Weighted Nadaraya-Watson estimation of conditional expected shortfall

Kengo Kato [*]

This version: July 13, 2009[†]

Abstract

This paper addresses the problem of nonparametric estimation of the conditional expected shortfall (CES) which has gained popularity in financial risk management. We propose a new nonparametric estimator of the CES. The proposed estimator is defined as a conditional counterpart of the sample average estimator of the unconditional expected shortfall, where the empirical distribution function is replaced by the weighted Nadaraya-Watson estimator of the conditional distribution function. We establish asymptotic normality of the proposed estimator under an α-mixing condition. The asymptotic results reveal that the proposed estimator has a good bias property. Simulation results illustrate the usefulness of the proposed estimator.

Keywords: α-mixing; conditional expected shortfall; nonparametric estimation; weighted Nadaraya-Watson estimation.
JEL classification numbers: C13, C14, G11.

1   Introduction

This paper addresses the problem of nonparametric estimation of the conditional expected shortfall (CES). The expected shortfall (ES) was first proposed by Artzner et al. (1997) as a risk measure and has gained popularity in financial risk management. Acerbi and Tasche (2002) showed that the ES of their definition is coherent in the sense of Artzner et al. (1999) for a general distribution; whereas the conventional value-at-risk (VaR) measure is not necessarily coherent because of the lack of subadditivity. In this sense, the ES has a theoretical advantage over the VaR.

So far, many authors have proposed various estimation methods for the ES and the CES. The estimation methods may be divided into three categories: parametric, semiparametric and nonparametric. For parametric and semiparametric methods for estimating the ES

---

[*]Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. Email: `kkato@hiroshima-u.ac.jp`
[†]The first version appeared on June 8, 2009. Some errors are corrected.

and the CES, we refer to McNeil et al. (2005). The simplest nonparametric estimator of the ES is the sample average of excessive losses larger than the VaR; see (4) below for the definition. Acerbi and Tasche (2002) mentioned the sample average estimator of the ES. Later, Scaillet (2004) proposed a kernel estimator of the ES as an alternative to the sample average estimator. Chen (2008) compared the sample average estimator with the kernel estimator of Scaillet (2004). He evaluated the variance of the kernel estimator up to the second order and found that the kernel smoothing does not necessarily improve the estimation accuracy. He remarked that the sample average estimator is good enough for estimating the ES.

Scaillet (2005) proposed a nonparametric estimator of the CES. His approach is plugging some nonparametric estimators of the conditional density and the conditional quantile into the expression of the CES. Specifically, he used the Nadaraya-Watson (NW) type double kernel estimator of the conditional density and the conditional quantile associated with the estimated conditional density. Later, based on the same approach, Cai and Wang (2008) proposed another nonparametric estimator of the CES. Instead of the NW double kernel estimator, they used the newly proposed weighted double kernel local linear estimator of the conditional density. Their estimator is advantageous over Scaillet's estimator in terms of the design adaptation. Peracchi and Tanase (2008) mentioned a different NW type kernel estimator of the CES which they called the fully nonparametric estimator. However, they did not investigate its asymptotic properties.

The purpose of this paper is to propose a new nonparametric estimator of the CES. The proposed estimator is defined as a conditional counterpart of the sample average estimator of the unconditional ES, where the empirical distribution function is replaced by the weighted Nadaraya-Watson (WNW) estimator of the conditional distribution function. We refer to the proposed estimator as the WNW estimator of the CES. Unlike the double kernel estimators of Scaillet (2005) and Cai and Wang (2008), the WNW estimator contains only a single smoothing. We establish asymptotic normality of the WNW estimator under an α-mixing condition and show that it is design adaptive at interior points of the support of the design distribution.

We compare the WNW estimator with the double kernel estimator of Cai and Wang (2008). As expected, the double kernel estimator contains an additional bias term carried over from the extra smoothing. In addition, by evaluating the variances of the terms which eventually contribute to the asymptotic variances of both estimators, we find that unlike the estimation of the conditional distribution function and the conditional quantile, the double smoothing does not necessarily lead to a higher order variance reduction in the estimation of the CES. Overall, the theoretical performance of the WNW estimator is comparable to that of the double kernel estimator despite the fact that the former is simpler than the latter. The conclusion is parallel to the aforementioned result of Chen (2008). We also conduct simulation experiments to study the finite sample performance of both estimators. In our limited simulation examples, the WNW estimator outperforms the double kernel estimator in the sense of the MSE.

The rest of the paper is organized as follows. In Section 2, we first introduce the

definition of the (C)ES and describe our method to estimate the CES. In Section 3, we investigate sampling properties of the proposed estimator. In Section 4, we compare the proposed estimator with the double kernel estimator of Cai and Wang (2008). In Section 5, we report a simulation study. The technical proofs are relegated to Appendix.

2 Preliminaries

2.1 Definition of CES

In this section, we first introduce the definition of the ES. Suppose we have a random variable Y. In financial risk management, Y is the negative log return of a portfolio. The ES at level $p \in (0,1)$ of Y is defined as

$$p^{-1}[\mathrm{E}[YI\{Y \geq q_{1-p}\}] + q_{1-p}\{p - \mathrm{P}(Y \geq q_{1-p})\}],$$

where q_{1-p} is the lower $(1-p)$-quantile of Y, namely, $q_{1-p} = \inf\{y : \mathrm{P}(Y \leq y) \geq 1 - p\}$. The present definition of the ES is due to Acerbi and Tasche (2002). They showed that the ES of their definition is coherent in the sense of Artzner et al. (1999) for a general distribution. If Y has a continuous distribution, the ES reduces to the conventional form

$$\mathrm{E}[Y|Y \geq q_{1-p}],$$

which coincides with the tail conditional expectation of Artzner et al. (1999).

The CES is defined analogously. Suppose we have a response variable Y and a explanatory variable X. A typical example of X is a lagged value of Y. Given $X = x_0$, the CES at level $p \in (0,1)$ of Y is defined as

$$m_p(x_0) = p^{-1}[\mathrm{E}[YI\{Y \geq q_{1-p}(x_0)\}|X = x_0] + q_{1-p}(x_0)\{p - \mathrm{P}(Y \geq q_{1-p}(x_0)|X = x_0)\}],$$

where x_0 is a design point, $F(y|x_0)$ is the conditional distribution function of Y given $X = x_0$ and $q_{1-p}(x_0)$ is the conditional lower $(1-p)$-quantile of Y given $X = x_0$, namely, $q_{1-p}(x_0) = \inf\{y : F(y|x_0) \geq 1 - p\}$. It is immediate to see that if $y \mapsto F(y|x_0)$ is continuous, $m_p(x_0)$ reduces to

$$m_p(x_0) = \mathrm{E}[Y|Y \geq q_{1-p}(x_0), X = x_0]. \tag{1}$$

Throughout this paper, we assume that the population conditional distribution is continuous so that $m_p(x_0)$ is of the form (1).

2.2 Proposed estimator

Let (Y_t, X_t), $t = 1, \ldots, n$ be observations of (Y, X). In this section, we introduce our method to estimate the CES and leave some remarks on it. Our basic idea to estimate $m_p(x_0)$ is described as follows. The first step is to estimate the conditional distribution function. Specifically, we use the WNW estimator:

$$\widehat{F}(y|x_0) = \frac{\sum_{t=1}^{n} p_t(x_0)K_h(X_t - x_0)I(Y_t \leq y)}{\sum_{t=1}^{n} p_t(x_0)K_h(X_t - x_0)}, \tag{2}$$

3

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(u) = h^{-1}K(u/h)$. The sequence of nonnegative weights $\{p_t(x_0)\}$ is chosen such that it maximizes $\sum_{t=1}^n \log\{p_t(x_0)\}$ subject to the constraints

$$\sum_{t=1}^n p_t(x_0) = 1, \ \sum_{t=1}^n p_t(x_0)(X_t - x_0)K_h(X_t - x_0) = 0. \tag{3}$$

The expression of $\{p_t(x_0)\}$ is simplified by introducing the Lagrange multiplier. Let $\widehat{\lambda}$ be the maximum point of

$$L_n(\lambda) = -\sum_{t=1}^n \log\left\{1 + \lambda(X_t - x_0)K_h(X_t - x_0)\right\}.$$

Then, each $p_t(x_0)$ is expressed as

$$p_t(x_0) = \frac{1}{n} \cdot \frac{1}{1 + \widehat{\lambda}(X_t - x_0)K_h(X_t - x_0)}.$$

See Fan and Yao (2005), pp.456-457 for a derivation of the expression. Then, the proposed estimator is defined as

$$\widehat{m}_p(x_0) = \frac{\sum_{t=1}^n p_t(x_0)K_h(X_t - x_0)Y_t I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}}{\sum_{t=1}^n p_t(x_0)K_h(X_t - x_0)I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}}.$$

where $\widehat{q}_{1-p}(x_0) = \inf\{y : \widehat{F}(y|x_0) \geq 1 - p\}$ is the lower $(1 - p)$-quantile of $\widehat{F}(y|x_0)$. We refer to $\widehat{m}_p(x_0)$ as the WNW estimator of the CES.

Clearly, the proposed estimator is a conditional counterpart of the sample average estimator of the unconditional ES which is defined as

$$\frac{\sum_{t=1}^n Y_t I(Y_t \geq \widehat{q}_{1-p})}{\sum_{t=1}^n I(Y_t \geq \widehat{q}_{1-p})}, \tag{4}$$

where $\widehat{q}_{1-p}$ is the lower $(1 - p)$-quantile of the empirical distribution function. Under the condition that the population distribution of Y is continuous, Chen (2008) showed asymptotic normality of the sample average estimator for geometric α-mixing processes.

The WNW estimator of the conditional distribution function was proposed by Hall et al. (1999). Later, Cai (2001, 2002) applied the WNW method to the estimation of the conditional mean and the conditional quantile. The WNW estimator is design adaptive due to the second property of (3) and is advantageous over the NW estimator in terms of the bias property. The WNW estimator of the conditional distribution function is favoravble since it lies between 0 to 1 and is nondecreasing in y, which means that the WNW estimator of the conditional distribution function is in itself a distribution function. The local linear estimator of the conditional distribution function does not necessarily meet these properties and is not suited to produce an estimator of the conditional quantile. Asymptotic properties of $\widehat{q}_{1-p}(x_0)$ were investigated by Cai (2002). He established asymptotic normality of $\widehat{q}_{1-p}(x_0)$ for α-mixing processes.

Instead of the WNW estimator, we may use the NW estimator of the conditional distribution function. We note that Peracchi and Tanase (2008) mentioned a NW type estimator of the CES which they call the fully nonparametric estimator. However, they did not investigate its asymptotic properties. Mimicking the proof of Theorem 1 below, it is not difficult to show that under the same conditions of Theorem 1, the NW estimator of $m_p(x_0)$ is $(nh)^{1/2}$-consistent and has the same asymptotic variance as $\widehat{m}_p(x_0)$ when x_0 is an interior point. However, the NW estimator has well known disadvantages that it is not design adaptive. In some cases of the design distribution, the bias of the NW estimator may become very large. Therefore, we recommend to use the WNW estimator.

3 Sampling properties

In this section, we investigate asymptotic properties of the WNW estimator of the CES. We note that the results of this section are not included in Cai (2001) since $Y_t I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}$ depends on the overall data. In order to justify our asymptotic theory, we consider some regularity conditions stated as follows:

(A1) The kernel function $K(\cdot)$ is a symmetric, bounded and compactly supported density function.

(A2) The process $\{(Y_t, X_t), t = 1, 2, \dots\}$ is stationary and α-mixing:

$$\alpha(j) := \sup_{i \geq 1} \sup_{A \in \mathcal{F}_1^i, B \in \mathcal{F}_{i+j}^\infty} |\mathrm{P}(A \cap B) - \mathrm{P}(A)\mathrm{P}(B)| \to 0,$$

where $\mathcal{F}_i^j$ is the σ-field generated by $\{(Y_k, X_k), k = i, \dots, j\}$ $(j \geq i)$. In addition,

$$\sum_{j=1}^\infty j^\gamma \{\alpha(j)\}^{1-2/\delta} < \infty$$

for some $\delta > 2$ and $\gamma > 1 - 2/\delta$.

(A3) For $j \geq 1$, let $g(x_1, x_{1+j}; j)$ denote the joint density of (X_1, X_{1+j}). Then, $g(x_1, x_{1+j}; j)$ is bounded in a neighborhood of $(x_1, x_{1+j}) = (x_0, x_0)$ uniformly over $j \geq 1$.

(A4) As $n \to \infty$, $h \to 0$ and $nh^3 \to \infty$. There exists a sequence of constants s_n satisfying $s_n \to \infty$ and $s_n = o((nh)^{1/2})$ such that

$$(n/h)^{1/2}\alpha(s_n) \to 0, \text{ as } n \to \infty.$$

(A5) The map $(x, x_{1+j}) \mapsto \mathrm{E}[Y_1^2 + Y_{1+j}^2 | X_1 = x_1, X_{1+j} = x_{1+j}]$ is bounded in a neighborhood of (x_0, x_0) uniformly over $j \geq 1$.

(A6) There exists $\delta^* > \delta$ such that the map $x \mapsto \mathrm{E}[|Y|^{\delta^*} | X = x]$ is bounded in a neighborhood of x_0, $\alpha(j) = O(j^{-\theta^*})$ for some $\theta^* \geq \delta^*\delta/\{2(\delta^* - \delta)\}$ and $n^{1/2-\delta/4}h^{\delta/\delta^*-1/2-\delta/4} = O(1)$.

(A7) The map $x \mapsto F(q_{1-p}(x_0)|x)$ is twice continuously differentiable in a neighborhood of x_0. In addition, the conditional density $f(y|x)$ of Y given $X = x$ satisfies the following conditions:

    (a) The map $x \mapsto f(q_{1-p}(x_0)|x)$ is positive and continuous at x_0.

    (b) There exists a constant $A_1 > 0$ such that $|f(y|x) - f(q_{1-p}(x_0)|x)| \leq A_1|y - q_{1-p}(x_0)|$ in a neighborhood of $(y, x) = (q_{1-p}(x_0), x_0)$.

(A8) The marginal density $g(x)$ of X is positive and continuously differentiable in a neighborhood of x_0.

We state some remarks on the conditions. The requirement of the bounded support of the kernel function can be relaxed at the expense of the lengthier proof. We impose the α-mixing condition, which covers many commonly used financial time series. See Section 2.6.1 of Fan and Yao (2005) and references therein. Chen (2008) imposed the geometric α-mixing condition, which is stronger than our α-mixing condition. Conditions (A1)-(A6) is standard in nonparametric regression. Actually, conditions (A1)-(A5) correspond to Conditions 1-3 in Masry and Fan (1997) except that we require that $nh^3 \to \infty$. The condition that $nh^3 \to \infty$ is not assumed in Cai (2001, 2002) and Cai and Wang (2008) but is required for the asymptotics of the WNW method; see the proof of Lemma 1. Condition (A6) corresponds to condition (B7) in Cai (2001). This condition is used to ensure the central limit theorem. Condition (A7) corresponds to conditions C1 and C2 in Cai (2002). We note that condition (A7) ensures that the conditional $(1 - p)$-quantile of Y given $X = x_0$ is determined uniquely and $m_p(x_0) = E[Y|Y \geq q_{1-p}(x_0)]$. From Theorem 3 of Cai (2002), under conditions (A1)-(A4) and (A7)-(A8), if $h = O(n^{-1/5})$, then as $n \to \infty$, $\widehat{q}_{1-p}(x_0) = q_{1-p}(x_0) + O_p\{(nh)^{-1/2}\}$.[1]

In the sequel, we use the following notations: $l_j(u|x) = E[Y^j I\{Y \geq u\}|X = x]$ for $j = 1, 2$; $l_1^{(a)}(u|x) = \partial^a l_1(u|x)/\partial x^a$ for $a = 1, 2$; $F^{(a)}(y|x) = \partial^a F(y|x)/\partial x^a$ for $a = 1, 2$; $\mu_j(K) = \int_{-\infty}^{\infty} u^j K(u)du$ for $j = 0, 2$. The next theorem is the main result of this paper. The proof of the theorem is relegated to Appendix A1.

Theorem 1. *Under conditions (A1)-(A8), if $h = O(h^{-1/5})$, the map $x \mapsto l_1(q_{1-p}(x_0)|x)$ is twice continuously differentiable in a neighborhood of x_0 and the map $x \mapsto l_2(q_{1-p}(x_0)|x)$ is continuous at x_0, then as $n \to \infty$,*

$$(nh)^{1/2}\left\{\widehat{m}_p(x_0) - m_p(x_0) - \frac{h^2\mu_2(K)}{2}B_p(x_0)\right\} \xrightarrow{d} N\left\{0, \mu_0(K^2)\sigma_p^2(x_0)/g(x_0)\right\},$$

where

$$B_p(x_0) = p^{-1}l_1^{(2)}(q_{1-p}(x_0)|x_0) + p^{-1}q_{1-p}(x_0)F^{(2)}(q_{1-p}(x_0)|x_0),$$
$$\sigma_p^2(x_0) = p^{-2}l_2(q_{1-p}(x_0)|x_0) - m_p^2(x_0) + (p^{-1} - 1)q_{1-p}(x_0)\{q_{1-p}(x_0) - 2m_p(x_0)\}.$$

---

[1]The author thinks that his proof may contain a technical gap. However, the conclusion of Theorem 3 in Cai (2002) still holds under those conditions; see Remark 1 in Appendix A1.

From the proof of Theorem 1, we obtain the representation

$$\widehat{m}_p(x_0) - m_p(x_0) = \frac{\sum_{t=1}^n p_t(x_0)K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\}I\{Y_t \geq q_{1-p}(x_0)\}}{p\sum_{t=1}^n p_t(x_0)K_h(X_t - x_0)}$$
$$- \{m_p(x_0) - q_{1-p}(x_0)\} + O_p\{(nh)^{-1}\}, \quad (5)$$

The first term of the right hand side of (5) is identical to the WNW regression estimator applied to the infeasible sample $\{(p^{-1}\{Y_t - q_{1-p}(x_0)\}I\{Y_t \geq q_{1-p}(x_0)\}, X_t), t = 1, \ldots, n\}$, which determines the asymptotic bias and variance of $\widehat{m}_p(x_0)$. The sample average estimator of the unconditional ES admits the similar representation; see equation (6) in Chen (2008). Unlike the estimation of the unconditional ES, the asymptotic variance of $\widehat{m}_p(x_0)$ is same as that for the independent observations because of the localization effect; see Section 5.3 of Fan and Yao (2005).

From Theorem 1, we see that the asymptotic mean squared error (AMSE) of $\widehat{m}_p(x_0)$ is

$$\text{AMSE}(x_0; p) = \frac{h^4}{4}\{\mu_2(K)B_p(x_0)\}^2 + \frac{\mu_0(K^2)}{nhg(x_0)}\sigma_p^2(x_0).$$

Minimizing the AMSE yields the optimal bandwidth

$$h_{\text{opt}}(x_0; p) = n^{-1/5}\left[\frac{\mu_0(K^2)\sigma_p^2(x_0)}{g(x_0)\{\mu_2(K)B_p(x_0)\}^2}\right]^{1/5}.$$

Unfortunately, $h_{\text{opt}}(x_0; p)$ is of a complicated form. The estimation of the unknown quantities in $h_{\text{opt}}(x_0; p)$ is feasible but complex. In practice, we suggest to use the the bootstrap approach described in Section 2.3 of Hall et al. (1999), which is simple to implement.

Theorem 1 reveals that the asymptotic bias of $\widehat{m}_p(x_0)$ is independent of the design density $g(\cdot)$, that is, $\widehat{m}_p(x_0)$ is design adaptive. The design adaptation of $\widehat{m}_p(x_0)$ is due to the second part of the property (3) of the weights $\{p_t(x_0)\}$. Meanwhile, it is not difficult to see that the asymptotic bias of the NW estimator of $m_p(x_0)$ is

$$\frac{h^2\mu_2(K)}{2}\left[B_p(x_0) + 2\frac{g'(x_0)}{g(x_0)}p^{-1}\{l_1^{(1)}(q_{1-p}(x_0)|x_0) + q_{1-p}(x_0)F^{(1)}(q_{1-p}(x_0)|x_0)\}\right].$$

Whereas, the asymptotic variance of the NW estimator is same as that of $\widehat{m}_p(x_0)$. It is well known that the dependence on the ratio $g'(x_0)/g(x_0)$ produces a large bias to the estimator in some cases. See Fan (1992) and Section 3.2.4 of Fan and Gijbels (1996). Thus, in terms of the bias property, the WNW estimator is advantageous over the NW estimator.

4 Double kernel estimator

In this section, we compare the WNW estimator with the double kernel estimator of Cai and Wang (2008). The basic idea of Cai and Wang (2008) is plugging some nonparametric estimators of $f(y|x_0)$ and $q_{1-p}(x_0)$ into the expression

$$m_p(x_0) = p^{-1}\int_{q_{1-p}(x_0)}^{\infty} yf(y|x_0)dy.$$

Specifically, combining the WNW method and the double kernel local linear method of Yu and Jones (1998), Cai and Wang (2008) proposed the weighted double kernel local linear estimator of $f(y|x)$:

$$\widetilde{f}(y|x_0) = \frac{\sum_{t=1}^n p_t(x_0)K_h(X_t - x_0)W_{h_0}(y - Y_t)}{\sum_{t=1}^n p_t(x_0)K_h(X_t - x_0)},$$

where $W(\cdot)$ is a symmetric kernel function and h_0 is a bandwidth associated with $W(\cdot)$. Let $\widetilde{F}(y|x_0)$ be the conditional distribution function with conditional density $\widetilde{f}(y|x_0)$:

$$\widetilde{F}(y|x_0) = \int_{-\infty}^y \widetilde{f}(u|x_0)du.$$

As an estimator of $q_{1-p}(x_0)$, Cai and Wang (2008) proposed to use the solution $\widetilde{q}_{1-p}(x_0)$ of the equation $\widetilde{F}(\widetilde{q}_{1-p}(x_0)|x_0) = 1 - p$, where we note that such $\widetilde{q}_{1-p}(x_0)$ always exists since $y \mapsto \widetilde{F}(y|x_0)$ is continuous. Let $\widetilde{m}_p(x_0)$ denote the corresponding plug-in estimator:

$$\widetilde{m}_p(x_0) = p^{-1} \int_{\widetilde{q}_{1-p}(x_0)}^\infty y\widetilde{f}(y|x_0)dy.$$

Cai and Wang (2008) established asymptotic normality of $\widetilde{m}_p(x_0)$ for α-mixing processes. We now compare the asymptotic biases and variances of $\widehat{m}_p(x_0)$ and $\widetilde{m}_p(x_0)$.

[1] Bias: From Theorem 4 of Cai and Wang (2008), the asymptotic bias of $\widetilde{m}_p(x_0)$ is

$$\text{bias}\{\widetilde{m}_p(x_0)\} : \frac{h^2\mu_2(K)}{2}B_p(x_0) + \frac{h_0^2\mu_2(W)}{2}p^{-1}f(q_{1-p}(x_0)|x_0).$$

It is clear that $\widetilde{m}_p(x_0)$ contains the additional bias term carried over from the y direction smoothing. Since the first order asymptotic variance of $\widetilde{m}_p(x_0)$ does not depend on h_0, it is reasonable to make h_0 small relative to h as suggested in Cai and Wang (2008). Although the additional term is theoretically negligible if we set $h_0 = o(h)$, it might affect the finite sample performance of $\widetilde{m}_p(x_0)$. When h_0 is small relative to h, the additional term improves the bias if $B_p(x_0)$ is negative; otherwise, it has an adverse effect on the estimation accuracy. In a practical situation, we do not know the sign of $B_p(x_0)$ in advance and there is an uncertainty about the effect of the y direction smoothing on the bias. The WNW estimator does not have such an uncertainty and may be preferable in this respect.

[2] Variance: From Theorem 4 of Cai and Wang (2008), the first order asymptotic variances of $\widehat{m}_p(x_0)$ and $\widetilde{m}_p(x_0)$ are same. However, for the estimation of the conditional distribution function and the conditional quantile, the y direction smoothing improves the variance in a higher order sense; see Azzalini (1981), Yu and Jones (1998), Cai and Roussas (1998), Chen and Tang (2005) and Remark 5 of Cai and Wang (2008), among others. Thus, we need to look at the variances of $\widehat{m}_p(x_0)$ and $\widetilde{m}_p(x_0)$ more carefully. Since it is difficult to evaluate the exact variances, we evaluate the variances of the terms which eventually contribute to the asymptotic variances of $\widehat{m}_p(x_0)$ and $\widetilde{m}_p(x_0)$. Although the approximation is formal, we believe that it reflects the true structure of the variance relationship between

$\widehat{m}_p(x_0)$ and $\widetilde{m}_p(x_0)$. From the proof of Theorem 1, the asymptotic variance of $\widehat{m}_p(x_0)$ comes from the variance of

$$I_1 = \frac{1}{npg(x_0)} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\} I\{Y_t \geq q_{1-p}(x_0)\},$$

where $b_t(x_0)$ is defines as (7). Meanwhile, from the proof of Theorem 4 in Cai and Wang (2008), the asymptotic variance of $\widetilde{m}_p(x_0)$ comes from the variance of

$$I_2 = \frac{1}{npg(x_0)} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\} G_{h_0}(Y_t - q_{1-p}(x_0)),$$

where $G(\cdot)$ is the distribution function of $W(\cdot)$ and $G_{h_0}(u) = G(u/h_0)$. Under certain regularity conditions, it is shown that

$$\text{var}(I_1) - \text{var}(I_2) = o\{(nh)^{-1}h_0\}. \tag{6}$$

For a derivation of (6), see Appendix A2. From (6), unlike the estimation of the conditional distribution function and the conditional quantile, the y direction smoothing does not improve the variance at the order of $(nh)^{-1}h_0$ for the estimation of the CES.

Overall, we may conclude that theoretical performance of the WNW estimator is comparable to that of the double kernel estimator despite the fact that the former is simpler than the latter. The conclusion is parallel to the result of Chen (2008) who studied the estimation on the unconditional ES. He compared the sample average estimator with the kernel estimator of Scaillet (2004) which contains the y direction smoothing. He evaluated the variance of the kernel estimator up to the second order and found that the y direction smoothing does not improve the variance at the second order while it produces an additional bias. He remarked that the simpler sample average estimator is good enough for estimating the ES. In the next section, we study the finite sample performance of both estimators.

5 Simulation study

In this section, we report simulation results which evaluate the finite sample performance of the WNW estimator and the double kernel (DK) estimator of Cai and Wang (2008). Specifically, we consider the following two models.

▶ Model 1 [ARCH(1)]: $Y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = 0.30 + 0.65 Y_{t-1}^2$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$, $X_t = Y_{t-1}$.

▶ Model 2: $Y_t = \sin(0.75 X_t) + 0.5\epsilon_t$, $X_t \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$, $X_t \perp\!\!\!\perp \epsilon_t$.

The ARCH process of Model 1 is shown to be α-mixing with exponetial decaying coefficients; see remark (viii) in Fan and Yao (2005), pp. 70. Thus, the process of Model 1 trivially satisfies the mixing conditions of Theorem 1. We note that Model 2 is due to Fan (1992). Figure 1 depicts the CES functions for both models where $p = 0.05$. Throughout
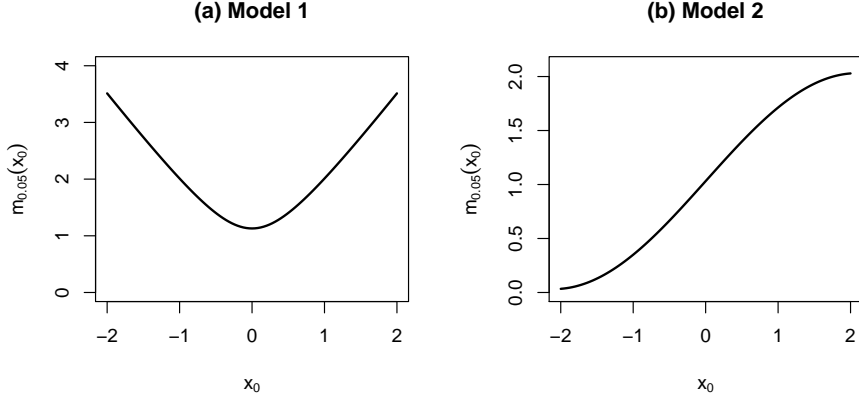
Figure 1: The CES functions for Models 1-2 where $p = 0.05$.

this section, we use the Gaussian kernel. The sample size n is 400 for each model. The number of repetitions is 10,000 for each simulation. The computational results of this section were obtained by using the matrix language Ox (Doornik, 2002). To generate ARCH processes, we used the G@RCH 4.2 package (Laurent and Peters, 2006). Figures 2-3 display biases, standard deviations and root mean squared errors of the WNW estimator and the DK estimator at $p = 0.05$ over a set of design points and a set of bandwidth values. Specifically, we choose 0 and the 0.75-quantile of X_t as design points for each model. In addition, we set $h = an^{-1/5}$, where a moves from 0.5 to 1.2 by 0.05 for Model 1 and from 0.6 to 1.3 by 0.05 for Model 2. As Cai and Wang (2008) noted, h_0 is not sensitive to $\widetilde{m}_p(x_0)$ as far as h_0 is sufficiently small relative to h. For simplicity, we set $h_0 = 0.1h$. Figures 2-3 show that $\widetilde{m}_p(x_0)$ contains an additional bias in comparison with $\widehat{m}_p(x_0)$, which corresponds to the observation in the previous section. More importantly, the MSE of $\widehat{m}_p(x_0)$ is smaller than that of $\widetilde{m}_p(x_0)$ over the set of bandwidth values for both models. Overall, the simulation results support the use of $\widehat{m}_p(x_0)$.

A   Appendix

A.1   Proof of Theorem 1

In this section, we provide a proof of Theorem 1. Before proving the theorem, we prepare some auxiliary lemmas.

Lemma 1. *Under conditions (A1)-(A4) and (A8), $\widehat{\lambda} = \lambda^* + O_p(h^3)$, where*

$$\lambda^* = h\mu_2(K)g'(x_0)/\{\mu_2(K^2)g(x_0)\}.$$

Proof. See the proof of Lemma 2 in Cai (2002). We remark that his evaluation of A_1 contains a minor error. The definition of A_1 is

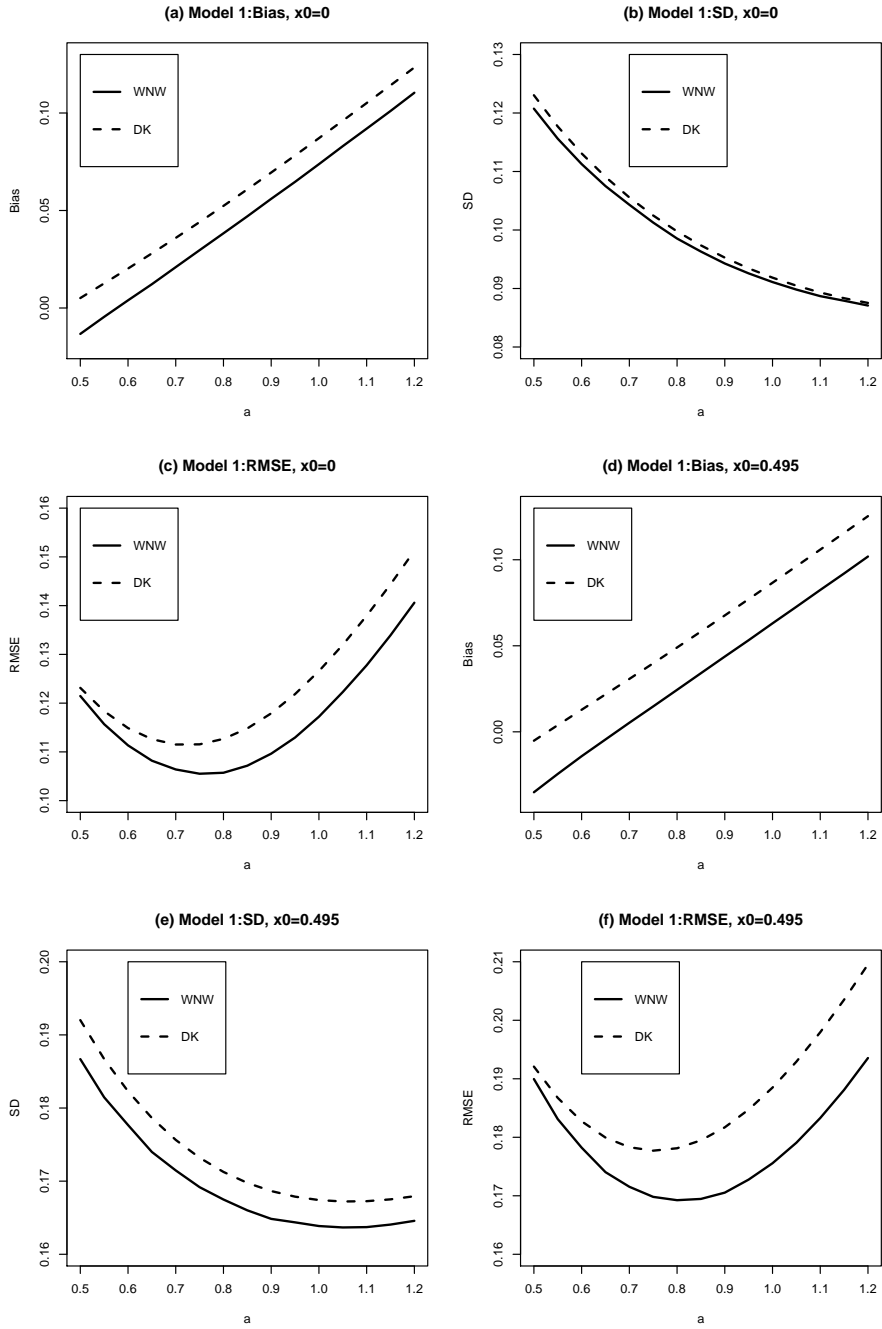$$A_1 = \frac{1}{n}\sum_{t=1}^{n}(X_t - x_0)K_h(X_t - x_0).$$

10

Figure 2: Biases, standard deviations (SD) and root mean squared errors (RMSE) of the WNW estimator and the DK estimator at $p = 0.05$ for Model 1, where $x_0 = \{0, 0.495\}$, $h = an^{-1/5}$ and $h_0 = 0.1h$
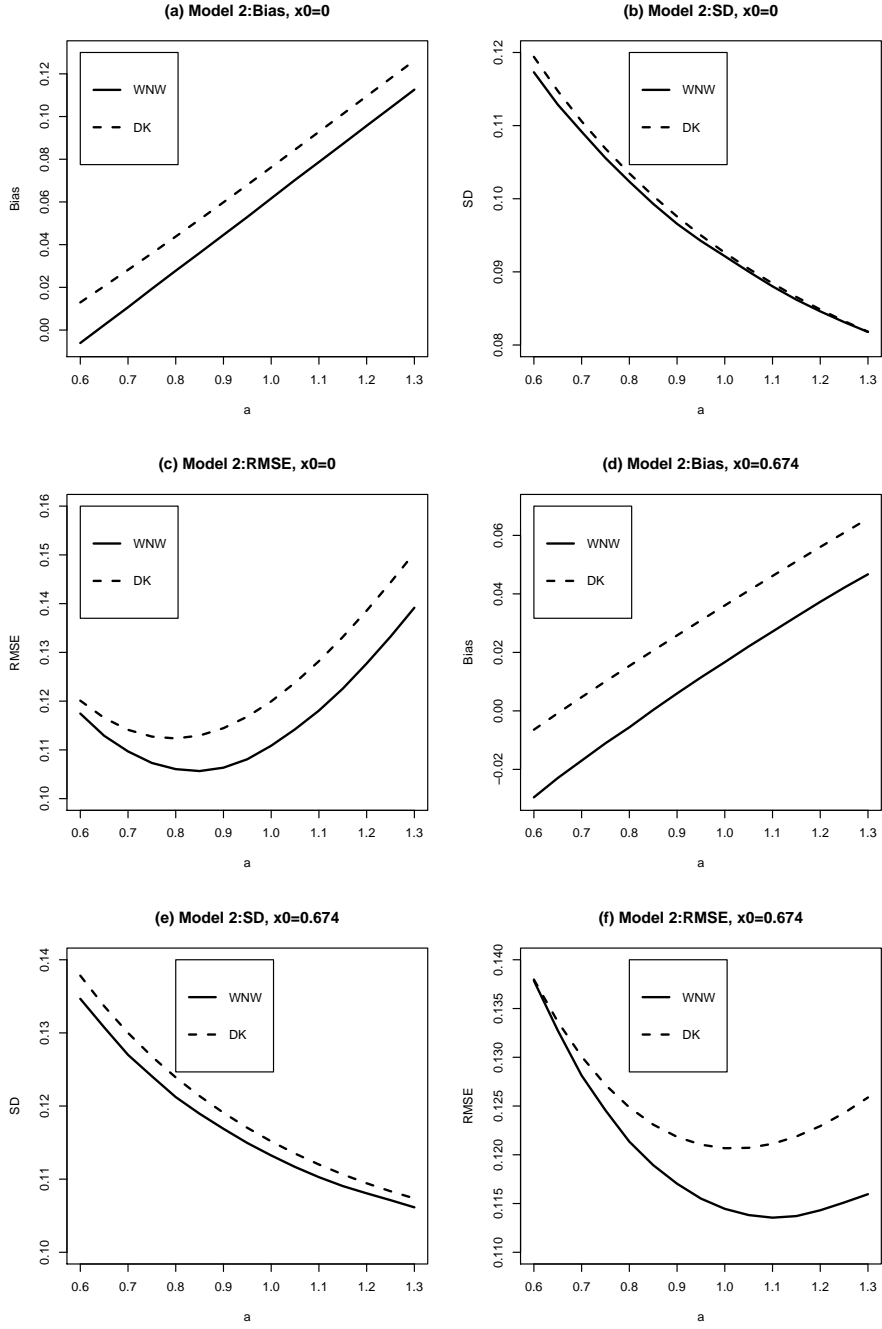
Figure 3: Biases, standard deviations (SD) and root mean squared errors (RMSE) of the WNW estimator and the DK estimator at $p = 0.05$ for Model 2, where $x_0 \in \{0, 0.679\}$, $h = an^{-1/5}$ and $h_0 = 0.1h$

A simple algebra yields that

$$
\mathrm{E}[A_1] = h \int_{-\infty}^{\infty} xK(x)g(x_0 + xh)dx = h \int_{-\infty}^{\infty} xK(x)\{g(x_0 + xh) - g(x_0)\}dx
$$
$$
= h^2 \mu_2(K)g'(x_0) + o(h^2).
$$

From Theorem 1 in Masry and Fan (1997), $\mathrm{var}(A_1) = O(n^{-1}h)$. By Chebyshev's inequality, for every $\epsilon > 0$, we have

$$
\mathrm{P}(|A_1 - \mathrm{E}[A_1]| > h^2\epsilon) \leq \frac{\mathrm{var}(A_1)}{h^4\epsilon^2} = O(n^{-1}h^{-3}) = o(1).
$$

Thus, we have $A_1 = \mathrm{E}[A_1] + o_p(h^2) = h^2\mu_2(K)g'(x_0) + o_p(h^2)$. $\qquad\square$

Put
$$
b_t(x_0) = \frac{1}{1 + \lambda^*(X_t - x_0)K_h(X_t - x_0)}. \tag{7}
$$
Since $\lambda^* = O(h)$, for large n, $b_t(x_0)$ is non-negative and bounded over all t. Note that for a sequence of non-negative random variables, say, $\{w_t\}$, we have

$$
|\textstyle\sum_{t=1}^{n} p_t(x_0)w_t - n^{-1}\sum_{t=1}^{n} b_t(x_0)w_t| \leq \sum_{t=1}^{n} |p_t(x_0) - n^{-1}b_t(x_0)||w_t|
$$
$$
\leq o_p(1)\textstyle\sum_{t=1}^{n} |n^{-1}b_t(x_0)w_t|
$$
$$
= \frac{o_p(1)}{n}\textstyle\sum_{t=1}^{n} b_t(x_0)w_t.
$$

Thus, we have $\sum_{t=1}^{n} p_t(x_0)w_t = \{1 + o_p(1)\}n^{-1}\sum_{t=1}^{n} b_t(x_0)w_t$. The same conclusion holds for a sequence of non-positive random variables. We will use this fact in the proofs of Lemma 2 and 3 below.

Lemma 2. *Under conditions (A1)-(A4) and (A8),*

$$
\sum_{t=1}^{n} p_t(x_0)\left(\frac{X_t - x_0}{h}\right)^j K_h(X_t - x_0) \xrightarrow{p} \mu_j(K)g(x_0), \ \ j = 0, 2.
$$

Proof. Since $(X_t - x_0)^j$ is non-negative if $j = 0, 2$, by Lemma 1, we have

$$
\sum_{t=1}^{n} p_t(x_0)\left(\frac{X_t - x_0}{h}\right)^j K_h(X_t - x_0) = \frac{1 + o_p(1)}{n}\sum_{t=1}^{n} b_t(x_0)\left(\frac{X_t - x_0}{h}\right)^j K_h(X_t - x_0).
$$

The rest of the proof is standard. See Theorem 1 of Masry and Fan (1997). $\qquad\square$

Lemma 3. *Under conditions (A1)-(A4) and (A7)-(A8), if $h = O(n^{-1/5})$, then as $n \to \infty$,*

$$
\sum_{t=1}^{n} p_t(x_0)K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\}I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}
$$
$$
= \sum_{t=1}^{n} p_t(x_0)K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\}I\{Y_t \geq q_{1-p}(x_0)\} + O_p\{(nh)^{-1}\}. \tag{8}
$$

13

Proof. Observe that

$$\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\}[I\{Y_t \geq \widehat{q}_{1-p}(x_0)\} - I\{Y_t \geq q_{1-p}(x_0)\}]$$

$$= \frac{1 + o_p(1)}{n} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\}[I\{Y_t \geq \widehat{q}_{1-p}(x_0)\} - I\{Y_t \geq q_{1-p}(x_0)\}].$$

$$=: \{1 + o_p(1)\}I_n,$$

where the first equality follows from Lemma 1 and the fact that $\{Y_t - q_{1-p}(x_0)\}[I\{Y_t \geq \widehat{q}_{1-p}(x_0)\} - I\{Y_t \geq q_{1-p}(x_0)\}]$ is non-positive. Thus, it suffices to show that

$$I_n = O_p\{(nh)^{-1}\}. \tag{9}$$

Let $Z_t = Y_t - q_{1-p}(x_0)$. Define

$$U_n(s) = \frac{1}{n} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0) Z_t[I\{Z_t \geq (nh)^{-1/2}s\} - I\{Z_t \geq 0\}], \ s \in \mathbb{R}.$$

We show that for each fixed $l > 0$,

$$\sup_{|s| \leq l} |U_n(s)| = O_p\{(nh)^{-1}\}. \tag{10}$$

Before proceeding to showing (10), we verify that (9) follows from (10). Since $\widehat{q}_{1-p}(x_0) = q_{1-p}(x_0) + O_p\{(nh)^{-1/2}\}$, for every $\epsilon > 0$, there exists a constant $l > 0$ such that for large n,

$$\mathrm{P}\left(|\widehat{q}_{1-p}(x_0) - q_{1-p}(x_0)| > (nh)^{-1/2}l\right) \leq \frac{\epsilon}{2}.$$

From (10), there exists a constant $M > 0$ such that for large n,

$$\mathrm{P}\left(\sup_{|t| \leq l} |U_n(t)| > (nh)^{-1}M\right) \leq \frac{\epsilon}{2}.$$

Thus, for large n,

$$\mathrm{P}\left(|I_n| > (nh)^{-1/2}M\right)$$

$$\leq \mathrm{P}\left(\sup_{|t| \leq l} |U_n(t)| > (nh)^{-1}M\right) + \mathrm{P}\left(|\widehat{q}_{1-p}(x_0) - q_{1-p}(x_0)| > (nh)^{-1/2}l\right)$$

$$\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

which implies (9).

It remains to show (10) for each fixed $l > 0$. Observe that

$$U_n(s) = \begin{cases} -n^{-1} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0) Z_t I\{0 \leq Z_t < (nh)^{-1/2}s\} & \text{if } s > 0, \\ n^{-1} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0) Z_t I\{(nh)^{-1/2}s \leq Z_t < 0\} & \text{if } s < 0. \end{cases}$$

14

Define

$$U_{1n}(s) = \frac{1}{n} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0) Z_t I\{0 \leq Z_t < (nh)^{-1/2}s\}, \ s > 0,$$

$$U_{2n}(s) = \frac{1}{n} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0) Z_t I\{(nh)^{-1/2}s \leq Z_t < 0\}, \ s < 0.$$

Then, it suffices to show that

$$\sup_{0 \leq s \leq l} U_{1n}(s) = O_p\{(nh)^{-1}\}, \quad \sup_{-l \leq s \leq 0} \{-U_{2n}(s)\} = O_p\{(nh)^{-1}\}.$$

We only prove the first part since the proof for the latter part is completely analogous. By the monotonicity of the indicator function, we have

$$\sup_{0 \leq s \leq l} U_{1n}(t) \leq \frac{1}{n} \sum_{t=1}^{n} b_t(x_0) K_h(X_t - x_0) Z_t I\{0 \leq Z_t < (nh)^{-1/2}l\}$$

$$= \bar{U}_{1n}.$$

Thus, it suffices to show that $\bar{U}_{1n} = O_p\{(nh)^{-1}\}$. Let $f_{Z|X}(z|x)$ denote the conditional density of Z_t given $X_t = x$. Clearly, $f_{Z|X}(z|x) = f(z + q_{1-p}(x_0)|x)$. From Condition (A7)-(b), there exists a constant $A_0 > 0$ such that $f_{Z|X}(z|x) \leq A_0$ in a neighborhood of $(z, x) = (0, x_0)$. By conditioning on X_t, we have

$$E[\bar{U}_{1n}] = E\left[b_t(x_0) K_h(X_t - x_0) \int_0^{(nh)^{-1/2}l} z f_{Z|X}(z|X_t) dz \right]$$

$$\leq A_0 l^2 (nh)^{-1} E[b_t(x_0) K_h(X_t - x_0)] = O\{(nh)^{-1}\}.$$

Since $\bar{U}_{1n}$ is a nonnegative random variable, (10) follows from Markov's inequality. $\square$

Lemma 4. *Under conditions (A1)-(A4) and (A8),*

$$\frac{\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0) I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}}{\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0)} = p + O_p\{(nh)^{-1}\}.$$

Proof. Form the definition of $\widehat{q}_{1-p}(x_0)$, we have

$$p \leq \frac{\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0) I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}}{\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0)} \leq p + \frac{\max_{1 \leq j \leq n} p_j(x_0) K_h(X_j - x_0)}{\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0)}.$$

Because of Lemmas 1-2 and condition (A1), we have

$$\frac{\max_{1 \leq j \leq n} p_j(x_0) K_h(X_j - x_0)}{\sum_{t=1}^{n} p_t(x_0) K_h(X_t - x_0)} = O_p\{(nh)^{-1}\}.$$

Therefore, we obtain the desired result. $\square$

Proof of Theorem 1. By Lemmas 2-4, we have

$$
\widehat{m}_p(x_0) - m_p(x_0)
$$
$$
= \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\} I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}}{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0) I\{Y_t \geq \widehat{q}_{1-p}(x_0)\}} - \{m_p(x_0) - q_{1-p}(x_0)\}
$$
$$
= \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Y_t - q_{1-p}(x_0)\} I\{Y_t \geq q_{1-p}(x_0)\}}{p \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)}[1 + O_p\{(nh)^{-1}\}]
$$
$$
- \{m_p(x_0) - q_{1-p}(x_0)\} + O_p\{(nh)^{-1}\}.
$$

Let $Z_t = \{Y_t - q_{1-p}(x_0)\} I\{Y_t \geq q_{1-p}(x_0)\}$ and $\mu(x) = \mathrm{E}[Z_t | X_t = x]$. Then, we have

$$
\frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0) Z_t}{p \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)} - \{m_p(x_0) - q_{1-p}(x_0)\}
$$
$$
= \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Z_t - \mu(x_0)\}}{p \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)}
$$
$$
= \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Z_t - \mu(X_t)\}}{p \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)}
$$
$$
+ \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{\mu(X_t) - \mu(x_0)\}}{p \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)}.
$$

The Taylor expansion yields that

$$
\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{\mu(X_t) - \mu(x_0)\}
$$
$$
= \frac{h^2 p B_p(x_0)}{2} \sum_{t=1}^n p_t(x_0) \left(\frac{X_t - x_0}{h}\right)^2 K_h(X_t - x_0)\{1 + o(1)\}
$$
$$
= \frac{h^2 \mu_2(K) p g(x_0)}{2} B_p(x_0) + o_p(h^2),
$$

where we have used the property (3) in the first equality and Lemma 2 in the second equality. Therefore, we obtain the expression

$$
(nh)^{1/2}\left\{\widehat{m}_p(x_0) - m_p(x_0) - \frac{h^2 \mu_2(K)}{2} B_p(x_0) + o_p(h^2)\right\}
$$
$$
= (nh)^{1/2}\frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Z_t - \mu(X_t)\}}{p \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)}[1 + O_p\{(nh)^{-1}\}] + O_p\{(nh)^{-1/2}\}
$$
$$
= (nh)^{1/2}\left[\frac{1}{pg(x_0)} \sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Z_t - \mu(X_t)\}\right]\{1 + o_p(1)\} + O_p\{(nh)^{-1/2}\},
$$

where the last equality follows from Lemma 2. By the Taylor expansion,

$$
\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)\{Z_t - \mu(X_t)\}
$$
$$
= \frac{1}{n} \sum_{t=1}^n b_t(x_0) K_h(X_t - x_0)\{Z_t - \mu(X_t)\} + (\widehat{\lambda} - \lambda^*) R_n, \quad (11)
$$

16

where

$$|R_n| \le \frac{O_p(1)}{n} \sum_{t=1}^{n} K_h(X_t - x_0)|Z_t - \mu(X_t)| = O_p(1).$$

Since $\widehat{\lambda} - \lambda^* = O_p(h^3)$ by Lemma 1, the second term of the right hand side of (11) is of order $O_p(h^3)$. Since $(nh)^{1/2}h^3 = (nh^5)^{1/2}h = O(h)$, it remains to show that

$$(nh)^{1/2}\left[\frac{1}{n}\sum_{t=1}^{n} b_t(x_0)K_h(X_t - x_0)\{Z_t - \mu(X_t)\}\right] \xrightarrow{d} N\left\{0, p^2 g(x_0)\mu_0(K^2)\sigma_p^2(x_0)\right\}.$$

This result can be proved by following the same line as that used in the proof of Theorem 3 in Masry and Fan (1997). We note that the expression of $\sigma^2(x) := \mathrm{E}[\{Z_t - \mu(X_t)\}^2|X_t = x]$ is derived as follows:

$$\begin{aligned}
\sigma^2(x) &= \mathrm{E}[\{Z_t - \mu(X_t)\}^2|X_t = x] \\
&= \mathrm{E}[Z_t^2|X_t = x] - (\mathrm{E}[Z_t|X_t = x])^2 \\
&= \left[l_2(q_{1-p}(x_0)|x) + \{1 - F(q_{1-p}(x_0)|x)\}q_{1-p}^2(x_0) - 2l_1(q_{1-p}(x_0)|x)q_{1-p}(x_0)\right] \\
&\quad - \left[l_1(q_{1-p}(x_0)|x) - \{1 - F(q_{1-p}(x_0)|x)\}q_{1-p}(x_0)\right]^2 \\
&= l_2(q_{1-p}(x_0)|x) - l_1^2(q_{1-p}(x_0)|x) \\
&\quad + F(q_{1-p}(x_0)|x)q_{1-p}(x_0)\left[\{1 - F(q_{1-p}(x_0)|x)\}q_{1-p}(x_0) - 2l_1(q_{1-p}(x_0)|x)\right].
\end{aligned}$$

Thus, the map $x \mapsto \sigma^2(x)$ is continuous at x_0, as required in Masry and Fan (1997). It is easy to see that $\sigma_p^2(x_0) = p^{-2}\sigma^2(x_0)$. However, we can not use the truncation argument to check the Lindeberg condition (3.10) in Masry and Fan (1997) under the present conditions. Alternatively, we employ Theorem 4.1 in Shao and Yu (1996) and condition (A6) to check the Lindeberg condition. Since the argument is similar to the proof of Theorem 1 in Cai (2001), we omit the detail for brevity. $\square$

Remark 1. The proof of Theorem 1 in Cai (2002) uses the following argument: Put $J_1 = \sum_{t=1}^{n}\{I(Y_t \le y) - F(y|x_0)\}p_t(x_0)K_h(X_t - x_0)$ and $J_2 = n^{-1}\sum_{t=1}^{n}\{I(Y_t \le y) - F(y|x_0)\}b_t(x_0)K_h(X_t - x_0)$. He claimed that since $p_t(x_0) = n^{-1}b_t(x_0)\{1 + o_p(1)\}$ uniformly over t, J_1 is written as $J_1 = \{1 + o_p(1)\}J_2$; see also the proofs of Theorem 1 in Cai (2001) and Theorem 1 in Cai and Wang (2008). However, this argument seems not rigorous since $|J_1 - J_2|$ is bounded by $o_p(1) \times n^{-1}\sum_{t=1}^{n}|I(Y_t \le y) - F(y|x_0)|K_h(X_t - x_0)$ and is not bounded by $o_p(1) \times |J_2|$. Nevertheless, the conclusion of the theorem still holds. To show asymptotic normality of J_1, expand J_1 as $J_1 = J_2 + R_n$ like (11) and evaluate the order of the reminder term R_n. Then, we may prove the conclusion of the theorem under the additional condition that $nh^3 \to \infty$ and $h = O(n^{-1/5})$. Theorem 3 of Cai (2002) is a direct consequence of Theorem 1 of the same paper and so still holds under the additional condition. On the other hand, such a modification may not be applied to the proof of Theorem 2 in Cai (2002), which states the boundary behavior of the WNW estimator of the conditional distribution function. So we do not pursue the boundary behavior of the proposed estimator in this paper.

A.2 Derivation of (6)

In this section, we derive (6). To do this, we need some additional regularity conditions stated as follows:

(B1) The kernel function $W(\cdot)$ is a symmetric, bounded and compactly supported density.

(B2) The process $\{(Y_t, X_t), t = 1, 2, \dots \}$ is stationary and α-mixing with

$$\sum_{j=1}^{\infty} \{\alpha(j)\}^{(1-2/\delta)/2} < \infty$$

for some $\delta > 2$.

(B3) The map $x \mapsto \mathrm{E}[|Y|^{\delta}|X = x]$ is bounded in a neighborhood of x_0, where δ is defined in (B2).

(B4) For each $j \geq 1$ and each (x_1, x_{1+j}) in a neighborhood of (x_0, x_0), the map $(u_1, u_{1+j}) \mapsto$ $\mathrm{E}[\{Y_1 - q_{1-p}(x_0)\}I(Y_1 \geq u_1)\{Y_{1+j} - q_{1-p}(x_0)\}I(Y_{1+j} \geq u_{1+j})|X_1 = x_1, X_{1+j} = x_{1+j}]$ is twice continuously differentiable in a neighborhood of $(q_{1-p}(x_0), q_{1-p}(x_0))$; the partial derivatives of the map are bounded by some constant not depending on j in the neighborhood of $(q_{1-p}(x_0), q_{1-p}(x_0))$ uniformly over (x_1, x_{1+j}) in the neighborhood of (x_0, x_0).

The substantial additional condition is (B4). The other conditions are more or less standard. Condition (B4) is high level; however, we put it for a technical convenience. We note that if the process is independent, condition (B4) is removed. The proof of Proposition 1 below basically follows the proof of Lemma 2 in Cai and Roussas (1998). A part of the proof overlaps the proof of Lemma 4 in Cai and Wang (2008), which is found in Cai and Wang (2006).

Proposition 1. *Suppose that conditions (A1), (A3), (A5), (A7)-(b), (A8) and (B1)-(B4) are satisfied. If $h, h_0 \to 0$ as $n \to \infty$, then (6) holds.*

Proof. Let $c_t(x_0) = b_t(x_0)K_h(X_t - x_0)$ and $Z_t = Y_t - q_{1-p}(x_0)$. By stationarity,

$$p^2 g^2(x_0) \mathrm{var}(I_2) = \frac{1}{n} \mathrm{var}\{c_1(x_0)Z_1 G_{h_0}(Z_1)\}$$

$$+ \frac{1}{n} \sum_{j=1}^{n-1} (1 - j/n) \mathrm{cov}\{c_1(x_0)Z_1 G_{h_0}(Z_1), c_{1+j}(x_0)Z_{1+j}G_{h_0}(Z_{1+j})\}.$$

Because of conditions (A1), (A7)-(b), (A8), (B1) and (B3), the Taylor expansion yields that

$$\mathrm{E}[c_1(x_0)Z_1 G_{h_0}(Z_1)] = \mathrm{E}[c_1(x_0)Z_1 I(Z_1 \geq 0)] + O(h_0^2),$$
$$\mathrm{E}[\{c_1(x_0)Z_1 G_{h_0}(Z_1)\}^2] = \mathrm{E}[\{c_1(x_0)Z_1\}^2 I(Z_1 \geq 0)] + O(h^{-1}h_0^2),$$

which implies that

$$\mathrm{var}\{c_1(x_0)Z_1 G_{h_0}(Z_1)\} = \mathrm{var}\{c_1(x_0)Z_1 I(Z_1 \geq 0)\} + o(h^{-1}h_0)$$

See also the proof of Lemma 4 in Cai and Wang (2006).

Define

$$
\begin{aligned}
a_j = \mathrm{cov}\{&c_1(x_0)Z_1 G_{h_0}(Z_1), c_{1+j}(x_0)Z_{1+j}G_{h_0}(Z_{1+j})\} \\
&- \mathrm{cov}\{c_1(x_0)Z_1 I(Z_1 \geq 0), c_j(x_0)Z_{1+j}I(Z_{1+j} \geq 0)\}, \ j \geq 1.
\end{aligned}
$$

It suffices to show

$$\sum_{j=1}^{n-1} |a_j| = o(h^{-1}h_0).$$

Because of conditions (B1) and (B4), the Taylor expansion yields that

$$
\begin{aligned}
\mathrm{E}[Z_1 G_{h_0}&(Z_1)Z_{1+j}G_{h_0}(Z_{1+j})|X_1 = x_1, X_{1+j} = x_{1+j}] \\
&= \mathrm{E}[Z_1 I(Z_1 \geq 0)Z_{1+j}I(Z_{1+j} \geq 0)|X_1 = x_1, X_{1+j} = x_{1+j}] + O(h_0^2),
\end{aligned}
$$

uniformly over (x_1, x_{1+j}) in a neighborhood of (x_0, x_0) and $j \geq 1$, which implies that

$$|a_j| \leq C_1 h_0^2, \tag{12}$$

for some constant $C_1 > 0$ not depending on j, where we have used conditions (A1), (A3) and (A5). From (A1), (A8) and (B3), Davydov's inequality (Fan and Yao, 2005, Proposition 2.5) implies that

$$
\begin{aligned}
|a_j| &\leq 16\{\alpha(j)\}^{1-2/\delta}(\mathrm{E}[|c_1(x_0)Z_1|^\delta])^{2/\delta} \\
&\leq C_2 h^{2/\delta-2}\{\alpha(j)\}^{1-2/\delta}, \tag{13}
\end{aligned}
$$

for some constant $C_2 > 0$ not depending on j. See the proof of Theorem 2 in Masry and Fan (1997). Combining (12) and (13), we have

$$
\begin{aligned}
|a_j| &= |a_j|^{1/2}|a_j|^{1/2} \\
&\leq C_3 h^{-1+1/\delta}h_0\{\alpha(j)\}^{(1-2/\delta)/2},
\end{aligned}
$$

where $C_3 = (C_1 C_2)^{1/2}$, which implies that

$$
\begin{aligned}
\sum_{j=1}^{n-1} |a_j| &\leq C_3 h^{-1+1/\delta}h_0 \sum_{j=1}^{\infty}\{\alpha(j)\}^{(1-2/\delta)/2} \\
&= o(h^{-1}h_0),
\end{aligned}
$$

where we have used condition (B2). Therefore, the proof is complete. $\qquad \square$

Acknowledgments

19

References

Acerbi, C. and Tasche, D. 2002. On coherence of expected shortfall. *Journal of Banking and Finance* 26: 1487-1503

Artzner, C., Delbaen, F., Eber, J.M. and Heath, D. 1997. Thinking coherently. *RISK* 10: 68-71.

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

Azzalini, A. 1981. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* 68: 326-328.

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. and Roussas, G.G. 1998. Efficient estimation of a distribution function under quadrant dependence. *Scandinavian Journal of Statistics* 25: 211-224.

Cai, Z. and Wang, X. 2006. "Nonparametric methods for estimating conditional VaR and expected shortfall." Wise Working Paper Series WISEWP0604, Wang Yanan Institute for Economic Studies, Xiamen University.

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

Chen, S.X. 2008. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* 6: 87-107.

Chen, S.X. and Tang, C.Y. 2005. Nonparametric inference of value at risk for dependent financial returns. *Journal of Financial Econometrics* 3: 227-255.

Doornik, J.A. 2002. *Object-Oriented Matrix Programming Using Ox*, 3rd edition. Timberlake Consultants Press, London.

Fan, J. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87: 998-1004.

Fan, J. and Gijbels, I. 1996. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.

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

Fan, J. and Yao, Q. 2005. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.

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

Laurent S. and Peters, J.-P. 2006. *G@RCH 4.2, Estimating and Forecasting ARCH Models.* Timberlake Consultants Press, London.

Masry, E. and Fan, J. 1997. Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24: 165-179.

McNeil, A.J., Frey, R. and Embrecht, P. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, Princeton.

Peracchi, F. and Tanase, A.V. 2008. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry* 24: 471-493.

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.

Shao, Q. and Yu, A. 1996. Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* 24: 2098-2127.

Yu, K. and Jones, M.C. 1998. Local linear quantile estimation. *Journal of the American Statistical Association* 93: 228-237.