Econometric modeling of risk measures: A selective review of the recent literature

3 authors:

Dingshi Tian
Zhongnan University of Economics and Law
4 PUBLICATIONS   **13** CITATIONS

SEE PROFILE

Zongwu Cai
University of Kansas
175 PUBLICATIONS   **4,210** CITATIONS

SEE PROFILE

Ying Fang
Xiamen University
40 PUBLICATIONS   **335** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Inference View project

Semiparametric Estimation of Conditional Autoregressive Expectile Models and Their Applications View project

Econometric modeling of risk measures: A selective review of the recent literature[*]

Dingshi Tian[a],   Zongwu Cai[b]   and   Ying Fang[a]

[a] Wang Yanan Institute for Studies in Economics, Department of Statistics, School of Economics, Ministry of Education Key Laboratory of Econometrics and Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen, Fujian 361005, China (*tiandingshi@gmail.com (Tian); yifst1@xmu.edu.cn (Fang)*)
[b] Department of Economics, University of Kansas, Lawrence, KS 66045, USA. (*caiz@ku.edu*)

October 30, 2018

Abstract

Since the financial crisis in 2008, the risk measures which are the core of risk management, have received increasing attention among economists and practitioners. In this review, the concentrate is on recent developments in the estimation of the most popular risk measures, namely, value at risk (VaR), expected shortfall (ES), and expectile. After introducing the concept of risk measures, the focus is on discussion and comparison of their econometric modeling. Then, parametric and nonparametric estimations of tail dependence are investigated. Finally, we conclude with insights into future research directions.

Keywords: Expectile; Expected Shortfall; Network; Nonparametric Estimation; Tail Dependence; Value at Risk

1   Introduction

Risk management plays a crucial role in financial institutions, such as banks, insurance industries, and investment funds, and at the core of risk management are the techniques employed to measure risk. Risk professionals have long been searching for a "good" risk measure. After Markowitz (1952), the variance of the profit and loss distributions became the dominating risk measure in finance. However, one shortcoming of this measure is that it requires the risk functional to be a random variable with finite variance. More importantly, the corresponding distributions are implicitly assumed to be approximately symmetric, in the sense that large gains and losses are equally penalized, which contradicts the fact that investors are usually more concerned about large losses than gains.

Different types of economic risks exist in the financial market, as has been progressively realized by financial and regulatory institutions. In 1996, the Basel Committee on Banking Supervision (BCBS) incorporated systematic risk as a supplement to credit risk and adopted a mandatory risk measure called the VaR to be calculated by all banks for each line of their balance sheets. Given a holding period with probability $1 - \tau$, VaR is defined as the possible maximum

1

loss, within a confidence level of $1 - \tau$. Since then, VaR has become one of the most widely used measures of market risk in risk management. The pioneering work on VaR was Morgan (1996), Jorion (1997), Duffie and Pan (1997). Among them, Morgan (1996) introduced the RiskMetrics methodology, which played an important role in the increasing popularity of VaR. However, VaR has long been a controversial topic. A first criticism is that when VaR is used as a risk measure, diversification does not necessarily reduce the risk. Perhaps more importantly, as criticized by Basak and Shapiro (2001), VaR may induce large losses because of its insensitivity to extreme loss. These drawbacks of VaR naturally elicit an important question that has contributed to the development of risk measurement: what properties should be expected from a good risk measure?

In the seminal work by Artzner et al. (1999), the desirable properties of risk measures were formalized in a set of axioms. A risk measure with the properties of translation invariance, subadditivity, positive homogeneity and monotonicity is called a *coherent* risk measure. To find such an oracle, Artzner et al. (1999) and Basak and Shapiro (2001) introduced ES, which is defined as the conditional expectation of the loss given that the loss exceeds the VaR level in a given time period, as an alternative to VaR. As ES remedies the aforementioned problems of VaR, it has replaced VaR in many institutions as a risk management tool. Specifically, the BCBS recommends financial institutions replace VaR with ES in their internal market risk models.

In addition to the coherent properties introduced above, backtesting and forecast verification are also important properties to consider. As risk measures are usually estimated using historical data, a risk measure must have the ability to verify and compare competing estimation procedures. In statistical decision theory, a risk measure where verification and comparison is possible is called an *elicitable* risk measure. In Embrechts and Hofert (2014) and Ziegel (2016), the connection of elicitability with backtesting and coherency, respectively, was considered. Unfortunately, Gneiting (2011) showed that ES is not elicitable, so an issue with direct backtesting of ES estimates may exist. With a view on the feasibility of the backtesting, expectile, first introduced by Newey and Powell (1987), has been suggested as a coherent and elicitable alternative to ES in recent studies. Expectile regression estimates are the solutions to the minimization of asymmetrically weighted mean squared errors. Owning to the quadratic loss function, expectile is sensitive to the extreme losses of a given distribution. Furthermore, Bellini et al. (2014) characterized generalized quantiles, which have positive homogeneity and convexity properties, and showed that expectile e_τ (with $\tau \geq \frac{1}{2}$) is the only generalized quantile that enjoys all the properties of a coherent risk measure.

The aforementioned risk measures assess the risk in an isolated perspective; however, the interaction between different individuals is also of great importance. The failure of Fannie Mae and Freddie Mac harmed the entire financial system and eventually led to the outbreak of the financial crisis, thereby calling for a study on measuring the financial co-movements between the agents in the financial markets. The copula approach is a natural way to measure the tail dependence between two institutions. More recently, the CoVaR approach was introduced by Adrian and Brunnermeier (2016) to measure tail dependence. CoVaR is defined as the VaR of an institution conditioned on another institution (or the whole system) being in distress. Finally, the financial network which includes quantitative information regarding interactions and the roles of institutions within the entire financial system, has become particularly popular for studying the system risk from an overall perspective. Therefore, in this paper, one section is used to discuss this topic.

The rest of this paper is organized as follows. In Section 2, the properties of some popular risk measures, namely, VaR, ES and expectile, are discussed and compared. Section 3 and Section 4 survey the practical evaluation of these risk measures in terms of parametric models and nonparametric models, respectively. Section 5 is devoted to investigating the recent developments in measuring tail dependence. Section 6 concludes the paper.

2 Risk Measures and Related Properties

This section introduces the definition of the risk measures discussed in this paper, namely, VaR, ES and expectile.

2.1 Popular Risk Measures

Since Markowitz (1952) stated the famous mean-variance framework in portfolio optimization, variance (standard deviation) had long been the dominant risk measure. The mean-variance framework was replaced by VaR, the most popular downside risk measure in recent years. Here, first, some notation is introduced. Let R_t be a random variable denoting the return of a portfolio at time t, and let $\tau \in (0,1)$ be the probability level.

Definition 2.1. *For a probability level τ, the **value at risk (VaR)** of a random variable R_t is defined as the minus τ-quantile of R_t:*

$$\mathrm{VaR}_\tau(R_t) = -q_\tau(R_t) = -\inf\{r|P(R_t \le r) \ge \tau\}. \tag{1}$$

VaR suffers from some drawbacks: it lacks subadditivity and is insensitive to extreme loss. To alleviate these problems inherent in VaR, Artzner et al. (1999) and Basak and Shapiro (2001) proposed ES as an alternative.

Definition 2.2. *For a probability level τ, the **expected shortfall (ES)** of a random variable R_t is defined as*

$$\mathrm{ES}_\tau(R_t) = -E[R_t|R_t \le -\mathrm{VaR}_\tau(R_t)]. \tag{2}$$

As ES has been criticized for issues related to backtesting, the coherent and elicitable risk measure, the expectile, was proposed by Bellini et al. (2014) and Ziegel (2016).

Definition 2.3. *For a probability level τ, the τ-**expectile** $e_\tau(R_t)$ of a random variable R_t is defined as*

$$e_\tau = \arg\min_{r \in \mathcal{R}} E[|\tau - I(R_t \le r)| \cdot |R_t - r|^2], \tag{3}$$

where $I(\cdot)$ denotes the indicator function.

Remark 1: *Most papers use returns to calculate VaR, and the return is defined as $R_t = \log(P_t) - \log(P_{t-1})$, where P_t denotes the price of a financial asset at time t. However, in practice, regulators or risk managers may be more concerned about the distribution of the losses, which is defined as $Y_t \equiv -R_t = \log(P_{t-1}) - \log(P_t)$. Moreover, in some cases, the model is easier to construct using losses rather than returns. For a probability level τ, the VaR of the loss variable Y_t is defined as*

$$\mathrm{VaR}_\tau(Y_t) = q_\tau(Y_t) = \inf\{y|P(Y_t \le y) \ge \tau\}, \tag{4}$$

and the related ES is defined as

$$\mathrm{ES}_\tau(Y_t) = E[Y_t|Y_t \ge \mathrm{VaR}_\tau(Y_t)]. \tag{5}$$

Therefore, in the following sections, the estimation of VaR and ES are discussed using either returns or losses according to the original paper.

2.2 Coherency and Elicitability

A risk measure can be interpreted as a function that maps a loss distribution or random variable to a capital amount. Artzner et al. (1999) proposed four axioms that bundle mathematical properties as possible criteria for the choice of such functions. The so-called coherent risk measures satisfy the following set of conditions.

Definition 2.4. *[Artzner et al. (1999)] Let $\mathcal{G}$ be a set of real-valued random variables (typically, the loss variables). The loss function $\rho(\cdot) : \mathcal{G} \to \mathcal{R}$ is a coherent risk measure if it satisfies*

1. *Translation invariance: $Y \in \mathcal{G}$, $c_0 \in \mathcal{R}$, then $\rho(Y + c_0) = \rho(Y) - c_0$;*

2. *Subadditivity: $Y_1, Y_2 \in \mathcal{G}$, then $\rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$;*

3. *Positive homogeneity: $\lambda \geq 0$, $Y \in \mathcal{G}$, then $\rho(\lambda Y) = \lambda \rho(Y)$;*

4. *Monotonicity: $Y_1, Y_2 \in \mathcal{G}$ with $Y_1 \leq Y_2$, then $\rho(Y_1) \leq \rho(Y_2)$.*

Translation invariance means that the risk measure decreases by c_0 if a sure amount c_0 is added to the position. Subadditivity is correlated with the concept of diversification and implies that the risk of a portfolio composed of two sub-portfolios is smaller than the sum of the risk of these two sub-portfolios. Positive homogeneity means that if the size of the portfolio is increased by a factor of λ while keeping the weights unchanged, the risk would increase by the same factor λ. Monotonicity means that assets with more negative random outcomes are more risky.

As mentioned above, one coherent risk measure is ES. Additionally, portfolio optimization is easier to implement with ES objectives than with VaR objectives, and the extreme losses are explicitly taken into account in the allocation process when ES is used. For further analysis of the advantages of using ES, please refer to Jondeau et al. (2007). However, Gneiting (2011) showed that ES is not elicitable, which is the property connected to comparing the performance of different forecast methods, so there could be an issue with direct backtesting of ES estimates. The concept of elicitability was first introduced by Osband (1985) and Lambert et al. (2008) and then by Gneiting (2011). To introduce elicitability, first, the definition of a consistent scoring function is recalled here.

Definition 2.5. *[Ziegel (2016)] Let $P \in \mathcal{P}$ be the distribution function of a real-valued random variable Y, where $\mathcal{P}$ is a class of probability measures on $\mathcal{R}$ with Borel sigma algebra, and let $\rho(\cdot)$ be a real-valued function on $\mathcal{P}$. A **scoring function** $s(\cdot, \cdot)$ is consistent for the functional $\rho(\cdot)$ relative to the class $\mathcal{P}$ if and only if for all $t \in \rho(P)$ and all $x \in \mathcal{R}$,*

$$E_P[s(t, Y)] \leq E_P[s(x, Y)], \tag{6}$$

*where x is a point estimate. Furthermore, $s(\cdot, \cdot)$ is **strictly consistent** if it is consistent and $E_P[s(t, Y)] = E_P[s(x, Y)]$ implies that $x \in \rho(P)$.*

Then, the definition of elicitability is introduced.

Definition 2.6. *[Ziegel (2016)] A function $\rho(\cdot)$ is **elicitable** relative to the class $\mathcal{P}$ if there exists a scoring function $s(\cdot, \cdot)$ that is strictly consistent for $\rho(\cdot)$ relative to $\mathcal{P}$.*

Expectile was shown to be elicitable by Bellini et al. (2014) and Ziegel (2016) and has been suggested as a coherent and elicitable alternative to ES. Ziegel (2016) also showed that expectile is indeed the only law-invariant and coherent elicitable risk measures. Expectile appears to be a perfect substitute for VaR and ES. However, expectile is not comonotonically additive from the so-called Kusuoka representation, see Emmer et al. (2015) and Ziegel (2016) for more discussions.

From the discussions above, it is clear that it is not easy to find an all-inclusive risk measure. Therefore, it turns to the estimation techniques of these risk measures and their further development is investigated.

3 Models for VaR and ES

Since Morgan (1996), RiskMetrics has been widely used in the estimation of VaR. However, in this review, conditional estimates of VaR and ES are of interest. Conditional estimators yield more accurate assessments of risk than their unconditional counterparts by modeling the stochastic nature of conditional volatility. The reader with interest in RiskMetrics and other unconditional parametric approach should refer to Morgan (1996), Hull and White (1998) for RiskMetrics; Butler and Schachter (1998), Gourieroux et al. (2000), and Chen and Tang (2005) for historical simulation and its variants; and McNeil (1997), Danielsson and de Vries (1997) for unconditional extreme value theory (EVT).

3.1 Parametric Models

By using parametric model to capture structure on the tails of the distribution, the following semiparametric approaches are free from estimating the complete distribution of returns. The first model is based on EVT. Second one is the quantile regression technique developed by Chernozhukov and Umantsev (2001) and Engle and Manganelli (2004). Finally, a regression approach for ES can be used too. Besides, it is worthy mentioning that Wang and Zhao (2016) proposed a semiparametric approach to estimate conditional VaR, which is characterized by a general parametric model with nonparametric noise distribution. The advantage of this paper is to combine the good merits of parametric and nonparametric approaches; see the paper by Wang and Zhao (2016) for details.

3.1.1 Extreme Value Theory

EVT was designed to model the tail behavior of random variables. The main idea of EVT is that, given a sufficiently large threshold u, returns less than u can be approximated by a generalized Pareto distribution (GPD). The GPD is particularly applicable for the estimation of VaR and ES, and it is defined as

$$\text{GPD}(w; \varsigma, \psi) = \begin{cases} 1 - \left(1 + \frac{\varsigma w}{\psi}\right)^{-1/\varsigma}, & \text{if } \varsigma \neq 0, \\ 1 - \exp\left(-\frac{w}{\psi}\right), & \text{if } \varsigma = 0, \end{cases} \tag{7}$$

where w denotes the exceedances over u, ς is a tail index representing the shapes of the tails of the distribution, and ψ is a scaling parameter. Considering the temporal evolution of volatility, McNeil and Frey (2000) generalized McNeil (1997) by proposing the GARCH-EVT model for the estimation of conditional VaR and ES. Their procedure to model the conditional volatility and the distribution of the tails separately is as follows. In the first step, they employ a GARCH model to filter the dependence in the return series, and the resulting residuals are i.i.d. if the GARCH model is correctly specified. The GARCH model adopted to filter out the first- and second-order dynamics is of the form

$$R_t = \mu_t + \sigma_t z_t, \ \mu_t = \mu + \varphi R_{t-1}, \ \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{8}$$

where $\{R_t\}_{t=1}^n$ is the return series, and some constraints of the parameters, such as $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$, are set to ensure positive volatility and covariance as a stationary process. After implementing the quasi-maximum likelihood (QML) for the estimation of the model, the residuals are standardized by $\widehat{z}_t = (R_t - \widehat{\mu}_t)/\widehat{\sigma}_t$, where $\widehat{\mu}_t$ and $\widehat{\sigma}_t$ are the obtained estimators. In the second step, the GPD is estimated based on all exceedances, which are defined as the realizations $\widehat{z}_t$ that are below a given threshold u. Defining $N_u = \sum_{t=1}^n I(\widehat{z}_t < u)$ as the number of exceedances, the standardized residuals are sorted and a sample of exceedances

is selected as $\{\widehat{z}_{(1)}, \ldots, \widehat{z}_{(N_u)}\}$. Then, $(\widehat{\varsigma}, \widehat{\psi})^\top$ is obtained by estimating the GPD using the exceedances. With $(\widehat{\varsigma}, \widehat{\psi})^\top$, after inverting the conditional cumulative distribution function (CDF) of the exceedances, the τ-quantile of random variable $\{R_t\}_{t=1}^n$ is obtained by

$$\widehat{q}_\tau(R_t) = \begin{cases} u + \frac{-\widehat{\psi}}{\widehat{\varsigma}}\left(\left(\frac{n}{N_u}\tau\right)^{-\widehat{\varsigma}} - 1\right), & \text{if } \varsigma \neq 0 \\ u + \widehat{\psi}\log(\frac{n}{N_u}\tau), & \text{if } \varsigma = 0 \end{cases}. \tag{9}$$

Finally, the forecasting of VaR and ES from t to $t+1$ is given by

$$\text{VaR}_{t+1,\tau} = -\left(\widehat{\mu}_{t+1} + \widehat{\sigma}_{t+1}\widehat{q}_\tau(R_t)\right), \tag{10}$$

$$\text{ES}_{t+1,\tau} = \frac{\text{VaR}_{t+1,\tau}}{1 - \widehat{\varsigma}} + \frac{\widehat{\psi} - \widehat{\varsigma}u}{1 - \widehat{\varsigma}}, \tag{11}$$

where $\widehat{\mu}_{t+1}$ and $\widehat{\sigma}_{t+1}$ are one-step forecasts of the expected return and variance, respectively, which can be calculated by $\widehat{\mu}_{t+1} = \widehat{\mu} + \widehat{\varphi}R_t$ and $\widehat{\sigma}_{t+1}^2 = \widehat{\omega} + \widehat{\alpha}\widehat{\epsilon}_t^2 + \widehat{\beta}\widehat{\sigma}_t^2$. Compared to unconditional EVT, an obvious improvement of the GARCH-EVT approach is the incorporation of expected return and volatility into the VaR changes.

A further study by Kuester et al. (2006) compared 13 conditional EVT models using NAS-DAQ Composite Index data and found that GARCH-EVT always performed better than other conditional models for probability level $0 < \tau < 0.1$, given that the GARCH-filtered innovations follow a normal or skew t-distribution. For a detailed discussion on EVT, refer to the survey paper of Rocco (2014).

3.1.2 Quantile Approach

We know that VaR is actually a quantile estimator, so the estimation of VaR is equal to the modeling of a quantile. The basic idea of the quantile regression approach is to model a given quantile of the distribution through time, and the technique of quantile regression was first introduced in Koenker and Bassett (1978) and further summarized in Koenker (2005). The τ-conditional quantile of R_t given $\mathcal{F}_{t-1}$, where $\mathcal{F}_t$ is the information set at time t, can be obtained by solving the following optimization problem

$$q_{t,\tau} = \arg\min_{y \in \mathcal{R}} E\{[\tau - I(R_t < r)](R_t - r)\,|\mathcal{F}_{t-1}\}. \tag{12}$$

Examples of conditional quantile models include Koenker and Zhao (1996)'s conditional quantile model, $q_{t,\tau}(\boldsymbol{\beta}) = \beta_{0,\tau} + \sum_{i=1}^p \beta_{i,\tau}R_{t-i} + e_t$ with $e_t = (\gamma_{0,\tau} + \sum_{j=1}^q \gamma_{j,\tau}|e_{t-j}|)\delta$, $p, q \geq 1$, in which δ is the τ-quantile of an i.i.d. random variable with mean zero and finite variance; and by taking $\mathbf{X}_t$ to be a proxy of $\mathcal{F}_{t-1}$, Taylor (1999)'s and Chernozhukov and Umantsev (2001)'s linear VaR, $q_{t,\tau}(\boldsymbol{\beta}) = \mathbf{X}_t^\top\boldsymbol{\beta}$ and quadratic VaR models, $q_{t,\tau}(\boldsymbol{\beta}) = \mathbf{X}_t^\top\boldsymbol{\beta} + \mathbf{X}_t^\top\boldsymbol{\Theta}\mathbf{X}_t$, in which $\boldsymbol{\Theta}$ is a coefficient matrix.

The CAViaR model, proposed by Engle and Manganelli (2004), is the most popular linear model. CAViaR aims to directly model the evolution of the quantile q_τ, and a generic CAViaR specification has the form:

$$R_t = q_{t,\tau}(\boldsymbol{\beta}) + \epsilon_{t,\tau}, \qquad Q_\tau(\epsilon_{t,\tau}|\mathbf{X}_t) = 0, \tag{13}$$

$$q_{t,\tau}(\boldsymbol{\beta}) = \beta_{0,\tau} + \sum_{i=1}^p \beta_{i,\tau}q_{t-i,\tau}(\boldsymbol{\beta}) + \sum_{j=1}^q \beta_{p+j,\tau}g(\mathbf{X}_{t-j}), \tag{14}$$

where $q_{t,\tau}(\boldsymbol{\beta})$ is the τ-conditional quantile of the return distribution, $Q_\tau(\epsilon_{t,\tau}|\mathbf{X}_t)$ is the τ-conditional quantile of $\epsilon_{t,\tau}$ given regressor $\mathbf{X}_t$, $\boldsymbol{\beta}_\tau = (\beta_{0,\tau}, \ldots, \beta_{p+q,\tau})$ is a vector of coefficients, and $g(\cdot)$ is possibly a nonlinear function. The three suggested specifications in Engle and Manganelli (2004) are as follows:

Symmetric absolute value model

$$q_{t,\tau}(\boldsymbol{\beta}) = \beta_{0,\tau} + \beta_{1,\tau}q_{t-1,\tau}(\boldsymbol{\beta}) + \beta_{2,\tau}|R_{t-1}|,$$

Asymmetric slope model

$$q_{t,\tau}(\boldsymbol{\beta}) = \beta_{0,\tau} + \beta_{1,\tau}q_{t-1,\tau}(\boldsymbol{\beta}) + \beta_{2,\tau}R_{t-1}^+ + \beta_{3,\tau}R_{t-1}^-,$$

Indirect GARCH(1,1) model

$$q_{t,\tau}^2(\boldsymbol{\beta}) = \beta_{0,\tau} + \beta_{1,\tau}q_{t-1,\tau}^2(\boldsymbol{\beta}) + \beta_{2,\tau}R_{t-1}^2,$$

where $R_t^+ = \max(R_t, 0)$ and $R_t^- = -\min(R_t, 0)$. In the first and third settings, the dynamic quantile responds symmetrically to past returns. By contrast, the asymmetric slope model considers asymmetric effects, which means that the responses to positive and negative returns can be different.

3.1.3 Regression Technique

An analogous approach to EVT was introduced by Cai et al. (2015), which proposed an easily implemented regression technique based on the proportional mean residual life (PMRL) regression model to estimate ES. The mean residual life function $m(y)$ is defined as

$$m(y) = E(Y_t - y|Y_t > y) = S^{-1}(y)\int_y^\infty S(v)\,dv, \tag{15}$$

where the loss variable Y_t is nonnegative with finite mean μ, $S(y) = 1 - F(y)$ is the corresponding survival function, with $F(y)$ denoting the CDF of Y_t. The above equation clearly shows that $m(y)$ is a function of $\mathrm{ES}_\tau(Y_t)$ if we set $y = \mathrm{VaR}_\tau(Y_t)$. Thus, to estimate ES_τ, we first need to estimate $m(y)$. The inversion formula to equation (15) gives

$$S(y) = \frac{m(0)}{m(y)}\exp\left(-\int_0^y m^{-1}(v)\,dv\right), \tag{16}$$

and the two survival functions $S_0(y)$ and $S_1(y)$ are said to have proportional mean residual life

$$S_1(y) = S_0(y)\left(\int_y^\infty S_0(v)\,dv/\mu_0\right)^{1/k-1}, \tag{17}$$

if the condition $m_1(y) = km_0(y)$ for all $y \geq 0$, $k > 0$ is satisfied. The technique in Maguluri and Zhang (1994) can be used to extend this condition to a regression context with explanatory variables $\mathbf{X}$, namely,

$$m(y|\mathbf{X}) = \exp(-\boldsymbol{\beta}^\top\mathbf{X})m_0(y). \tag{18}$$

Then, one has

$$S(y|\mathbf{X}) = S_0(y)\left(\frac{1}{\mu_0}\int_y^\infty S_0(v)dv\right)^{\exp(\boldsymbol{\beta}^\top\mathbf{X})-1}, \tag{19}$$

where $S_0(y)$ is the baseline survival function, and $\mu_0 = E[\exp(\boldsymbol{\beta}^\top\mathbf{X})Y|\mathbf{X}]$ denotes the corresponding baseline mean. Oakes and Dasu (2003) proposed a class of weighted ratio estimators to estimate the above PMRL model. Suppose that the time series sample $\{(Y_t, \mathbf{X}_t)\}_{t=1}^n$ follows a population satisfying (18); then, β can be estimated by solving

$$\frac{\frac{1}{n}\sum_{t=1}^n \mathbf{X}_t Y_t \exp(\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_t)}{\frac{1}{n}\sum_{t=1}^n Y_t \exp(\widehat{\boldsymbol{\beta}}^\top\mathbf{X}_t)} - \frac{1}{n}\sum_{t=1}^n \mathbf{X}_t = 0. \tag{20}$$

If we set $S_0(y) = \exp(y/\mu_0)$ to be the baseline survival function, then (20) is the true maximum likelihood equation of the exponential regression model. $\widehat{m}(y|\mathbf{X})$ is readily estimated with the quasi-likelihood estimator given by (20); then, the ES estimates $\widehat{ES}_\tau$ with an estimated $\widehat{VaR}_\tau$ can be obtained.

3.2 Nonparametric Models

The aforementioned approaches assume that the tail of the distribution is well characterized by certain parametric models; therefore, they all enjoy good interpretation and easy implementation. However, these methods may also suffer from the model misspecification problem. Furthermore, as noted by Cai and Wang (2008), nonparametric modeling is appealing in several aspects. First, little or no prior information of the functionals is needed in the nonparametric approach; thus, the model misspecification problem is avoided. Besides, nonparametric modeling can provide insight for further parametric fitting.

3.2.1 Nonparametric Approach

To estimate the VaR and ES, Cai (2002), Wu et al. (2008), and Cai and Wang (2008) considered nonparametric modeling based on kernel smoothing to estimate the conditional distribution function.

Assume that $\{(Y_t, \mathbf{X}_t)\}_{t=1}^n$ are stationary time series, where Y_t is a risk or loss variable, and $\mathbf{X}_t$ is a vector of covariates that can include both economic variables and lagged variables of Y_t. The VaR of Y_t given $\mathbf{X}_t = \mathbf{x}$ can be formulated as $q_\tau(\mathbf{x}) = S^{-1}(\tau|\mathbf{x})$, where the survival function $S(y|\mathbf{x}) = 1 - F(y|\mathbf{x})$ and $F(y|\mathbf{x})$ is the conditional CDF of Y given $\mathbf{X} = \mathbf{x}$. Then, the nonparametric estimator of VaR can be constructed as $\widehat{q}_\tau(\mathbf{x}) = \widehat{S}^{-1}(\tau|\mathbf{x})$; thus, the ES estimator $\widehat{\mathrm{ES}}_\tau(\mathbf{x})$ is easily obtained using the plugging-in method,

$$\widehat{\mathrm{ES}}_\tau(\mathbf{x}) = \frac{1}{\tau} \int_{\widehat{q}_\tau(\mathbf{x})}^{\infty} y \widehat{f}(y|\mathbf{x}) \, dy, \tag{21}$$

where $f(y|\mathbf{x})$ is the conditional CDF of Y_t given $\mathbf{X}_t = \mathbf{x}$. Therefore, the key point to estimate VaR and ES is to estimate the conditional probability distribution function (PDF) and CDF.

In the nonparametric setting, the kernel type of the nonparametric estimation of conditional quantiles has serious drawbacks: the asymptotic bias cannot be adaptive, and the boundary effect requires boundary modification. Many different types of local linear estimators have been proposed to overcome these drawbacks. One approach using a "check" function, such as a robustified local linear smoother, was proposed by Fan et al. (1994). An alternative procedure first estimates the CDF using the "double-kernel" local linear technique proposed by Fan et al. (1996) and then inverts the obtained CDF estimator to estimate the conditional quantile. This approach is called Yu and Jones's estimator, see Yu and Jones (1997, 1998) for a detailed comparison of these two methods.

However, as noted by Hall et al. (1999), although the local linear estimators of the Yu and Jones's type have some attractive properties, such as no boundary effects, design adaptation, and mathematical efficiency, they also have the disadvantage of producing conditional distribution function estimators that are not constrained to be monotone increasing or to lie between zero and one; see Cai (2002) for details. Despite their large bias and boundary effects, the Nadaraya-Watson (NW) methods are superior because they have the properties of positivity and monotonicity. To combine these two approach, Hall et al. (1999) proposed a weighted NW (WNW) estimator, which is designed to possess the superior property of the NW estimator that it is always a distribution function and to preserve the properties of the local linear methods, such as bias reduction and no boundary effect. Cai (2002) then established the weak consistency and asymptotic normality of the WNW conditional distribution estimator for α-mixing time series at both the boundary and interior points. The WNW estimator of the conditional distribution function of Y_t given $\mathbf{X}_t = \mathbf{x}$ is defined as

$$\widehat{F}_w(y|\mathbf{x}) = \sum_{t=1}^{n} W_{c,t}(\mathbf{x}, h) I(Y_t \leq y), \tag{22}$$

and the weights $W_{c,t}(\mathbf{x}, h)$ are given by

$$W_{c,t}(\mathbf{x}, h) = \frac{p_t(\mathbf{x})W_h(\mathbf{x} - \mathbf{X}_t)}{\sum_{t=1}^{n} p_t(\mathbf{x})W_h(\mathbf{x} - \mathbf{X}_t)}, \tag{23}$$

where $W_h(\cdot) = W(\cdot/h)/h$ with $W(\cdot)$ denoting the kernel function and h being bandwidth, and $\{p_t(\mathbf{x})\}_{t=1}^{n}$ denotes the a series of weighted functions such that $p_t(\mathbf{x}) \geq 0$. The optimal weights $p_t(x)$ can be obtained by solving the following optimization problem,

$$\max \sum_{t=1}^{n} \log(p_t(\mathbf{x})) \ \ s.t. \ \ \sum_{t=1}^{n} p_t(\mathbf{x}) = 1, \ \ \text{and} \ \ \sum_{t=1}^{n}(\mathbf{X}_t - \mathbf{x})p_t(\mathbf{x})W_h(\mathbf{x} - \mathbf{X}_t) = 0. \tag{24}$$

Following Cai (2002), Wu et al. (2008) considered using the kernel method to estimate conditional quantiles for both short-range and long-range dependent process. The Bahadur representations and central limit theorems for the NW estimators of those processes were established via the NW methods. Moreover, conditional on the historical information or a state process, the asymptotic theory for the kernel estimates of the VaR was also proposed. Cai and Wang (2008) moved one step further and combined the double kernel local linear technique of Yu and Jones (1998) and the WNW method of Cai (2002) to propose the "weighted double kernel local linear" (WDKLL) estimator. In their paper, they started with nonparametric estimation of the conditional PDF and CDF. For a given symmetric kernel $K(\cdot)$, as $h_0 \to 0$, it is readily seen that $E[K_{h_0}(y - Y_t)|\mathbf{X}_t = \mathbf{x}] \approx f(y|\mathbf{x})$ with $K_{h_0}(\cdot) = K(\cdot/h_0)/h_0$, which means that $Y_t^*(y) = K_{h_0}(y - Y_t)$ can be regarded as an initial estimator of $f(y|\mathbf{x})$ smoothing in the y direction. Then, the local linear technique is applied to the observed variable $Y_t^*(y)$ versus $\mathbf{X}_t$, which leads to

$$\sum_{t=1}^{n} [Y_t^*(y) - a - b(\mathbf{X}_t - \mathbf{x})]^2 W_{h_1}(\mathbf{X}_t - \mathbf{x}), \tag{25}$$

where the bandwidth $h_1 = h_1(n) > 0$ at this stage satisfies $nh_1 \to \infty$ as $n \to \infty$. Minimizing (25) with respect to a and b, a locally weighted least squares estimator of $f(y|\mathbf{x})$ is expressed as

$$\widehat{f_{ll}}(y|\mathbf{x}) = \sum_{t=1}^{n} W_{ll,t}(\mathbf{x}, h_1)Y_t^*(y), \tag{26}$$

where the weights are given by

$$W_{ll,t}(\mathbf{x}, h_1) = \frac{[S_{n,2}(\mathbf{x}) - (\mathbf{x} - \mathbf{X}_t)S_{n,1}(\mathbf{x})]W_{h_1}(\mathbf{x} - \mathbf{X}_t)}{S_{n,0}(\mathbf{x})S_{n,2}(\mathbf{x}) - S_{n,1}^2(\mathbf{x})}, \tag{27}$$

with $S_{n,j}(\mathbf{x}) = \sum_{t=1}^{n} W_{h_1}(\mathbf{x} - \mathbf{X}_t)(\mathbf{X}_t - \mathbf{x})^j$. From Fan and Gijbels (1996), it follows that $W_{ll,t}(\mathbf{x}, h_1)$ satisfies the following discrete moments condition,

$$\sum_{t=1}^{n} W_{ll,t}(\mathbf{x}, h_1)(\mathbf{X}_t - \mathbf{x})^j = \begin{cases} 1, & \text{if } j = 0 \\ 0, & \text{otherwise} \end{cases}, \tag{28}$$

for $0 \leq j \leq 1$. Furthermore, by integrating $\widehat{f_{ll}}(y|\mathbf{x})$, the double kernel local linear estimator of $F(y|\mathbf{x})$ is constructed as

$$\widehat{F_{ll}}(y|\mathbf{x}) = \int_{-\infty}^{y} \widehat{f_{ll}}(v|\mathbf{x})\, dv = \sum_{t=1}^{n} W_{ll,t}(\mathbf{x}, h_1)G_{h_0}(y - Y_t), \tag{29}$$

where $G_{h_0}(\cdot) = G(\cdot/h_0)$ with $G(\cdot)$ denoting the CDF of $K(\cdot)$. Clearly, $\widehat{F}_{ll}(y|\mathbf{x})$ satisfies some basic properties of a distribution function, such as differentiability with respect to y, $\widehat{F}_{ll}(-\infty|\mathbf{x}) = 0$, and $\widehat{F}_{ll}(\infty|\mathbf{x}) = 1$.

Since both WNW estimators $\widehat{F}_w(y|\mathbf{x})$ and double kernel local linear estimator $\widehat{F}_{ll}(y|\mathbf{x})$ have attractive properties, by combining them under a unified framework, the WDKLL estimator is obtained as follows

$$\widehat{f}_c(y|\mathbf{x}) = \sum_{t=1}^{n} W_{c,t}(\mathbf{x}, h_1) Y_t^*(y), \tag{30}$$

where $W_{c,t}(\mathbf{x}, h_1)$ is given in (23), and

$$\widehat{F}_c(y|\mathbf{x}) = \int_{-\infty}^{y} \widehat{f}_c(v|\mathbf{x})\, dv = \sum_{t=1}^{n} W_{c,t}(\mathbf{x}, h_1) G_{h_0}(y - Y_t). \tag{31}$$

Clearly, both $\widehat{f}_c(y|\mathbf{x})$ and $\widehat{F}_c(y|\mathbf{x})$ satisfy the conditions for the PDF and CDF. Therefore, the nonparametric estimator of VaR can be constructed as $q_\tau(y|\mathbf{x}) = \widehat{S}_c^{-1}(\tau|\mathbf{x})$, where $\widehat{S}_c(y|\mathbf{x}) = 1 - \widehat{F}_c(y|\mathbf{x})$, and ES estimator is obtained in a similar way as

$$\widehat{\mathrm{ES}}_\tau(\mathbf{x}) = \frac{1}{\tau} \sum_{t=1}^{n} W_{c,t}(\mathbf{x}, h_1)[Y_t \bar{G}_{h_0}(\widehat{q}_\tau(\mathbf{x}) - Y_t) + h_0 G_{1,h_0}(\widehat{q}_\tau - Y_t)], \tag{32}$$

where $\bar{G}(z) = 1 - G(z)$ and $G_{1,h_0}(z) = G_1(z/h_0)$ with $G_1(z) = \int_z^\infty v K(v)\, dv$. The proposed nonparametric estimators enjoy asymptotic normality and consistency at both the boundary and interior points.

3.2.2 Extreme Value Theory

Volatility clustering is a typical feature of the return series. More importantly, it is difficult to distinguish the difference between volatility clustering of a stationary process and time-varying volatility due to nonstationarity with unit root. Therefore, to accommodate the potential of violating stationary assumption caused by erratic changes, Chavez-Demoulin et al. (2014) extended the classical peak over threshold (POT) method to fit time-varying volatility in the estimation of VaR, which is applicable to both stationary and nonstationary time series. Specifically, in the first step, on the basis of the conditional probability principle, the conditional distribution $F_{y|\mathcal{F}_{t-1}}(\cdot)$ of loss variable Y_t in the upper tail is decomposed as

$$P(Y_t > y|\mathcal{F}_{t-1}) = P(Y_t > u|\mathcal{F}_{t-1})P(Y_t - u > y - u|Y_t > u, \mathcal{F}_{t-1}), \tag{33}$$

where u is a predetermined threshold with $u < y$. Then, the weekly counts $N_u(l)$ of the losses above the threshold u during week l are modeled independently with the excess W_t captured by a negative loss that exceeds u at time t. Following POT, $N_u(l)$ follows a Poisson distribution with parameter λ_l^w and $W_t \sim \mathrm{GPD}(w_t; \varsigma_t, \psi_t)$. Note that the parameters of the Poisson and GPD distributions are allowed to vary over time, which is different from the classical POT approach. They further assumed that these parameters are realizations of independent hidden processes, which can be modeled by Laplace innovations in addition to maximizing a-posteriori estimation. Thus, smooth coefficient estimators with occasionally abrupt temporal changes can be obtained. Specifically, weekly parameters $\eta_l^w = \log \lambda_l^w$ and $\varphi_t = \log \psi_t$ are assumed to follow a temporal first-order Markov process with Laplace innovations $\eta_{l+1}^w|\eta_l^w = \mathrm{Laplace}(\eta_l^w, \iota_1)$ and $\varphi_{t+1}|\varphi_t = \mathrm{Laplace}(\varphi_t, \iota_2)$, respectively. Denote $\mathrm{Laplace}(\eta, \iota)$ as a Laplace distribution with location parameter η and scale parameter $\iota > 0$, where ι reflects the degree of the change.

A small ι implies abrupt and frequent changes and vice versa for a large ι. Bayes' theorem can be applied to obtain the log-posterior distributions of η^w and φ as

$$L(\eta_1^w, \ldots, \eta_L^w; n_u(1), \ldots, n_u(L)) = \sum_{l=1}^{L} (n_u(l)\eta_l^w - \exp \eta_l^w) - \gamma_1 \sum_{l=2}^{L} |\eta_l^w - \eta_{l-1}^w|, \quad (34)$$

and

$$L(\varphi_1, \ldots, \varphi_{n_u}, \varsigma; w_1, \ldots, w_{n_u}) = \sum_{t=1}^{n_u} \left(-\varphi_t - (1 + 1/\varsigma) \log\{1 + \varsigma w_t \exp(-\varphi_t)\} \right) \\ - \gamma_2 \sum_{t=1}^{n_u-1} |\varphi_{t+1} - \varphi_t|, \quad (35)$$

given the Poisson count $n_u(1), \ldots, n_u(L)$. An iterated dual mode (IDM) algorithm is employed here to compute the smooth maximum a-posteriori estimates of $(\eta_1^w, \ldots, \eta_L^w)$, ς and $(\varphi_1, \ldots, \varphi_{n_u})$, which is easy to implement and has guaranteed convergence in this situation.

In the above model, the Poisson parameters are weekly estimated, and the GPD parameters are estimated at the times of exceedance using the nonparametric Bayesian smoothing introduced above. To obtain daily parameter estimates, homogeneity is assumed within each week, so one has $\widehat{\lambda}_t = \widehat{\lambda}_l^w/5$, with piecewise constancy between days of exceedances for ς of the GPD, i.e.,

$$\widehat{\varsigma}_t = \widehat{\varsigma}_{t_i} \quad for \ t_i \le t < t_{i+1}, \ i = 1, \ldots, n_u. \quad (36)$$

Clearly, not only the point estimates of $\{\widehat{\theta}_t = (\widehat{\lambda}_t, \widehat{\varsigma}_t, \widehat{\psi})\}_{t=1}^{T}$ and $(\widehat{\gamma}_1, \widehat{\gamma}_2)$ but also a measure of uncertainty by means of the predictive distributions are provided. Given the estimated Poisson parameter $\widehat{\eta}_L^w$ at week L, one can estimate the distribution of Poisson parameter η_{L+1}^w at week $L + 1$ by

$$\eta_{L+1}^w | \eta_L^w = \widehat{\eta}_L^w \sim \text{Laplace}(\widehat{\eta}_L^w, \widehat{\gamma}_1). \quad (37)$$

Meanwhile, the GPD parameter φ_{T+1} at time $T + 1$ can be estimated by

$$\varphi_{T+1} | \varphi_T = \widehat{\varphi}_T \sim \begin{cases} \delta_{\widehat{\varphi}_T}, & \text{with } 1 - \widehat{P}(X_T \ge u), \\ \text{Laplace}(\widehat{\varphi}_T, \widehat{\gamma}_2), & \text{with } \widehat{P}(X_T \ge u), \end{cases} \quad (38)$$

where δ_φ is the Dirac mass at φ. After smoothing the data up to time T, we take the best bias-variance trade-off estimates $\widehat{\lambda}_T$ and $\widehat{\varphi}_T$ and insert them into (37) and (38) to obtain the approximate distribution of $\theta_{T+1} | \theta_T = \widehat{\theta}_T = (\widehat{\lambda}_T, \widehat{\varsigma}_T, \widehat{\psi})$. Finally, one has

$$\widehat{\text{VaR}}_\tau(Y_T) = F_{\theta_{T+1}|\widehat{\theta}_T}^{-1}(\tau), \quad (39)$$

where

$$F_{\theta_{T+1}|\widehat{\theta}_T}(Y) = 1 - \{1 - \exp(-\lambda_{T+1}|\widehat{\lambda}_T)\}(1 + \widehat{\psi}\frac{Y - u}{\varsigma_{T+1}|\widehat{\varsigma}_T})^{-1/\widehat{\psi}}.$$

ES within this conditional setup can be obtained by following the same procedure as that in POT.

By combining a nonparametric method and EVT, Martins-Filho et al. (2018) proposed a two-stage estimation procedure for VaR and ES. Assume that the loss variable $\{Y_t\}$ follows a location-scale process formulated as

$$Y_t = m(\mathbf{X}_t) + h^{1/2}(\mathbf{X}_t)\epsilon_t, \quad (40)$$

as in Yao and Tong (1996), where $m(\mathbf{X}_t)$ and $h(\mathbf{X}_t)$ are unknown functions of $\mathbf{X}_t$, and $\{\epsilon_t\}$ is an i.i.d. innovation process with $E(\epsilon_t) = 0$ and $\text{Var}(\epsilon_t) = 1$. The distribution function $F_\epsilon(\cdot)$ of

ϵ_t belongs to the maximum domain of attraction of a Fréchet distribution with parameter $-1/\kappa$, denoted by $F_\epsilon(\cdot) \in D(\mathbb{F}_{-1/\kappa})$. Under this setting, one has

$$
\begin{aligned}
\text{VaR}_\tau(\mathbf{x}) &\equiv q_\tau(\mathbf{x}) = m(\mathbf{x}) + h^{1/2}(\mathbf{x})q_\tau(\epsilon_t), \\
\text{ES}_\tau(\mathbf{x}) &\equiv E[Y_t|Y_t > q_\tau(\mathbf{x})] = m(\mathbf{x}) + h^{1/2}(\mathbf{x})E(\epsilon_t|\epsilon_t > q_\tau(\epsilon_t)),
\end{aligned}
\tag{41}
$$

where $q_\tau(\mathbf{x})$ is the conditional τ-quantile of Y_t given $\mathbf{X}_t = \mathbf{x}$ that is associated with distribution function $F_{y|\mathbf{X}}(\cdot)$, and $q_\tau(\epsilon_t)$ denotes the τ-quantile of ϵ_t. In the first stage, the local linear approach is employed for the estimation of $m(\mathbf{x})$, and the obtained estimator is denoted as $\widehat{m}(\mathbf{x})$. The procedure proposed by Fan and Yao (1998) is followed for the estimation of $h(\mathbf{x})$. Specifically, the residual sequence is obtained by $Y_t^* = Y_t - \widehat{m}(\mathbf{x})$; then,

$$
(\widehat{a}(\mathbf{x}), \widehat{b}(\mathbf{x})) = \arg\min_{a,b} \sum_{t=1}^n \left(Y_t^{*2} - a - b^\top(\mathbf{X}_t - \mathbf{x})\right)^2 K_h(\mathbf{X}_t - \mathbf{x}).
\tag{42}
$$

With estimators $\widehat{m}(\mathbf{x})$ and $\widehat{h}(\mathbf{x}) = \widehat{a}(\mathbf{x})$, a sequence of standardized nonparametric residuals $\{\epsilon_t\}$ can be generated by

$$
\widehat{\epsilon}_t = \begin{cases} (Y_t - \widehat{m}(\mathbf{X}_t))/\widehat{h}^{1/2}(\mathbf{X}_t), & \text{if } \widehat{h}(\mathbf{X}_t) > 0, \\ 0, & \text{otherwise,} \end{cases}
\tag{43}
$$

for $t = 1, \ldots, n$. In the second stage, to estimate VaR_τ and ES_τ, these residuals are used to construct estimators for $q_\tau(\epsilon_t)$ and $E(\epsilon_t|\epsilon_t > q_\tau(\epsilon_t))$. To this end, Theorem 7 in Pickands (1975), i.e., if and only if $F_\epsilon(\cdot) \in D(\mathbb{F}_{-1/\kappa})$, for $\kappa < 0$ and some function $\psi(\zeta) > 0$ with $\zeta \in \mathcal{R}$, is employed to obtain

$$
\lim_{\zeta \to \infty} \sup_{\zeta + w < \infty} |F_{\zeta,\epsilon_t}(w) - \mathbb{G}(w; \varsigma, \psi(\zeta))| = 0,
\tag{44}
$$

where $F_{\zeta,\epsilon}(w) = [F_\epsilon(w + \zeta) - F_\epsilon(\zeta)]/[1 - F_\epsilon(\zeta)]$ and $\mathbb{G}(w; \varsigma, \psi) \equiv 1 - (1 - \varsigma w/\psi)^{1/\varsigma}$ with $0 < w < \infty$ is a GPD. The main idea of this theorem is that when $F_\epsilon(\cdot) \in D(\mathbb{F}_{-1/\kappa})$, the extreme upper tail of ϵ_t is uniformly close to a GPD. As in Davis and Resnick (1984) and Smith (1987), (44) is used to motivate the estimation of $q_\tau(\epsilon_t)$. After some algebra,

$$
q_\tau(\epsilon_t) \approx q_{\tau_N}(\epsilon_t) + \frac{\psi(q_{\tau_N}(\epsilon_t))}{\kappa}\left(1 - \left(\frac{n}{N}(1 - \tau)\right)^\kappa\right),
\tag{45}
$$

where $\tau_N = 1 - N/n$ with $N \to \infty$ and $N/n \to 0$. Following Smith (1987), κ and $\psi(q_{\tau_N}(\epsilon_t))$ can be estimated using the approximation provided by $\mathbb{G}(w; \varsigma, \psi(q_{\tau_N}(\epsilon_t)))$ to obtain $\widehat{q}_\tau(\epsilon_t)$. For the estimation of $E(\epsilon_t|\epsilon_t > q_\tau(\epsilon_t))$, the exceedances over the quantile $q_\tau(\epsilon_t)$ are assumed to be distributed as $g(w_1; \varsigma_1, \psi_1) = \psi_1^{-1}(1 - \varsigma_1 w_1/\psi_1)^{1/(\varsigma_1 - 1)}$, which is the density function associated with the GPD. Then,

$$
\widehat{E}(\epsilon_t|\epsilon_t > q_\tau(\epsilon_t)) = \widehat{q}_\tau(\epsilon_t)/[1 + \widehat{\varsigma}_1],
\tag{46}
$$

where $\widehat{\varsigma}_1$ is the estimator of ς_1 obtained in the last step. Finally, VaR_τ and ES_τ can be estimated by plugging $\widehat{q}_\tau(\epsilon_t)$ and $\widehat{E}(\epsilon_t|\epsilon_t > q_\tau(\epsilon_t))$ into (41).

3.2.3 Quantile Approach

Nonparametric quantile regression is one of the most popular research areas in econometrics. Among the vast number of research papers, refer to Honda (2000, 2004), Kim (2007), Cai and Xu (2008), Cai and Xiao (2012) and the references therein for the varying coefficient quantile model; Zou and Yuan (2008), Kai et al. (2010) and Kai et al. (2011) for composite quantile regression; and Wu et al. (2010), Kong and Xia (2012) and Fan et al. (2017) for the single index quantile model. Overall, a detailed discussion of these models can be found in Koenker (2005) and thus the details are omitted.

4 Expectile Models

In this section, we survey the existing literature for the estimation of expectile models.

4.1 Parametric Models

To extend the CAViaR model to accommodate the expectile framework, Kuan et al. (2009) proposed CARE models, which allow expectiles to be estimated in a dynamic context based on some special types of autoregressive processes. The τ-conditional expectile of return variable R_t given information set $\mathcal{F}_{t-1}$, can be obtained by solving the following optimization problem

$$e_{t,\tau} = \arg\min_{r \in \mathcal{R}} E\{|\tau - I(R_t < r)|(R_t - r)^2 \,|\, \mathcal{F}_{t-1}\}, \tag{47}$$

and a generic CARE model has the form,

$$R_t = e_{t,\tau} + \epsilon_{t,\tau}, \quad \text{Expec}_\tau(\epsilon_{t,\tau}|\mathcal{F}_{t-1}) = 0, \text{ and } e_{t,\tau} = \beta_{0,\tau} + \sum_{i=1}^{p} \beta_{i,\tau} g_i(R_{t-i}), \tag{48}$$

where $\text{Expec}_\tau(\epsilon_{t,\tau}|\mathcal{F}_{t-1})$ denotes the τ-conditional expectile of $\epsilon_{t,\tau}$ given $\mathcal{F}_{t-1}$, and $g_i(\cdot)$ is a nonlinear function. Two model specifications of the CARE approach exist. The first specification considers the asymmetric effects on tail expectiles and includes the magnitudes of positive and negative lagged returns in the model:

$$e_{t,\tau} = \beta_{0,\tau} + \beta_{1,\tau} R_{t-1} + \beta_{2,\tau}(R_{t-1}^+)^2 + \beta_{3,\tau}(R_{t-1}^-)^2. \tag{49}$$

Alternatively, $|R_{t-1}|$ is considered to represent the magnitude of R_{t-1} in the second specification as

$$e_{t,\tau} = \beta_{0,\tau} + \beta_{1,\tau} R_{t-1}^+ + \beta_{2,\tau} R_{t-1}^-. \tag{50}$$

Kuan et al. (2009) defined the expectile as expectile-based value at risk (EVaR), $\text{EVaR}(\tau) = |e_\tau|$ with $\tau < 0.5$. EVaR is considered to be a downside risk measure, and further an intuition for τ is given. If EVaR is taken as a margin (capital requirement), then the probability level τ can be understood as the relative cost of the expected margin shortfall. The model estimation employs the asymmetric least squares proposed by Newey and Powell (1987), and the asymptotic results are extended to allow for stationary and weekly dependent data.

Another independent work based on a parametric setup is Taylor (2008). As noted by Kuan et al. (2009), this paper is concerned mainly with the determination of quantile-based VaR based on expectiles; thus, their models are the same as the CAViaR models of Engle and Manganelli (2004). More recently, Xu et al. (2018) considered the potential time-varying parameter property and employed an adaptive method to fit a parametric expectile model for quantifying tail risk dynamics.

4.2 Nonparametric and Semiparametric Models

As noted by Cai and Xiao (2012), a linear regression model may not be sufficiently flexible to capture the underlying complex dependence structure in some practical applications. An increasing number of researchers is focusing on nonparametric or semiparametric expecitle models. Xie et al. (2014) proposed a varying coefficient expectile model. Afterwards, Cai et al. (2018) extended their work to consider a more general case, i.e., a partially varying coefficient expectile model.

4.2.1 A Varying Coefficient Expectile Model

The general form of the varying coefficient expectile model proposed by Xie et al. (2014) is

$$e_{t,\tau}(\mathbf{X}_t, U_t) = \boldsymbol{\beta}_\tau^\top(U_t)\mathbf{X}_t, \tag{51}$$

where $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots, X_{pt})$ are risk factors, U_t is a single effect modifying risk factor and $\boldsymbol{\beta}_\tau(U_t) = (\boldsymbol{\beta}_{1,\tau}(U_t), \boldsymbol{\beta}_{2,\tau}(U_t), \ldots, \boldsymbol{\beta}_{p,\tau}(U_t))^\top$ is a vector of smooth varying-coefficient functions of U_t. The introduction of U_t considers current information and can also include exogenous variables containing economic and market information as risk factors. To estimate $\boldsymbol{\beta}(\cdot)$, the iterative weighted local least squares (IWLLS), similar to that in Newey and Powell (1987) and Yao and Tong (1996), is employed. Furthermore, the consistency and asymptotic normality of the proposed estimator are established. Unfortunately, note that the aforementioned models in (49), (50) and (51) do not include any lagged variable of $e_{\tau,t}$.

4.2.2 A Partially Varying-Coefficient Expectile Model

A purely nonparametric expectile model may suffer from the curse of dimensionality. To address this problem, Cai et al. (2018) proposed a partially varying coefficient expectile model,

$$e_{\tau,t}(\mathbf{X}_t, U_t) = \boldsymbol{\alpha}_\tau^\top \mathbf{X}_{t1} + \boldsymbol{\beta}_\tau^\top(U_t)\mathbf{X}_{t2}, \tag{52}$$

where $\mathbf{X}_t = (\mathbf{X}_{t,1}^\top, \mathbf{X}_{t,2}^\top)^\top \in \mathcal{R}^{p+q}$ is a vector of risk factors, U_t is called the smooth variable, $\boldsymbol{\alpha}_\tau \in \mathcal{R}^p$ are constant coefficients and $\boldsymbol{\beta}_\tau(\cdot) = (\beta_{1,\tau}(\cdot), \ldots, \beta_{q,\tau}(\cdot))^\top$ is a vector of smooth varying-coefficient functions that are twice continuously differentiable. This new model adopts a partially linear form, in which some coefficients are assumed to be constant while other coefficients are allowed to depend on smoothing variables selected by economic theories or stylized facts. The model is quite flexible such that it includes both models in Kuan et al. (2009) and Xie et al. (2014) as special cases. The new model not only shares all the merits of a fully varying-coefficient model but also achieves more efficient estimation for the parametric coefficient part. More importantly, model (52) allows $\mathbf{X}_t$ to include the lagged variables of $e_{\tau,t}$.

The typical method used for estimating parameters in a semiparametric model is to employ the so-called profile least squares type method for classical regression models. But it might not be suitable in quantile/expectile setting due to lack of an explicit mathematical expression; see Cai and Xiao (2012) for details. Therefore, the three-stage semiparametric procedure of Cai and Xiao (2012) is employed to overcome this problem. In the first stage, $\boldsymbol{\alpha}_\tau$ is treated as a function of U_t, and the model simplifies to a purely nonparametric expectile model, which can be estimated via the IWLLS approach. However, in this stage, only local information is used in the estimation of $\boldsymbol{\alpha}_\tau$, which it is actually a global parameter. Thus, in the second stage, the average method is implemented to take advantage of the full sample information, and a $\sqrt{n}$-consistent estimator for $\boldsymbol{\alpha}_\tau$ is obtained as

$$\widetilde{\boldsymbol{\alpha}}_\tau = \frac{1}{n} \sum_{t=1}^n \widehat{\boldsymbol{\alpha}}_\tau(U_t). \tag{53}$$

Finally, the estimated partial expectile residual $R_t^* = R_t - \widetilde{\boldsymbol{\alpha}}_\tau^\top \mathbf{X}_{t1}$ is used to acquire a feasible local linear functional coefficient estimation of $\boldsymbol{\beta}_\tau(\cdot)$. Consistent estimators for $\boldsymbol{\alpha}_\tau$ and $\boldsymbol{\beta}_\tau(\cdot)$ are obtained through the three-stage procedure, and the asymptotic properties of the proposed estimator are established.

5 Models for Characterizing Tail Dependence

5.1 Copula Approach

After the financial crisis, people realized that the failure of one institution may endanger the whole financial system. Tail dependence, which measures financial co-movement, aroused the interests of many researchers and is widely accepted to be asymmetric, nonlinear and time varying. Copula models, which allow for flexible modeling of nonlinear dependence structures, have been advocated as alternatives to correlation-based models to measure tail dependence. Among the vast number of papers aiming at using copula models to forecast the VaR, Siburg et al. (2015) considered employing copulas calibrated on the basis of nonparametric sample estimators of the coefficient of lower tail dependence (LTD). Assume that R_1 and R_2 are returns variables with continuous marginal distribution functions $F_1(\cdot)$ and $F_2(\cdot)$. Then, LTD is defined as the limit of the conditional probability that R_1 is less than or equal to the τ-th quantile of F_1 given that R_2 is less than or equal to the τ-th quantile of F_2 as τ goes to 0. Specifically, one has

$$\text{LTD} = \lim_{\tau \to 0^+} P\big(R_1 \leq F_1^{-1}(\tau) | R_2 \leq F_2^{-1}(\tau)\big). \tag{54}$$

From the definition above, LTD can be interpreted as the limiting likelihood of two financial assets to crash simultaneously. As a copula is always an appropriate tool to model the dependence, according to Sklar's theorem, there exists a unique copula $C(\cdot, \cdot)$ such that

$$P(R_1 \leq r_1, R_2 \leq r_2) = C(F_1(r_1), F_2(r_2)). \tag{55}$$

Normally, if the copula is known to come from a parametric copula family, the canonical maximum-likelihood (CML) estimator can be used. However, when the copula family is mis-specified and the CML estimator is no longer optimal, the choice of the optimal copula parametric family remains an open empirical question. A nonparametric estimator of the LTD was employed in this paper to improve the accuracy of the copula model. To this end, LTD can be re-written as

$$\text{LTD} = \lim_{\tau \to 0^+} C(\tau, \tau)/\tau. \tag{56}$$

Now, let $\{R_{1t}, R_{2t}\}_{t=1}^n$ be an i.i.d. sample of $\{R_1, R_2\}$. The specific nonparametric estimator for LTD, introduced by Schmidt and Stadtmüller (2006), is given by

$$\text{LTD}_n = \frac{1}{k} \sum_{t=1}^n I(R(K_{1t}) \leq k, R(K_{2t}) \leq k), \tag{57}$$

where $I(\cdot, \cdot)$ is indicator function, $R(K_{1t})$ and $R(K_{2t})$ are the ranks of observations R_{1t} and R_{2t} respectively, and $k \in \{1, \ldots, n\}$ is chosen by a plateau-finding algorithm. The procedure of forecasting the bivariate portfolio VaR is as follows. First, financial data are filtered with a GARCH(1,1) model to obtain approximately i.i.d. samples of standardized residuals. Considering that LTD is commonly used to characterize the dependence in the financial market data, here they opted the easily implemented Clayton copula, which is denoted by $C_\theta(\cdot, \cdot)$ with parameter θ. Then, the parameters of the marginal models are estimated via maximum-likelihood, and the copula parameter θ is obtained by converting the nonparametric estimator of LTD into the parameter of the Clayton copula. Finally, an out-of-sample version of the simulation procedures laid out by Nikoloulopoulos et al. (2012) is used to obtain the forecasters for VaR and ES.

5.2 Network and CoVaR Procedure

Another way to model the tail dependence is the CoVaR approach, which was first introduced by Adrian and Brunnermeier (2016), one of the most popular approaches of modeling systemic

risks. The CoVaR of institution i relative to institution j is defined as the VaR of institution i conditioned on institution j being in distress. A related risk measure is the marginal expected shortfall (MES) proposed by Acharya et al. (2012) and Acharya et al. (2017), which tracks the sensitivity of an individual firm's return to a market-wide extreme event. To develop a unified framework for conceptualizing and empirically measuring this type of connectedness between institutions, Diebold and Yılmaz (2014) proposed the general framework of network approach. Connectedness, in their paper, is based on assessing shares of forecast error variation in various locations (firms, markets, countries, etc.) due to shocks arising everywhere. This notion is intimately related to the economic explanation of variance decomposition, in which we decompose the forecast error variance of variable i into parts attributed to other variables in the system. In particular, connectedness can be modeled by the function $C(X, H, A(L))$, where X is the set of objects interested, H is the prediction horizon for the variance decompositions, and $A(L)$ are the dynamics. This connectedness function refers to a population, so we need to find estimated approximating models $\widehat{C}(X, H, M(\widehat{\theta}_t))$, in which $M(\widehat{\theta}_t)$ is a dynamic model approximating $A(L)$. Meanwhile, the connectedness table D (variance decomposition matrix), which is very similar to the adjacency matrix except that it is not filled simply with 0-1 entries and it is asymmetric, is introduced to describe the links between individuals in the system. As the connected table is asymmetric, the links between objects are directed, so we need to classify them into the From-degrees and To-degrees. From-degrees measure the exposures of individual firms to system shocks from the network, whereas To-degrees measure the contributions of individual firms to system network events. As CoVaR and MES are all aimed at tracking the associations between individual firms and the overall market movement, and they are weighted and directional similar to the connectedness measurement, Diebold and Yılmaz (2014) unified these two measures and proposed that they are closely related to different directional aggregations of a certain weighted directed network.

To study the degree of tail dependence among various institutions, namely, financial network system risk, Hautsch et al. (2014) considered applying a two-stage quantile regression approach to account for a company's interconnectedness within the financial sector. In the first step, the firms' specific VaRs are modeled by firm characteristics, macro state variables, and tail risk spillover effects captured by loss exceedances. This model framework rules out the possibility that the identified risk connections result from common risk factors. A well known least absolute shrinkage and selection operator technique is used to shrink the high-dimensional covariates to a feasible number of relevant risk connections. In the second step, to measure a firm's system impact, a value-weighted index of the financial sector is constructed and then regressed with the firm's VaR estimated in step 1 and other control variables, such as pre-identified company-specific risk drivers and the macro variables. If the firm's VaR has a significant and nonnegative marginal effect on the VaR of the system, the company is said to be systemically relevant. In this model, even if individual risk remains constant, the systemic impact of a firm can vary over time due to varying market or balance sheet conditions, which remedies a major drawback of the typical CoVaR approach that it can change only through the channel of individual VaRs.

Considering the potential nonlinear dependence structure in the study of network risk, Härdle et al. (2016) proposed a single index model (SIM) to estimate system interconnectedness across financial institutions in a high-dimensional framework. Their estimation procedure can be illustrated by three steps. In the first step, as in Adrian and Brunnermeier (2016), the VaR is estimated for each financial institution in the system through linear quantile regression,

$$Y_{it} = \alpha_{i,\tau} + \boldsymbol{\gamma}_{i,\tau}\mathbf{M}_{t-1} + \epsilon_{it}, \tag{58}$$

where $\mathbf{M}_{t-1}$ is a vector of macro state variables. In this way, the τ-VaR for institution i at time t can be estimated as $\widehat{\text{VaR}}_{it,\tau} = \widehat{\alpha}_{i,\tau} + \widehat{\boldsymbol{\gamma}}_{i,\tau}\mathbf{M}_{t-1}$. In the second step, to construct a risk interdependence network, the SIM for quantile regression is employed. This model captures the nonlinear dependence structure because it is based on a SIM quantile variable selection

technique, more precisely,

$$Y_{it} = g(\boldsymbol{\beta}_{i|\mathbf{z}_i}^\top \mathbf{Z}_{it}) + \epsilon_{it}, \tag{59}$$

where $g(\cdot)$ is an unknown function, $\boldsymbol{\beta}_{i|\mathbf{z}_i} = \{\boldsymbol{\beta}_{i|-i}, \boldsymbol{\beta}_{i|M}, \boldsymbol{\beta}_{i|B_i}\}$ is a vector of coefficients, and $\mathbf{Z}_{it} = \{\mathbf{Y}_{-i,t}, \mathbf{M}_{t-1}, \mathbf{B}_{i,t-1}\}$ is a vector of covariates, in which $\mathbf{Y}_{-i,t}$ includes log returns of all financial institutions except for institution i, and $\mathbf{B}_{i,t-1}$ are the firm's characteristics. The minimum average contrast estimation approach with penalization outlined in Fan et al. (2017) is employed to estimate the shape of the unknown function $g(\cdot)$ and coefficients $\boldsymbol{\beta}_{i|\mathbf{z}_i}$. Then, the CoVaR of institution i given other institutions in distress can be estimated as

$$\widehat{\text{CoVaR}}_{i|\alpha,\widetilde{\mathbf{z}}_{it}} = \widehat{g}(\widehat{\boldsymbol{\beta}}_{i|\widetilde{\mathbf{z}}_i}^\top \widetilde{\mathbf{Z}}_{it}), \tag{60}$$

where $\widehat{g}(\cdot)$ is the estimated link function, $\widehat{\boldsymbol{\beta}}_{i|\widetilde{\mathbf{z}}_i}^\top$ is the estimator for $\boldsymbol{\beta}_{i|\mathbf{z}_i}$, and $\widetilde{\mathbf{Z}}_{it} = \{\widehat{\text{VaR}}_{-i,t,\tau},\ \mathbf{M}_{t-1}, \mathbf{B}_{i,t-1}\}$ with $\widehat{\text{VaR}}_{-i,t,\tau}$ denoting the VaR of all institutions except institution i estimated in step 1. An index $\widehat{D}_{i|\widetilde{\mathbf{z}}_i}$ was also proposed to measure the marginal effect of covariates evaluated at $\mathbf{Z}_{it} = \widetilde{\mathbf{Z}}_{it}$, which is formulated as follows,

$$\widehat{D}_{i|\widetilde{\mathbf{z}}_i} = \frac{\partial \widehat{g}(\widehat{\boldsymbol{\beta}}_{i|\mathbf{z}_i}^\top \mathbf{Z}_{it})}{\partial \mathbf{Z}_{it}}\Big|_{\mathbf{Z}_{it}=\widetilde{\mathbf{Z}}_{it}} = \widehat{g}'(\widehat{\boldsymbol{\beta}}_{i|\widetilde{\mathbf{z}}_i}^\top \widetilde{\mathbf{Z}}_{it})\widehat{\boldsymbol{\beta}}_{i|\widetilde{\mathbf{z}}_i}. \tag{61}$$

This index can capture spillover effects across financial institutions and further characterize their evolution as a system represented by a network.

The CAViaR model gained considerable attention during the past few years, and the development in both theory and application has occurred mainly in the univariate setup. To study the degree of tail interdependence among different random variables, White et al. (2015) proposed a multivariate regression quantile model that extends the multi-quantile model CAViaR model of White et al. (2008) to the multivariate case. Let $q_{j,t,\tau}$ be the τ-conditional quantile of the j-th random variable Y_{jt} given $\mathcal{F}_{t-1}$ for $j = 1, \ldots, m$ with $m \geq 2$. A simple version of their proposed structure is as follows,

$$q_{j,t,\tau} = \mathbf{X}_t^\top \boldsymbol{\beta}_j + \sum_{k=1}^m b_{jk} q_{k,t-1,\tau} \ \text{ for } m \geq 2, \ \text{ equivalently, } \ \mathbf{q}_{t,\tau} = \boldsymbol{\beta}\mathbf{X}_t + \mathbf{B}\,\mathbf{q}_{t-1,\tau}, \tag{62}$$

where $\mathbf{X}_t$ are covariates with the coefficients $\boldsymbol{\beta}^\top = (\beta_1, \ldots, \beta_m)$, $\mathbf{q}_{t,\tau}^\top = (q_{1,t,\tau}, \ldots, q_{m,t,\tau})$, and $\mathbf{B} = (b_{jk})_{m \times m}$ is the matrix of coefficients of the lagged variables $\mathbf{q}_{t-1,\tau}$ to capture the network system risk. The above model has a VAR structure and is called VAR for VaR. The tail co-dependence between two random variables Y_i and Y_j is captured by the off-diagonal coefficients b_{ij} and b_{ji} in $\mathbf{B}$, and the model reduces to the CAViaR of Engle and Manganelli (2004) if $b_{ij} = 0$ for all $i \neq j$. Furthermore, this model is particularly applicable in many aspects; see White et al. (2015) for more discussions. As the seriousness of risk contagion across financial institutions has been brought to the forefront by the financial crisis, VAR for VaR appears to be particularly suitable to measure the financial spillover effect. Another potential applications of this model might include that quantile impulse-response function can be employed to assess the resilience and persistence of financial institutions to shocks through the overall index and that model (62) can be extended to capture the dynamic network system risk.

6 Conclusion

After the outbreak of the recent financial crisis, the need of financial institutions for accurate risk measures has never been greater. Since the pioneering work of Markowitz (1952), which

proposed variance as the measure of risk, VaR, ES and expectile have become the standard measures of market risk employed by regulators and financial institutions. As the nature of risks is changing over time (dynamic change), methods for measuring these risks must adapt to the current environment. To accommodate the demands of special interests, a vast array of parametric and nonparametric models of these risk measures are proposed. Hopefully, this selective overview provides a better picture of this important field in finance and econometrics.

However, the whole story of methods for the estimation of risk measures is far from complete. New innovative techniques are needed in the framework of high-dimensional and/or high-frequency data for estimating the tail risk. Moreover, as risk contagion is becoming the main concern in the industry, one should put greater emphasis on network analysis and the CoVaR of system risk under different types of dependence structures. More importantly, how to capture the dynamic change in various risks is of great interest in both scientific research and financial practitioners. Finally, expectile has proved to be superior to VaR and ES in some aspects; thus, more critical analysis of the relative merits and potential estimation procedures accustomed to specific regression techniques are required. The measurement of risk developed to confront and address the challenges in the frontiers of technological innovation and scientific research and will continue to advance based on emerging societal needs in the future.

References

Acharya, V. V., Engle, R., and Richardson, M. (2012). Capital shortfall: a new approach to ranking and regulating systemic risks. *American Economic Review*, 102(3):59–64.

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *Review of Financial Studies*, 30(1):2–47.

Adrian, T. and Brunnermeier, M. K. (2016). CoVaR. *American Economic Review*, 106(7):1705–1741.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Basak, S. and Shapiro, A. (2001). Value-at-Risk-based risk management: optimal policies and asset prices. *Review of Financial Studies*, 14(2):371–405.

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

Butler, J. and Schachter, B. (1998). Estimating Value-at-Risk with a precision measure by combining kernel estimation with historical simulation. *Review of Derivatives Research*, 1:371–390.

Cai, Z. (2002). Regression quantiles for time series. *Econometric theory*, 18(1):169–192.

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.*

Cai, Z., Su, J., and Sufianti (2015). A regression analysis of expected shortfall. *Statistics and Its Interface*, 8(3):295–303.

Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, 147(1):120–130.

Cai, Z. and Xiao, Z. (2012). Semiparametric quantile regression estimation in dynamic models with partially varying coefficients. *Journal of Econometrics*, 167(2):413–425.

Cai, Z. and Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 103(484):1595–1608.

Chavez-Demoulin, V., Embrechts, P., and Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, 181(1):44–52.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of Value-at-Risk for dependent financial returns. *Journal of Financial Econometrics*, 3(2):227–255.

Chernozhukov, V. and Umantsev, L. (2001). Conditional Value-at-Risk: aspects of modeling and estimation. *Empirical Economics*, 26(1):271–292.

Danielsson, J. and de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4(2):241–257.

Davis, R. and Resnick, S. (1984). Tail estimates motivated by extreme value theory. *Annals of Statistics*, 12(4):1467–1487.

Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, 4(3):7–49.

Embrechts, P. and Hofert, M. (2014). Statistics and quantitative risk management for banking and insurance. *Annual Review of Statistics and Its Application*, 1:493–514.

Emmer, S., Kratz, M., and Tasche, D. (2015). What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk*, 18(2):31–60.

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications*. Chapman & Hall/CRC Press, New York.

Fan, J., Hu, T.-C., and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics*, 21(4):433–446.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660.

Fan, J., Yao, Q., and Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, 83(1):189–206.

Fan, Y., Härdle, W. K., Wang, W., and Zhu, L. (2017). Single-index-based CoVaR with very high-dimensional covariates. *Journal of Business & Economic Statistics*, 36(2):212–226.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gourieroux, C., Laurent, J.-P., and Scaillet, O. (2000). Sensitivity analysis of values at risk. *Journal of Empirical Finance*, 7(3):225–245.

Hall, P., Wolff, R. C., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94(445):154–163.

Härdle, W. K., Wang, W., and Yu, L. (2016). TENET: tail-event driven network risk. *Journal of Econometrics*, 192(2):499–513.

Hautsch, N., Schaumburg, J., and Schienle, M. (2014). Financial network systemic risk contributions. *Review of Finance*, 19(2):685–738.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Annals of the Institute of Statistical Mathematics*, 52(3):459–470.

Honda, T. (2004). Quantile regression in varying coefficient models. *Journal of Statistical Planning and Inference*, 121(1):113–125.

Hull, J. C. and White, A. D. (1998). Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives*, 5(3):9–19.

Jondeau, E., Poon, S.-H., and Rockinger, M. (2007). *Financial modeling under non-Gaussian distributions*. Spring-Verlag, Berlin.

Jorion, P. (1997). *Value at Risk: A New Benchmark for Measuring Derivatives Risk*. Irwin Professional Publishers, New York.

Kai, B., Li, R., and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(1):49–69.

Kai, B., Li, R., and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Annals of Statistics*, 39(1):305–332.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *Annals of Statistics*, 35(1):92–108.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press, New York.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Koenker, R. and Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12(5):793–813.

Kong, E. and Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, 28(4):730–768.

Kuan, C.-M., Yeh, J.-H., and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the conditional autoregressive expectile models. *Journal of Econometrics*, 150(2):261–270.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-Risk prediction: a comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Lambert, N. S., Pennock, D. M., and Shoham, Y. (2008). Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138. Chicago, IL, USA.

Maguluri, G. and Zhang, C.-H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society. Series B(Methodological)*, 56(3):477–489.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Martins-Filho, C., Yao, F., and Torero, M. (2018). Nonparametric estimation of conditional Value-at-Risk and expected shortfall based on extreme value theory. *Econometric Theory*, 34(1):23–67.

McNeil, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin*, 27(1):117–137.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3):271–300.

Morgan, J. P. (1996). *Value at Risk*. RiskMetrics Technical Document, New York.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2012). Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics & Data Analysis*, 56(11):3659–3673.

Oakes, D. and Dasu, T. (2003). Inference for the proportional mean residual life model. *Lecture Notes-Monograph Series*, 43:105–116.

Osband, K. (1985). *Providing Incentives for Better Cost Forecasting*. PhD thesis, University of California, Berkeley.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131.

Rocco, M. (2014). Extreme value theory in finance: a survey. *Journal of Economic Surveys*, 28(1):82–108.

Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scandinavian Journal of Statistics*, 33(2):307–335.

Siburg, K. F., Stoimenov, P., and Weiß, G. N. (2015). Forecasting portfolio-Value-at-Risk with nonparametric lower tail dependence estimates. *Journal of Banking & Finance*, 54:129–140.

Smith, R. L. (1987). Estimating tails of probability distributions. *Annals of Statistics*, 15(3):1174–1207.

Taylor, J. W. (1999). A quantile regression approach to estimating the distribution of multiperiod returns. *Journal of Derivatives*, 7(1):64–78.

Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252.

Wang, C.-S. and Zhao, Z. (2016). Conditional Value-at-Risk: semiparametric estimation and inference. *Journal of Econometrics*, 195(1):86–103.

White, H. L., Kim, T.-H., and Manganelli, S. (2008). Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR. ECB Working Paper, Department of Economics, University of California at San Diego.

White, H. L., Kim, T. H., and Manganelli, S. (2015). VAR for VaR: measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187(1):169–188.

Wu, T. Z., Yu, K., and Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*, 101(7):1607–1621.

Wu, W. B., Yu, K., and Mitra, G. (2008). Kernel conditional quantile estimation for stationary processes with application to conditional Value-at-Risk. *Journal of Financial Econometrics*, 6(2):253–270.

Xie, S., Zhou, Y., and Wan, A. T. (2014). A varying-coefficient expectile model for estimating value at risk. *Journal of Business & Economic Statistics*, 32(4):576–592.

Xu, X., Mihoci, A., and Härdle, W. K. (2018). lcare - localizing conditional autoregressive expectiles. *Journal of Empirical Finance*, 48:198–220.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, 6(2-3):273–292.

Yu, K. and Jones, M. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics & Data Analysis*, 25(2):159–166.

Yu, K. and Jones, M. (1998). Local linear quantile regression. *Journal of the American statistical Association*, 93(441):228–237.

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, 26(4):901–918.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics*, 36(3):1108–1126.