Single-index expectile models for estimating conditional value at risk and expected shortfall

Rong Jiang[1];Xueping Hu[2]; Keming Yu[*3]

1 Department of Statistics, Donghua University, Shanghai, China

2 School of Mathematics and Computational Science, Anqing Normal University, Anqing, China

3 Brunel University, Uxbridge, Middlesex UB8 3PH, London, UK. Correspondence author

April 19, 2020

Abstract

This article develops a single-index approach for modeling the expectile-based value at risk (EVaR). EVaR has an advantage over the conventional quantile-based VaR (QVaR) of being more sensitive to the magnitude of extreme losses. EVaR can also be used for calculating QVaR and expected shortfall (ES) by exploiting the one-to-one mapping from expectiles to quantiles, and the relationship between VaR and ES. We develop an asymmetric least squares technique for estimating the unknown regression parameter and link function in a single-index model, and establish the asymptotic normality of the resultant estimators. Simulation studies and real data applications are conducted to illustrate the finite sample performance of the proposed methods.

Keywords: Single-index model; Expectile regression; Value at risk.

1 Introduction

In financial time series, expectiles emerge as alternative to popular risk measures such as value at risk (VaR) and expected shortfall (ES), see Ziegel (2014). VaR is a popular measure to evaluate the market risk of a portfolio. VaR identifies the loss that is likely to be exceeded by a specified probability that generally ranges between 0.90 and 0.99 over a defined period. VaR is therefore a quantile of the portfolio loss distribution; however, the use of VaR is not without criticism. It is generally agreed that VaR has three major drawbacks. First, it focuses exclusively on the lower tail of the distribution, and hence it conveys only a small slice of the information about the loss distribution. Second, it lacks subadditivity. That means the VaR of a portfolio can be larger than the sum of the individual VaRs, which contradicts the conventional wisdom that diversification reduces risk. Third, VaR tells us nothing about the magnitude of the loss as it accounts only for the probabilities of the losses but not their sizes. In light of these shortcomings of the VaR, Artzner et al. (1999) proposed to measure portfolio risk by expected shortfall (ES) instead. The ES is defined as the conditional expectation of a loss given that the loss is larger than the VaR. Contrary to VaR, ES provides information on the magnitude of the loss beyond the VaR level and is subadditive. However, the calculation of ES can be an intricate computational exercise due to the lack of a closed form formula (Yuan and Wong, 2010).

Taylor (2008) developed a simple way to calculate the ES associated with a given expectile based VaR (EVaR) estimate. EVaR proposed by Kuan et al. (2009) is an alternative to the quantile-based VaR (QVaR) as a downside risk measure. The population α-quantile of Y, $q_\alpha(Y)$, which is obtained by minimizing the quantile-check function

$$q_\alpha(Y) = \arg\min_m E\left[|\alpha - I\{Y < m\}| \cdot |Y - m|\right].$$

Newey and Powell (1987) adopted the asymmetric concept from quantile regression in a smooth manner and proposed asymmetric least squares (ALS) estimation, from which expectile originates. The τ-expectile of Y, $\tau \in (0,1)$, is defined as the value of $Q_\tau(Y)$ which minimizes the following loss function

$$Q_\tau(Y) = \arg\min_m E\left[|\tau - I\{Y < m\}| \cdot (Y - m)^2\right].$$

By comparing the above two loss functions, we can see that expectiles are more sensitive to the extreme values of the data than quantile estimates which are based on absolute errors. Specially, when $\tau = \alpha = 0.5$, the $Q_\tau(Y)$ and $q_\alpha(Y)$ are reduced to the mean and median of Y, respectively. This feature makes the EVaR correspondingly more sensitive to the scale of losses than the conventional QVaR. Moreover, expectile estimates and their covariances are easier to compute, reasonably efficient under normality conditions (Efron, 1991; Schnabel and Eilers, 2009), and can always be calculated regardless of the quantile level, while the empirical quantiles can be undefined at the extreme tails. It has been shown that for a given distribution, there is an one-to-one mapping between quantiles and expectiles (Efron, 1991; Yao and Tong, 1996). In view of this, Efron (1991) proposed to estimate the α-quantile by the expectile for which there is $100\alpha\%$ of sample observations lying below it. As pointed out by Kuan et al. (2009), this one-to-one relationship permits the interpretation of EVaR as a flexible QVaR, in the sense that its tail probability is determined not a priori but by the underlying distribution. Kuan et al. (2009) also showed that the asymmetric parameter in the weighted mean squared errors may be interpreted as the relative cost of the expected margin shortfall which represents prudentiality, and the EVaR is thus a risk measure under a given level of prudentiality.

In 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 which allow for asymmetric dynamic effects of the magnitude of positive and negative lagged returns on tail expectiles. 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. Cai et al. (2018) proposed a three-stage estimation procedure for a class of dynamic expectile models with partially varying coefficients. In such a 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 to parametric models, nonparametric and semiparametric methods can provide more flexibility and capture parameter heterogeneity and nonlinearity. The single-index approach has in particular the advantage of mitigating the risk of misclassifying the link function over existing parametric regression models and circumventing the curse of dimensionality that afflicts the estimation of many nonparametric and semiparametric regression models. These important advantages make the single-index approach a widely accepted modeling approach. The literature of single-index models is extensive. However, the single-index approach to the estimation of expectiles remains heretofore unexamined. In this study, let the τ-conditional expectile of Y given X

be modeled by the single-index model:

$$Q_\tau(Y|X) = g_0(X^\top \gamma_0), \tag{1.1}$$

where Y is the univariate response and X is a vector of p-dimensional covariates. The function $g_0(\cdot)$ is unspecified, nonparametric smoothing function. γ_0 is the unknown index vector coefficient, and for the sake of identifiability (Lin and Kulasekera, 2007), we assume that $\|\gamma_0\| = 1$ and that the first component of γ_0 is positive. Here $\|\cdot\|$ denotes the Euclidean norm. Note that $g_0(\cdot)$ and γ_0 may be dependent on τ. For notational simplicity, τ is dropped in $g_0(\cdot)$ and γ_0, whenever there is confusion. We call our model (1.1) single-index expectile model. It encompasses the CARE model of Kuan et al. (2009) as a special case.

Overall, this study offers a novel approach and makes the following two key contributions:

(1) We consider a new class of conditional dynamic expectile models: single-index expectile models.

(2) The proposed model and its inferential procedures are applied to assess the conditional VaR and ES.

The remainder of the paper is organized as follows. In Section 2, we introduce the ALS procedure for model (1.1). In Section 3, we introduce using expectile to estimate conditional VaR and ES. Both the simulation examples and the applications of two real datasets are given in Section 4 to illustrate the proposed procedures. Finally, Section 5 concludes the paper. All the conditions and technical proofs are relegated to the Appendix.

2 Asymmetric least squares estimation

2.1 Methodology

We set Y be asset returns, X be risk factors that may include lagged returns or other economic and financial factors. We assume that all series $\{X_t, Y_t\}_{t=1}^T$ in (X, Y) are strictly stationary processes satisfying the strong mixing (α-mixing) condition with finite two moments. Theoretically, the true parameter vector γ_0 in model (1.1) solves the following minimization problem:

$$\gamma_0 = \arg \min_{\|\gamma\|=1,\gamma_1>0} E\left[\rho_\tau\left(Y - g_0(X^\top\gamma)\right)\right], \tag{2.1}$$

where $\rho_\tau(\lambda) = |\tau - I(\lambda < 0)| \cdot \lambda^2$ for $\tau \in (0,1)$ is the loss function and the superscript $(\top)$ denotes the transpose of a vector or matrix. The right-hand side of (2.1) is the expected loss which can be equivalently written as

$$E\left[\rho_\tau\left(Y - g_0(X^\top\gamma)\right)\right] = E\left\{E\left[\rho_\tau\left(Y - g_0(X^\top\gamma)\right) \mid X^\top\gamma\right]\right\}. \tag{2.2}$$

We estimate the index function $g_0(\cdot)$ in (2.1) by local linear smoothing. For $X_t^\top\gamma$ "close" to u, $g(X_t^\top\gamma)$ can be approximated linearly by

$$g(X_t^\top\gamma) \approx g(u) + g'(u)(X_t^\top\gamma - u) \equiv a + b(X_t^\top\gamma - u), \tag{2.3}$$

where $a \equiv g(u)$ and $b \equiv g'(u)$. Following (2.3), $E\left[\rho_\tau\left(Y - g_0(X^\top\gamma)\right) \mid X^\top\gamma = u\right]$ can be approximated by

$$\sum_{t=1}^{T} \rho_\tau\left\{Y_t - a - b(X_t^\top\gamma - u)\right\} K\left(\frac{X_t^\top\gamma - u}{h}\right),$$

where $K(\cdot)$ is the kernel function and h is the bandwidth. By averaging on u, the empirical approximation of (2.2) is

$$\sum_{t'=1}^{T}\sum_{t=1}^{T} \rho_\tau\left\{Y_t - a_{t'} - b_{t'}\left(X_t - X_{t'}\right)^\top\gamma\right\} K_{t,t'}, \tag{2.4}$$

where $K_{t,t'} = K_h\left(\left(X_t - X_{t'}\right)^\top\gamma\right) / \sum_{t=1}^{T} K_h\left(\left(X_t - X_{t'}\right)^\top\gamma\right)$ and $K_h(\cdot) = K(\cdot/h)/h$. By (2.1), (2.2) and (2.4), the expectile regression estimate of γ_0 is

$$\hat{\gamma} = \arg \min_{\|\gamma\|=1,\gamma_1>0} \sum_{t'=1}^{T}\sum_{t=1}^{T} \rho_\tau\left\{Y_t - a_{t'} - b_{t'}\left(X_t - X_{t'}\right)^\top\gamma\right\} K_{t,t'}. \tag{2.5}$$

Since in (2.5), $a_{t'}$ and $b_{t'}$, $t' = 1, \ldots, T$ are unknown, minimization of (2.5) should be done by iteratively solving two simple problems, one with respect to $a_{t'}$ and $b_{t'}$, $t' = 1, \ldots, T$, and the other with respect to γ. The estimation procedure for estimating γ_0 is stated as follows:

Step 0. (Initialization step): Obtain initial $\hat{\gamma}^0$ from the minimum average variance estimation (MAVE) in Xia and Härdle (2006), which is an ALS estimate with $\tau = 0.5$.

Step 1. Given $\hat{\gamma}^0$, obtain $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^T$ by solving a series of the following minimisations

$$\min_{(a_{t'},b_{t'})} \sum_{t'=1}^{T} \rho_\tau \left\{Y_t - a_{t'} - b_{t'}(X_t - X_{t'})^\top \hat{\gamma}^0\right\} K_h \left((X_t - X_{t'})^\top \hat{\gamma}^0\right),$$

with the bandwidth h chosen optimally.

Step 2. Given $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^T$ in Step 1, obtain $\hat{\gamma}$ by solving

$$\min_{\|\gamma\|=1,\gamma_1>0} \sum_{t'=1}^{T} \sum_{t=1}^{T} \rho_\tau \left\{Y_t - \hat{a}_{t'} - \hat{b}_{t'} \left(X_t - X_{t'}\right)^\top \gamma\right\} K_{t,t'}.$$

with $K_{t,t'}$ evaluated at $\hat{\gamma}^0$ and h from Step 1.

Step 3. Repeat Steps 1 and 2 until convergence.

After obtaining the estimator $\hat{\gamma}$ of γ_0 in model (1.1) by the above estimation procedure, we can estimate $g_0(\cdot)$ in model (1.1) at any given point u by $\hat{g}(u \mid \hat{\gamma}) = \hat{a}$, where

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \sum_{t=1}^{T} \rho_\tau \left\{Y_t - a - b(X_t^\top \hat{\gamma} - u)\right\} K_h \left(X_t^\top \hat{\gamma} - u\right).$$

Remark 2.1: For the selection of h, the multifold cross-validation criterion proposed by Cai et al. (2000) is used to the model (1.1). The main idea behind this approach is that since the classical cross-validation may not work well for time series data in the literature. The multifold cross validation criterion is attentive to the structure of stationary time series data. Let l and H be two positive integers and the window l satisfies $T > lH$. First, with the H sub-series of length $T-ql$, $q = 1, \ldots, H$, the unknown functions are estimated. Then it computes, based on the estimated model, the one-step forecast errors of other subseries each of length l. Specifically, the optimal bandwidth is obtained by minimizing the average asymmetric mean squared error (AAMSE),

$$AAMSE(h) = \sum_{q=1}^{H} AAMSE_q(h),$$

where $AAMSE_q(h) = l^{-1} \sum_{t=T-ql+1}^{T-ql+l} \rho_\tau \left(Y_t - \hat{g}(X_t^\top \hat{\gamma})\right)$. It is worth noting that bandwidth is rescaled for different sample size according to the optimal rate $h = O(T^{-1/5})$, and one can take $l = [T/10]$ and $H = 4$ in practical implementations as suggested in Cai et al. (2000).

2.2 Asymptotic properties

Let $F_Y(\cdot|u)$ be the distribution function of Y conditional on u. Denote by $f_{U_0}(\cdot)$ the marginal density function of $U_0 = X^\top\gamma_0$. We state the asymptotic normality for $\hat{g}(u|\hat{\gamma})$ and $\hat{\gamma}$ in the following theorems with proofs in appendix.

Theorem 2.1. Suppose that conditions (C1)-(C5) given in the Appendix hold. If $T \to \infty$, and $h = cT^{-1/5}$ with $0 < c < \infty$, then for an interior point u of the support of $f_{U_0}(\cdot)$,

$$\sqrt{Th}\left\{\hat{g}(u|\hat{\gamma}) - g_0(u) - \frac{1}{2}g_0''(u)\mu_2 h^2\right\} \xrightarrow{L} N\left(0, \nu_0 f_{U_0}^{-1}(u)S^{-2}(u)\Sigma(u)\right),$$

where $\xrightarrow{L}$ stands for convergence in distribution, $\mu_2 = \int u^2K(u)du$, $\nu_0 = \int K^2(u)du$, and

$$S(u) = 2\left[\tau\{1 - F_Y(g_0(u)|u)\} + (1-\tau)F_Y(g_0(u)|u)\right],$$
$$\Sigma(u) = 4E\left[|\tau - I\{Y \le g_0(u)\}|(Y - g_0(u))\right]^2.$$

Theorem 2.2. Under the same conditions as in Theorem 2.1, then

$$\sqrt{T}(\hat{\gamma} - \gamma_0) \xrightarrow{L} N(\mathbf{0}, \tilde{S}^{-1}\tilde{\Sigma}\tilde{S}^{-1}),$$

where $\tilde{S}^{-1}$ is the generalized inverse of $\tilde{S}$, and

$$\tilde{S} = 2E\left\{\left|\tau - I\{Y \le g_0(X^\top\gamma_0)\}\right|\{g_0'(X^\top\gamma_0)\}^2(X - E(X|X^\top\gamma_0))(X - E(X|X^\top\gamma_0))^\top\right\},$$
$$\tilde{\Sigma} = 4E\left[\left\{|\tau - I\{Y \le g_0(X^\top\gamma_0)\}|(Y - g_0(X^\top\gamma_0))g_0'(X^\top\gamma_0)\right\}^2(X - E(X|X^\top\gamma_0))(X - E(X|X^\top\gamma_0))^\top\right]$$

2.3 Variance estimation

We consider estimation of the covariance matrix of $\hat{g}(u|\hat{\gamma})$ and $\hat{\gamma}$. By the proofs of Theorems 2.1 and 2.2 in the Appendix, we have

$$\hat{S}(u) = \frac{2}{Th}\sum_{t=1}^{T}\left|\tau - I\left\{Y_t \le \hat{g}(X_t^\top\hat{\gamma})\right\}\right|K_h(X_t^\top\hat{\gamma} - u) \xrightarrow{P} f_{U_0}(u)S(u),$$
$$\hat{\Sigma}(u) = \frac{4}{Th}\sum_{t=1}^{T}\left[\left|\tau - I\left\{Y_t \le \hat{g}(X_t^\top\hat{\gamma})\right\}\right|\left(Y_t - \hat{g}(X_t^\top\hat{\gamma})\right)K_h(X_t^\top\hat{\gamma} - u)\right]^2 \xrightarrow{P} \nu_0 f_{U_0}(u)\Sigma(u),$$
$$\tilde{S}^* = \frac{2}{T}\sum_{t'=1}^{T}\sum_{t=1}^{T}\left|\tau - I\left\{Y_t \le \hat{g}(X_t^\top\hat{\gamma})\right\}\right|\hat{g}'^2(X_{t'}^\top\hat{\gamma})(X_t - X_{t'})(X_t - X_{t'})^\top\hat{K}_{t,t'} \xrightarrow{P} \tilde{S},$$
$$\tilde{\Sigma}^* = \frac{4}{T}\sum_{t'=1}^{T}\sum_{t=1}^{T}\left[\left|\tau - I\left\{Y_t \le \hat{g}(X_t^\top\hat{\gamma})\right\}\right|\left(Y_t - \hat{g}(X_t^\top\hat{\gamma})\right)\hat{g}'(X_{t'}^\top\hat{\gamma})\hat{K}_{t,t'}\right]^2(X_t - X_{t'})(X_t - X_{t'})^\top \xrightarrow{P} \tilde{\Sigma},$$

where $\hat{K}_{t,t'} = K_h\left((X_t - X_{t'})^\top \hat{\gamma}\right) / \sum_{t=1}^{\top} K_h\left((X_t - X_{t'})^\top \hat{\gamma}\right)$. Thus, consistent estimators of the asymptotic covariance matrix of $\hat{g}(u, \hat{\gamma})$ and $\hat{\gamma}$ are $\hat{S}^{-1}(u)\hat{\Sigma}(u)\hat{S}^{-1}(u)$ and $\tilde{S}^{*-1}\tilde{\Sigma}^*\tilde{S}^{*-1}$, respectively. We usually call them sandwich estimators. Based on the Theorems 2.1-2.2 and above results, we can conduct statistical inference for $g_0(\cdot)$ and γ_0, such as confidence interval and hypotheses testing.

2.4 An algorithm for computing estimates

In this section, we use an iterative asymmetric least squares (IALS) approach in conjunction for obtaining the estimates of γ_0 and $g_0(\cdot)$. The IALS algorithm is detailed as following:

Step 0. Obtain initial $\hat{\gamma}^0$ and $\{\hat{a}^0_{t'}, \hat{b}^0_{t'}\}_{t'=1}^{T}$ by MAVE.

Step 1. Given $\hat{\gamma}^0$ and $\{\hat{a}^0_{t'}, \hat{b}^0_{t'}\}_{t'=1}^{T}$, obtain $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^{T}$ with an iterative solution,

$$\begin{pmatrix} \hat{a}_{t'} \\ \hat{b}_{t'} \end{pmatrix} = \left[\sum_{t=1}^{T} w_\tau(\hat{e}_{t,t'}) K_h\left((X_t - X_{t'})^\top \hat{\gamma}^0\right) Z_{t,t'} Z_{t,t'}^\top\right]^{-1} \sum_{t=1}^{T} w_\tau(\hat{e}_{t,t'}) K_h\left((X_t - X_{t'})^\top \hat{\gamma}^0\right) Z_{t,t'} Y_t,$$

where $w_\tau(e) = |\tau - I\{e \leq 0\}|$, $\hat{e}_{t,t'} = Y_t - \hat{a}^0 - \hat{b}^0 (X_t - X_{t'})^\top \hat{\gamma}^0$ and $Z_{t,t'} = (1, (X_t - X_{t'})\hat{\gamma}^0)^\top$.

Step 2. Given $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^{T}$ obtained in Step 1, obtain $\hat{\gamma}$ with an iterative solution,

$$\hat{\gamma} = \left[\sum_{t'=1}^{T}\sum_{t=1}^{T} w_\tau(\bar{e}_{t,t'}) \bar{K}_{t,t'} \hat{b}^2_{t'} (X_t - X_{t'})(X_t - X_{t'})^\top\right]^{-1} \sum_{t'=1}^{T}\sum_{t=1}^{T} w_\tau(\bar{e}_{t,t'}) \bar{K}_{t,t'} \hat{b}_{t'} (X_t - X_{t'})(Y_t - \hat{a}_{t'}), \tag{2.6}$$

where $\bar{e}_{t,t'} = Y_t - \hat{a}_{t'} - \hat{b}_{t'}(X_t - X_{t'})^\top \hat{\gamma}^0$ and $\bar{K}_{t,t'} = K_h\left((X_t - X_{t'})^\top \hat{\gamma}^0\right) / \sum_{t=1}^{T} K_h\left((X_t - X_{t'})^\top \hat{\gamma}^0\right)$.

Step 3. Repeat Steps 1 and 2 until convergence.

Step 4. After obtaining the estimator $\hat{\gamma}$, we can obtain $\hat{g}(u \mid \hat{\gamma}) = \hat{a}$ at any given point u with an iterative solution,

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \left[\sum_{t=1}^{T} w_\tau(\tilde{e}_t) K_h\left(X_t^\top \hat{\gamma} - u\right) \tilde{Z}_t \tilde{Z}_t^\top\right]^{-1} \sum_{t=1}^{T} w_\tau(\tilde{e}_t) K_h\left(X_t^\top \hat{\gamma} - u\right) \tilde{Z}_t Y_t, \tag{2.7}$$

where $\tilde{e}_t = Y_t - \hat{a}^0 - \hat{b}^0 (X_t^\top \hat{\gamma} - u)$ and $\tilde{Z}_t = (1, (X_t \hat{\gamma} - u))^\top$. $\hat{a}^0$ and $\hat{b}^0$ can obtain by MAVE.

3 Using expectile to estimate conditional VaR and ES

In this section, we introduce using expectile to estimate conditional VaR and ES.

3.1 Using expectile to estimate conditional VaR

In this section, we use expectiles as estimators of quantiles. This was first proposed by Efron (1991) who was attracted by the computational simplicity of ALS relative to quantile regression. The proposal involves using, as an estimator of the α quantile, the expectile for which the proportion of in-sample observations lying below the expectile is α. This is based on the fact that, for each τ expectile, there is a corresponding α quantile, though τ is typically not equal to α. The existence of a one-to-one mapping from expectiles to quantiles is supported by the theoretical work of Yao and Tong (1996). Thus, we set, as estimator of the α quantile, the τ expectile for which the proportion of in-sample observations lying below the expectile is α. Then, for model (1.1), instead of calculating the quantile-based conditional VaR (QCVaR) at a predetermined α level, a more sensible strategy is to compute the expectile-based conditional VaR (ECVaR) at a corresponding τ as following (see Taylor, 2008):

$$QCVaR(\alpha|X) = ECVaR(\tau|X) = |Q_\tau(Y|X)| = \left|g_0(X^\top \gamma_0)\right|.$$

Then, after obtaining the estimators $\hat{\gamma}$ of γ_0 by (2.6), we can obtain the $\hat{g}(X^\top \hat{\gamma})$ for any given X by (2.7) with an iterative solution. Thus, the estimation of conditional VaR is

$$\hat{Q}CVaR(\alpha|X) = \hat{E}CVaR(\tau|X) = \left|\hat{g}(X^\top \hat{\gamma})\right|.$$

3.2 Using expectile to estimate conditional ES

As mentioned in Section 1, ES overcomes certain weaknesses of VaR and is becoming a widely used downside risk measure. Now, expectiles can also be used to calculate ES by using the relationship between VaR and ES, and that between expectiles and quantiles. We consider the minimization of the function $E\left[|\tau - I\{Y < m\}| \cdot (Y-m)^2 \mid X\right]$ over m

for model (1.1). It is straightforward to show that the solution $Q_\tau(Y|X) = g_0(X^\top\gamma_0)$ of the minimization satisfies following expression:

$$\frac{1-2\tau}{\tau}E\left[\left(Y - g_0(X^\top\gamma_0)\right)I\{Y < g_0(X^\top\gamma_0)\} \mid X\right] = g_0(X^\top\gamma_0) - E[Y|X]. \tag{3.1}$$

This suggests a link between expectiles and ES. Expression (3.1) can be rewritten as

$$\begin{aligned}&E(Y \mid Y < g_0(X^\top\gamma_0), X)\\ =&\left(1 + \frac{\tau}{(1-2\tau)F_Y(g_0(X^\top\gamma_0) \mid X)}\right)g_0(X^\top\gamma_0) - \frac{\tau}{(1-2\tau)F_Y(g_0(X^\top\gamma_0) \mid X)}E[Y|X].\end{aligned}$$

This expression provides a formula for the conditional ES of the quantile α that coincides with the τ expectile. Referring to the relationship between expectile and quantile, we have $F_Y(g_0(X^\top\gamma_0) \mid X) = \alpha$. Hence, we can obtain

$$ES(\alpha|X) = E(Y \mid Y < g_0(X^\top\gamma_0), X) = \left(1 + \frac{\tau}{(1-2\tau)\alpha}\right)g_0(X^\top\gamma_0) - \frac{\tau}{(1-2\tau)\alpha}E[Y|X]. \tag{3.2}$$

If Y is defined to be $E[Y|X] = 0$, expression (3.2) becomes a simple form as following

$$ES(\alpha|X) = \left(1 + \frac{\tau}{(1-2\tau)\alpha}\right)g_0(X^\top\gamma_0). \tag{3.3}$$

This expression relates the conditional ES associated with the α-quantile to the corresponding τ-expectile under a conditional zero-mean distribution. The expression applies to the ES in the lower tail of the distribution; the corresponding upper tail expression may be obtained by replacing α and τ by $(1-\alpha)$ and $(1-\tau)$, respectively. Similar to section 3.1, we can obtain the estimation of (3.3) as

$$\hat{ES}(\alpha|X) = \left(1 + \frac{\tau}{(1-2\tau)\alpha}\right)\hat{g}(X^\top\hat{\gamma}).$$

4 Numerical studies

In this section, we first use Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures and then demonstrate the application of the proposed methods with the daily data of the S&P500 return series. The programs are written in R and are available upon request from the authors.

4.1 Simulation example

In this example, we focus on the finite sample performance of the proposed model and its IALS estimators. The data are generated from the following model:

$$Y_t = 2\exp\left\{-3\left(X_t^\top\gamma\right)^2\right\} + \sigma\left(X_t^\top\gamma\right)\varepsilon_t, \tag{4.1}$$

where $X_t^\top\gamma = Y_{t-1}\gamma_1 + Y_{t-2}\gamma_2$, $\gamma = (\gamma_1, \gamma_2)^\top = (1,1)^\top/\sqrt{2}$ and ε_t is i.i.d. $N(0,1)$. In this section, sample sizes are considered to be $T = 200, 500$ and simulations are repeated 100 times for each of given sample sizes. The Quartic kernel function $K(u) = \frac{15}{16}(1-u^2)I(|u| \le 1)$ is used for local linear smoothing. Homoscedastic and heteroscedastic models are considered by Case 1: $\sigma\left(X_t^\top\gamma\right) \equiv 0.5$ and Case 2: $\sigma\left(X_t^\top\gamma\right) \equiv 0.5\sqrt{1+\sin\left(X_t^\top\gamma\right)}$, respectively. We consider $\tau = 0.10, 0.25, 0.50, 0.75, 0.90$. The corresponding τ-level expectiles $Q_\tau(\varepsilon_t)$ of ε_t are -0.86,-0.44, 0, 0.44, and 0.86, respectively. This means that these $Q_\tau(\varepsilon_t)$'s are the solutions to $E\left[\rho_\tau(\varepsilon_t - Q)\right] = 0$ under the stated error distribution of ε_t. The expectile regression model is $Q_\tau(Y_t|X_t) = 2\exp\left\{-3\left(X_t^\top\gamma\right)^2\right\} + \sigma\left(X_t^\top\gamma\right)Q_\tau(\varepsilon_t)$.

To measure the performance of single-index coefficient estimate $\hat{\gamma}$, the mean and standard deviation (SD) of Bias are reported. The Bias of $\hat{\gamma}_j$ is defined by $\hat{\gamma}_j - \gamma_j$(true-value). The mean and SD of Bias are summarized in Tables 1 and 2. From Tables 1 and 2, one can see that the estimators are close to the true value, because the Bias are all very small.

The performances of $\hat{g}(u)$ with $u = 0.2, 0.3, 0.4$ are assessed by taking the relative absolute error (RAE):

$$RAE(u) = \left|\frac{\hat{g}(u) - g_0(u)}{g_0(u)}\right|.$$

The mean and standard deviation (in parentheses) of RAE are summarized in Table 3. From Table 3, one can see that the estimators are close to the true value, because of the small RAE. Moreover, Figure 1 illustrates the estimated $\hat{g}(\cdot)$ along with the data of $g_0(\cdot)$ under Case 2 and $T = 500$, showing that the proposed method performs well.

Tables 1-3 also report results relating to the performance of the sandwich method for constructing standard errors. In the Tables 1-3, SD is the standard deviation of $\hat{\gamma}$ or $\hat{g}(u)$ across 100 replications. The mean and standard deviation of 100 estimated standard deviation of $\hat{\gamma}$ or $\hat{g}(u)$ by the sandwich method (see Section 2.3), denoted by ESD and ESD_{sd}, respectively. ESD provides information on the accuracy of the sandwich method.

The results show that the differences of SD and ESD are generally small.

Finally, we evaluate the sensitivity of ECVaR and QCVaR to catastrophic events. We do so in the case ε_t's are obtained from an independent $N(0, 1/\sqrt{0.99})$ distribution with probability 0.99 or from $N(c, 1/\sqrt{0.01})$ with probability 0.01. We consider the following two cases: $\tau = \alpha = 0.05$ and $\tau = \alpha = 0.01$. In both cases, we let $c = [-50, -1]$, $T = 200$ and $\sigma\left(X_t^\top \gamma\right) \equiv 1$. Figure 2 depicts the ECVaR and QCVaR,

$$ECVaR_\tau = \arg\min_m \sum_{t=1}^{T} \left[|\tau - I\{\hat{\varepsilon}_t < m\}| \cdot (\hat{\varepsilon}_t - m)^2 \right],$$
$$QCVaR_\alpha = \arg\min_m \sum_{t=1}^{T} \left[|\alpha - I\{\hat{\varepsilon}_t < m\}| \cdot |\hat{\varepsilon}_t - m| \right],$$

where $\hat{\varepsilon}_t = Y_t - \hat{g}(X_t^\top \hat{\gamma})$, $\hat{g}(\cdot)$ and $\hat{\gamma}$ are based on $\tau = 0.5$. One can observe that the ECVaR is very sensitive to the change of values of c, while the QCVaR does not change when the values of c increase under $\tau = \alpha = 0.05$. For the case of $\tau = \alpha = 0.01$, both ECVaR and QCVaR vary with c. However, the change of ECVaR is relatively larger than that of QCVaR for each c. The results here are similar to those obtained in Kuan et al. (2009), Xie et al. (2014) and Cai et al. (2018).

4.2 Real data example 1: Boston housing data.

As an illustration, we now apply the proposed methodology to Boston housing data. Harrison and Rubinfeld (1978) firstly studied this housing data and estimated the demand for clean air. This data contain 506 observations on 14 variables, and is available in the MASS library in R or http://lib.stat.cmu.edu/datasets/boston. The dependent variable of interest is medv, the median value of owner-occupied homes in \$1000s, and the other thirteen variables are statistical measurements on the 506 census tracts in suburban Boston from the 1970 census.

Many regression studies have used this data set and found potential relationship between medv and RM, TAX, PTRATIO, LSTAT (see Wu, et al., 2010 and Jiang, et al., 2016). In this study, we focus on the following four covariates:

RM: average number of rooms per dwelling;

TAX: full-value property tax (in dollar) per \$10,000;

PTRATIO: pupil-teacher ratio by town;

LSTAT: lower status of the population (percent).

We follow previous studies and take logarithmic transformations on TAX and LSTAT. The dependent variable is centered around zero. In this study, the single-index expectile model (1.1) is used to fit the data for five different expectile levels τ=0.10, 0.25, 0.50, 0.75, 0.90. We apply our methodology in Section 2 to analyze this data. The estimated 10th, 25th, 50th, 75th, 90th expectiles and their 95% pointwise confidence intervals are shown in Figure 3 together with scatter plots of medv and the estimated indices. These plots suggest that the estimated conditional expectile function provides a good fit to the data.

4.3 Real data example 2: S&P500 index data.

To illustrate the practical usefulness of application of our proposed expectile model, a daily data of S&P500 index between January 4, 2010 and November 23, 2018 with 2240 observations is downloaded from Yahoo Finance. The daily returns are computed as 100 times the difference of the log of the prices, that is, $Y_t = 100 \log(p_t/p_{t-1})$, where p_t is the daily price. Table 4 collects the summary statistics of the daily returns. It is shown that the sample mean is close to zero with slightly negatively skewed, and by the histogram and estimated density in Figure 4, we can see that the data is approximate to skewed normal distribution, which gives a motivation to use expectile rather than quantile model. Figure 4 also gives the time series plot for S&P500, and it shows that extreme values mainly occur during 2011, and the return series is less volatile from 2012 to 2015.

We let the τ-level expectile be modeled by

$$Q_\tau\left(Y_t \mid Y_{t-1}^+, Y_{t-1}^-, Y_{t-2}^+, Y_{t-2}^-\right) = g_\tau\left(Y_{t-1}^+\gamma_{\tau,1} + Y_{t-1}^-\gamma_{\tau,2} + Y_{t-2}^+\gamma_{\tau,1} + Y_{t-2}^-\gamma_{\tau,2}\right), \tag{4.2}$$

and denote the corresponding single-index model as SI(2). The first 1500 observations from 2010 to 2015 are used for model estimation and the remaining 740 observations are reserved for the out-of-sample evaluation. For comparison purpose, we consider the following SQ(2) and ABS(2) parametric models used in Kuan et al. (2009):

SQ(2) model:

$$Y_t = a_{0,\tau} + a_{1,\tau}Y_{t-1} + b_{1,\tau}(Y_{t-1}^+)^2 + \beta_{1,\tau}(Y_{t-1}^-)^2 + b_{2,\tau}(Y_{t-2}^+)^2 + \beta_{2,\tau}(Y_{t-2}^-)^2 + \varepsilon_{t,\tau}.$$

ABS(2) model:

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

where $v^+ = \max(v, 0)$ and $v^- = \max(-v, 0)$. The index coefficient estimates and the standard deviation of index coefficient (Sd) are present in Table 5. It shows that Y_{t-1}^{-} and Y_{t-2}^{-} are more important than Y_{t-1}^{+} and Y_{t-2}^{+}. We also calculate the in-sample τ_{in} and out-of-sample τ_{out} tail probabilities for the estimated expectiles in Table 6. These probabilities are almost all bigger than given τ. This suggests that, when the index of prudentiality τ is our concern, the QCVaR will likely underestimate value at risk at the same level. Note that the out-of-sample tail probabilities are all smaller than their in-sample counterparts. This may be explained by the fact that both indices are less volatile in the out-of-sample period, as can be seen from Figure 4. Among the three models, the SI(2) model yields the smallest $|\tau_{out} - \tau_{in}|/\tau_{in}$ for any given τ except $\tau = 0.01$. This may be taken as an indication that the SI(2) model produces more stable estimates than the SQ(2) and ABS(2) models.

Finally, we apply the conditional VaR and ES estimated from the SI(2) model for the last 740 observations. As discussed in Section 3, $ECVaR(\tau|X)$ is identical to $QCVaR(\alpha|X)$ at the τ-level. We consider $\alpha = 0.05$, where the corresponding $\tau = 0.0158$. Figure 5 presents the estimated 5% QCVaR and 5% ES based on the SI(2) model for the last 740 out-of-sample observations of S&P500 index data, together with the actual observation-s. Furthermore, to compare the calculation time of the conditional VaR and ES using expectile and quantile methods, we present the computational running time for the full implementation of these methods with the SI(2) mode. The $QCVaR(\alpha|X)$ can be obtained by the quantile regression (Wu et al., 2010), and we can calculate the $ES(\alpha|X)$ as the average of less than $QCVaR(\alpha|X)$. Due to the need to estimate the value of τ, the estimation of expectile method would appear to be more computationally demanding than quantile method. However, it is interesting to note that, even with this extra task, the computational running time of expectile method (38.8702 seconds) was less than for the quantile method (110.2207 seconds). This is because the ALS minimization is somewhat less challenging than the quantile regression minimization. Taylor (2008) reached similar conclusions.

5 Conclusion

Model-based risk measurers in finance such as quantile regression and expectile regression models have received increasing attention among economists and practitioners. Nonparametric regression and semiparametric regression models could provide insight for parametric regression models and avoid model misspecification, but may face the curse of dimensionality phenomena that arise when analyzing and organizing data in high-dimensional analysis of large panel of economic and financial data which can be generated and stored with cheaper cost in this era of big data. It is well-known that single-index models can effectively avoid curse of dimensionality and are powerful tools for dimension reduction and semiparametric modeling. In this paper a single-index expectile model is first proposed and then an estimation procedure is employed to estimate the index coefficients and index function. Our simulation results re-confirm the fact that expectile models are more sensitive to extreme values than quantile models. Using the S&P500 return series, the proposed single-index expectile model outperforms the existing parametric models. For future works, it is interesting to consider expectile model for massive dataset.

Acknowledgment

The two anonymous referees provided numerous valuable comments that improve the manuscript. This research is supported by the National Natural Science Foundation of China (No. 11801069, No. 11261048) and the Shanghai Sailing Program (No. 17YF1400800).

Appendix

To establish the asymptotic properties of the proposed estimators, the following technical conditions are imposed.

C1. The function $g_0(\cdot)$ has a continuous and bounded second-order derivative.

C2. The kernel $K(\cdot) \geq 0$ has a compact support and its first derivative is bounded. It satisfies $\int_{-\infty}^{+\infty} K(u)du = 1$, $\int_{-\infty}^{+\infty} uK(u)du = 0$, $\int_{-\infty}^{+\infty} u^2K(u)du < \infty$, and $|\int_{-\infty}^{+\infty} u^jK^2(u)du| < \infty, j = 0, 1, 2$.

C3. The conditional density function of Y given u, $f_Y(y|u)$ is continuous and bounded away from 0 and ∞ in u for each y.

C4. The density function of $\mathbf{X}^\top\gamma$ is positive and uniformly continuous for γ in a neighborhood of γ_0. Further the density of $\mathbf{X}^\top\gamma_0$ ($f_{U_0}(\cdot)$) is continuous and bounded away from 0 and ∞ on its support. $f_{U_0,U_l}(\cdot,\cdot)$ is the density of (U_0, U_l) with $f_{U_0,U_l}(\cdot,\cdot)$ bounded.

C5. All process in $\{Y_t, X_t\}$ are α-mixing such that $\sum_l l^{\bar{c}}[\alpha(l)]^{1-2/\delta} \leq \infty$ for some $\delta > 2$ and $\bar{c} > 1 - 2/\delta$. Moreover, there exists a sequence of positive integers s_T such that $s_T \to \infty$, $s_T = o\left(\sqrt{Th}\right)$, and $\sqrt{T/h}\alpha(s_T) \to 0$ as $T \to \infty$.

Remark A.1: Conditions C1-C5 are standard conditions, which are commonly used in single-index models and expectile models, see Xia and Hӑrdle (2006), Wu et al., (2010), and Xie et al., (2014).

Lemma 1. For any $x, y \in R$, and $\tau \in (0,1)$,

$$|\rho_\tau(x+y) - \rho_\tau(x) - \rho'_\tau(x)y| \leq 4y^2,$$
$$|\rho'_\tau(x+y) - \rho'_\tau(x) - \rho''_\tau(x)y| \leq 4|y|,$$

where $\rho'_\tau(x) = \partial^2\rho_\tau(x)/\partial x^2$ for $x \neq 0$, and $\rho''_\tau(0) = 0$.

Proof. See Yao and Tong (1996).

Proof of Theorem 2.1. Note that

$$\sqrt{Th}\{\hat{g}(u|\hat{\gamma}) - g_0(u)\} = \sqrt{Th}\{\hat{g}(u|\hat{\gamma}) - \hat{g}(u|\gamma_0)\} + \sqrt{Th}\{\hat{g}(u|\gamma_0) - g_0(u)\},$$

where $\hat{g}(\cdot|\gamma_0)$ is a local linear estimator of $g_0(\cdot)$ when the index coefficient γ_0 is known. $\hat{g}(u|\hat{\gamma}) - \hat{g}(u|\gamma_0)$ can be shown $o_p(h^2)$. The details are given below. For given u,

$$(\hat{g}(u|\hat{\gamma}), \hat{g}'(u|\hat{\gamma})) = \arg\min_{(a,b)} \sum_{t=1}^{T} \rho_\tau\{Y_t - a - b(X_t^\top\hat{\gamma} - u)\}K\left(\frac{X_t^\top\hat{\gamma} - u}{h}\right),$$
$$(\hat{g}(u|\gamma_0), \hat{g}'(u|\gamma_0)) = \arg\min_{(a,b)} \sum_{t=1}^{T} \rho_\tau\{Y_t - a - b(X_t^\top\gamma_0 - u)\}K\left(\frac{X_t^\top\gamma_0 - u}{h}\right).$$

Denote

$$\begin{aligned}
&\bar{\theta}^* = \sqrt{Th}\{\hat{g}(u|\hat{\gamma}) - g_0(u), h(\hat{g}'(u|\hat{\gamma}) - g_0'(u))\}^\top,\\
&\bar{\theta}^{**} = \sqrt{Th}\{\hat{g}(u|\gamma_0) - g_0(u), h(\hat{g}'(u|\gamma_0) - g_0'(u))\}^\top,\\
&Z_t^* = (1, (X_t^\top\hat{\gamma} - u)/h)^\top, Z_t^{**} = (1, (X_t^\top\gamma_0 - u)/h)^\top,\\
&Y_t^* = Y_t - g_0(u) - g_0'(u)(X_t^\top\hat{\gamma} - u), Y_t^{**} = Y_t - g_0(u) - g_0'(u)(X_t^\top\gamma_0 - u),\\
&K_t^* = K\left(\left[X_t^\top\hat{\gamma} - u\right]/h\right), K_t^{**} = K\left(\left[X_t^\top\gamma_0 - u\right]/h\right).
\end{aligned}$$

Thus $\bar{\theta}^*$ and $\bar{\theta}^{**}$ minimize

$$\begin{aligned}
L_T^*(\theta) &= \sum_{t=1}^T \left[\rho_\tau(Y_t^* - \theta^\top Z_t^*/\sqrt{Th}) - \rho_\tau(Y_t^*)\right] K_t^*,\\
L_T^{**}(\theta) &= \sum_{t=1}^T \left[\rho_\tau(Y_t^{**} - \theta^\top Z_t^{**}/\sqrt{Th}) - \rho_\tau(Y_t^{**})\right] K_t^{**}.
\end{aligned}$$

Write

$$L_T^*(\theta) = E[L_T^*(\theta)|\mathcal{X}] - (Th)^{-1/2}\left(\sum_{t=1}^T \rho_\tau'(Y_t^*)Z_t^{*\top}K_t^* - E[\rho_\tau'(Y_t^*)|U_t]Z_t^{*\top}K_t^*\right)\theta + R_T^*(\theta), \tag{A.1}$$

where $\mathcal{X}$ is the σ-field generated by $\{X_t^\top\hat{\gamma}\}_{t=1}^T$, $U_t = X_t^\top\hat{\gamma}$, and

$$\begin{aligned}
&E[L_T^*(\theta)|\mathcal{X}]\\
=&-(Th)^{-1/2}\sum_{t=1}^T E[\rho_\tau'(Y_t^*)|U_t]\theta^\top Z_t^* K_t^* + (2Th)^{-1}\theta^\top\left(\sum_{t=1}^T K_t^* E[\rho_\tau''(Y_t^*)|U_t]Z_t^* Z_t^{*\top}\right)\theta(1 + o_p(1)),
\end{aligned}$$

and

$$\frac{1}{Th}\sum_{t=1}^T K_t^* E[\rho_\tau''(Y_t^*)|U_t]Z_t^* Z_t^{*\top} = E\left[h^{-1}K_t^* E[\rho_\tau''(Y_t^*)|U_t]Z_t^* Z_t^{*\top}\right] + o_p(1).$$

By Condition C4, we can obtain $f_U(\cdot) = f_{U_0}(\cdot)(1+o(1))$, where $U = X^\top\hat{\gamma}$, and by Condition C2, we can obtain

$$\begin{aligned}
&E\left[h^{-1}K_t^* E[\rho_\tau''(Y_t^*)|U_t]Z_t^* Z_t^{*\top}\right]\\
=&2f_{U_0}(u\left[\tau\{1 - F_Y(g_0(u) \mid u)\} + (1-\tau)F_Y(g_0(u) \mid u)\right]\begin{pmatrix} 1 & 0\\ 0 & \mu_2\end{pmatrix}(1 + o_p(1)),
\end{aligned}$$

thus

$$(Th)^{-1}\sum_{t=1}^T K_t^* E[\rho_\tau''(Y_t^*)|U_t]Z_t^* Z_t^{*\top} = S^*(1 + o_p(1)),$$

where $S^* = 2f_{U_0}(u)\left[\tau\{1 - F_Y(g_0(u) \mid u)\} + (1-\tau)F_Y(g_0(u) \mid u)\right] \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix}$. Thus,

$$E[L_T^*(\theta)|\mathcal{X}] = -(Th)^{-1/2}\sum_{t=1}^{T} E[\rho'_\tau(Y_t^*)|U_t]\theta^\top Z_t^* K_t^* + \frac{1}{2}\theta^\top S^*\theta(1 + o_p(1)). \qquad (A.2)$$

For $R_T^*(\theta)$ defined in (A.1), similar to the idea of proof Lemma B.2 in Xie et al. (2014), we can obtain $R_T^*(\theta) = o_p(1)$. Thus, substitute (A.2) into (A.1), we have

$$L_T^*(\theta) = \frac{1}{2}\theta^\top S^*\theta + W_T^{*\top}\theta + o_p(1),$$

where $W_T^* = -(Th)^{-1/2}\sum_{t=1}^{T}\rho'_\tau(Y_t^*)Z_t^*K_t^*$. By applying the convexity lemma (Pollard, 1991) and the quadratic approximation lemma (Fan and Gijbels, 1996), the minimizer of $L_T^*(\theta^*)$ can be expressed as

$$\bar\theta^* = -\{S^*\}^{-1}W_T^* + o_p(1).$$

$\bar\theta^{**}$ can be shown similarly as $\bar\theta^{**} = -\{S^*\}^{-1}W_T^{**} + o_p(1)$, where $W_T^{**} = -(Th)^{-1/2}\sum_{t=1}^{T}\rho'_\tau(Y_t^{**})Z_t^{**}K_t^{**}$. So, by the definitions of $\bar\theta^*$ and $\bar\theta^{**}$, we have

$$\begin{aligned}\sqrt{Th}\{\hat g(u|\hat\gamma) - \hat g(u|\gamma_0)\} &= f_{U_0}^{-1}(u)S(u)^{-1}(Th)^{-1/2}\sum_{t=1}^{T}\left[\rho'_\tau(Y_t^*)K_t^* - \rho'_\tau(Y_t^{**})K_t^{**}\right] \\ &= 2f_{U_0}^{-1}(u)S(u)^{-1}(Th)^{-1/2}\sum_{t=1}^{T}\left|\tau - I(Y_t^{**} < 0)\right|\left(Y_t^*K_t^* - Y_t^{**}K_t^{**}\right),\end{aligned}$$

where $S(u) = 2\left[\tau\{1 - F_Y(g_0(u) \mid u)\} + (1-\tau)F_Y(g_0(u) \mid u)\right]$ and the last equality is due to the fact that Y_t^* has the same sign as Y_t^{**} a.s. when $\|\hat\gamma - \gamma_0\| = O_p(T^{-1/2})$. Thus, we can obtain

$$E\left[\sqrt{Th}\{\hat g(u|\hat\gamma) - \hat g(u|\gamma_0)\}\right]^2 = O\left(h^{-1}E\left[Y_t^*K_t^* - Y_t^{**}K_t^{**}\right]^2\right) = O(o(1)) = o(1),$$

which also implies $E\left[\sqrt{Th}\{\hat g(u|\hat\gamma) - \hat g(u|\gamma_0)\}\right] = o(1)$. Thus, $\sqrt{Th}\{\hat g(u|\hat\gamma) - \hat g(u|\gamma_0)\} = o_p(1)$ according to its first and second moment. Therefore, $\hat g(u|\hat\gamma) - \hat g(u|\gamma_0) = o_p(h^2)$ under condition $h = cT^{-1/5}$ with $0 < c < \infty$. Following, we need to prove that

$$\sqrt{Th}\{\hat g(u|\gamma_0) - g_0(u) - \frac{1}{2}g_0''(u)\mu_2 h^2\} \xrightarrow{L} N\left(0, \nu_0 f_{U_0}^{-1}(u)S^{-2}(u)\Sigma(u)\right).$$

The details are given below. By $\bar\theta^{**} = -\{S^*\}^{-1}W_T^{**} + o_p(1)$, we have

$$\sqrt{Th}(\hat g(u|\gamma_0) - g_0(u)) = \frac{1}{\sqrt{Th}}f_{U_0}^{-1}(u)S^{-1}(u)\tilde W_T + o_p(1),$$

where $\tilde{W}_T = \sum_{t=1}^{T} \rho'_\tau(Y_t^{**})K_t^{**}$. The asymptotic normality of $\tilde{W}_T/\sqrt{T}$ is based on the Cramér-World method. Denote that

$$\tilde{W}_T^* = \frac{1}{\sqrt{Th}}(\tilde{W}_T - E\tilde{W}_T) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{\sqrt{h}}\left\{\rho'_\tau(Y_t^{**})K_t^{**} - E\left[\rho'_\tau(Y_t^{**})K_t^{**}\right]\right\} \equiv \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}\bar{W}_t,$$

where $\bar{W}_t = \frac{1}{\sqrt{h}}\left\{\rho'_\tau(Y_t^{**})K_t^{**} - E\left[\rho'_\tau(Y_t^{**})K_t^{**}\right]\right\}$. Let us partition $\{1, 2, \ldots, T\}$ into $2k+1$ subsets with large blocks of size $r = r_T$ and small blocks of size $s = s_T$. Set

$$k = \left[\frac{T}{r_T + s_T}\right], \tag{A.3}$$

where $[\cdot]$ is an integer function. Similar to Lemma 1, we obtain

$$Var(\bar{W}_0) = \nu_0 f_{U_0}(u)\Sigma(u)(1 + o(p)), \tag{A.4}$$

where $\Sigma(u) = 4E\left[|\tau - I(Y \le g_0(u))|(Y - g_0(u))\right]^2$, and

$$\sum_{l=0}^{T-1}\left|cov(\bar{W}_0, \bar{W}_l)\right| = o(1). \tag{A.5}$$

Now, define the random variables, $\eta_j = \sum_{i=j(r+s)}^{j(r+s)+r-1}\bar{W}_i$, $\varphi_j = \sum_{i=j(r+s)+r}^{(j+1)(r+s)}\bar{W}_i$, and $\psi_k = \sum_{i=k(r+s)}^{T-1}\bar{W}_i$ for $0 \le j \le k-1$. Then, we have

$$\tilde{W}_T^* = \frac{1}{\sqrt{T}}\left\{\sum_{j=0}^{k-1}\eta_j + \sum_{j=0}^{k-1}\varphi_j + \psi_k\right\}. \tag{A.6}$$

Condition C5 implies that there is a sequence of positive constants $c_T \to \infty$ such that $c_T s_T = o(\sqrt{Th})$ and $\sqrt{Th^{-1}}\alpha(s_T) \to 0$. Define the large-block size $r_T = \left[\sqrt{Th}/c_T\right]$ and the small-block size s_T. Then, as $T \to \infty$, we can obtain

$$s_T/r_T \to 0, \quad r_T/k \to 0, \quad r_T/\sqrt{Th} \to 0, \tag{A.7}$$

and $(T/r_T)\alpha(s_T) \to 0$. By the properties of stationary, (A.4) and (A.5), we have

$$\begin{aligned}
E\left[\sum_{j=0}^{k-1}\varphi_j\right]^2 =& \sum_{j=0}^{k-1}Var(\varphi_j) + 2\sum_{0\le t<l\le k-1}cov(\varphi_t, \varphi_l) \\
=& ks_T[Var(\bar{W}_0) + o_p(1)] + O\left(T\sum_{l=0}^{T-1}\left|cov(\bar{W}_0, \bar{W}_l)\right|\right) \\
=& O(ks_T) + o(T).
\end{aligned}$$

Hence by (A.3) and (A.7), we have

$$T^{-1}E\left[\sum_{j=0}^{k-1}\varphi_j\right]^2 = O(ks_T T^{-1}) + o(1) = o(1). \tag{A.8}$$

Similarly, we can show that

$$T^{-1}E\left[\psi_k\right]^2 \leq T^{-1}(T-k(s_T+r_T))Var(\bar{W}_0)+2\sum_{l=1}^{T-1}\left|cov(\bar{W}_0,\bar{W}_l)\right| \leq T^{-1}(s_T+r_T)Var(\bar{W}_0)+o(1) = o(1). \tag{A.9}$$

Then, by (A.6), (A.8) and (A.9), we have

$$\tilde{W}_T^* = \frac{1}{\sqrt{T}}\sum_{j=0}^{k-1}\eta_j + o_p(1).$$

By Lemma 1.1 of Volkonskii and Rozanov (1959), we have

$$\left|E\left[exp\left(it\sum_{j=0}^{k-1}\eta_j\right)\right] - \prod_{j=0}^{k-1}E\left[exp\left(it\eta_j\right)\right]\right| \leq 16(T/r_T)\alpha(s_T) \to 0. \tag{A.10}$$

Also, follows by using (A.4)-(A.6), we can obtain

$$T^{-1}\sum_{j=0}^{k-1}E[\eta_j^2] = \nu_0 f_{U_0}(u)\Sigma(u) + o(1). \tag{A.11}$$

Finally, by the definition of $\bar{W}_t$, we have

$$\frac{1}{\sqrt{T}}\max_{0\leq j\leq p-1}|\eta_j| \leq O(r_T/\sqrt{Th}) = o(1),$$

and for any constant $\tilde{c} > 0$, as $T \to \infty$

$$\max_{0\leq j\leq p-1} P\left(|\eta_j| \geq \tilde{c}\sqrt{T\nu_0 f_{U_0}(u)}\Sigma^{1/2}(u)\right) \to 0,$$

thus,

$$T^{-1}\sum_{j=0}^{k-1}E\left[\eta_j^2 I\left(|\eta_j| \geq \tilde{c}\sqrt{T}\nu_0 f_{U_0}(u)\Sigma(u)\right)\right] \to 0. \tag{A.12}$$

(A.10) states that the summands η_j are asymptotically independent, and (A.11) and (A.12) are the standard Lindeberg-Feller conditions for the asymptotic normality of $\tilde{W}_T^*$ under the independent setup, such as

$$\tilde{W}_T^* \xrightarrow{L} N(0, \nu_0 f_{U_0}(u)\Sigma(u)).$$

Moreover, we can obtain

$$E[\sqrt{Th}(\hat{g}(u|\gamma_0) - g_0(u))] = \frac{1}{2}g_0''(u)\mu_2 h^2(1 + o(1)),$$

$$Var[\sqrt{Th}(\hat{g}(u|\gamma_0) - g_0(u))] = \nu_0 f_{U_0}^{-1}(u)S^{-2}(u)\Sigma(u)(1 + o(1)).$$

This completes the proof.

Proof of Theorem 2.2. Given $\{\hat{a}_{t'}, \hat{b}_{t'}\}_{t'=1}^T$, minimize the following to obtain $\hat{\gamma}$

$$\sum_{t'=1}^T \sum_{t=1}^T \rho_\tau \left\{Y_t - \hat{a}_{t'} - \hat{b}_{t'} \left(X_t - X_{t'}\right)^\top \gamma\right\} K_{t,t'}.$$

Write $\hat{\gamma}^* = \sqrt{T}(\hat{\gamma} - \gamma_0)$, then $\hat{\gamma}^*$ minimizes the following,

$$Q_T^*(\gamma^*) = \sum_{t'=1}^T \sum_{t=1}^T \left[\rho_\tau \left(\tilde{Y}_{t,t'}^* - \hat{b}_{t'} X_{t,t'}^\top \gamma^*/\sqrt{T}\right) - \rho_\tau \left(\tilde{Y}_{t,t'}^*\right)\right] K_{t,t'},$$

where $\tilde{Y}_{t,t'}^* = Y_t - \hat{a}_{t'} - \hat{b}_{t'} X_{t,t'}^\top \gamma_0$ and $X_{t,t'} = X_t - X_{t'}$. Note that $Q_T^*(\gamma^*)$ can be write as

$$Q_T^*(\gamma^*) = E\left[Q_T^*(\gamma^*)\right] - \frac{1}{\sqrt{T}} \sum_{t'=1}^T \sum_{t=1}^T K_{t,t'} \left\{\rho_\tau'(\tilde{Y}_{t,t'}^*)\hat{b}_{t'} X_{t,t'}^\top - E\left[\rho_\tau'(\tilde{Y}_{t,t'}^*)\hat{b}_{t'} X_{t,t'}^\top\right]\right\} \gamma^* + o_p(1).$$

$$\begin{aligned}
E\left[Q_T^*(\gamma^*)\right] =& \sum_{t'=1}^T \sum_{t=1}^T \left\{E\left[\rho_\tau \left(\tilde{Y}_{t,t'}^* - \hat{b}_{t'} X_{t,t'}^\top \gamma^*/\sqrt{T}\right)\right] - E\left[\rho_\tau \left(\tilde{Y}_{t,t'}^*\right)\right]\right\} K_{t,t'} \\
=& -\frac{1}{\sqrt{T}} \sum_{t'=1}^T \sum_{t=1}^T E\left[\rho_\tau' \left(\tilde{Y}_{t,t'}^*\right) \hat{b}_{t'} X_{t,t'}^\top \gamma^*\right] K_{t,t'} \\
&+ \frac{1}{2T}\gamma^{*\top} \sum_{t'=1}^T \sum_{t=1}^T E\left[\rho_\tau'' \left(\tilde{Y}_{t,t'}^*\right) \hat{b}_{t'}^2 X_{t,t'} X_{t,t'}^\top\right] K_{t,t'}\gamma^* + o_p(1).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
Q_T^*(\gamma^*) =& -\frac{1}{\sqrt{T}} \left[\sum_{t'=1}^T \sum_{t=1}^T K_{t,t'}\rho_\tau'(\tilde{Y}_{t,t'}^*)\hat{b}_{t'} X_{t,t'}^\top\right] \gamma^* \\
&+ \frac{1}{2T}\gamma^{*\top} \left[\sum_{t'=1}^T \sum_{t=1}^T E\left[\rho_\tau'' \left(\tilde{Y}_{t,t'}^*\right) \hat{b}_{t'}^2 X_{t,t'} X_{t,t'}^\top\right] K_{t,t'}\right] \gamma^* + o_p(1) \\
=& -\frac{1}{\sqrt{T}} \left[\sum_{t'=1}^T \sum_{t=1}^T K_{t,t'}\rho_\tau'(\tilde{Y}_{t,t'}^*)\hat{b}_{t'} X_{t,t'}^\top\right] \gamma^* + \frac{1}{2}\gamma^{*\top}\tilde{S}\gamma^* + o_p(1).
\end{aligned}$$

The proof of the asymptotic normality of $\hat{\gamma}$ relies on quadratic approximation and literally follow a similar logic as in the proof of Theorem 2.1,

$$\sqrt{T}(\hat{\gamma} - \gamma_0) = \tilde{S}^{-1}\frac{1}{\sqrt{T}} \sum_{t'=1}^T \sum_{t=1}^T \rho_\tau'(\tilde{Y}_{t,t'}^*)\hat{b}_{t'} X_{t,t'}^\top K_{t,t'} + o_p(1) \xrightarrow{L} N(0, \tilde{S}^{-1}\tilde{\Sigma}\tilde{S}^{-1}),$$

thus, the proof of Theorem 2.2 is finished.

References

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9, 203-228.

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

Cai, Z., Fang, Y., and Tian, D. (2018). Assessing tail risk using expectile regressions with partially varying coefficients. Submitted to Journal of Management Science and Engineering.

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

Fan, J. and Gijbels, I. (1996). Local polynomial modeling and its applications, London: Chapman and Hall.

Jiang, R., Qian, W.M. and Zhou. Z.G. (2016). Weighted composite quantile regression for single-index models. Journal of Multivariate Analysis 148, 34-48.

Harrison, D., and Rubinfeld, D.L. (1978). Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management, 5, 81-102.

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

Lin, W. and Kulasekera, K.B. (2007). Identifiability of single-index models and additive-index models. Biometrika, 94, 496-501.

Newey, W.K. and 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.

Schnabel, S.K. and Eilers, P.H.C. (2009). Optimal expectile smoothing. Computational Statistics and Data Analysis, 53, 4168-4177.

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

Volkonskii, V.A. and Rozanov, Y.A. (1959). Some limit theorems for random functions. Theory of Probability and Its Applications, 4, 178-197.

Wu, T.Z., Yu, K. and Yu, Y. (2010). Single-index quantile regression. Journal of Multivariate analysis, 101, 1607-1621.

Xia, Y. and Härdle, W. (2006). Semi-parametric estimation of partially linear single-index models. Journal of Multivariate Analysis, 97, 1162-1184.

Xie, S.Y., Zhou, Y. and Wan, A. (2014). A varying-coefficient expectile model for estimating value at risk. Journal of Business & Economic Statistics, 32, 576-592.

Yao, Q. and Tong, H. (1996). Asymmetric least square regression estimation: a nonparametric approach. Journal of Nonparametric Statistics, 6, 273-292.

Yuan, M.L. and Wong, M.C.W. (2010). Analytical VaR and expectile shortfall for quadratic porfolios. Journal of Derivatives, 7, 33-44.

Ziegel, J. (2014). Coherence and elicitability. Mathematical Finance, 26, 901-918.