Article

Assessing Tail Risk Using Expectile Regressions with Partially Varying Coefficients

Zongwu Cai [1], Ying Fang [2,*] and Dingshi Tian [2]

[1] Department of Economics, University of Kansas, 1460 Jayhawks Blowhard, Lawrence, KS 66045, USA; caiz@ku.edu

[2] Wang Yanan Institute for Studies in Economics, Department of Statistics, School of Economics, Ministry of Education Key Laboratory of Econometrics and Fujian Key Laboratory of Statistical Science, Xiamen University, 422 Siming South Road, Xiamen, Fujian 361005, China; tiandingshi@gmail.com

* Correspondence: yifst1@xmu.edu.cn or yifst1@gmail.com

Abstract: To characterize heteroskedasticity, nonlinearity, and asymmetry in tail risk, this study investigates a class of conditional (dynamic) expectile models with partially varying coefficients in which some coefficients are allowed to be constants, but others are allowed to be unknown functions of random variables. A three-stage estimation procedure is proposed to estimate both the parametric constant coefficients and nonparametric functional coefficients. Their asymptotic properties are investigated under a time series context, together with a new simple and easily implemented test for testing the goodness of fit of models and a bandwidth selector based on newly defined cross-validatory estimation for the expected forecasting expectile errors. The proposed methodology is data-analytic and of sufficient flexibility to analyze complex and multivariate nonlinear structures without suffering from the curse of dimensionality. Finally, the proposed model is illustrated by simulated data, and applied to analyzing the daily data of the S&P500 return series.

1. Introduction

How to properly assess tail risk is one of the most important and challenging tasks in financial risk management. Expectile, as an alternative risk measure to value at risk (VaR), has received more attentions in recent years. VaR denotes the loss that is likely to be exceeded by a specified probability level, which is actually the quantile of a portfolio loss distribution. However, in cases where the size of extreme losses matters—for example, the occurrence of catastrophic events—VaR becomes a very conservative tail risk measure because a quantile-based risk measure depends only on the probability of the occurrence of extreme losses rather than the magnitude of extreme losses. Expectile, first introduced by Newey and Powell (1987), can rectify such an

undesirable situation by defining a risk measure based on the minimization of asymmetrically weighted mean square errors. Moreover, expectile has more merits compared with other popular risk measures in several ways. For example, expectile is considered to be a better alternative to both VaR and expected shortfall because it shares desirable properties of coherence and elicitability (See, for example, the papers by Bellini et al. (2014), Bellini and Valeria (2015), and Ziegel (2016) for details). Another advantage is that expectile is easier computed than VaR and expected shortfall, which is attractive in applications. Finally, since there exists a one-to-one mapping between quantiles and expectiles, as argued in Efron (1991), Jones (1994), and Yao and Tong (1996)), and the link between VaR and expected shortfall as addressed in Taylor (2008), expectile can be used to calculate both VaR and expected shortfall.

By virtue of the aforementioned advantages of expectile, there has been an increasing number of studies devoted to developing conditional expectile models in recent years. For example, Kuan et al. (2009) proposed a class of conditional autoregressive expectile (CARE) models that allow for asymmetric dynamic effects of the magnitude of positive and negative lagged returns on tail expectiles, while De Rossi and Harvey (2009) proposed applying a modified state space signal extraction algorithm to estimate time-varying expectiles, which may offer an alternative method to that in Kuan et al. (2009). Recently, Xie et al. (2014) enriched the conditional dynamic expectile model by including variables reflecting current information of investment environment and adopting a varying coefficient setup. In this way, a varying coefficient setup allows the conditional expectile model to be linear in some components with their coefficients determined by unknown functions of other variables. Compared with the aforementioned parametric models, a varying coefficient model can provide more flexibility and capture parameter heterogeneity and nonlinearity. Furthermore, a varying coefficient model can accommodate structural information by choosing smoothing variables and alleviate the curse of dimensionality by adopting an additive structure (See, for example, Cai et al. (2000) for more details).

In this study, inspired by the empirical studies on characterizing heteroskedasticity and nonlinearity, as well as asymmetry, in assessing the tail risk of asset returns for S&P500, we consider a new class of conditional dynamic expectile models with partially varying coefficients. This new model adopts a partially linear form, in which some coefficients are assumed to be constant, while other coefficients are allowed to depend on some smoothing variables selected by economic theories or stylized facts. It is flexible enough to include both models in Kuan et al. (2009) and Xie et al. (2014) as special cases. Particularly, it shares not only all merits of a fully varying coefficient model, but also can achieve more efficient estimation for the parametric coefficient part. Different from a fully varying coefficient model in Xie et al. (2014), the partially linear setup leads itself to a three-stage estimation procedure. The first stage is to fit a fully varying coefficient model, the second stage helps achieve the estimation of constant parameters with a parametric convergence rate, and the third stage re-estimates the varying coefficients by using the estimates at the second stage. Now, an important statistical question in fitting model (1) arises if the coefficient functions are actually varying, or more generally, if a parametric model fits the given data. This amounts to testing whether the coefficient functions are constant or in a certain parametric form. To this end, a simple constancy test is developed to test varying coefficients to see if they really depend on particular economic variables. Finally, the proposed model and its inferential procedures are applied to find a suitable expectile model to assess the tail risk of daily returns of S&P500 and the detailed analyses are reported in Section 3.

The rest of the paper is organized as follows: Section 2 introduces the new model and proposes the estimation method. The asymptotic properties of the proposed estimators are investigated in Section 2 too,

together with a simple and fast algorithm for bandwidth selection and smoothing variable selection. A simple and easily implemented test is proposed for testing whether functional coefficients are really changing or not. Monte Carlo experiments and empirical analysis results of a real data examples are reported in Section 3. Finally, Section 4 concludes the paper. All technical proofs are deferred to the appendix.

2. Expectile Models with Partially Varying Coefficients

2.1. Model Setup

Assume that $(Y_t, \mathbf{U}_t, \mathbf{X}_t), t = 1, 2, \ldots, n$ is a sequence of strictly stationary random vectors. The τ-th conditional expectile of Y_t, given $\mathbf{U}_t = \mathbf{u}$ and $\mathbf{X}_t = \mathbf{x}$, is then defined by

$$e_\tau(\mathbf{u}, \mathbf{x}) = \arg\min_{\xi \in \mathcal{R}} E\{Q_\tau(Y_t - \xi) | \mathbf{U}_t = \mathbf{u}, \mathbf{X}_t = \mathbf{x}\},$$

where $Q_\tau(v) = |\tau - I(v \leq 0)| v^2$ with $I(\cdot)$ denoting the indicator function. This study considers the τ-th conditional expectile of Y_t, given $\mathbf{U}_t$ and $\mathbf{X}_t$, with a partially varying coefficient framework as

$$e_\tau(\mathbf{U}_t, \mathbf{X}_t) = \mathbf{a}_\tau^\top \mathbf{X}_{t,1} + \mathbf{b}_\tau^\top(\mathbf{U}_t) \mathbf{X}_{t,2}, \tag{1}$$

where $\mathbf{X}_t = (\mathbf{X}_{t,1}^\top, \mathbf{X}_{t,2}^\top)^\top \in \mathcal{R}^{p+q}$, and $\mathbf{U}_t$ is a smoothing variable. Here, both $\mathbf{X}_t$ and $\mathbf{U}_t$ are allowed to include the past returns of Y_t so that the model is dynamic. Without loss of generality, it is assumed that $\mathbf{U}_t = U_t$ is a scalar variable for simplicity. Moreover, $\mathbf{a}_\tau = (a_{1,\tau}, \ldots, a_{p,\tau})^\top$ denotes a vector of constant coefficients of $\mathbf{X}_{t,1}$, and $\mathbf{b}_\tau(\cdot) = (b_{1,\tau}(\cdot), \ldots, b_{q,\tau}(\cdot))^\top$ is a vector of functional coefficients of $\mathbf{X}_{t,2}$. For simplicity, τ is dropped in $\mathbf{a}_\tau$ and $\mathbf{b}_\tau(\cdot)$ from now on if it does not cause any confusion.

The above model is general enough to include some existing expectile models as special cases. For example, the CARE model proposed by Kuan et al. (2009) can be considered a special case of a partially varying coefficient expectile model, where the coefficients of the intercept term and past returns are constant, but the coefficients of the magnitude of past return, measured either by the square of past returns or by the absolute value of past returns, are varying, depending on whether the past returns are positive or negative. Moreover, if the constant coefficients, $\mathbf{a}$, are not included, the above model becomes a fully varying coefficient expectile model, as in Xie et al. (2014).

2.2. Estimation Procedures

2.2.1. Three-Stage Estimation Procedure

Similar to the quantile model with partially varying coefficients in Cai and Xiao (2012), the well-known estimation method in Robinson (1988) or profile least squares estimation approach in Speckman (1988) for classical semiparametric regression estimation approach may not be applied to estimating $\mathbf{a}$ and $\mathbf{b}(\cdot)$ due to the fact that the expectile model does not have explicit normal equations. Therefore, estimation of a partially varying coefficient model is not trivial compared with a fully varying coefficient model, as in Xie et al. (2014). To estimate $\mathbf{a}$ and $\mathbf{b}(\cdot)$, the following estimation procedures are proposed. First, $\mathbf{a}$ is considered a function of U_s

for $1 \leq s \leq n$, and then, based on the local constant approximation, $\mathbf{a}(U_s)$ can be estimated by minimizing the following locally weighted loss function:

$$\min_{a,b} \sum_{t=1}^{n} Q_\tau \left(Y_t - \mathbf{a}^\top(U_s)\mathbf{X}_{t,1} - \mathbf{b}^\top(U_s)\mathbf{X}_{t,2} \right) K_{h_1}(U_t - U_s),$$

where $K(\cdot)$ is a kernel function, $K_{h_1}(x) = K(x/h_1)/h_1$, and h_1 denotes the bandwidth used at this step, which controls the smoothness; it satisfies $h_1 = h_1(n) \to 0$ and $n\,h_1^2 \to \infty$. The local constant estimator for $\mathbf{a}(U_s)$ is obtained, denoted by $\hat{\mathbf{a}}(U_s)$. To improve estimation efficiency for $\mathbf{a}$ by using full sample information, at the second stage, one can take a simple average method for $\hat{\mathbf{a}}(U_s)$, which is given by

$$\tilde{\mathbf{a}} = \tilde{\mathbf{a}}_\tau = \frac{1}{n} \sum_{s=1}^{n} \hat{\mathbf{a}}(U_s),$$

which is shown in Theorem 1 (later), in which the above estimator is $\sqrt{n}$-consistent and asymptotically normally distributed.

Finally, $\mathbf{b}(\cdot)$ is re-estimated by using the partial expectile residual, $Y_{t1}^* = Y_t - \tilde{\mathbf{a}}^\top \mathbf{X}_{t,1}$, where $\tilde{\mathbf{a}}$ is a $\sqrt{n}$-consistent estimator of $\mathbf{a}$, obtained possibly from the second stage. Thus, for the given grid point, u_0, the estimator of $\mathbf{b}(u_0)$ can be obtained by the following minimization problem using local linear approximation of $\mathbf{b}(U_t)$ at the grid point u_0:

$$\min_{b,b'} \sum_{t=1}^{n} Q_\tau \left(Y_{t1}^* - \mathbf{b}^\top(u_0)\mathbf{X}_{t,2} - \mathbf{b}'^\top(u_0)\mathbf{X}_{t,2}(U_t - u_0) \right) K_{h_2}(U_t - u_0),$$

where h_2 denotes the bandwidth at this stage and $\mathbf{b}'(\cdot)$ is the first order derivative of $\mathbf{b}(\cdot)$. The local linear estimator of $\mathbf{b}(u_0)$ is denoted by $\tilde{\mathbf{b}}(u_0)$.

2.2.2. Bandwidth Selection

Bandwidth selection is always a challenging issue for any semiparametric model in real applications. For the proposed three-stage estimation procedure, the model needs to select bandwidths, h_1, at the first stage, and h_2 at the third stage. There is no existing theory available in the literature on how to select h_1 optimally at the first stage. However, our simulation results show that the estimation of $\mathbf{b}(u_0)$ is not sensitive to the choice of h_1 as long as the first stage estimation is under-smoothed. For the selection of h_2 at the last stage, the multifold cross-validation criterion proposed by Cai et al. (2000) for the mean regression model is extended to the proposed expectile model, briefly described below. The main idea behind this approach is that, since the classical cross-validation may not work well for time series data in the literature, this simple and quick procedure is attentive to the structure of stationary time series data.

Let m and Q be two positive integers. The window, l, satisfies $n > lQ$. First, with the Q sub-series of length $n - ql$ ($q = 1, \ldots, Q$), the unknown functions are estimated. Based on the estimated model, the one-step forecasting errors of the length-l time series of the next section are computed. Specifically, the optimal bandwidth is obtained by minimizing the average asymmetric mean squared error (AAMSE),

$$\text{AAMSE}(h_2) = \sum_{q=1}^{Q} \text{AAMSE}_q(h_2), \tag{2}$$

where for $1 \leq q \leq Q$,

$$\text{AAMSE}_q(h_2) = \frac{1}{m} \sum_{t=n-ql+1}^{n-ql+l} Q_\tau \left(Y_t - \tilde{\mathbf{a}}_\tau^\top \mathbf{X}_{t,1} - \tilde{\mathbf{b}}_\tau^\top(U_t)\mathbf{X}_{t,2} \right).$$

It is worth noting that bandwidth is rescaled for different sample size according to the optimal rate $h_2 = O(n^{-1/5})$, and one can take $l = [0.1n]$ and $Q = 4$ in practical implementations, as suggested in Cai et al. (2000). Note that a similar idea to the above selection procedure outlined in equation (2) was adopted in Xie et al. (2014) too.

2.2.3. Smoothing Variable Selection

Choosing an appropriate smoothing variable, U_t, is of great importance in applying functional coefficient models. To this end, economic theory or knowledge on the real data can be helpful. Nevertheless, if without any prior information, some data-driven model selection methods, such as Akaike information criterion, cross validation and other criteria are also suggested. Here, an easily implemented approach is proposed as follows. The first is to select a potential set of U_t based on theory or existing models, and then the optimal U_t is obtained when it reaches the minimum AAMSE value defined in equation (2). In the empirical study conducted in Section 3.2, the practical implementation of this approach is presented.

2.3. Large Sample Theory

In this section, asymptotic properties for both the proposed estimators, $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}(u_0)$, are presented, respectively. Moreover, to improve its estimation efficiency, some weighted average estimators for $\mathbf{a}$ are addressed. Finally, a simple test on testing constancy is developed. It is shown to have an asymptotical Chi-square distribution.

2.3.1. Notations and Assumptions

Note that some notations are defined here and used throughout the paper. First, $f_u(\cdot)$ denotes the marginal density of U_t, and $f_{y|u,x}(\cdot)$ and $F_{y|u,x}(\cdot)$ are the conditional density function and distribution function of Y_t, given U_t and $\mathbf{X}_t$, respectively. Moreover, we define $\mathbf{\Omega}(u) = E[\mathbf{X}_t\mathbf{X}_t^\top|U_t = u]$, $\mathbf{\Gamma}^*(u) = E[Q_\tau'(Y_t - e_\tau(U_t,\mathbf{X}_t))^2\mathbf{X}_t\mathbf{X}_t^\top|U_t = u]$, $\mathbf{\Gamma}(u) = 2E[(\tau\{1 - F_{y|u,x}(e_\tau(U_t,\mathbf{X}_t))\} + (1-\tau)F_{y|u,x}(e_\tau(U_t,\mathbf{X}_t)))\mathbf{X}_t\mathbf{X}_t^\top|U_t = u]$, $\mathbf{\Omega}_2(u) = E[\mathbf{X}_{t,2}\mathbf{X}_{t,2}^\top|U_t = u]$, and $\mathbf{\Gamma}_2(u) = 2E[(\tau\{1 - F_{y|u,x}(e_\tau(U_t,\mathbf{X}_t))\} + (1-\tau)F_{y|u,x}(e_\tau(U_t,\mathbf{X}_t)))\mathbf{X}_{t,2}\mathbf{X}_{t,2}^\top|U_t = u]$, where $\mu_j = \int u^j K(u)\,\mathrm{d}u$ and $\nu_j = \int u^j K^2(u)\,\mathrm{d}u$ for $j \geq 0$.

Now, assumptions are presented here for deriving asymptotic results. Note that these assumptions given in the paper are sufficient conditions, but not necessarily the weakest ones.

Assumption A. (A1) $\mathbf{b}(u)$ *is twice continuously differentiable in u. Further,* $f_u(u)$ *is continuously differentiable and has a support,* $\{u : 0 < F_u(u) < 1\}$, *and* $f_{y|u,x}(\cdot)$ *is bounded and satisfies the Lipschitz condition.*

(A2) *The kernel function,* $K(\cdot)$, *is a bounded nonnegative symmetric function with compact support.*

(A3) *The bandwidth,* h_1, *satisfies* $h_1 \to 0$ *and* $nh_1^2 \to \infty$.

(A4) $|f(u, v|\mathbf{x}_0, \mathbf{x}_l)| \leq C < \infty$, for all $l \geq 1$, where $f(u, v|\mathbf{x}_0, \mathbf{x}_l)$ is the conditional density of (U_0, U_l), given $(\mathbf{X}_0, \mathbf{X}_l)$.

(A5) The process $\{(Y_t, \mathbf{X}_t, \mathbf{U}_t)\}_{t=1}^n$ is β-mixing with the mixing coefficient, $\beta(\cdot)$, satisfying $\sum_{k=1}^\infty k^2 [\beta(k)]^{\delta/(1+\delta)} < \infty$ for some $\delta > 0$.

(A6) $E||\mathbf{X}_t||^{2(\delta+1)} < \infty$ for some $\delta > 0$. Further, functions $\mathbf{\Omega}(u_0)$ and $\mathbf{\Gamma}(u_0)$ are continuous in a neighbourhood of u_0, with their inverse functions being uniformly bounded.

Assumption B. (B1) $E|\mathbf{X}_{t,2}|^{2(\delta^*+1)} < \infty$ for $\delta^* > \delta$, $\mathbf{\Omega}_2(u_0)$ and $\mathbf{\Gamma}_2(u_0)$ and their inverse functions are uniformly bounded.

(B2) There exists a sequence of positive integers, s_n, when $n \to \infty$, to satisy $s_n \to \infty$, $s_n = o(\sqrt{nh_2})$, and $\sqrt{n/h_2}\beta(s_n) \to 0$.

(B3) As $n \to \infty$, $h_2 = h_2(n) \to 0$, $n^{1/2-\delta/4} h_2^{\delta/\delta^* - \delta/4 - 1/2} = O(1)$, and $h_1/h_2 = o(1)$.

Remark 1 (Discussions of Assumptions). *First, Assumptions* A1–A4 *are standard in the nonparametric literature. Assumption A5 is also used in Cai and Xiao (2012), and is stronger than Assumptions A3–A4 in Xie et al. (2014), which ensures the second-stage estimator, $\tilde{\mathbf{a}}$, to be $\sqrt{n}$-consistent. $E||\mathbf{X}_t||^{2(\delta+1)} < \infty$ in Assumption A6 is generally required to ensure that $1/n \sum_{t=1}^n \mathbf{X}_t \mathbf{X}_t^\top \to E(\mathbf{X}_t \mathbf{X}_t^\top)$ for a mixing process. The boundedness of the inverse function of $\mathbf{\Omega}(u_0)$, $\mathbf{\Omega}_2(u_0)$, $\mathbf{\Gamma}(u_0)$ and $\mathbf{\Gamma}_2(u_0)$ are the necessary and sufficient conditions for the model identification at the first and third stages. For the same reason, Assumption B1 is required for the estimation at the third stage. To satisfy both Assumptions A5 and B2, a sufficient condition for the mixing coefficient, $\beta(n)$, is provided. Suppose that $h_2 = O(n^{-\rho})$ for some $0 < \rho < 1$, $s_n = (nh_2/\log n)^{1/2}$, and $\beta(n) = O(n^{-d})$. Then, Assumption A5 is satisfied if $d > 3/\delta(\delta+1)$, and Assumption B2 is satisfied for $d > (1+\rho)/(1-\rho)$. Hence, if $\beta(n) = O(n^{-d})$ and $d > \max\{3/\delta(\delta+1), (1+\rho)/(1-\rho)\}$, both conditions are satisfied. Assumption B3 is a technical condition. Clearly, if $\delta > 3$, or if $\delta < \delta^* \leq 1 + 1/(3-\delta)$ is satisfied when $2 < \delta < 3$, Assumption B3 is automatically satisfied (See, for example, Cai et al. (2000) and Cai (2002a) for more details).*

2.3.2. Asymptotic Properties

Let us first provide the asymptotic properties of the constant coefficients estimator, $\tilde{\mathbf{a}}$. To simplify presentation, the asymptotic result is stated here only with all technical details relegated to the appendix. The main idea of the proof is that, under certain conditions, $\hat{\mathbf{a}}(U_t)$ can be expressed as a linear estimator plus a higher order term. In such a way, the average estimator, $\tilde{\mathbf{a}}$, can be formulated as a U-statistic plus a higher order term, and then the asymptotic normality can be obtained by applying the central limit theorem of a U-statistic (See, for example, Dette and Spreckelsen (2004)). Now, more notations are needed. $\varphi(z_t, z_t) = Q'_\tau(Y_t - e_\tau(U_t, \mathbf{X}_t))$, $\mathbf{e}_1^\top = (\mathbf{I}_p, \underline{0}_{p\times q})$, $\mathbf{I}_p$ denotes the p-dimensional identity matrix, $\underline{0}_{p\times q}$ represents the $p \times q$ zero matrix, $\mathbf{\Gamma}'(\cdot)$ and $f'_u(\cdot)$ are the first order derivatives of $\mathbf{\Gamma}(\cdot)$ and $f_u(\cdot)$, respectively; and $\mathbf{b}''_\tau(\cdot)$ stands for the second order derivative of $\mathbf{b}_\tau(\cdot)$. Next, the asymptotic normality of $\tilde{\mathbf{a}}_\tau$ is stated in the following theorem.

Theorem 1. *Suppose Assumption* A *holds. Then,*

$$\sqrt{n}[\tilde{\mathbf{a}}_\tau - \mathbf{a}_\tau - B_a] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Sigma}_a), \tag{3}$$

where the asymptotic bias term is $B_a = \frac{1}{2} B_1^* h_1^2$ *with*

$$B_1^* = \frac{\mu_2}{\mu_0} \mathbf{e}_1^\top E \left[2 \left(\mathbf{\Gamma}^{-1}(U_t) \mathbf{\Gamma}'(U_t) + f_u'(U_t)/f_u(U_t) \right) \begin{pmatrix} 0 \\ \mathbf{b}_\tau'(U_t) \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{b}_\tau''(U_t) \end{pmatrix} \right],$$

and the asymptotic variance is given by

$$\begin{aligned} \mathbf{\Sigma}_a = \frac{1}{\mu_0^2} \Big\{ &E[\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_t) \mathbf{\Gamma}^*(U_t) \mathbf{\Gamma}^{-1}(U_t) \mathbf{e}_1] \\ &+ 2 \sum_{t=1}^\infty Cov\big(\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_1) \mathbf{X}_1 \varphi(z_1, z_1), \mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_{t+1}) \mathbf{X}_{t+1} \varphi(z_{t+1}, z_{t+1}) \big) \Big\}. \end{aligned}$$

From Theorem 1, one can observe that the estimator of constant coefficients has a parametric convergence rate. When $nh_1^4 \to 0$, the asymptotic bias term, B_a, in Theorem 1 converges to 0, so that

$$\sqrt{n}[\tilde{\mathbf{a}}_\tau - \mathbf{a}_\tau] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Sigma}_a),$$

which implies that to obtain $\tilde{\mathbf{a}}_\tau$, one needs to use the under-smoothing technique in the sense that $nh_1^4 \to 0$.

Remark 2. *It is possible to improve the estimation efficiency of $\tilde{\mathbf{a}}$ by using a weighted average method. Since the estimation of $\mathbf{a}$ might be influenced by the tail behaviour of the distribution of U_t, similar to Cai and Masry (2000), one can use a trimming function, $w_t = I(U_t \in \mathcal{U})$, with a compact set, $\mathcal{U} \in \mathcal{R}$, which leads to the following weighted average estimator:*

$$\mathbf{a}^w = \left[\sum_{t=1}^n w_t \right]^{-1} \sum_{t=1}^n w_t \hat{\mathbf{a}}(U_t).$$

Following Cai and Fan (2000), a general weighted average estimator can be given by

$$\check{\mathbf{a}} = \left[\sum_{t=1}^n W(U_t) \right]^{-1} \sum_{t=1}^n W(U_t) \hat{\mathbf{a}}(U_t).$$

A more efficient estimator can be obtained by choosing an optimal weighting function. Under certain regularity conditions, when $\varphi(z_t, z_t)$ is a martingale difference sequence, it can be showed that

$$\sqrt{n}[\check{\mathbf{a}} - \mathbf{a}] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \check{\mathbf{\Sigma}}_a),$$

where $\check{\mathbf{\Sigma}}_a = 1/\mu_0^2 E^{-1}[W(U_t)] E[W(U_t) \mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_t) \mathbf{\Gamma}^*(U_t) \mathbf{\Gamma}^{-1}(U_t) \mathbf{e}_1 W(U_t)] E^{-1}[W(U_t)]$. If the weighting function is chosen as

$$W_{opt}(U_t) = \mu_0^2 [\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_t) \mathbf{\Gamma}^*(U_t) \mathbf{\Gamma}^{-1}(U_t) \mathbf{e}_1]^{-1},$$

it is easy to show that the corresponding asymptotic variance is optimal, given by

$$\check{\mathbf{\Sigma}}_{a,opt} = E^{-1}[W_{opt}(U_t)], \tag{4}$$

which may be consistently estimable.

Next, the asymptotic properties for $\tilde{\mathbf{b}}(u_0)$ must be derived. To this end, some additional notations are needed. Let $\boldsymbol{\Gamma}_2^*(u) = E[Q_\tau^{'2}(Y_t - e_\tau(U_t, \mathbf{X}_t))\mathbf{X}_{t,2}\mathbf{X}_{t,2}^\top | U_t = u]$, $\boldsymbol{\Sigma}(u) = f_u(u)\mathrm{diag}\{\nu_0, \nu_2\} \otimes \boldsymbol{\Gamma}_2^*(u)$. Now, we have the following theorem:

Theorem 2. *Suppose Assumptions* A *and* B *hold. Then,*

$$\sqrt{nh_2}\left[\tilde{\mathbf{b}}(u_0) - \mathbf{b}(u_0) - \frac{\mu_2 h_2^2}{2\mu_0}\mathbf{b}''(u_0)\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Sigma}_b(u_0)), \tag{5}$$

where $\boldsymbol{\Sigma}_b(u_0) = \nu_0/(f_u(u_0)\mu_0^2)\boldsymbol{\Gamma}_2^{-1}(u_0)\boldsymbol{\Gamma}_2^*(u_0)\boldsymbol{\Gamma}_2^{-1}(u_0)$.

The estimator has the same asymptotic result, as in Xie et al. (2014). It is worth emphasizing that the asymptotic result is *oracle* in the sense that the asymptotic result in Theorem 2 is exactly the same as that for the case in which **a** would be known.

2.4. Inference

Now, our focus is on how to test constancy on varying coefficients, $\mathbf{b}(\cdot)$. A constancy test is usually of interest because one may need to know whether the varying coefficients depend on particular smoothing variables or not. Since the selection of smoothing variables is determined by economic theories, the constancy test here serves as a vehicle to test underlying economic theories. To this end, we consider a null hypothesis given by

$$H_0 : \mathbf{b}(u) = \mathbf{b}_0, \quad \text{for some unknown } \mathbf{b}_0.$$

In light of Cai and Xiao (2012), it is easy to show that

$$\|\sqrt{nh_2}\widehat{\boldsymbol{\Sigma}}_b^{-1/2}(u_j)(\tilde{\mathbf{b}}(u_j) - \widehat{\mathbf{b}}_0)\|^2 \xrightarrow{\mathcal{L}} \chi^2(q),$$

where $\{u_j\}_{j=1}^{m_u}$ is a sequence of m_u distinct points within the domain of U_t, $\widehat{\mathbf{b}}_0$ is the estimator under the null hypothesis, and $\chi^2(q)$ denotes a Chi-distribution with degrees of freedom, q, with the dimension of $\mathbf{X}_{t2}$. Hence, a simple and easily implemented test statistic, T_n, given below, has a limiting Chi-square distribution under the null:

$$T_n = \sum_{1 \le j \le m_u} \|\sqrt{nh_2}\widehat{\boldsymbol{\Sigma}}_b^{-1/2}(u_j)(\tilde{\mathbf{b}}(u_j) - \widehat{\mathbf{b}}_0)\|^2 \xrightarrow{\mathcal{L}} \chi^2(m_u q), \tag{6}$$

which is slightly different from that in Cai and Xiao (2012), proposing the use of the maximum rather than summation in equation (6). To calculate T_n, one needs to find a consistent estimator of $\boldsymbol{\Sigma}_b(u_0)$. As it is the upper left $q \times q$ matrix of $\mathbf{D}^{-1}(u_0)\boldsymbol{\Sigma}(u_0)\mathbf{D}^{-1}(u_0)$, where $\mathbf{D}(u) = f_u(u)\,\mathrm{diag}\{\mu_0, \mu_2\} \otimes \boldsymbol{\Gamma}_2(u)$. Then, $\hat{\boldsymbol{\Sigma}}_b(u_0)$ can be calculated using the easily implemented estimators as shown below. To this end, we define

$$\hat{\mathbf{D}}(u_0) = \frac{1}{nh_2}\sum_{t=1}^n K\left(\frac{U_t - u_0}{h_2}\right) Q_\tau''(Y_{t1}^* - \hat{\eta}(u_0, U_t, \mathbf{X}_t))\mathbf{Z}_{t,2}^*\mathbf{Z}_{t,2}^{*\top}$$

as an estimator of $\mathbf{D}(u_0)$, and

$$\hat{\boldsymbol{\Sigma}}(u_0) = \frac{1}{nh_2} \sum_{t=1}^{n} K^2 \left(\frac{U_t - u_0}{h_2} \right) Q_\tau'^2 (Y_{t1}^* - \hat{\eta}(u_0, U_t, \mathbf{X}_t)) \mathbf{Z}_{t,2}^* \mathbf{Z}_{t,2}^{*\top}$$

as an estimator of $\boldsymbol{\Sigma}(u_0)$, where $\hat{\eta}(u_0, U_t, \mathbf{X}_{t,2}) = \hat{\boldsymbol{\beta}}^\top(u_0) \mathbf{Z}_{t,2}$, $\boldsymbol{\beta}(u) = (\mathbf{b}^\top(u), \mathbf{b}'(u)^\top)^\top$, $\mathbf{Z}_{t,2} = (\mathbf{X}_{t,2}^\top, \mathbf{X}_{t,2}^\top(U_t - u_0))^\top$ and $\mathbf{Z}_{t,2}^* = (\mathbf{X}_{t,2}^\top, \mathbf{X}_{t,2}^\top(U_t - u_0)/h_2)^\top$. The consistency of both estimators is shown in Lemmas in the appendix.

Remark 3. *The testing procedure given by equation (6) is an asymptotic test. It has the advantage that its limiting distribution is free of any nuisance parameter. Alternatively, a Bootstrap-based test of (6) can be applied to improve finite sample performance. Clearly, another issue related to the proposed test is the choice of finite distinct points, $\{u_j\}_{j=1}^{m}$. In practice, one may consider, for example, choosing certain quartiles. In some applications, different choices of $\{u_j\}_{j=1}^{m_u}$ may potentially lead to different conclusions in a finite sample. Thus, it would be desirable to consider all points, u, on the domain of U_t so that some L_p-type tests may be constructed. Of course, investigating the properties of those test statistics warrants further future research.*

3. Simulation Studies and an Empirical Example

3.1. Simulation Studies

In this section, two simulated examples are used to illustrate the finite sample performance of the proposed model and its estimators. To measure the performance, the medians and standard deviations of the root mean squared errors (RMSEs) are reported. The RMSE for $\tilde{b}_j(\cdot)$ is defined by

$$\text{RMSE}_{b_j} = \left[\frac{1}{G} \sum_{k=1}^{G} \left\{ \tilde{b}_j(u_k) - b_j(u_k) \right\}^2 \right]^{1/2}, \quad 1 \le j \le q,$$

where $\{u_k\}_{k=1}^{G}$ are grid points within the domain of U_t. For RMSE_{a_j} of $\tilde{a}_j$, it is just the absolute deviation error, that is, $\text{RMSE}_{a_j} = |\tilde{a}_j - a_j|$. For each simulated example, sample sizes are considered to be $n = 200, 400$, and 800; simulations are repeated 500 times for each given sample size. Different probability levels are considered to be $\tau = 0.25, 0.50$, and 0.75. When generating the series of Y_t, the initial value is set to be zero and the first 100 observations are dropped to reduce the impact from the initial value. For the bandwidth used at the first step and by following the idea in Cai (2002b) and Cai and Xiao (2012), h_1 is set to be $d_1 n^{-1/10} h_0$ so that it is under-smoothing, where $h_0 = n^{-1/5}$, and $d_1 > 0$ is a constant. The bandwidth at the third stage is selected based on the modified multifold cross-validation criterion given in Cai et al. (2000). To be specific, h_2 is taken to $d_2 n^{-1/5}$, where d_2 ranges from 0.005 to 0.2, and we then choose the optimal d_2 to minimize $\text{AAMSE}(h_2)$ in (2).

For each simulated example, we compare the degree of sensitivity to tail events in quantile and expectile models by modeling extreme tail events similar to that in Kuan et al. (2009) and Xie et al. (2014). To be specific, two cases to model extreme values in the tail are considered. The ε_t is generated independently from either $N(0, 1/\sqrt{1-P})$ with probability $1-P$ or $N(c, 1/\sqrt{P})$ with probability P, where Case 1: $P = 0.01$ and

$\tau = \theta = 0.05$, and Case 2: $P = \tau = \theta = 0.01$. Here, θ denotes the probability level of quantile regression, c is set to take values from -1 to -50, and $n = 800$ for both cases.

Example 1. *The data generating process (DGP) is given by*

$$Y_t = a_1 Y_{t-1} + b_1(U_t) Y_{t-2} + \varepsilon_t, \qquad t = 1, \dots, n,$$

where $a_1 = 0.5$, $b_1(U_t) = -0.75 + 0.5\cos(\sqrt{2}\pi U_t)$, U_t *is generated from a Uniform* $(-1, 1)$, *and* ε_t *is i.i.d.* $N(0, 1)$. *In this example, the expectile model is given by* $e_\tau(Y_t) = e_\tau(\varepsilon_t) + a_1 Y_{t-1} + b_1(U_t) Y_{t-2}$.

Table 1. Median and standard deviation (in parentheses) of the RMSE values for Example 1

n	d_1	$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.75$	
		RMSE_{a_1}	RMSE_{b_1}	RMSE_{a_1}	RMSE_{b_1}	RMSE_{a_1}	RMSE_{b_1}
200	0.5	0.0511	0.1381	0.0533	0.1403	0.0604	0.1385
		(0.0485)	(0.0553)	(0.0448)	(0.0502)	(0.0494)	(0.0481)
	1	0.0508	0.1408	0.0531	0.1319	0.0561	0.1395
		(0.0495)	(0.0484)	(0.0413)	(0.0496)	(0.0481)	(0.0502)
	2	0.0533	0.1343	0.0479	0.1356	0.0588	0.1358
		(0.0491)	(0.0517)	(0.0453)	(0.0479)	(0.0480)	(0.0487)
400	0.5	0.0387	0.1044	0.0373	0.1015	0.0382	0.1034
		(0.0372)	(0.0338)	(0.0318)	(0.0345)	(0.0328)	(0.0350)
	1	0.0388	0.1070	0.0335	0.1028	0.0407	0.1086
		(0.0343)	(0.0328)	(0.0305)	(0.0346)	(0.0153)	(0.0342)
	2	0.0380	0.1050	0.0332	0.1028	0.0384	0.1055
		(0.0339)	(0.0350)	(0.0294)	(0.0318)	(0.0336)	(0.0367)
800	0.5	0.0264	0.0836	0.0247	0.0835	0.0291	0.0858
		(0.0254)	(0.0238)	(0.0213)	(0.0244)	(0.0246)	(0.0239)
	1	0.0264	0.0836	0.0257	0.0814	0.0282	0.0842
		(0.0254)	(0.0238)	(0.0220)	(0.0228)	(0.0234)	(0.0250)
	2	0.0256	0.0822	0.0244	0.0828	0.0263	0.0838
		(0.0237)	(0.0258)	(0.0215)	(0.0239)	(0.0228)	(0.0242)

Table 1 reports the medians and standard deviations (in parentheses) of RMSE values of $\tilde{a}_1$ (RMSE_{a_1}) and $\tilde{b}_1$ (RMSE_{b_1}) for all cases. First, one can see that the medians and standard deviations of RMSE values for all cases decrease as the sample size increases. For example, when $\tau = 0.50$, the median and standard deviation of the RMSE_{a_1} values for $n = 400$ are 0.037 and 0.032, respectively, and they decrease to 0.023 and 0.021, respectively, when the sample size is doubled. Clearly, the same pattern for RMSE_{b_1} can be observed too. Indeed, when $\tau = 0.50$ and the sample size is 400, the median is 0.102 and the corresponding standard deviation is 0.035. When the sample size increases to 800, the median and its standard deviation decrease to 0.084 and 0.024, respectively. Furthermore, Table 1 also reports the impact of different values of h_1 on the estimation of $\tilde{a}_1$. When h_1 is under-smoothed, different choices of d_1 in a reasonable range have very little impact on the estimation performance of $\tilde{a}_1$. For example, when $\tau = 0.25$ and the sample size is 800, the medians of RMSE_{a_1} are 0.0264, 0.0264, and 0.0256 when d_1 takes values of 0.5, 1, and 2, respectively. The standard deviations are almost same for different values of d_1.

Figure 1 depicts the degree of sensitivity of quantile and expectile models to catastrophic events in the tail. Figure 1(a) reports the results in Case 1 where $P = 0.01$ and $\tau = \theta = 0.05$. One can observe that the expectile model is very sensitive to the change of values of c, while the quantile model does not change when the values of c increase. In the Case 2, where $P = \tau = \theta = 0.01$, both vary with c. However, the change of expectile is relatively larger than that of the quantile for each c. The results here are similar to those obtained in Kuan et al. (2009) and Xie et al. (2014).

(a) Case 1: $P=0.01$, $\tau=0.05$, $\theta=0.05$ (b) Case 2: $P=0.01$, $\tau=0.01$, $\theta=0.01$
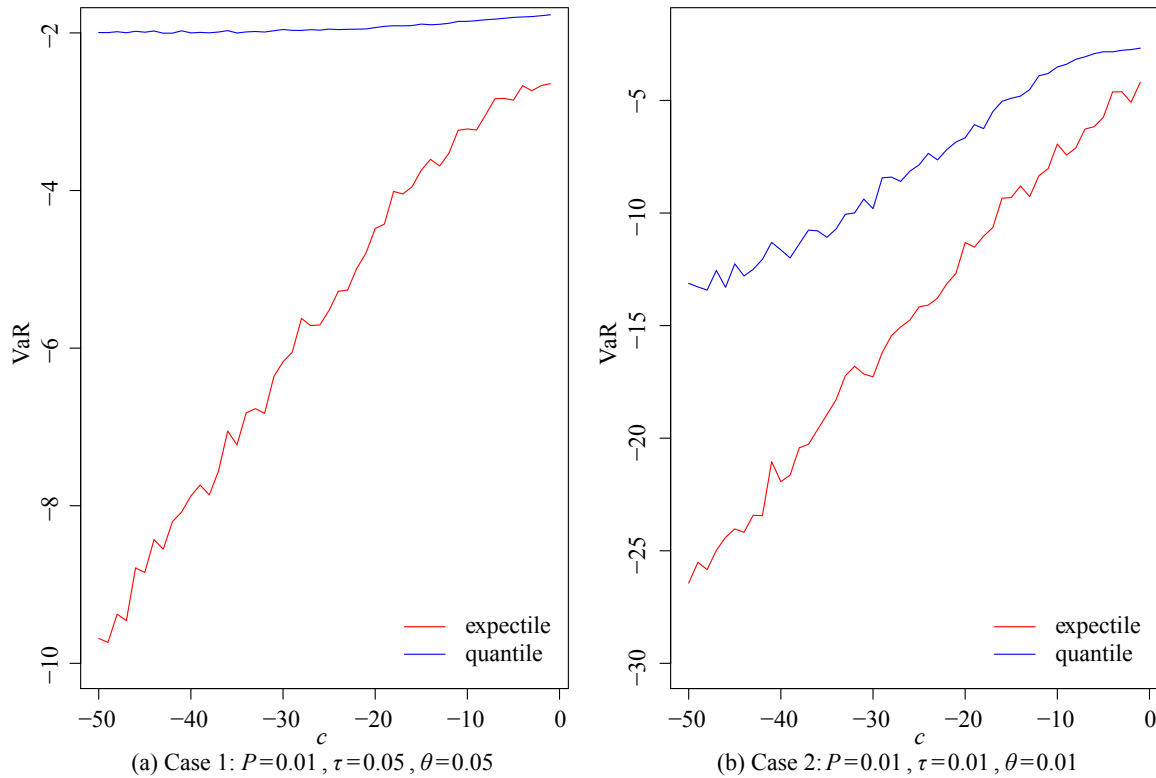
Figure 1. The sensitivity of quantile and expectile to extreme event for Example 1.

Example 2. *In this example, the following DGP is considered:*

$$Y_t = a_1 X_{t,1} + b_1(U_t)X_{t,2} + \sigma(U_t)\varepsilon_t, \qquad t = 1, \ldots, n,$$

where $a_1 = 0.5$, $b_1(U_t) = \cos(\sqrt{2}\pi U_t)$, and $\sigma(U_t) = \exp(-4(U_t - 1)^2) + \exp(-5(U_t - 2)^2)$. Here, $X_{t,1}$ and $X_{t,2}$ are generated from $X_{t,1} = 0.75X_{t-1,1} + v_{t,1}$ and $X_{t,2} = -0.5X_{t-1,2} + v_{t,2}$, respectively, with $v_{t,1} \sim i.i.d.N(0,1)$ and $v_{t,2} \sim i.i.d.N(0,1/4)$; U_t is generated from $U_t = 0.5U_{t-1} + v_{t,3}$ with $v_{t,3} \sim i.i.d.N(0,1)$; and $\varepsilon_t \sim i.i.d.N(0,1/4)$. The corresponding expectile regression model is then given by $e_\tau(Y_t|X_{t,1}, X_{t,2}, U_t) = e_\tau(\varepsilon_t)\sigma(U_t) + a_1 X_{t,1} + b_1(U_t)X_{t,2}$.

Table 2 reports the medians and the standard deviations (in parentheses) of the RMSE values for Example 2. First, one can observe that in all cases, both the medians and the standard deviations of RMSE values decrease as the sample size increases. For example, when $\tau = 0.5$, the medians of RMSE_{a_1} and RMSE_{b_1} values decrease from 0.0198 and 0.1324 to 0.0057 and 0.0684, respectively, when the sample size increases from 200 to 400. The

standard deviations decrease from 0.0331 and 0.0514 to 0.0052 and 0.0200, respectively, when the sample size increases from 200 to 800. One can see that RMSE_{a_1} value shrinks to zero quicker than RMSE_{b1} value because the former has a parametric convergence rate, while the latter has only a nonparametric rate. Similar to Figure 1, Figure 2 reports the magnitude of the sensitivity to catastrophic events for quantile and expectile models for Example 2 for two cases: Case 1 and Case 2. The results are quite similar to those observed in Figure 1. In conclusion, the expectile model is much more sensitive to the values of c than the quantile models in both simulated examples. This indicates clearly that the expectile seems to be a better risk measure than the quantile for the case of the occurrence of extreme events.

Table 2. Median and standard deviation (in parentheses) of the RMSE values for Example 2

n	$\tau = 0.25$		$\tau = 0.50$		$\tau = 0.75$	
	RMSE_{a_1}	RMSE_{b_1}	RMSE_{a_1}	RMSE_{b_1}	RMSE_{a_1}	RMSE_{b_1}
200	0.0207	0.1365	0.0198	0.1324	0.0207	0.1383
	(0.0273)	(0.0535)	(0.0331)	(0.0514)	(0.0259)	(0.0565)
400	0.0094	0.0999	0.0104	0.0925	0.0105	0.0984
	(0.0273)	(0.0535)	(0.0110)	(0.0322)	(0.0136)	(0.0336)
800	0.0061	0.0699	0.0057	0.0684	0.0058	0.0716
	(0.0056)	(0.0235)	(0.0052)	(0.0200)	(0.0056)	(0.0219)

(a) Case 1: P=0.01, τ=0.05, θ=0.05   (b) Case 2: P=0.01, τ=0.01, θ=0.01
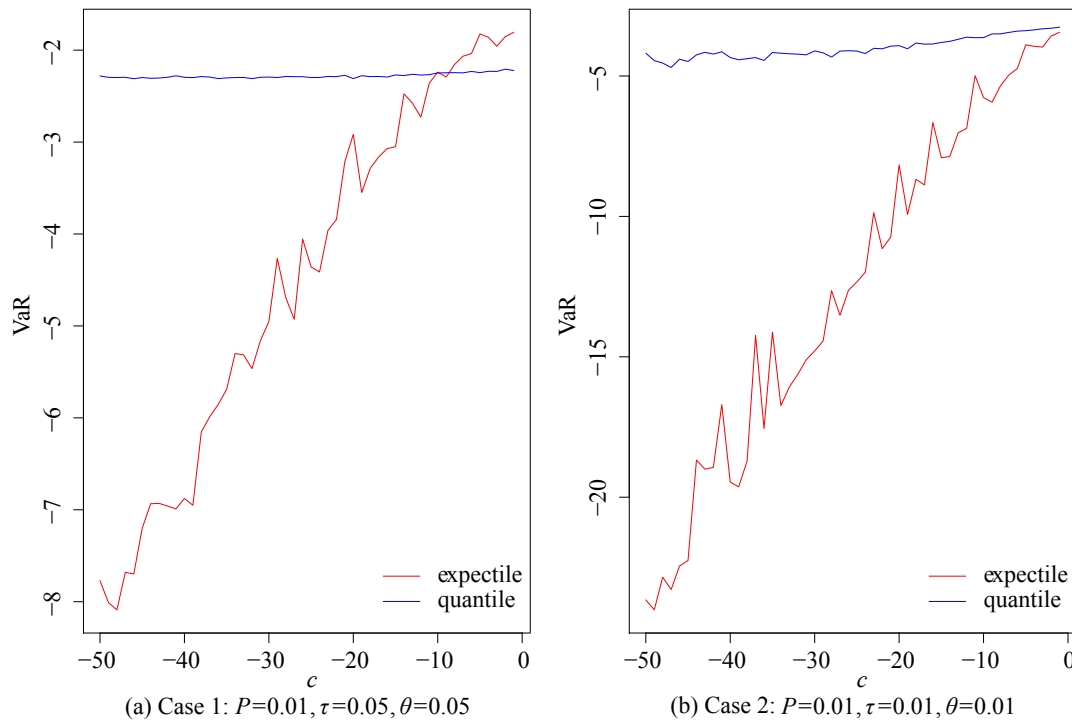
Figure 2. The sensitivity of quantile and expectile to extreme event for Example 2.

3.2. An Empirical Example

To illustrate the practical usefulness of the application of our proposed expectile model, we consider the daily data of S&P500 from January 4, 2010 to December 7, 2017 with 2000 observations in total. The data are downloaded from Yahoo Finance. The daily returns are computed as the difference of the log transformation of the index, multiplying by 100, that is, $Y_t = 100 \log(p_t/p_{t-1})$, where p_t is the daily price. Table 3 reports the summary statistics of the return series. Clearly, one can see from Table 3 that the sample mean is close to zero, but the distribution is slightly negatively skewed, which motivates us to use the expectile model than the quantile model. Figure 3 gives the time series plot for S&P500, and it clearly shows that extreme values mainly occur during 2010–2012, that is, the period of economic recovery from the financial crisis. However, the return series is less volatile from 2012 to 2015.

Table 3. Summary statistics of return series

Mean	Min	Median	Max	S.Dev.	Skew.	Kurt.
0.0425	−6.8958	0.0535	4.6317	0.9320	−0.4621	4.7855
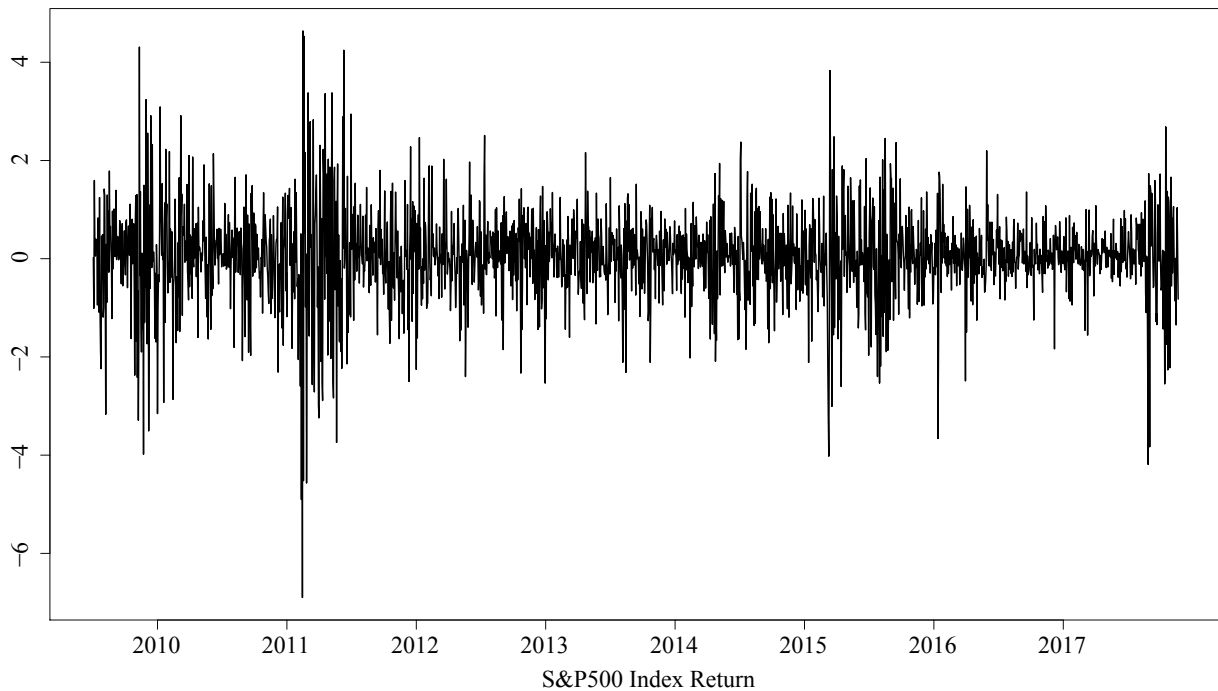
Figure 3. Time series plot of stock return series: S&P500

To model the aforementioned financial data, Kuan et al. (2009) proposed the ABS(2) model and the SQ(2) model, given by

$$e_{t,\tau} = a_{0,\tau} + \delta_{1,\tau}Y_{t-1}^+ + \lambda_{1,\tau}Y_{t-1}^- + \delta_{2,\tau}Y_{t-2}^+ + \lambda_{2,\tau}Y_{t-2}^-,$$

where $v^+ = \max(v,0)$ and $v^- = \max(-v,0)$, and

$$e_{t,\tau} = a_{0,\tau} + a_{1,\tau}Y_{t-1} + b_{1,\tau}(Y_{t-1}^+)^2 + \gamma_{1,\tau}(Y_{t-1}^-)^2 + b_{2,\tau}(Y_{t-2}^+)^2 + \gamma_{2,\tau}(Y_{t-2}^-)^2,$$

Respectively. They have an ability to capture asymmetric properties in the tail risk for financial data, where as Xie et al. (2014) proposed a fully varying coefficient (VC) model to fit the exchange rate data, defined as

$$e_\tau(X_t, U_t) = b_{0,\tau}(U_t) + b_{1,\tau}(U_t)Y_{t-1} + b_{2,\tau}(U_t)Y_{t-2},$$

which, unfortunately, is unable to characterize the asymmetric effect as emphasized in ABS(2) and SQ(2). To capture the asymmetric effects by generalizing the models considered in Kuan et al. (2009) and Xie et al. (2014), the following model is proposed:

$$e_\tau(Y_t, U_t) = b_{0,\tau}(U_t) + b_{1,\tau}(U_t)Y_{t-1}^+ + b_{2,\tau}(U_t)Y_{t-1}^- + b_{3,\tau}(U_t)Y_{t-2}^+ + b_{4,\tau}(U_t)Y_{t-2}^-. \qquad (7)$$

Before estimating the functional coefficients in (7), two issues are addressed. The first question is how to choose U_t. As mentioned in Section 2.2.3, choosing U_t in the above model is of importance in real applications. Unfortunately, Xie et al. (2014) did not provide any theory on how to choose U_t empirically or economically. In this empirical study, due to lack of physical background on how to choose U_t, U_t is selected to be the lagged variable of Y_t, say Y_{t-1} or Y_{t-2}. The optimal choice of U_t is determined based on the data-driven method introduced in Section 2.2.3. From the AAMSE results presented in Table 4, $U_t = Y_{t-1}$ is selected.

Table 4. AAMSE values of choosing the smoothing variable.

	S&P 500			
AAMSE	0.005	0.01	0.05	0.1
$U_t = Y_{t-1}$	0.2492	0.3143*	0.6329*	0.8588*
$U_t = Y_{t-2}$	0.2399*	0.3339	0.6474	0.8610

Note: * denotes that the corresponding AAMSE value is smaller

Table 5. P-values of constancy tests for the VC model in equation (7).

τ	0.005	0.01	0.05	0.1
$b_{0,\tau}$	0.0005	0.0000	0.0000	0.0000
$b_{1,\tau}$	0.0031	0.0000	0.0000	0.0000
$b_{2,\tau}$	0.9410	0.0000	0.0000	0.0000
$b_{3,\tau}$	0.8520	0.2826	0.9693	0.9987
$b_{4,\tau}$	0.0119	0.0000	0.0001	0.0210

The second issue is whether the fully varying coefficient model given in equation (7) is appropriate. To this end, a constancy test is conducted to determine which coefficients are truly varying. Table 5 reports the testing results for all coefficients under four expectile levels, $\tau = 0.005, 0.01, 0.05,$ and 0.1, respectively, and one cannot reject the null hypothesis of constancy for $b_{3,\tau}(\cdot)$ in all cases.

Therefore, given the evidences in Tables 4 and 5, the following partially varying coefficient (PVC) expectile model is investigated:

$$e_\tau(X_t, U_t) = a_{0,\tau}Y_{t-2}^+ + b_{0,\tau}(U_t) + b_{1,\tau}(U_t)Y_{t-1}^+ + + b_{2,\tau}(U_t)Y_{t-1}^- + b_{3,\tau}(U_t)Y_{t-2}^-, \qquad (8)$$

$U_t = Y_{t-1}$ is termed as PVC model hereafter.

Next, Figures 4–5 depict the estimated curves for functional coefficients in the PVC model for $\tau = 0.005$, 0.01, 0.05, and 0.1, respectively. To investigate the asymmetry effects of positive and negative returns, the figures of the estimated functional coefficients in pairs are presented to get a clear insight. For example, one can see clearly that the effects of Y_{t-1}^+ and Y_{t-1}^-, measured by $b_{1,\tau}(Y_{t-1})$ (Figure 4(a)) and $b_{2,\tau}(Y_{t-1})$ (Figure 4(b)), show obvious asymmetric effects. When $Y_{t-1} \in (-1, 0)$, $b_{1,\tau}(Y_{t-1})$ are positive in all case, while $b_{2,\tau}(Y_{t-1})$ are negative under the same circumstance. However, when $Y_{t-1} \in (0, 1)$, the effects of Y_{t-1}^+ and Y_{t-1}^- are nearly symmetric, although the magnitude may not be exactly identical. In Figure 5, it is clear that the constant coefficient, Y_{t-2}^+, and functional coefficient, Y_{t-2}^-, measured by $a_{0,\tau}$ and $b_{3,\tau}(Y_{t-1})$, are both negative, which is consistent with the findings in Kuan et al. (2009), although the time periods are different. It is also worthy to mention that the functional coefficient, $b_{3,\tau}(Y_{t-1})$, achieves its minimal value when Y_{t-1} almost equals to zero.
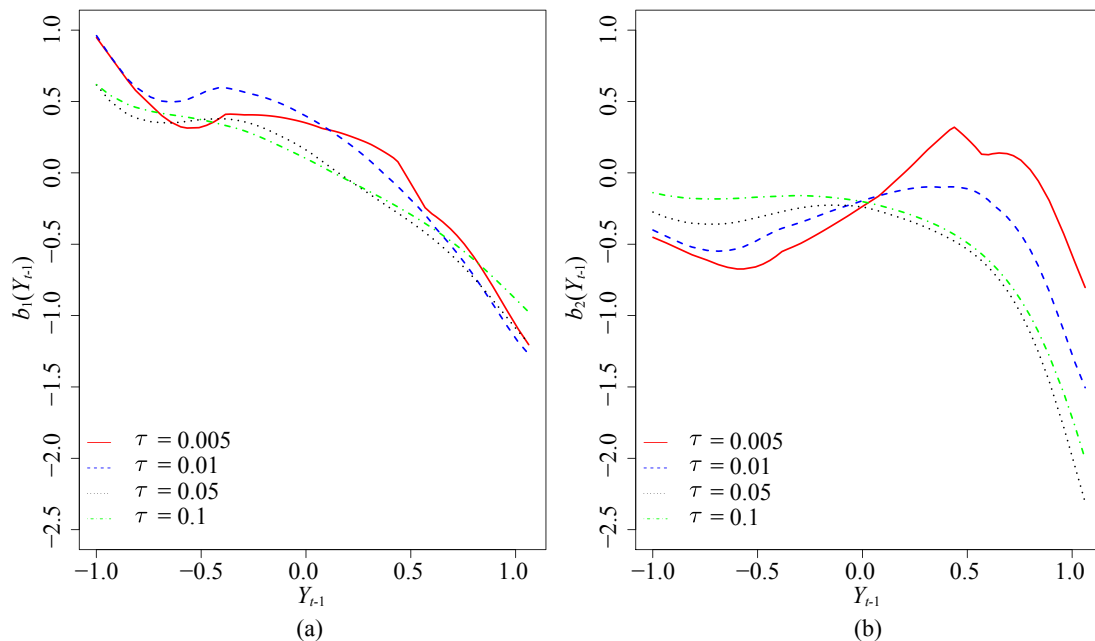
Figure 4. Functional coefficients $b_{1,\tau}(\cdot)$ (a) and $b_{2,\tau}(\cdot)$ (b) for S&P500 in model (8).

Finally, to compare the relative performance of these three models in terms of predictive ability, all models are estimated on rolling windows of length $N = 1500$. As discussed in Campbell (2007) and references therein, when assessing the accuracy of forecasting models for VaR, one needs to consider evaluation procedures other than violation measures. Here, we employ the Murphy diagram introduced in Ehm et al. (2016), which plots the expected scores for competing expectile forecasters. The expected score is calculated using the score function as

$$S(e_\tau, Y) = \frac{1}{n} \sum_{t=1}^{n} S_{\tau,\omega}(e_{\tau,t}, Y_t),$$

where $e_{\tau,t}$ is the one-step expectile forecaster for a rolling sample of $\{Y_{t-1}\}$, and $S_{\tau,\omega}(e_{\tau,t}, Y_t)$ is given by

$$S_{\tau,\omega}(e_{\tau,t}, Y_t) = |I(Y_t < e_{\tau,t}) - \tau|\{(Y_t - \omega)^+ - (e_{\tau,t} - \omega)^+ - (Y_t - e_{\tau,t})I(\omega < e_{\tau,t})\}$$

$$= \begin{cases} (1-\tau)(\omega - Y_t), & \text{if } Y_t \leq \omega < e_{\tau,t}, \\ \tau(Y_t - \omega), & \text{if } e_{\tau,t} \leq \omega < Y_t, \\ 0, & \text{otherwise.} \end{cases}$$
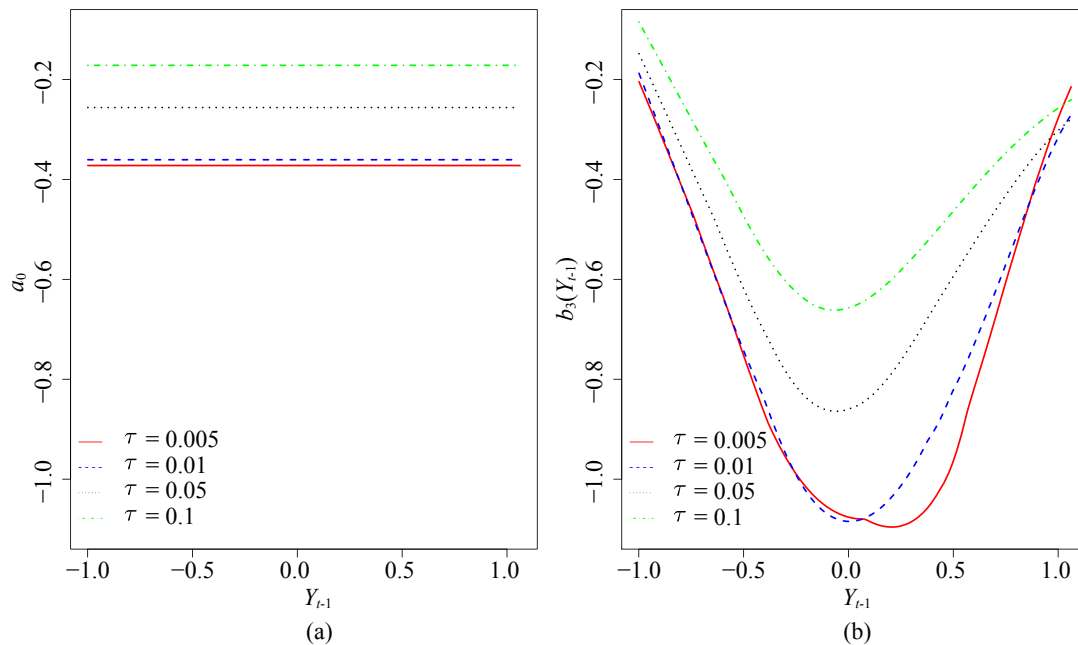

Figure 5. Coefficient $a_{0,\tau}$ **(a)** and functional coefficient $b_{3,\tau}(\cdot)$ **(b)** in model **(8)**.

To estimate the model in equation (8), the normal kernel function is used for local linear estimation. The methods introduced in Section 2.2.2 are employed to choose bandwidths at the third stage. The values of expectile forecasters for three models under each case of $\tau = 0.005, 0.01, 0.05$, and 0.1 are displayed in Figure 6. Moreover, Figure 7 plots the Murphy diagram for the forecasters of the three models under various τs. It demonstrates that the PVC model outperforms the other two models under all cases. The numerical results suggest that our PVC model is a better alternative model to ABS and SQ models for the given dataset.

4. Conclusion

First, a class of dynamic expectile models with partially varying coefficients is proposed and a three-stage estimation procedure is employed to estimate both the constant and varying coefficients. Then, it shows that the constant coefficient estimator has a parametric convergence rate, while the varying coefficient estimator has a nonparametric rate. We also propose weighted average estimators for constant coefficients for further improving estimation efficiency. Moreover, a simple test statistic is derived to testing the constancy of varying coefficients. Our simulation results re-confirm that expectile models are more sensitive to extreme values than quantile models. Using the S&P500 return series, the proposed expectile model with partially varying coefficients outperforms other existing models in most cases. For future works, it is interesting to consider an expectile model including the lag of expectile term, which constitutes an analog of CAViaR models under
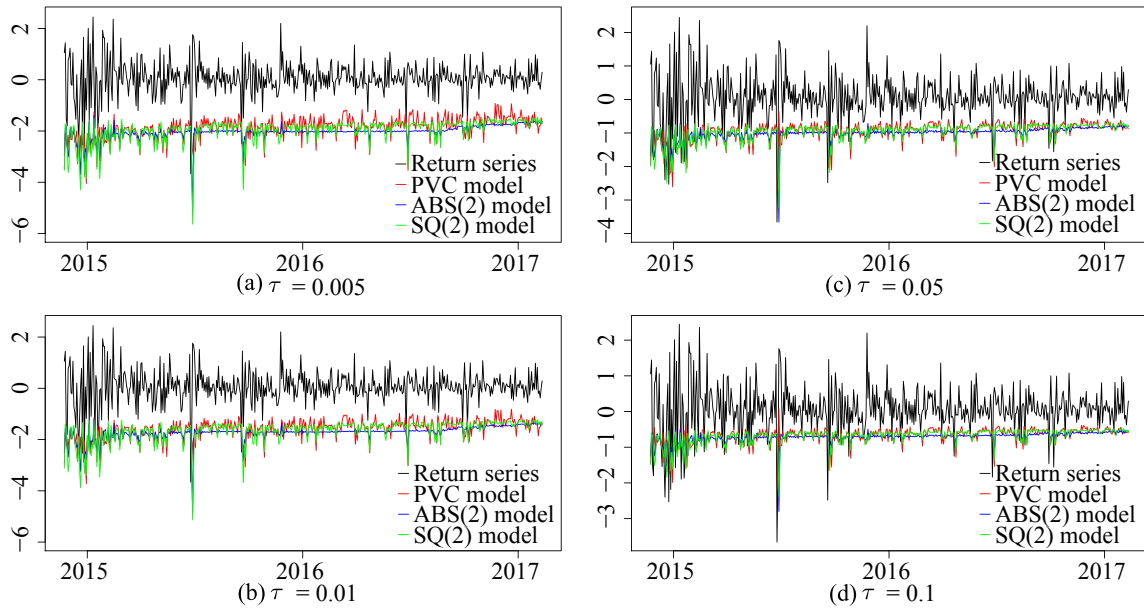
Figure 6. **Expectile forecasters of three models under four cases: (a)** $\tau = 0.005$**, (b)** $\tau = 0.01$**, (c)** $\tau = 0.05$**, (d)** $\tau = 0.1$**.**
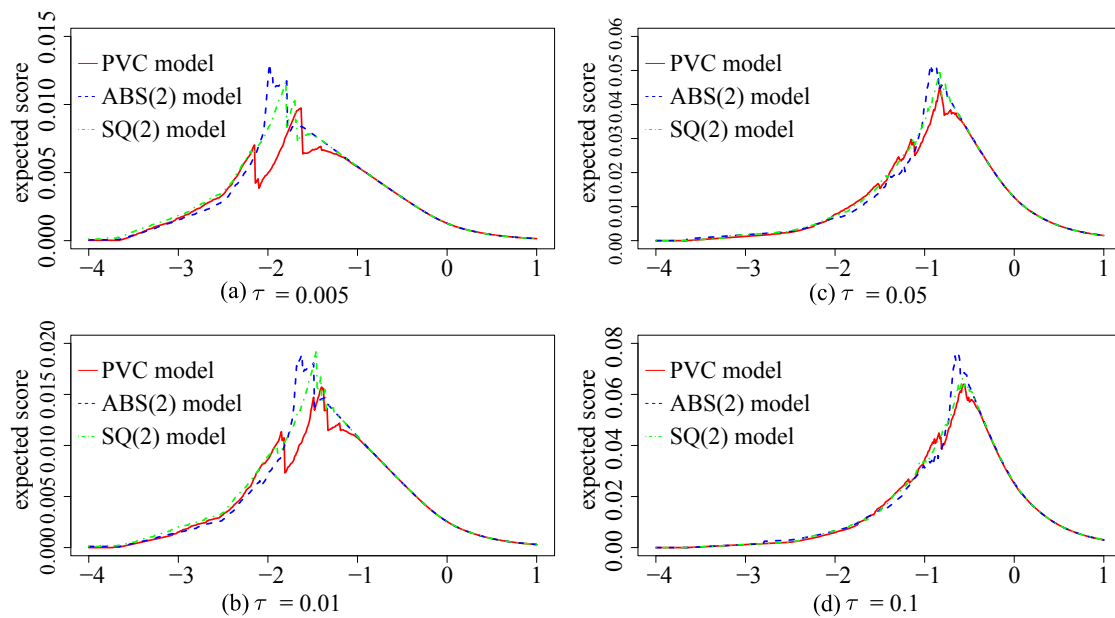

Figure 7. **Murphy diagram for the forecasters of three models under four cases: (a)** $\tau = 0.005$**, (b)** $\tau = 0.01$**, (c)** $\tau = 0.05$**, (d)** $\tau = 0.1$**.**

expectile setting. Moreover, developing a general specification test on varying coefficients based on the proposed expectile models with partially varying coefficients could be of great importance.

Appendix A   Mathematical Proofs

Appendix A.1 Notations and Definitions

In this section, some additional notations and definitions are introduced and used in the following sections. Let $z_t = (U_t, \mathbf{X}_t, Y_t)$, define $\mathbf{S}(z_t) = \mu_0 f_u(U_t)\mathbf{\Gamma}(U_t)$, $\mathrm{M}(z_t) = \mathbf{X}_t$, and $\mathrm{Z}(u_0, z_t) = Q'_\tau(\tilde{Y}_t)\mathrm{M}(z_t)K_{h_1}(U_t - u_0)$, where $\tilde{Y}_t = Y_t - \mathbf{a}^\top(u_0)\mathbf{X}_{1,t} - \mathbf{b}^\top(u_0)\mathbf{X}_{2,t}$. Define

$$\hat{\boldsymbol{\theta}} = \sqrt{nh_1}\{\hat{a}_1(u_0) - a_1(u_0), \ldots, \hat{a}_p(u_0) - a_p(u_0), \; \hat{b}_1(u_0) - b_1(u_0), \ldots, \hat{b}_q(u_0) - b_q(u_0)\}^\top.$$

Then, $\hat{\mathbf{a}}^\top(u_0)\mathbf{X}_{1,t} + \hat{\mathbf{b}}^\top(u_0)\mathbf{X}_{2,t} = \mathbf{a}(u_0)\mathbf{X}_{1,t} + \mathbf{b}(u_0)\mathbf{X}_{2,t} + \hat{\boldsymbol{\theta}}^\top\mathbf{X}_t/\sqrt{nh_1}$ and $\hat{\boldsymbol{\theta}}$ minimizes the following objective function

$$\Psi_n(\boldsymbol{\theta}) \equiv \sum_{t=1}^n [Q_\tau(\tilde{Y}_t - \boldsymbol{\theta}^\top\mathbf{X}_t/\sqrt{nh_1}) - Q_\tau(\tilde{Y}_t)]K((U_t - u_0)/h_1).$$

Proof of Theorem 1.

To establish the asymptotic result of $\tilde{\mathbf{a}}$, the first step is to derive the the local Bahadur representation for the estimators obtained from the first stage. By Lemma A.1, together with the convexity theorem in Pollard (1991), $\hat{\boldsymbol{\theta}}$ can be explicitly expressed as

$$\hat{\boldsymbol{\theta}} = \mathbf{S}^{-1}(u_0)\mathbf{W}_n/\sqrt{nh_1} + o_p(1) \tag{A-1}$$

uniformly for $\boldsymbol{\theta}$ in compact set of $\mathcal{K}_1$, where $\mathbf{W}_n = \sum_{t=1}^n Q'_\tau(\tilde{Y}_t)K(\frac{U_t - u_0}{h_1})\mathbf{X}_t$. It follows from (A-1) that for any u_0 under Assumption A,

$$\hat{\mathbf{a}}(u_0) - \mathbf{a}(u_0) \approx \frac{1}{n}\sum_{t=1}^n \mathbf{e}_1^\top\mathbf{S}^{-1}(u_0)Q'_\tau(\tilde{Y}_t)K((U_t - u_0)/h_1)\mathbf{X}_t = \frac{1}{n}\sum_{t=1}^n \mathbf{e}_1^\top\mathbf{S}^{-1}(u_0)\mathrm{Z}(u_0, z_t).$$

Next, the leave-one-out method is used to obtain the following formula for each point U_s,

$$\hat{\mathbf{a}}(U_s) - \mathbf{a}(U_s) \approx \frac{1}{n}\sum_{t \neq s}^n \mathbf{e}_1^\top\mathbf{S}^{-1}(z_s)\mathrm{Z}(z_s, z_t).$$

Hence,

$$\tilde{\mathbf{a}} - \mathbf{a} = \frac{1}{n}\sum_{t=1}^n [\hat{\mathbf{a}}(U_t) - \mathbf{a}(U_t)] \approx \frac{2}{n^2}\sum_{1 \leq s < t \leq n}^n \mathbf{e}_1^\top\mathbf{S}^{-1}(z_s)\mathrm{Z}(z_s, z_t)$$

$$= \frac{1}{n^2}\sum_{1 \leq s < t \leq n}^n [\mathbf{e}_1^\top\mathbf{S}^{-1}(z_s)\mathrm{Z}(z_s, z_t) + \mathbf{e}_1^\top\mathbf{S}^{-1}(z_t)\mathrm{Z}(z_t, z_s)] = \frac{n-1}{2n}\mathbb{U}_n,$$

where with $h_n(z_s, z_t) = \mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)\mathbf{Z}(z_s, z_t) + \mathbf{e}_1^\top \mathbf{S}^{-1}(z_t)\mathbf{Z}(z_t, z_s)$,

$$\mathbb{U}_n = \frac{2}{n(n-1)} \sum_{1 \le s < t \le n} h_n(z_s, z_t).$$

To derive the asymptotic properties for $\tilde{\mathbf{a}}$, it suffices to show that $\mathbb{U}_n$ is a U-statistics with non-degenerate dependent kernel, $h_n(z_s, z_t)$. Applying a Hoeffding decomposition as in Lee (1990), one has

$$\mathbb{U}_n = \gamma_n + 2H_n^{(1)} + H_n^{(2)}, \tag{A-2}$$

where $H_n^{(1)} = \sum_{t=1}^n h_n^{(1)}(z_t)/n$, $H_n^{(2)} = \sum_{t \ne s} h_n^{(2)}(z_s, z_t)/n(n-1)$, and $\gamma_n = E[h_n(z_s, z_t)]$ with $h_n^{(1)}(v)$ and $h_n^{(2)}(v, w)$ defined by $h_n^{(1)}(v) = E(h_n(v, z_t)) - \gamma_n$, and $h_n^{(2)}(v, w) = h_n(v, w) - E(h_n(v, z_t)) - E(h_n(z_s, w)) + \gamma_n$.

Lemma A.1. *Under Assumptions A, as $n \to \infty$, one has,*

$$\Psi_n(\boldsymbol{\theta}) = \frac{1}{2}\boldsymbol{\theta}^\top \mathbf{S}(u_0)\boldsymbol{\theta} - \frac{1}{\sqrt{nh_1}}\mathbf{W}_n^\top \boldsymbol{\theta} + r_n(\boldsymbol{\theta}),$$

where $\mathbf{S}(u_0) = f_u(u_0)\mu_0\mathbf{\Gamma}(u_0)$ and $\sup_{\boldsymbol{\theta} \in \mathcal{K}_1}|r_n(\boldsymbol{\theta})| = o_p(1)$ for any compact set $\mathcal{K}_1$.

Lemma A.2. *Under Assumptions A,*

$$\begin{aligned}
C_n \equiv \max\Big\{ & \sup_{s \ne t, i \ne j, t \ne j} E|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta}, \sup_{s \ne t, i \ne j, t \ne j} E^{1\otimes}|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta} \\
& \sup_{s \ne t, i \ne j, t \ne j} E^{3\otimes}|h_n(z_s, z_t)h_n(z_s, z_j)|^{1+\delta}, \sup_{s \ne t, i \ne j, t \ne j} E^{2\otimes}|h_n(z_s, z_t)h_n(z_s, z_j)|^{1+\delta} \Big\} \\
& = O(h_1^{-2(1+\delta)}),
\end{aligned}$$

where $E^{1\otimes}, E^{2\otimes}$, and $E^{3\otimes}$ denote the expectations with respect to the measures $P_{z_{s_1}} \otimes P_{z_{s_2}, z_{s_3}, z_{s_4}}$, $P_{z_{s_1}, z_{s_2}} \otimes P_{z_{s_3}}$, and $P_{z_{s_1}, z_{s_2}, z_{s_3}} \otimes P_{z_{s_4}}$ for $s_1 < s_2 < s_3 < s_4$, respectively.

Lemma A.3. *Under Assumptions A, as $n \to \infty$, one has,*
(a) $E|h_n^{(1)}(z_s)|^4 = O(1)$,
(b) $E|h_n^{(2)}(z_s, z_t)|^2 = o(h_1^{-1})$.

Lemma A.4. *Under Assumption A, as $n \to \infty$, then,*
(a) $\gamma_n = B_1^* h_1^2 + o(h_1^2)$,
(b) $n\text{Var}(H_n^{(1)}) = \Sigma_a + o(1)$,
where $\Sigma_a \equiv \Sigma_a^* + 2\sum_{s=1}^{n-1} \text{Cov}(h_n^{(1)}(z_1), h_n^{(1)}(z_{s+1}))$.

Our proof uses Theorem 2 in Dette and Spreckelsen (2004) to establish the asymptotic results of the proposed estimator. It is easy to find that the kernel, $h_n(z_s, z_t)$, satisfies the assumptions of Theorem 2 in Dette and Spreckelsen (2004). Thus, we must check the other conditions, such as (17) and (18), in Dette and

Spreckelsen (2004). The condition in (17) in Dette and Spreckelsen (2004) is checked and proved in Lemma A.2, and the proof of the condition in (18) in Dette and Spreckelsen (2004) is given in Lemma A.3. Then, one has

$$\frac{\mathbb{U}_n - E^{\otimes}(\mathbb{U}_n)}{\sqrt{\mathrm{Var}(\mathbb{U}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1),$$

which implies that

$$\frac{\mathbb{U}_n - \gamma_n}{\sqrt{\mathrm{Var}(H_n^{(1)}(z_t))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

The asymptotic normality follows from Lemma A.4 and equation (A-2) that

$$\sqrt{n}\left[\tilde{\mathbf{a}} - \mathbf{a} - \frac{1}{2}\gamma_n\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Sigma}_a),$$

which completes the proof of Theorem 1.

Proof of Theorem 2.

To simplify notation, define

$$\hat{\boldsymbol{\vartheta}} = \sqrt{nh_2}\{\hat{b}_1(u_0) - b_1(u_0), \ldots, \hat{b}_q(u_0) - b_q(u_0), h_2(\hat{b}_1'(u_0) - b_1'(u_0)), \ldots, h_2(\hat{b}_q'(u_0) - b_q'(u_0))\}^{\top},$$

which minimizes the following function

$$\Phi_n(\boldsymbol{\vartheta}) \equiv \Phi_n(\boldsymbol{\vartheta}; \tau, \mathbf{X}, U, u_0) = \sum_{t=1}^{n}\{Q_\tau(\tilde{Y}_t^* - \boldsymbol{\vartheta}^{\top}\mathbf{Z}_{t,2}^*/\sqrt{nh_2}) - Q_\tau(\tilde{Y}_t^*)\}K(\frac{U_t - u_0}{h_2}), \qquad \text{(A-3)}$$

where $\tilde{Y}_t^* = Y_{t1}^* - \boldsymbol{\beta}_\tau^{\top}(u_0)\mathbf{Z}_{t,2}$ with $\boldsymbol{\beta}_\tau(u) = (\mathbf{b}^{\top}(u), \mathbf{b}'^{\top}(u))^{\top}$, $\mathbf{Z}_{t,2} = (\mathbf{X}_{t,2}^{\top}, \mathbf{X}_{t,2}^{\top}(U_t - u_0))^{\top}$, and $\mathbf{Z}_{t,2}^* = (\mathbf{X}_{t,2}^{\top}, \mathbf{X}_{t,2}^{\top}(U_t - u_0)/h_2)^{\top}$. Note that τ is dropped from $\boldsymbol{\beta}_\tau(u)$ afterwards. Then, the following two lemmas are provided to establish the asymptotic properties of $\hat{\mathbf{b}}(u_0)$.

Lemma A.5. *Under Assumptions A and B, as $n \to \infty$, the following results hold true:*
(a) $\Phi_n(\boldsymbol{\vartheta}) = \frac{1}{2}\boldsymbol{\vartheta}^{\top}\mathbf{D}(u_0)\boldsymbol{\vartheta} - \frac{1}{\sqrt{nh_2}}\mathbf{G}_n^{\top}\boldsymbol{\vartheta} + R_n(\boldsymbol{\vartheta}),$
(b) $R_n^*(\boldsymbol{\vartheta}) = o_p(1),$
(c) $\sup_{\boldsymbol{\vartheta}\in\mathcal{K}_2}|R_n(\boldsymbol{\vartheta})| = o_p(1),$
where $\mathbf{G}_n = \sum_{t=1}^{n} Q_\tau'(\tilde{Y}_t^*)K(\frac{U_t - u_0}{h_2})\mathbf{Z}_{t,2}^*.$

Lemma A.6. *Under Assumptions A and B, as $n \to \infty$, one has*

$$\frac{1}{\sqrt{nh_2}}\left[\mathbf{G}_n - \frac{nh_2^3}{2}f_u(u_0)\begin{pmatrix}\boldsymbol{\Gamma}_2(u_0)\mathbf{b}''(u_0)\mu_2 \\ 0\end{pmatrix} + o(nh_2^3)\right] \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}(u_0)).$$

From the convexity lemma of Pollard (1991) and Lemma A.5, the minimizer $\hat{\boldsymbol{\vartheta}}$ can be expressed as

$$\hat{\boldsymbol{\vartheta}} = \mathbf{D}^{-1}(u_0)\mathbf{G}_n/\sqrt{nh_2} + o_p(1)$$

uniformly for $\boldsymbol{\vartheta} \in \mathcal{K}_2$, which is a compact set of $\boldsymbol{\vartheta}$. From the above equation, we have

$$\sqrt{nh_2}\mathbf{H}(\hat{\boldsymbol{\beta}}(u_0) - \boldsymbol{\beta}(u_0)) = \mathbf{D}^{-1}(u_0)\mathbf{G}_n / \sqrt{nh_2} + o_p(1),$$

where $\mathbf{H} = \mathbf{I}_q \otimes \mathrm{diag}(1, h_2)$ is the selection matrix. Together with Lemma A.6, Theorem 2 is proved.

Appendix A.2 Proofs of Lemmas

Proof of Lemma A.1.

It follows from the same procedure as that used in the proof of Lemma A.5.

Proof of Lemma A.2.

It is easy to find that the kernel, $h_n(z_s, z_t)$, satisfies the assumptions of Theorem 2 in Dette and Spreckelsen (2004). Thus, the remaining condition needs to be checked, namemly, the condition in (17) of Dette and Spreckelsen (2004). To this end, κ is chosen to satisfy $1/\kappa + 1/\iota = 1$, where $1 < \iota < 2/(1+\delta)$. It follows from the Hölder's inequality that

$$E|h_n(z_s, z_t)h_n(z_s, z_t)|^{1+\delta} \leq [E|h_n(z_s, z_t)|^{\kappa(1+\delta)}]^{\frac{1}{\kappa}}[E|h_n(z_s, z_t)|^{\iota(1+\delta)}]^{\frac{1}{\iota}}.$$

By the C_r-inequality, one obtains

$$
\begin{aligned}
E|h_n(z_s, z_t)|^{\kappa(1+\delta)} &= E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)Z(z_s, z_t) + \mathbf{e}_1^\top \mathbf{S}^{-1}(z_t)Z(z_t, z_s)|^{\kappa(1+\delta)} \\
&\leq C\{E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)Z(z_s, z_t)|^{\kappa(1+\delta)} + E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_t)Z(z_t, z_s)|^{\kappa(1+\delta)}\} \\
&\leq CE|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)Z(z_s, z_t)|^{\kappa(1+\delta)} \\
&= CE|\mathbf{e}_1^\top \mathbf{S}^{-1}(U_s)\varphi(z_s, z_t)K(\frac{U_t - U_s}{h_1})\mathbf{Z}_t^{*\top}|^{\kappa(1+\delta)} = O(h_1^{-\kappa(1+\delta)}).
\end{aligned}
$$

In a similar way, it follows that $E|h_n(z_s, z_t)|^{\iota(1+\delta)} = O(h_1^{-\iota(1+\delta)})$. Then,

$$\sup_{s\neq t, i\neq j, t\neq j} E|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta} = O(h_1^{-2(1+\delta)})$$

can be easily shown. By the same token, one can show that

$$\sup_{s\neq t, i\neq j, t\neq j} E^{1\otimes}|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta} = O(h_1^{-2(1+\delta)}),$$

$$\sup_{s\neq t, s\neq j, t\neq j} E^{3\otimes}|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta} = O(h_1^{-2(1+\delta)}),$$

and

$$\sup_{s\neq t, i\neq j, t\neq j} E^{2\otimes}|h_n(z_s, z_t)h_n(z_s, z_j)|^{1+\delta} = O(h_1^{-2(1+\delta)}).$$

Therefore,

$$
\begin{aligned}
C_n = \max\{ &\sup_{s\neq t, i\neq j, t\neq j} E|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta}, \sup_{s\neq t, i\neq j, t\neq j} E^{1\otimes}|h_n(z_s, z_t)h_n(z_i, z_j)|^{1+\delta} \\
&\sup_{s\neq t, i\neq j, t\neq j} E^{2\otimes}|h_n(z_s, z_t)h_n(z_s, z_j)|^{1+\delta}, \sup_{s\neq t, i\neq j, t\neq j} E^{2\otimes}|h_n(z_s, z_t)h_n(z_s, z_j)|^{1+\delta}\} \\
= &\, O(h_1^{-2(1+\delta)}),
\end{aligned}
$$

so that the condition in (17) of Theorem 2 in Dette and Spreckelsen (2004) is satisfied.

Proof of Lemma A.3.

Note that $E[\mathbf{X}_t \varphi(z_t, z_t) K_{h_1}(U_t - u_0)] = 0$ according to the first order condition. Then,

$$
\begin{aligned}
&E[Z(u_0, z_t)] \\
&= E[\mathbf{X}_t Q_\tau'(Y_t - e_\tau(U_t, \mathbf{X}_t) + \mathbf{X}_{t,2}^\top[\mathbf{b}(U_t) - \mathbf{b}(u_0)]) K_{h_1}(U_t - u_0)] \\
&= E\Big[\boldsymbol{\Gamma}(U_t) K_{h_1}(U_t - u_0) \begin{pmatrix} 0 \\ \mathbf{b}'(u_0)(U_t - u_0) + \frac{1}{2}\mathbf{b}''(u_0)(U_t - u_0)^2 \end{pmatrix}\Big](1 + o(1)) \\
&= \Big\{ E\Big[\boldsymbol{\Gamma}(U_t) \begin{pmatrix} 0 \\ \mathbf{b}'(u_0)(U_t - u_0) \end{pmatrix} K_{h_1}(U_t - u_0)\Big] \\
&\quad + E\Big[\boldsymbol{\Gamma}(U_t) \begin{pmatrix} 0 \\ \frac{1}{2}\mathbf{b}''(u_0)(U_t - u_0)^2 \end{pmatrix} K_{h_1}(U_t - u_0)\Big]\Big\}(1 + o(1)).
\end{aligned}
\tag{A-4}
$$

For the first term on the right hand side of (A-4), one can obtain

$$
\boldsymbol{\Gamma}(u_0 + uh) = \boldsymbol{\Gamma}(u_0) + \boldsymbol{\Gamma}'(u_0)uh + o(h) \quad and \quad f_u(u_0 + uh) = f_u(u_0) + f_u'(u_0)uh + o(h)
$$

by Taylor expansion. Thus,

$$
\begin{aligned}
&E[\boldsymbol{\Gamma}(U_t)\mathbf{b}'(u_0)(U_t - u_0) K_{h_1}(U_t - u_0)] \\
&= \int \boldsymbol{\Gamma}(u_0 + uh)(U_t - u_0)K(u)f_u(u_0 + uh)\mathbf{b}'(u_0)\, \mathrm{d}u \\
&= h_1 \int \big[\boldsymbol{\Gamma}(u_0)f_u(u_0) + \boldsymbol{\Gamma}'(u_0)f_u(u_0)uh_1 + \boldsymbol{\Gamma}(u_0)f_u'(u_0)uh_1\big]\mathbf{b}'(u_0)uK(u)\, \mathrm{d}u(1 + o(1)) \\
&= \mu_2 h_1^2[\boldsymbol{\Gamma}'(u_0)f_u(u_0) + \boldsymbol{\Gamma}(u_0)f_u'(u_0)]\mathbf{b}'(u_0).
\end{aligned}
$$

For the second term on the right hand side of (A-4), one has

$$
E[\boldsymbol{\Gamma}(U_t)\frac{1}{2}\mathbf{b}''(u_0)(U_t - u_0)^2 K_{h_1}(U_t - u_0)] = \frac{\mu_2 h_1^2}{2}\boldsymbol{\Gamma}(u_0)f_u(u_0)\mathbf{b}''(u_0)(1 + o(1)).
$$

Therefore,

$$
E[Z(u_0, z_t)] = \frac{\mu_2 h_1^2}{2}\Big[2\big(\boldsymbol{\Gamma}'(u_0)f_u(u_0) + \boldsymbol{\Gamma}(u_0)f_u'(u_0)\big)\begin{pmatrix} 0 \\ \mathbf{b}'(u_0) \end{pmatrix} + \boldsymbol{\Gamma}(u_0)f_u(u_0)\begin{pmatrix} 0 \\ \mathbf{b}''(u_0) \end{pmatrix}\Big](1 + o(1)).
$$

It follows from the definition of $h_n(z_s, z_t)$ and (A-2) that

$$
\begin{aligned}
E[h_n(v, z_t)] &= E[\mathbf{e}_1^\top \mathbf{S}^{-1}(v) Z(v, z_t)] + E[\mathbf{e}_1^\top \mathbf{S}^{-1}(z_t) Z(z_t, v)] \\
&= E[\mathbf{e}_1^\top \mathbf{S}^{-1}(z_t) Z(z_t, v)] + o(h_1^2) \\
&= E[\mathbf{e}_1^\top \mathbf{S}^{-1}(z_t) \varphi(z_t, v) \mathbf{M}(v) K(\frac{v - U_t}{h_1})] + o(h_1^2) \\
&= \mathbf{e}_1^\top \mathbf{S}^{-1}(v) \varphi(v, v) \mathbf{M}(v) f_u(v) + o(h_1).
\end{aligned}
$$

Then, it is readily seen that

$$
h_n^{(1)}(v) = \mathbf{e}_1^\top \mathbf{S}^{-1}(v) \varphi_\tau(v, v) \mathbf{M}(v) f_u(v) + o(h_1),
$$

where $f_u(\cdot)$ is the density function of U_t, and

$$
h_n^{(2)}(v, w) = h_n(v, w) - \mathbf{e}_1^\top \mathbf{S}^{-1}(v) \varphi_\tau(v, v) \mathbf{M}(v) f_u(v) - \mathbf{e}_1^\top \mathbf{S}^{-1}(w) \varphi_\tau(w, w) \mathbf{M}(w) f_u(w) + o(1).
$$

Therefore,

$$
\begin{aligned}
E|h_n^{(1)}(z_s)|^4 &= \frac{1}{\mu_0^4} E|\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_s) \varphi(z_s, z_s) \mathbf{X}_s|^4 + o(h_1^2) \\
&\leq C E|\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_s) \mathbf{X}_s \mathbf{X}_s^\top \mathbf{\Gamma}^{-1}(U_s) \mathbf{e}_1|^2 \leq C,
\end{aligned}
$$

and

$$
\begin{aligned}
E|h_n(z_s, z_t)|^2 &= E[\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s) Z(z_s, z_t) + \mathbf{e}_1^\top \mathbf{S}^{-1}(z_t) Z(z_t, z_s)]^2 \\
&\leq C E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s) Z(z_s, z_t)|^2 \\
&\leq C E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s) Z(z_s, z_t) Z^\top(z_s, z_t) \mathbf{S}^{-1}(z_s) \mathbf{e}_1| \\
&\leq C \mathbf{e}_1^\top E[E\{\mathbf{S}^{-1}(z_s) Z(z_s, z_t) Z^\top(z_s, z_t) \mathbf{S}^{-1}(z_s)\}] \mathbf{e}_1 \\
&= C \mathbf{e}_1^\top E[\mathbf{S}^{-1}(z_s) \int \varphi^2(z_s, z_t) K_{h_1}^2(U_t - U_s) \mathbf{X}_t \mathbf{X}_t^\top dF(z_t) \mathbf{S}^{-1}(z_s)] \mathbf{e}_1 = O(h_1^{-1}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E|h_n^{(2)}(z_s, z_t)|^2 &= C E|h_n(z_s, z_t) - e_1^\top \mathbf{S}^{-1}(z_s) \varphi_\tau(z_s, z_s) \mathbf{M}(z_s) f_u(z_s) \\
&\quad - e_1^\top \mathbf{S}^{-1}(z_t) \varphi_\tau(z_t, z_t) \mathbf{M}(z_t) f(z_t)|^2 + o(1) \\
&= C\{E|h_n(z_s, z_t)|^2 + E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s) \varphi_\tau(z_s, z_s) \mathbf{M}(z_s) f_u(z_s)|^2 \\
&\quad + E|\mathbf{e}_1^\top \mathbf{S}^{-1}(z_t) \varphi_\tau(z_t, z_t) \mathbf{M}(z_t) f(z_t)|^2\} + o(1) \\
&= C\{E|h_n(z_s, z_t)|^2 + \frac{1}{\mu_0^2} E|\mathbf{e}_1^\top (\mathbf{\Gamma}^{-1}(U_s)) \varphi_\tau(z_s, z_s) \mathbf{X}_s|^2 \\
&\quad + \frac{1}{\mu_0^2} E|\mathbf{e}_1^\top (\mathbf{\Gamma}^{-1}(U_t)) \varphi_\tau(z_t, z_t) \mathbf{X}_s|^2\} + o(1) \\
&= C E|h_n(z_s, z_t)|^2 + C_1 = O(h_1^{-1}).
\end{aligned}
$$

Clearly, Lemma A.3 is established.

Proof of Lemma A.4.

It is easy see from Lemma A.3 and (A-2) that

$$
\begin{aligned}
\gamma_n &= \int\int h_n(z_s, z_t)\mathrm{d}F(z_s)\mathrm{d}F(z_t)\\
&= \int\int [\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)Z(z_s, z_t) + \mathbf{e}_1^\top \mathbf{S}^{-1}(z_t)Z(z_t, z_s)]\mathrm{d}F(z_s)\mathrm{d}F(z_t)\\
&= 2\int\int \mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)Z(z_s, z_t)\mathrm{d}F(z_s)\mathrm{d}F(z_t)\\
&= \frac{\mu_2 h_1^2}{\mu_0}\mathbf{e}_1^\top E\left[2\Big(\mathbf{\Gamma}^{-1}(U_s)\mathbf{\Gamma}'(U_s) + f_u'(U_s)/f_u(U_s)\Big)\begin{pmatrix}0\\\mathbf{b}'(U_s)\end{pmatrix} + \begin{pmatrix}0\\\mathbf{b}''(U_s)\end{pmatrix}\right](1 + o(1))\\
&= B_1^* h_1^2 + o(h_1^2),
\end{aligned}
$$

which completes the proof of (a). For (b), as $E[h_n^{(1)}(z_s)] = 0$ holds, it is easy to show that

$$
\begin{aligned}
\mathrm{Var}(h_n^{(1)}(z_s)) &= E[\mathbf{e}_1^\top \mathbf{S}^{-1}(z_s)\varphi(z_s, z_s)\mathrm{M}(z_s)f(z_s)]^2 + o(h_1^2)\\
&= \frac{1}{\mu_0^2}E[\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_s)\mathbf{X}_s\mathbf{X}_s^\top \mathbf{\Gamma}^{-1}(U_s)\mathrm{Q}_\tau'^2(Y_s - \mathbf{e}_\tau(U_s, \mathbf{X}_s))\mathbf{e}_1] + o(h_1^2)\\
&= \frac{1}{\mu_0^2}E[\mathbf{e}_1^\top \mathbf{\Gamma}^{-1}(U_s)E(\mathrm{Q}_\tau'^2(Y_s - \mathbf{e}_\tau(U_s, \mathbf{X}_s))\mathbf{X}_s\mathbf{X}_s^\top)\mathbf{\Gamma}^{-1}(U_s)\mathbf{e}_1] + o(h_1^2)\\
&= \mathbf{\Sigma}_a^* + o(h_1^2),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Cov}(h_n^{(1)}(z_1), h_n^{(1)}(z_{s+1})) &= E[h_n^{(1)}(z_1)h_n^{(1)}(z_{s+1})]\\
&= \frac{1}{\mu_0^2}E[\mathbf{e}_1^\top (\mathbf{\Gamma}^{-1}(U_1)\mathbf{X}_1\mathbf{X}_{s+1}^\top \mathbf{\Gamma}^{-1}(U_{s+1})\varphi(z_1, z_1)\varphi(z_{s+1}, z_{s+1})\mathbf{e}_1] + o(h_1^2)\\
&= \mathrm{Cov}(w_1, w_{s+1}) + o(1) \le C\beta(s).
\end{aligned}
$$

Using the above results together with properties of stationarity, one obtains

$$
\begin{aligned}
n\mathrm{Var}(H_n^{(1)}) &= \frac{1}{n}\sum_{s=1}^n \mathrm{Var}(h_n^{(1)}(z_s)) + 2\sum_{s=1}^{n-1}(1 - \frac{s}{n})\mathrm{Cov}(h_n^{(1)}(z_1), h_n^{(1)}(z_{s+1}))\\
&= \mathbf{\Sigma}_a^* + 2\sum_{s=1}^{n-1}\mathrm{Cov}(h_n^{(1)}(z_1), h_n^{(1)}(z_{s+1})) + o(1) = \mathbf{\Sigma}_a + o(1),
\end{aligned}
$$

and Lemma A.4 holds.

Proof of Lemma A.5.

Write $\eta(U_t, \mathbf{X}_{t,2}) = \mathbf{b}^\top(U_t)\mathbf{X}_{t,2}$. Applying Taylor expansion leads to

$$\eta(U_t, \mathbf{X}_{t,2}) = \eta(u_0, U_t, \mathbf{X}_{t,2}) + \frac{1}{2}\sum_{j=1}^q b_j''(u_0)X_{tj,2}(U_t - u_0)^2 + o(h_2^2),$$

for u in $|u - u_0| < h_2$, in which $X_{tj,2}$ is the j-th element of $\mathbf{X}_{t,2}$. Let $\phi(v|u, \mathbf{x}) = E[Q_\tau(Y_{t1}^* - \eta(U_t, \mathbf{X}_{t,2}) + v)|U_t = u, \mathbf{X}_{t,2} = \mathbf{x}]$. Denote $\partial\phi(v|u,\mathbf{x})/\partial v$ and $\partial^2\phi(v|u,\mathbf{x})/\partial v^2$ by $\phi'(v|u,\mathbf{x})$ and $\phi''(v|u,\mathbf{x})$, respectively. It is worth mentioning that $\Phi_n(\boldsymbol{\vartheta})$ is also convex in $\boldsymbol{\vartheta}$ and it can be re-written as

$$\Phi_n(\boldsymbol{\vartheta}) = E[\Phi_n(\boldsymbol{\vartheta})|U_t, \mathbf{X}_{t,2}] - \frac{1}{\sqrt{nh_2}}\sum_{t=1}^n \left\{ Q_\tau'(\tilde{Y}_t^*)\mathbf{Z}_{t,2}^* K(\frac{U_t - u_0}{h_2}) \right.$$
$$\left. - E[Q_\tau'(\tilde{Y}_t^*)|U_t, \mathbf{X}_{t,2}]\mathbf{Z}_{t,2}^* K(\frac{U_t - u_0}{h_2}) \right\}^\top \boldsymbol{\vartheta} + R_n^*(\boldsymbol{\vartheta}). \tag{A-5}$$

Let us deal with the first term on right hand side of (A-5). For this purpose, it follows from equation (A-3) that

$$E[\Phi_n(\boldsymbol{\vartheta})|U_t, \mathbf{X}_{t,2}] \tag{A-6}$$
$$= \sum_{t=1}^n \left[\phi\left(\eta(U_t, \mathbf{X}_{t,2}) - \eta(u_0, U_t, \mathbf{X}_{t,2}) - \frac{\boldsymbol{\vartheta}^\top \mathbf{Z}_{t,2}^*}{\sqrt{nh_2}} \Big| U_t, \mathbf{X}_{t,2}\right) \right.$$
$$\left. - \phi\left(\eta(U_t, \mathbf{X}_{t,2}) - \eta(u_0, U_t, \mathbf{X}_{t,2}) | U_t, \mathbf{X}_{t,2}\right) \right] K(\frac{U_t - u_0}{h_2})$$
$$= -\sum_{t=1}^n \phi'\left(\eta(U_t, \mathbf{X}_{t,2}) - \eta(u_0, U_t, \mathbf{X}_{t,2}) | U_t, \mathbf{X}_{t,2}\right) \frac{\boldsymbol{\vartheta}^\top \mathbf{Z}_{t,2}^*}{\sqrt{nh_2}} K(\frac{U_t - u_0}{h_2})$$
$$+ \frac{1}{2}\sum_{t=1}^n \phi''\left(\eta(U_t, \mathbf{X}_{t,2}) - \eta(u_0, U_t, \mathbf{X}_{t,2}) | U_t, \mathbf{X}_{t,2}\right) (\frac{\boldsymbol{\vartheta}^\top \mathbf{Z}_{t,2}^*}{\sqrt{nh_2}})^2 K(\frac{U_t - u_0}{h_2})(1 + o_p(1))$$
$$= -\frac{1}{\sqrt{nh_2}}\sum_{t=1}^n E[Q_\tau'(\tilde{Y}_t^*)|U_t, \mathbf{X}_{t,2})]\mathbf{Z}_{t,2}^{*\top} K(\frac{U_t - u_0}{h_2})\boldsymbol{\vartheta}$$
$$+ \frac{1}{2nh_2}\boldsymbol{\vartheta}^\top \left\{ \sum_{t=1}^n K(\frac{U_t - u_0}{h_2}) E[Q_\tau''(Y_t - e_\tau(U_t, \mathbf{X}_t))|U_t, \mathbf{X}_{t,2}]\mathbf{Z}_{t,2}^* \mathbf{Z}_{t,2}^{*\top} \right\} \boldsymbol{\vartheta}(1 + o_p(1)).$$

For the second term on the right hand side of (A-5),

$$E\left[K(\frac{U_t - u_0}{h_2}) E[Q_\tau''(Y_t - e_\tau(U_t, \mathbf{X}_t))|U_t, \mathbf{X}_{t,2}]\mathbf{Z}_{t,2}^* \mathbf{Z}_{t,2}^{*\top} \right]$$
$$= 2E\left[K(\frac{U_t - u_0}{h_2})\{\tau[1 - F_{y|x,u}(e_\tau(U_t, \mathbf{X}_t))] + (1 - \tau)F_{y|x,u}(e_\tau(U_t, \mathbf{X}_t))\}\mathbf{Z}_{t,2}^* \mathbf{Z}_{t,2}^{*\top} \right]$$
$$= h_2 f_u(u_0) \begin{pmatrix} \mu_0 & 0 \\ 0 & \mu_2 \end{pmatrix} \otimes \boldsymbol{\Gamma}_2(u_0)(1 + o_p(1)).$$

Then, by ergodicity, one has

$$\frac{1}{nh_2}\sum_{t=1}^{n}K(\frac{U_t-u_0}{h_2})Q''_\tau(Y_t-e_\tau(U_t,\mathbf{X}_t))\mathbf{Z}^*_{t,2}\mathbf{Z}^{*\top}_{t,2}\to f_u(u_0)\begin{pmatrix}\mu_0 & 0\\ 0 & \mu_2\end{pmatrix}\otimes\mathbf{\Gamma}_2(u_0). \tag{A-7}$$

Using Lemma A.6(b) (see below), together with (A-5), (A-6), and (A-7), (a) holds true, where $R_n(\boldsymbol{\vartheta})=o_p(1)$ for each fixed $\boldsymbol{\vartheta}$. To prove (b), note that

$$R_n^*(\vartheta)=\sum_{t=1}^{n}\{\mathbf{V}_t-E(\mathbf{V}_t|U_t,\mathbf{X}_{t,2})\},$$

where

$$\mathbf{V}_t=\sum_{t=1}^{n}\{Q_\tau(\tilde{Y}_t^*-\boldsymbol{\vartheta}^\top\mathbf{Z}^*_{t,2}/\sqrt{nh_2})-Q_\tau(\tilde{Y}_t^*)+Q'_\tau(\tilde{Y}_t^*)\boldsymbol{\vartheta}^\top\mathbf{Z}^*_{t,2}/\sqrt{nh_2}\}K(\frac{U_t-u_0}{h_2}).$$

Since $ER_n^*(\boldsymbol{\vartheta})=0$,

$$ER_n^{*2}(\boldsymbol{\vartheta})=nE\mathbf{V}_t^2+2\sum_{s=1}^{n-1}(n-s)\mathrm{Cov}(\mathbf{V}_1,\mathbf{V}_{s+1}),$$

and by Lemma A.2 in Yao and Tong (1996),

$$E\mathbf{V}_t^2\le 16E[(\boldsymbol{\vartheta}^\top\mathbf{Z}^*_{t,2}/\sqrt{nh_2})^4K^2(\frac{U_t-u_0}{h_2})]=O(\frac{1}{n^2h_2}).$$

Let $d_n\to\infty$ be a sequence of positive integers, such that $d_nh_2(n)\to0$, and define

$$J_1=\sum_{s=1}^{d_n-1}n|\mathrm{Cov}(\mathbf{V}_1,\mathbf{V}_{s+1})|\quad\text{and}\quad J_2=\sum_{s=d_n}^{n-1}n|\mathrm{Cov}(\mathbf{V}_1,\mathbf{V}_{s+1})|.$$

By Cauchy–Schwartz inequality and stationarity, for $s<d_n$,

$$|\mathrm{Cov}(\mathbf{V}_1,\mathbf{V}_{s+1})|\le CE\mathbf{V}_t^2=O(\frac{1}{n^2h_2}),$$

so that

$$J_1=nd_nO(\frac{1}{n^2h_2})=o(\frac{1}{nh_2^2}).$$

Next, the upper bound of J_2 is derived. Using Davydov's inequality (See Corollary A.2 in Hall and Heyde (1980)), it is easy to obtain

$$|\mathrm{Cov}(\mathbf{V}_1,\mathbf{V}_{s+1})|\le C[\beta(s)]^{1-2/\delta}[E|\mathbf{V}_t|^\delta]^{2/\delta}.$$

Using Lemma A.2 in Yao and Tong (1996) again, one has

$$E|\mathbf{V}_t|^\delta\le 4^\delta E[(\frac{\boldsymbol{\vartheta}^\top Z^*_{t,2}}{\sqrt{nh_2}})^{2\delta}K^\delta(\frac{U_t-u_0}{h_2})]$$
$$\le C\frac{h_2^{\delta-1}}{(nh_2)^\delta}E[(\boldsymbol{\vartheta}^\top\mathbf{Z}^*_{t,2})^{2\delta}K(\frac{U_t-u_0}{h_2})]$$

$$\leq C\frac{h_2}{(nh_2)^{\delta}} = Cn^{-\delta}h_2^{1-\delta},$$

which implies that

$$J_2 \leq nCn^{-2}h_2^{\frac{2}{\delta}-2} \sum_{s=d_n}^{\infty} \beta^{1-2/\delta}(s)$$

$$\leq Cn^{-1}h_2^{\frac{2}{\delta}-2}d_n^{-c} \sum_{s=d_n}^{\infty} s^c \beta^{1-2/\delta}(s) = o(\frac{1}{nh_2}),$$

by choosing d_n, such that $h_2^{1-2/\delta}d_n^c = O(1)$ for $\delta > 2$, so that $d_n h_2 \to 0$ is satisfied. Consequently, $ER_n^{*2}(\theta) = o(\frac{1}{nh_2^2})$ and

$$P(|R_n^*(\theta)| > \epsilon) \leq \frac{ER_n^{*2}(\theta)}{\epsilon^2} = o(1),$$

which completes the proof of (b).

The fact that $\mathbf{G}_n$ is stochastically bounded, together with the convex function, $\Phi_n(\boldsymbol{\vartheta}) \xrightarrow{p} \frac{1}{2}\boldsymbol{\vartheta}^{\top}\mathbf{D}(u_0)\boldsymbol{\vartheta} - \frac{1}{\sqrt{nh_2}}\mathbf{G}_n^{\top}\boldsymbol{\vartheta}$, implies that

$$\sup_{\boldsymbol{\vartheta}\in\mathcal{K}_2} |R_n(\boldsymbol{\vartheta})| = o_p(1),$$

for any compact set, $\mathcal{K}_2$, which follows from the convexity lemma in Pollard (1991). This completes the proof of Lemma A.5.

Proof of Lemma A.6.

Since $E[\mathbf{Z}_{t,2}^* Q_\tau'(Y_{t1}^* - \eta(U_t, \mathbf{X}_{t,2}))K(\frac{U_t-u_0}{h_2})] = 0$, one has

$$E[\mathbf{Z}_{t,2}^* Q_\tau'(\tilde{Y}_t^*)K(\frac{U_t - u_0}{h_2})]$$

$$= E[\mathbf{Z}_{t,2}^* Q_\tau''(\tilde{Y}_t^*)K(\frac{U_t - u_0}{h_2})\mathbf{X}_{t,2}^{\top}\{\frac{1}{2}\mathbf{b}''(u_0)(U_t - u_0)^2\}](1 + o(1))$$

$$= E\left[\begin{pmatrix} 1 \\ \frac{U_t-u_0}{h_2} \end{pmatrix} \otimes 2\{\tau(1 - F_{y|x,u}(e_\tau(U_t, \mathbf{X}_t))) + (1-\tau)F_{y|x,u}(e_\tau(U_t, \mathbf{X}_t))\}\mathbf{X}_{t,2}\mathbf{X}_{t,2}^{\top} \right.$$

$$\left. K(\frac{U_t - u_0}{h_2})\frac{1}{2}\mathbf{b}''(u_0)(U_t - u_0)^2\right](1 + o(1))$$

$$= \frac{h_2^3}{2}f_u(u_0)\begin{pmatrix} \boldsymbol{\Gamma}_2(u_0)\mathbf{b}''(u_0)\mu_2 \\ 0 \end{pmatrix}(1 + o(1)), \tag{A-8}$$

and

$$E[\mathbf{Z}_{t,2}^* Q_\tau'(\tilde{Y}_t^*)K(\frac{U_t - u_0}{h_2})]^2$$

$$= E\left[\begin{pmatrix} 1 & \frac{U_t-u_0}{h_2} \\ \frac{U_t-u_0}{h_2} & (\frac{U_t-u_0}{h_2})^2 \end{pmatrix} \otimes K^2(\frac{U_t - u_0}{h_2})Q_\tau'^2(\tilde{Y}_t^*)\mathbf{X}_{t,2}\mathbf{X}_{t,2}^{\top}\right](1 + o(1))$$

$$= h_2 f_u(u_0) \begin{pmatrix} v_0 & 0 \\ 0 & v_2 \end{pmatrix} \otimes \mathbf{\Gamma}_2^*(u_0)(1 + o(1)) = h_2 \mathbf{\Sigma}(u_0)(1 + o(1)). \tag{A-9}$$

Next, the basic idea of proving the asymptotic normality is to employ the classical large-block and small-block technique, which partition the $\{1, \ldots, n\}$ into $2k_n + 1$ subsets with large block of size $r = r_n$ and small block of size $s = s_n$, where

$$k = k_n = \left[\frac{n}{r_n + s_n} \right].$$

Then, the Cramer–Wold device is used to derive the asymptotic normality of $\mathbf{G}_n$ for any unit vector, $d \in R^{2q}$. To this end, define

$$\begin{aligned}
P_n &\equiv \frac{d^\top}{\sqrt{nh_2}}(\mathbf{G}_n - E\mathbf{G}_n) \\
&= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{d^\top}{\sqrt{h_2}} \left[\mathbf{Z}_{t,2}^* Q_\tau'(\tilde{Y}_t^*) K\left(\frac{U_t - u_0}{h_2}\right) - E\{\mathbf{Z}_{t,2}^* Q_\tau'(\tilde{Y}_t^*) K\left(\frac{U_t - u_0}{h_2}\right)\} \right] \\
&\stackrel{\triangle}{=} \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} H_{n,t}^*.
\end{aligned}$$

From (A-8) and (A-9), it is easy to show that

$$\text{Var}(H_{n,t}^*) = \mathbf{\Sigma}_p(u_0)(1 + o(1)), \tag{A-10}$$

where $\mathbf{\Sigma}_p(u_0) = d^\top \mathbf{\Sigma}(u_0)d$ and

$$\sum_{s=1}^{n-1} |\text{Cov}(H_{n,0}^*, H_{n,s}^*)| = o(1). \tag{A-11}$$

For $0 \le j \le k - 1$, define the following three random variables

$$\eta_j = \sum_{i=j(r+s)}^{j(r+s)+r-1} H_{n,i}^*, \quad \xi_j = \sum_{i=j(r+s)+r}^{(j+1)(r+s)} H_{n,i}^*, \quad \text{and} \quad \zeta_k = \sum_{i=k(r+s)}^{n-1} H_{n,i}^*.$$

Then,

$$P_n = \frac{1}{\sqrt{n}} \left(\sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \xi_j + \zeta_k \right) = \frac{1}{\sqrt{n}}(P_{n,1} + P_{n,2} + P_{n,3}).$$

To establish the asymptotic result of P_n, Theorem 18.4.1 of Ibragimov and Linnik (1971) is employed. To this end, the following conditions must be checked

$$\frac{1}{n} E[P_{n,2}]^2 \to 0, \quad \frac{1}{n} E[P_{n,3}]^2 \to 0, \tag{A-12}$$

$$\left| E[\exp(itP_{n,1})] - \prod_{j=0}^{k-1} E[\exp(it\eta_j)] \right| \to 0, \tag{A-13}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E(\eta_j^2) \to \mathbf{\Sigma}_p(u_0), \tag{A-14}$$

and

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left[\eta_j^2 I\{|\eta_j| \geq \epsilon\sqrt{n\Sigma_p(u_0)}\}\right] \to 0 \tag{A-15}$$

for every $\epsilon > 0$. We first prove (A-12) and consider the large block sizes. Assumption (B2) implies that there is a sequence of positive constant, $a_n \to \infty$, such that

$$a_n s_n = o(\sqrt{nh_2(n)}),$$

and

$$(nh_2^{-1})^{1/2}\beta(s_n) \to 0.$$

Defining the large-block size, $r_n = [(nh_2)^{1/2}/a_n]$, and the small-block size, s_n, shows that

$$s_n/r_n \to 0, \quad r_n/n \to 0, \quad r_n(nh_2)^{-1/2} \to 0 \tag{A-16}$$

as $n \to \infty$, and

$$(n/r_n)\beta(s_n) \to 0.$$

It follows from the stationarity and equations (A-10) and (A-11) that

$$E[P_{n,2}]^2 = \sum_{j=0}^{k-1} \text{Var}(\xi_j) + 2\sum_{0 \leq i < j \leq k-1} \text{Cov}(\xi_i, \xi_j) \equiv I_1 + I_2,$$

in which

$$I_1 = k\text{Var}(\xi_0) = k\text{Var}(\sum_{i=r}^{r+s} H_{n,i}^*) = ks_n[\Sigma_p(u_0) + o(1)] = O(ks_n).$$

Next, I_2 is considered. Let $r_j^* = j(r_n + s_n)$, then $r_j^* - r_i^* \geq r_n$ for all $j > i$. Thus,

$$|I_2| \leq 2\sum_{0 \leq i < j \leq k-1} \sum_{j_1=1}^{s_n} \sum_{j_2=1}^{s_n} |\text{Cov}(P_{n,r_i^*+r_n+j_1}, P_{n,r_j^*+r_n+j_2})|$$

$$\leq 2\sum_{j_1=1}^{n-r_n} \sum_{j_2=j_1+r_n}^{n} |\text{Cov}(P_{n,j_1}, P_{n,j_2})| \leq 2n\sum_{j=r_n+1}^{n} |\text{Cov}(P_{n,1}, P_{n,j})| = o(n). \tag{A-17}$$

It is straight forward that from (A-16) and (A-17), one can obtain

$$\frac{1}{n}E[P_{n,2}^2] = O(ks_n n^{-1}) + o(1) = o(1). \tag{A-18}$$

In the same way, the stationarity and equations (A-10) and (A-16) imply that

$$\text{Var}(P_{n,3}) = \text{Var}\left(\sum_{j=1}^{n-k(r_n+s_n)} P_{n,j}\right) = O(n - k(r_n + s_n)) = o(n). \tag{A-19}$$

Thus, combining (A-18) and (A-19), (A-12) is established. To prove (A-13), applying Lemma A.1 of Volkonskii and Rozanov (1959) (See also Ibragimov and Linnik (1971)) leads to

$$\left| E[\exp(itP_{n,1})] - \prod_{j=0}^{k_n-1} E[\exp(it\eta_j)] \right| \le 16(n/r_n)\beta(s_n) \to 0.$$

Then,

$$\frac{1}{n}\sum_{j=0}^{k-1} E(\eta_j^2) = \frac{k}{n}E(\eta_1^2) = \frac{kr_n}{n} \cdot \frac{1}{r_n}\mathrm{Var}(\sum_{j=1}^{r_n} P_{n,j}) \to \boldsymbol{\Sigma}_p(u_0),$$

so that (A-14) is proved.

Finally, an application of Theorem 4.1 of Shao and Yu (1996) and Assumption B implies that both (A-15) and

$$E\left[\eta_j^2 I\left(|\eta_j| \ge \epsilon\sqrt{n\boldsymbol{\Sigma}_p(u_0)}\right)\right] \le Cn^{1-\delta/2}E(|\eta_j|^\delta) \le Cn^{1-\delta/2}r_n^{\delta/2}(E(|H_{n,0}^*|^{\delta^*}))^{\delta/\delta^*} \qquad \text{(A-20)}$$

hold true. Since

$$\begin{aligned}E(|H_{n,0}^*|^{\delta^*}) &\le Ch_2^{-\delta^*/2}E(\|\mathbf{G}_n\|^{\delta^*}) \\ &\le Ch_2^{-\delta^*/2}E[\|\mathbf{Z}_t^*\mathbf{Q}_\tau'(\tilde{Y}_t^*)K(\frac{U_t-u_0}{h_2})\|^{\delta^*}] \le Ch_2^{1-\delta^*/2}.\end{aligned} \qquad \text{(A-21)}$$

Thus, by (A-20) and (A-21),

$$E\left[\eta_j^2 I\{|\eta_j| \ge \epsilon\sqrt{n\boldsymbol{\Sigma}_p(u_0)}\right] \le Cn^{1-\delta/2}r_n^{\delta/2}h_2^{(2-\delta^*)\delta/(2\delta^*)}.$$

Therefore, by Assumption B and the definition of r_n, one has

$$\frac{1}{n}\sum_{j=0}^{k-1} E\left[\eta_j^2 I\{|\eta_j| \ge \epsilon\sqrt{n\boldsymbol{\Sigma}_p(u_0)}\right] \le Ca_n^{1-\delta/2}n^{1/2-\delta/4}h_2^{\delta/\delta^*-\delta/4-1/2} = o(1),$$

because $a_n \to \infty$. Finally, because of (A-12)–(A-15), one can use Theorem 18.4.1 of Ibragimov and Linnik (1971) to show that

$$P_n \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}_p(u_0)),$$

which completes the proof of Lemma A.6.

References

Bellini, F., Klar, B., Müller, A., Gianin, E.R., 2014. Generalized quantiles as risk measures. Insurance: Mathematics and Economics 54, 41–48.

Bellini, F., Valeria, B., 2015. On elicitable risk measures. Quantitative Finance 15, 725–733.

Cai, Z., 2002a. Regression quantiles for time series. Econometric Theory 18, 169–192.

Cai, Z., 2002b. Two-step likelihood estimation procedure for varying-coefficient models. Journal of Multivariate Analysis 82, 189–209.

Cai, Z., Fan, J., 2000. Average regression surface for dependent data. Journal of Multivariate Analysis 75, 112–142.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941–956.

Cai, Z., Masry, E., 2000. Nonparametric estimation of additive nonlinear ARX time series: local linear fitting and projections. Econometric Theory 16, 465–501.

Cai, Z., Xiao, Z., 2012. Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. Journal of Econometrics 167, 413–425.

Campbell, S.D., 2007. A review of backtesting and backtesting procedures. Journal of Risk 9, 1–17.

De Rossi, G., Harvey, A., 2009. Quantiles, expectiles and splines. Journal of Econometrics 152, 179–185.

Dette, H., Spreckelsen, I., 2004. Some comments on specification tests in nonparametric absolutely regular processes. Journal of Time Series Analysis 25, 159–172.

Efron, B., 1991. Regression percentiles using asymmetric squared error loss. Statistica Sinica 1, 93–125.

Ehm, W., Gneiting, T., Jordan, A., Krüger, F., 2016. Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings. Journal of the Royal Statistical Society: Series B 78, 505–562.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic press, New York.

Ibragimov, I., Linnik, I.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff, Groningen.

Jones, M.C., 1994. Expectiles and M-quantiles are quantiles. Statistics & Probability Letters 20, 149–153.

Kuan, C.M., Yeh, J.H., Hsu, Y.C., 2009. Assessing value at risk with CARE, the conditional autoregressive expectile models. Journal of Econometrics 150, 261–270.

Lee, J., 1990. U-statistics: Theory and Practice. Marcel Dekker, New York.

Newey, W.K., Powell, J.L., 1987. Asymmetric least squares estimation and testing. Econometrica 55, 819–847.

Pollard, D., 1991. Asymptotics for least absolute deviation regression estimators. Econometric Theory 7, 186–199.

Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.

Shao, Q.M., Yu, H., 1996. Weak convergence for weighted empirical processes of dependent sequences. Annals of Probability 24, 2 098–2127.

Speckman, P., 1988. Kernel smoothing partial linear models. Journal of Royal Statistical Society: Series B 50, 413–426.

Taylor, J.W., 2008. Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics 6, 231–252.

Volkonskii, V., Rozanov, Y.A., 1959. Some limit theorems for random functions. Theory of Probability & Its Applications 4, 178–197.

Xie, S., Zhou, Y., Wan, A.T., 2014. A varying-coefficient expectile model for estimating Value at Risk. Journal of Business & Economic Statistics 32, 576–592.

Yao, Q., Tong, H., 1996. Asymmetric least squares regression estimation: a nonparametric approach. Journal of Nonparametric Statistics 6, 273–292.

Ziegel, J.F., 2016. Coherence and elicitability. Mathematical Finance 26, 901–918.