

This article was downloaded by: [City University of Hong Kong Library]

On: 30 October 2014, At: 19:17

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Business & Economic Statistics

Publication details, including instructions for authors and subscription information:

<http://amstat.tandfonline.com/loi/ubes20>

A Varying-Coefficient Expectile Model for Estimating Value at Risk

Shangyu Xie^a, Yong Zhou^{bc} & Alan T. K. Wan^d

^a RCAF and School of Banking and Finance, University of International Business and Economics, Beijing, China ()

^b Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

^c School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China ()

^d Department of Management Sciences, City University of Hong Kong, Hong Kong S. A. R., China ()

Accepted author version posted online: 07 May 2014. Published online: 28 Oct 2014.

To cite this article: Shangyu Xie, Yong Zhou & Alan T. K. Wan (2014) A Varying-Coefficient Expectile Model for Estimating Value at Risk, Journal of Business & Economic Statistics, 32:4, 576-592, DOI: [10.1080/07350015.2014.917979](https://doi.org/10.1080/07350015.2014.917979)

To link to this article: <http://dx.doi.org/10.1080/07350015.2014.917979>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://amstat.tandfonline.com/page/terms-and-conditions>

A Varying-Coefficient Expectile Model for Estimating Value at Risk

Shangyu XIE

RCAF and School of Banking and Finance, University of International Business and Economics, Beijing, China
(xshyu@amss.ac.cn)

Yong ZHOU

Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, and
School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, China
(yzhou@amss.ac.cn)

Alan T. K. WAN

Department of Management Sciences, City University of Hong Kong, Hong Kong S. A. R., China
(Alan.Wan@cityu.edu.hk)

This article develops a nonparametric varying-coefficient 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. Previous studies on conditional EVaR estimation only considered parametric autoregressive model set-ups, which account for the stochastic dynamics of asset returns but ignore other exogenous economic and investment related factors. Our approach overcomes this drawback and allows expectiles to be modeled directly using covariates that may be exogenous or lagged dependent in a flexible way. Risk factors associated with profits and losses can then be identified via the expectile regression at different levels of prudentiality. We develop a local linear smoothing technique for estimating the coefficient functions within an asymmetric least squares minimization set-up, and establish the consistency and asymptotic normality of the resultant estimator. To save computing time, we propose to use a one-step weighted local least squares procedure to compute the estimates. Our simulation results show that the computing advantage afforded by this one-step procedure over full iteration is not compromised by a deterioration in estimation accuracy. Real data examples are used to illustrate our method. Supplementary materials for this article are available online.

KEY WORDS: Asymmetric squared error loss; α -mixing; Expected shortfall; Local linear smoothing; One-step weighted least squares; Value at risk.

1. INTRODUCTION

Value at risk (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 exer-

cise due to the lack of a closed form formula (Yuan and Wong 2010).

Focusing on the third of the above-mentioned drawbacks of the VaR, Kuan, Yeh, and Hsu (2009) proposed an expectile-based VaR (EVaR) as an alternative to the quantile-based VaR (QVaR) as a downside risk measure. Expectile regression estimates are obtained by minimizing an asymmetrically weighted sum of the squared errors (Aigner, Amemiya, and Poirier 1976; Newey and Powell 1987). Expectile estimates are thus more sensitive to the extreme values of the data than quantile estimates which are based on absolute errors. 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

© 2014 American Statistical Association
Journal of Business & Economic Statistics
October 2014, Vol. 32, No. 4

DOI: 10.1080/07350015.2014.917979

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jbes.

level, while the empirical quantiles can be undefined at the extreme tails. It has been shown that for a given distribution, there is a one-to-one mapping between quantiles and expectiles (Efron 1991; Jones 1994; Yao and Tong 1996). In view of this, Efron (1991) proposed that the τ -quantile be estimated by the expectile for which there is $100\tau\%$ of sample observations lying below it. As pointed out by Kuan, Yeh, and Hsu (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, Yeh, and Hsu (2009) 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 prudence, and the EVaR is thus a risk measure under a given level of prudence. As pointed out by Taylor (2008), there is an algebraic link between EVaR and ES under a given distribution of returns. This relationship permits a simple calculation of the ES associated with a given EVaR estimate.

Another contribution of Kuan, Yeh, and Hsu's (2009) study is the development of a class of conditional autoregressive expectile (CARE) models, which allow expectiles to be estimated in a dynamic context where past returns influence present returns based on some special types of autoregressive processes. The allowance for asymmetric effects of positive and negative returns on the tail expectiles is a notable feature of the CARE models. Kuan, Yeh, and Hsu (2009) also established the asymptotic properties of the asymmetric least squares (ALS) estimator of the expectiles. Some of Kuan and coworker's results extend those of Newey and Powell (1987) to allow for stationary and weakly dependent data. The main disadvantage of the CARE approach is that it considers only the stochastic dynamics of returns but ignores current information of the investment environment, such as the state of the economy and the financial environment at the time. It is possible to extend the CARE models to include both past observed returns and exogenous variables containing economic and market information as risk factors, and the objective of this article is to take steps in this direction. Unlike the work of Kuan, Yeh, and Hsu (2009) which is based on a parametric set-up, we adopt the semiparametric varying-coefficient approach (Cleveland, Grosse, and Shyu 1991; Hastie and Tibshirani 1993). Nonparametric and semiparametric methods have the general benefits of allowing the data to self-adjust instead of assuming a priori a functional form. The varying-coefficient approach, which allows coefficients to vary with other variables, has in particular the advantage of circumventing the curse of dimensionality that afflicts the estimation of many nonparametric and semiparametric models. It also allows dynamic patterns and interactions of the covariates to be modeled in a flexible way. Due to these important advantages, the varying-coefficient approach is now a widely accepted modeling approach. The literature of varying-coefficient models is extensive; for a general appreciation of its scope and purpose, see Fan and Huang (2005) and Fan and Zhang (2008). Recent developments of the varying-coefficient approach emphasize time series analysis (Cai, Fan, and Yao 2000; Cai and Xu 2008), longitudinal analysis (Fan and Zhang 2000; Fan and Li 2004), and survival analysis (Fan, Lin, and Zhou 2006; Cai et al. 2007, 2008). To the best of our knowledge, Honda (2004), Kim (2007), and Cai and Xu (2008) are the only existing studies that consider varying-coefficient

models for conditional quantiles; Honda (2004) and Cai and Xu (2008) used local polynomials to estimate conditional quantiles with varying-coefficients, while Kim (2007) proposed an estimation methodology based on polynomial splines. However, the varying-coefficient approach to the estimation of expectiles remains heretofore unexamined. We call our model the varying-coefficient expectile regression model. It generalizes the functional coefficient autoregressive (FAR) model of Chen and Tsay (1993), and also encompasses the CARE model of Kuan, Yeh, and Hsu (2009) as a special case. Our model estimation is based on the ALS method using local linear smoothing along the lines of Yao and Tong (1996). It is shown that the proposed estimator possesses the desirable large sample properties of consistency and asymptotic normality. We also develop an efficient algorithm to facilitate the computation of the EVaR estimates.

The plan of the article is as follows. Section 2 describes the varying-coefficient expectile model, and develops an ALS-based nonparametric approach for model estimation. The same section also discusses an iterative weighted local least squares approach together with a one-step algorithm for computing the coefficient estimates. Section 3 derives the asymptotic properties of the estimator and develops an estimator of the variance. Section 4 reports results of a simulation study and empirical applications. The technical proofs of theorems are presented in the Appendix.

2. MODEL SPECIFICATION AND ESTIMATION

2.1 Model Framework and an ALS Nonparametric Estimation Approach

Let (Y_t, \mathbf{X}_t, U_t) , $t = 1, 2, \dots, T$, be a sequence of strictly stationary random vectors, each having the same distribution as (Y, \mathbf{X}, U) . We let Y be asset returns, $\mathbf{X} = (X_1, X_2, \dots, X_p)$ be risk factors that may include lagged returns or other economic and financial factors, and U be a single effect modifying risk factor. We assume that all series in (Y, \mathbf{X}, U) are strictly stationary processes satisfying the strong mixing (α -mixing) condition and $E(Y^2) < \infty$. The population θ -expectile of Y is the parameter $v_\theta(Y)$ that minimizes the function $E[Q_\theta(Y - v)]$, where

$$Q_\theta(z) = \begin{cases} (1 - \theta)z^2, & z \leq 0 \\ \theta z^2, & z > 0 \end{cases} \quad (2.1)$$

is the asymmetric squared error loss function, and $\theta \in (0, 1)$ is an asymmetric parameter that controls the degree of loss asymmetry. Note that $v_\theta(Y)$ is different from the τ -quantile of Y , $q_\tau(Y)$, which is obtained by minimizing $E[\rho_\tau(Y - q)]$, with

$$\rho_\tau(z) = \begin{cases} (\tau - 1)z, & z \leq 0 \\ \tau z, & z > 0 \end{cases} \quad (2.2)$$

being the asymmetric absolute error loss function, and τ the corresponding asymmetric parameter. Owing to the squared error loss function (2.1), expectiles are more sensitive to the tails of the distribution than quantiles. Clearly, when $\theta = \tau = 0.50$, the θ -expectile and τ -quantile of Y reduce to the mean and median of Y , respectively.

In this study, we let the θ -conditional expectile of Y be modeled by the varying-coefficient model

$$v_\theta(Y|\mathbf{X}, U) = \mathbf{X}^T \mathbf{a}_\theta(U), \tag{2.3}$$

where $\mathbf{a}_\theta(U) = (a_{1,\theta}(U), \dots, a_{p,\theta}(U))^T$ is a vector of smooth varying-coefficient functions of U , and $a_{j,\theta}(U)$'s, $j = 1, \dots, p$, may be dependent on θ . For notational simplicity, we write $a_{j,\theta}(U)$ simply as $a_j(U)$ hereafter, whenever there is confusion. The varying-coefficient model set-up is flexible in that the responses are linearly associated with a set of covariates, but their regression coefficients can vary with U . This framework also overcomes the curse of dimensionality as $a_j(U)$'s are all low-dimensional functions. The coefficient function $a_j(U)$ thus characterizes the manner in which the relationship between a risk factor X_j 's and asset returns Y changes as the level of the effect modifying risk factor U changes. We refer to θ -conditional expectile as the EVaR with θ -level of prudence. As mentioned in Section 1, the θ -level EVaR may be converted to QVaR's with tail probabilities that depend on the underlying distribution. Now, given (2.3) and the definition of expectile, the estimators of $a_j(U)$'s are the solutions to the optimization problem

$$\min_{\mathbf{a}(u)} E \left\{ Q_\theta \left(Y - \sum_{j=1}^p a_j(U) X_j \right) \mid \mathbf{X} = x, U = u \right\}, \tag{2.4}$$

with the loss function being the asymmetric squared error loss defined in (2.1). Clearly, when $\theta = 0.50$, (2.4) reduces to the standard least squares objective function.

When $a_j(U)$'s are parameters rather than functions, and all of X_j 's are lagged values of Y or their transformations, the varying-coefficient expectile model reduces to the CARE models of Kuan, Yeh, and Hsu (2009). Also, when $p = 1$ and $X_1 \equiv 1$, model (2.3) reduces to the ordinary nonparametric expectile regression model studied by Yao and Tong (1996). Another interesting special case of (2.3) is the following functional-coefficient autoregressive (FAR) model introduced by Chen and Tsay (1993):

$$Y_t = a_1(Y_{t-d}^*)Y_{t-1} + \dots + a_p(Y_{t-d}^*)Y_{t-p} + \varepsilon_t, \tag{2.5}$$

where ε_t is a sequence of iid random variables distributed independently of Y_{t-i} for any $i > 0$, $Y_{t-d}^* = (Y_{t-d}, \dots, Y_{t-d+k})^T$, $i_j > 0$, $j = 1, \dots, k$. The FAR model is a very inclusive model that contains the threshold autoregressive (TAR) model (Tong 1983, 1990), the exponential autoregressive (EXPAR) model (Haggan and Ozaki 1981), and the smooth transition AR (STAR) model (Granger and Teräsvirta 1993) as special cases. To show that (2.5) satisfies the strong mixing condition, suppose that $a_j(\cdot)$ can be written as $a_j(\cdot) = \delta_j(\cdot) + \gamma_j(\cdot)$, with $\gamma_j(Y_{t-d}^*) \| Y_{t-d}^* \|$ being bounded on R^d , and $\delta_j(\cdot) \leq c_j$ such that all roots of $\lambda^p - c_1 \lambda^{p-1} - \dots - c_p = 0$ lie inside the unit circle. From Theorem 1.2 of Chen and Tsay (1993), if the density of ε is always positive, the Markov chain $\{Y_t\}$ generated by model (2.5) is geometrically ergodic, which implies that Y_t is a strong mixing process (Pham 1986).

We estimate the coefficient functions in (2.3) by local linear smoothing. Denote $\mathbf{a} = (a_1(\cdot), \dots, a_p(\cdot))^T$ and $\mathbf{b} = (b_1, \dots, b_p)^T = (a_1'(\cdot), \dots, a_p'(\cdot))^T$, and assume that $a_j(u)$, $j = 1, \dots, p$, is twice continuously differentiable so that it can be

approximated locally as $a_j(u) \approx a_j + b_j(u - u_0)$ in the neighborhood of a point u_0 . Write $\boldsymbol{\beta} = (\mathbf{a}^T, \mathbf{b}^T)^T$. For a given θ -level, the local linear regression estimator $\hat{\boldsymbol{\beta}}_\theta = (\hat{\mathbf{a}}_\theta^T, \hat{\mathbf{b}}_\theta^T)^T$ is the estimator that minimizes

$$\frac{1}{T} \sum_{t=1}^T Q_\theta(Y_t - \boldsymbol{\beta}^T \mathbf{Z}_t) K((U_t - u_0)/h), \tag{2.6}$$

where $\mathbf{Z}_t = (\mathbf{X}_t^T, \mathbf{X}_t^T(U_t - u_0))^T$, and $K(\cdot)$ is a kernel function and $h = h_T$ a bandwidth. Minimizing (2.6) with respect to $\boldsymbol{\beta}$ yields the following estimating equation:

$$\frac{1}{T} \sum_{t=1}^T L_\theta(Y_t - \boldsymbol{\beta}^T \mathbf{Z}_t) \mathbf{Z}_t^* K((U_t - u_0)/h) = 0, \tag{2.7}$$

where $L_\theta(z) = 2z \cdot |\theta - I(z \leq 0)|$ is the derivative of the loss function $Q_\theta(z)$ defined in (2.1), $\mathbf{Z}_t^* = (\mathbf{X}_t^T, \mathbf{X}_t^T(U_t - u_0)/h)^T = \mathbf{H}^{-1} \mathbf{Z}_t$, and $\mathbf{H} = I_p \otimes \text{diag}(1, h)$, with \otimes denoting the Kronecker product, and I_p being a p -dimensional identity matrix. When $p = 1$ and $X_1 \equiv 1$, the solution of (2.7) reduces to a special case of the local M-estimator given in Cai and Ould-Said (2003).

2.2 An Algorithm for Computing Estimates

Many different approaches can be used to solve Equation (2.7) and obtain estimates for the unknown parameters $\boldsymbol{\beta}$. Here, we use an iterative weighted local least squares (IWLLS) approach in conjunction with a one-step algorithm for obtaining the estimates. IWLLS approaches similar to ours have been used in other studies of expectiles (e.g., Newey and Powell 1987; Yao and Tong 1996). The one-step algorithm has the advantage of saving computing time. To describe the IWLLS approach, let us rewrite the estimating Equation (2.7) as

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T |\theta - I_{\{Y_t - \boldsymbol{\beta}^T \mathbf{Z}_t \leq 0\}}| K_t \mathbf{Z}_t^* (Y_t - \boldsymbol{\beta}^T \mathbf{Z}_t) \\ & \equiv \frac{1}{T} \sum_{t=1}^T w(e_t(\theta); \theta) K_t \mathbf{Z}_t^* e_t(\theta) = 0, \end{aligned}$$

where $K_t = K((U_t - u_0)/h)$, and $w(e_t(\theta); \theta) = |\theta - I_{\{e_t(\theta) \leq 0\}}|$ with $e_t(\theta) = Y_t - \boldsymbol{\beta}^T \mathbf{Z}_t$. Thus, the IWLLS estimator may be expressed as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_\theta(u_0) &= \left[\sum_{t=1}^T w(e_t(\theta); \theta) K_t \mathbf{Z}_t \mathbf{Z}_t^T \right]^{-1} \\ & \times \left[\sum_{t=1}^T w(e_t(\theta); \theta) K_t \mathbf{Z}_t Y_t \right], \end{aligned} \tag{2.8}$$

where $e_t(\theta) = Y_t - \mathbf{Z}_t^T \boldsymbol{\beta}_\theta(u_0)$. Although (2.8) does not provide a closed-form solution, it can be used as a basis for obtaining an iterative solution. Let $\hat{\boldsymbol{\beta}}_\theta^j(u_0)$ denote the IWLLS estimate obtained from the j th iteration. The IWLLS estimate $\hat{\boldsymbol{\beta}}_\theta^{j+1}(u_0)$

of the next iteration is thus

$$\widehat{\beta}_\theta^{j+1}(u_0) = \left[\sum_{t=1}^T w(\widehat{e}_t(\theta); \theta) K_t \mathbf{Z}_t \mathbf{Z}_t^T \right]^{-1} \times \left[\sum_{t=1}^T w(\widehat{e}_t(\theta); \theta) K_t \mathbf{Z}_t Y_t \right], \quad (2.9)$$

where $\widehat{e}_t(\theta) = Y_t - \mathbf{Z}_t^T \widehat{\beta}_\theta^j(u_0)$, provided that the matrix inverse in (2.9) exists. Iteration continues until convergence is achieved. A natural initial value for the iteration is the least squares estimator obtained by setting $\theta = 0.50$ because this estimator is easy to compute.

The computational burden involved in applying this iterative procedure to estimate the entire coefficient function $\widehat{\beta}_\theta(u)$ is likely to be enormous as it entails the calculation of estimates at hundreds of points even over a small interval. One way to reduce the computational burden is to resort to the following one-step algorithm that has been used in several other studies of varying-coefficient models (e.g., Cai, Fan, and Li 2000; Cai et al. 2007). Let $\widetilde{\beta}_0(u_0)$ be an initial estimate at any given point u_0 . A one-step application of (2.9) yields the estimator

$$\widehat{\beta}_{OS}(u_0) = \left[\sum_{t=1}^T w(\widetilde{e}_t(\theta); \theta) K_t \mathbf{Z}_t \mathbf{Z}_t^T \right]^{-1} \times \left[\sum_{t=1}^T w(\widetilde{e}_t(\theta); \theta) K_t \mathbf{Z}_t Y_t \right], \quad (2.10)$$

where $\widetilde{e}_t(\theta) = Y_t - \mathbf{Z}_t^T \widetilde{\beta}_0(u_0)$. The choice of a good initial estimate is the key to the success of this one-step algorithm. Following Fan and Chen (1999), it is not difficult to show that a sufficient condition for the estimator resulting from the one-step algorithm to share the same asymptotic properties as that obtained using the fully iterative procedure is $\mathbf{H}(\widetilde{\beta}_0 - \beta_0) = O_p(h^2 + (Th)^{-1/2})$. As pointed out by Cai et al. (2007), if this sufficient condition is not satisfied, a multistep estimator that repeatedly applies the one-step algorithm k times should be used instead, in which case the sufficient condition is relaxed to $\mathbf{H}(\widetilde{\beta}_0 - \beta_0) = O_p\{(h^2 + (Th)^{-1/2})^{1/k}\}$. We follow Cai, Fan, and Li's (2000) strategy by first computing the IWLLS estimates (based on full iterations) at a few fixed points. These estimates are then used as initial values for their nearest grid points, and at each of these grid points we compute the one-step estimate based on (2.10). Then we use the newly computed one-step estimates as initial values for their nearest grid points to compute the one-step estimates at these other points. We propagate until the one-step estimates at all grid points are obtained. For example, in our simulation study, our aim is to estimate the coefficient functions at $n_{\text{grid}} = 200$ grid points. To do so we first compute the IWLLS estimates at five distant points: $u_{20}, u_{60}, u_{100}, u_{140}, u_{180}$. We then use, for instance, $\widehat{\beta}(u_{60})$ as an initial value for calculating the estimates $\widehat{\beta}(u_{59})$ and $\widehat{\beta}(u_{61})$ based on the one-step estimator formula (2.10), and subsequently proceed to use these estimates as initial values for calculating the one-step estimates $\widehat{\beta}(u_{58})$ and $\widehat{\beta}(u_{62})$, and so on. We continue this process until the one-step estimates at all the points in the neighborhood of u_{60} , say, the points between u_{40}

and u_{79} , are all calculated. In Section 4, we will show that estimates resulting from this one-step procedure are as efficient as those obtained from IWLLS based on full iterations.

2.3 Bandwidth Selection

Various bandwidth selection techniques have been developed for nonparametric regression. Here, we adopt the method proposed by Cai, Fan, and Yao (2000), which may be regarded as a modified multifold cross-validation criterion that takes into consideration the structure of stationary time series data.

Let m and H be two positive integers such that $n > mH$. The method first uses H subseries, each of length $T - km$ ($k = 1, \dots, H$), to estimate the unknown coefficient functions. Then it computes, based on the estimated models, the one-step forecast errors of other subseries each of length m . Specifically, we select the optimal bandwidth $h = h_{\text{opt}}$ that minimizes average mean squared (AMS) error

$$\text{AMS}(h) = \sum_{k=1}^H \text{AMS}_k(h), \quad (2.11)$$

where

$$\text{AMS}_k(h) = \frac{1}{m} \sum_{T-km+1}^{T-km+m} Q_\theta \left\{ Y_t - \sum_{j=1}^p \widehat{a}_{j,k}(U_t) X_{t,j} \right\}, \quad k = 1, \dots, H, \quad (2.12)$$

and $\widehat{a}_{j,k}(\cdot)$'s are computed based on the subsample $\{(U_t, \mathbf{X}_t, Y_t), 1 \leq t \leq T - km\}$. In practice, the choices of m and H usually do not have a large impact on the value of h chosen by this method as long as mH is reasonably large so that the forecast errors are stable. We will apply this bandwidth selection method in the simulation and real data analysis of Section 4.

2.4 Using Expectile to Estimate VaR and ES

The θ -level expectile is the EVaR at θ -level of prudence. As mentioned previously, for any given distribution and θ , there is always a τ such that the θ -level EVaR equals the τ -level QVaR. This special feature makes it possible to use EVaR to compute QVaR. Specifically, let $F(y)$ be the distribution function of asset returns Y , and for any $\tau \in (0, 1)$, let $\theta(\tau)$ be the expectile level such that $v_{\theta(\tau)}(Y) = q_\tau(Y)$. Yao and Tong (1996) showed that $\theta(\tau)$ and $q_\tau(Y)$ are related via the formula

$$\theta(\tau) = \frac{\tau q_\tau(Y) - \int_{-\infty}^{q_\tau(Y)} y dF(y)}{E[Y] - 2 \int_{-\infty}^{q_\tau(Y)} y dF(y) - (1 - 2\tau) q_\tau(Y)}. \quad (2.13)$$

Note that (2.13) depends on the density of Y as well as the values of Y beyond the quantile level. Thus, while it is possible for two assets under different returns distributions to have the same QVaR, their EVaR's will likely be different due to the dependence of expectiles on $F(y)$ and the extreme values of Y . Indeed, a key difference between QVaR and EVaR is that quantiles provide information only on the distribution of Y , but not their values. Kuan, Yeh, and Hsu (2009) gave an example where two assets with one having higher risk than the other can have the same QVaR at a common quantile level, but vastly different

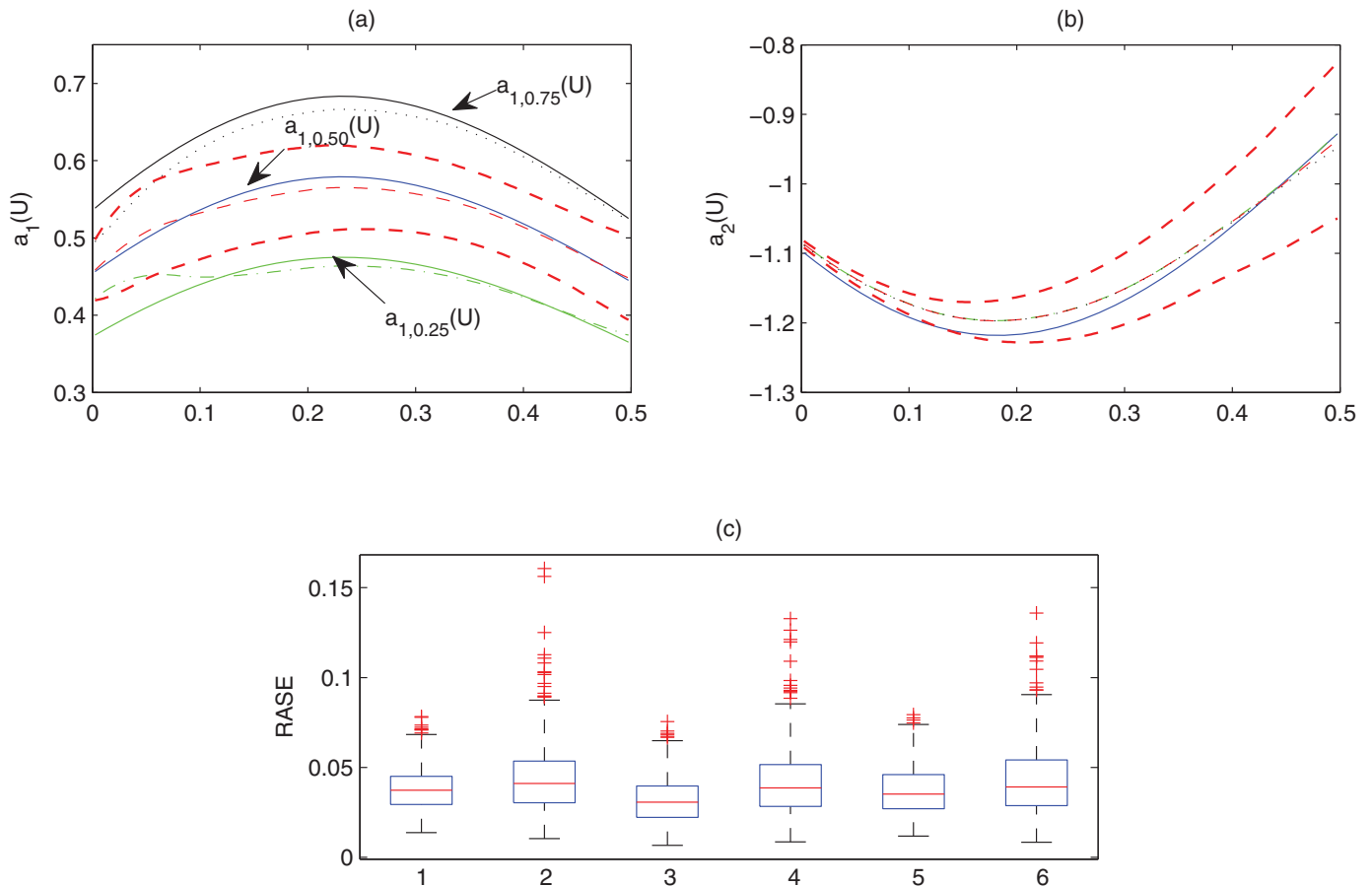


Figure 1. Parts (a) and (b) give the plots of the estimated coefficient functions for three expectiles: $\theta = 0.25$ (dashed-dotted curve), $\theta = 0.50$ (dashed curve), and $\theta = 0.75$ (dotted curve) alongside the true coefficient functions (solid curve) under Design 1. In the case of $\theta = 0.50$, the pointwise confidence bands are provided by the thick dashed curves. (c) provides the boxplots of RASE for $a_1(u)$ and $a_2(u)$ for $\theta = 0.25$ (Plots 1 and 2), $\theta = 0.50$ (Plots 3 and 4), and $\theta = 0.75$ (Plots 5 and 6) under Design 1.

EVaRs at a common expectile level. That is, the risk exposure to the tail events are different. Thus, instead of calculating the QVaR at a predetermined τ level as is commonly done, a more sensible strategy is to compute the EVaR at a given θ , then let the data reveal the corresponding tail probability and the QVaR at that level. Figure 1 and Table 1 of Kuan, Yeh, and Hsu (2009) provide information on the correspondence between θ and τ based on (2.13). For many distributions, when $\tau < (>)0.50$, $\theta < (>)\tau$ and θ is larger (smaller) for distributions with thicker tails.

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. Taylor (2008) provided a formula linking these quantities. He showed that the first-order condition resulting from the minimization of $E[Q_\theta(Y - \nu)]$ over ν may be written as

$$E(Y|Y < \nu_\theta(Y)) = \left(1 + \frac{\theta}{(1 - 2\theta)F(\nu_\theta(Y))}\right) \nu_\theta(Y) - \frac{\theta}{(1 - 2\theta)F(\nu_\theta(Y))} E(Y).$$

Table 1. Efficiency comparison of the IWLLS and one-step estimators under Design 1

T	$\theta = 0.25$				$\theta = 0.50$				$\theta = 0.75$			
	IWLLS		One-step		IWLLS		One-step		IWLLS		One-step	
	RASE ₁ (S.E.)	RASE ₂ (S.E.)	RASE ₁ (S.E.)	RASE ₂ (S.E.)	RASE ₁ (S.E.)	RASE ₂ (S.E.)	RASE ₁ (S.E.)	RASE ₂ (S.E.)	RASE ₁ (S.E.)	RASE ₂ (S.E.)	RASE ₁ (S.E.)	RASE ₂ (S.E.)
200	0.0466	0.0501	0.0466	0.0501	0.0410	0.0471	0.0410	0.0471	0.0486	0.0479	0.0486	0.0479
	0.0187	0.0278	0.0188	0.0278	0.0186	0.0270	0.0186	0.0270	0.0201	0.0286	0.0202	0.0286
400	0.0396	0.0438	0.0396	0.0439	0.0326	0.0411	0.0326	0.0411	0.0394	0.0416	0.0394	0.0416
	0.0143	0.0219	0.0144	0.0219	0.0143	0.0202	0.0143	0.0202	0.0151	0.0213	0.0151	0.0213

Now, from the relationship between expectile and quantile, we have $F(v_\theta(Y)) = \tau$. Hence, the above expression becomes

$$ES(\tau) = \left(1 + \frac{\theta}{(1-2\theta)\tau}\right)v_\theta(Y) - \frac{\theta}{(1-2\theta)\tau}E(Y). \quad (2.14)$$

This expression relates the ES associated with the τ -quantile to the corresponding θ -expectile. The expression applies to the ES in the lower tail of the distribution; the corresponding upper tail expression may be obtained by replacing τ and θ by $(1-\tau)$ and $(1-\theta)$, respectively. Taylor's (2008) empirical results showed that the ES computed based on expectiles are very close to those obtained by other methods. The conditional $ES(\tau|\cdot)$ can be obtained by substituting $v_\theta(Y)$ by $v_\theta(Y|\cdot)$.

3. ASYMPTOTIC PROPERTIES AND VARIANCE ESTIMATION

The purpose of this section is to explore the asymptotic properties of the estimators $\widehat{\beta}'_\theta$ s. In particular, we prove that the $\widehat{\beta}'_\theta$'s are consistent and asymptotically normal. We also provide an empirical method for estimating the variance of the estimator.

3.1 Asymptotic Properties

Recall that $\beta_0(u_0) = (a_1(u_0), \dots, a_p(u_0), b_1(u_0), \dots, b_p(u_0))^T$, and $\widehat{\beta}(u_0) = (\widehat{a}_1(u_0), \dots, \widehat{a}_p(u_0), \widehat{b}_1(u_0), \dots, \widehat{b}_p(u_0))^T$. Now, let $\eta(u, \mathbf{X}) = \sum_{j=1}^p a_j(u)X_{ij}$, $\mu_l = \int u^l K(u)du$, and $v_l = \int u^l K^2(u) du$, $l = 0, 1, \dots$. Further, denote $f_U(\cdot)$ as the marginal density of U , $\mathbf{a}'(u)$ and $\mathbf{a}''(u)$ the first and second derivatives of $\mathbf{a}(u)$, respectively. We have the following theorems:

Theorem 1. (Consistency) Assume that all series in $\{(Y_t, \mathbf{X}_t, U_t)\}$ are strongly stationary processes satisfying the α -mixing condition. Let $\widehat{\beta}_\theta$ be the solution to the estimating Equation (2.7) for any given $\theta \in (0, 1)$. Suppose that conditions (A.1) through (A.6) in Appendix are satisfied. Then we have, as $T \rightarrow \infty$,

$$\mathbf{H}(\widehat{\beta}_\theta - \beta_0) \rightarrow 0$$

in probability.

Corollary 1. Under the conditions of Theorem 1, we have

$$\widehat{\mathbf{a}}_\theta(u) \rightarrow \mathbf{a}(u)$$

in probability for any $u \in \mathcal{U}$, with \mathcal{U} being the support set of u .

Theorem 2. (Asymptotic Normality) Assume that all of the conditions stated in Theorem 1 are satisfied. Assume also that $F_{X,U}(\cdot)$ is the distribution of Y conditional on (\mathbf{X}, U) . Then we have, as $T \rightarrow \infty$,

$$\begin{aligned} &\sqrt{Th} \left[\mathbf{H}\{\widehat{\beta}_\theta(u_0) - \beta_0(u_0)\} - \frac{h^2}{2(\mu_0\mu_2 - \mu_1^2)} \right. \\ &\quad \times \left. \begin{pmatrix} \mu_2^2 - \mu_1\mu_3 \\ \mu_0\mu_3 - \mu_1\mu_2 \end{pmatrix} \mathbf{a}''(u_0) \right] + o(h^2) \\ &\xrightarrow{\mathcal{L}} N(0, \mathbf{D}(\theta, u_0)^{-1} \boldsymbol{\Sigma}(\theta, u_0) \mathbf{D}(\theta, u_0)^{-1}), \end{aligned} \quad (3.1)$$

where

$$\mathbf{D}(\theta, u_0) = f_U(u_0) \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \boldsymbol{\Gamma}(\theta, u_0),$$

$$\boldsymbol{\Sigma}(\theta, u_0) = f_U(u_0) \begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix} \otimes \mathbf{A}(\theta, u_0),$$

$$\mathbf{A}(\theta, u_0) = E(Q_\theta^2(\eta(u_0, \mathbf{X}))\mathbf{X}\mathbf{X}^T | U = u_0),$$

and

$$\begin{aligned} \boldsymbol{\Gamma}(\theta, u_0) &= 2E(\{[1 - F_{X,U}(\eta(u_0, \mathbf{X}))] \\ &\quad + (1-\theta)F_{X,U}(\eta(u_0, \mathbf{X}))\}\mathbf{X}\mathbf{X}^T | U = u_0). \end{aligned}$$

Corollary 2. Under the conditions of Theorem 2, we have

$$\begin{aligned} &\sqrt{Th} \left[\{\widehat{\mathbf{a}}_\theta(u_0) - \mathbf{a}_0(u_0)\} - \frac{(\mu_2^2 - \mu_1\mu_3)h^2}{2(\mu_0\mu_2 - \mu_1^2)} \mathbf{a}''(u_0) + o(h^2) \right] \\ &\xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}), \end{aligned} \quad (3.2)$$

where

$$\boldsymbol{\Sigma} = \frac{c_0^2 v_0 + 2c_0 c_1 v_1 + c_1^2 v_2}{f_U(u_0)} \boldsymbol{\Gamma}^{-1}(\theta, u_0) \mathbf{A}(\theta, u_0) \boldsymbol{\Gamma}^{-1}(\theta, u_0),$$

$$c_0 = \mu_2/(\mu_0\mu_2 - \mu_1^2) \text{ and } c_1 = -\mu_1/(\mu_0\mu_2 - \mu_1^2).$$

Corollaries 1 and 2 focus on the the estimator of the functional coefficient vector, $\widehat{\mathbf{a}}_\theta$, and are special cases of Theorems 1 and 2, respectively. By Corollary 1, $\widehat{\mathbf{a}}_\theta$ converges to the true parameter vector \mathbf{a}_0 when the sample size is sufficiently large and is thus consistent. By Corollary 2, $\widehat{\mathbf{a}}_\theta$ follows an asymptotic normal distribution. The result of Corollary 2 is useful for constructing confidence intervals of the unknowns.

Remark 1. When $\theta = 0.50$, the asymmetric squared error loss reduces to a symmetric squared error loss, and accordingly the expectile regression degenerates to the conditional mean regression. For this special case, our results are the same as those of Cai, Fan, and Li (2000), who considered the estimation of the conditional mean function using a varying-coefficient approach.

Remark 2. The conditional expectile for a given level of θ may be calculated using the estimated coefficients by virtue of Equation (2.3). This expectile measure is also the EVaR, which may be converted to a τ -level QVaR under an assumed distribution using the one-to-one relationship between expectiles and quantiles. The ES corresponding to a given EVaR may also be calculated using Taylor's (2008) procedure.

3.2 Variance Estimation

We consider estimation of the covariance matrix of $\widehat{\beta}_\theta$ by the sandwich method. To do so, we have to first estimate $\mathbf{D}(\theta, u_0)$ and $\boldsymbol{\Sigma}(\theta, u_0)$. By sample analogy, we have

$$\widehat{\mathbf{D}}(\theta, u_0) = \frac{1}{Th} \sum_{t=1}^T \mathbf{Z}_t^{*\otimes 2} Q_\theta''(Y_t - \mathbf{Z}_t^T \widehat{\beta}_\theta) K_t \quad (3.3)$$

and

$$\widehat{\Sigma}(\theta, u_0) = \frac{1}{Th} \sum_{t=1}^T Q_{\theta}^2(Y_t - \mathbf{Z}_t^T \widehat{\beta}_{\theta}) K_t^2 \mathbf{Z}_t^{*\otimes 2}, \quad (3.4)$$

where $a^{\otimes 2} = aa^T$, and $Q'_{\theta}(\cdot)$ and $Q''_{\theta}(\cdot)$ are, respectively, the first and second derivatives of the function $Q_{\theta}(\cdot)$ with respect to θ . We show in Lemmas A.1 and A.2 in the Appendix that $\widehat{\mathbf{D}}(\theta, u_0)$ and $\widehat{\Sigma}(\theta, u_0)$ are consistent estimators of their respective unknowns. That is,

$$\widehat{\mathbf{D}}(\theta, u_0) \xrightarrow{P} f_U(u_0) \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma(\theta, u_0)$$

and

$$\widehat{\Sigma}(\theta, u_0) \xrightarrow{P} f_U(u_0) \begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix} \otimes \mathbf{A}(\theta, u_0).$$

Thus, a consistent estimator of the sandwich covariance matrix of $\widehat{\beta}_{\theta}$ is

$$\widehat{\mathbf{D}}(\theta, u_0)^{-1} \widehat{\Sigma}(\theta, u_0) \widehat{\mathbf{D}}(\theta, u_0)^{-1}. \quad (3.5)$$

The asymptotic covariance matrix of $\mathbf{a}(u_0)$ is the $p \times p$ northwest submatrix of (3.5). The formulas (3.3) and (3.4) provide a simple way to estimate variance that avoids the estimation of the density function $f_U(\cdot)$ in $\mathbf{D}(\theta, u_0)$ and $\Sigma(\theta, u_0)$, which is required if one uses the alternative plug-in method.

4. SIMULATION STUDIES AND REAL DATA EXAMPLES

4.1 Simulated Studies

In this section, we focus on the finite sample properties and identify, in the context of simulation experiments, estimation and inference properties of the proposed methodology. Information is provided relating to (i) the comparative performance of the estimator based on the one-step algorithm described in Section 2.2 and that based on full iterations, and (ii) the accuracy of the proposed standard error formula (3.5). While these results are specific to the collection of particular simulation experiments analyzed, the sampling evidence does provide an indication of the type of relative performance that can occur over a range of scenarios. We use the root average squared error (RASE) to judge the quality of the parameter estimates. The RASE is measured by

$$\text{RASE}_j = \left[n_{\text{grid}}^{-1} \sum_{k=1}^{n_{\text{grid}}} \{\widehat{a}_j(u_k) - a_j(u_k)\}^2 \right]^{1/2}, \quad j = 0, 1, 2, \dots, p \quad (4.1)$$

where u_k 's, $k = 1, \dots, n_{\text{grid}}$, are the grid points at which the coefficient functions $a_j(u)$'s are estimated. In each case we report estimation sampling performance relating to the empirical median and the standard deviation of the RASE across 500 repetitions. We consider sample sizes of $T = 200, 400, 800$, and use the Epanechnikov kernel function $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ for local linear smoothing. The method discussed in Section 2.3 is used to select the bandwidth.

Our application of the one-step estimation algorithm involves the following:

- Step 1. Set $n_{\text{grid}} = 200$ and divide the interval of U into $\{u_1, u_2, \dots, u_{200}\}$.
- Step 2. Solve (2.7) by letting $\theta = 0.50$, and use the least squares estimate $\widehat{\beta}_0(u_0)$ as initial values for IWLLS estimation at the following points: $u_0 = u_{20}, u_{60}, u_{100}, u_{140}, u_{180}$.
- Step 3. Compute the IWLLS estimates at these five points and denote the estimates as $\widehat{\beta}_{\theta}(u_{20}), \widehat{\beta}_{\theta}(u_{60}), \widehat{\beta}_{\theta}(u_{100}), \widehat{\beta}_{\theta}(u_{140}),$ and $\widehat{\beta}_{\theta}(u_{180})$.
- Step 4. Calculate, using formula (2.10), the one-step estimator at u_{K-1} and u_{K+1} using $\widehat{\beta}_{\theta}(u_K)$ obtained in Step 3 as an initial value, $K = 20, 60, 100, 140, 180$, and then use the resultant estimates as the initial values for calculating $\widehat{\beta}(u_{K-2})$ and $\widehat{\beta}(u_{K+2})$ based on formula (2.10). Continue this process until all the one-step estimates at u_{n_k} for $n_k = K - 20, \dots, K + 19$ are computed.

We consider the following three experimental designs. Similar designs were used by Cai, Fan, and Yao (2000) and Cai and Xu (2008) in their simulation studies.

Design 1. Varying-coefficient EXPAR model:

$$Y_t = a_1^*(Y_{t-1})Y_{t-1} + a_2(Y_{t-1})Y_{t-2} + 2a_1^*(Y_{t-1})Y_{t-1}I(Y_{t-1} > 0)\varepsilon_t,$$

where $a_1^*(u) = 0.138 + (0.316 + 0.982u)e^{-3.89u^2}$, $a_2(u) = -0.437 - (0.659 + 1.260u)e^{-3.89u^2}$, and ε_t is iid $N(0, 0.2^2)$. We consider $\theta = 0.25, 0.50, 0.75$. The corresponding θ -level expectiles $v_{\theta}(\varepsilon)$ of ε_t are $-0.09, 0,$ and 0.09 , respectively. This means these $v_{\theta}(\varepsilon)$'s are the solutions to $E\{L_{\theta}(\varepsilon - v)\} = 0$ under the stated error distribution of ε_t . The expectile regression model is $v_{\theta}(Y_t|Y_{t-1}, Y_{t-2}) = (1 + 2v_{\theta}(\varepsilon)I(Y_{t-1} > 0))a_1^*(Y_{t-1})Y_{t-1} + a_2(Y_{t-1})Y_{t-2} \equiv a_1(Y_{t-1})Y_{t-1} + a_2(Y_{t-1})Y_{t-2}$. The optimal bandwidth, chosen by minimizing the average mean squared error (2.11) for $\theta = 0.50$, is $h_{\text{opt}} = 0.1365$ for $T = 400$, and $h_{\text{opt}} = 0.1507$ for $T = 800$.

Design 2. Varying-coefficient TAR model:

$$Y_t = a_1(Y_{t-2})Y_{t-1} + a_2(Y_{t-2})Y_{t-2} + \varepsilon_t,$$

where $a_1(u) = 0.4I(u \leq 1) - 0.8I(u > 1)$, $a_2(u) = -0.6I(u \leq 1) + 0.2I(u > 1)$, and ε_t is iid $N(0, 1)$. As in Design 1, we consider $\theta = 0.25, 0.50, 0.75$, with the corresponding θ -level expectiles $v_{\theta}(\varepsilon)$ of ε_t being $-0.44, 0,$ and 0.44 , respectively under the stated distribution of ε_t . The expectile regression model is $v_{\theta}(Y_t|Y_{t-1}, Y_{t-2}) = a_1(Y_{t-2})Y_{t-1} + a_2(Y_{t-2})Y_{t-2} + v_{\theta}(\varepsilon)$. The optimal bandwidths are chosen in the same manner as those under Design 1. Note that an equivalent expression of the varying-coefficient TAR model is the following four-covariate parametric linear model:

$$Y_t = \beta_1 I(Y_{t-2} \leq 1)Y_{t-1} + \beta_2 I(Y_{t-2} > 1)Y_{t-1} + \beta_3 I(Y_{t-2} \leq 1)Y_{t-2} + \beta_4 I(Y_{t-2} > 1)Y_{t-2} + \varepsilon_t \equiv \beta_1 X_{t,1}^* + \beta_2 X_{t,2}^* + \beta_3 X_{t,3}^* + \beta_4 X_{t,4}^* + \varepsilon_t,$$

Table 2. Accuracy of the proposed method under Design 1

T	u ₀	θ = 0.25		θ = 0.50		θ = 0.75		
		SD	SE(SE _{std})	SD	SE(SE _{std})	SD	SE(SE _{std})	
400	a ₁ (u ₀)	0.20	0.0376	0.0296 (0.0094)	0.0349	0.0286 (0.0084)	0.0366	0.0297 (0.0085)
		0.30	0.0349	0.0263 (0.0079)	0.0325	0.0255 (0.0072)	0.0336	0.0268 (0.0076)
		0.40	0.0272	0.0246 (0.0051)	0.0261	0.0237 (0.0044)	0.0278	0.0248 (0.0049)
	a ₂ (u ₀)	0.20	0.0233	0.0171 (0.0050)	0.0221	0.0166 (0.0043)	0.0233	0.0172 (0.0043)
		0.30	0.0339	0.0269 (0.0080)	0.0322	0.0261 (0.0073)	0.0338	0.0272 (0.0074)
		0.40	0.0563	0.0395 (0.0122)	0.0532	0.0384 (0.0116)	0.0564	0.0399 (0.0122)
800	a ₁ (u ₀)	0.20	0.0271	0.0223 (0.0064)	0.0254	0.0213 (0.0057)	0.0270	0.0226 (0.0064)
		0.30	0.0231	0.0194 (0.0051)	0.0217	0.0185 (0.0046)	0.0228	0.0194 (0.0048)
		0.40	0.0187	0.0182 (0.0034)	0.0177	0.0173 (0.0029)	0.0185	0.0182 (0.0031)
	a ₂ (u ₀)	0.20	0.0176	0.0130 (0.0035)	0.0168	0.0125 (0.0031)	0.0173	0.0132 (0.0034)
		0.30	0.0224	0.0201 (0.0050)	0.0213	0.0192 (0.0046)	0.0223	0.0202 (0.0051)
		0.40	0.0404	0.0302 (0.0086)	0.0375	0.0288 (0.0078)	0.0386	0.0301 (0.0081)

where $(\beta_1, \beta_2, \beta_3, \beta_4) = (0.4, -0.8, -0.6, 0.2)$. The parameters $\beta_j, j = 1, \dots, 4$ can be estimated by the one-step algorithm described in Kuan, Yeh, and Hsu (2009).

Design 3. Varying-coefficient model with exogenous regressors:

$$Y_t = a_1(U_t)Y_{t-1} + a_2(U_t)Y_{t-2} + \varepsilon_t,$$

where $a_1(U_t) = \sin(\sqrt{2}\pi U_t), a_2(U_t) = \cos(\sqrt{2}\pi U_t), U_t$'s are generated from an independent Uniform(0,3) distribution, and ε_t 's are obtained from an independent $N(0, 1/\sqrt{1-P})$ distribution with probability $1-P$ or from $N(c, 1/\sqrt{P})$ with probability P . This choice of error distribution, which follows Kuan, Yeh, and Hsu (2009), enables the examination of the relative performance of QVaR and EVaR in the presence of catastrophic losses. The expectile regression is $v_\theta(Y_t|Y_{t-1}, Y_{t-2}, U_t) = a_1(U_t)Y_{t-1} + a_2(U_t)Y_{t-2} + v_\theta(\varepsilon_t)$. We set $c = -5$ and $P = 0.05$. The optimal bandwidths are chosen in the same manner as under the previous two designs.

Table 1 compares the one-step and IWLLS methods under the set-up of the varying-coefficient EXPAR model in Design 1. It reports the median and the standard deviation (shown in brackets) of the RASE's. As seen from the table, the one-step and iterative methods exhibit very comparable sampling performance, and in all cases the use of the one-step method as an alternative to the IWLLS method does not result in any loss of efficiency. Figure 1(a) and (b) depicts graphically the estimated coefficient functions $a_1(U)$ and $a_2(U)$ by the one-step estimator for the three expectiles $\theta = 0.25$ (dashed-dotted curve), 0.50 (dashed curve), and 0.75 (dotted curve) for $T = 400$ with $h_{opt} = 0.1365$. In the case of $\theta = 0.50$, the 95% point-wise confidence bands of the coefficients without bias corrections are also depicted (thick dashed curves). In all cases, the estimated coefficient functions are very close to the true coefficient function (solid curve). Figure 1(c) gives the boxplots of the RASE for the three expectiles. These plots provide a measure of uncertainty associated with the one-step estimator of the coefficient functions. Note that the RASEs and their volatility are smaller for expectile level $\theta = 0.50$ than for $\theta = 0.25, 0.75$ due to the larger amount of data available at $\theta = 0.50$.

Table 2 reports results relating to the performance of the sandwich method for constructing standard errors under the EXPAR model of Design 1. In the table, SD is the standard deviation of a given estimated coefficient function across 500 replications. The mean and standard deviation of 500 estimated standard errors, denoted by SE and SE_{std}, respectively, provide information on the accuracy of the sandwich method based on formula (3.5). To conserve space, the results are presented for the points $u_0 = 0.20, 0.30, 0.40$. We find that the standard errors typically underestimate the true standard deviation, although the differences are generally slight, being of a magnitude within two standard deviations of the simulation errors. As expected, the bias decreases as the number of local data points Th increases.

All of the general comments above apply to Designs 2 and 3 in broad terms; in particular, we have not found any discernable difference in the efficiency achieved by the one-step and full iterative methods. On the other hand, in most cases, the sandwich tends to slightly underestimate the true standard errors of the estimates. Tables 3 and 4 summarize these results. Figures 2 and 3 provide the plots of the estimated coefficient functions and boxplots of the RASE for the case of $T = 400$ with the bandwidths being chosen by minimizing the average mean squared error (2.11). For brevity, the comparison results between the one-step and IWLLS estimators under Designs 2 and 3 are not shown but they are largely similar to those reported under Design 1.

As mentioned above, the varying-coefficient TAR model under Design 2 has an equivalent parametric representation. To assess the performance of the proposed varying-coefficient method compared to the parametric method, we consider the following ratio of the AMS defined in (2.12) for $m = T$ and $k = 1$:

$$RAMS = \frac{AMS_0}{AMS_1},$$

where $AMS_0 = \frac{1}{T} \sum_{t=1}^T Q_\theta\{Y_t - \sum_{j=1}^4 \hat{\beta}_j X_{t,j}^*\}$, with $\hat{\beta}_j$'s being the estimates based on the parametric model obtained using the one-step algorithm of Kuan, Yeh, and Hsu (2009), and $AMS_1 = \frac{1}{T} \sum_{t=1}^T Q_\theta\{Y_t - \sum_{j=1}^p \hat{a}_j(U_t) X_{t,j}\}$. For the chosen θ values of 0.25, 0.50, and 0.75, the RAMS takes on 0.9334, 0.9620, and 1.0078, respectively. These ratios are all very close to one, implying that the proposed varying-coefficient method is

Table 3. Accuracy of the proposed method under Design 2

T	u ₀	θ = 0.25		θ = 0.50		θ = 0.75		
		SD	SE(SE _{std})	SD	SE(SE _{std})	SD	SE(SE _{std})	
400	a ₁ (u ₀)	0.75	0.1412	0.1192 (0.0308)	0.1350	0.1168 (0.0253)	0.1421	0.1165 (0.0271)
		1.00	0.1645	0.1317 (0.0375)	0.1600	0.1306 (0.0348)	0.1710	0.1302 (0.0370)
		1.25	0.1536	0.1208 (0.0365)	0.1434	0.1192 (0.0328)	0.1468	0.1176 (0.0331)
	a ₂ (u ₀)	0.75	0.2154	0.1861 (0.0403)	0.2051	0.1786 (0.0334)	0.2165	0.1822 (0.0365)
		1.00	0.1896	0.1569 (0.0367)	0.1791	0.1525 (0.0335)	0.1871	0.1576 (0.0392)
		1.25	0.1541	0.1262 (0.0313)	0.1501	0.1229 (0.0277)	0.1576	0.1247 (0.0285)
800	a ₁ (u ₀)	0.75	0.0963	0.0878 (0.0185)	0.0953	0.0853 (0.0169)	0.1031	0.0887 (0.0190)
		1.00	0.1069	0.0978 (0.0217)	0.1008	0.0955 (0.0198)	0.1046	0.0988 (0.0224)
		1.25	0.1157	0.0888 (0.0215)	0.1064	0.0858 (0.0201)	0.1073	0.0871 (0.0194)
	a ₂ (u ₀)	0.75	0.1475	0.1347 (0.0251)	0.1399	0.1296 (0.0238)	0.1480	0.1364 (0.0277)
		1.00	0.1273	0.1144 (0.0231)	0.1197	0.1100 (0.0212)	0.1262	0.1155 (0.0232)
		1.25	0.1095	0.0914 (0.0181)	0.0992	0.0878 (0.0155)	0.1011	0.0910 (0.0165)

as efficient as the parametric method. This result is encouraging given that the (true) parametric form of the model is practically unknown.

We also evaluate the sensitivity of EVaR and QVaR to catastrophic events. We do so in the context of Design 3, and consider the following two cases: Case (a): $P = 0.01$ and $\tau = \theta = 0.05$; Case (b): $P = \tau = \theta = 0.01$. In both cases, we let $c = [-1, -50]$ and $T = 400$ and the number of replications be 500. Figure 4 plots the EVaR and QVaR for the estimated errors $\hat{\varepsilon}_t = Y_t - \hat{a}_1(U_t)Y_{t-1} - \hat{a}_2(U_t)Y_{t-1}$. Note that although the coefficient estimators $\hat{a}_1(\cdot)$ and $\hat{a}_2(\cdot)$ are obtained based on $\theta = 0.50$, they can also be obtained using other values of θ as under Design 2, the coefficients $a_1(\cdot)$ and $a_2(\cdot)$ do not depend on θ . The figures show very similar results to those shown in Figure 2 of Kuan, Yeh and Hsu (2009). Specifically, When $P < \tau$, EVaR varies with c , but the corresponding QVaR is relatively insensitive to the magnitude of extreme losses based on the error distribution. Although QVaR changes with c when $P \geq \tau$, its magnitude is smaller than that of the EVaR for all c . These suggest that the EVaR is a more sensitive risk measurement to catastrophic losses than the QVaR.

4.2 Real Data Examples

Example 1.

This example is based on 2348 daily closing bid prices of the Euro in terms of the U.S. dollar between January 1, 2004, and December 31, 2012. We denote the bid price as p_t , and compute the daily returns Y_t as 100 times the difference of the log of prices, that is, $Y_t = 100 \log(p_t/p_{t-1})$.

We let the θ -level expectile be modeled by

$$v_\theta(Y_t|Y_{t-1}, Y_{t-2}, U_t) = a_{0,\theta}(U_t) + a_{1,\theta}(U_t)Y_{t-1} + a_{2,\theta}(U_t)Y_{t-2}, \tag{4.2}$$

and denote the corresponding varying-coefficient model as VC(2). Fan, Yao, and Cai (2003) and Cai and Xu (2008) also considered the modeling of exchange rate data by a varying-coefficient approach, although they did not apply their methods to expectile estimation. Following these authors, we let $U_t = p_{t-1}/M_t - 1$ be the effect modifier, where $M_t = \sum_{j=1}^L p_{t-j}/L$ is a moving average of time and can be used to proxy the trend at time, and L is the total number of periods in the moving

Table 4. Accuracy of the proposed method under Design 3 for $P = 0.05$ and $c = -5$

T	u ₀	θ = 0.25		θ = 0.50		θ = 0.75		
		SD	SE(SE _{std})	SD	SE(SE _{std})	SD	SE(SE _{std})	
400	a ₁ (u ₀)	0.75	0.0805	0.0424 (0.0335)	0.0623	0.0346 (0.0242)	0.0538	0.0302 (0.0181)
		1.50	0.0786	0.0423 (0.0305)	0.0599	0.0351 (0.0221)	0.0514	0.0309 (0.0180)
		2.25	0.0772	0.0429 (0.0302)	0.0593	0.0349 (0.0214)	0.0522	0.0304 (0.0177)
	a ₂ (u ₀)	0.75	0.0880	0.0434 (0.0319)	0.0692	0.0361 (0.0234)	0.0618	0.0318 (0.0190)
		1.50	0.0814	0.0436 (0.0340)	0.0649	0.0364 (0.0243)	0.0598	0.0316 (0.0197)
		2.25	0.0876	0.0441 (0.0321)	0.0695	0.0361 (0.0226)	0.0628	0.0311 (0.0183)
800	a ₁ (u ₀)	0.75	0.0550	0.0333 (0.0262)	0.0403	0.0254 (0.0180)	0.0339	0.0216 (0.0142)
		1.50	0.0494	0.0322 (0.0225)	0.0365	0.0251 (0.0148)	0.0319	0.0215 (0.0114)
		2.25	0.0502	0.0317 (0.0223)	0.0375	0.0249 (0.0154)	0.0328	0.0216 (0.0123)
	a ₁ (u ₀)	0.75	0.0569	0.0331 (0.0232)	0.0450	0.0259 (0.0168)	0.0397	0.0223 (0.0145)
		1.50	0.0600	0.0323 (0.0228)	0.0449	0.0255 (0.0161)	0.0392	0.0221 (0.0131)
		2.25	0.0509	0.0308 (0.0218)	0.0382	0.0246 (0.0154)	0.0334	0.0215 (0.0126)

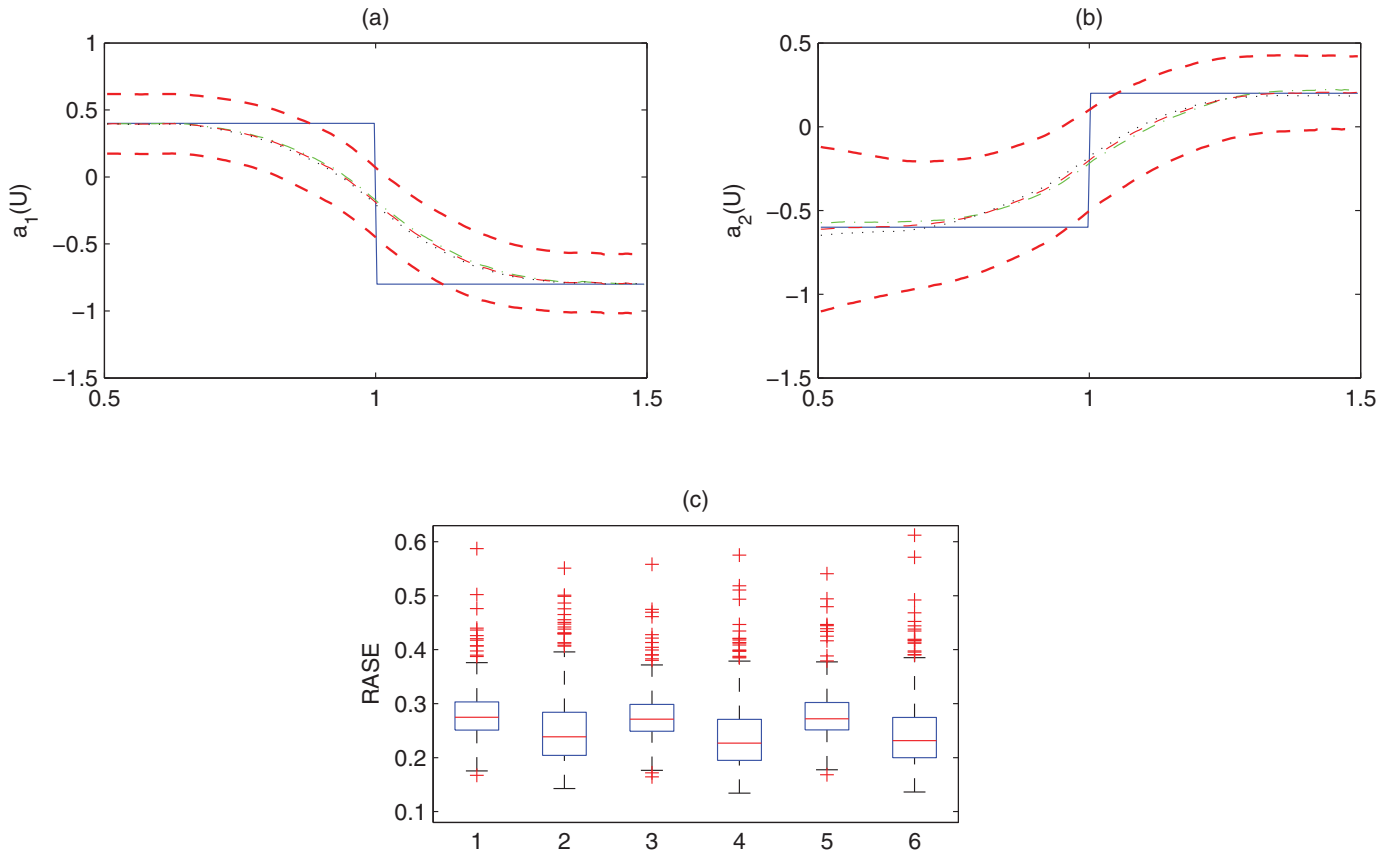


Figure 2. Parts (a) and (b) give the plots of the estimated coefficient functions for three expectiles: $\theta = 0.25$ (dashed-dotted curve), $\theta = 0.50$ (dashed curve), and $\theta = 0.75$ (dotted curve) alongside the true coefficient functions (solid curve) under Design 2. In the case of $\theta = 0.50$, the pointwise confidence bands are provided by the thick dashed curves. (c) provides the boxplots of RASE for $a_1(u)$ and $a_2(u)$ for $\theta = 0.25$ (Plots 1 and 2), $\theta = 0.50$ (Plots 3 and 4), and $\theta = 0.75$ (Plots 5 and 6) under Design 2.

average. We choose $L = 10$ to account for a two-week period. As discussed in Fan, Yao, and Cai (2003) and Cai and Xu (2008), this choice of U_t is a moving average technical trading rule commonly used in finance, and M_t a proxy for the trend at time $t - 1$.

We use the first 1500 observations for model estimation and the remaining observations for out-of-sample forecast evaluation, and adopt the Epanechnikov kernel function $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ for local linear smoothing. For bandwidth selection we set $\theta = 0.01, 0.05, 0.50, 0.95$, and use the method described in Section 2.3 with $m = [0.1 \times T] = 150$ and $H = 4$; we choose $m = [0.1 \times T]$ instead of $m = 1$ to alleviate computational burden. By minimizing the AMS in (2.11), we obtain optimal bandwidth values of $h_{opt} = 0.023, 0.028, 0.032, 0.031$ corresponding to the aforementioned θ values.

For comparison purposes, we also consider the following SQ(2) and ABS(2) parametric models used in Kuan's (2009) empirical analysis:

SQ(2) model:

$$Y_t = a_0(\theta) + a_1(\theta)Y_{t-1} + b_1(\theta)(Y_{t-1}^+)^2 + \gamma_1(Y_{t-1}^-)^2 + b_2(\theta)(Y_{t-2}^+)^2 + \gamma_2(Y_{t-2}^-)^2 + e_t(\theta)$$

ABS(q) model:

$$Y_t = a_0(\theta) + \delta_1(\theta)Y_{t-1}^+ + \lambda_1(\theta)Y_{t-1}^- + \delta_2(\theta)Y_{t-2}^+ + \lambda_2(\theta)Y_{t-2}^- + e_t(\theta).$$

Table 5 reports results corresponding to $\theta = 0.01, 0.05, 0.10$, where τ_{in} and τ_{out} are the in-sample and out-of-sample tail

Table 5. Comparison results for Example 1

	$\theta = 10\%$			$\theta = 5\%$			$\theta = 1\%$		
	VC(2)	SQ(2)	ABS(2)	VC(2)	SQ(2)	ABS(2)	VC(2)	SQ(2)	ABS(2)
τ_{in}	17.13%	17.07%	17.00%	10.53%	10.53%	10.73%	3.47%	3.07%	3.20%
τ_{out}	19.23%	19.45%	19.45%	13.28%	14.13%	13.39%	3.40%	3.93%	4.04%
$ \tau_{out} - \tau_{in} /\tau_{in}$	12.27%	13.95%	14.40%	26.11%	34.18%	24.75%	1.90%	28.22%	26.20%
Out-of-sample θ	11.46%	12.41%	11.68%	5.72%	6.50%	5.99%	0.73%	0.85%	0.91%

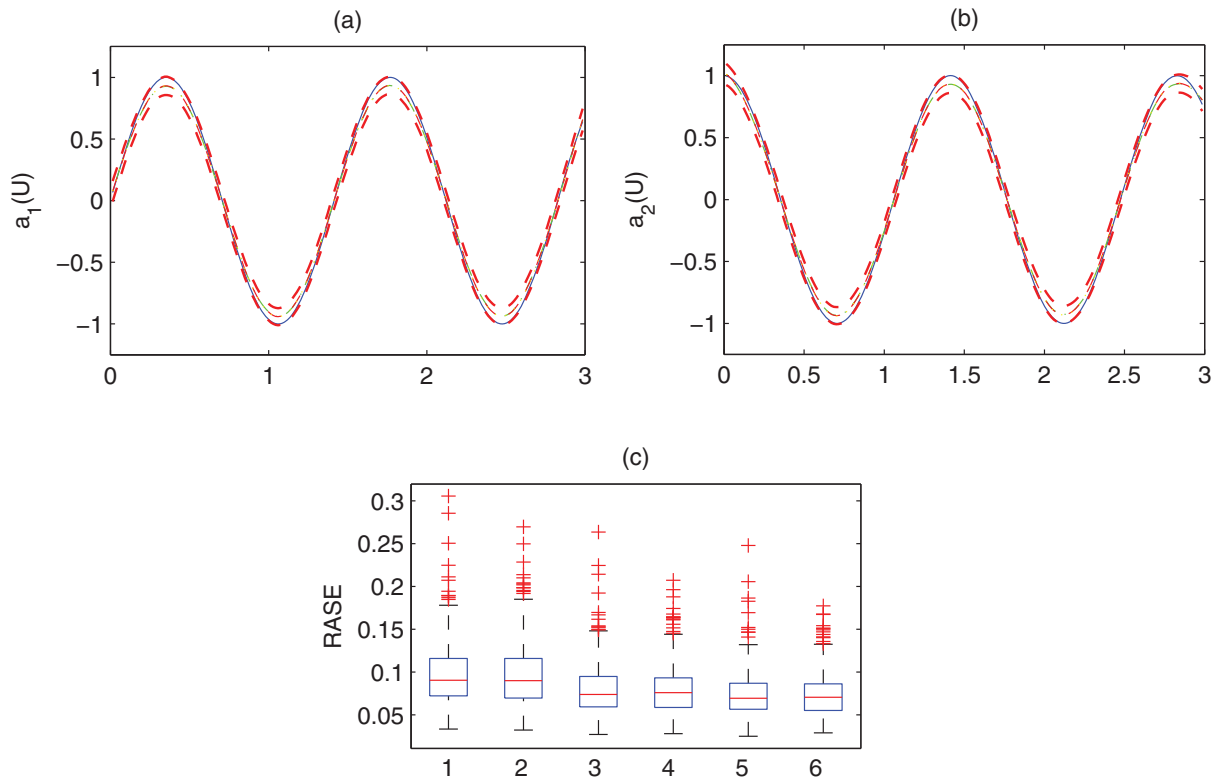


Figure 3. Parts (a) and (b) give the plots of the estimated coefficient functions for three expectiles: $\theta = 0.25$ (dashed-dotted curve), $\theta = 0.50$ (dashed curve) and $\theta = 0.75$ alongside the true coefficient functions (solid curve) under Design 3. In the case of $\theta = 0.50$, the pointwise confidence bands are provided by the thick dashed curves. (c) provides the boxplots of RASE for $a_1(u)$ and $a_2(u)$ for $\theta = 0.25$ (Plots 1 and 2), $\theta = 0.50$ (Plots 3 and 4), and $\theta = 0.75$ (Plots 5 and 6) under Design 3 with $P = 0.05$ and $c = -5$.

probabilities for the estimated expectiles, respectively. We also consider larger values of θ , but as the results are qualitatively similar to those reported in Table 5 we omit them for brevity. We observe from the table that for any given θ value, which represents prudence, all three models invariably produce larger τ_{in} and τ_{out} than θ , and larger τ_{out} than τ_{in} . The former observation indicates that the QVaR, if calculated at the same level

as the desired level of prudence, will likely underestimate value at risk. The latter observation is of no surprise, given the higher level of volatility of Y_t in the out-of-sample period, as reflected in Figure 5(a). That said, among the three models, the VC(2) model yields the smallest $|\tau_{out} - \tau_{in}|/\tau_{in}$ for any given θ . This may be taken as an indication that the VC(2) model produces more stable estimates than the SQ(2) and ABS(2) models.

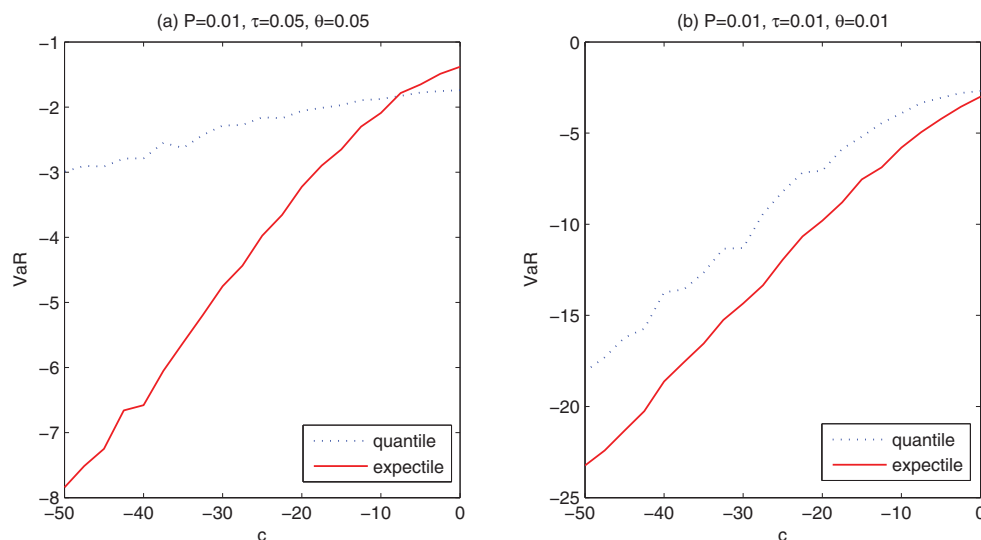


Figure 4. The catastrophic loss sensitivity of QVaR and EVaR.

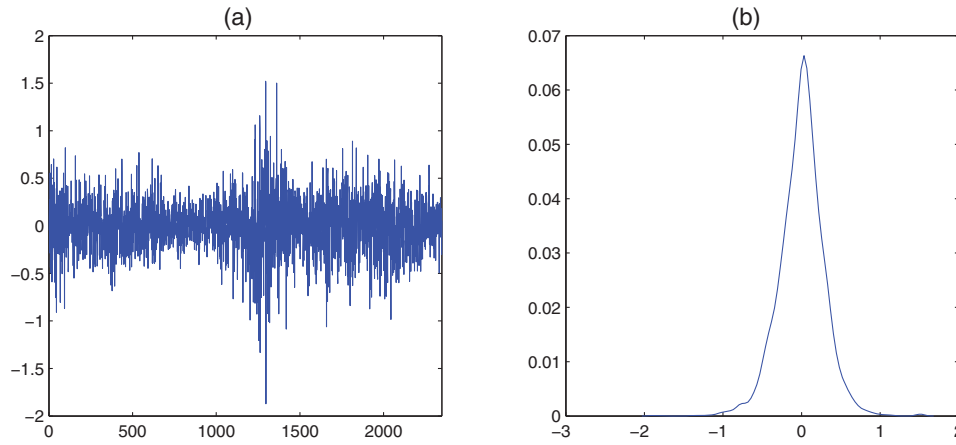


Figure 5. (a) Scatterplot of Y_t ; (b) the estimated density function of Y_t for Example 1.

Comparing the τ_{in} and τ_{out} values shown in Table 5 against the tail probabilities under various distributions for different values of θ as shown in Table 1 of Kuan, Yeh, and Hsu (2009), it appears that the returns series Y_t is best characterized by a Student's t distribution with 10 degrees of freedom. Of the three models, the VC(2) produces the best out-of-sample θ es-

timates when $\theta = 0.05$ and 0.10 , but the worst estimate when $\theta = 0.01$.

Figure 6(a)–(c) presents plots of the estimated coefficient functions for $\theta = 0.01$ (solid curve), $\theta = 0.05$ (dashed curve), $\theta = 0.50$ (dashed-dotted curve), and $\theta = 0.95$ (dotted curve). We observe from Figure 6(a) that the four expectile curves are

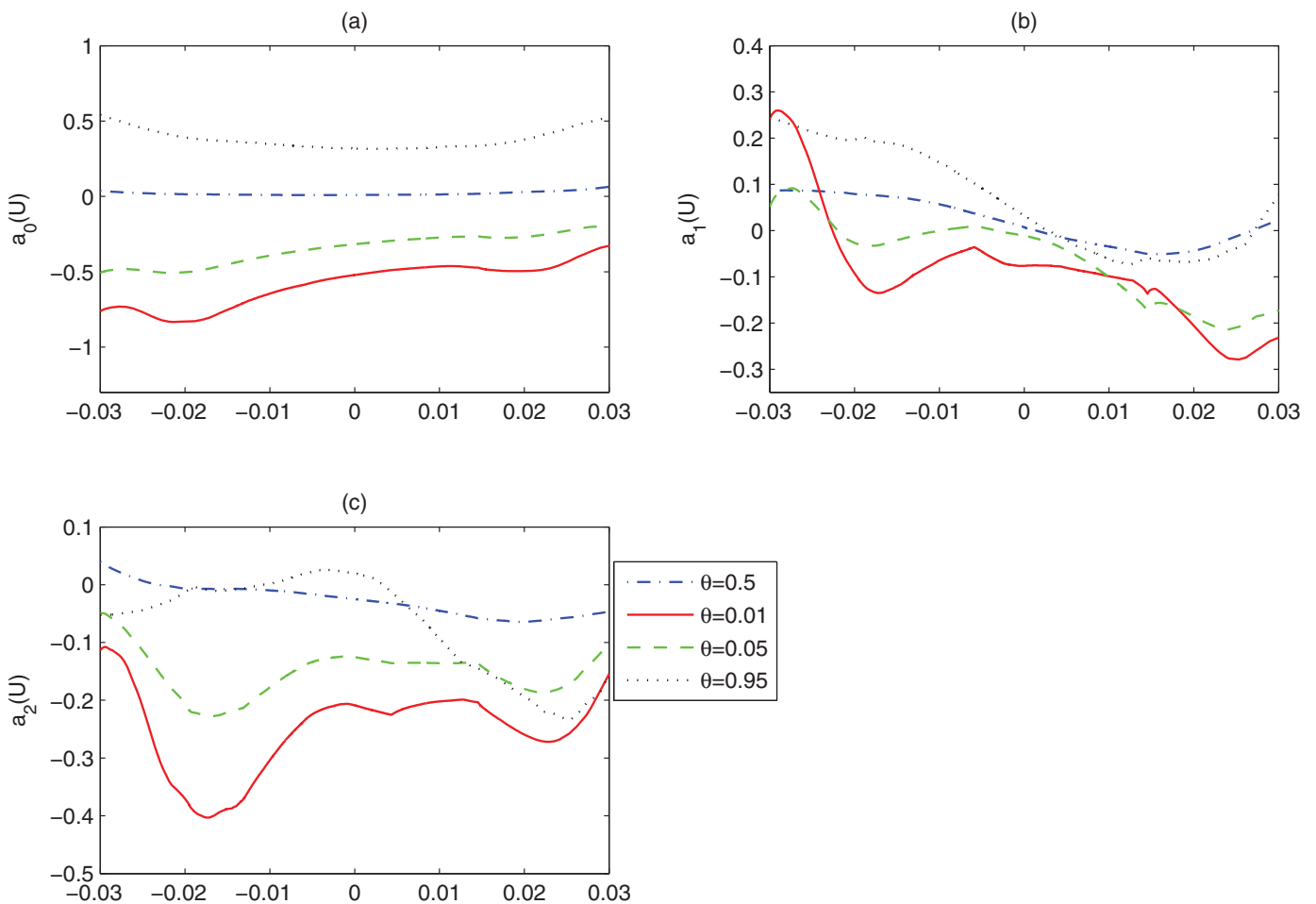


Figure 6. The estimated coefficient functions for three expectiles: $\theta = 0.01$ (solid curve), $\theta = 0.05$ (dashed curve), $\theta = 0.50$ (dashed-dotted curve), and $\theta = 0.95$ (dotted curve) for Example 1.

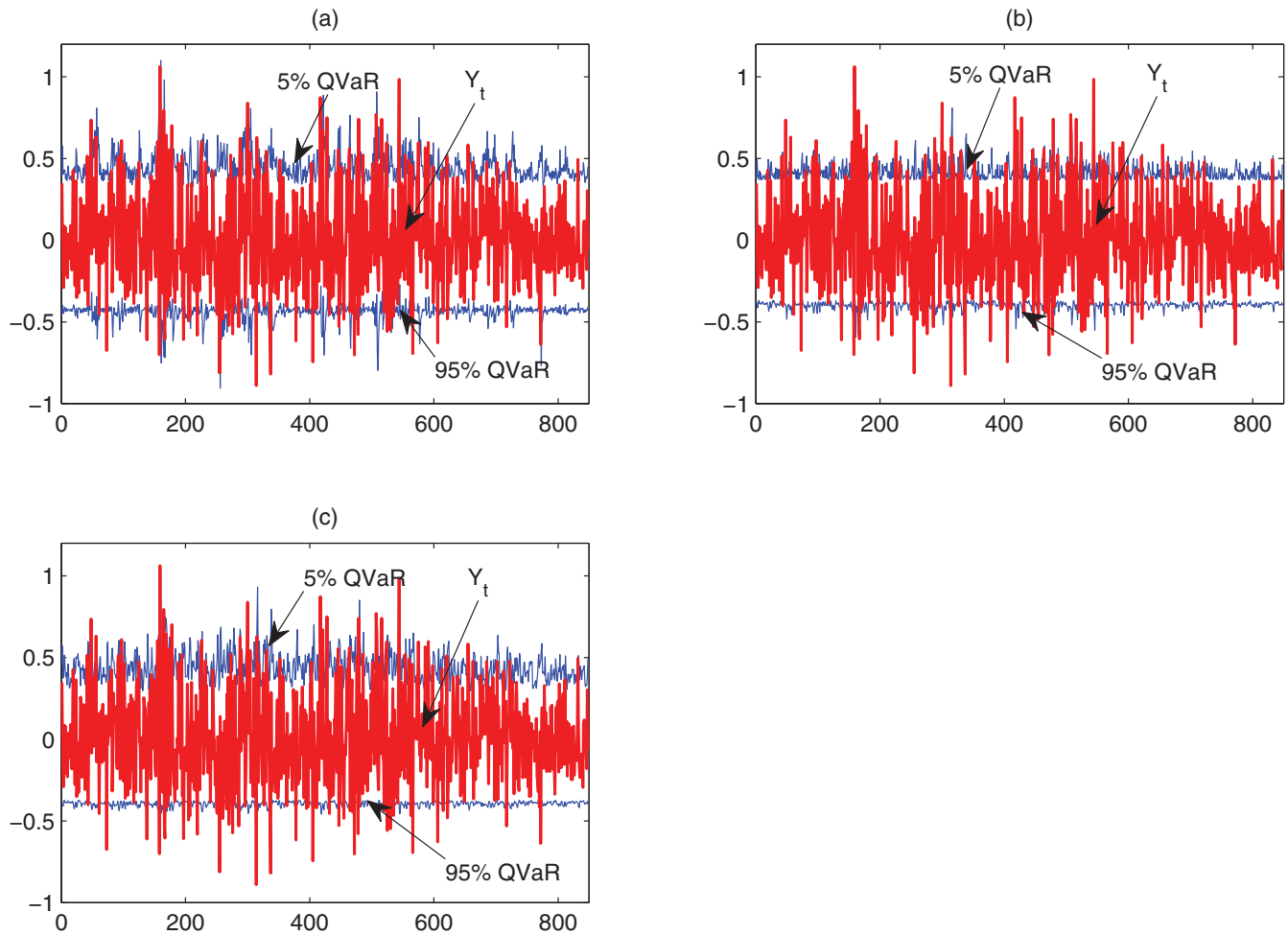


Figure 7. Plot of Y_t and the estimated 5% and 95% VaR based on the (a) VC(2), (b) SQ(2), and (c) ABS(2) models for the 848 out-of-sample observations of Example 1.

almost parallel to one another, and the estimated functions $\hat{a}_{0.05}(\cdot)$ and $\hat{a}_{0.95}(\cdot)$ exhibit a mild quadratic symmetry at $U = 0$. This implies that the intercept function $\hat{a}_{0,\theta}(\cdot)$ depends on θ . On the other hand, Figure 6(b) and (c) shows that for $\theta = 0.01$ and $\theta = 0.05$, the estimated functions have a slight W-shape, and for $\theta = 0.95$, the estimated functions behave like an asymmetrical parabola. Indeed, the strong asymmetry of some of the estimated functions is a notable feature of the results. For example, Figure 6(b) shows that $\hat{a}_{0.95}(\cdot)$ has a large positive value when $U < -0.25$, but a smaller negative value when $U > 0.2$. As

well, the estimated functions at different θ levels frequently intersect one another. Furthermore, Figure 6(b) and (c) reveals that Y_{t-1} and Y_{t-2} mostly have a negative impact on the conditional expectile except for levels of θ higher than 0.5.

We also apply the VaRs estimated from the three models to construct forecast intervals for the last 848 observations. This entails the application of the correspondence between expectile and quantile as described in Section 2.4. Specifically, we first apply kernel smoothing to estimate the density function $f(y)$ using the first 1500 observations. This allows us to calculate $E(Y)$, $q_\tau(Y)$, and $\int_{-\infty}^{q_\tau(Y)} y dF(y)$ in (2.13) and the corresponding expectile levels $\theta(\tau)$ for different τ 's. Then we estimate $v_\theta(Y|\cdot)$, the conditional EVaR at the various expectile levels, by the varying-coefficient model. As discussed previously, $v_\theta(Y|\cdot)$ is identical to $q_\tau(Y|\cdot)$ at the τ -level. For example, when $\tau = 0.05$, $\theta(0.05) = 0.0216$ (based on $h_{opt} = 0.023$), and when $\tau = 0.95$, $\theta(95\%) = 0.9801$ (based on $h_{opt} = 0.031$). Figure 7 (a)–(c) presents the estimated 5% and 95% VaR using the VC(2), SQ(2), and ABS(2) models, respectively, together with the actual observations. The estimated 90% prediction interval of the VC(2) model contains 87.26% of the observations. This is closer to the predetermined 90% and higher

Table 6. Descriptive statistics for Example 2

	Mean	St.dev.	t-stat	Min	Max	Size
SHCOMP	-0.0088	0.8487	-0.3821	-4.0199	3.9236	1358
Monday	0.0467	1.0362	1.6608**	-3.7427	3.3958	253
Tuesday	-0.1192	0.7821	-5.6165***	-4.0199	1.7777	278
Wednesday	0.0512	0.8480	2.2250**	-2.9195	3.1921	279
Thursday	-0.0556	0.7835	-2.6151***	-2.9357	3.8598	277
Friday	0.0387	0.7736	1.8435**	-2.3590	3.9236	271

NOTE: “***” indicates significance at the 1% level; “**” indicates significance at 5% level.

Table 7. (Out-of-sample) 95% prediction interval for Example 2

Date	True value	Prediction interval	Date	True value	Prediction interval
2012.8.20	-0.1632	(-2.3280 1.4136)	2012.9.07	1.5763	(-0.7385 1.4195)
2012.8.21	0.2325	(-2.0648 1.0674)	2012.9.10	0.1453	(-2.2653 1.6579)
2012.8.22	-0.2170	(-1.9747 1.7835)	2012.9.11	-0.2927	(-2.7469 1.0321)
2012.8.23	0.1103	(-1.7477 1.6779)	2012.9.12	0.1227	(-2.0284 1.3820)
2012.8.24	-0.4331	(-1.5740 1.4513)	2012.9.13	-0.3316	(-1.6709 1.7300)
2012.8.27	-0.7622	(-2.1518 1.8912)	2012.9.14	0.2763	(-0.8578 1.3851)
2012.8.28	0.3670	(-2.1233 1.0751)	2012.9.17	-0.9373	(-2.2697 1.6482)
2012.8.29	-0.4193	(-2.0212 1.5464)	2012.9.18	-0.3979	(-2.5024 1.0798)
2012.8.30	-0.1210	(-1.7136 1.6946)	2012.9.19	0.1744	(-2.0077 1.5826)
2012.8.31	0.2459	(-1.5611 1.4764)	2012.9.20	-0.9125	(-1.6980 1.7017)
2012.9.04	-0.3281	(-1.4903 0.9183)	2012.9.21	0.0397	(-1.5574 1.6511)
2012.9.05	-0.1270	(-2.0101 1.5741)	2012.9.24	0.1391	(-2.3343 1.2837)
2012.9.06	0.3024	(-1.8279 1.6235)			

than the corresponding 84.32% and 85.14% achieved by the SQ(2) and ABS(2)-based prediction intervals.

Example 2. In this example, we use dummy variables representing the five weekdays and the returns of the S&P index as

exogenous variables to model the returns of the Shanghai Stock Exchange Composite (SHCOMP) Index. Our data are based on daily observations between January 4, 2007 and September 24, 2012, totaling 1358 observations. We denote p_t and p_t^* as the daily closing values of the SHCOMP and S&P indices, and

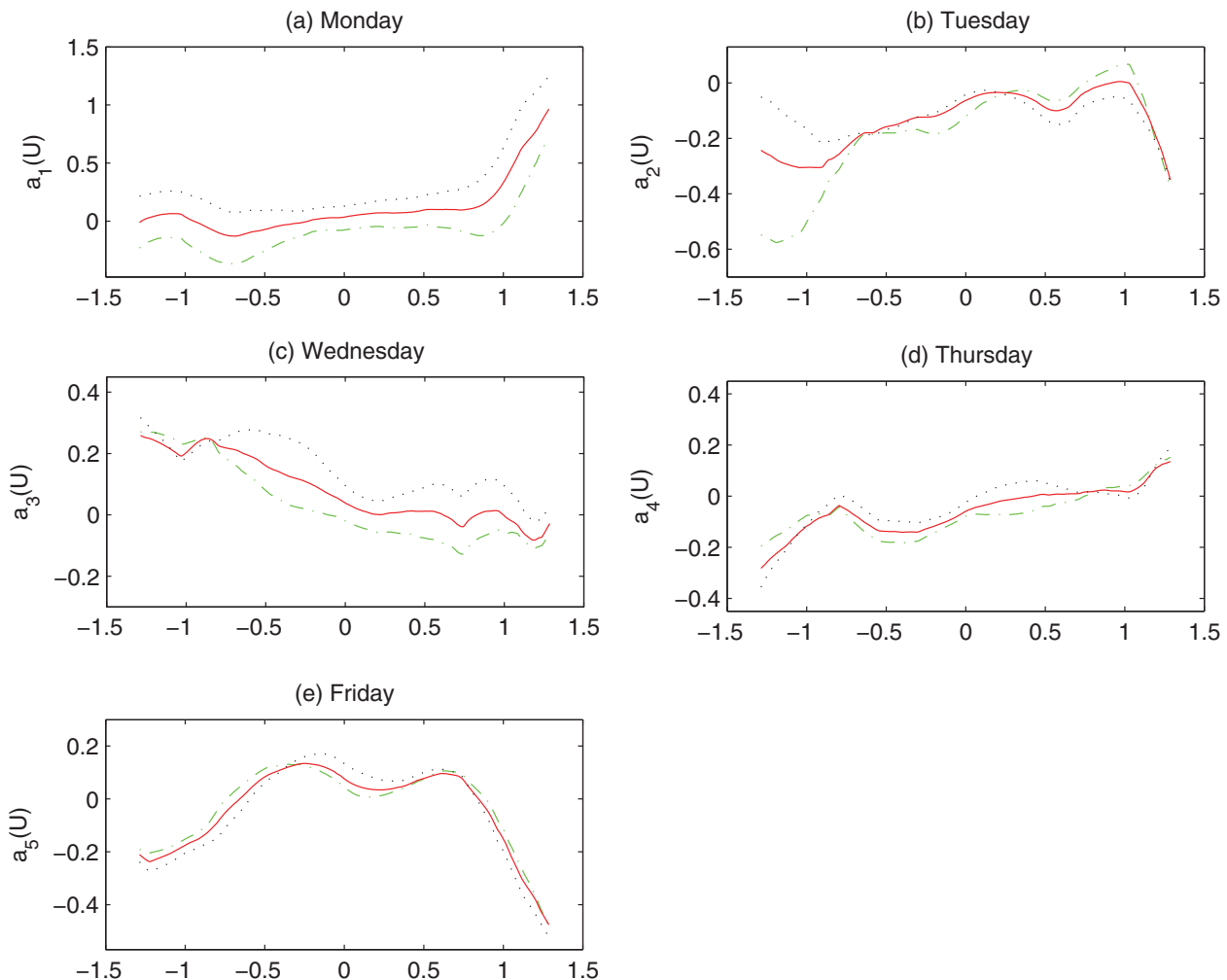


Figure 8. The estimated coefficient functions for three expectiles: $\theta = 0.25$ (dashed-dotted curve), $\theta = 0.50$ (solid curve), and $\theta = 0.75$ (dotted curve) for Example 2.

define the corresponding daily returns as $Y_t = 100 \log(p_t/p_{t-1})$ and $U_t = 100 \log(p_t^*/p_{t-1}^*)$. We let the expectile of Y_t be modeled by

$$\begin{aligned} v_\theta(Y_t|D1_t, D2_t, D3_t, D4_t, D5_t, U_{t-1}) \\ = a_{0,\theta}(U_{t-1}) + a_{1,\theta}(U_{t-1})D1_t + a_{2,\theta}(U_{t-1})D2_t \\ + a_{3,\theta}(U_{t-1})D3_t + a_{4,\theta}(U_{t-1})D4_t + a_{5,\theta}(U_{t-1})D5_t, \end{aligned} \tag{4.4}$$

where $D_i, i = 1, \dots, 5$, are the dummy variables for the five weekdays. The inclusion of these dummy variables allows the assessment of the day-of-the-week effect. The descriptive statistics and t -test results presented in Table 6 show that Monday, Wednesday, and Friday are generally characterized by significant positive returns, while the opposite is observed for Tuesday and Thursday. We use the first 1333 observations for estimation and the remaining 25 observations for forecast evaluation. As in the last example, we use the Epanechnikov kernel function $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ for local linear smoothing, and the bandwidth selection method described in Section 2.3 with $m = [0.1 \times T] = 133$ and $H = 4$, resulting in a selected bandwidth of $h_{opt} = 0.45$.

Figure 8(a)–(e) presents the estimated coefficient functions $a_j(U_t), j = 1, \dots, 5$ for $\theta = 0.25, 0.50$, and 0.75 . In all cases the day-of-the-week effect on Y_t varies with the S&P500 returns. This is most noticeable with the Wednesday and Friday effects, as shown in Figure 8(a) and (e). Also, Figure 8(b) and (d) reveals negative Tuesday and Thursday effects, while Figure 8(e) reveals a mild “M-shaped” nonlinear Friday effect. The estimated coefficient functions generally do not exhibit any significant variations across the three expectiles. Moreover, Figure 8(a) and (c) shows that the day-of-the-week effects on daily returns can be different at different levels of returns. For example, there is a positive Monday effect for high daily returns ($\theta = 0.75$), but a negative Monday effect for low daily returns ($\theta = 0.25$) when $U_t < 1.0$.

Table 7 gives the 95% prediction intervals of the last 25 observations. By setting $\tau = 0.025$ and $\tau = 0.0975$, we have $\theta(0.025) = 0.0062$ and $\theta(0.975) = 0.9907$ based on formula (2.13) that relates expectiles to quantiles. We construct these intervals using the method as described in the previous example. Table 7 shows that in 24 out of 25 cases the prediction intervals contain the true observations. This indicates that the proposed method is effective.

5. CONCLUDING REMARKS

We have developed the varying-coefficient expectile regression model which allows expectiles to be modeled using covariates in a flexible manner. We used a local linear smoothing technique for the estimation of coefficient functions and adopted a one-step weighted local least squares algorithm for computing the estimates. Our results provided evidences that EVaRs are more sensitive than QVaRs to catastrophic events which are often subjects of concern to practitioners and policy makers. Our numerical results revealed that the proposed varying-coefficient expectile model frequently outperforms the existing parametric

models in terms of estimation accuracy and the proposed one-step algorithm works well. It is hoped that the method presented in this article will find applications in future empirical studies. More recent work of the authors is looking to develop other model specifications and downside risk measures that can better assess the dynamic behavior of tail risk.

APPENDIX A: ASSUMPTIONS AND PROOFS OF TECHNICAL RESULTS

Assumptions:

- (A.1) $a_j(u)$ is twice continuously differentiable in $u \in \mathcal{U}, j = 0, 1, \dots, p$.
- (A.2) The kernel function $K(\cdot)$ is a bounded nonnegative symmetric function with compact support.
- (A.3) All processes in $\{(Y_t, \mathbf{X}_t, U_t)\}$ are α -mixing such that $\sum_l l^c [\alpha(l)]^{1-2/\delta} \leq \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$.
- (A.4) As $T \rightarrow \infty, h_T \rightarrow 0$ and $Th_T^2 \rightarrow \infty$. Moreover, there exists a sequence of positive integers s_T such that $s_T \rightarrow \infty, s_T = o((Th_T)^{1/2})$, and $(Th_T^{-1})^{1/2} \alpha(s_T) \rightarrow 0$ as $T \rightarrow \infty$.
- (A.5) Let $f_{X,U}(y)$ be the conditional density of Y given $\mathbf{X} = x$ and $U = u$. Assume that $f_{X,U}(y) \leq M$.
- (A.6) Let $f(\mu, \nu|x_0, x_t)$ be the conditional density of (U_0, U_t) given $(\mathbf{X}_0, \mathbf{X}_t)$. Assume that $|f(\mu, \nu|x_0, x_t)| \leq M < \infty$, for all $l \geq 1$.

Recall that $\mathbf{Z}_t = (\mathbf{X}_t^T, \mathbf{X}_t^T(U_t - u_0))^T, \mathbf{Z}_t^* = (\mathbf{X}_t^T, \mathbf{X}_t^T(U_t - u_0)/h)^T, K_t = K((U_t - u_0)/h), \boldsymbol{\beta}_0(u_0) = (a_1(u_0), \dots, a_p(u_0), a'_1(u_0), \dots, a'_p(u_0))^T = (\mathbf{a}^T, \mathbf{b}^T)^T$, and $\widehat{\boldsymbol{\beta}} = (\widehat{\mathbf{a}}^T, \widehat{\mathbf{b}}^T)^T$. Write $\bar{\eta}(u_0, U_t, \mathbf{X}_t) = \sum_{j=1}^p (a_j(u_0) + a'_j(u_0)(U_t - u_0))X_{tj} = \boldsymbol{\beta}_0^T \mathbf{Z}_t$, and $\eta(u, \mathbf{X}_t) = \sum_{j=1}^p a_j(u)X_{tj}$. Applying a Taylor series expansion to $\eta(u, \mathbf{X}_t)$ with respect to u around $|u - u_0| < h$, we obtain

$$\begin{aligned} \eta(u, \mathbf{X}_t) = \bar{\eta}(u_0, U_t, \mathbf{X}_t) + \frac{1}{2} \sum_{j=1}^p a''_j(u_0)X_{tj}(U_t - u_0)^2 \\ + o(h^2). \end{aligned} \tag{A.1}$$

For purposes of technical convenience, we reparameterize the estimation function (2.6) as

$$\begin{aligned} \widehat{\boldsymbol{\xi}} = \sqrt{Th}(\widehat{a}_1(u_0) - a_1(u_0), \dots, \widehat{a}_p(u_0) - a_p(u_0), \\ h(\widehat{b}_1(u_0) - b_1(u_0)), \dots, h(\widehat{b}_p(u_0) - b_p(u_0)))^T. \end{aligned}$$

Then

$$\sum_{j=1}^p (\widehat{a}_j + \widehat{b}_j(U_t - u_0))X_{tj} = \boldsymbol{\beta}_0^T \mathbf{Z}_t + \widehat{\boldsymbol{\xi}}^T \mathbf{Z}_t^*/\sqrt{Th},$$

and $\widehat{\boldsymbol{\xi}}$ minimizes the function

$$\begin{aligned} G_T(\boldsymbol{\xi}) \equiv G_T(\boldsymbol{\xi}; \theta, \mathbf{X}, U, u_0) \\ = \sum_{t=1}^T \{Q_\theta(Y_t^* - \boldsymbol{\xi}^T \mathbf{Z}_t^*/\sqrt{Th}) - Q_\theta(Y_t^*)\} K_t, \end{aligned} \tag{A.2}$$

where $Y_t^* = Y_t - \boldsymbol{\beta}_0^T \mathbf{Z}_t = Y_t - \bar{\eta}(u_0, U_t, \mathbf{X}_t)$. Note that $G_T(\boldsymbol{\xi})$ is convex in $\boldsymbol{\xi}$.

The proof of Theorem 1 is based on an approximation of the objective function in (A.2) by the quadratic function given in Lemma A.1. It can then be shown that the estimator $\widehat{\boldsymbol{\xi}}$ shares the same asymptotic behavior as the minimizer of the quadratic function which is asymptotically normal. The convexity lemma (Pollard 1991) plays a key role in the approximation. This proof technique is similar to those used

by Fan, Hu, and Truong (1994) and Yao and Tong (1996) for studying nonparametric regression with iid observations and nonparametric expectile regression under the α -mixing condition.

Lemma A.1. Assume that conditions (A.1) through (A.6) are satisfied. Then we have, as $T \rightarrow \infty$,

$$G_T(\xi) = \frac{1}{2}\xi^T \mathbf{D}(\theta, u_0)\xi - \frac{1}{\sqrt{Th}}W_T^T \xi + R_T(\xi),$$

where $W_T = \sum_{i=1}^T Q'_\theta(Y_i^*)K_i \mathbf{Z}_i^*$, $\sup_{\xi \in \mathcal{K}} |R_T(\xi)| = o_p(1)$ for any compact set \mathcal{K} .

Proof of Lemma A.1: The detailed proof, which is very similar to the proof of Theorem 2 given by Fan, Hu, and Truong (1994), is available from the online supplementary file.

Lemma A.2. Assume that conditions (A.1) through (A.6) are satisfied. Then we have, as $T \rightarrow \infty$,

$$\frac{1}{\sqrt{Th}} \left[W_T - \frac{h^2}{2} f_U(u_0) \begin{pmatrix} a''(u_0)\mu_2 \\ a''(u_0)\mu_3 \end{pmatrix} \otimes \mathbf{\Gamma}(\theta, u_0) + o(h^2) \right] \xrightarrow{\mathcal{L}} N(0, \mathbf{\Sigma}(\theta, u_0)), \quad (\text{A.3})$$

where $\mathbf{A}(\theta, u_0) = E(Q_\theta'^2(\eta(u_0, \mathbf{X}))\mathbf{X}\mathbf{X}^T | U = u_0)$,

$$\mathbf{\Sigma}(\theta, u_0) = f_U(u_0) \begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix} \otimes \mathbf{A}(\theta, u_0),$$

and

$$\mathbf{\Gamma}(\theta, u_0) = 2E(\{\theta[1 - F_{X,U}(\eta(u_0, \mathbf{X}))] + (1 - \theta)F_{X,U}(\eta(u_0, \mathbf{X}))\}\mathbf{X}\mathbf{X}^T | U = u_0).$$

Proof of Lemma A.2: The detailed proof, which is very similar to the proof of Theorem 2 of Cai, Fan, and Li (2000), is available from the online supplementary file. The proof uses the small-block and large-block technique and the Cramér-Wold method.

Proofs of Theorem 1 and Theorem 2. From the convexity lemma of Pollard (1991), the minimizer $\hat{\xi}$ of the convex function $G_T(\xi)$ converges in probability to the minimizer $-\mathbf{D}^{-1}(\theta, u_0)W_T/\sqrt{Th}$ of the convex function $(1/2)\xi^T \mathbf{D}(\theta, u_0)\xi - \frac{1}{\sqrt{Th}}W_T^T \xi$. Now, from Lemma A.1, $\hat{\xi}$ can be explicitly expressed as

$$\hat{\xi} = -\mathbf{D}^{-1}(\theta, u_0)W_T/\sqrt{Th} + o_p(1) \quad (\text{A.4})$$

uniformly for $\xi \in \mathcal{K}$ which is a compact set of ξ . From (A.4), we have

$$H(\hat{\beta}_\theta - \beta) = -\mathbf{D}^{-1}(\theta, u_0)W_T/Th + o_p(1) \quad (\text{A.5})$$

uniformly for $u_0 \in \mathcal{U}$. Using Lemma A.2 in conjunction with (A.5) proves Theorem 1. Theorem 2 is an immediate consequence of (A.5) and Lemma A.2 and Slutsky's theorem. \square

SUPPLEMENTARY MATERIALS

The supplementary file contains proofs of Lemmas A.1 and A.2.

ACKNOWLEDGMENTS

The authors are grateful for the financial support provided by grants from the following bodies: National Natural Science Foundation of China (No. 71203025 (Xie), No. 71271128 (Zhou)), Strategic Research Grant Scheme, City University of Hong Kong (No. 7008134 (Wan)), National Natural Science Funds for Distinguished Young Scholar (No. 70825004 (Zhou)), NCMIS (Zhou), and State Key Program of National Natural Science Foundation of China (No. 71331006 (Xie and Zhou)). The

authors thank the Editor, Associate Editor, and several referees for their constructive comments. All remaining errors, if any, are of the authors.

[Received November 2012. Revised March 2014.]

REFERENCES

- Aigner, D. J., Amemiya, T., and Poirier, D. J. (1976), "On the Estimation of Production Frontiers: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," *International Economic Review*, 17, 377–396. [576]
- Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999), "Coherent Measures of Risk," *Mathematical Finance*, 9, 203–228. [576]
- Cai, J., Fan, J., Jiang, J., and Zhou, H. (2008), "Partially Linear Hazard Regression With Varying-Coefficients for Multivariate Survival Data," *Journal of the Royal Statistical Society, Series B*, 70, 141–158. [577]
- Cai, Z., Fan, J., and Li, R. (2000), "Efficient Estimation and Inferences for Varying Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902. [579,581,591]
- Cai, Z., Fan, J., and Yao, Q. (2000), "Functional-Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–955. [577,579,582]
- Cai, J., Fan, J., Zhou, H., and Zhou, Y. (2007), "Hazard Models With Varying Coefficients for Multivariate Failure Time Data," *The Annals of Statistics*, 35, 324–354. [577,579]
- Cai, Z., and Ould-Said, E. (2003), "Local M-Estimator for Nonparametric Time Series," *Statistics and Probability Letters*, 65, 433–449. [578]
- Cai, Z., and Xu, X. (2008), "Nonparametric Quantile Estimation for Dynamic Smooth Coefficient Models," *Journal of the American Statistical Association*, 103, 1595–1608. [577,582,584]
- Chen, R., and Tsay, R. S. (1993), "Functional Coefficient Autoregressive Models," *Journal of American Statistical Association*, 88, 298–308. [577,578]
- Cleveland, W. S., Grosse, E., and Shyu, W. M. (1991), "Local Regression Models," in *Statistical Models in S*, eds. J. M. Chambers and T. Hastie, Pacific Grove, CA: Wadsworth and Brooks/Cole, pp. 309–376. [577]
- Efron, B. (1991), "Regression Percentiles Using Asymmetric Squared Error Loss," *Statistica Sinica*, 1, 93–125. [576]
- Fan, J., and Chen, J. (1999), "One-Step Local Quasi-Likelihood Estimation," *Journal of the Royal Statistical Society, Series B*, 61, 927–943. [579]
- Fan, J., Hu, T. C., and Truong, Y. K. (1994), "Robust Non-Parametric Function Estimation," *Scandinavian Journal of Statistics*, 21, 433–446. [591]
- Fan, J., and Huang, T. (2005), "Profile Likelihood Inference on Semiparametric Varying-Coefficient Partially Linear Models," *Bernoulli*, 11, 1031–1057. [577]
- Fan, J., and Li, R. (2004), "New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis," *Journal of the American Statistical Association*, 99, 710–723. [577]
- Fan, J., Lin, H., and Zhou, Y. (2006), "Local Partial-Likelihood Estimation for Lifetime Data," *The Annals of Statistics*, 34, 290–325. [577]
- Fan, J., Yao, Q., and Cai, Z. (2003), "Adaptive Varying-Coefficient Linear Models," *Journal of the Royal Statistical Society, Series B*, 65, 57–80. [584]
- Fan, J., and Zhang, J. T. (2000), "Functional Linear Models for Longitudinal Data," *Journal of the Royal Statistical Society, Series B*, 62, 303–322. [577]
- Fan, J., and Zhang, W. (2008), "Statistical Methods With Varying Coefficient Models," *Statistics and its Interface*, 1, 179–195. [577]
- Granger, C., and Teräsvirta, T. (1993), *Modeling Nonlinear Economic Relationships*, Oxford: Oxford University Press. [578]
- Haggan, V., and Ozaki, T. (1981), "Modeling Nonlinear Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model," *Biometrika*, 68, 189–196. [578]
- Hastie, T. J., and Tibshirani, R. (1993), "Varying Coefficient Model," *Journal of the Royal Statistical Society, Series B*, 55, 757–796. [577]
- Honda, T. (2004), "Quantile Regression in Varying Coefficient Models," *Journal of Statistical Planning and Inference*, 121, 113–125. [577]
- Jones, M. C. (1994), "Expectiles and M-Quantiles," *Statistics and Probability Letters*, 20, 149–153. [577]
- Kim, M. O. (2007), "Quantile Regression With Varying-Coefficients," *The Annals of Statistics*, 35, 92–108. [577]
- Kuan, C., Yeh, J., and Hsu, Y. (2009), "Assessing Value at Risk With CARE, the Conditional Autoregressive Expectile Models," *Journal of Econometrics*, 150, 261–270. [576,577,578,579,583,587]
- Newey, W. K., and Powell, J. L. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 819–847. [576,577,578]

- Pham, D. T. (1986), "The Mixing Properties of Bilinear and Generalized Random Coefficient Autoregressive Models," *Stochastic Processes and Their Applications*, 23, 291–300. [578]
- Pollard, D. (1991), "Asymptotics for Least Absolute Deviation Regression Estimations," *Econometric Theory*, 70, 186–199. [590,591]
- Schnabel, S. K., and Eilers, P. H. C. (2009), "Optimal Expectile Smoothing," *Computational Statistics and Data Analysis*, 53, 4168–4177. [576]
- Taylor, J. W. (2008), "Estimating Value at Risk and Expected Shortfall Using Expectiles," *Journal of Financial Econometrics*, 6, 231–252. [577,580,581]
- Tong, H. (1983), *Threshold Models in Nonlinear Time Series Analysis*, Heidelberg: Springer. [578]
- (1990), *Non-linear Time Series: A Dynamical System Approach*, New York: Oxford University Press. [578]
- Yao, Q., and Tong, H. (1996), "Asymmetric Least Squared Regression Estimation: A Nonparametric Approach," *Journal of Nonparametric Statistics*, 6, 273–292. [577,578,579,591]
- Yuan, M. L., and Wong, M. C. W. (2010), "Analytical VaR and Expected Shortfall for Quadratic Portfolios," *Journal of Derivatives*, 7, 33–44. [576]