



# Nonparametric estimation of conditional VaR and expected shortfall

Zongwu Cai<sup>a,b,\*</sup>, Xian Wang<sup>a</sup>

<sup>a</sup> Department of Mathematics and Statistics and Department of Economics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

<sup>b</sup> Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, Fujian 361005, China

## ARTICLE INFO

### Article history:

Available online 11 September 2008

### JEL classification:

C14  
D81  
G10  
G22  
G31

### Keywords:

Boundary effects  
Empirical likelihood  
Expected shortfall  
Local linear estimation  
Nonparametric smoothing  
Value-at-risk  
Weighted double kernel

## ABSTRACT

This paper considers a new nonparametric estimation of conditional value-at-risk and expected shortfall functions. Conditional value-at-risk is estimated by inverting the weighted double kernel local linear estimate of the conditional distribution function. The nonparametric estimator of conditional expected shortfall is constructed by a plugging-in method. Both the asymptotic normality and consistency of the proposed nonparametric estimators are established at both boundary and interior points for time series data. We show that the weighted double kernel local linear conditional distribution estimator has the advantages of always being a distribution, continuous, and differentiable, besides the good properties from both the double kernel local linear and weighted Nadaraya–Watson estimators. Moreover, an ad hoc data-driven fashion bandwidth selection method is proposed, based on the nonparametric version of the Akaike information criterion. Finally, an empirical study is carried out to illustrate the finite sample performance of the proposed estimators.

Published by Elsevier B.V.

## 1. Introduction

Value-at-risk (hereafter, VaR) and expected shortfall (ES) have become two popular measures of market risk associated with an asset or portfolio of assets, during the last decade. In particular, VaR has been chosen by the Basel Committee on Banking Supervision as the benchmark of risk measurement for capital requirements. Both VaR and ES have been used by financial institutions for asset management and minimization of risk, and have been rapidly developed as analytic tools to assess riskiness of trading activities.<sup>1</sup> In terms of the formal definition, VaR is simply a quantile of the loss distribution (future portfolio values) over a prescribed holding period (e.g., 2 weeks) at a given confidence level, while ES is the expected loss, given that the loss is at least as large as some given VaR. It is well known from Artzner et al. (1999), that ES is a coherent risk measure satisfying homogeneity, monotonicity, risk-free condition or translation invariance, and subadditivity,

\* Corresponding author at: Department of Mathematics and Statistics and Department of Economics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. Tel.: +1 7046872650; fax: +1 7046876415.

E-mail address: [zca@uncc.edu](mailto:zca@uncc.edu) (Z. Cai).

<sup>1</sup> See, to name just a few, Morgan (1996), Duffie and Pan (1997), Jorion (2001) and Duffie and Singleton (2003) for financial background, statistical inferences, and various applications.

while VaR is not coherent, because it does not satisfy subadditivity. As advocated by Artzner et al. (1999), ES is preferred in practice due to its better properties, although VaR is widely used in applications.

Measures of risk might depend on the state of the economy, since economic and market conditions vary from time to time. This requires risk managers to focus on the conditional distributions of profit and loss, which take a full survey of current information on the investment environment, such as macroeconomic, financial, and political environments, in forecasting future market values, volatilities, and correlations. Moreover, it is well documented that VaR is expected to increase as the past returns become very negative, because one bad day increases the probability of the next day. Similarly, very good days also increase VaR; see Duffie and Singleton (2003) and Engle and Manganelli (2004). Therefore, VaR could depend on the past returns in some way. On the other hand, as pointed out by Duffie and Singleton (2003) and Engle and Manganelli (2004), not only are the prices of the underlying market indices changing randomly over time, but the portfolio itself is changing, as is the volatility of prices, the credit qualities of counterparties, and so on. Hence, an appropriate risk analytical tool or methodology should be allowed to adapt to varying market conditions, and to reflect the latest available information in a time series setting rather than the iid framework. Most of the existing risk management literature has concentrated on unconditional

distributions and the iid setting, although there have been some studies on conditional distributions and time series data.<sup>2</sup>

The main focus of this paper is on the nonparametric estimation of conditional value-at-risk (CVaR), and conditional expected shortfall (CES) functions where the conditional information contains economic and market (exogenous) variables and past observed returns.<sup>3</sup> Most of studies in the literature and applications are limited to parametric models, such as all standard industry models as CreditRisk<sup>+</sup>, CreditMetrics, CreditPortfolio View and the model proposed by the KMV corporation and others.<sup>4</sup> Parametric models for CVaR and CES are the most efficient if the underlying functions are correctly specified. However, a misspecification may cause serious bias, and model constraints may distort the underlying distributions. It is well known that nonparametric modeling is appealing in several aspects. One of those is that little or no restrictive prior information on functionals is needed. Another advantage is that it allows a wide range of data dependence, which makes it adaptable in the context of financial losses. Further, nonparametric modeling may provide useful insight for further parametric fitting.

This paper has the following contributions to the literature. The first one is to propose a new nonparametric approach to estimate CVaR and CES. In essence, our estimator for CVaR is based on inverting a newly proposed estimator of conditional distribution function. The estimator for CES is found by plugging in the estimated conditional probability density function (PDF) and the estimated CVaR function. More precisely, our newly proposed estimator combines the weighted Nadaraya–Watson (WNW) method of Cai (2002) and the double kernel local linear technique of Yu and Jones (1998), termed as “weighted double kernel local linear” (WDKLL) estimator. It is shown that, analogous to Scaillet (2005), Cai (2002) and Yu and Jones (1998), the proposed estimators overcome the so-called boundary effects. The second contribution is to establish the asymptotic properties for the WDKLL estimators of conditional PDF and cumulative distribution function (CDF) for  $\alpha$ -mixing time series at both boundary and interior points. It is therefore shown that the WDKLL method enjoys the same convergence rates as those of the double kernel local linear estimator and the WNW estimator. It is also demonstrated that the WDKLL CDF estimators have desired sampling properties at both boundary and interior points. Finally, we show that the WDKLL estimator of CVaR has the following three properties: it exists always due to the WDKLL estimator of CDF being a distribution function, it inherits the differentiability from the WDKLL estimator of CDF, and it possesses the same asymptotic properties as the standard local linear estimator does; see Fan and Gijbels (1996).

The rest of the paper is organized as follows. In Section 2, we present the detailed motivations and formulations for the new nonparametric estimation procedures for estimating conditional PDF, CDF, VaR and ES. We establish the asymptotic properties of these nonparametric estimators at both boundary and interior

points with some comparisons in Section 3. Together with an easily implemented data-driven method for selecting the bandwidth based on the nonparametric Akaike information criterion (AIC), a Monte Carlo simulation study and an empirical application to a stock index return are presented in Section 4. Finally, some lemmas and the derivations of theorems are given in Section 5, and the Appendix contains the technical proofs of certain lemmas.

## 2. Nonparametric Estimating Procedures

Assume that the observed data  $\{(X_t, Y_t); 1 \leq t \leq n\}$ ,  $X_t \in \mathcal{R}^d$ , are available and they are observed from a stationary time series model. Here  $Y_t$  is the risk or loss variable which can be the negative logarithm of return (log loss) and  $X_t$  is allowed to include both economic and market (exogenous) variables and the lagged variables of  $Y_t$ . Note that our presentation is only for the case that  $d = 1$ , and the proposed methodology and theory continue to hold for  $d > 1$  with more complex notations.

Clearly, CVaR  $v_p(x)$  can be expressed as  $v_p(x) = S^{-1}(p | x)$ , where  $S(y | x) = 1 - F(y | x)$  and  $F(y | x)$  is the conditional CDF of  $Y_t$  given  $X_t = x$ . The nonparametric estimation of  $v_p(x)$  can be constructed as  $\widehat{v}_p(x) = \widehat{S}^{-1}(p | x)$ , where  $\widehat{S}^{-1}(p | x)$  is a nonparametric estimation of  $S^{-1}(p | x)$ . CES  $\mu_p(x)$  is formulated as

$$\mu_p(x) = E[Y_t | Y_t \geq v_p(x), X_t = x] = \int_{v_p(x)}^{\infty} yf(y | x)dy/p,$$

where  $f(y | x)$  is the conditional PDF of  $Y_t$  given  $X_t = x$ . To estimate  $\mu_p(x)$ , one can use the plugging-in method as

$$\widehat{\mu}_p(x) = \int_{\widehat{v}_p(x)}^{\infty} y\widehat{f}(y | x)dy/p, \tag{1}$$

where  $\widehat{v}_p(x)$  is a nonparametric estimation of  $v_p(x)$ , and  $\widehat{f}(y | x)$  is a nonparametric estimation of  $f(y | x)$ . But the bandwidths used to estimate  $\widehat{v}_p(x)$  and  $\widehat{f}(y | x)$  might not be necessarily the same.

First, we start with our nonparametric estimators for conditional PDF and CDF. It is noted for a given symmetric kernel  $K(\cdot)$ ,

$$E\{K_{h_0}(y - Y_t) | X_t = x\} = f(y | x) + \frac{h_0^2}{2}\mu_2(K)f^{2,0}(y | x) + o(h_0^2) \approx f(y | x), \text{ as } h_0 \rightarrow 0, \tag{2}$$

where  $K_{h_0}(u) = K(u/h_0)/h_0$ ,  $\mu_2(K) = \int_{-\infty}^{\infty} u^2K(u)du$ ,  $f^{2,0}(y | x) = \partial^2/\partial y^2 f(y | x)$ , and  $\approx$  denotes an approximation by ignoring the higher terms  $O(h_0^j)$  for  $j \geq 2$  if  $h_0 = o(h)$ , where  $h$  is the smoothing bandwidth in the  $x$  direction (see (3)). We can see that  $Y_t^*(y) = K_{h_0}(y - Y_t)$  can be regarded as an initial estimate of  $f(y | x)$  smoothing in the  $y$  direction. Thus, the left hand side of (2) can be regarded as a nonparametric regression of the observed variable  $Y_t^*(y)$ , versus  $X_t$  and the local linear (or polynomial) fitting scheme can be applied here. This leads to the locally weighted least squares regression problem:

$$\sum_{t=1}^n \{Y_t^*(y) - a - b(X_t - x)\}^2 W_h(x - X_t), \tag{3}$$

where  $W(\cdot)$  is a kernel function and bandwidth  $h = h(n) > 0$  satisfies  $h \rightarrow 0$  and  $n h \rightarrow \infty$  as  $n \rightarrow \infty$ . Note that (3) involves two kernels  $K(\cdot)$  and  $W(\cdot)$ . This is the reason for calling it “double kernel”.

Minimizing (3) with respect to  $a$  and  $b$ , we obtain the locally weighted least squares estimator of  $f(y | x)$ , which is  $\widehat{a}$ . From Fan and Gijbels (1996) or Fan et al. (1996), this estimator can be re-expressed as a linear estimator as

$$\widehat{f}_l(y | x) = \sum_{t=1}^n W_{l,t}(x, h)Y_t^*(y),$$

<sup>2</sup> See Chernozhukov and Umanstev (2001), Cai (2002), Fan and Gu (2003), Engle and Manganelli (2004), Cai and Xu (2008), Scaillet (2005) and Cosma et al. (2007) for conditional models, and Duffie and Pan (1997), Artzner et al. (1999), Rockafellar and Uryasev (2000), Acerbi and Tasche (2002), Frey and McNeil (2002), Scaillet (2004), Chen and Tang (2005) and Chen (2006) for unconditional models.

<sup>3</sup> Note that CVaR defined here is essentially the conditional quantile or quantile regression of Koenker and Bassett (1978), rather than CVaR defined in some risk management literature; see, e.g., Rockafellar and Uryasev (2000) and Jorion (2001). Moreover, ES here is called TailVaR in Artzner et al. (1999). Finally, both ES and CES have been known and popular for decades among actuary sciences and insurance industries. See the book by Embrechts et al. (1997) for the excellent review on this subject and the papers by McNeil (1997), Scaillet (2005) and Chen (2006).

<sup>4</sup> See Chernozhukov and Umanstev (2001), Frey and McNeil (2002), Engle and Manganelli (2004), and references therein on parametric models in practice and Fan and Gu (2003) and references therein for semiparametric models.

where with  $S_{n,j}(x) = \sum_{t=1}^n W_h(x - X_t)(X_t - x)^j$ , the weights  $\{W_{ll,t}(x, h)\}$  are given by

$$W_{ll,t}(x, h) = [S_{n,2}(x) - (x - X_t)S_{n,1}(x)]W_h(x - X_t) \times [S_{n,0}(x)S_{n,2}(x) - S_{n,1}^2(x)]^{-1}.$$

Clearly,  $\{W_{ll,t}(x, h)\}$  satisfy the so-called discrete moments conditions as follows:

$$\sum_{t=1}^n W_{ll,t}(x, h)(X_t - x)^j = \delta_{0,j} = \begin{cases} 1 & \text{if } j = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

for  $0 \leq j \leq 1$ ; see Equation 3.12 in Fan and Gijbels (1996, p. 63). Note that the estimator  $\hat{f}_{ll}(y | x)$  can range outside  $[0, \infty)$ . The double kernel local linear estimator of  $F(y | x)$  is constructed by integrating  $\hat{f}_{ll}(y | x)$

$$\hat{F}_{ll}(y | x) = \int_{-\infty}^y \hat{f}_{ll}(y | x) dy = \sum_{t=1}^n W_{ll,t}(x, h)G_{h_0}(y - Y_t),$$

where  $G(\cdot)$  is the distribution function of  $K(\cdot)$  and  $G_{h_0}(u) = G(u/h_0)$ . Clearly,  $\hat{F}_{ll}(y | x)$  is differentiable with respect to  $y$  and satisfies  $\hat{F}_{ll}(-\infty | x) = 0$  and  $\hat{F}_{ll}(\infty | x) = 1$ .

Although Yu and Jones (1998) showed that the double kernel local linear estimator  $\hat{F}_{ll}(y | x)$  has some attractive properties, it is not constrained to neither fall between zero and one, nor be monotonically increasing, which is not good for estimating CVaR if the inverting method is used. In both respects, the NW method is superior, despite its rather large bias and boundary effects. Here we need to mention that the boundary effect might cause a problem for estimating  $v_p(x)$  since it only concerns the tail probability. So, Hall et al. (1999) and Cai (2002) proposed the WNW estimator, which is designed to possess the superior properties of local linear method and to preserve the property that the NW estimator is always a distribution function. The WNW estimator of the conditional distribution  $F(y | x)$  of  $Y_t$  given  $X_t = x$  is defined by

$$\hat{F}_{c1}(y | x) = \sum_{t=1}^n W_{c,t}(x, h)I(Y_t \leq y), \quad (5)$$

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

$$W_{c,t}(x, h) = p_t(x)W_h(x - X_t) \left[ \sum_{t=1}^n p_t(x)W_h(x - X_t) \right]^{-1}, \quad (6)$$

and  $\{p_t(x)\}$  is chosen to be  $p_t(x) = n^{-1}\{1 + \lambda(X_t - x)W_h(x - X_t)\}^{-1} \geq 0$ . Here,  $\lambda$  is a function of the data and  $x$  and is uniquely defined by maximizing the logarithm of the empirical likelihood  $L_n(\lambda) = -\sum_{t=1}^n \log\{1 + \lambda(X_t - x)W_h(x - X_t)\}$  subject to the constraints  $p_t(x) \geq 0$ ,  $\sum_{t=1}^n p_t(x) = 1$ , and the discrete moments conditions

$$\sum_{t=1}^n W_{c,t}(x, h)(X_t - x)^j = \delta_{0,j} \quad (7)$$

for  $0 \leq j \leq 1$ . In implementation, Cai (2002) recommended using the Newton–Raphson scheme to find the root of equation  $L'_n(\lambda) = 0$ . It can be shown easily that  $\hat{F}_{c1}(y | x)$  must lie between 0 and 1 and be monotonic in  $y$  but not continuous (and of course, indifferentiable) in  $y$ .

To accommodate all of the above attractive properties of both estimators  $\hat{F}_{ll}(y | x)$  and  $\hat{F}_{c1}(y | x)$  under a unified framework, we propose the following nonparametric estimators for conditional PDF  $f(y | x)$  and its conditional CDF  $F(y | x)$ , termed as “weighted double kernel local linear estimation”,

$$\hat{f}_c(y | x) = \sum_{t=1}^n W_{c,t}(x, h)Y_t^*(y),$$

where  $W_{c,t}(x, h)$  is given in (6), and

$$\hat{F}_c(y | x) = \int_{-\infty}^y \hat{f}_c(y | x) dy = \sum_{t=1}^n W_{c,t}(x, h)G_{h_0}(y - Y_t). \quad (8)$$

Note that if  $p_t(x)$  in (6) is a constant for all  $t$ , or  $\lambda = 0$ , then  $\hat{f}_c(y | x)$  becomes the classical NW type double kernel estimator used by Scaillet (2005). However, Scaillet (2005) adopted a single bandwidth for smoothing in both the  $y$  and  $x$  directions. Clearly,  $\hat{f}_c(y | x)$  is a PDF and  $\hat{F}_c(y | x)$  is a CDF and differentiable in  $y$ .

It is worth pointing out that the differentiability of the estimated CDF can make the asymptotic analysis much easier for the nonparametric estimators of CVaR and CES (see Section 3, and the proof of Theorem 4) and reduces the asymptotic variance (or asymptotic mean squared error (AMSE)) in a higher order sense (see Remark 3 and (15)); see Cai and Roussas (1998) and Chen and Tang (2005).

Finally, we now are ready to formulate our nonparametric estimators for  $v_p(x)$  and  $\mu_p(x)$ . In view of (8),  $v_p(x)$  is estimated by  $\hat{v}_p(x) = \hat{S}_c^{-1}(p | x)$ , where  $\hat{S}_c(y | x) = 1 - \hat{F}_c(y | x)$ . Note that  $\hat{v}_p(x)$  always exists, and is uniquely determined since  $\hat{S}_c(p | x)$  is a survival function itself. Plugging  $\hat{v}_p(x)$  and  $\hat{f}_c(y | x)$  into (1), we obtain the nonparametric estimation of  $\mu_p(x)$ ,

$$\hat{\mu}_p(x) = p^{-1} \sum_{t=1}^n W_{c,t}(x, h) [Y_t \bar{G}_{h_0}(\hat{v}_p(x) - Y_t) + h_0 G_{1,h_0}(\hat{v}_p(x) - Y_t)], \quad (9)$$

where  $\bar{G}(u) = 1 - G(u)$ ,  $G_{1,h_0}(u) = G_1(u/h_0)$ , and  $G_1(u) = \int_u^\infty vK(v)dv$ .

### 3. Statistical properties

A time series model is often assumed to follow a certain linear time series model, such as an autoregressive and moving average process. Here we consider a more general structure—the  $\alpha$ -mixing process, which includes many linear and nonlinear time series models as special cases.<sup>5</sup> The asymptotic results are derived under the  $\alpha$ -mixing assumption.

Next, we introduce some notations and list the regularity conditions for the asymptotic properties. Define  $\alpha(K) = \int_{-\infty}^\infty uK(u)\bar{G}(u)du$  and  $\mu_j(W) = \int_{-\infty}^\infty u^j W(u)du$ . Also, for any  $j \geq 0$ ,  $l_j(u | v) = E[Y_t^j I(Y_t \geq u) | X_t = v] = \int_u^\infty y^j f(y | v) dy$ , and  $l_j^{a,b}(u | v) = \frac{\partial^{ab}}{\partial u^a \partial v^b} l_j(u | v)$ . Clearly,  $l_0(u | v) = S(u | v)$  and  $l_1(v_p(x) | x) = p\mu_p(x)$ . Finally,  $l_j^{1,0}(u | v) = -u^j f(u | v)$  and  $l_j^{2,0}(u | v) = -[u^j f^{1,0}(u | v) + j u^{j-1} f(u | v)]$ .

**Assumption A.** A1. For any fixed  $x$  and  $y$ ,  $0 < F(y | x) < 1$ . The marginal PDF of  $X_t$ ,  $g(x) > 0$ , is continuous at  $x$ .  $F(y | x)$  has continuous second order derivative with respect to both  $x$  and  $y$ .

A2. Both kernels  $K(\cdot)$  and  $W(\cdot)$  are symmetric, bounded, and compactly supported density.

A3.  $h \rightarrow 0$  and  $nh \rightarrow \infty$ , and  $h_0 \rightarrow 0$  and  $nh_0 \rightarrow \infty$ , as  $n \rightarrow \infty$ .

A4. Let  $g_{1,t}(\cdot, \cdot)$  be the joint density of  $X_1$  and  $X_t$  for  $t \geq 2$ . Assume that  $|g_{1,t}(x_1, x_t) - g(x_1)g(x_t)| \leq M < \infty$  for all  $x_1$  and  $x_t$ .

A5. The process  $\{X_t, Y_t\}$  is a stationary  $\alpha$ -mixing with the mixing coefficient satisfying  $\alpha(t) = O(t^{-(2+\delta_1)})$  for some  $\delta_1 > 0$ .

A6.  $nh^{1+2/\delta_1} \rightarrow \infty$ , where  $\delta_1$  is given in Assumption A5.

<sup>5</sup> See, e.g., Pötscher and Prucha (1997), Cai (2002, 2007), Carrasco and Chen (2002), Meitz and Saikkonen (2004) and Chen and Tang (2005).

A7.  $h_0 = o(h)$ .

**Assumption B.** B1. Assume that  $E(|Y_t|^{\delta_2} | X_t = x) \leq M_3 < \infty$  for some  $\delta_2 > 2$  in a neighborhood of  $x$ .

B2. Assume that  $|g_{1,t}(y_1, y_t | x_1, x_t)| \leq M_1 < \infty$  for all  $t \geq 2$ , where  $g_{1,t}(y_1, y_t | x_1, x_t)$  be the conditional PDF of  $Y_1$  and  $Y_t$  given  $X_1 = x_1$  and  $X_t = x_t$ .

B3. Assume that the mixing coefficient of the  $\alpha$ -mixing process  $\{(X_t, Y_t)\}_{t=-\infty}^{\infty}$  satisfies  $\sum_{t \geq 1} t^{a_1} \alpha^{1-2/\delta_2}(t) < \infty$  for some  $a_1 > 1 - 2/\delta_2$ , where  $\delta_2$  is given in Assumption B1.

B4. Assume that there exists a sequence of integers  $s_n > 0$  such that  $s_n \rightarrow \infty$ ,  $s_n = o((nh)^{1/2})$ , and  $(n/h)^{1/2} \alpha(s_n) \rightarrow 0$ , as  $n \rightarrow \infty$ .

B5. There exists  $\delta_3 > \delta_2$  such that  $E(|Y_t|^{\delta_3} | X_t = x) \leq M_4 < \infty$  in a neighborhood of  $x$ ,  $\alpha(t) = O(t^{-\theta_1})$ , where  $\theta_1 \geq \delta_3 \delta_2 / \{2(\delta_3 - \delta_2)\}$ ,  $n^{1/2-\delta_2/4} h^{\delta_2/\delta_3-1/2-\delta_2/4} = O(1)$ , and  $\delta_2$  is given in Assumption B1.

**Remark 1.** Note that Assumptions A1–A5 and B1–B5 are used commonly in the literature of time series data (see, e.g. Fan and Gijbels (1996) and Cai (2001)).  $\alpha$ -mixing imposed in Assumption A5 is weaker than  $\beta$ -mixing in Hall et al. (1999), and  $\rho$ -mixing in Fan et al. (1996). Because A6 is satisfied by the bandwidths of optimal size (i.e.,  $h = O(n^{-1/5})$ ) if  $\delta_1 > 1/2$ , we do not concern ourselves with such refinements. Indeed, Assumptions A1–A6 are also required in Cai (2002). Assumption A7 means that when the initial step bandwidth is chosen as small as possible, the bias from the initial step can be ignored. Since the common technique – truncation approach – for time series data is not applicable to our setting (see, e.g. Fan and Gijbels (1996)), the purpose of Assumption B5 is to use the moment inequality. If  $\alpha(\cdot)$  decays geometrically, then Assumptions B4 and B5 are satisfied automatically. Note that Assumptions B3, B4 and B5 are stronger than Assumptions A5 and A6. This is not surprising, because as higher moments are involved, a faster decaying rate of  $\alpha(\cdot)$  is required. Finally, Assumptions B1–B5 are also imposed in Cai (2001).

First, we embrace the investigation of the asymptotic behaviors of  $\widehat{f}_c(y | x)$ , including the asymptotic normality stated in the following theorem.

**Theorem 1.** Under Assumptions A1–A6 with  $h$  in A3 and A6 replaced by  $h_0 h$ , we have

$$\sqrt{nh_0 h} [\widehat{f}_c(y | x) - f(y | x) - B_f(y | x)] \rightarrow N(0, \sigma_f^2(y | x)),$$

where the asymptotic bias is  $B_f(y | x) = h^2 \mu_2(W) f^{0,2}(y | x)/2 + h_0^2 \mu_2(K) f^{2,0}(y | x)/2$ , and the asymptotic variance is  $\sigma_f^2(y | x) = \mu_0(K^2) \mu_0(W^2) f(y | x)/g(x)$ .

**Remark 2.** The asymptotic results for  $\widehat{f}_c(y | x)$  in Theorem 1 are similar to those for  $\widehat{f}_{ll}(y | x)$  in Fan et al. (1996) for  $\rho$ -mixing sequence, which is stronger than  $\alpha$ -mixing, but as mentioned earlier,  $\widehat{f}_{ll}(y | x)$  is not always a PDF. The asymptotic bias and variance are intuitively expected. The bias comes from the approximations in both the  $x$  and  $y$  directions, and the variance is from the local conditional variance of the density  $f(y | x)$ .

Next, we study the asymptotic behaviors for  $\widehat{S}_c(y | x)$  at both interior and boundary points.

**Theorem 2.** Under Assumptions A1–A6, we have

$$\sqrt{nh} [\widehat{S}_c(y | x) - S(y | x) - B_S(y | x)] \rightarrow N(0, \sigma_S^2(y | x)),$$

where the asymptotic bias is  $B_S(y | x) = h^2 \mu_2(W) S^{0,2}(y | x)/2 - h_0^2 \mu_2(K) f^{1,0}(y | x)/2$ , and the asymptotic variance is  $\sigma_S^2(y | x)$

$x) = \mu_0(W^2) S(y | x)[1 - S(y | x)]/g(x)$ . In particular, if Assumption A7 holds true, then,

$$\begin{aligned} \sqrt{nh} \left[ \widehat{S}_c(y | x) - S(y | x) - \frac{h^2}{2} \mu_2(W) S^{0,2}(y | x) \right] \\ \rightarrow N(0, \sigma_S^2(y | x)). \end{aligned}$$

**Remark 3.** Note that the asymptotic results for  $\widehat{S}_c(y | x)$  in Theorem 2 are analogous to those for  $\widehat{S}_{ll}(y | x) = 1 - \widehat{F}_{ll}(y | x)$  in Yu and Jones (1998) for iid data, but  $\widehat{F}_{ll}(y | x)$  is not always a distribution function. A comparison of  $B_S(y | x)$  with the asymptotic bias for WNW estimator  $\widehat{S}_{c1}(y | x)$  reveals an extra term  $h_0^2 f^{1,0}(y | x) \mu_2(K)/2$  in  $B_S(y | x)$  due to the vertical smoothing in the  $y$  direction. Also, there is an extra term in the asymptotic variance (see (15)). These extra terms carried over from the initial estimate can be ignored if Assumption A7 is satisfied. Clearly, the second term on the right hand in (15) can be negative (say,  $\alpha(K) = -9/70$  for  $K(u) = 0.75(1 - u^2)_+$ ) so that the asymptotic variance (or AMSE) can be reduced in a higher order sense due to smoothing in the  $y$  direction; see Cai and Roussas (1998) and Chen and Tang (2005).

**Remark 4.** From Theorem 2, the asymptotic mean squared error of  $\widehat{S}_c(y | x)$  is  $AMSE = O(h^4 + 1/nh)$  if  $h_0 = o(h)$ . By minimizing AMSE derived from Theorem 2, and taking  $h_0 = o(h)$ , we obtain the optimal bandwidth given by  $h_{opt,S}(y | x) = O(n^{-1/5})$ . Therefore, the optimal rate of the AMSE of  $\widehat{S}_c(y | x)$  is  $n^{-4/5}$ .

Now we establish the asymptotic properties for  $\widehat{S}_c(y | x)$  at boundary points. Following Fan and Gijbels (1996), we take  $W(\cdot)$  to have support  $[-1, 1]$  and  $g(\cdot)$  to have support  $[0, 1]$ . Without loss of generality, we consider the left boundary point  $x = ch$ ,  $0 < c < 1$ . Then, under Assumptions A1–A7, we can show that

$$\sqrt{nh} [\widehat{S}_c(y | ch) - S_c(y | ch) - B_{S,c}(y)] \rightarrow N(0, \sigma_{S,c}^2(y)), \quad (10)$$

where the asymptotic bias is  $B_{S,c}(y) = h^2 \beta_0(c) S^{0,2}(y | 0+) / [2\beta_1(c)]$ , the asymptotic variance is  $\sigma_{S,c}^2(y) = \beta_2(0) S(y | 0+) [1 - S(y | 0+)] / [\beta_1^2(c) g(0+)]$ ,  $g(0+) = \lim_{z \downarrow 0} g(z)$ ,  $\beta_0(c) = \int_{-1}^c \frac{u^2 W(u)}{1 - \lambda_c u W(u)} du$ ,  $\beta_j(c) = \int_{-1}^c \frac{W^j(u)}{(1 - \lambda_c u W(u))^j} du$  for  $1 \leq j \leq 2$ , and  $\lambda_c$  is the root of equation  $L_c(\lambda) = 0$  with  $L_c(\lambda) = \int_{-1}^c \frac{u W(u)}{1 - \lambda u W(u)} du$ . The proof of (10) is similar to that for Theorem 2 in Cai (2002) and thus omitted. Theorem 2 and (10) reflect that the advantages of the WKDLL estimator are the same as those for standard local linear estimator.

By the differentiability of  $\widehat{S}_c(\widehat{v}_p(x) | x)$ , we use the Taylor expansion, and ignore the higher terms to obtain

$$\widehat{S}_c(\widehat{v}_p(x) | x) = p \approx \widehat{S}_c(v_p(x) | x) - \widehat{f}_c(v_p(x) | x) (\widehat{v}_p(x) - v_p(x)). \quad (11)$$

Then, by Theorem 1,

$$\begin{aligned} \widehat{v}_p(x) - v_p(x) &\approx [\widehat{S}_c(v_p(x) | x) - p] / \widehat{f}_c(v_p(x) | x) \\ &\approx [\widehat{S}_c(v_p(x) | x) - p] / f(v_p(x) | x). \end{aligned}$$

As an application of Theorem 2, we can establish the following theorem for the asymptotic normality of  $\widehat{v}_p(x)$ , but the proof is omitted, since it is similar to that for Theorem 2.

**Theorem 3.** Under Assumptions A1–A6, we have

$$\sqrt{nh} [\widehat{v}_p(x) - v_p(x) - B_v(x)] \rightarrow N(0, \sigma_v^2(x)),$$

where the asymptotic bias is  $B_v(x) = B_S(v_p(x) | x)/f(v_p(x) | x)$  and the asymptotic variance is  $\sigma_v^2(x) = \mu_0(W^2)p(1 - p)/[g(x)f^2(v_p(x) | x)]$ . In particular, if Assumption A7 holds, then,

$$\sqrt{nh} \left[ \widehat{v}_p(x) - v_p(x) - \frac{h^2}{2} \frac{S^{0,2}(v_p(x) | x)}{f(v_p(x) | x)} \mu_2(W) \right] \rightarrow N(0, \sigma_v^2(x)).$$

**Remark 5.** First, we have  $\widehat{v}_p(x) - v_p(x) = O_p(h^2 + h_0^2 + (nh)^{-1/2})$  as a consequence of Theorem 3, so that  $\widehat{v}_p(x)$  is a consistent estimator of  $v_p(x)$  with a convergence rate  $\sqrt{nh}$ . Also, note that the asymptotic results for  $\widehat{v}_p(x)$  in Theorem 3 are akin to those for  $\widehat{v}_{ll,p}(x) = \widehat{S}_{ll}^{-1}(p | x)$  in Yu and Jones (1998) for iid data. But in the bias term of Theorem 3, the quantity  $S^{0,2}(v_p(x) | x)/f(v_p(x) | x)$  replaces  $v_p''(x)$ , which is in the bias term of the “check” function type local linear estimator in Yu and Jones (1998) for iid data and Cai and Xu (2008) for time series. This is not surprising, since the bias comes only from the approximation. The former utilizes the approximation of conditional distribution function, but the latter uses the approximation of conditional VaR function.

**Remark 6.** Theorem 3 implies that the AMSE of  $\widehat{v}_p(x)$  is given by  $AMSE = O(h^4 + 1/nh)$  if  $h_0 = o(h)$  and its optimal rate is  $n^{-4/5}$  for the optimal bandwidth  $h_{opt,v}(x) = O(n^{-1/5})$ . Theorems 2 and 3 conclude that  $\widehat{S}_c(y | x)$  and  $\widehat{v}_p(x)$  are insensitive to the choice of  $h_0$ , if  $h_0 = o(h)$ . This makes the selection of bandwidths much easier in practice. Also, by comparing  $h_{opt,v}(x)$  with  $h_{opt,S}(y | x)$ , it turns out that  $h_{opt,v}(x)$  is  $h_{opt,S}(y | x)$  evaluated at  $y = v_p(x)$ . Therefore, the best choice of the bandwidth for estimating  $S_c(y | x)$  can be used for estimating  $v_p(x)$ .

**Remark 7.** Similar to (10), one can establish the asymptotic result at boundaries for  $v_p(x)$  as follows. Under Assumption A7,

$$\sqrt{nh} [\widehat{v}_p(ch) - v_p(ch) - B_{v,c}] \rightarrow N(0, \sigma_{v,c}^2),$$

where the asymptotic bias is  $B_{v,c} = h^2 \beta_2(c) S^{0,2}(v_p(0+) | 0+) / [2\beta_1(c)f(v_p(0+) | 0+)]$  and the asymptotic variance is  $\sigma_{v,c}^2 = \beta_0(0)p[1 - p] / [\beta_1^2(c)f^2(v_p(0+) | 0+)g(0+)]$ . Clearly,  $\widehat{v}_p(x)$  inherits all good properties from the estimator  $\widehat{S}_c(y | x)$ .

Now, we embark on examining the asymptotic behavior for  $\widehat{\mu}_p(x)$  at both interior and boundary points. First, we establish the asymptotic normality for  $\widehat{\mu}_p(x)$  when  $x$  is an interior point.

**Theorem 4.** Under Assumption A1–A4 and B2–B5, we have

$$\sqrt{nh} [\widehat{\mu}_p(x) - \mu_p(x) - B_\mu(x)] \rightarrow N(0, \sigma_\mu^2(x)),$$

where  $B_\mu(x) = B_{\mu,0}(x) + h_0^2 \mu_2(K) p^{-1} f(v_p(x) | x) / 2$  with  $B_{\mu,0}(x) = \mu_2(W) B_0(x)$  and  $B_0(x) = h^2 [I_1^{0,2}(v_p(x) | x) - v_p(x) S^{0,2}(v_p(x) | x)] / (2p)$ , and  $\sigma_\mu^2(x) = \mu_0(W^2) \sigma_*^2(x) / g(x)$  with  $\sigma_*^2(x) = p^{-2} I_2(v_p(x) | x) - \mu_p^2(x) + (1/p - 1) v_p(x) \{v_p(x) - 2\mu_p(x)\}$ . In particular, if Assumption A7 holds true, then,

$$\sqrt{nh} [\widehat{\mu}_p(x) - \mu_p(x) - B_{\mu,0}(x)] \rightarrow N(0, \sigma_\mu^2(x)).$$

**Remark 8.** Consequently, Theorem 4 concludes that  $\widehat{\mu}_p(x) - \mu_p(x) = O_p(h^2 + h_0^2 + (nh)^{-1/2})$ , so that  $\widehat{\mu}_p(x)$  is a consistent estimator of  $\mu_p(x)$  with a convergence rate  $\sqrt{nh}$ . Further, note that Scaillet (2005) did not provide an expression for the asymptotic bias term such as  $B_\mu(x)$  in the first result, or  $B_{\mu,0}(x)$  in the second conclusion in Theorem 4. Clearly, the second term in the asymptotic bias expression is carried over from the  $y$  direction

smoothing at the initial step and is negligible if Assumption A7 is satisfied. Clearly, Assumption A7 implies that  $B_\mu(x)$  becomes  $B_{\mu,0}(x)$ .

**Remark 9.** Similar to Remark 6, we can derive the AMSE of  $\widehat{\mu}_p(x)$  and the optimal bandwidth from Theorem 4. As expected, the optimal rate of the AMSE of  $\widehat{\mu}_p(x)$  is  $n^{-4/5}$ .

Finally, we offer the asymptotic results for  $\widehat{\mu}_p(x)$  at the left boundary point  $x = ch$ . Under Assumption A7,

$$\sqrt{nh} [\widehat{\mu}_p(ch) - \mu_p(ch) - B_{\mu,c}] \rightarrow N(0, \sigma_{\mu,c}^2),$$

where  $B_{\mu,c} = \beta_2(c) B_0(0+) / \beta_1(c)$  and  $\sigma_{\mu,c}^2 = \beta_0(0) \sigma_*^2(0+) / [g(0+) \beta_1^2(c)]$ . The proof of the above result can be carried out by using the second assertion in Lemma 1, and following the proof of Theorem 4, and it is thus omitted. Next, we consider the comparison of the performance of the WDKLL estimation  $\widehat{\mu}_p(x)$  with the NW type kernel estimator as in Scaillet (2005). To this effect, it is not very difficult to derive the asymptotic results for the NW type kernel estimator, but the proof is omitted. See Scaillet (2005) for the results at the interior point. Under some regularity conditions, it can be shown tediously that the asymptotic bias term for the NW type kernel estimator is of the order  $h$ , by comparing to the order  $h^2$  for the WDKLL estimate (see  $B_{\mu,c}$ ) at the left boundary  $x = ch$ . This shows that the WDKLL estimate does not suffer from boundary effects, but the NW type kernel estimator estimate does. This is another advantage of using the WDKLL estimator rather than the WW type kernel estimator.

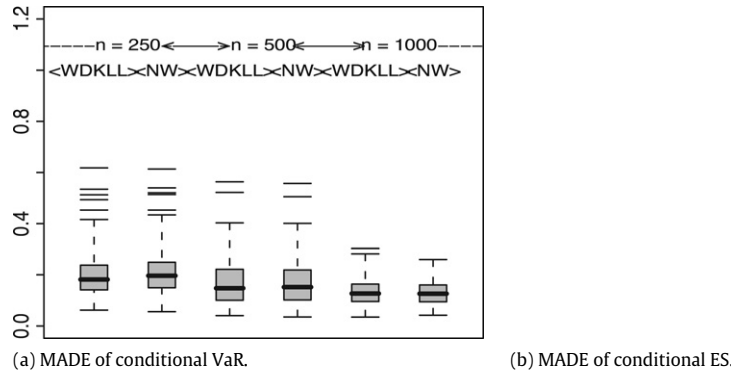
#### 4. Empirical applications

To illustrate the proposed methods, we consider a simulated example and a real data example on a stock index return. Throughout this section, the Epanechnikov kernel  $K(u) = 0.75(1 - u^2)_+$  is used, and the bandwidth selector described next is used.

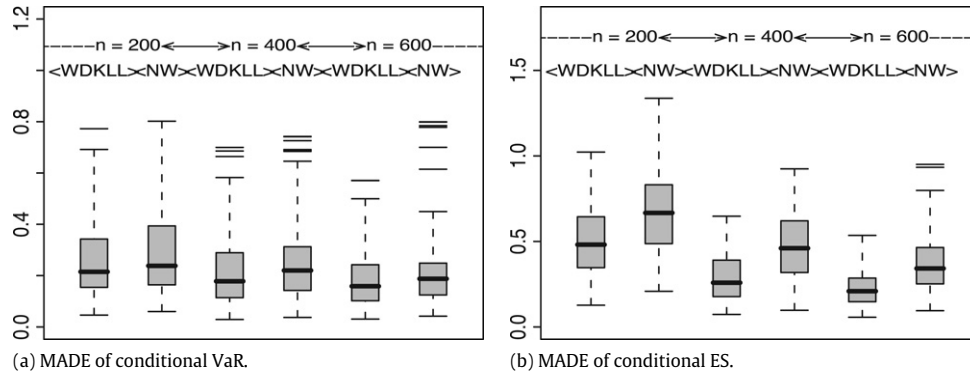
##### 4.1. Bandwidth selection

With the basic model at hand, one must address the important bandwidth selection issue, as the quality of the curve estimates depends sensitively on the choice of the bandwidth. For practitioners, it is desirable to have an easily implemented data-driven rule. However, almost nothing has been done so far about this problem in the context of estimating  $v_p(x)$  and  $\mu_p(x)$ , although there are some results available in the literature in other contexts for some specific purposes.

As indicated earlier, the choice of the initial bandwidth  $h_0$  is insensitive to the final estimation, but it needs to be specified. First, we use a very simple idea to choose  $h_0$ . As mentioned previously, the WNW method involves only one bandwidth in estimating the conditional distribution and VaR. Because the WNW estimate is a linear smoother (see (5)), we recommend using the optimal bandwidth selector, the so-called nonparametric AIC proposed by Cai and Tiwari (2000), to select the bandwidth, called  $\widehat{h}$ . Then we take  $0.1 \times \widehat{h}$  or smaller as the initial bandwidth  $h_0$ . For given  $h_0$ , we can select  $h$  as follows. According to (8),  $\widehat{F}_c(\cdot | \cdot)$  is a linear estimator, so that the nonparametric AIC selector can be applied here to the selection of the optimal bandwidth for  $\widehat{F}_c(\cdot | \cdot)$ , denoted by  $h_s$ . As mentioned at the end of Remark 6, the bandwidth for  $\widehat{v}_p(x)$  is the same as that for  $\widehat{F}_c(\cdot | \cdot)$ , so that it is simple to take  $h_s$  as  $h_v$ . From (9),  $\widehat{\mu}_p(x)$  also is a linear estimator for given  $\widehat{v}_p(x)$ . Therefore, by the same token, the nonparametric AIC selector is applied to selecting  $h_\mu$  for  $\widehat{\mu}_p(x)$ . This simple approach is used in our implementation in the next sections.



**Fig. 1.** Simulation results for Model I in Example 1 when  $p = 0.05$ . (a) Boxplots of MADE for both WDKLL and NW conditional VaR estimates. (b) Boxplots of MADE for both WDKLL and NW conditional ES estimates.



**Fig. 2.** Simulation results for Model II in Example 1 when  $p = 0.05$ . (a) Boxplots of MADE for both WDKLL and NW conditional VaR estimates. (b) Boxplots of MADE for both WDKLL and NW conditional ES estimates.

4.2. A simulated example

In the simulated example, we demonstrate the finite sample performance of the estimators in terms of the mean absolute deviation error (MADE). For example, the MADE for  $\hat{\mu}_p(x)$  is defined as  $\mathcal{E}_{\mu_p} = \sum_{k=1}^m |\hat{\mu}_p(x_k) - \mu_p(x_k)|/m$ , where  $\{x_k\}_{k=1}^m$  are the pre-determined regular grid points. Similarly, we can define the MADE for  $\hat{v}_p(x)$ , denoted by  $\mathcal{E}_{v_p}$ .

**Example 1.** In this simulated example, we consider the following two models.

**ModelI:** ARCH(1) model with  $X_t = Y_{t-1}$

$$Y_t = 0.01 + 0.62X_t + \sigma_t \varepsilon_t, \quad \sigma_t^2 = 0.15 + 0.65\sigma_{t-1}^2,$$

$$\varepsilon_t \sim N(0, 1).$$

**ModelIII:** Multivariate model with  $X_t$  including two lagged variables  $X_{t1} = Y_{t-1}$  and  $X_{t2} = Y_{t-2}$

$$Y_t = m(X_t) + \sigma(X_t)\varepsilon_t,$$

where  $m(x) = 0.63x_1 - 0.47x_2$ ,  $\sigma^2(x) = 0.5 + 0.23x_1^2 + 0.3x_2^2$ , and  $\varepsilon_t \sim N(0, 1)$ . For Model I, we consider three sample sizes:  $n = 250, 500$ , and  $1000$  and the experiment is repeated 500 times for each sample size. The MADEs are computed for each sample size and each replication. Due to limited space, we present only the boxplots of the 500 MADEs for the CVaR estimates in Fig. 1(a) and the CES estimates in Fig. 1(b). It is visually verified from Fig. 1(a) and (b) that both WDKLL and NW estimations become stable as the sample size increases. Furthermore, it is obvious that the MADEs for the WDKLL estimator are smaller than those for the NW estimator, and therefore it indicates that our WDKLL estimator has smaller bias than that for the NW estimator. Overall, the performance of the WDKLL estimator should be better than that of the NW estimator.

For Model II, three sample sizes:  $n = 200, 400$ , and  $600$ , are considered. For each sample size, we replicate the design 500 times. Here, we only present the boxplots of the 500 MADEs for the CVaR and CES estimates in Fig. 2. Fig. 2(a) displays the boxplots of the 500  $\mathcal{E}_{v_p}$ -values for the WDKLL and NW estimates of CVaR, while the boxplots of the 500  $\mathcal{E}_{\mu_p}$ -values for the WDKLL and NW estimates of CES are given in Fig. 2(b). From Fig. 2(a) and (b), we can observe that the estimations become stable as the sample size increases for both the WDKLL and NW estimators. This is in line with our asymptotic theory that the proposed estimators are consistent. Moreover, the WDKLL estimator has smaller bias than that for the NW estimator.

4.3. A Real Example

**Example 2.** Now we illustrate our proposed methodology, by considering a real data set on Dow Jones Industrials (DJI) index returns. We take a sample of 1801 daily prices from DJI index, from November 3, 1998 to January 3, 2006, and compute the daily returns as 100 times the difference of the log of prices. Let  $Y_t$  be the daily negative log return (log loss) of DJI index and  $X_t$  be the first lagged variable of  $Y_t$ . The estimators proposed in this paper are used to estimate the 5% CVaR and CES functions. The estimation results are shown in Fig. 3 for the 5% CVaR estimate given in Fig. 3(a) and the 5% CES estimate given in Fig. 3(b). Both the CVaR and CES estimates exhibit a U-shape, which corresponds to the so-called “volatility smile”. Therefore, the risk tends to be lower when the lagged log loss of DJI is close to the empirical average, and larger otherwise. We can also observe the curves are asymmetric. This may indicate that the DJI index is more likely to fall if there were a loss within the last day than if there was a same amount of positive return.

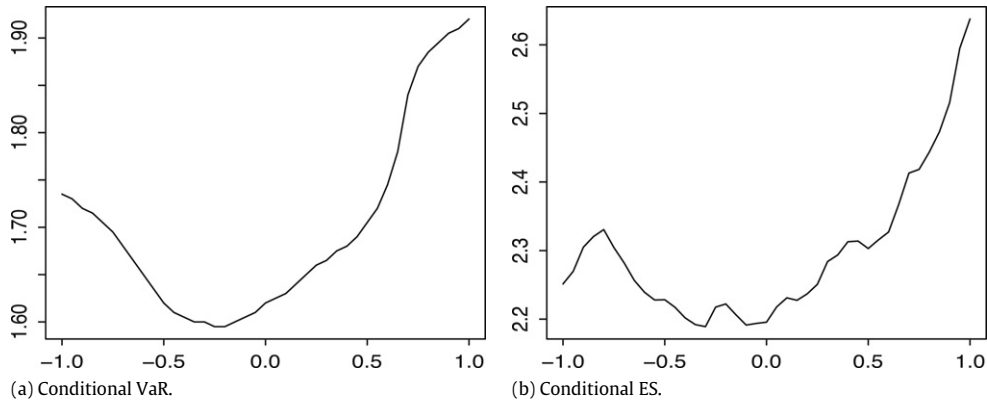


Fig. 3. (a) 5% conditional VaR estimate for DJI index. (b) 5% conditional ES estimate for DJI index.

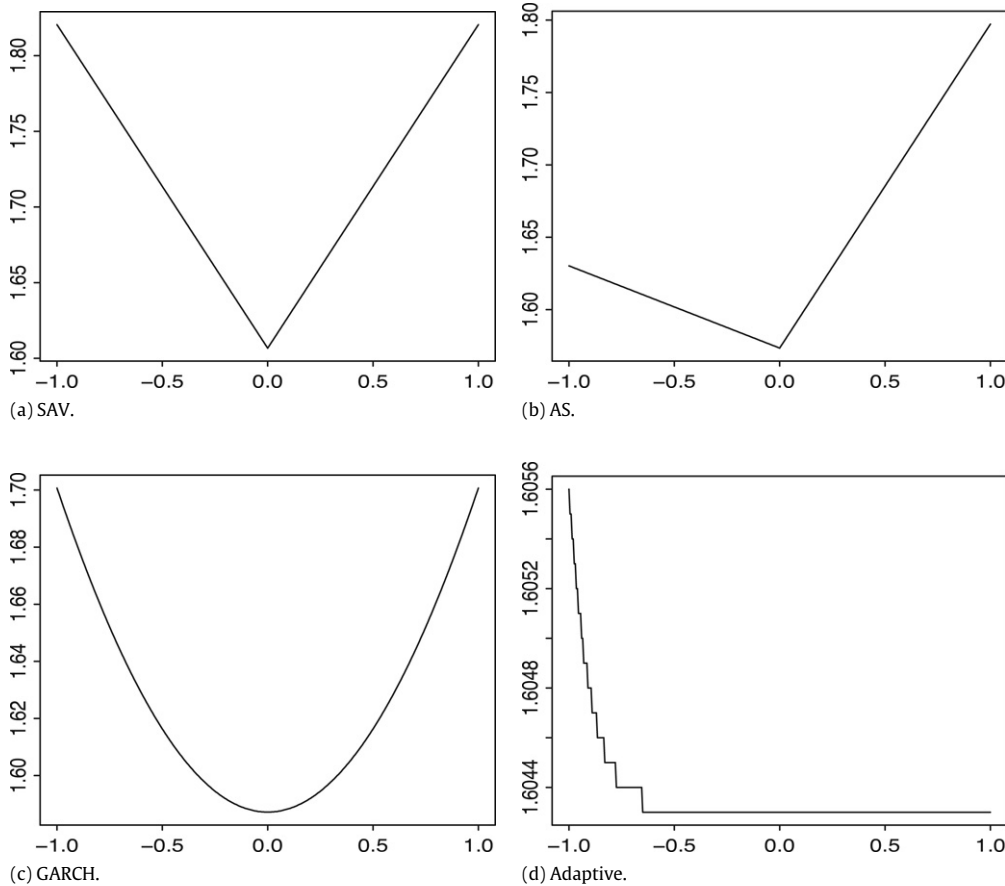


Fig. 4. 5% CAViAR news impact curve for DJI for (a) Symmetric absolute value, (b) Asymmetric slope, (c) Indirect GARCH, and (d) Adaptive.

As mentioned in Section 1, we may gain some insight for parametric fittings based on nonparametric modeling. To build a link between the parametric model and the nonparametric models, we follow the conditional autoregressive value at risk (CAViAR) estimator proposed by Engle and Manganelli (2004) and estimate the CVAiR using the following four parametric CAViAR specifications.

- A. Symmetric absolute value (SAV):  $f_t(\beta) = \beta_1 + \beta_2 f_{t-1}(\beta) + \beta_3 |y_{t-1}|$ ,
- B. Asymmetric slope (AS):  $f_t(\beta) = \beta_1 + \beta_2 f_{t-1}(\beta) + \beta_3 (y_{t-1})^+ + \beta_4 (y_{t-1})^-$ ,
- C. Indirect GARCH(1,1) (GARCH):  $f_t(\beta) = (\beta_1 + \beta_2 f_{t-1}^2(\beta) + \beta_3 y_{t-1}^2)^{1/2}$ ,

D. Adaptive:  $f_t(\beta) = \beta f_{t-1}^2(\beta) + \beta \{ [1 + \exp(G[y_{t-1} - v_{t-1}(\beta)])]^{-1} - p \}$ ,

where  $f_t(\beta) \equiv f_t(x_{t-1}, \beta_p)$  denotes the  $p$ -quantile of returns at time  $t$ ,  $G$  is some positive finite number, and  $p$  is the risk level. Fig. 4 displays 5% news impact curve for CVAiR, where the news impact curve shows how VaR changes as available information set  $X$  (past return) varies, i.e. the curve of  $v_p(x)$ . We also perform the dynamic quantile (DQ) test proposed by Engle and Manganelli (2004), a specification test for the particular CAViAR process and can be useful for model selection; see Engle and Manganelli (2004) for detailed description.

Based on the results of the DQ tests, all four specifications cannot be rejected by both in-sample and out-of sample DQ tests. This suggests that all four specifications are appropriate for 5%

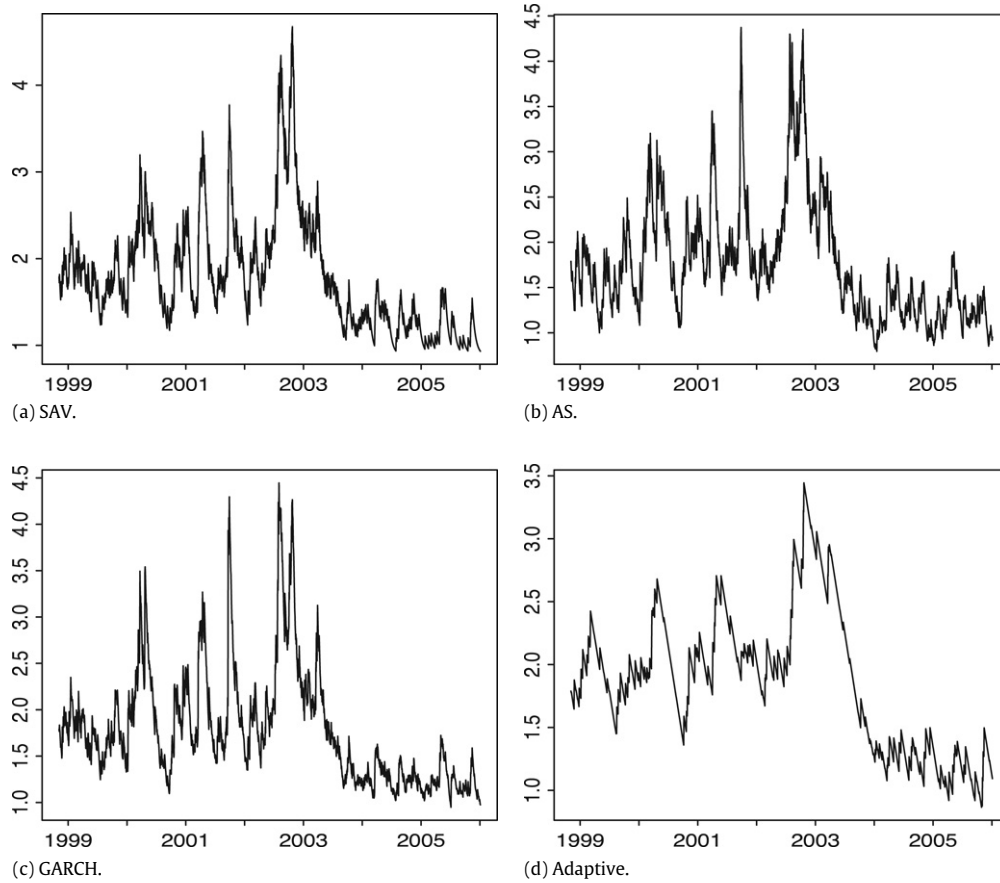


Fig. 5. 5% estimated CVaR based on CAViaR for DJI for (a) Symmetric absolute value, (b) Asymmetric slope, (c) Indirect GARCH, and (d) Adaptive.

CVaR estimation of the DJI index based on the DQ tests. Results of the DQ tests are not reported here due to the space limitation. Fig. 5 presents the 5% CAViaR estimators for the DJI index under the four parametric settings. It is obvious that it is a very hard task in choosing an appropriate setting, since there is little difference among the four specifications. But, by comparing Figs. 3 and 4, one can observe that our WDKLL estimators for CVaR have the similar pattern to the CAViaR estimator in Fig. 4(b), which suggests that positive and negative returns have an asymmetric impact on VaR. Therefore, the asymmetric slope model may be an appropriate specification for the estimation of CVaR for this data set.

5. Proofs of Theorems

In this section, we present only the brief proofs of Theorem 1–Theorem 4; see Cai and Wang (2006) for details. First, we list two lemmas. The proof of Lemma 1 can be found in Cai (2002) and the proof of Lemma 2 is relegated to the Appendix.

Lemma 1. Under Assumptions A1–A5, we have

$$\lambda = -h\lambda_0\{1 + o_p(1)\} \text{ and } p_t(x) = n^{-1}b_t(x)\{1 + o_p(1)\},$$

where  $\lambda_0 = \mu_2(W)g'(x)/[2\mu_2(W^2)g(x)]$  and  $b_t(x) = [1 - h\lambda_0(X_t - x)W_h(x - X_t)]^{-1}$ . Further, we have  $p_t(ch) = n^{-1}b_t^c(ch)\{1 + o_p(1)\}$ , where  $b_t^c(x) = [1 + \lambda_c(X_t - x)K_h(x - X_t)]^{-1}$ .

Lemma 2. Under Assumptions A1–A5, we have, for any  $j \geq 0$ ,

$$J_j = n^{-1} \sum_{t=1}^n c_t(x) \left(\frac{X_t - x}{h}\right)^j = g(x)\mu_j(W) + O_p(h^2),$$

where  $c_t(x) = b_t(x)W_h(x - X_t)$ .

Before we start providing the main steps for proofs of the theorems. First, it follows from Lemmas 1 and 2 that

$$\begin{aligned} W_{c,t}(x, h) &\approx \frac{b_t(x)W_h(x - X_t)}{\sum_{t=1}^n b_t(x)W_h(x - X_t)} \approx n^{-1}g^{-1}(x)b_t(x)W_h(x - X_t) \\ &= \frac{c_t(x)}{ng(x)}. \end{aligned} \tag{12}$$

Now we embark on the proofs of the theorems.

Proof of Theorem 1. By (7), we decompose  $\hat{f}_c(y | x) - f(y | x)$  into three parts as follows

$$\hat{f}_c(y | x) - f(y | x) \equiv I_1 + I_2 + I_3, \tag{13}$$

where  $I_1 = \sum_{t=1}^n \varepsilon_{t,1} W_{c,t}(x, h)$  with  $\varepsilon_{t,1} = Y_t^*(y) - E(Y_t^*(y) | X_t)$ ,

$$I_2 = \sum_{t=1}^n [E(Y_t^*(y) | X_t) - f(y | X_t)]W_{c,t}(x, h) \text{ and}$$

$$I_3 = \sum_{t=1}^n [f(y | X_t) - f(y | x)]W_{c,t}(x, h).$$

An application of the Taylor expansion, (7) and (12), and Lemmas 1 and 2 gives

$$\begin{aligned} I_3 &= \sum_{t=1}^n \frac{1}{2} f^{0,2}(y | x) W_{c,t}(x, h) (X_t - x)^2 + o_p(h^2) \\ &= \frac{h^2}{2} \mu_2(W) f^{0,2}(y | x) + o_p(h^2). \end{aligned}$$

By (2) and following the same steps as in the proof of Lemma 2, we have

$$I_2 = \frac{h_0^2 \mu_2(K)}{2g(x)} n^{-1} \sum_{t=1}^n f^{2,0}(y | X_t) c_t(x) + o_p(h_0^2 + h^2)$$

$$= \frac{h_0^2}{2} \mu_2(K) f^{2,0}(y | x) + o_p(h_0^2 + h^2).$$

Thus, (13) becomes

$$\sqrt{nh_0h} [\widehat{f}_c(y | x) - f(y | x) - B_f(y | x) + o_p(h^2 + h_0^2)] = \sqrt{nh_0h} I_1$$

$$= g^{-1}(x) I_4 \{1 + o_p(1)\} \rightarrow N(0, \sigma_y^2(y | x)),$$

where  $I_4 = \sqrt{h_0h/n} \sum_{t=1}^n \varepsilon_{t,1} c_t(x)$ . This, together with Lemma 3 in the Appendix, therefore, proves the theorem.  $\square$

**Proof of Theorem 2.** Similar to (13), we have

$$\widehat{S}_c(y | x) - S(y | x) \equiv I_5 + I_6 + I_7, \tag{14}$$

where with  $\varepsilon_{t,2} = \bar{G}_{h_0}(y - Y_t) - E(\bar{G}_{h_0}(y - Y_t) | X_t)$ ,  $I_5 = \sum_{t=1}^n \varepsilon_{t,2} W_{c,t}(x, h)$ ,

$$I_6 = \sum_{t=1}^n [E\{\bar{G}_{h_0}(y - Y_t) | X_t\} - S(y | X_t)] W_{c,t}(x, h) \quad \text{and}$$

$$I_7 = \sum_{t=1}^n [S(y | X_t) - S(y | x)] W_{c,t}(x, h).$$

Similar to the analysis of  $I_2$ , by the Taylor expansion, (7), and Lemmas 1 and 2, we have

$$I_7 = \sum_{t=1}^n \frac{1}{2} S^{0,2}(y | x) W_{c,t}(x, h) (X_t - x)^2 + o_p(h^2)$$

$$= \frac{h^2}{2} \mu_2(W) S^{0,2}(y | x) + o_p(h^2).$$

Since  $E[\bar{G}_{h_0}(y - Y_t) | X_t = x] = S(y | x) - \frac{h_0^2}{2} \mu_2(K) f^{1,0}(y | x) + o(h_0^2)$ , and following the same arguments as in the proof of Lemma 2, we have

$$I_6 = -\frac{h_0^2 \mu_2(K)}{2g(x)} n^{-1} \sum_{t=1}^n f^{1,0}(y | X_t) c_t(x) + o_p(h_0^2 + h^2)$$

$$= -\frac{h_0^2}{2} \mu_2(K) f^{1,0}(y | x) + o_p(h_0^2 + h^2).$$

Therefore, by (14),

$$\sqrt{nh} [\widehat{S}_c(y | x) - S(y | x) - B_S(y | x) + o_p(h^2 + h_0^2)] = \sqrt{nh} I_5.$$

Clearly, to accomplish the proof of theorem, it suffices to establish the asymptotic normality of  $\sqrt{nh} I_5$ . To this end, first, we compute  $\text{Var}(\varepsilon_{t,2} | X_t = x)$ . Note that  $E[\bar{G}_{h_0}^2(y - Y_t) | X_t = x] = S(y | x) + 2h_0\alpha(K)f(y | x) + O(h_0^2)$ , which implies that

$$\text{Var}(\varepsilon_{t,2} | X_t = x) = S(y | x)[1 - S(y | x)] + 2h_0\alpha(K)f(y | x) + o(h_0).$$

This, together with the fact that

$$\text{Var}(\varepsilon_{t,2} c_t(x)) = E[c_t^2(x) E\{\varepsilon_{t,2}^2 | X_t\}] = E[c_t^2(x) \text{Var}(\varepsilon_{t,2} | X_t)]$$

leads to

$$h\text{Var}\{\varepsilon_{t,2} c_t(x)\} = \mu_0(W^2)g(x) [S(y | x)\{1 - S(y | x)\} + 2h_0\alpha(K)f(y | x)] + o(h_0).$$

Now, since  $|\varepsilon_{t,2}| \leq 1$ , by following the same arguments as those used in the proofs of Lemmas 2 and 3 in the Appendix (or Lemma 1

and Theorem 1 in Cai (2002)), we can show, although tediously, that

$$\text{Var}(I_8) = \sigma_y^2(y | x)g^2(x) + 2\mu_0(W^2)h_0\alpha(K)f(y | x)g(x) + o(h_0), \tag{15}$$

where  $I_8 = \sqrt{h/n} \sum_{t=1}^n \varepsilon_{t,2} c_t(x)$ , and  $\sqrt{nh} I_5 = g^{-1}(x) I_8 \{1 + o_p(1)\} \rightarrow N(0, \sigma_y^2(y | x))$ . This completes the proof of Theorem 2.  $\square$

**Proof of Theorem 4.** Similar to (11), we use the Taylor expansion, and ignore the higher terms to obtain

$$\int_{\widehat{v}_p(x)}^{\infty} y K_{h_0}(y - Y_t) dy$$

$$\approx \int_{v_p(x)}^{\infty} y K_{h_0}(y - Y_t) dy - v_p(x) K_{h_0}(v_p(x) - Y_t) [\widehat{v}_p(x) - v_p(x)]$$

$$= Y_t \bar{G}_{h_0}(v_p(x) - Y_t) - v_p(x) K_{h_0}(v_p(x) - Y_t) [\widehat{v}_p(x) - v_p(x)] + h_0 G_{1,h_0}(v_p(x) - Y_t).$$

Plugging the above into (9) leads to

$$p\widehat{\mu}_p(x) \approx \widehat{\mu}_{p,1}(x) + I_9, \tag{16}$$

where  $\widehat{\mu}_{p,1}(x) = \sum_{t=1}^n W_{c,t}(x, h) Y_t \bar{G}_{h_0}(v_p(x) - Y_t) - v_p(x) \widehat{f}_c(v_p(x) | x) [\widehat{v}_p(x) - v_p(x)]$ , which will be shown later to be the source of both the asymptotic bias and variance, and  $I_9$  contributes only the asymptotic bias and has the form of  $h_0 \sum_{t=1}^n W_{c,t}(x, h) G_{1,h_0}(v_p(x) - Y_t)$ . From (11), (8) and (12),

$$\widehat{\mu}_{p,1}(x) \approx g^{-1}(x) n^{-1} \sum_{t=1}^n \varepsilon_{t,3} c_t(x) + \sum_{t=1}^n W_{c,t}(x, h) E\{\zeta_t(x) | X_t\}$$

$$\equiv \widehat{\mu}_{p,2}(x) + \widehat{\mu}_{p,3}(x),$$

where  $\zeta_t(x) = [Y_t - v_p(x)] \bar{G}_{h_0}(v_p(x) - Y_t) + p v_p(x)$  and  $\varepsilon_{t,3} = \zeta_t(x) - E\{\zeta_t(x) | X_t\}$ . Next, we derive the asymptotic bias and variance for  $\widehat{\mu}_{p,1}(x)$ . Indeed, we will show that asymptotic bias of  $\widehat{\mu}_p(x)$  comes from both  $\widehat{\mu}_{p,3}(x)$  and  $I_9$ , and the asymptotic variance for  $\widehat{\mu}_{p,1}(x)$  is only from  $\widehat{\mu}_{p,2}(x)$ . First, we consider  $\widehat{\mu}_{p,3}(x)$ . Now, it is easy to see by the Taylor expansion that

$$E[Y_t \bar{G}_{h_0}(v_p(x) - Y_t) | X_t = v]$$

$$= \int_{-\infty}^{\infty} K(u) du \int_{v_p(x) - h_0 u}^{\infty} y f(y | v) dy$$

$$= I_1(v_p(x) | v) - \frac{h_0^2}{2} \mu_2(K) [v_p(x) f^{1,0}(v_p(x) | v) + f(v_p(x) | x)] + o(h_0^2),$$

which leads to

$$\zeta(v) = E[\zeta_t(x) | X_t = v]$$

$$= A(v_p(x) | v) - \frac{h_0^2}{2} \mu_2(K) f(v_p(x) | v) + o(h_0^2), \tag{17}$$

where  $A(v_p(x) | v) = I_1(v_p(x) | v) - v_p(x)[S(v_p(x) | v) - p]$ . It is easy to verify that  $A(v_p(x) | v) = E\{[Y_t - v_p(x)] I(Y_t \geq v_p(x)) | X_t = v\} + p v_p(x)$ ,  $A(v_p(x) | x) = p \mu_p(x)$ , and  $A^{0,2}(v_p(x) | x) = I_1^{0,2}(v_p(x) | x) - v_p(x) S^{0,2}(v_p(x) | x)$ . Therefore, by (17), the Taylor expansion, and (7),

$$\widehat{\mu}_{p,3}(x) = \sum_{t=1}^n W_{c,t}(x, h) \zeta(X_t)$$

$$= \zeta(x) + \frac{1}{2} \zeta''(x) \sum_{t=1}^n W_{c,t}(x, h) (X_t - x)^2 + o_p(h^2).$$

Further, by Lemmas 1 and 2,

$$\begin{aligned} \widehat{\mu}_{p,3}(x) &= p\mu_p(x) + \frac{h^2}{2}\mu_2(W)A^{0.2}(v_p(x) | x) \\ &\quad - \frac{h_0^2}{2}\mu_2(K)f(v_p(x) | x) + o_p(h_0^2). \end{aligned}$$

This, in conjunction with Lemma 4 in the Appendix, concludes that

$$\widehat{\mu}_{p,3}(x) + I_9 = p[\mu_p(x) + B_\mu(x)] + o_p(h^2 + h_0^2),$$

which, together with (16), implies that

$$\widehat{\mu}_{p,1}(x) - p[\mu_p(x) + B_\mu(x)] = \widehat{\mu}_{p,2}(x) + o_p(h^2 + h_0^2)$$

and

$$\widehat{\mu}_p(x) - \mu_p(x) - B_\mu(x) = p^{-1}\widehat{\mu}_{p,2}(x) + o_p(h^2 + h_0^2).$$

Finally, by Lemma 5 in the Appendix, we have

$$\begin{aligned} \sqrt{nh} [\widehat{\mu}_p(x) - \mu_p(x) - B_\mu(x) + o_p(h^2 + h_0^2)] \\ = \frac{1}{pg(x)} I_{10} \{1 + o_p(1)\} \rightarrow N(0, \sigma_\mu^2(x)), \end{aligned}$$

where  $I_{10} = \sqrt{h/n} \sum_{t=1}^n \varepsilon_{t,3} c_t(x)$ . Thus, we have proven the theorem.  $\square$

**Acknowledgments**

We thank Songxi Chen, Jiti Gao and Olivier Scaillet and the guest editors as well as two anonymous referees for their constructive and helpful comments and suggestions. This research was supported, in part, by the National Science Foundation grant DMS-0404954 and funds provided by the University of North Carolina at Charlotte.

**Appendix. Proofs of Lemmas**

In this section, we present the proofs of Lemmas 2–5. Note that we use the same notation as in Sections 2–5. Also, throughout this appendix, we denote a generic constant by  $C$ , which may take different values at different appearances.

**Proof of Lemma 2.** Let  $\xi_t = c_t(x)(X_t - x)^j/h^j$ . It is easy to verify by the Taylor expansion that

$$E(J_j) = E(\xi_t) = \int \frac{v^j W(v)g(x - hv)}{1 + h\lambda_0 v W(v)} dv = g(x)\mu_j(W) + O(h^2), \tag{A.1}$$

and  $E(\xi_t^2) = h^{-1} \int v^{2j} W^2(v)g(x - hv)[1 + h\lambda_0 v W(v)]^{-2} dv = O(h^{-1})$ . Also, by the stationarity, a straightforward manipulation yields

$$n\text{Var}(J_j) = \text{Var}(\xi_1) + \sum_{t=2}^n l_{n,t} \text{Cov}(\xi_1, \xi_t), \tag{A.2}$$

where  $l_{n,t} = 2(n - t + 1)/n$ . The second term on the right hand side of (A.2) can be decomposed into two terms as follows

$$\sum_{t=2}^n |\text{Cov}(\xi_1, \xi_t)| = \sum_{t=2}^{d_n} (\dots) + \sum_{t=d_n+1}^n (\dots) \equiv J_{j1} + J_{j2}, \tag{A.3}$$

where  $d_n = O(h^{-1/(1+\delta_1/2)})$ . For  $J_{j1}$ , it follows by Assumption A4 that  $|\text{Cov}(\xi_1, \xi_t)| \leq C$ , so that  $J_{j1} = O(d_n) = o(h^{-1})$ . For  $J_{j2}$ , Assumption A2 implies that  $|(X_t - x)^j W_h(x - X_t)| \leq Ch^{j-1}$ , so that  $|\xi_t| \leq Ch^{-1}$ . Then, it follows from the Davydov's inequality (see, e.g., Theorem 17.2.1 of Ibragimov and Linnik (1971) that  $|\text{Cov}(\xi_1, \xi_{t+1})| \leq Ch^{-2}\alpha(t)$ , which, together with Assumption A5,

implies that  $J_{j2} \leq Ch^{-2} \sum_{t \geq d_n} \alpha(t) \leq Ch^{-2} d_n^{-(1+\delta_1)} = o(h^{-1})$ . This, together with (A.2) and (A.3), implies that  $\text{Var}(J_j) = O((nh)^{-1}) = o(1)$ . This completes the proof of the lemma.  $\square$

**Lemma 3.** Under Assumption A1–A6 with  $h$  in A3 and A6 replaced by  $hh_0$ , we have

$$I_4 = \sqrt{\frac{h_0 h}{n}} \sum_{t=1}^n \varepsilon_{t,1} c_t(x) \rightarrow N(0, \sigma_f^2(y | x)g^2(x)).$$

**Proof.** It follows by using the same lines as those used in the proof of Lemma 2 and Theorem 1 in Cai (2002), omitted.  $\square$

**Lemma 4.** Under Assumptions A1–A6, we have

$$\begin{aligned} I_9 &= h_0 \sum_{t=1}^n W_{c,t}(x, h)G_{1,h_0}(v_p(x) - Y_t) \\ &= h_0^2 \mu_2(K)f(v_p(x) | x) + o_p(h_0^2). \end{aligned}$$

**Proof.** Define  $\xi_{t,1} = c_t(x)G_{1,h_0}(v_p(x) - Y_t)$ . Then, by Lemma 1,  $I_9 = I_{10}\{1 + o_p(1)\}$ , where  $I_{10} = g^{-1}(x)h_0 \sum_{t=1}^n \xi_{t,1}/n$ . Similar to (A.1), we can show that

$$E(\xi_{t,1}) = h_0 \mu_2(K)f(v_p(x) | x)g(x) + O(h_0 h^2),$$

and

$$\begin{aligned} E(\xi_{t,1}^2) &= E[b_t^2(x)W_h^2(x - X_t)E\{G_{1,h_0}^2(v_p(x) - Y_t) | X_t\}] \\ &= O(h_0/h), \end{aligned}$$

so that  $\text{Var}(\xi_{t,1}) = O(h_0/h)$ . By following the same arguments in the derivation of  $\text{Var}(J_j)$  in Lemma 2, one can show that  $\text{Var}(I_{10}) = O((nh)^{-1}) = o(1)$ . This proves the lemma.  $\square$

**Lemma 5.** Under Assumptions A1–A4 and B2–B5, we have

$$I_{10} = \sqrt{\frac{h}{n}} \sum_{t=1}^n \varepsilon_{t,3} c_t(x) \rightarrow N(0, p^2 g^2(x)\sigma_\mu^2(x)).$$

**Proof.** It follows by using the same lines as those used in the proof of Lemma A.1 and Theorem 1 in Cai (2001), omitted.  $\square$

**References**

Acerbi, C., Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487–1503.  
 Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9, 203–228.  
 Cai, Z., 2001. Weighted Nadaraya–Watson regression estimation. Statistics and Probability Letters 51, 307–318.  
 Cai, Z., 2002. Regression quantiles for time series data. Econometric Theory 18, 169–192.  
 Cai, Z., 2007. Trending time varying coefficient time series models with serially correlated errors. Journal of Econometrics 137, 163–188.  
 Cai, Z., Roussas, G.G., 1998. Efficient estimation of a distribution function under quadrant dependence. Scandinavian Journal of Statistics 25, 211–224.  
 Cai, Z., Tiwari, R.C., 2000. Application of a local linear autoregressive model to BOD time series. Environmetrics 11, 341–350.  
 Cai, Z., Wang, X., 2006. Nonparametric estimation of conditional VaR and expected shortfall. Department of Mathematics and Statistics, University of North Carolina at Charlotte. Working paper.  
 Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. Journal of the American Statistical Association 103.  
 Carrasco, M., Chen, X., 2002. Mixing and moments properties of various GARCH and stochastic volatility models. Econometric Theory 18, 17–39.  
 Chen, S.X., 2006. Nonparametric estimation of expected shortfall. Department of Statistics, Iowa State University. Working paper.  
 Chen, S.X., Tang, C.Y., 2005. Nonparametric inference of value at risk for dependent financial returns. Journal of Financial Econometrics 3, 227–255.

- Chernozhukov, V., Umanstev, L., 2001. Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* 26, 271–292.
- Cosma, A., Scaillet, O., von Sachs, R., 2007. Multivariate wavelet-based shape preserving estimation for dependent observations. *Bernoulli* 13, 301–329.
- Duffie, D., Pan, J., 1997. An overview of value at risk. *Journal of Derivatives* 4, 7–49.
- Duffie, D., Singleton, K.J., 2003. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, Princeton.
- Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. *Modeling Extremal Events For Finance and Insurance*. Springer-Verlag, New York.
- Engle, R.F., Manganelli, S., 2004. CAViaR: Conditional autoregressive value at risk by regression quantile. *Journal of Business and Economics Statistics* 22, 367–381.
- Fan, J., Gijbels, I., 1996. *Local Polynomial Modeling and Its Applications*. Chapman and Hall, London.
- Fan, J., Gu, J., 2003. Semiparametric estimation of value-at-risk. *Econometrics Journal* 6, 261–290.
- Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* 83, 189–206.
- Frey, R., McNeil, A.J., 2002. VaR and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *Journal of Banking and Finance* 26, 1317–1334.
- Hall, P., Wolff, R.C.L., Yao, Q., 1999. Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* 94, 154–163.
- Ibragimov, I.A., Linnik, Yu.V., 1971. *Independent and Stationary Sequences of Random Variables*. Walters–Noordhoff, Groningen, The Netherlands.
- Jorion, P., 2001. *Value at Risk*, 2nd Edition. McGraw-Hill, New York.
- Koenker, R., Bassett, G.W., 1978. Regression quantiles. *Econometrica* 46, 33–50.
- McNeil, A., 1997. Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27, 117–137.
- Meitz, M., Saikkonen, P., 2004. Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Economics and Finance* 573, Stockholm School of Economics. Working paper.
- Morgan, J.P., 1996. *Risk Metrics–Technical Documents*, 4th ed. New York.
- Pötscher, B.M., Prucha, I.R., 1997. *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer-Verlag, Berlin.
- Rockafellar, R., Uryasev, S., 2000. Optimization of conditional value-at-risk. *Journal of Risk* 2, 21–41.
- Scaillet, O., 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* 14, 115–129.
- Scaillet, O., 2005. Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal* 74, 639–660.
- Yu, K., Jones, M.C., 1998. Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.