products $C_k^i B_k$ and treat them all as individual parameters, the model becomes linear-in-parameter and can be computed based on linear programming. Ignoring these algebraic restrictions may result in less efficient sampling performance, but it serves as a simple and useful procedure that facilitates the derivation of limit theory as well as faster computation in practice. Of course, additional estimation steps based on such linear estimators can be further considered for the purpose to achieve improved finite sample performance. See for example, Section 4.3 below for a brief discussion in this direction. The price we pay here is that we now have $K = (t_0 + 1)k^2 + k$ instead of $2k^2 + k$ parameters. We note that, when $p_0(x)$ is a constant, each term $\Phi_k(\tau)' C_k^i B_k \Psi_k(x_{t-1})$ in (15) consists of one constant term (multiplied by $\Phi_k(\tau)$). When we treat all $C_k^i B_k$ as individual parameters, the first column of each $C_k^i B_k$ are not identified. Hence we set the first columns of $C_k^i B_k$ to zero for all $i \geq 1$, leaving total $K = (t_0 + 1)k^2 - t_0 k$ number of parameters. For simplicity of presentation, we will continue to use the notation of $C_k^i B_k$ with the understanding that its first column are zero when $p_0(x)$ is a constant. This point will become clearer in our discussion on identification issue below.

For better exposition, denote $\beta_K = \text{vec}(C_k^{t_0} B_k, \ldots, C_k B_k, B_k)$ the K-dimensional vector consisting of all coefficients to be estimated, and rewrite (15) as

$$q_t^{(k)}(\tau) = \beta_K' Z_{k,\tau}(X_t), \tag{16}$$

for $t = t_0 + 1, \ldots, T$, with the corresponding K-dimensional vectors:

$$Z_{k,\tau}(X_t) = (\Psi_k(x_{t-t_0})', \ldots, \Psi_k(x_t)')' \otimes \Phi_k(\tau),$$

which depends on $X_t = (x_t, x_{t-1}, \ldots, x_{t-t_0})$. As a result, we define the coefficient vector β_K in the approximation quantile $q_t^{(k)}(\tau)$ as

$$\beta_K^Q = \underset{\beta}{\operatorname{argmin}}\ E \int_0^1 \rho_\tau(y - \beta' Z_{k,\tau}(X)) d\tau. \tag{17}$$

Observe that β_K defined by (16) is obtained via the best L^2 approximation of the underlying true H and g functions in the truncated space, while β_K^Q defined by (17) is derived through the best approximation of the underlying true population relationship of y_t and its past information in model (2), measured through the integrated check function loss. However, these two approximations have certain relationship as given by Lemmas 1 and 2 in Belloni et al. (2019) and its supplementary material. In particular, the approximation error in L^2 sense is an upper bound of that in the check function sense under some conditions on the conditional density function of y_t. Therefore, from now on we shall use β_K^Q as the approximation coefficient but still use the notation β_K instead of β_K^Q for notation brevity.

We estimate β_K by

$$\hat{\beta}_K = \underset{\beta}{\operatorname{argmin}}\ L_T(\beta), \quad L_T(\beta) := \frac{1}{T-t_0}\sum_{t=t_0+1}^{T}\int_0^1 \rho_\tau(y_t - \beta' Z_{k,\tau}(X_t))d\tau, \tag{18}$$

and denote the resulting estimate by $\hat{\beta}_K = (\widehat{C_k^{t_0}B_k}, \ldots, \widehat{C_k B_k}, \hat{B}_k)$. Evidently, we can have $\hat{B}_k$ as estimator of B_k directly from $\hat{\beta}_K$. Let $\hat{D}_k := \widehat{C_k^{t_0}B_k} + \cdots + \widehat{C_k B_k} + \hat{B}_k$. If $\hat{D}_k$ is invertible and $C^{t_0+1} \approx 0$, we can use $\hat{C}_k = I - \hat{B}_k\hat{D}_k^{-1}$ as the estimator of C_k.[1]

Then the estimators of $H(u,v)$ and $g(u,x)$ are given by

$$\hat{H}(u,v) = \Phi_k(u)'\hat{C}_k\Phi_k(v), \qquad \hat{g}(u,x) = \Phi_k(u)'\hat{B}_k\Psi_k(x). \tag{19}$$

We shall study the asymptotic behaviors of these estimators in the next section.

3.2. Identification

Our model is in an additive nonparametric form, appropriate restrictions are needed to ensure identification. See, for example, Dong and Linton (2018) and the reference therein for related discussions on this issue in additive nonparametric literature.

Specifically, consider the case that $p_0(x)$ is a constant. Observe that, using Eq. (10), iterations of (2) yields

$$\begin{aligned}
q_t(\tau) &= \int_0^1\int_0^1 H(\tau,u)H(u,v)q_{t-2}(v)dvdu + \int_0^1 H(\tau,u)\left[b_0(u) + \sum_{j=1}^{\infty} b_j(u)p_j(y_{t-2})\right]du \\
&\quad + \sum_{j=1}^{\infty} b_j(\tau)p_j(y_{t-1}) + b_0(\tau) \\
&= \int_0^1\int_0^1 H(\tau,u)H(u,v)q_{t-2}(v)dvdu + \sum_{j=1}^{\infty}\left[\int_0^1 H(\tau,u)b_j(u)\right]du\ p_j(y_{t-2}) \\
&\quad + \sum_{j=1}^{\infty} b_j(\tau)p_j(y_{t-1}) + \int_0^1 H(\tau,u)b_0(u)du + b_0(\tau) \\
&= \cdots + \int_0^1 H(\tau,u)b_0(u)du + b_0(\tau) + \cdots \\
&= \cdots + b_0(\tau) + \mathcal{H}b_0(\tau) + \mathcal{H}^2 b_0(\tau) + \cdots
\end{aligned}$$

where $\mathcal{H}b_0(\tau) = \int H(\tau,u)b_0(u)du$ as in (3).

If we treat each $b_0^{(j)}(\tau) \equiv \mathcal{H}^j b_0(\tau)$ as individual functions (similar to treating $C^k B$ as individual parameters), the identification issue raises. Hence we impose the following condition.

Identification condition: For each $\tau \in [0,1]$,

$$\mathcal{H}b_0(u) = \int_0^1 H(\tau,u)b_0(u)du = 0, \quad \text{where} \quad b_0(u) = \int_{\mathcal{X}} g(u,x)\pi(x)dx. \tag{20}$$

Let $\mathbf{b}_0' = (b_{00}, b_{10}, \ldots)$. It is easy to see that (20) is equivalent to $C\,\mathbf{b}_0 = 0$. This form is exactly the same as setting the first columns of $C^i B$ to zero in the estimation procedure in Section 3.1.

When $p_0(x) \not\equiv c$, the first term in (9) is no longer $b_0(\tau)$, rather it is $b_0(\tau)p_0(x)$. Accordingly, we do not need any identification condition at all, since the variation of $p_0(x)$ with x can help us identify $b_0(\tau)$ in the iteration of model (2). Moreover, it is worth pointing out that when the functional space is $L^2(\mathcal{X})$ and $\mathcal{X}$ is an unbounded interval such as $\mathcal{X} = \mathbb{R}$ or $\mathcal{X} = \mathbb{R}^+$, then $p_0(x) \not\equiv c$ and hence no identification condition is needed.

[1] In most cases $\hat{D}_k$ is dominated by the last term $\hat{B}_k$ in the sum. Since B_k is invertible for most of well-behaved functions $g(\tau,x)$, $\hat{D}_k$ is likely invertible. However, when $\hat{D}_k$ is indeed not invertible, alternative estimator of C_k can be used. For example, let $A_0 = B_k$, $A_1 = C_k B_k$, …, $A_{t_0} = C_k^{t_0} B_k$. We have t_0 matrix equations: $(I - C_k)A_i = A_i - A_{i+1}$, $i = 0, \ldots, t_0 - 1$. If $A_0 = B_k$ has rank d : $k/\sqrt{t_0} \le d \le k$, then we can solve $I - C_k$ from the estimated matrices A_i. In addition, even if the above condition is not satisfied, numerically we may find out the estimates of C_k and B_k by nonlinear optimization targeting directly on them, as illustrated in the simulation section.

4. Asymptotic theory

4.1. Assumptions

Here we first provide some conditions and notations needed in the development for the theoretical properties of the estimators. Let $\alpha = (\alpha_1, \alpha_2)$ be a vector with nonnegative integers α_1 and α_2, and let $|\alpha| = \alpha_1 + \alpha_2$; define $\partial^\alpha = \partial^{\alpha_1}\partial^{\alpha_2}$ the partial derivative operator. Let $C(\Omega)$ be the set of all functions continuous on $\Omega \subset \mathbb{R}^2$ and define $C^m(\Omega) = \{f \in C(\bar{\Omega}) \,|\, \partial^\alpha f \in C(\bar{\Omega}) \text{ for all } |\alpha| \leq m\}$ where $\bar{\Omega}$ is the closure of Ω.

Assumption A.

(A.1) $H(u, v) \in C^m(D)$, $D = [0, 1]^2$, for some $m > 0$.
(A.2) If $\mathcal{X} = \mathbb{R}$, assume that $g(u, x) \in C^m([0, 1])$ for each fixed x and $x^{m-i} g_x^{(i)}(u, x)$ belongs to Hilbert space $L^2(\mathbb{R})$ for all $0 \leq i \leq m$ $(m > 1)$ and each fixed $u \in [0, 1]$. If $\mathcal{X} = [a, b]$ a bounded interval, assume that $g(u, x) \in C^m([0, 1] \times [a, b])$.
(A.3) The matrix C in (11) satisfies $\|C\|_F < 1$.

In (A.1) and (A.2) we impose smoothness on certain orders of the partial derivatives for $H(\cdot, \cdot)$ and $g(\cdot, \cdot)$. They ensure the orthogonal series expansions converge at a sufficiently fast rate so the bias due to truncation is of a smaller order. For notation convenience we use m to represent the differential order for the bivariate functions $H(u, v)$ and $g(u, x)$ though different orders for them may be needed. The concrete requirements for m are shown in Assumption C below. Notice that in (A.2) we allow the second argument of $g(u, x)$ to admit unbounded support for a broader range of applications. Apart from the identification condition, the main difference for functions with unbounded support versus bounded support in our case is that the convergence rate of the orthogonal expansion for the unbounded case is slower than that for the bounded case, in view of Appendix C of Dong et al. (2016) and Lemma B.1 in the supplement of this paper. Condition (A.3), $\lceil H \rceil = \|C\|_F < 1$, ensures the convergence of the representations stated in Theorem 1.

Assumption B.

(B.1) The conditional density $f_t(y|\mathcal{F}_{t-1})$ of y_t given $\mathcal{F}_{t-1}$ is bounded from above uniformly over y and t; meanwhile, $f_t(q_t(\tau)|\mathcal{F}_{t-1})$ is bounded away from zero uniformly over t; the derivative of $y \mapsto f_t(y|\mathcal{F}_{t-1})$ is continuous and bounded in absolute value from above uniformly over y and t.
(B.2) Let

$$\Sigma = E\left[\psi_t(\beta_K)\psi_t(\beta_K)'\right], \quad \text{with} \quad \psi_t(u) := \int_0^1 Z_{k,\tau}(X_t)(\mathbb{I}(y_t \leq u' Z_{k,\tau}(X_t)) - \tau) d\tau. \tag{21}$$

Suppose that the matrix Σ has maximum eigenvalue $\lambda_{\max} = O(K^\nu)$ and minimum eigenvalue $\lambda_{\min} = O(K^{-\nu})$ with $\nu > 0$ where K is the dimension of Σ. In addition, suppose that $\Omega = E[Z_{k,\tau}(X_t)Z_{k,\tau}(X_t)']$ has eigenvalues bounded above from infinity and below from zero uniformly in τ and T; $\sup_\tau E\|Z_{k,\tau}(X_t)\|^4 = O(K^2)$.
(B.3) The matrix $D_k := C_k^{t_0} B_k + \cdots + C_k B_k + B_k$ is invertible.

The conditions in Assumption (B.1) are standard in the literature (e.g. Belloni et al., 2019). Assumption (B.2) imposes different conditions on the eigenvalues of Σ and Ω. As shown by Proposition 2.1 in Belloni et al. (2015) and the discussion in Belloni et al. (2019, p.9), under mild conditions on the density of the covariate, the matrix Ω has bounded eigenvalues above from infinity and below from zero because of the use of orthogonal bases in the functional spaces. On the other hand, it is reasonable to allow the matrix Σ to have its maximum eigenvalue to diverge and its minimum eigenvalue to decay with certain rates similar to Lemma 8 in Chang et al. (2015, p 296) and Assumption C in Dong et al. (2021). We note that this assumption is much weaker than that typically used in most of the sieve literature where the bounded eigenvalue assumption is frequently used (see Ai and Chen, 2007; Chang et al., 2015; Belloni et al., 2015; Dong et al., 2016). The concrete requirements on ν will be given in Assumption C below. The order requirement of the fourth moment for the vector $Z_{k,\tau}(X_t)$ is fulfilled if each component has a bounded fourth moment. The assumption (B.3) ensures the inversibility of $\widehat{D}_k$ with high probability for the estimation of C_k.

Assumption C. Suppose that both k and t_0 diverge with T, and that

(C.1) $\sqrt{T} k^{-m/2} = o(1)$ for some $m > 3 + 2\nu$;
(C.2) $\log(T) K^2 / \sqrt{T} = o(1)$ where $K = (t_0 + 1)k^2 + k$;
(C.3) $K^\nu(\|C\|^{t_0} + t_0 k^{-(m-1)/2}) = o(1)$ and $t_0^{-\nu/2} k^{1-\nu} \to \infty$, where ν is given in Assumption B.

The condition (C.1) is an undersmoothing condition that controls the truncation residues in the derivation of the asymptotic normality below. Notice that the condition $m > 3 + 2\nu$ is due to the unbounded support assumption for $g(u, x)$ in x and the divergence order of the maximum eigenvalue proposed in Assumption B. When the support of $g(u, x)$ in x is bounded, the condition on m can actually be weakened to $m \geq 2 + \nu$. For notational simplicity, we use the stronger condition for both cases here. Similar undersmoothing conditions are often used, such as that in Corollary 3.1 of Chen and Christensen (2015, p.454) and Comment 4.3 of Belloni et al. (2015). Notice also that the conditions (C.2)–(C.3) are technical requirements used in the lemmas in Appendix B to derive the consistency of the estimators of the coefficients and in the derivation of the Bahadur representation. In particular, the condition $K^{\nu}\|C\|^{t_0} = o(1)$ is quite weak on t_0 since $\|C\|^{t_0}$ decays exponentially in t_0. Thus t_0 is allowed to increase relatively slowly for the condition to hold, hence one does not need to use a very large t_0 in practice. All these are comparable with the literature such as Assumption 7 in Horowitz and Lee (2005) and Theorems 1 and 2 in Belloni et al. (2019).

4.2. Asymptotic results

Theorem 2 (*Consistency*). *Suppose that the sequence $\{y_t\}$ is strictly stationary and the identification condition (20) holds. Under Assumptions A–C, we have*

$$\|\hat{\beta}_K - \beta_K\| \lesssim_P \sqrt{K/T},$$

as $T \to \infty$. Consequently, when $g(u,x) \in L^2([0,1]\times[a,b],\pi(x))$, we have

$$\|\hat{H}(u,v) - H(u,v)\| = O_P(\sqrt{k^2/T} + k^{-(2m-1)/2}),$$
$$\|\hat{g}(u,x) - g(u,x)\| = O_P(\sqrt{k^2/T} + k^{-(2m-1)/2}),$$

and for any $t > t_0$,

$$\hat{q}_t(\tau) - q_t(\tau) = O_P(\|C\|^{t_0+1} + t_0 k^{-(2m-1)/2} + t_0 k^3 T^{-1/2});$$

When $g(u,x) \in L^2([0,1]\times\mathbb{R})$, we have $\|\hat{g}(u,x) - g(u,x)\| = O_P(\sqrt{k^2/T} + k^{-(m-1)/2})$, $\hat{q}_t(\tau) - q_t(\tau) = O_P(\|C\|^{t_0+1} + t_0 k^{-(m-1)/2} + t_0 k^3 T^{-1/2})$, while $\|\hat{H}(u,v) - H(u,v)\|$ has the same order as in the first case.

The proof is given in Appendix A. The theorem shows the convergence rate of the estimator of the coefficients in orthogonal expansions in (16) and the convergence rates of the estimators for the nonparametric functions in norm sense. The later basically consists of two parts, the variance term ($\sqrt{K/T}$) and the bias term ($k^{-(2m-1)/2}$ for bounded support and $k^{-(m-1)/2}$ for unbounded support). When the bias and the variance terms have the same decay rate, the estimators attain their optimal rates.

To present the asymptotic distribution of the estimators, define

$$J = E\left[\int_0^1 f_t(q_t(\tau)|\mathcal{F}_{t-1}) Z_{k,\tau}(X_t) Z_{k,\tau}(X_t)' d\tau\right], \tag{22}$$

where $f_t(\cdot|\cdot)$ is given in Assumption B. In addition, let $\mathcal{P}_B$ be the $k^2 \times K$ selection matrix such that $\mathcal{P}_B \beta_K = \mathrm{vec}(B_k)$. That is, $\mathcal{P}_B$ draws the k^2 elements from the K-dimensional vector β_K to form $\mathrm{vec}(B_k)$. Hence $\mathcal{P}_B J^{-1}\Sigma J^{-1}\mathcal{P}_B'$ results in a $k^2 \times k^2$ matrix with elements selected from the $K \times K$ matrix $J^{-1}\Sigma J^{-1}$. Similarly, let $\mathcal{P}_D$ be the $k^2 \times K$ selection matrix such that $\mathcal{P}_D \beta_K = \mathrm{vec}(D_k)$.

Theorem 3 (*Normality*). *Suppose Assumptions A–C hold. Also, suppose $\{y_t\}$ is stationary and α-mixing with coefficients $\alpha(i)$ satisfying $\sum_{i=1}^{\infty} i^c [\alpha(i)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$. We have, as $T \to \infty$, for any $(u,x) \in [0,1]\times\mathcal{X}$,*

$$\frac{\sqrt{T}}{\sigma_{1k}}(\hat{g}(u,x) - g(u,x)) \to_D N(0,1), \tag{23}$$

and for any $(u,v) \in [0,1]^2$,

$$\frac{\sqrt{T}}{\sigma_{2k}}(\hat{H}(u,v) - H(u,v)) \to_D N(0,1), \tag{24}$$

where

$$\sigma_{1k}^2 = m_k(u,x)'\mathcal{P}_B J^{-1}\Sigma J^{-1}\mathcal{P}_B' m_k(u,x), \text{ and } \sigma_{2k}^2 = P_k(u,x)' J^{-1}\Sigma J^{-1} P_k(u,x),$$

in which $m_k(u,x) = \Phi_k(u) \otimes \Psi_k(x)$ is a k^2 dimensional vector, and

$$P_k(u,v) = \mathcal{P}_D'[\Phi_k(u)'(I - C_k) \otimes D_k^{-1}\Phi_k(v)] - \mathcal{P}_B'[\Phi_k(u)' \otimes D_k^{-1}\Phi_k(v)]$$

is a K dimensional vector. Here J and Σ are defined in (22).

Moreover, for $t > t_0$, as $T \to \infty$,

$$\frac{\sqrt{T}}{\sigma_{3kt}}(\hat{q}_t(\tau) - q_t(\tau)) \to_D N(0,1), \tag{25}$$

where $\sigma_{3kt}^2 = Z_{k,\tau}(X_t)' J^{-1}\Sigma J^{-1} Z_{k,\tau}(X_t)$.

The proof is relegated to Appendix A. We note that, unlike Theorem 2, the asymptotic normality results do not depend on whether the support $\mathcal{X}$ is bounded or unbounded. This is because the results are first established for unbounded support $\mathcal{X}$, but it also covers the bounded $\mathcal{X}$ case using the fact that the convergence of $\hat{g}(u,x)$ to $g(u,x)$ is faster than that on unbounded $\mathcal{X}$, as established in Theorem 2.

Remark 4. Note that the condition on the mixing coefficients enables us to use the small-big-block argument to derive asymptotic normality. The convergence rates in Theorem 3 are common when sieve method is used. In (23), the convergence rate of $\hat{g}(u,x)$ is $\sigma_{1k}/\sqrt{T}$. Here, given the conditions in Assumption B.1 and B.2, the matrix J^{-1} has maximum eigenvalue bounded above from infinite, and further due to the condition on the eigenvalues of the matrix Σ in Assumptions B and C, the quantity σ_{1k} depends on the parameter ν, so that it is between $c_1K^{-\nu/2}\|\Phi_k(u)\otimes\Psi_k(x)\|$ and $c_2K^{\nu/2}\|\Phi_k(u)\otimes\Psi_k(x)\|$ with some positive constants c_1 and c_2. Furthermore, the magnitude of the quantity $\|\Phi_k(u)\otimes\Psi_k(x)\|$ depends on whether the support of x is bounded or unbounded. In particular, when the support is bounded, $\|\Phi_k(u)\otimes\Psi_k(x)\| = O(k)$. The rate is the same as that in Theorem 2 when combining with the undersmoothing condition in Assumption C. See equations (2.10) and (2.11) in Dong et al. (2021) for the calculation of $\|\Phi_k(u)\otimes\Psi_k(x)\|$. When the eigenvalues of Σ is bounded, $\nu=0$, the convergence rate of $\hat{g}_k(u,x)$ is then proportional to $k/\sqrt{T}$, which is comparable with the literature such as Theorem 3.1 in Dong and Linton (2018). When $\nu\neq 0$, the convergence rate is affected by a power of K, namely, the rate is between $c_1kK^{-\nu/2}/\sqrt{T}$ and $c_2kK^{\nu/2}/\sqrt{T}$.

The convergence rate of (24) is $\sigma_{2k}/\sqrt{T}$, which is similar to that of (23), except $\mathcal{P}_B' m_k(u,x)$ is replaced by $P_k(u,v)$ in σ_{2k}^2. Interestingly, the rate of $\hat{q}_t(\tau)$ in (25) is $\sigma_{3kt}/\sqrt{T}$ where σ_{3kt} depends on $Z_{k,\tau}(X_t)$ hence it has the subscript t. Similar to σ_{1k}, by Assumption B, the order of magnitude of σ_{3kt} is between $c_1K^{-\nu/2}\|Z_{k,\tau}(X_t)\|$ and $c_2K^{\nu/2}\|Z_{k,\tau}(X_t)\|$ with some positive constants c_1 and c_2. In the special case where $\nu=0$ the convergence rate in (25) is proportional to $\|Z_{k,\tau}(X_t)\|/\sqrt{T} = O_P(\sqrt{K/T})$. The normality in (25) can be viewed as a self-normalized version of central limit theorem for $q_t(\tau)$.

In order to facilitate statistic inference, we need to construct consistent estimators for the matrices J and Σ. To this end, we follow the literature such as Powell (1984) and Belloni et al. (2019, p. 8) to define

$$\begin{aligned}\hat{\Sigma} &= \frac{1}{T-t_0}\sum_{t=t_0+1}^{T}\psi_t(\hat{\beta}_K)\psi_t(\hat{\beta}_K)',\\ \hat{J} &= \frac{1}{2h(T-t_0)}\sum_{t=t_0+1}^{T}\int_0^1 I(|y_t - Z_{k,\tau}(X_t)'\hat{\beta}_K| < h)Z_{k,\tau}(X_t)Z_{k,\tau}(X_t)'d\tau,\end{aligned} \tag{26}$$

where $h=h_T\to 0$ as $T\to\infty$ plays the role of bandwidth, $I(\cdot)$ is the indicator function and $\psi_t(\cdot)$ is defined in (21).

Corollary 1. *Under the conditions in Theorem 3, as $T\to\infty$ we have*

$$\|\hat{\Sigma}-\Sigma\| \to_P 0, \quad \textit{and} \quad \|\hat{J}-J\| \to_P 0. \tag{27}$$

The proof of the corollary is relegated to Appendix B. It is clear that with the availability of $\hat{\Sigma}$ and $\hat{J}$ the consistent estimates of σ_i^2, $i=1,2,3$, can be easily obtained. Hence, statistical inference of the unknown parameters can then be conducted based on these estimates. For example, we may test cross-quantile dependence based on $H(\tau,u)$. More specifically, we may consider the null hypothesis $\mathrm{H}_0 : H(\tau,u)=0$, corresponding to "there is no impact of uth quantile of past distribution on the τth quantile of current distribution", based on the statistic

$$\iota_T(\tau,u) = \frac{\sqrt{T}}{\hat{\sigma}_{2k}}\hat{H}_k(\tau,u),$$

which is asymptotically normal under the null hypothesis.

One important thing is that the estimators $\hat{H}(\tau,u)$ and $\hat{g}(\tau,u)$ in (19) may not satisfy the identification condition (20) when $p_0(x)$ is a constant in the finite support case, even though the constraint is built in for the unrestricted estimation of β_K. A simple alleviation is to calibrate the estimate $\hat{b}_0(u)$, because the estimators are asymptotically consistent. Note that $b_0(u)=\Phi(u)'\mathbf{b}_0$ where $\mathbf{b}_0$ is defined below (20), hence we define $\hat{b}_0(u)=\Phi_k(u)'\hat{\mathbf{b}}_{0k}$ where $\hat{\mathbf{b}}_{0k}$ is a vector of the k elements of the first column of $\hat{B}_k$. When $\hat{C}_k$ is of full rank, we simply let $\hat{\mathbf{b}}_{0k}^{\text{iden}}=0$, and then $\hat{b}_0^{\text{iden}}(u)=0$. And when $\hat{C}_k$ is not of full rank, let $\bar{C}$ be the sub-matrix from $\hat{C}_k$ with full row rank and whose rank is the same as $\hat{C}_k$. Define

$$\hat{\mathbf{b}}_{0k}^{\text{iden}} = [I-\bar{C}'(\bar{C}\bar{C}')^{-1}\bar{C}]\hat{\mathbf{b}}_{0k}, \tag{28}$$

and then define $\hat{b}_0^{\text{iden}}(u)=\Phi_k(u)'\hat{\mathbf{b}}_{0k}^{\text{iden}}$. Notice that in the second case, $\hat{\mathbf{b}}_{0k}^{\text{iden}}$ is the residual of the projection of $\hat{\mathbf{b}}_{0k}$ onto $\bar{C}$, hence $\bar{C}\hat{\mathbf{b}}_{0k}^{\text{iden}}=0$, from which we have $\hat{C}_k\hat{\mathbf{b}}_{0k}^{\text{iden}}=0$.

Corollary 2. *Under conditions in Theorems 2 and 3, $\|\hat{b}_0^{iden}(u)-b_0(u)\|\to 0$, and for any $u\in(0,1)$ and $x\in\mathcal{X}$, as $T\to\infty$,*

$$\frac{\sqrt{T}}{\sigma_{1k}}(\hat{g}^{iden}(u,x)-g(u,x)) \to_D N(0,1), \tag{29}$$

where $\hat{g}^{iden}(u,x)=\hat{b}_0^{iden}(u)+\Phi_k(u)'\hat{B}_{-1,k}\Psi_k(x)$, $\hat{B}_{-1,k}$ is the $\hat{B}_k$ deleting the first column, and σ_{1k} is the same as in Theorem 3.

4.3. Further discussions

The FQAR model is nonlinear. A regression estimation directly based on Eqs. (13) is highly nonlinear in the parameters since $q_{t-1}(\cdot)$ is latent and depends on the unknown parameters nonlinearly. Estimation of (C_k, B_k) based on such a nonlinear optimization is complicated both computationally and theoretically. For this reason, we consider in Section 3 a linear sieve estimator that ignores the algebraic operation for C and B, and treat all matrices $C^i B$ as different unknown matrices.

Based on the estimator proposed in Section 3, additional steps can be performed to take into consideration of the restrictions implied in the power functions, and further extensions can be constructed to obtain estimators with improved properties. In this subsection, we briefly discuss a two step sieve estimator based on the estimator in Section 3. This two-step procedure can be easily implemented in practice.

Let $\hat{q}_t(\cdot)$ be an estimator of $q_t(\cdot)$ based on the procedure proposed in Section 3, with the monotone transformation if needed. We consider the following second step regression of y_t on the given $\hat{q}_t(\cdot)$, $\Phi_k(\cdot)$, and $\Psi_k(x_t)$:

$$\hat{\theta} = \left(\check{C}_k, \check{B}_k\right) = \underset{C_k, B_k}{\operatorname{argmin}} \sum_t \int_0^1 \rho_\tau \left(y_t - \Phi_k(\tau)' C_k \int_0^1 \Phi_k(v) \hat{q}_{t-1}(v) dv - \Phi_k(\tau)' B_k \Psi_k(x_t) \right) d\tau.$$

The computation of this optimization step is fast through linear programming. This two-step estimator serves as an intermediate solution between the linear-in-parameter estimator and the completely nonlinear estimator. In principle, one could also consider a multi-step iteration.

The limiting distribution of the two-step estimator $\left(\check{C}_k, \check{B}_k\right)$ will be affected by the first step estimation. In particular, under Assumptions A–C, notice that

$$\hat{q}_{t-1}(v) = \hat{\beta}_K' Z_{k,v}(X_{t-1}) = \sum_{j=0}^{t_0} \Phi_k(v)' \hat{C}_k^j \hat{B}_k \Psi_k(x_{t-j-1}).$$

Denote $\theta = \operatorname{vec}\left(C_k, B_k\right) = [\operatorname{vec}\left(C_k\right)', \operatorname{vec}\left(B_k\right)']' = \left[\theta_I', \theta_{II}'\right]'$, and let

$$\xi_{k,t-1} = \int_0^1 \Phi_k(v) q_{t-1,k}(v) dv, \ \tilde{\xi}_{k,t-1} = \int_0^1 \Phi_k(v) \hat{q}_{t-1}(v) dv,$$

and

$$\check{Z}_{k,t-1} = \int_0^1 Z_{k,v}(X_{t-1}) \otimes \Phi_k(v) dv, \ W_{k,\tau}(X_t) = \left(\xi_{k,t-1}', \Psi_k(x_t)'\right)' \otimes \Phi_k(\tau).$$

We define

$$\Gamma_{10} = -\mathrm{E}\left[\int_0^1 f_{y_t}\left(q_t(\tau)\right) W_{k,\tau}(X_t) W_{k,\tau}(X_t)^\top d\tau\right],$$

and

$$\Gamma_{20} = -\mathrm{E}\int_0^1 \left[f_{y_t}\left(q_t(\tau)\right)\right] W_{k,\tau}(X_t) \left[\theta_I' \left(\check{Z}_{k,t-1}^\top \otimes \Phi_k(\tau)\right)\right] d\tau$$

where $\theta_I = \operatorname{vec}(C_k)$. Then the two-step estimators has the following asymptotic expansion

$$\begin{aligned} &\sqrt{T}\left(\hat{\theta} - \theta^*\right) \\ &= \Gamma_{10}^{-1} \frac{1}{\sqrt{T}} \sum_t \int_0^1 \left[\tau - 1\left(y_t \le q_t(\tau)\right)\right] W_{k,\tau}(X_t) d\tau + \Gamma_{10}^{-1} \Gamma_{20} J^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \psi_t(\beta_K) + o_p(1), \end{aligned}$$

where the second term, $\Gamma_{10}^{-1}\Gamma_{20}J^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\beta_K)$, captures the effect from the first stage estimation, where ψ and J are defined in (22) and (21).

5. Monte Carlo experiments

In this section, Monte Carlo experiments are conducted to examine the finite sample performance of the proposed FQAR estimators. The first part of the study shows the estimation performance of C_k and B_k as well as the kernel functions $H(u, v)$ and $g(u, x)$. The second part illustrates the performance of the estimated quantile functions, $\hat{q}_t(\tau)$, and the one-step-ahead and multi-step-ahead forecasting errors of the quantile functions.

We generate the data from the FQAR model (2) starting with $q_0(\tau) \equiv 0$ for $\tau \in [0, 1]$. The observed data was generated via $y_t = q_t(U)$ where U is a random number drawn from a uniform distribution on $[0, 1]$, and $x_t = y_{t-1}$ with the initial value $y_0 = 0$. We consider three designs of the kernel functions:

Experiment 1 (Multiplicative form):

Let $H(u, v) = h_1(u) h_2(v)$, $g(u, x) = g_1(u) g_2(x)$, where

$$\begin{cases} h_1(u) = u - \cos(u) - 1 \\ h_2(v) = \sin(v) \end{cases}, \quad \begin{cases} g_1(u) = 1 - e^{-u} \\ g_2(x) = \sqrt{x + 1} \end{cases}, \quad u, v \in [0, 1], \quad x \in [-1, 1]. \tag{30}$$

Table 1
Means and standard deviations (in brackets) of estimation errors (×10).

	Experiment 1			Experiment 2			Experiment 3		
$k\backslash T$	T = 500	T = 1000	T = 5000	T = 500	T = 1000	T = 5000	T = 500	T = 1000	T = 5000
Panel A: Estimation error of C									
k = 2	3.28(2.23)	2.06(1.16)	0.78(0.44)	4.10(3.49)	2.72(1.70)	1.48(0.62)	1.08(1.95)	0.80(1.32)	0.42(0.75)
k = 3	1.46(0.69)	0.94(0.37)	0.41(0.16)	1.31(0.59)	1.04(0.31)	0.72(0.11)	0.05(0.04)	0.04(0.03)	0.02(0.01)
k = 4	1.10(0.53)	0.73(0.36)	0.32(0.08)	0.85(0.34)	0.64(0.20)	0.45(0.05)	0.03(0.02)	0.02(0.01)	0.01(0.00)
k = 5	0.87(0.34)	0.56(0.19)	0.24(0.08)	0.67(0.22)	0.50(0.11)	0.40(0.04)	0.01(0.00)	0.01(0.00)	0.01(0.00)
Panel B: Estimation error of B									
k = 2	1.06(0.63)	0.67(0.30)	0.43(0.13)	2.96(0.79)	2.71(0.35)	2.57(0.12)	0.16(0.02)	0.16(0.01)	0.15(0.01)
k = 3	0.71(0.28)	0.46(0.17)	0.27(0.10)	1.16(0.13)	1.13(0.09)	1.08(0.04)	0.51(0.23)	0.53(0.18)	0.53(0.07)
k = 4	0.92(0.37)	0.62(0.26)	0.35(0.17)	0.43(0.12)	0.34(0.10)	0.28(0.05)	0.37(0.01)	0.40(0.06)	0.41(0.02)
k = 5	0.72(0.27)	0.50(0.19)	0.24(0.09)	0.41(0.15)	0.31(0.11)	0.20(0.06)	0.08(0.01)	0.07(0.01)	0.06(0.00)

Notes: We simulate the data and estimate the coefficient matrices with 100 repetitions. The performance of estimation is evaluated by the Frobenius norm over k, $E = \frac{1}{k}\|X_k - \hat{X}_k\|$, where X_k is the coefficient matrix, i.e. C_k or B_k here. The results presented in this table are scaled by 10.

The multiplicative form settings of $H(u,v)$ and $g(u,x)$ indicate the expanded coefficient matrices C and B are degenerate, i.e. $rank(C) = rank(B) = 1$, with

$$c_{ij} = \int_0^1 \int_0^1 h(u,v)\varphi_i(u)\varphi_j(v)dudv = \int_0^1 h_1(u)\varphi_i(u)du \int_0^1 h_2(v)\varphi_j(v)dv,$$

$$b_{ij} = \int_0^1 \int_{\mathcal{X}} g(u,x)\varphi_i(u)p_j(x)\pi(x)dudx = \int_0^1 g_1(u)\varphi_i(u)du \int_{\mathcal{X}} g_2(x)p_j(x)\pi(x)dx,$$

which means that any rows or columns of C and B have multiple relationship with each other. The above setting satisfies Assumption A with $\lceil H \rceil = \|C\| < 1$.

Experiment 2 (Convolution form):

Let $H(u,v) = h_1(u-v)h_2(v)$, $g(u,x) = g_1(u-x)g_2(x)$, where

$$\begin{cases} h_1(u-v) = (u-v)^3 \\ h_2(v) = e^v \end{cases}, \quad \begin{cases} g_1(u-x) = \sin(u-x) \\ g_2(x) = 1-x^2 \end{cases}, \quad u,v \in [0,1], \quad x \in [-1,1]. \tag{31}$$

The convolution form settings of $H(u,v)$ and $g(u,x)$ ensure the expanded coefficient matrices C and B are full rank so that they are invertable. The above settings also satisfy Assumption A and $\lceil H \rceil = \|C\| < 1$.

Experiment 3 (Truncated form):

We also consider a truncated form design of $H(u,v)$ and $g(u,x)$ as follows

$$H(u,v) = \Phi_k(u)C_k\Phi_k(v), \quad g(u,x) = \Phi_k(u)B_k\Psi(x), \quad u,v \in [0,1], \quad x \in [-a,a]. \tag{32}$$

where k is the truncation parameter, and $\Phi_k(u)$ and $\Psi_k(x)$ are the first k elements of $\Phi(u)$ and $\Psi(x)$. For $\Phi(u)$, we adopt a trigonometric sequence $\{\varphi_j(x), j \geq 0\}$:

$$\varphi_0(u) \equiv 1, \quad \varphi_j(u) = \sqrt{2}\cos(\pi jx), \quad j \geq 0. \tag{33}$$

For $\Psi(x)$, we use an extended Chebyshev polynomial sequence $\{p_j(x), j \geq 0\}$ with reference distribution $\pi(x) = \frac{2}{\pi\sqrt{1-x^2}}$:

$$p_0(x) \equiv 1/\sqrt{2}, \quad p_j(x) = \cos(j\arccos(x/a)), \quad j \geq 0. \tag{34}$$

It is known that $\{\varphi_j(x), j \geq 0\}$ is a complete orthogonal sequence in $L^2[0,1]$ and $\{p_j(x), j \geq 0\}$ is an orthogonal system associated with Hilbert space $L^2([-a,a],\pi(x))$. Here we set $a = 0.15$, $k = 5$ and C_k and B_k be the estimated values from the empirical study on S&P 500 index shown in Section 6.

Typical realizations of the simulated series y_t under each experimental setting are shown in Figure C.1 in Appendix C. Some (marginal) summary statistics are shown in Table C.1. The series generated under three experiment designs are all stationary according to the ADF tests. The data produced in the first two experiments exhibit higher mean and volatility and display significant autocorrelations, while that in the third experiment is more leptokurtic and shows weak autocorrelations.

We then use the proposed seive method to estimate $H(u,v)$ and $g(u,x)$. For a given truncation parameter k, we use the estimation procedure proposed in Section 4.3 to estimate $\hat{C}_k$ and $\hat{B}_k$, and consequently obtain the estimates of the kernel function $H(u,v)$ and the intercept function $g(u,x)$. In the experiments, we use $t_0 = 5$. Figure C.2 and Figure C.3 in Appendix C respectively show the estimated kernel functions $H(u,v)$ and intercept functions $g(u,x)$ with $k = 5$, compared with the true ones, using the three realizations shown in Figure C.1 with sample size $T = 500$. We can observe that the shape of the functions are quite close to the true ones for both $H(u,v)$ and $g(u,x)$, while the relative error appears to be larger for $H(u,v)$ compared with that for $g(u,v)$.

To illustrate the estimation performance and the impacts of using different k and T, we estimate $\hat{C}_k$ and $\hat{B}_k$ under different settings of tuning parameters with 100 repetitions and report the summary statistics of the estimation errors in Table 1. From the table, it is evident that the estimation errors of C_k become smaller as k and T increase, indicating the convergence progress discussed in Section 4. The errors in Experiment 3 decrease much more noticeably as k increases, which could be attribute to the finite-dimensional true function designs with $k = 5$ as illustrated in (32). As for the estimation errors of B_k, the overall trend is similar to those of C_k for different values of k and T. While there are some discrepancies in Experiment 3, it is worth noting that the results of $k = 5$ and $T = 500$ also yield considerably smaller errors for B_k.

Next, we focus on the estimation and forecast performance of the quantile functions. We first discuss the estimation procedure and then formulate the one-step-ahead and multi-step-ahead forecasting procedures, assuming $x_t = y_{t-1}$. Once we have obtained $\widehat{C}_k$ and $\widehat{B}_k$, the kernel function and intercept function can be estimated as $\widehat{H}(u, v) = \Phi_k(u)'\widehat{C}_k\Phi(u)$ and $\widehat{g}(u, x) = \Phi(u)'\widehat{C}_k\Psi(x)$, and then following (15), the quantile function can be estimated as:

$$\widehat{q}_t(\tau) = \Phi_k(\tau)'\widehat{C}_k^{t_0}\widehat{B}_k\Psi_k(y_{t-t_0-1}) + \cdots \Phi_k(\tau)'\widehat{C}_k\widehat{B}_k\Psi_k(y_{t-2}) + \Phi_k(\tau)'\widehat{B}_k\Psi_k(y_{t-1}), \tag{35}$$

$t = t_0 + 1, \ldots, T$.

The one-step ahead forecast and multi-step ahead forecast of the quantile functions at time t are based on the available information set $\mathcal{F}_t$. Conditional on $\mathcal{F}_t$, we write the estimated matrices at time t as $\widehat{C}_k|_t$ and $\widehat{B}_k|_t$, and the estimated kernel function and intercept function as $\widehat{H}(u, v)|_t$ and $\widehat{g}(u, x)|_t$. Then, the one-step-ahead forecast of the quantile function is

$$\begin{aligned}\widehat{q}_{t(+1)}(\tau) &= \int_0^1 \widehat{H}(\tau, u)|_t q_t(u)du + \widehat{g}(\tau, y_t)|_t \\ &= \Phi_k(\tau)'\widehat{C}_k^{t_0}|_t\widehat{B}_k|_t\Psi_k(y_{t-t_0}) + \cdots \Phi_k(\tau)'\widehat{C}_k|_t\widehat{B}_k|_t\Psi_k(y_{t-1}) + \Phi_k(\tau)'\widehat{B}_k|_t\Psi_k(y_t).\end{aligned} \tag{36}$$

The multi-step-ahead forecast of the quantile function requires the input of future data (e.g. the unobserved $x_{t+2} = y_{t+1}$). Here we use the median under the predicted quantile function as the predicted value (e.g. $\hat{y}_{t(+1)} = \hat{q}_{t(+1)}(0.5)$). Specifically, for $h = 1, 2, \ldots$, recursively we use

$$\widehat{q}_{t(+h)}(\tau) = \Phi_k(\tau)'\widehat{C}_k^{t_0}|_t\widehat{B}_k|_t\Psi_k(\widehat{y}_{t(+h-t_0-1)}) + \cdots \Phi_k(\tau)'\widehat{C}_k|_t\widehat{B}_k|_t\Psi_k(\widehat{y}_{t(+h-2)}) + \Phi_k(\tau)'\widehat{B}_k|_t\Psi_k(\widehat{y}_{t(+h-1)}), \tag{37}$$

where $\widehat{y}_{t(+h-i)} = y_{t+h-i}$ if $h \le i$ and $\widehat{y}_{t(+h-i)} = \widehat{q}_{t(+h-i)}(0.5)$ if $h > i$.

We first conduct the one-step-ahead rolling forecast of $q_t(\tau)$ following (36) via rolling estimation of C_k and B_k. Figure C.4 shows the dynamics of the true and predicted quantile functions for the case when $k = 5$. Despite the different types of quantile functions generated by the three experiments, the results demonstrate that the predicted quantile functions closely resemble the true ones across all three experiments. We also observe that almost all the predicted quantile functions display monotonicity. Thus, the rearrangement procedure has very little impact on the results.

On the evaluation of forecasting errors, we adopt the mean squared (prediction) error (MSE) and symmetric mean absolute percent error (SMAPE) for evaluating the performance of the one-step-ahead and multi-step-ahead forecasting results of the quantile functions:

$$MSE(h) = \frac{1}{T - h - T_0 + 1}\sum_{t=T_0}^{T-h}\int_0^1 \left(\widehat{q}_{t(+h)}(\tau) - q_{t+h}(\tau)\right)^2 d\tau, \tag{38}$$

$$SMAPE(h) = \frac{1}{T - h - T_0 + 1}\sum_{t=T_0}^{T-h}\int_0^1 \frac{\left|\widehat{q}_{t(+h)}(\tau) - q_{t+h}(\tau)\right|}{\left|\widehat{q}_{t(+h)}(\tau)\right| + \left|q_{t+h}(\tau)\right|} d\tau, \tag{39}$$

where $[T_0 + 1, T]$ is the out-of-sample interval. Here we fix the length $T - T_0 = 100$. We use SMAPE instead of MAPE (which only has $|q_{t+h}(\tau)|$ in the denominator) to avoid large distortions when $|q_t(\tau)|$ is close to 0 for some τ. SMAPE takes value between 0 to 100%.

Table 2 reports one-step-ahead forecast errors of $\widehat{q}_{t(+1)}(\tau)$ with different settings of k and T. It is seen that, as the sample size T and the truncation order k increase, both the MSE and SMAPE of the predictions become smaller. Note that the results of the MSE in Experiment 3 have been scaled by 10^5, and hence they are much smaller than those of the other two experiments. This is due to the small range of the observed data generated in Experiment 3, as shown in Table C.1. The SMAPE of Experiment 3 are larger than the other two experiments. As k increases, the prediction errors in Experiment 3 decrease significantly, possibly due to the finite-dimensional nature of the true functions in design (32).

To further investigate the forecasting errors of $\widehat{q}_t(\tau)$ at different τ in $[0, 1]$, we compute the MSE and SMAPE for each individual τ without averaging over τ as in (38) and (39). Figs. 1 and 2 show the forecasting errors of $\widehat{q}_t(\tau)$ with different τ varying from 0 to 1 by MSE and SPAPE, respectively. In general, it can be observed that the errors decrease as the increase of k and T, both for MSE and SMAPE. The effect of k is more pronounced for Experiment 3, as seen in Table 2. Since the data in all three experiments exhibit negative skewness as shown in Table C.1, larger MSE values are observed in the left tails, particularly in the first two experiments. On the other hand, a peak in SMAPE can be seen around $\tau = 0.3$ to $\tau = 0.5$ in Experiment 1 and Experiment 3. This is mainly due to the fact that the true and predicted quantiles at these τ values are close to zero, resulting in larger SMAPE values as defined by (39).

Finally, we compare the multi-step-ahead forecast results of the FQAR model with the GARCH model (with constant mean). For quantile forecast via the GARCH model, we assume conditional normality of the innovation term and compute the quantile function as

$$\hat{q}_{t(+h)}(\tau) = \hat{\mu} + \hat{\sigma}_{t(+h)}\Phi^{-1}(\tau), \quad \tau \in [0, 1] \tag{40}$$

Table 2
Mean and standard deviations (in brackets) of one-step-ahead forecast errors.

	Experiment 1			Experiment 2			Experiment 3		
$k \backslash T$	$T = 500$	$T = 1000$	$T = 5000$	$T = 500$	$T = 1000$	$T = 5000$	$T = 500$	$T = 1000$	$T = 5000$
Panel A: Mean Squared Error (MSE)									
$k = 2$	1.51(5.55)	0.46(0.39)	0.36(0.17)	3.13(5.21)	2.21(2.73)	2.02(2.87)	13.3(34.6)	5.59(13.8)	1.88(3.68)
$k = 3$	0.77(0.26)	0.25(0.23)	0.13(0.03)	0.99(1.47)	0.77(0.54)	0.67(0.44)	0.51(0.59)	0.44(0.35)	0.33(0.12)
$k = 4$	0.30(0.42)	0.25(0.73)	0.06(0.04)	0.35(0.38)	0.16(0.29)	0.08(0.04)	0.09(0.02)	0.04(0.02)	0.02(0.02)
$k = 5$	0.35(0.40)	0.23(0.39)	0.05(0.03)	0.26(0.29)	0.14(0.12)	0.06(0.04)	0.07(0.13)	0.04(0.09)	0.01(0.02)
Panel B: Symmetric Mean Absolute Percent Error (SMAPE)									
$k = 2$	2.41(1.14)	1.99(0.84)	1.73(0.62)	2.89(1.62)	2.81(1.32)	2.76(1.17)	3.94(1.93)	3.79(1.86)	3.13(1.22)
$k = 3$	2.04(1.05)	1.55(0.59)	1.25(0.20)	2.12(0.86)	2.08(0.86)	1.98(0.78)	2.71(0.24)	2.65(0.12)	2.62(0.09)
$k = 4$	1.57(0.89)	1.24(0.88)	0.79(0.33)	1.47(1.40)	1.08(0.70)	0.74(0.37)	1.56(0.69)	1.36(0.51)	0.93(0.33)
$k = 5$	1.73(0.86)	1.36(0.85)	0.91(0.36)	1.39(0.96)	1.15(0.74)	0.74(0.39)	1.53(0.76)	1.26(0.65)	0.73(0.37)

Notes: We simulate the data and conduct the rolling one-step-ahead forecast of the quantile functions with 100 out-of-sample results. The MSE results in the table are scaled by 100 for Experiments 1 and 2, and by 10^5 for Experiment 3. The SMAPE results are scaled by 10.

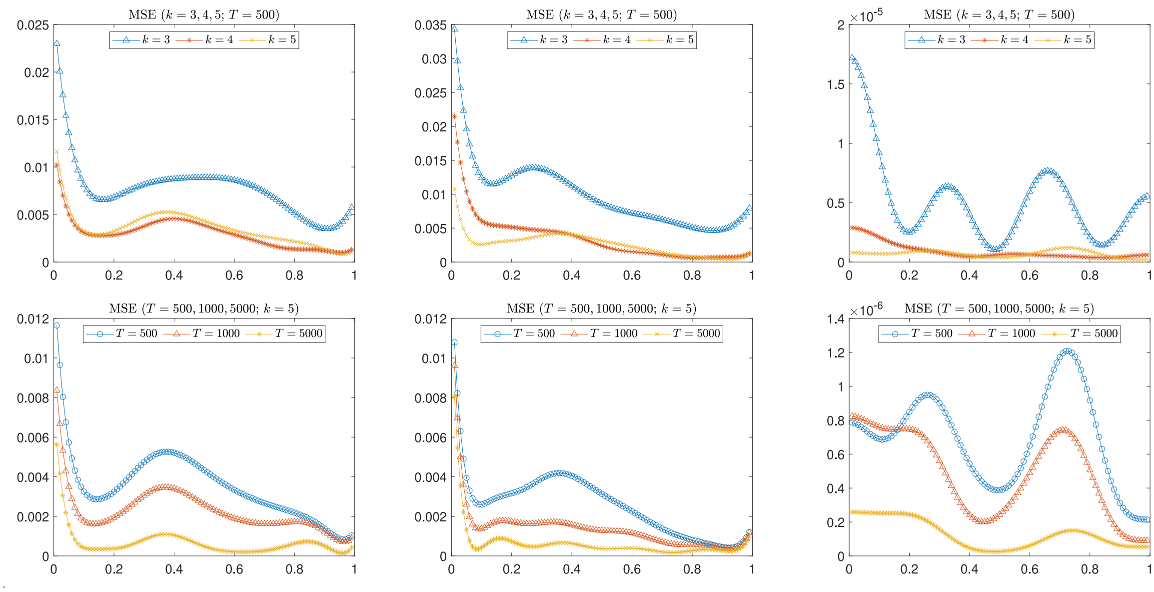

Fig. 1. MSE of $\hat{q}_t(\tau)$ with τ varying from 0 to 1 Left (Experiment 1); Middle (Experiment 2); Right (Experiment 3).

Table 3
Multi-step-ahead forecast errors.

Model	Experiment 1			Experiment 2			Experiment 3		
	$h = 1$	$h = 5$	$T = 20$	$h = 1$	$h = 5$	$h = 20$	$h = 1$	$h = 5$	$h = 20$
Panel A: Mean Squared Error (MSE)									
GARCH	0.82(0.64)	0.58(0.36)	0.61(0.37)	9.48(6.70)	9.15(6.09)	9.17(6.43)	0.23(0.19)	0.26(0.27)	0.42(0.77)
FQAR($k = 3$)	0.70(2.60)	0.74(1.11)	0.77(1.01)	0.99(1.47)	7.69(6.96)	8.15(5.61)	0.51(0.59)	0.56(0.70)	0.63(0.83)
FQAR($k = 5$)	0.35(0.42)	0.47(0.50)	0.49(0.50)	0.26(0.29)	9.72(12.3)	9.87(11.4)	0.07(0.13)	0.17(0.34)	0.34(0.75)
Panel B: Symmetric Mean Absolute Percent Error (SMAPE)									
GARCH	2.31(0.93)	2.02(0.81)	2.19(0.80)	5.06(1.85)	5.19(1.68)	5.40(1.84)	1.92(0.33)	1.94(0.34)	2.01(0.42)
FQAR($k = 3$)	2.04(1.05)	2.04(0.88)	2.11(1.02)	2.12(0.86)	4.69(1.89)	4.81(1.83)	2.71(0.24)	2.73(0.22)	2.75(0.25)
FQAR($k = 5$)	1.73(0.86)	1.80(0.92)	1.80(0.92)	1.39(0.96)	4.72(2.47)	5.29(2.30)	1.53(0.76)	1.77(0.79)	1.91(0.89)

Notes: We simulate the data and conduct the rolling h-step-ahead forecast of the quantile functions with 100 out-of-sample results. The MSE results in the table are scaled by 100 for Experiments 1 and 2, and by 10^5 for Experiment 3. The SMAPE results are scaled by 10.

where $\Phi^{-1}(\tau)$ is the inverse normal cumulative distribution function, $\hat{\mu}$ is estimated with the sample mean, and $\hat{\sigma}_{t(+h)}$ is the h-step-ahead forecast of conditional volatility under the estimated GARCH model.

Table 3 shows the comparison of the multi-step-ahead forecast errors of the quantile functions obtained via the GARCH and FQAR model. The advantage of the FQAR model compared with the GARCH model is significant for $h = 1$, both in mean and standard deviations of MSE and SMAPE, while the difference between the two models become smaller for longer forecasting horizon.

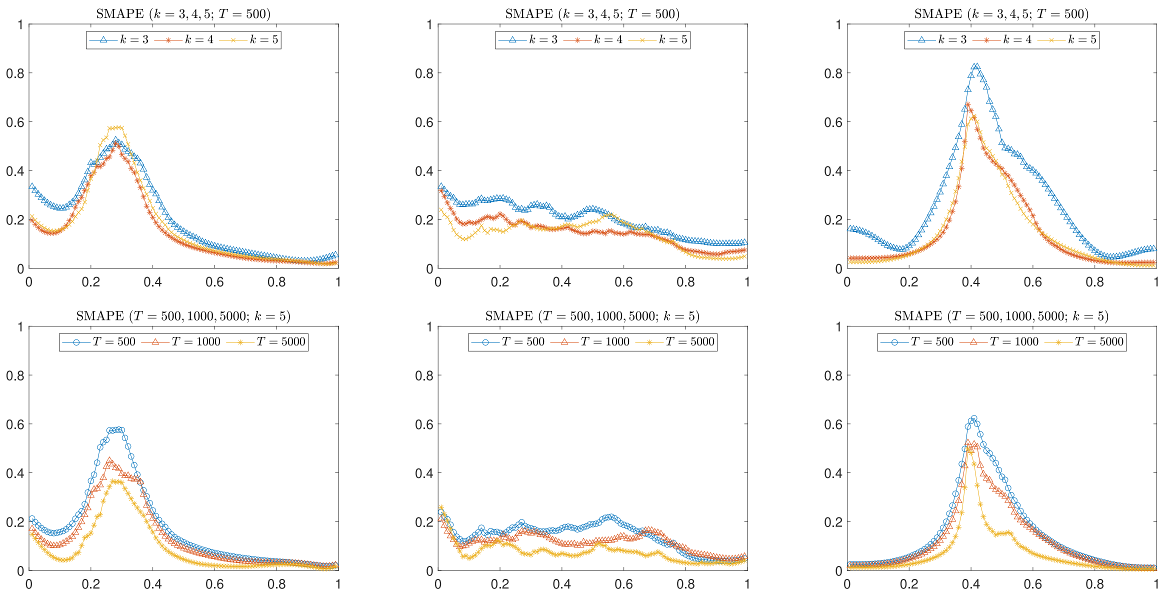

Fig. 2. SMAPE of $\hat{q}_t(\tau)$ with τ varying from 0 to 1 Left (Experiment 1); Middle (Experiment 2); Right (Experiment 3).

6. Empirical application

There is a large literature on empirical study of distributional dynamics in finance. Since the introduction of the Value-at-Risk (VaR), risk measures based on the left tail distributional information have become a key input for making financial decisions (Jorion, 2006). There is also an accumulated empirical literature arguing that information of the past distributions affect the future distribution, including the empirical studies on risk forecasts, spillover of risks, and cross-quantile microprudential dynamics — see, e.g. Beirne et al. (2010), De Nicolò. et al. (2014), Yahya et al. (2020), Muhammad et al. (2020), Pham (2021), and Dai et al. (2022).

In this section, we use the S&P 500 index data as an illustrative example to demonstrate how the distribution is affected by the previous distribution in a functional manner, and then we apply the proposed FQAR model and method to study the distributional dynamics and assess its empirical performance. Specifically we use the S&P 500 index data from July 2, 2012 to June 30, 2022 with sample size 2353. Let y_t be the log return (difference of the log price) at the day of $t : 2 \le t \le 2353$. Let $x_t = y_{t-1}$, $t = 3, \dots, 2353$, so that the FQAR model is built using data y_t, $t = 3, \dots, 2353$. Figure D.1 in Appendix D shows the log return series.

We first show how the quantiles of $\{y_t\}$ are influenced by the whole distribution of $\{y_{t-l}\}$ based on the cross-quantilogram method proposed by Han et al. (2016), where we have modified the framework to study the temporal dynamic pattern here. Consider a measure of dependence between two events $\{y_t \le q_t(\tau_1)\}$ and $\{y_{t-l} \le q_t(\tau_2)\}$ for an arbitrary pair (τ_1, τ_2) with $\tau_1, \tau_2 \in (0, 1)$ and an integer l denoting the lag length. The cross-quantilogram between two events $\{y_t \le q_t(\tau_1)\}$ and $\{y_{t-l} \le q_t(\tau_2)\}$ is defined by

$$\rho_{\tau_1,\tau_2}(l) = \frac{E\left[\psi_{\tau_1}(y_t - q_t(\tau_1))\psi_{\tau_2}(y_{t-l} - q_{t-l}(\tau_2))\right]}{\sqrt{E\left[\psi^2_{\tau_1}(y_t - q_t(\tau_1))\right]}\sqrt{E\left[\psi^2_{\tau_2}(y_{t-l} - q_{t-l}(\tau_2))\right]}} \tag{41}$$

where $\psi_\tau(u) = 1[u < 0] - \tau$, and $1(\cdot)$ is the indicator function. The cross-quantilogram captures the quantile dependence between y_t and y_{t-l}, hence the relationship between $q_t(\tau_1)$ and $q_{t-l}(\tau_2)$ for any pairs of τ_1 and τ_2 in $(0, 1)$.

Here we focus on the quantile dependence between $\{y_t\}$ and $\{y_{t-1}\}$. To estimate the sample cross-quantilogram, we employ a stationary bootstrap procedure following the methodology proposed by Han et al. (2016). Subsequently, the relationship between $\{y_t \le q_t(\tau_1)\}$ and $\{y_{t-1} \le q_t(\tau_2)\}$ is visualized through the heat map plot of the cross-quantilogram, as adopted in Yahya et al. (2020), Pham (2021), and Dai et al. (2022). Fig. 3 presents the cross-quantilogram heat maps from y_{t-1} to y_t of the S&P series under consideration. The y-axis represents the quantiles of y_t, and the x-axis represents the quantiles of y_{t-1}. Fig. 3 shows that $q_t(\tau)$ is not only influenced by $q_{t-1}(\tau)$ at the same quantile level τ, but also exhibits significantly influence across all the quantiles of y_{t-1}, especially in the tail region of y_t. This result strengthens the assumption made in the FQAR model (1) and (2) where we assume that the quantile of y_t is determined by the entire range of quantiles in y_{t-1}, which cannot be adequately captured by traditional models such as GARCH or other existing models.

Next, we present the estimation results and empirical performance of the FQAR model. Following the same approach as in the simulation study, we first estimate $\widehat{C}_k$ and $\widehat{B}_k$ and consequently obtain the estimates of kernel function $H(u, v)$ and the intercept function $g(u, x)$. We expand $H(u, v)$ and $g(u, x)$ based on the trigonometric sequence $\{\varphi_j(x), j \ge 0\}$ for $u, v \in [0, 1]$ and the extended Chebyshev polynomial sequence $\{p_j(x), j \ge 0\}$ for $x \in [-a, a]$ defined in (33) and (34). Here we choose $a = 0.15$ to cover the range

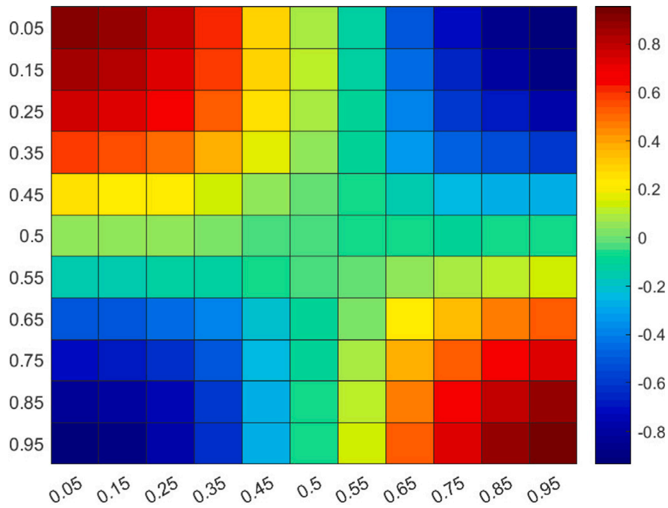

Fig. 3. Cross-quantilogram heat map between quantiles of y_t (y-axis) and y_{t-1} (x-axis).

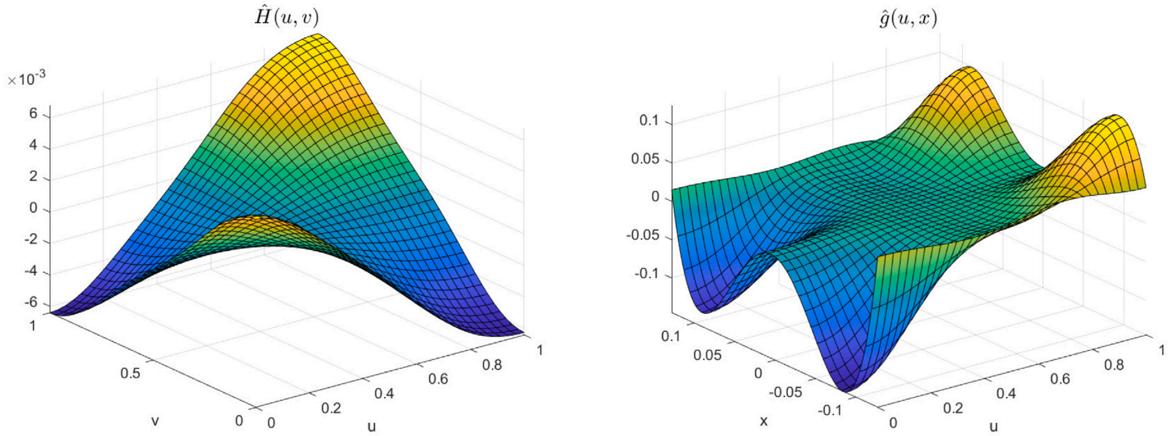

Fig. 4. Estimated kernel and intercept functions.

of data, $[-0.1277, 0.0897]$, instead of using $a = 1$. In practice a tight coverage range is preferred. Otherwise, the variation in x/a may be small, resulting in small variation in $p_j(x_t)$ and the degeneration of the design matrix in Eq. (16), causing identification issue in the estimation of $\widehat{C}_k$ and $\widehat{B}_k$.

The estimated H and g functions are shown in Fig. 4 with truncation parameter $k = 5$. For a fixed τ, $H(\tau, v)$ represents the impact of the past quantiles $q_{t-1}(v)$ on the current quantile $q_t(\tau)$, and $g(\tau, v)$ indicates how the recent observations $x_t = y_{t-1}$ affect the current quantile $q_t(\tau)$, according to model (2). The estimated results of H and g serve as the real input in the third simulation experiment in Section 5.

To investigate how $q_t(\tau)$ is affected by the entire range of past $q_{t-1}(v)$ for $v \in [0, 1]$ in the FQAR model, we plot the estimated kernel and intercept functions with selected $\tau = 0.05, 0.1, 0.5, 0.9, 0.95$ in Fig. 5. From the figure it is seen that $q_t(\tau)$ is not only determined by $q_{t-1}(\tau)$, but also the entire range of $q_{t-1}(v)$ for $v \in [0, 1]$, while the structure is different for different τ. In particular, the lower tail of the current distribution is positively correlated with the lower tail and negatively correlated with the upper tail of the past distribution, as observed in the results of $\tau = 0.05$ or 0.1. The upper tail of the current distribution exhibits an opposite result for $\tau = 0.05$ or 0.1. The median ($\tau = 0.5$) is relatively hard to predict as $\widehat{H}(u, v)$ is close to 0 for $\tau = 0.5$, while it also shows weakly dependence on $q_{t-1}(v)$ for relatively small or large values of v. These findings align with the results illustrated in Fig. 3, which further validate the importance of allowing the quantiles of $\{y_t\}$ to be influenced by the entire distribution of $\{y_{t-l}\}$, as proposed in model (1) and (2). In addition, we may also observe from the results of $\widehat{g}(u, x)$ that the tail quantiles, $q_t(u)$ for $u = 0.05, 0.1, 0.9, 0.95$, are more significantly influenced by the change of x_t (or y_{t-1}).

On the monotonicity of the estimated quantile functions, there are only two non-monotone estimated functions (out of 2352 of them) in this example, less than 0.1%. Figure D.2 shows the two violations. It is seen that the violation is very minor and local.

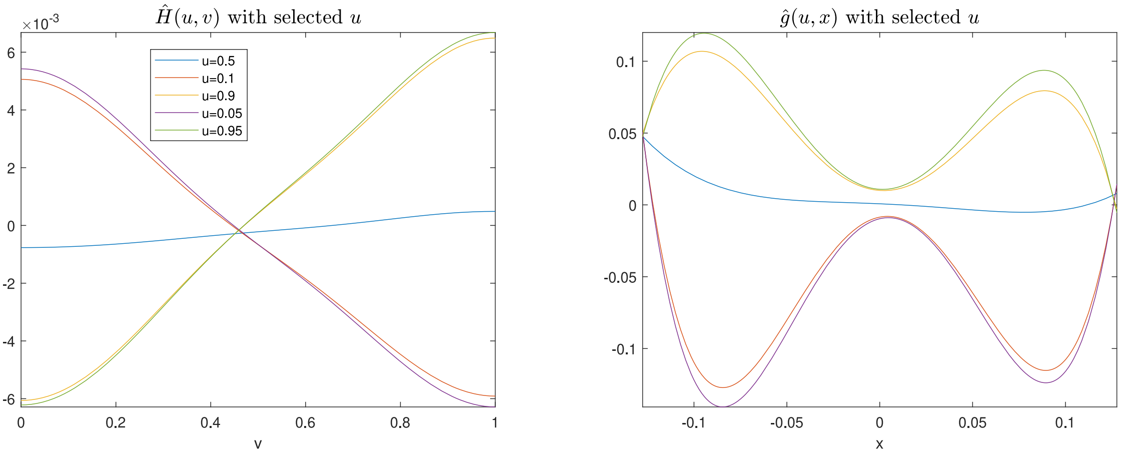

Fig. 5. Estimated kernel and intercept functions with selected u.

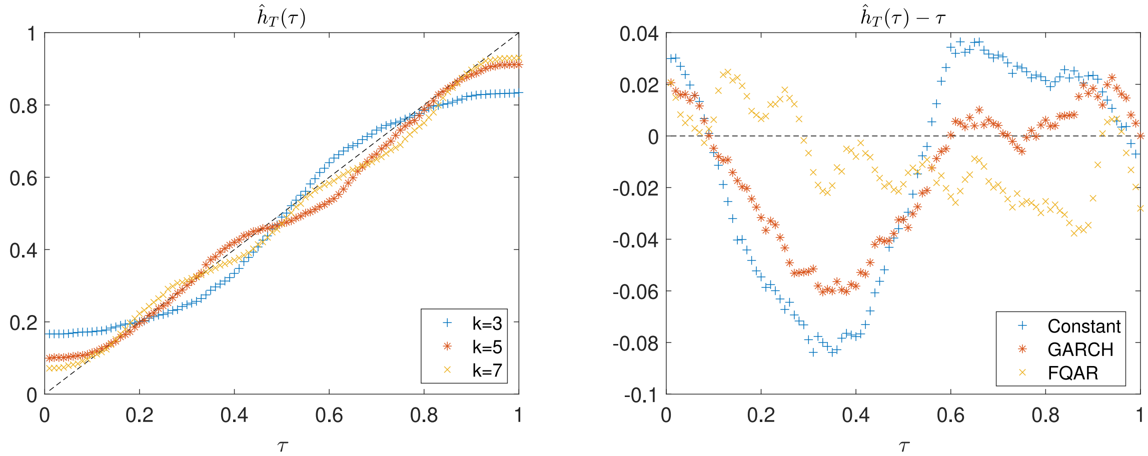

Fig. 6. Estimated hit functions.

Therefore, the rearrangement procedure has minimal influence on the adjustment of the quantiles and subsequently has little impact on the estimation errors.

We then choose $T_0 = [T/2]= 1177$ and implement a rolling forecast of the conditional quantile function $\hat{q}_{t(+1)}(\tau)$ for $t = 1177,\ldots,2352$. Figure D.3 shows $\hat{q}_{t(+1)}(0.05)$, $\hat{q}_{t(+1)}(0.5)$ and $\hat{q}_{t(+1)}(0.95)$, along with the observed return. $\hat{q}_{t(+1)}(0.5)$ is the predicted conditional median or the center of the distribution. The yellow line ($\tau = 0.05$) and purple line ($\tau = 0.95$) can be viewed as the predicted 5% and 95% Value-at-Risk (VaR). It is seen that the two lines cover most of the observed data points.

To further investigate the empirical performance of the FQAR model, we compute the sample mean of the hit function defined as:

$$\hat{h}_T(\tau) = \frac{1}{T-T_0}\sum_{t=T_0}^{T-1} 1\left(y_{t+1} \leq \hat{q}_{t(+1)}(\tau)\right), \quad \tau \in (0,1)$$

where $1(\cdot)$ is the indicator function. If $\hat{q}_{t(+1)}(\tau)$ is correctly predicted, $\hat{h}_T(\tau)$ should lie on a 45° line.

The left panel of Fig. 6 depicts the empirical hit function $\hat{h}_T(\tau)$ of FQAR model with different values truncation parameter $k = 3$, 5 and 7. It is obvious that $\hat{h}_T(\tau)$ lies closer to the 45° line as k increases. The bias is positive in the left tail and negative in the right tail, while the bias is much smaller for larger k, consistent with the numerical evidence presented in Section 5. The right panel of Fig. 6 compares the performance between the GARCH and FQAR models. For the GARCH models, the quantiles are estimated based on Eq. (40), with a constant mean (estimated using the sample mean) and GARCH variances. The FQAR model is based on the truncation parameter $k = 10$. We also added the performance of a simple i.i.d. Gaussian model with constant mean and variance estimated by their sample versions, labeled as 'Constant' in the figure. In order to enhance the visual clarity of the results, we represent the errors by plotting $\hat{h}_T(\tau) - \tau$ in the right panel, which clearly demonstrates that the errors of FQAR model are significantly smaller than those of the traditional models, especially for the quantiles near the median. The superior performance of the FQAR model can be attributed to its more flexible model specification that allows a general functional relationship between the current distribution and its past. In contrast, the GARCH model used here assumes that the relationship between the current (conditional) distribution and its past is captured by the relationship between the current conditional variance and its past.

7. Conclusion

Recognizing the fact that the (conditional) distributions of economic or financial variables can have important impact on the future outcomes, we introduce a Functional Quantile Autoregression (FQAR) model. This model is designed to explore the impact of past distributional information on the current and future behavior of a time series. The flexibility of our proposed model lies in its functional dependence structure on the underlying dynamic mechanism, allowing for the consideration of various types of distributional information that may affect present or future outcomes. It is worth noting that, in contrast to the existing literature on functional autoregressive models relying on functionally observed data, our approach focuses on traditional time series.

The proposed FQAR model complements traditional time series models and expands the modeling options for time series analysis. In contrast to existing literature on time series that incorporates selected summary indices of past distributions – in particular moments (e.g. skewness and/or kurtosis) and tail measures (e.g. VaR or ES) – into empirical models, the FQAR model allows the current conditional quantile function to be associated with the entire past conditional distribution in a functional autoregressive form, as well as past observations or other covariates.

There are several avenues for further investigation and extension of the FQAR model and the proposed sieve estimation method. We anticipate that the model introduced in this paper will find broad applicability in the examination of distributional dynamics across various applied areas.

Appendix. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2024.105765.

References

Adrian, T., Boyarchenko, N., Giannone, D., 2019. Vulnerable growth. Amer. Econ. Rev. 109, 1263–1289.
Ai, C., Chen, X., 2007. Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. J. Econometrics 141, 5–43.
Barlow, R., Bartholomew, D., Bremner, J., Brunk, H., 1972. Statistical Inference Under Order Restrictions. Wiley, New York.
Beirne, J., Caporale, G.M., Schulze-Ghattas, M., Spagnolo, N., 2010. Global and regional spillovers in emerging stock markets: A multivariate GARCH-in-mean analysis. Emerg. Mark. Rev. 11 (3), 250–260.
Belloni, A., Chernozhukov, V., Chetverikov, D., Fernandez-Val, I., 2019. Conditional quantile processes based on series or many regressors. J. Econometrics 213, 4–29.
Belloni, A., Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. J. Econometrics 186, 345–366.
Bosq, D., 2000. Linear Processes in Function Spaces: Theory and Applications. In: Lecture Notes in Statistics, Springer, New York.
Brockwell, P., Davis, R., 1991. Time Series: Theory and Method. Springer Series in Statistics, Springer-Verlag, New York.
Cardot, H., Crambes, C., Sarda, P., 2005. Quantile regression when the covariates are functions. Nonparametric Statistics 17, 841–856.
Chang, J., Chen, S., Chen, X., 2015. High dimensional generalized empirical likelihood for moment restrictions with dependent data. J. Econometrics 185, 283–304.
Chang, M., Chen, X., Schorfheide, F., 2021. Heterogeneity and Aggregate Fluctuations. Working Paper.
Chang, M., Schorfheide, F., 2022. On the Effects of Monetary Policy Shocks on Earnings and Consumption Heterogeneity. Working Paper.
Chen, X., Christensen, T., 2015. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. J. Econometrics 188, 447–465.
Chen, K., Müller, H., 2012. Conditional quantile analysis when covariates are functions, with application to growth data. J. R. Stat. Soc. Ser. B Stat. Methodol 74, 67–89.
Chen, R., Tsay, R., 1993. Nonlinear additive ARX models. J. Amer. Statist. Assoc. 88, 955–967.
Chernozhukov, V., Fernádex-Val, I., Galichon, A., 2009. Improving point and interval estimators of monotone functions by rearrangement. Biometrika 96, 559–575.
Chernozhukov, V., Fernádex-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78, 1093–1125.
Coibion, O., Gorodnichenko, Y., Kueng, L., Silvia, J., 2017. Innocent bystanders? Monetary policy and inequality. J. Monetary Econ. 88, 70–89.
Dai, Z., Zhu, J., Zhang, X., 2022. Time-frequency connectedness and cross-quantile dependence between crude oil, Chinese commodity market, stock market and investor sentiment. Energy Econ. 114, 106226.
Daouia, A., Park, B., 2013. On projection-type estimators of multivariate isotonic functions. Scand. J. Stat. 40, 363–386.
de Castro, L., Costa, B.N., Galvao, A.F., Zubelli, J.P., 2023. Conditional quantiles: An operator-theoretical approach. Bernoulli 29, 2392–2416.
De Nicolò., G., Gamba, A., Lucchetta, M., 2014. Microprudential regulation in a dynamic model of banking. Rev. Financ. Stud. 27 (7), 2097–12138.
Dette, H., Neumeyer, N., Pilz, K., 2006. A simple nonparametric estimator of a strictly monotone regression function. Bernoulli 12, 469–490.
Dong, C., Gao, J., stheim, D.T., 2016. Estimation for single-index and partially linear single-index integrated models. Ann. Statist. 44, 425–453.
Dong, C., Linton, O., 2018. Additive nonparametric models with time variable and both stationary and nonstationary regressors. J. Econometrics 207, 212–236.
Dong, C., Linton, O., Peng, B., 2021. A weighted sieve estimator for nonparametric time series models with nonstationary variables. J. Econometrics 222, 909–932.
Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. J. Bus. Econom. Statist. 22, 367–381.
Ferraty, A., Vieu, P., 2005. Conditional quantiles for dependent functional data with application to the climatic phenomenon. Sankhyā 378–398.
Han, H., Linton, O., Oka, T., Whang, Y.J., 2016. The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. J. Econometrics 193 (1), 251–270.
Hörmann, S., Kokoszka, P., 2012. Functional Time Series. In: Handbook of Statistics, vol. 30, Elsevier, UK.
Horowitz, J., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. J. Amer. Statist. Assoc. 100, 1238–1249.
Huang, J.Z., Yang, L., 2004. Identification of non-linear additive autoregressive models. J. R. Stat. Soc. Ser. B Stat. Methodol. 66 (2), 463–477.
Jorion, P., 2006. Risk management for hedge funds with position information. Available at SSRN 969988.

Kato, K., 2012. Estimation in functional linear quantile regression. Ann. Statist. 40, 3018–3136.
Koenker, R., Xiao, Z., 2006. Quantile autoregression. J. Am. Statist. Assoc. 101, 980–990.
Liu, X., Xiao, H., Chen, R., 2016. Convolutional autoregressive models for functional time series. J. Econometrics 194, 263–282.
Millick, H., Padhan, H., Mahalik, M., 2019. Does skewed pattern of income distribution matter for the environmental quality? Evidence from selected BRICS economies with an application of quantile-on-quantile regression (QQR) approach. Energy Policy 129, 120–131.
Muhammad, S., Long, X., Salman, M., Dauda, L., 2020. Effect of urbanization and international trade on CO2 emissions across 65 belt and road initiative countries. Energy 196, 117102.
Müller, H.G., Stadtmüller, U., 2005. Generalized functional linear models. Ann. Statist. 33, 774–805.
Park, J.Y., Qian, J., 2012. Functional regression of continuous state distributions. J. Econometrics 167, 397–412.
Patton, A.J., Ziegel, J.F., Chen, R., 2019. Dynamic semiparametric models for expected shortfall (and value-at-risk). J. Econometrics 211 (2), 388–413.
Pham, L., 2021. Frequency connectedness and cross-quantile dependence between green bond and green equity markets. Energy Econ. 98, 105257.
Powell, J., 1984. Least absolute deviations estimation for the censored regression model. J. Econometrics 25, 303–325.
Sim, N., Zhou, H., 2015. Oil prices, US stock return, and the dependence between their quantiles. J. Bank. Financ. 55, 1–8.
Taylor, A., Lay, D., 1980. Introduction to Functional Analysis. John Wiley & Sons, New York.
Westling, T., van der Laan, M.J., Carone, M., 2020. Correcting an estimator of a multivariate monotone function with isotonic regression. Electron. J. Stat. 14 (2), 3032–3069.
White, H., Kim, T., Manganelli, S., 2010. Modeling autoregressive conditional skewness and kurtosis with multi-quantile caviar. In: Volatility and Time Series Econometrics: Essays in Honor of Robert Engle. Oxford University Press.
Yahya, M., Ghosh, S., Kanjilal, K., Dutta, A., Uddin, G.S., 2020. Evaluation of cross-quantile dependence and causality between nonferrous metals and clean energy indexes. Energy 202, 117777.
Zhao, Z., Zhang, Z., Chen, R., 2019. Modeling maxima with autoregressive conditional Fréchet model. J. Econometrics 207, 325–351.