WISE WORKING PAPER SERIES

WISEWP0604

Nonparametric Methods for Estimating Conditional VaR and Expected Shortfall

Zongwu Cai And Xian Wang

July 30, 2006

厦门大学王亚南经济研究院
The Wang Yanan Institute for Studies in Economics, Xiamen University

Nonparametric Methods for Estimating Conditional VaR and Expected Shortfall*

Zongwu Cai[a,b,c] and Xian Wang[a]

[a]Department of Mathematics & Statistics and Department of Economics,
University of North Carolina at Charlotte, Charlotte, NC 28223, USA

[b]Wang Yanan Institute for Studies in Economics, Xiamen University, China.

[c]College of Economics and Management, Shanghai Jiaotong University, China

This version: July 30, 2006

In this article we propose a new nonparametric estimation method to estimate the conditional value-at-risk and expected shortfall functions based on the weighted double kernel local linear estimator of the conditional distribution function. The conditional value-at-risk is estimated by inverting the estimated conditional distribution function. The nonparametric estimator of the conditional expected shortfall is constructed by a plugging-in method. First, we establish the asymptotic normality and weak consistency of the weighted double kernel local linear estimator of the conditional distribution for time series data at both boundary and interior points. Also, we also show that the weighted double kernel local linear conditional distribution estimator not only preserves the bias, variance, and more importantly, automatic good boundary behavior properties of the double kernel local linear estimator and the weighted Nadaraya-Watson estimator, but also has the additional advantages of being always a distribution itself, continuity, and differentiability. Secondly, we show that the proposed weighted double kernel local linear estimators for both the conditional value-at-risk and expected shortfall are weakly consistent and normally distributed under the time series context at both boundary and interior points. Moreover, an automatic 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 performance of the proposed estimators.

JEL classification: C14; D81; G10; G22; G31

Keywords: Bandwidth selection; Boundary effects; Coherent risk measurements; Empirical likelihood; Expected shortfall; Local liner estimation; Nonparametric smoothing; Quantile regression; Time series; Value-at-risk; Weighted double kernel.

---

*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. *Correspondence author*: Tel. (704) 687-2650. Fax: (704) 687-6415. *E-mail address*: zcai@uncc.edu (Z. Cai).

1 Introduction

The value-at-risk (hereafter, VaR) and expected shortfall (ES) have become two popular measures on market risk associated with an asset or a portfolio of assets during the recent decade. In particular, VaR has been chosen by the Basle Committee on Banking Supervision as the benchmark of risk measures for capital requirements and both of them have been used by financial institutions for asset managements and minimization of risk as well as have been developed rapidly as analytic tools to assess riskiness of trading activities. See, to name just a few, Morgan (1996), Duffie and Pan (1997), Jorion (2001, 2003), and Duffie and Singleton (2003) for the financial background, statistical inferences, and various applications. 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 quantile of the loss distribution (e.g., VaR). It is well known from Artzner, Delbaen, Eber and Heath (1999) that ES is a coherent risk measure such as it satisfies the four axioms: homogeneity (increasing the size of a portfolio by a factor should scale its risk measure by the same factor), monotonicity (a portfolio must have greater risk if it has systematically lower values than another), risk-free condition or translation invariance (adding some amount of cash to a portfolio should reduce its risk by the same amount), and subadditivity (the risk of a portfolio must be less than the sum of separate risks or merging portfolios cannot increase risk). VaR satisfies homogeneity, monotonicity, and risk-free condition but is not sub-additive. See Artzner, *et al.* (1999) for details.

As advocated by Artzner, *et al.* (1999), ES is preferred 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 need to focus on the conditional distributions of profit and loss, which take full account of current information about the investment environment (macroeconomic and financial as well as political) in forecasting future market values, volatilities, and correlations. As pointed out by Duffie and Singleton (2003), not only are the prices of the underlying market indices changing randomly over time, the portfolio itself is changing, as are the volatilities of prices, the credit qualities of counterparties, and so on. On the other hand, one would expect the VaR to increase as the past returns become very negative, because one bad day makes the probability of the next somewhat greater. Similarly, very good days also increase the VaR, as would be the case for volatility models. Therefore, VaR could depend on the past returns in someway. 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 the conditional distributions and time series data. For more background, see Chernozhukov and Umanstev (2001), Cai (2002), Fan and Gu (2003), Engle and Manganelli (2004), Cai and Xu (2005), and Scaillet (2005), and references therein 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 among others for unconditional models. Also, most of studies in the literature and applications are limited to parametric models, such as all standard industry models like CreditRisk$^+$, CreditMetrics, CreditPortfolio View and the model proposed by the KMV corporation. 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.

The main focus of this paper is on the conditional value-at-risk (CVaR) and conditional expected shortfall (CES) and is to propose a new nonparametric estimation procedure to estimate CVaR and CES functions where the conditional information is allowed to contain economic and market (exogenous) variables and the past observed returns. Parametric models for CVaR and CES can be most efficient if the underlying functions are correctly specified. See Chernozhukov and Umanstev (2001) for a polynomial type regression model and Engle and Manganelli (2004) for a GARCH type parametric model for CVaR based on regression quantile. However, a misspecification may cause serious bias and model constraints may distort the underlying distributions. A nonparametric modeling is appealing in several aspects. One of the advantages for nonparametric modeling is that little or no restrictive prior information on functionals is needed. Further, it may provide useful insight for further parametric fitting.

This paper proposes a new nonparametric approach to estimate CVaR and CES. In essence, our estimator for CVaR is based on inverting a newly proposed estimator of the conditional distribution function for time series data and the estimator for CES is by a plugging-in method based on plugging in the estimated conditional probability density function and the estimated CVaR function. Note that they are analogous to the estimators studied by Scaillet (2005) by using the Nadaraya-Watson (NW) type double kernel (smoothing in both the y and x directions) estimation, and Cai (2002) by utilizing the weighted Nadaraya-Watson (WNW) kernel type technique to avoid the so-called boundary effects as well as Yu and Jones (1998) by employing the double kernel local linear method. More precisely, our newly proposed estimator combines the 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.

The paper consists of two themes. The first part is devoted to establishing the asymptotic properties for the WDKLL estimators of the conditional probability density function (PDF) and cumulative distribution function (CDF) for the α-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 of Yu and Jones (1998) and the WNW estimator of Cai (2002). It is also shown that the WDKLL estimators have desired sampling properties at both boundary and interior points of the support of the design density, which seems to be seminal. Secondly, we derive the WDKLL estimator of CVaR by inverting the WDKLL conditional distribution estimator and the WDKLL estimator of CES by plugging in the WDKLL PDF and CVaR estimators. We show that the WDKLL CVaR estimator exists always due to the WDKLL CDF being a distribution function itself and that it inherits all better properties from the WDKLL CDF estimator; that is, the WDKLL CDF is a CDF and differentiable and it possess the asymptotic properties such as design adaption, avoiding boundary effects, and mathematical efficiency.

Note that CVaR defined here is essentially the conditional quantile or quantile regression of Koenker and Bassett (1978), based on the conditional distribution, rather than CVaR defined in some risk management literature (see, e.g., Rockafellar and Uryasev, 2000; Jorion, 2001, 2003) which is what we call ES here. Also, note that the ES here is called TailVaR in Artzner, *et al.* (1999). Moreover, as aforementioned, CVaR can be regarded as a special case of quantile regression. See Cai and Xu (2005) for the state-of-the-art about current research on nonparametric quantile regression, including CVaR. Further, note that both ES and CES have been known for decades among actuary sciences and they are very popular in insurance industry. Indeed, they have been used to assess risk on a portfolio of potential claims, and to design reinsurance treaties. See the book by Embrechts, Kluppelberg, and Mikosch (1997) for the excellent review on this subject and the papers by McNeil (1997), Hürlimann (2003), and Scaillet (2005). Finally, ES or CES is also closely related to other applied fields such as the mean residual life function in reliability and the biometric function in biostatistics. See Oakes and Dasu (1990) and Cai and Qian (2000) and references therein.

The plan of the paper is as follows. Section 2 provides a blueprint for the basic notations and concepts. In Section 3, we present the detailed motivations and formulations for the new nonparametric estimation procedures for estimating the conditional PDF, CDF, VaR and ES. We establish the asymptotic properties of these nonparametric estimators at both boundary and interior points with a comparison in Section 4. Together with a convenient and efficient data-driven method for selecting the bandwidth based on the nonparametric Akaike information criterion (AIC), Monte Carol simulation studies and empirical applications on several stock index returns are presented in Section 5. Finally, the derivations of the theorems are given in Section 6 with some lemmas and Appendix contains the technical proofs of certain lemmas needed in the proofs of the theorems in Section 6.

2 Framework

Assume that the observed data $\{(X_t, Y_t); 1 \le t \le n\}$, $X_t \in \Re^d$, 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 and also it can be a vector. But, for the expositional purpose, we only consider the case that X_t is a scalar ($d = 1$). Note that the proposed methodologies and their theory for the univariate case ($d = 1$) continue to hold for multivariate situations ($d > 1$). Extension to the case $d > 1$ involves no fundamentally new ideas. Note that models with large d are often not practically useful due to "curse of dimensionality".

We now turn to considering the nonparametric estimation of the conditional expected shortfall $\mu_p(x)$, which is defined as $\mu_p(x) = E[Y_t \,|\, Y_t \ge \nu_p(x), X_t = x]$, where $\nu_p(x)$ is the conditional value-at-risk, which is defined as the solution of $P(Y_t \ge \nu_p(x) \,|\, X_t = x) = S(\nu_p(x) \,|\, x) = p$ or expressed as $\nu_p(x) = S^{-1}(p \,|\, x)$, where $S(y \,|\, x)$ is the conditional survival function of Y_t given $X_t = x$; $S(y \,|\, x) = 1 - F(y \,|\, x)$, and $F(y \,|\, x)$ is the conditional cumulative distribution function. It is easy to see that

$$\mu_p(x) = \int_{\nu_p(x)}^{\infty} y\, f(y \,|\, x)\, dy/p,$$

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

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

where $\widehat{\nu}_p(x)$ is a nonparametric estimation of $\nu_p(x)$ and $\widehat{f}(y \,|\, x)$ is a nonparametric estimation of $f(y \,|\, x)$.

Note that Scaillet (2005) used the NW type double kernel method to estimate $f(y \,|\, x)$ first, due to Roussas (1969), denoted by $\widetilde{f}(y \,|\, x)$, and then estimated $\nu_p(x)$ by inverting the estimated conditional survival function, denoted by $\widetilde{\nu}_p(x)$, and finally estimated $\mu_p(x)$ by plugging $\widetilde{f}(y \,|\, x)$ and $\widetilde{\nu}_p(x)$ into (1), denoted by $\widetilde{\mu}_p(x)$, where $\widetilde{\nu}_p(x) = \widetilde{S}^{-1}(y \,|\, x)$ and $\widetilde{S}(y \,|\, x) = \int_y^{\infty} \widetilde{f}(u \,|\, x) du$. But, it is well documented (see, e.g., Fan and Gijbels, 1996) that the NW kernel type procedures have serious drawbacks: the asymptotic bias involves the design density so that they can not be adaptive, and they have boundary effects so that they require boundary modifications. In particular, the boundary effect might cause a big problem for estimating CVaR $\nu_p(x)$ since it is only concerned with the tail probability. The question now is how to estimate CVaR $\nu_p(x)$ and the conditional density function $f(y \,|\, x)$ efficiently and optimally so that we can estimate $\mu_p(x)$ well. Therefore, we need to address this issue in the next section.

3 Nonparametric Estimating Procedures

We start with the nonparametric estimators for the conditional density function and its distribution function first and then turn to discussing the nonparametric estimators for the conditional VaR and ES functions.

3.1 Estimation of Conditional PDF and CDF

There are several methods available for estimating CVaR $\nu_p(x)$ and conditional density $f(y \mid x)$ in the literature, such as kernel and nearest-neighbor. To name just a few, see Lejeune and Sarda (1988), Troung (1989), Samanta (1989), and Chaudhuri (1991) for iid errors, Roussas (1969) and Roussas (1991) for Markovian processes, and Troung and Stone (1992) and Boente and Fraiman (1995) for mixing sequences. To attenuate these drawbacks of the kernel type estimators mentioned in Section 2, recently, some new methods of estimating conditional quantiles have been proposed. The first one, a more direct approach, by using the "check" function such as the robustified local linear smoother, was provided by Fan, Hu, and Troung (1994) and further extended by Yu and Jones (1997, 1998) for iid data. A more general nonparametric setting was explored by Cai and Xu (2005) for time series data. This modeling idea was initialed by Koenker and Bassett (1978) for linear regression quantiles and Fan, Hu, and Troung (1994) for nonparametric models. See Cai and Xu (2005) and references therein for more discussions on models and applications. An alternative procedure is first to estimate the conditional distribution function by using double kernel local linear technique of Fan, Yao, and Tong (1996) and then to invert the conditional distribution estimator to produce an estimator of a conditional quantile or CVaR. Yu and Jones (1997, 1998) compared these two methods and suggested that the double kernel local linear would be better.

To make a connection between the conditional density (distribution) function and nonparametric regression problem, it is noted by the standard kernel estimation theory (see, e.g., Fan and Gijbles, 1996) that for a given symmetric density function $K(\cdot)$,

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

where $K_{h_0}(u) = K(u/h_0)/h_0$, $\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u) du$, $f^{2,0}(y \mid x) = \partial^2/\partial y^2 f(y \mid x)$, and $\approx$ denotes an approximation by ignoring the higher terms. Note that $Y_t^*(y) = K_{h_0}(y - Y_t)$ can be regarded as an initial estimate of $f(y \mid x)$ smoothing in the y direction. Also, note that this approximation ignores the higher order terms $O(h_0^j)$ for $j \geq 2$, since they are negligible if $h_0 = o(h)$, where h is the bandwidth used in smoothing in the x direction (see (3) below). Therefore, the smoothing in the y direction is not important in the context of this subject so that intuitively, it should be under-smoothed. Thus, the left hand side of (2) can be regraded as a nonparametric regression of the observed variable $Y_t^*(y)$ versus X_t and the local linear (or polynomial) fitting scheme of Fan and Gijbles (1996) can be applied to here. This leads us to consider the following 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 $h = h(n) > 0$ is the bandwidth satisfying $h \to 0$ and $n\,h \to \infty$ as $n \to \infty$, which controls the amount of smoothing used in the estimation. Note that (3) involves two kernels $K(\cdot)$ and $W(\cdot)$. This is the reason of calling "double kernel". Minimizing the locally weighted least squares with respect to a and b, we obtain the locally weighted least squares estimator of $f(y\,|\,x)$, denoted by $\widehat{f}(y\,|\,x)$, which is $\widehat{a}$. From Fan and Gijbels (1996) or Fan, Yao and Tong (1996), $\widehat{f}(y\,|\,x)$ can be re-expressed as a linear estimator form as

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

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

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

Clearly, $\{W_{ll,t}(x, h)\}$ satisfy the so-called discrete moments conditions as follows: for $0 \leq j \leq 1$,

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

based on the least squares theory; see (3.12) of Fan and Gijbels (1996, p.63). Note that the estimator $\widehat{f}_{ll}(y\,|\,x)$ can range outside $[0,\ \infty)$. The double kernel local linear estimator of

$F(y \mid x)$ is constructed (see (8) of Yu and Jones (1998)) by integrating $\widehat{f}_{ll}(y \mid x)$

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

where $G(\cdot)$ is the distribution function of $K(\cdot)$ and $G_{h_0}(u) = G(u/h_0)$. Clearly, $\widehat{F}_{ll}(y \mid x)$ is continuous and differentiable with respect to y with $\widehat{F}_{ll}(-\infty \mid x) = 0$ and $\widehat{F}_{ll}(\infty \mid x) = 1$. Note that the differentiability of the estimated distribution function can make the asymptotic analysis much easier for CVaR and CES (see later).

Although Yu and Jones (1998) showed that the double kernel local linear estimator has some attractive properties such as no boundary effects, design adaptation, and mathematical efficiency (see, e.g., Fan and Gijbels, 1996), it has the disadvantage of producing conditional distribution function estimators that are not constrained either to lie between zero and one or to be monotone increasing, which is not good for estimating CVaR if the inverting method is used. In both these respects, the NW method is superior, despite its rather large bias and boundary effects. The properties of positivity and monotonicity are particularly advantageous if the method of inverting conditional distribution estimator is applied to produce an estimator of the conditional quantile or CVaR. To overcome these difficulties, Hall, Wolff, and Yao (1999) and Cai (2002) proposed the WNW estimator based on an empirical likelihood principle, which is designed to possess the superior properties of local linear methods such as bias reduction and no boundary effects, and to preserve the property that the NW estimator is always a distribution function, although it might require more computational efforts since it requires estimating and optimizing additional weights aimed at the bias correction. Cai (2002) discussed the asymptotic properties of the WNW estimator at both interior and boundary points for the mixing time series under some regularity assumptions and showed that the WNW estimator has a better performance than other competitors. See Cai (2002) for details.

The WNW estimator of the conditional distribution $F(y \mid x)$ of Y_t given $X_t = x$ is defined by

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

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

$$W_{c,t}(x,h) = \frac{p_t(x)\, W_h(x - X_t)}{\sum_{t=1}^{n} p_t(x)\, W_h(x - X_t)}, \tag{6}$$

and $\{p_t(\mathbf{x})\}$ is chosen to be $p_t(x) = n^{-1}\,\{1 + \lambda\,(X_t - x)\, W_h(x - X_t)\}^{-1} \ge 0$ with λ, a function of data and x, 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 $\sum_{t=1}^{n} p_t(x) = 1$ and the discrete moments conditions in (4); that is,

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

for $0 \le j \le 1$. Also, see Cai (2002) for details on this aspect. In implementation, Cai (2002) recommended using the Newton-Raphson scheme to find the root of equation $L_n'(\lambda) = 0$. Note that $0 \le \widehat{F}_{c1}(y\,|\,x) \le 1$ and it is monotone in y. But $\widehat{F}_{c1}(y\,|\,x)$ is not continuous in y and of course, not differentiable in y either. Note that under regression setting, Cai (2001) provided a comparison of the local linear estimator and the WNW estimator and discussed the asymptotic minimax efficiency of the WNW estimator.

To accommodate all nice properties (monotonicity, continuity, differentiability, and lying between zero and one) and the attractive asymptotic properties (design adaption, avoiding boundary effects, and mathematical efficiency, see Cai (2002) for detailed discussions) of both estimators $\widehat{F}_{ll}(y\,|\,x)$ and $\widehat{F}_{c1}(y\,|\,x)$ under a unified framework, we propose the following nonparametric estimators for the conditional density function $f(y\,|\,x)$ and its conditional distribution function $F(y\,|\,x)$, termed as *weighted double kernel local linear estimation*,

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

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

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

Note that if $p_t(x)$ in (6) is a constant for all t, or $\lambda = 0$, then $\widehat{f}_c(y\,|\,x)$ 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, $\widehat{f}_c(y\,|\,x)$ is a probability density function so that $\widehat{F}_c(y\,|\,x)$ is a cumulative distribution function (monotone, $0 \le \widehat{F}_c(y\,|\,x) \le 1$, $\widehat{F}_c(-\infty|\,x) = 0$, and $\widehat{F}_c(\infty\,|\,x) = 1$). Also, $\widehat{F}_c(y\,|\,x)$ is continuous and differentiable in y. Further, as expected, it will be shown that like $\widehat{F}_{c1}(y\,|\,x)$, $\widehat{F}_c(y\,|\,x)$ has the attractive properties such as no boundary effects, design adaptation, and mathematical efficiency.

3.2 Estimation of Conditional VaR and ES

We now are ready to formulate the nonparametric estimators for $\nu_p(x)$ and $\mu_p(x)$. To this end, from (8), $\nu_p(x)$ is estimated by inverting the estimated conditional survival distribution $\widehat{S}_c(y\,|\,x) = 1 - \widehat{F}_c(y\,|\,x)$, denoted by $\widehat{\nu}_p(x)$ and defined as $\widehat{\nu}_p(x) = \widehat{S}_c^{-1}(p\,|\,x)$. Note that $\widehat{\nu}_p(x)$ always exists since $\widehat{S}_c(p\,|\,x)$ is a survival function itself. Plugging-in $\widehat{\nu}_p(x)$ and $\widehat{f}_c(y\,|\,x)$ into (1), we obtain the nonparametric estimation of $\mu_p(x)$,

$$\begin{aligned}
\widehat{\mu}_p(x) &= p^{-1} \int_{\widehat{\nu}_p(x)}^{\infty} y\, \widehat{f}_c(y\,|\,x)\, dy = p^{-1} \sum_{t=1}^{n} W_{c,t}(x,h) \int_{\widehat{\nu}_p(x)}^{\infty} y\, K_{h_0}(y - Y_t)dy \\
&= p^{-1} \sum_{t=1}^{n} W_{c,t}(x,h) \left[Y_t\, \bar{G}_{h_0}(\widehat{\nu}_p(x) - Y_t) + h_0\, G_{1,h_0}(\widehat{\nu}_p(x) - Y_t) \right],
\end{aligned} \tag{9}$$

where $\bar{G}(u) = 1 - G(u)$, $G_{1,h_0}(u) = G_1(u/h_0)$, and $G_1(u) = \int_u^{\infty} v\, K(v)dv$. We next discuss the sampling properties of the proposed estimators.

4 Distribution Theory

4.1 Assumptions

The errors in a time series model are usually assumed to follow certain linear time series models such as an autoregressive and moving average (ARMA) process. Here we consider a more general structure – the α-mixing process, which includes many linear and nonlinear time series models as special cases. The asymptotic results here are derived under the α-mixing assumption, which is popular and common for controlling dependence in dynamic econometrics and finance with exogenous or lagged variables; see Pötscher and Prucha (1997) and Cai (2002, 2003) for more details. Finally, Carrasco and Chen (2002) showed that some generalized autoregressive conditional heteroscedastic (GARCH) and stochastic volatility models are strong mixing under some mild conditions. See Chen and Tang (2005) for more examples.

Before we proceed with the asymptotic properties of the proposed nonparametric estimators, we first list all assumptions needed for the asymptotic theory, although some of them might not be the weakest possible. Note that proofs of the asymptotic results presented in this section may be found in Section 6 with some lemmas and their detailed proofs relegated to Appendix. First, we introduce some notation. Let $\alpha(K) = \int_{-\infty}^{\infty} u\, K(u)\, \bar{G}(u) du$ and $\mu_j(W) = \int_{-\infty}^{\infty} u^j\, W(u) du$. Also, for any $j \geq 0$, write

$$l_j(u \,|\, v) = E[Y_t^j\, I(Y_t \geq u) \,|\, X_t = v] = \int_u^{\infty} y^j\, f(y \,|\, v) dy, \qquad l_j^{a,b}(u \,|\, v) = \frac{\partial^{ab}}{\partial u^a \partial v^b}\, l_j(u \,|\, v),$$

and $l_j^{a,b}(\nu_p(x) \,|\, x) = l_j^{a,b}(u \,|\, v)\Big|_{u=\nu_p(x), v=x}$. Clearly, $l_0(u \,|\, v) = S(u \,|\, v)$ and $l_1(\nu_p(x) \,|\, x) = p\, \mu_p(x)$. Finally, $l_j^{1,0}(u \,|\, v) = -u^j\, f(u \,|\, v)$ and $l_j^{2,0}(u \,|\, v) = -[u^j\, f^{1,0}(u \,|\, v) + j\, u^{j-1}\, f(u \,|\, v)]$. We now list the following regularity conditions.

Assumption A:

A1. For fixed y and x, $0 < F(y \,|\, x) < 1$, $g(x) > 0$, the marginal density of X_t, and is continuous at x, and $F(y\,|\,x)$ has continuous second order derivative with respect to both x and y.

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

A3. $h \to 0$ and $n\,h \to \infty$, and $h_0 \to 0$ and $n\,h_0 \to \infty$, as $n \to \infty$.

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

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

A6. $n\,h^{1+2/\delta} \to \infty$.

A7. $h_0 = o(h)$.

Assumption B:

B1. Assume that $E\left(|Y_t|^{\delta}\,|\,X_t = u\right) \leq M_3 < \infty$ for some $\delta > 2$, in a neighborhood of x.

B2. Assume that $|g_{1,t}(y_1,\,y_2\,|x_1,\,x_2)| \leq M_1 < \infty$ for all $t \geq 2$, where $g_{1,t}(y_1,\,y_2\,|\,x_1,\,x_2)$ be the conditional density of Y_1 and Y_t given $X_1 = x_1$ and $X_t = x_2$.

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

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

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

Remark 1. Note that Assumptions A1 - A5 and B1 - B5 are used commonly in the literature of time series data (see, e.g., Masry and Fan, 1997, Cai, 2001). Note that α-mixing imposed in Assumption A5 is weaker than β-mixing in Hall, Wolff, and Yao (1999) and ρ-mixing in Fan, Yao, and Tong (1996). Because A6 is satisfied by the bandwidths of optimal size (i.e., $h \approx n^{-1/5}$) if $\delta > 1/2$, we do not concern ourselves with such refinements. Indeed, Assumptions A1 - A6 are also required in Cai (2002). Assumption A7 means that the initial step bandwidth should be chosen as small as possible so that 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., Masry and Fan, 1997), the purpose of Assumption B5 is to use the moment inequality. If $\alpha(t)$ 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 the higher moments involved, the faster decaying rate of $\alpha(\cdot)$ is required. Finally, Assumptions B1 - B5 are also imposed in Cai (2001).

4.2 Asymptotic Properties for $\widehat{f}_c(y\,|\,x)$ and $\widehat{S}_c(y\,|\,x)$

First, we investigate the asymptotic behavior of $\widehat{f}_c(y\,|\,x)$ and we have the following asymptotic normality for $\widehat{f}_c(y\,|\,x)$.

Theorem 1: *Under Assumptions A1 - A5 and B1 - B4 with h in A3, B3, and B4 replaced by $h_0\, h$, we have*

$$\sqrt{n\, h_0\, h}\ \left[\widehat{f}_c(y\,|\,x) - f(y\,|\,x) - B_f(y\,|\,x) + o_p(h^2 + h_0^2)\right] \to N\left\{0,\ \sigma_f^2(y\,|\,x)\right\},$$

where the asymptotic bias is

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

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 $\widehat{f}_c(y \mid x)$ in Theorem 1 are similar to those for $\widehat{f}_{ll}(y \mid x)$ in Fan, Yao, and Tong (1996) for the ρ-mixing sequence, which is stronger than α-mixing, but as mentioned early, $\widehat{f}_{ll}(y \mid x)$ is not always a probability density function. The asymptotic bias and variance are intuitively expected. The bias comes from the approximations in both x and y directions and the variance is from the local conditional variance in the density estimation setting, which is $f(y \mid x)$.

Next, we study the asymptotic behaviors for $\widehat{S}_c(y \mid x)$ at both interior and boundary points. Similar to Theorem 1 for $\widehat{f}_c(y \mid x)$, we have the following asymptotic normality for $\widehat{S}_c(y \mid x)$.

Theorem 2: *Under Assumptions A1 - A6, we have*

$$\sqrt{n\,h}\,\left[\widehat{S}_c(y \mid x) - S(y \mid x) - B_S(y \mid x) + o_p(h^2 + h_0^2)\right] \to N\left\{0,\, \sigma_S^2(y \mid x)\right\},$$

where the asymptotic bias is given by

$$B_S(y \mid x) = \frac{h^2}{2}\,\mu_2(W)\,S^{0,2}(y \mid x) - \frac{h_0^2}{2}\,\mu_2(K)\,f^{1,0}(y \mid x),$$

and the asymptotic variance is $\sigma_S^2(y \mid x) = \mu_0(W^2)\,S(y \mid x)\,[1 - S(y \mid x)]/g(x)$. *In particular, if Assumption A7 holds true, then,*

$$\sqrt{n\,h}\,\left[\widehat{S}_c(y \mid x) - S(y \mid x) - \frac{h^2}{2}\,\mu_2(W)\,S^{0,2}(y \mid x) + o_p(h^2)\right] \to N\left\{0,\, \sigma_S^2(y \mid x)\right\}.$$

Remark 3: Note that the asymptotic results for $\widehat{S}_c(y \mid x)$ in Theorem 2 are analogous to those for $\widehat{S}_{ll}(y \mid x) = 1 - \widehat{F}_{ll}(y \mid x)$ in Yu and Jones (1998) for iid data, but as mentioned previously, $\widehat{F}_{ll}(y \mid x)$ is not always a distribution function. A comparison of $B_s(y \mid x)$ with the asymptotic bias for $\widehat{S}_{c1}(y \mid x)$ (see Theorem 1 in Cai (2002)), it reveals that there is an extra term $\frac{h_0^2}{2}\, f^{1,0}(y \mid x)\,\mu_2(K)$ in the asymptotic bias expression $B_s(y \mid x)$ due to the vertical smoothing in the y direction. Also, there is an extra term in the asymptotic variance (see

(20)). These extra terms are carried over from the initial estimate but they can be ignored if the bandwidth at the initial step is taken to be a higher order than the bandwidth at the smoothing step.

Remark 4: It is important to examine the performance of $\widehat{S}_c(y\,|\,x)$ by considering the asymptotic mean squared error (AMSE). Theorem 2 concludes that the AMSE of $\widehat{S}_c(y\,|\,x)$ is

$$\begin{aligned}\text{AMSE}\left(\widehat{S}_c(y\,|\,x)\right) &= \frac{\{h^2\,\mu_2(W)\,S^{0,2}(y\,|\,x) - h_0^2\,\mu_2(K)\,f^{1,0}(y\,|\,x)\}^2}{4} \\ &\quad + \frac{1}{n\,h}\,\frac{\mu_0(W^2)\,S(y\,|\,x)\,[1-S(y\,|\,x)]}{g(x)}. \qquad (10)\end{aligned}$$

By minimizing AMSE in (10) and taking $h_0 = o(h)$, therefore, we obtain the optimal bandwidth given by

$$h_{opt,S}(y\,|\,x) = \left[\frac{\mu_0(W^2)\,S(y\,|\,x)\,[1-S(y\,|\,x)]}{\{\mu_2(W)\,S^{0,2}(y\,|\,x)\}^2\;g(x)}\right]^{1/5}\;n^{-1/5}.$$

Therefore, the optimal rate of the AMSE of $\widehat{S}_c(y\,|\,x)$ is $n^{-4/5}$.

As for the boundary behavior of the WDKLL estimator, we can follow Cai (2002) to establish a similar result for $\widehat{S}_c(y\,|\,x)$ like Theorem 2 in Cai (2002). Without loss of generality, we consider the left boundary point $x = c\,h$, $0 < c < 1$. From Fan, Hu, and Troung (1994), we take $W(\cdot)$ to have support $[-1,\ 1]$ and $g(\cdot)$ to have support $[0,\ 1]$. Then, under Assumptions A1 - A7, by following the same proof as that for Theorem 2 and using the second assertion in Lemma 1, although not straightforward, we can show that

$$\sqrt{n\,h}\,\left[\widehat{S}_c(y\,|\,c\,h) - S_c(y\,|\,c\,h) - B_{S,c}(y) + o_p(h^2)\right]\;\to N\left(0,\,\sigma^2_{S,c}(y)\right), \qquad (11)$$

where the asymptotic bias term is given by $B_{S,c}(y) = h^2\,\beta_0(c)\,S^{0,2}(y\,|\,0+)/[2\,\beta_1(c)]$ and the asymptotic variance is $\sigma^2_{S,c}(y) = \beta_2(0)\,S(y\,|\,0+)[1-S(y\,|\,0+)]/[\beta_1^2(c)\,g(0+)]$ with $g(0+) = \lim_{z\downarrow 0} g(z)$,

$$\beta_0(c) = \int_{-1}^{c}\frac{u^2\,W(u)}{1-\lambda_c\,u\,W(u)}\,du, \qquad \beta_j(c) = \int_{-1}^{c}\frac{W^j(u)}{\{1-\lambda_c\,u\,W(u)\}^j}\,du, \quad 1\le j\le 2,$$

and λ_c being the root of equation $L_c(\lambda) = 0$

$$L_c(\lambda) = \int_{-1}^{c} \frac{u\,W(u)}{1 - \lambda\,u\,W(u)}\,du.$$

Note that the proof of (11) is similar to that for Theorem 2 in Cai (2002) and omitted. Theorem 2 and (11) reflect two of the major advantages of the WKDLL estimator: (a) the asymptotic bias does not depend on the design density $g(x)$, and indeed it is dependent only on the simple conditional distribution curvature $S^{0,2}(y\,|\,x)$ and conditional density curvature $f^{1,0}(y\,|\,x)$; and (b) it has an automatic good behavior at boundaries. See Cai (2002) for the detailed discussions.

Finally, we remark that if the point 0 were an interior point, then, (11) would hold with $c = 1$, which becomes Theorem 2. Therefore, Theorem 2 shows that the WKDLL estimation has the automatic good behavior at boundaries without the need of the boundary correction.

4.3 Asymptotic Properties for $\widehat{\nu}_p(x)$ and $\widehat{\mu}_p(x)$

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

$$\widehat{S}_c(\widehat{\nu}_p(x)\,|\,x) = p \approx \widehat{S}_c(\nu_p(x)\,|\,x) - \widehat{f}_c(\nu_p(x)\,|\,x)\,(\widehat{\nu}_p(x) - \nu_p(x)), \tag{12}$$

then,

$$\widehat{\nu}_p(x) - \nu_p(x) \approx [\widehat{S}_c(\nu_p(x)\,|\,x) - p]/\widehat{f}_c(\nu_p(x)\,|\,x) \approx [\widehat{S}_c(\nu_p(x)\,|\,x) - p]/f(\nu_p(x)\,|\,x)$$

by Theorem 1. As an application of Theorem 2, we can establish the following theorem for the asymptotic normality of $\widehat{\nu}_p(x)$ but the proof is omitted since it is similar to that for Theorem 2.

Theorem 3: *Under Assumptions A1 - A6, we have*

$$\sqrt{n\,h}\;\left[\widehat{\nu}_p(x) - \nu_p(x) - B_\nu(x) + o_p(h^2 + h_0^2)\right] \to N\left\{0,\ \sigma_\nu^2(x)\right\},$$

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

$$\sqrt{n\,h}\,\left[\widehat{\nu}_p(x) - \nu_p(x) - \frac{h^2}{2}\,\frac{S^{0,2}(\nu_p(x) \,|\, x)}{f(\nu_p(x) \,|\, x)}\,\mu_2(W) + o_p(h^2)\right] \to N\left\{0,\ \sigma_\nu^2(x)\right\}.$$

Remark 5: First, as a consequence of Theorem 3, $\widehat{\nu}_p(x) - \nu_p(x) = O_p\left(h^2 + h_0^2 + (n\,h)^{-1/2}\right)$ so that $\widehat{\nu}_p(x)$ is a consistent estimator of $\nu_p(x)$ with a convergence rate. Also, note that the asymptotic results for $\widehat{\nu}_p(x)$ in Theorem 3 are akin to those for $\widehat{\nu}_{ll,p}(x) = \widehat{S}_{ll}^{-1}(p \,|\, x)$ in Yu and Jones (1998) for iid data. But in the bias term of Theorem 3, the quantity $S^{0,2}(\nu_p(x) \,|\, x)/f(\nu_p(x) \,|\, x)$, involving the second derivative of the conditional distribution function with respect to x, replaces $\nu_p''(x)$, the second derivative of the conditional VaR function itself, 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 (2005) for time series. See Cai and Xu (2005) for details. This is not surprising since the bias comes only from the approximation. The former utilizes the approximation of the conditional distribution function but the later uses the approximation of the conditional VaR function. Finally, Theorems 2 and 3 imply that if the initial bandwidth h_0 is chosen small as possible such as $h_0 = o(h)$, the final estimates of $S(y \,|\, x)$ and $\nu_p(x)$ are not sensitive to the value of h_0 as long as it satisfies $h_0 = o(h)$. This makes the selection of bandwidths much easier in practice, which will be elaborated later (see Section 5.1).

Remark 6: Similar to Remark 5, we can derive the asymptotic mean squared error for $\widehat{\nu}_p(x)$. By following Yu and Jones (1998), Theorem 3 and (20) (given in Section 6) imply that the AMSE of $\widehat{\nu}_p(x)$ is given by

$$\begin{aligned}\text{AMSE}\,(\widehat{\nu}_p(x)) \;=\; & \frac{\left\{h^2\, S^{0,2}(\nu_p(x) \,|\, x)\,\mu_2(W) - h_0^2\, f^{1,0}(\nu_p(x) \,|\, x)\,\mu_2(K)\right\}^2}{4\, f^2(\nu_p(x) \,|\, x)} \\ & + \frac{1}{n\,h}\,\frac{\mu_0(W^2)\,[p(1-p) + 2\,h_0\, f(\nu_p(x) \,|\, x)\,\alpha(K)]}{f^2(\nu_p(x) \,|\, x)\, g(x)}. \end{aligned} \tag{13}$$

Note that the above result is similar to that in Theorem 1 in Yu and Jones (1998) for the double kernel local linear conditional quantile estimator. But, a comparison of (13) with Theorem 3 in Cai (2002) for the WNW estimator reveals that (13) has two extra terms (negligible if $h_0 = o(h)$) due to the vertical smoothing in the y direction, as mentioned previously. By minimizing AMSE in (13) and taking $h_0 = o(h)$, therefore, we obtain the optimal bandwidth given by

$$h_{opt,\nu}(x) = \left[\frac{\mu_0(W^2)\, p(1-p)}{\{\mu_2(W)\, S^{0,2}(\nu_p(x)\,|\,x)\}^2\; g(x)} \right]^{1/5} \; n^{-1/5}.$$

Therefore, the optimal rate of the AMSE of $\widehat{\nu}_p(x)$ is $n^{-4/5}$. By comparing $h_{opt,\nu}(x)$ with $h_{opt,S}(y\,|\,x)$, it turns out that $h_{opt,\nu}(x)$ is $h_{opt,\nu}(y\,|\,x)$ evaluated at $y = \nu_p(x)$. Therefore, the best choice of the bandwidth for estimating $S_c(y\,|\,x)$ can be used for estimating $\nu_p(x)$.

Remark 7: Similar to (11), one can establish the asymptotic result at boundaries for $\nu_p(x)$ as follows, one can show that if $h_0 = o(h)$,

$$\sqrt{n\,h}\, \left[\widehat{\nu}_p(c\,h) - \nu_p(c\,h) - B_{\nu,c} + o_p(h^2)\right] \;\rightarrow N\left(0,\, \sigma^2_{\nu,c}\right),$$

where the asymptotic bias is $B_{\nu,c} = h^2\beta_2(c)S^{0,2}(\nu_p(0+)|0+)/[2\beta_1(c)f(\nu_p(0+)|0+)]$ and the asymptotic variance is $\sigma^2_{\nu,c} = \beta_0(0)\,p\,[1-p]/[\beta_1^2(c)\,f^2(\nu_p(0+)\,|\,0+)\,g(0+)]$. Clearly, $\widehat{\nu}_p(x)$ inherits all good properties from the WDKLL conditional distribution estimator $S_c(y\,|\,x)$. Note that the above result can be established by using the second assertion in Lemma 1 and following the same lines along with those used in the proof of Theorem 2 and omitted.

Finally, we examine the asymptotic behavior for $\widehat{\mu}_p(x)$ at both interior and boundary points. First, we establish the following theorem for the asymptotic normality for $\widehat{\mu}_p(x)$.

Theorem 4: *Under Assumptions A1 - A5 and B1 - B5, we have*

$$\sqrt{n\,h}\, \left[\widehat{\mu}_p(x) - \mu_p(x) - B_\mu(x) + o_p(h^2 + h_0^2)\right] \rightarrow N\left\{0,\, \sigma^2_\mu(x)\right\},$$

where the asymptotic bias is $B_\mu(x) = B_{\mu,0}(x) + \frac{h_0^2}{2}\,\mu_2(K)\,p^{-1} f(\nu_p(x)\,|\,x)$ *with*

$$B_{\mu,0}(x) = \frac{h^2}{2}\,\mu_2(W)\,p^{-1}\,\left[l_1^{0,2}(\nu_p(x)\,|\,x) - \nu_p(x)\,S^{0,2}(\nu_p(x)\,|\,x)\right],$$

and the asymptotic variance is

$$\sigma_\mu^2(x) = \frac{\mu_0(W^2)}{p\,g(x)}\,\left[p^{-1}\,l_2(\nu_p(x)\,|\,x) - p\,\mu_p^2(x) + (1-p)\,\nu_p(x)\,\{\nu_p(x) - 2\,\mu_p(x)\}\right].$$

In particular, if Assumption A7 holds true, then,

$$\sqrt{n\,h}\,\left[\widehat{\mu}_p(x) - \mu_p(x) - B_{\mu,0}(x) + o_p(h^2)\right] \to N\left\{0,\ \sigma_\mu^2(x)\right\}.$$

Remark 8: First, Theorem 4 concludes that $\widehat{\mu}_p(x) - \mu_p(x) = O_p\left(h^2 + h_0^2 + (n\,h)^{-1/2}\right)$ so that $\widehat{\mu}_p(x)$ is a consistent estimator of $\mu_p(x)$ with a convergence rate. Also, note that the asymptotic results in Theorem 4 imply that $\widehat{\mu}_p(x)$ is a consistent estimator for $\mu_p(x)$ with a convergence rate $\sqrt{nh}$. Further, note that although the asymptotic variance $\sigma_\mu^2(x)$ is the same as in Scaillet (2005) for $\widetilde{\mu}_p(x)$, Scaillet (2005) did not provide an expression for the asymptotic bias term like $B_\mu(x)$ 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 it is negligible if $h_0 = o(h)$. If $h_0 = o(h)$, then $B_\mu(x)$ becomes $B_{\mu,0}(x)$.

Remark 9: Like Remark 5, the AMSE for $\widehat{\mu}_p(x)$ can be derived in the same manner. It follows from Theorem 4 that the AMSE of $\widehat{\mu}_p(x)$ is given by

$$\text{AMSE}\,(\widehat{\mu}_p(x)) = \frac{1}{n\,h}\,\sigma_\mu^2(x) + \left\{B_{\mu,0}(x) + \frac{h_0^2}{2}\,\mu_2(K)\,p^{-1}\,f(\nu_p(x)\,|\,x)\right\}^2. \tag{14}$$

If $h_0 = o(h)$, minimizing AMSE in (14) with respect to h yields the optimal bandwidth given by

$$h_{opt,\mu}(x) = \left[\frac{\sigma_\mu(x)}{\mu_2(W)\,p^{-1}\,\left\{l_1^{0,2}(\nu_p(x)\,|\,x) - \nu_p(x)\,S^{0,2}(\nu_p(x)\,|\,x)\right\}}\right]^{2/5} n^{-1/5}.$$

Therefore, 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 $x = c\,h$. By the same fashion, one can show that if $h_0 = o(h)$,

$$\sqrt{n\,h}\,\left[\widehat{\mu}_p(c\,h) - \mu_p(c\,h) - B_{\mu,c} + o_p(h^2)\right] \;\to N\left(0,\, \sigma^2_{\mu,c}\right),$$

where the asymptotic bias is

$$B_{\mu,c} = h^2\beta_2(c)\,p^{-1}\,\left[l_1^{0,2}(\nu_p(0+)\,|\,0+) - \nu_p(0+)\,S^{0,2}(\nu_p(0+)\,|\,0+)\right]/[2\beta_1(c)],$$

and the asymptotic variance is

$$\sigma^2_{\mu,c} = \frac{\beta_0(0)}{p\,\beta_1^2(c)\,g(0+)}\,\left[p^{-1}\,l_2(\nu_p(0+)\,|\,0+) - p\,\mu_p^2(0+) + (1-p)\,\nu_p(0+)\,\{\nu_p(0+) - 2\,\mu_p(0+)\}\right].$$

Note that the proof of the above result can be carried over by using the second assertion in Lemma 1 and following the same lines along with those used in the proof of Theorem 4 and omitted. Next, we consider the comparison of the performance of the WDKLL estimation $\widehat{\mu}_p(x)$ with the NW type kernel estimator $\widetilde{\mu}_p(x)$ 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 since it is along the same line with the proof of Theorem 2. See Scaillet (2005) for the results at the interior point. Under some regularity conditions, it can be shown although tediously (see Cai (2002) for details) that at the left boundary $x = c\,h$, the asymptotic bias term for the NW type kernel estimator $\widetilde{\mu}_p(x)$ is of the order h by comparing to the order h^2 for the WDKLL estimate (see $B_{\mu,c}$ above). 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 the WDKLL estimator over the WW type kernel estimator $\widetilde{\mu}_p(x)$.

5 Empirical Examples

To illustrate the methods proposed earlier, we consider two simulated examples and two real data examples on stock index returns. Throughout this section, the Epanechnikov kernel $K(u) = 0.75(1-u^2)_+$ is used and bandwidths are selected as described in the next section.

5.1 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 a convenient and effective data-driven rule. However, almost nothing has been done so far about this problem in the context of estimating CVaR $\nu_p(x)$ and CES $\mu_p(x)$ 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 not very sensitive 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 Akaike information criterion proposed by Cai and Tiwari (2000), to select the bandwidth, called $\widetilde{h}$. Then we take $0.1 \times \widetilde{h}$ or smaller as the initial bandwidth h_0. For the 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 of Cai and Tiwari (2000) can be applied here to select the optimal bandwidth for $\widehat{F}_c(\cdot|\cdot)$, denoted by h_S. As mentioned at the end of Remark 6, the bandwidth for $\widehat{\nu}_p(x)$ is the same as that for $\widehat{F}_c(\cdot|\cdot)$ so that it is simply to take h_S as h_ν. From (9), $\widehat{\mu}_p(x)$ is a linear estimator too. Therefore, by the same token, the nonparametric AIC selector is applied to selecting h_μ for $\widehat{\mu}_p(x)$. This is used in our implementation in the next sections.

5.2 Simulated Examples

In the simulated examples, we demonstrate the performance of the estimators in terms of the mean absolute deviation error (MADE). For example, for the conditional expected shortfall function, the MADE is defined as

$$\mathcal{E}_{\mu_p} = \frac{1}{n_0} \sum_{k=1}^{n_0} |\widehat{\mu}_p(x_k) - \mu_p(x_k)|,$$

where $x_k, k = 1, \cdots, n_0$ are the regular grid points. Similarly, we can define the MADE for the conditional value-at-risk function, denoted by $\mathcal{E}_{\nu_p}$.

Example 1. We use a simulated example of an ARCH type model with $X_t = Y_{t-1}$

$$Y_t = 0.9 \sin(2.5X_t) + \sigma(X_t)\varepsilon_t,$$

where $\sigma^2(x) = 0.8\sqrt{1.2 + x^2}$ and $\{\varepsilon_t\}$ are i.i.d. $N(0, 1)$. We consider three sample sizes: $n = 250$, $n = 500$, and $n = 1000$. The 5% WDKLL and NW estimation of conditional value-at-risk and expected shortfalls are computed and 500 replications are performed for each sample size. We compute the mean absolute deviation errors for each sample size. The results are summarized in Figures 1 and 2. For each n, the 500 $\mathcal{E}_{\nu_p}$ values of WDKLL estimation and the 500 $\mathcal{E}_{\nu_p}$ values of NW estimation of conditional VaR are plotted in Figure 1(d) in the form of boxplots. We can observe that the estimation becomes stable as the sample size increases for both WDKLL and NW estimators. This is in line with our asymptotic theory that the proposed estimators are consistent. It is obvious that the MADEs of WDKLL estimator are smaller than the MADEs of NW estimator. This indicates that our WDKLL estimator has smaller bias than NW estimator.

Figures $1(a) - (c)$, respectively, display the true conditional VaR functions $\nu_p(x) = 0.9 \sin(2.5X_t) + \sigma(x)\Phi^{-1}(1 - p)$ in solid lines, where $\Phi(\cdot)$ is the standard normal distribution function. Also, the dashed lines represent the proposed WDKLL estimates of conditional VaR from a typical sample. The dotted lines represent the NW estimates of conditional VaR from a typical sample. The typical sample is selected in such a way that its $\mathcal{E}_{\nu_p}$ value equals to the median in the 500 replications. It is obvious that the both WDKLL and NW estimates give the best fit of true conditional VaR function when $n = 500$. The performance of WDKLL is better than NW estimator, especially in the boundary.

In Figures 2$(a)-(c)$, the solid lines show the true conditional ES functions $\mu_p(x) = 0.9\sin(2.5X_t)p + \sigma(x)\mu_1(\Phi^{-1}(1-p))$, where $\mu_1(t) = \int_t^\infty u\phi(u)du$ and $\phi(\cdot)$ is the standard normal distribution density function. The dashed lines are the proposed WDKLL estimates of conditional ES from the typical sample. The dotted lines are the NW estimates of conditional ES from the typical sample. The typical sample is selected in such a way that its $\mathcal{E}_{\mu_p}$ value equals to the median in the 500 replications. For each n, boxplots of the 500 $\mathcal{E}_{\mu_p}$ values of conditional ES are plotted in Figure 2(d) for both WDKLL and NW estimates. We can conclude that conditional ES estimators have a similar performance as that for conditional VaR estimators. The estimation becomes stable as the sample size increases and the estimated curves perform better as n increases.

The 1% WDKLL and NW estimates of conditional VaR and ES are computed under the same setting. The results are displayed in Figures 3 and 4. Results similar to those for the 5% estimates can be observed. But it is not surprising to see that the performance of 1% conditional VaR and conditional ES estimates is not good as that for the 5% estimates.

Example 2. In the above example, we only consider the case that X_t is a scalar. In this example, we consider the multivariate situation, i.e. X_t consists of two lagged variables: Y_{t-1} and Y_{t-2}. The model is shown below:

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

where $m(X_t) = 0.63Y_{t-1} - 0.47Y_{t-2}$ and $\sigma^2(X_t) = 0.5 + 0.23Y_{t-1}^2 + 0.3Y_{t-2}^2$. $\{\varepsilon_t\}$ are generated from $N(0,1)$. Three sample sizes: $n = 200$, $n = 400$, and $n = 600$, are considered here. For each sample size, we replicate the design 500 times. Here we only present the boxplots of the MADE for the conditional VaR and ES estimates in Figure 5. Figures 5(a) display boxplots of the 500 $\mathcal{E}_{\nu_p}$ values of WDKLL and the 500 $\mathcal{E}_{\nu_p}$ NW estimates of conditional VaR. Figures 5(b) display boxplots of the 500 $\mathcal{E}_{\mu_p}$ values of WDKLL and the 500 $\mathcal{E}_{\mu_p}$ values NW estimates of conditional ES. From Figures 5(a) and (b), it is visually verified that both WDKLL and

NW estimation become stable as the sample size increases and the performance of WDKLL estimator is better than the performance of NW estimator.

5.3 Real Examples

Example 3. Now we consider data on Dow Jones Industrials (DJI) index returns. We take a sample of 1801 daily prices from DJI indices, 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 and X_t be the first lagged variable of Y_t. The estimators proposed in this paper are used to estimate the 5% conditional VaR and ES functions. The estimation results are shown in Figure 6. Figure 6(a) shows the 5% conditional VaR estimates and Figure 6(b) shows the 5% conditional ES estimates. Both conditional VaR and ES estimates exhibit a U-shape, which is close 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 that the curves are asymmetric. This may indicate that the DJI is more likely to fall down if there was a loss within the last day than there was a same amount positive return.

Example 4. We apply the proposed methods to estimate the conditional value-at-risk and expected shortfall of the International Business Machine Co. (NYSE: IBM) security returns. The data are one-day prices recorded from March 1, 1996 to April 6, 2005. We use the same method to calculate the daily returns as in Example 3. In order to estimate the value-at-risk of a stock return, generally, the information set X_t may contain a market index of corresponding capitalization and type, the industry index, and the lagged values of stock return. For this example, Y_t is the log loss of IBM, and we choose two variables as information set for the sake of simplicity. Let X_t be the first lagged variable of Y_t and lagged daily negative log return of Dow Jones Industrials (DJI) index. Our main results from the estimation of the model are summarized in Figure 7. Figures 7(a) and (b) show the surfaces of conditional VaR and ES estimators of IBM returns. For the fixed value of IBM log loss, we know conditional VaR and ES are the functions of DJI log loss. Figures 7(c) and (e) depict the conditional VaR and ES curves for three different values of negative log-return of IBM stock (-0.275, -0.025, 0.325). For the fixed value of DJI log loss, we conclude that conditional VaR and ES are the functions of IBM log loss. Figures 7(d) and (f) display the conditional VaR and ES curves for three different values of negative log-return of DJI stock (-0.225, 0.025, 0.425).

From Figures 7(c) - (f), we can observe that most of these curves are U-shaped. This is consistent with the results observed in Example 3. Also, we need to notice that these three curves in each figure are not parallel. This implies that the effects of lagged IBM and lagged DJI variables on the risk of IBM are mixing. Let us examine Figure 7(d). When past IBM log loss is around -0.2 these three curves are close to each other. It seems that DJI has fewer effects (bring less information) on CVaR around this value. On the other hand, DJI has more effects when IBM log loss is far from this value.

6 Proofs of Theorems

In this section, we present the proofs of Theorems 1 - 4. First, we list two lemmas.

Lemma 1: *Under Assumptions A1 - A5, we have*

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

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*

$$p_t(c\,h) = n^{-1}\,b_t^c(c\,h)\,\{1 + o_p(1)\},$$

where $b_t^c(x) = [1 + \lambda_c\,(X_t - x)\,K_h(x - X_t)]^{-1}$.

Proof: See the proofs of Lemmas 2 and 3 in Cai (2002), omitted. □

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 $c_t(x) = b_t(x)\, W_h(x - X_t)$.

Proof: See Appendix. □

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

$$W_{c,t}(x,h) \approx \frac{b_t(x)\, W_h(x - X_t)}{\sum_{t=1}^{n} b_t(x)\, W_h(x - X_t)} \approx n^{-1}\, g^{-1}(x)\, b_t(x)\, W_h(x - X_t) = \frac{c_t(x)}{n\, g(x)}. \tag{15}$$

Now we embark on the proofs of theorems.

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

$$\widehat{f}_c(y \,|\, x) - f(y \,|\, x) \equiv I_1 + I_2 + I_3, \tag{16}$$

where with $\varepsilon_{t,1} = Y_t^*(y) - E(Y_t^*(y) | X_t)$,

$$I_1 = \sum_{t=1}^{n} \varepsilon_{t,1}\, W_{c,t}(x,h), \quad I_2 = \sum_{t=1}^{n} [E(Y_t^*(y)|X_t) - f(y|X_t)]\, W_{c,t}(x,h),$$

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), (15), and Lemmas 1 and 2 gives

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

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

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

Therefore,

$$I_2 + I_3 = \frac{h^2}{2}\,\mu_2(W)\,f^{0,2}(y\,|\,x) + \frac{h_0^2}{2}\,\mu_2(K)\,f^{2,0}(y\,|\,x) + o_p(h^2 + h_0^2) = B_f(y\,|\,x) + o_p(h^2 + h_0^2).$$

Thus, (16) becomes

$$\begin{aligned} & \sqrt{nh_0h}\;\left[\widehat{f}_c(y\,|\,x) - f(y\,|\,x) - B_f(y\,|\,x) + o_p(h^2 + h_0^2)\right] = \sqrt{nh_0h}\;I_1 \\ = \;\; & g^{-1}(x)\;I_4\,\{1 + o_p(1)\} \to N\left\{0,\;\sigma_f^2(y\,|\,x)\right\}, \end{aligned}$$

where $I_4 = \sqrt{h_0\,h/n}\,\sum_{t=1}^n \varepsilon_{t,1}\,c_t(x)$. This, together with Lemma 3 in Appendix, therefore, proves the theorem. □

Proof of Theorem 2: Similar to (16), we have

$$\widehat{S}_c(y\,|\,x) - S(y\,|\,x) \equiv I_5 + I_6 + I_7, \tag{17}$$

where with $\varepsilon_{t,2} = \bar{G}_{h_0}(y - Y_t) - E(\bar{G}_{h_0}(y - Y_t)|X_t)$,

$$I_5 = \sum_{t=1}^n \varepsilon_{t,2}\,W_{c,t}(x,h), \quad I_6 = \sum_{t=1}^n [E\{\bar{G}_{h_0}(y - Y_t)\,|\,X_t\} - S(y|X_t)]\,W_{c,t}(x,h),$$

and

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

By analogy with the analysis of I_2, the Taylor expansion, (7), and Lemmas 1 and 2, we have

$$\begin{aligned} I_7 &= \sum_{t=1}^n \frac{1}{2} S^{0,2}(y\,|\,x)\,W_{c,t}(x,h)\,(X_t - x)^2 + o_p(h^2) \\ &= \frac{1}{2} S^{0,2}(y\,|\,x)\,g^{-1}(x)\,n^{-1}\,\sum_{t=1}^n c_t(x)\,(X_t - x)^2 + o_p(h^2) \\ &= \frac{h^2}{2}\,\mu_2(W)\,S^{0,2}(y\,|\,x) + o_p(h^2). \end{aligned}$$

To evaluate I_6, first, we consider the following

$$\begin{aligned} & E[\bar{G}_{h_0}(y - Y_t)\,|\,X_t = x] = \int_{-\infty}^{\infty} K(u)\,S(y - h_0\,u\,|\,x)du \\ = \;\; & S(y\,|\,x) + \frac{h_0^2}{2}\,\mu_2(K)\,S^{2,0}(y\,|\,x) + o(h_0^2) \\ = \;\; & S(y\,|\,x) - \frac{h_0^2}{2}\,\mu_2(K)\,f^{1,0}(y\,|\,x) + o(h_0^2). \end{aligned} \tag{18}$$

By (18) and following the same arguments as in the proof of Lemma 2, we have

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

Therefore,

$$I_6 + I_7 = \frac{h^2}{2}\,\mu_2(W)\,S^{0,2}(y\,|\,x) - \frac{h_0^2}{2}\,\mu_2(K)\,f^{1,0}(y\,|\,x) + o_p(h^2+h_0^2) = B_S(y\,|\,x) + o_p(h^2+h_0^2),$$

so that by (17),

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

Clearly, to accomplish the proof of theorem, it suffices to establish the asymptotic normality of $\sqrt{n\,h}\,I_5$. To this end, first, we compute $\mathrm{Var}(\varepsilon_{t,2}\,|\,X_t = x)$. Note that

$$\begin{aligned} & E[\bar{G}_{h_0}^2(y - Y_t)\,|\,X_t = x] = \int_{-\infty}^{\infty} \bar{G}_{h_0}^2(y-u)\,f(u\,|\,x)du \\ =\ & \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(u_1)\,K(u_2)\,S(\max(y - h_0\,u_1, y - h_0\,u_2)\,|\,x)du_1du_2 \\ =\ & S(y\,|\,x) + 2\,h_0\,\alpha(K)\,f(y\,|\,x) + O(h_0^2), \end{aligned} \tag{19}$$

which, in conjunction with (18), implies that

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

This, together with the fact that

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

leads to

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

Now, since $|\varepsilon_{t,2}| \le 1$, by following the same arguments as those used in the proofs of Lemma 2 and 3 in Appendix (or Lemma 1 and Theorem 1 in Cai (2002)), we can show although tediously that

$$\mathrm{Var}(I_8) = \sigma_S^2(y\,|\,x)\,g^2(x) + 2\,\mu_0(W^2)\,h_0\,\alpha(K)\,f(y\,|\,x)\,g(x) + o(h_0), \tag{20}$$

where $I_8 = \sqrt{h/n}\,\sum_{t=1}^n \varepsilon_{t,2}\, c_t(x)$, and

$$\sqrt{n\,h}\, I_5 = g^{-1}(x)\, I_8\, \{1 + o_p(1)\} \to N\left\{0,\ \sigma_S^2(y\,|\,x)\right\}.$$

This completes the proof of Theorem 2. □

Proof of Theorem 4: Similar to (12), we use the Taylor expansion and ignore the higher terms to obtain

$$\int_{\widehat{\nu}_p(x)}^{\infty} y\, K_{h_0}(y - Y_t)dy \approx \int_{\nu_p(x)}^{\infty} y\, K_{h_0}(y - Y_t)dy - \nu_p(x)\, K_{h_0}(\nu_p(x) - Y_t)\ [\widehat{\nu}_p(x) - \nu_p(x)]$$
$$= Y_t \bar{G}_{h_0}(\nu_p(x) - Y_t) - \nu_p(x)\, K_{h_0}(\nu_p(x) - Y_t)\ [\widehat{\nu}_p(x) - \nu_p(x)] + h_0\, G_{1,h_0}(\nu_p(x) - Y_t).$$

Plugging the above into (9) leads to

$$p\,\widehat{\mu}_p(x) \approx \widehat{\mu}_{p,1}(x) + I_9, \tag{21}$$

where

$$\widehat{\mu}_{p,1}(x) = \sum_{t=1}^n W_{c,t}(x,h)\, Y_t \bar{G}_{h_0}(\nu_p(x) - Y_t) - \nu_p(x) \widehat{f}_c(\nu_p(x)|x)[\widehat{\nu}_p(x) - \nu_p(x)],$$

which will be shown later to be the source of both the asymptotic bias and variance, and

$$I_9 = h_0 \sum_{t=1}^n W_{c,t}(x,h)\, G_{1,h_0}(\nu_p(x) - Y_t),$$

which will be shown to contribute only the asymptotic bias (see Lemma 4 in Appendix). From (12) and (8),

$$\widehat{f}_c(\nu_p(x)\,|\,x)\,[\widehat{\nu}_p(x) - \nu_p(x)] \approx \sum_{t=1}^n W_{c,t}(x,h)\{\bar{G}_{h_0}(\nu_p(x) - Y_t) - p\}.$$

Therefore, by (15),

$$\begin{aligned}
\widehat{\mu}_{p,1}(x) &= \sum_{t=1}^n W_{c,t}(x,h)\,[\{Y_t - \nu_p(x)\}\bar{G}_{h_0}(\nu_p(x) - Y_t) - p\,\nu_p(x)] \\
&= \sum_{t=1}^n W_{c,t}(x,h)\,\varepsilon_{t,3} + \sum_{t=1}^n W_{c,t}(x,h)\, E\{\zeta_t(x)\,|\,X_t\} \\
&\approx g^{-1}(x)\, n^{-1} \sum_{t=1}^n \varepsilon_{t,3}\, c_t(x) + \sum_{t=1}^n W_{c,t}(x,h)\, E\{\zeta_t(x)\,|\,X_t\} \\
&\equiv \widehat{\mu}_{p,2}(x) + \widehat{\mu}_{p,3}(x),
\end{aligned}$$

where $\zeta_t(x) = [Y_t - \nu_p(x)]\, \bar{G}_{h_0}(\nu_p(x) - Y_t) + p\,\nu_p(x)$ 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 $\widehat{\mu}_{p,3}(x)$ and together with 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

$$\begin{aligned} & E[Y_t\, \bar{G}_{h_0}(\nu_p(x) - Y_t) \,|\, X_t = v] = \int_{-\infty}^{\infty} K(u) du \int_{\nu_p(x) - h_0\, u}^{\infty} y\, f(y \,|\, v) dy \\ = & \int_{-\infty}^{\infty} l_1(\nu_p(x) - h_0\, u \,|\, v)\, K(u) du = l_1(\nu_p(x) \,|\, v) + \frac{h_0^2}{2}\, \mu_2(K)\, l_1^{2,0}(\nu_p(x) \,|\, v) + o(h_0^2) \\ = & \; l_1(\nu_p(x) \,|\, v) - \frac{h_0^2}{2}\, \mu_2(K)\, \left[\nu_p(x)\, f^{1,0}(\nu_p(x) \,|\, v) + f(\nu_p(x) \,|\, x)\right] + o(h_0^2), \end{aligned}$$

which, in conjunction with (18), leads to

$$\zeta(v) = E[\zeta_t(x) \,|\, X_t = v] = A(\nu_p(x) \,|\, v) - \frac{h_0^2}{2}\, \mu_2(K)\, f(\nu_p(x) \,|\, v) + o(h_0^2), \tag{22}$$

where $A(\nu_p(x)|v) = l_1(\nu_p(x) \,|\, v) - \nu_p(x)\, [S(\nu_p(x) \,|\, v) - p]$. It is easy to verify that $A(\nu_p(x)|v) = E[\{Y_t - \nu_p(x)\}\, I(Y_t \geq \nu_p(x)) \,|\, X_t = v] + p\,\nu_p(x)$, $A(\nu_p(x)|x) = p\,\mu_p(x)$, and $A^{0,2}(\nu_p(x)|x) = l_1^{0,2}(\nu_p(x) \,|\, x) - \nu_p(x)\, S^{0,2}(\nu_p(x) \,|\, x)$. Therefore, by (22), the Taylor expansion, and (7), $\widehat{\mu}_{p,3}(x)$ becomes

$$\widehat{\mu}_{p,3}(x) = \sum_{t=1}^{n} W_{c,t}(x,h)\, \zeta(X_t) = \zeta(x) + \frac{1}{2}\, \zeta''(x) \sum_{t=1}^{n} W_{c,t}(x,h)\, (X_t - x)^2 + o_p(h^2).$$

Further, by Lemmas 1 and 2,

$$\begin{aligned} \widehat{\mu}_{p,3}(x) &= \zeta(x) + \frac{h^2}{2}\, \mu_2(W)\, \zeta''(x) + o_p(h^2) \\ &= p\,\mu_p(x) + \frac{h^2}{2}\, \mu_2(W)\, A^{0,2}(\nu_p(x) \,|\, x) - \frac{h_0^2}{2}\, \mu_2(K)\, f(\nu_p(x) \,|\, x) + o_p(h_0^2). \end{aligned}$$

This, in conjunction with Lemma 4, concludes that

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

so that by (21),

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

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 Appendix, we have

$$\sqrt{n\,h}\left[\widehat{\mu}_p(x) - \mu_p(x) - B_\mu(x) + o_p(h^2 + h_0^2)\right] = \frac{1}{p\,g(x)}\,I_{10}\,\{1 + o_p(1)\} \to N\left\{0,\, \sigma_\mu^2(x)\right\},$$

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

Appendix: Proofs of Lemmas

In this section, we present the proofs of Lemmas 2, 3, 4, and 5. Note that we use the same notation as in Sections 2 - 6. 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

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

and

$$E(\xi_t^2) = h^{-1}\int \frac{v^{2j}\,W^2(v)\;g(x - h\,v)}{[1 + h\,\lambda_0\,v\,W(v)]^2}\,dv = O(h^{-1}).$$

Also, by the stationarity, a straightforward manipulation yields

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

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

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

where $d_n = O(h^{-1/(1+\delta/2)})$. For J_{j1}, it follows by Assumption A4 that $|\mathrm{Cov}(\xi_1,\, \xi_t)| \le C$, so that $J_{j1} = O(d_n) = o(h^{-1})$. For J_{j2}, Assumption A2 implies that $|(X_t - x)^j\,W_h(x - X_t)| \le$

$C\,h^{j-1}$, so that $|\xi_t| \le C\,h^{-1}$. Then, it follows from the Davydov's inequality (see, e.g., Theorem 17.2.1 of Ibragimov and Linnik (1971)) that $|\text{Cov}(\xi_1,\, \xi_{t+1})| \le C\,h^{-2}\,\alpha(t)$, which, together with Assumption A5, implies that

$$J_{j2} \le C\,h^{-2} \sum_{t \ge d_n} \alpha(t) \le C\; h^{-2}\, d_n^{-(1+\delta)} = o(h^{-1}).$$

This, together with (A.2) and (A.3), therefore implies that $\text{Var}(J_j) = O((n\,h)^{-1}) = o(1)$. This completes the proof of the lemma. □

Lemma 3: *Under Assumptions A1 - A6, we have*

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

Proof: It follows by using the same lines as those used in the proof of Lemma 2 and Theorem 1 in Cai (2002), omitted. The outline is described as follows. First, similar to the proof of Lemma 2, it is easy to see that

$$\text{Var}(I_4) = h_0\,h\,\text{Var}(\varepsilon_{t,1}\, c_t(x)) + h_0\,h \sum_{t=2}^{n} l_{n,t}\, \text{Cov}(\varepsilon_{1,1}\, c_1(x),\, \varepsilon_{t,1}\, c_t(x)). \qquad \text{(A.4)}$$

Next, we compute $\text{Var}(\varepsilon_{t,1}\,|\,X_t = x)$. Note that

$$h_0\, E[Y_t^*(y)^2\,|\,X_t = x] = \int_{-\infty}^{\infty} K^2(u)\, f(y - h_0 u\,|\,x) du = \mu_0(K^2)\, f(y\,|\,x) + O(h_0^2),$$

which, together with the fact that

$$\text{Var}(\varepsilon_{t,1}\, c_t(x)) = E\left[c_t^2(x)\, E\{\varepsilon_{t,1}^2\,|\,X_t\}\right] = E\left[c_t^2(x)\,\text{Var}(\varepsilon_{t,1}\,|\,X_t)\right]$$

and (2), implies that

$$h\,h_0\,\text{Var}(\varepsilon_{t,1}\, c_t(x)) = \mu_0(K^2)\,\mu_0(W^2)\, f(y\,|\,x)\, g(x) + O(h_0^2) = \sigma_f^2(y\,|\,x)\, g^2(x) + O(h_0^2).$$

As for the second term on the right hand side of (A.4), similar to (A.3), it is decomposed into two summons. By using Assumptions A4 and B2 for the first summon and using the Davydov's inequality and Assumption A5 to the second summon, we can show that the second term on the right hand side of (A.4) goes to zero as n goes to infinity. Thus, $\text{Var}(I_4) \to \sigma_f^2(y\,|\,x)\,g^2(x)$ by (A.4). To show the normality, we employ Doob's small-block and large-block technique (see, e.g., Ibragimov and Linnik, 1971, p. 316). Namely, partition $\{1,\,\ldots,\,n\}$ into $2\,q_n+1$ subsets with large-block of size $r_n = \lfloor (n\,h)^{1/2} \rfloor$ and small-block of size $s_n = \lfloor (n\,h)^{1/2}/\log n \rfloor$, where $q_n = \lfloor n/(r_n+s_n) \rfloor$ with $\lfloor x \rfloor$ denoting the integer part of x. By following the same steps as in the proof of Theorem 1 in Cai (2002), we can accomplish the rest of proofs: the summands for the large-blocks are asymptotically independent, two summands for the small-blocks are asymptotically negligible in probability, and the standard Lindeberg-Feller conditions hold for the summands for the large-blocks. See Cai (2002) for details. So, the proof of the lemma is complete. □

Lemma 4: *Under Assumptions A1 - A6, we have*

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

Proof: Define $\xi_{t,1} = c_t(x)\,G_{1,h_0}(\nu_p(x)-Y_t)$. 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),

$$\begin{aligned} E\left(\xi_{t,1}\right) &= E\left[c_t(x)\,E\left\{G_{1,h_0}(\nu_p(x)-Y_t)\,|\,X_t\right\}\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{K(u)\,W(v)\,u\,S(\nu_p(x)-h_0\,u)\,|\,x)\,g(x-h\,v)}{1+h\,\lambda_0\,v\,W(v)}\,dudv \\ &= h_0\,\mu_2(K)\,f(\nu_p(x)\,|\,x)\,g(x) + O(h_0\,h^2), \end{aligned}$$

and

$$E(\xi_{t,1}^2) = E\left[b_t^2(x)\,W_h^2(x-X_t)\,E\left\{G_{1,h_0}^2(\nu_p(x)-Y_t)\,|\,X_t\right\}\right] = O(h_0/h),$$

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 $\text{Var}(I_{10}) = O((n\,h)^{-1}) = o(1)$. This proves the lemma. □

Lemma 5: *Under Assumptions A1 - A5 and B1 - B5, we have*

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

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. The main idea is as follows. First, similar to the proof of Lemmas 2 and 3, we will show by Assumptions B1 - B3 that

$$\mathrm{Var}(I_{10}) \to p^2 \sigma_\mu^2(x)\, g^2(x). \tag{A.5}$$

Finally, we need to compute $\mathrm{Var}(\varepsilon_{t,3}\, c_t(x))$. Since

$$\mathrm{Var}(\varepsilon_{t,3}\, c_t(x)) = E\left[c_t^2(x)\, E\{\varepsilon_{t,3}^2 \,|\, X_t\}\right] = E\left[c_t^2(x)\, \mathrm{Var}(\zeta_t(x) \,|\, X_t)\right],$$

then, we first need to calculate $\mathrm{Var}(\zeta_t(x) \,|\, X_t)$. To this effect, by (22),

$$\begin{aligned} &\mathrm{Var}(\zeta_t(x) \,|\, X_t = v) = \mathrm{Var}[(Y_t - \nu_p(x))\, \bar{G}_{h_0}(\nu_p(x) - Y_t) \,|\, X_t = v] \\ =\ & E\left[(Y_t - \nu_p(x))^2 \bar{G}_{h_0}^2(\nu_p(x) - Y_t)|X_t = v\right] - [l_1(\nu_p(x)|v) - \nu_p(x) S(\nu_p(x)|v)]^2 + O(h_0^2). \end{aligned}$$

Similar to (19),

$$\begin{aligned} & E[(Y_t - \nu_p(x))^2\, \bar{G}_{h_0}^2(\nu_p(x) - Y_t) \,|\, X_t = v] = \int_{-\infty}^{\infty} G_{h_0}^2(\nu_p(x) - y)\, (y - \nu_p(x))^2\, f(y \,|\, v) dy \\ =\ & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(u_1)\, K(u_2)\, \tau(\max(\nu_p(x) - h_0\, u_1, \nu_p(x) - h_0\, u_2) \,|\, v) du_1 du_2 \\ =\ & \tau(\nu_p(x) \,|\, v) - 2\, h_0\, \tau^{1,0}(\nu_p(x) \,|\, v)\, \alpha(K) + O(h_0^2) = \tau(\nu_p(x) \,|\, v) + O(h_0^2) \end{aligned}$$

since $\tau^{1,0}(\nu_p(x) \,|\, v) = 0$, where $\tau(u \,|\, v) = l_2(u \,|\, v) - 2\, \nu_p(x) l_1(u \,|\, v) + \nu_p^2(x) S(u \,|\, v)$. Therefore,

$$\mathrm{Var}(\zeta_t(x) \,|\, X_t = v) = \mathrm{Var}[(Y_t - \nu_p(x)) I(Y_t \geq \nu_p(x)) \,|\, X_t = v] + O(h_0^2),$$

and

$$h\, \mathrm{Var}(\varepsilon_{t,3}\, c_t(x)) = \mu_0(W^2)\, \mathrm{Var}[(Y_t - \nu_p(x)) I(Y_t \geq \nu_p(x)) \,|\, X_t = x]\, g(x) + o(1).$$

Similar to Lemmas 2 and 3, clearly, we have,

$$\mathrm{Var}(I_{10}) = h\, \mathrm{Var}(\varepsilon_{t,3}\, c_t(x)) + h \sum_{t=2}^{n} l_{n,t}\, \mathrm{Cov}(\varepsilon_{1,3}\, c_1(x),\, \varepsilon_{t,3}\, c_t(x)),$$

and the first term on right hand side of the above equation converges to $p^2\, \sigma_\mu^2(x)\, g^2(x)$. As for the second term on the right hand side of the above equation, similar to (A.3), it is decomposed into two summons. By using Assumptions A4 and B2 for the first summon and using the Davydov's inequality and Assumption A5 to the second summon, we can show that the second term on the right hand side of the above equation goes to zero as n goes to infinity. Thus, (A.5) holds. To show the normality, we employ Doob's small-block and large-block technique (see, e.g., Ibragimov and Linnik, 1971, p. 316). Namely, partition $\{1, \ldots, n\}$ into $2\, q_n + 1$ subsets with large-block of size r_n and small-block of size s_n, where s_n is given in Assumption B4, $q_n = \lfloor n/(r_n + s_n) \rfloor$ with $\lfloor x \rfloor$ denoting the integer part of x, and $r_n = \lfloor (n\, h)^{1/2}/\gamma_n \rfloor$ with γ_n satisfying followings: γ_n is a sequence of positive numbers $\gamma_n \to \infty$ such that $\gamma_n\, s_n/\sqrt{n\, h} \to 0$ and $\gamma_n\, (n/h)^{1/2} \alpha(s_n) \to 0$ by Assumption B4. By following the same steps as in the proof of Theorem 1 in Cai (2001), we can accomplish the rest of proofs: the summands for the large-blocks are asymptotically independent, two summands for the small-blocks are asymptotically negligible in probability, and the standard Lindeberg-Feller conditions hold for the summands for the large-blocks. See Cai (2001) for details. Therefore, the lemma is proved. □

References

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

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

Boente, G. and R. Fraiman (1995). Asymptotic distribution of smoothers based on local means and local medians under dependence. *Journal of Multivariate Analysis*, **54**, 77-90.

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. (2003). Trending time varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, in press.

Cai, Z. and L. Qian (2000). Local estimation of a biometric function with covariate effects. In *Asymptotics in Statistics and Probability* (M. Puri, ed), 47-70.

Cai, Z. and R.C. Tiwari (2000). Application of a local linear autoregressive model to BOD time series. *Environmetrics*, **11**, 341-350.

Cai, Z. and X. Xu (2005). Nonparametric quantile estimations for dynamic smooth coefficient models. *Working Paper*, Department of Mathematics and Statistics, University of North Carolina at Charlotte. Revised for *Journal of the American Statistical Association*.

Carrasco, M. and X. Chen (2002). Mixing and moments properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 17-39.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of statistics*, **19**, 760-777.

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

Chernozhukov, V. and L. Umanstev (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, **26**, 271-292.

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

Duffie, D. and K.J. Singleton (2003). *Credit Risk: Pricing, Measurement, and Management*. Princeton: Princeton University Press.

Embrechts, P., C. Kluppelberg, and T. Mikosch (1997). *Modeling Extremal Events For Finance and Insurance*. New York: Springer-Verlag.

Engle, R.F. and S. Manganelli (2004). CAViaR: conditional autoregressive value at risk by regression quantile. *Journal of Business and Economics Statistics*, **22**, 367-381.

Fan, J. and I. Gijbels (1996). *Local Polynomial Modeling and Its Applications*. London: Chapman and Hall.

Fan, J. and J. Gu (2003). Semiparametric estimation of value-at-risk. *Econometrics Journal*, **6**, 261-290.

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

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

Frey, R. and A.J. McNeil (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., R.C.L. Wolff, and Q. Yao (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, **94**, 154-163.

Hürlimann, W. (2003). A Gaussian exponential approximation to some compound Poisson distributions. *ASTIN Bulletin*, **33**, 41-55.

Ibragimov, I.A. and Yu. V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Groningen, the Netherlands: Walters-Noordhoff.

Jorion, P. (2001). *Value at Risk*, 2nd Edition. New York: McGraw-Hill.

Jorion, P. (2003). *Financial Risk Manager Handbook*, 2nd Edition. New York: John Wiley.

Koenker, R. and G.W. Bassett (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

Lejeune, M.G. and P. Sarda (1988). Quantile regression: a nonparametric approach. *Computational Statistics and Data Analysis*, **6**, 229-281.

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

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

Morgan, J.P. (1996). *Risk Metrics - Technical Documents*, 4th Edition, New York.

Oakes, D. and T. Dasu (1990). A note on residual life. *Biometrika*, **77**, 409-410.

Pötscher, B.M. and I.R. Prucha (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Berlin: Springer-Verlag.

Rockafellar, R. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21-41.

Roussas, G.G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *The Annals of Mathematical Statistics*, **40**, 1386-1400.

Roussas, G.G. (1991). Estimation of transition distribution function and its quantiles in Markov processes: Strong consistency and asymptotic normality. In G.G. Roussas (ed.), *Nonparametric Functional Estimation and related Topics*, pp. 443-462. Amsterdam: Kluwer Academic.

Samanta, M. (1989). Nonparametric estimation of conditional quantiles. *Statistics and Probability Letters*, **7**, 407-412.

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.

Troung, Y.K. (1989). Asymptotic properties of kernel estimators based on local median. *The Annals of Statistics*, **17**, 606-617.

Troung, Y.K. and C.J. Stone (1992) Nonparametric function estimation involving time series. *The Annals of Statistics* 20, 77-97.

Yu, K. and M.C. Jones (1997). A comparison of local constant and local linear regression quantile estimation. *Computational Statistics and Data Analysis*, **25**, 159-166.

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

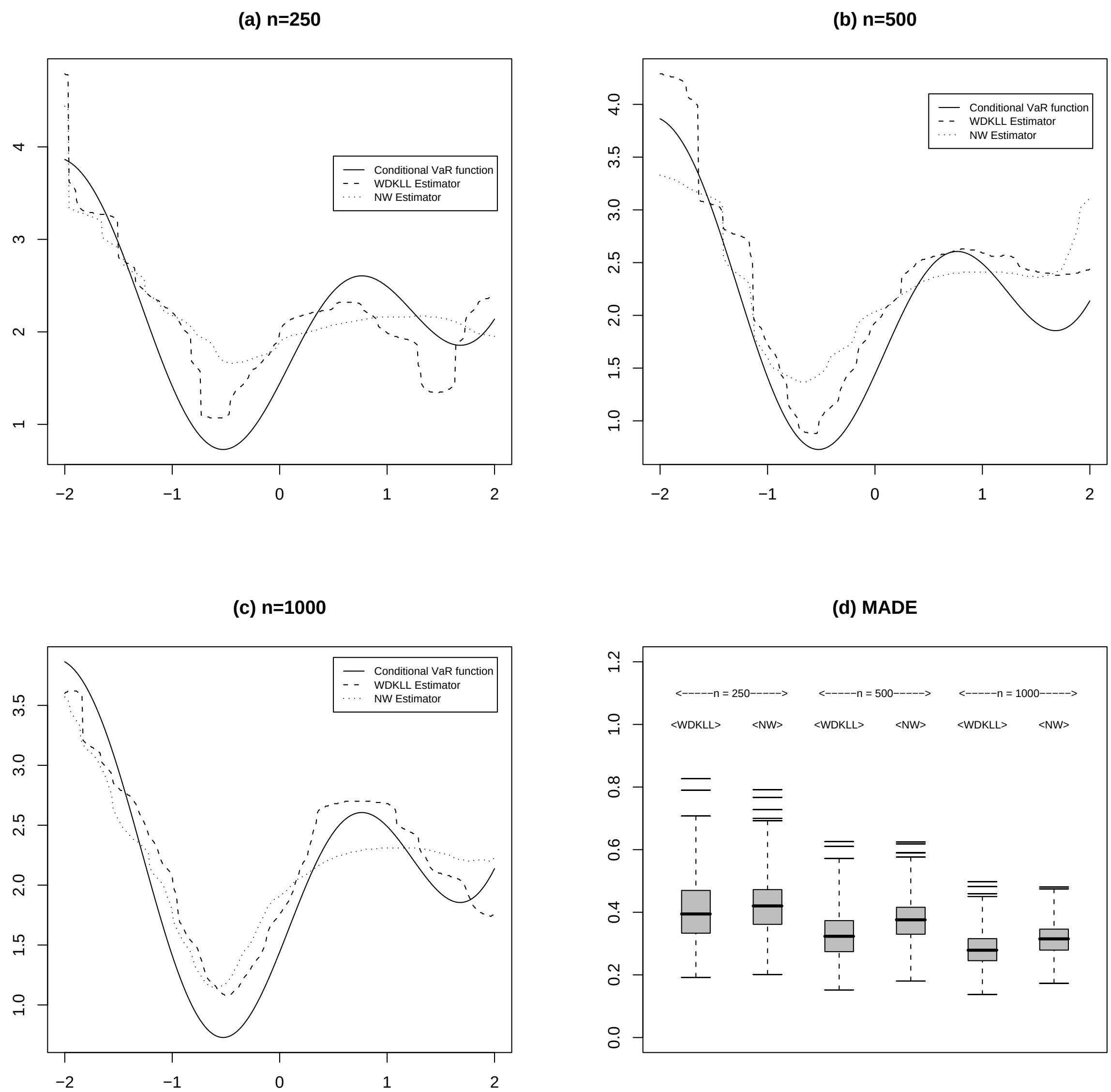

Figure 1: Simulation results for Example 1 when $p = 0.05$. Displayed in (a) - (c) are the true conditional VaR functions (in solid lines), the estimated WDKLL conditional VaR functions (in dashed lines), and the estimated NW conditional VaR functions (in dotted lines)for $n = 250$, 500 and 1000, respectively. Boxplots of the 500 MADE values for both WDKLL and NW estimation of the conditional VaR are plotted in (d).

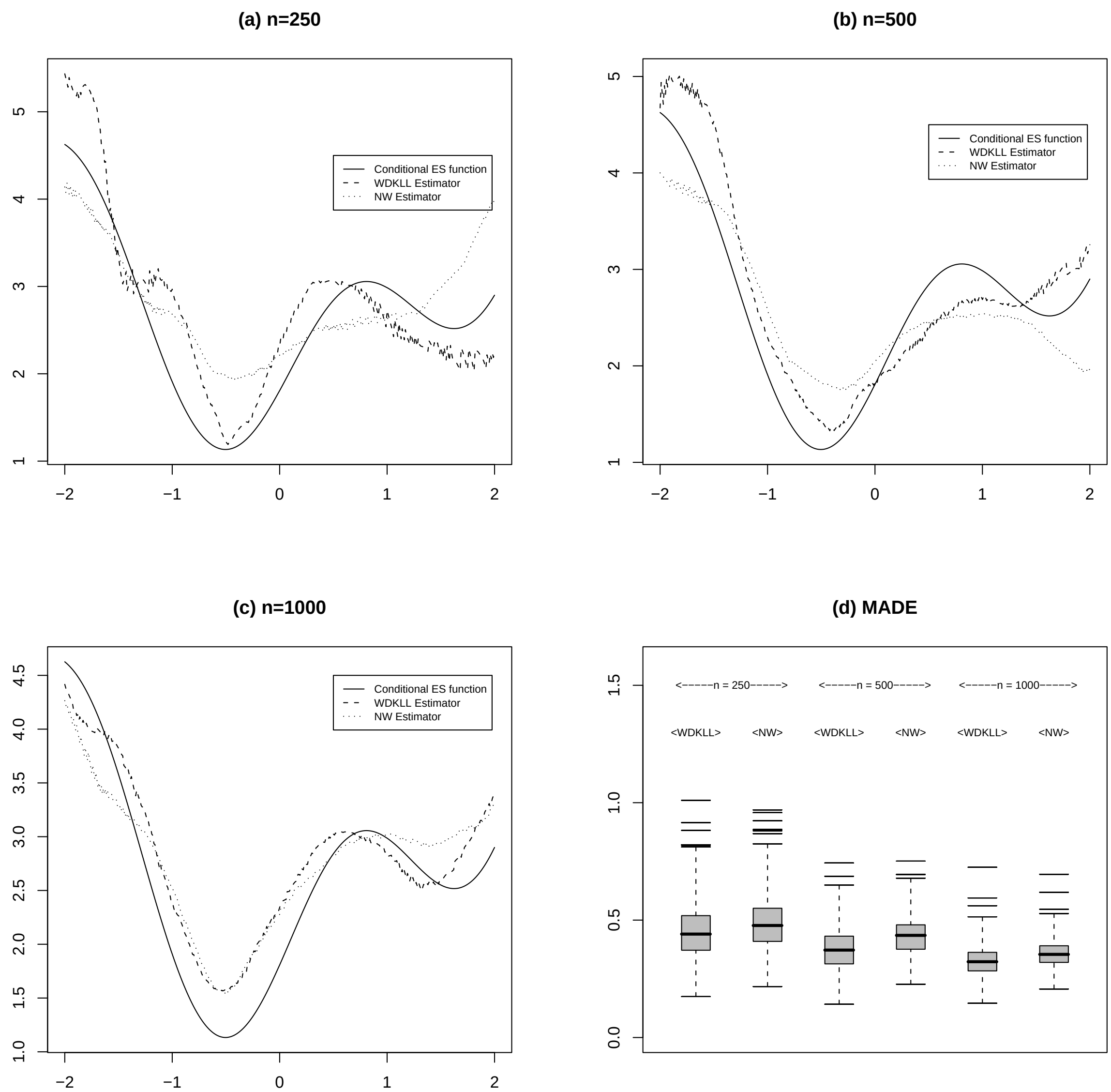

Figure 2: Simulation results for Example 1 when $p = 0.05$. Displayed in (a) - (c) are the true conditional ES functions (in solid lines), the estimated WDKLL conditional ES functions (in dashed lines), and the estimated NW conditional ES functions (in dotted lines) for $n = 250$, 500 and 1000, respectively. Boxplots of the 500 MADE values for both WDKLL and NW estimation of the conditional ES are plotted in (d).

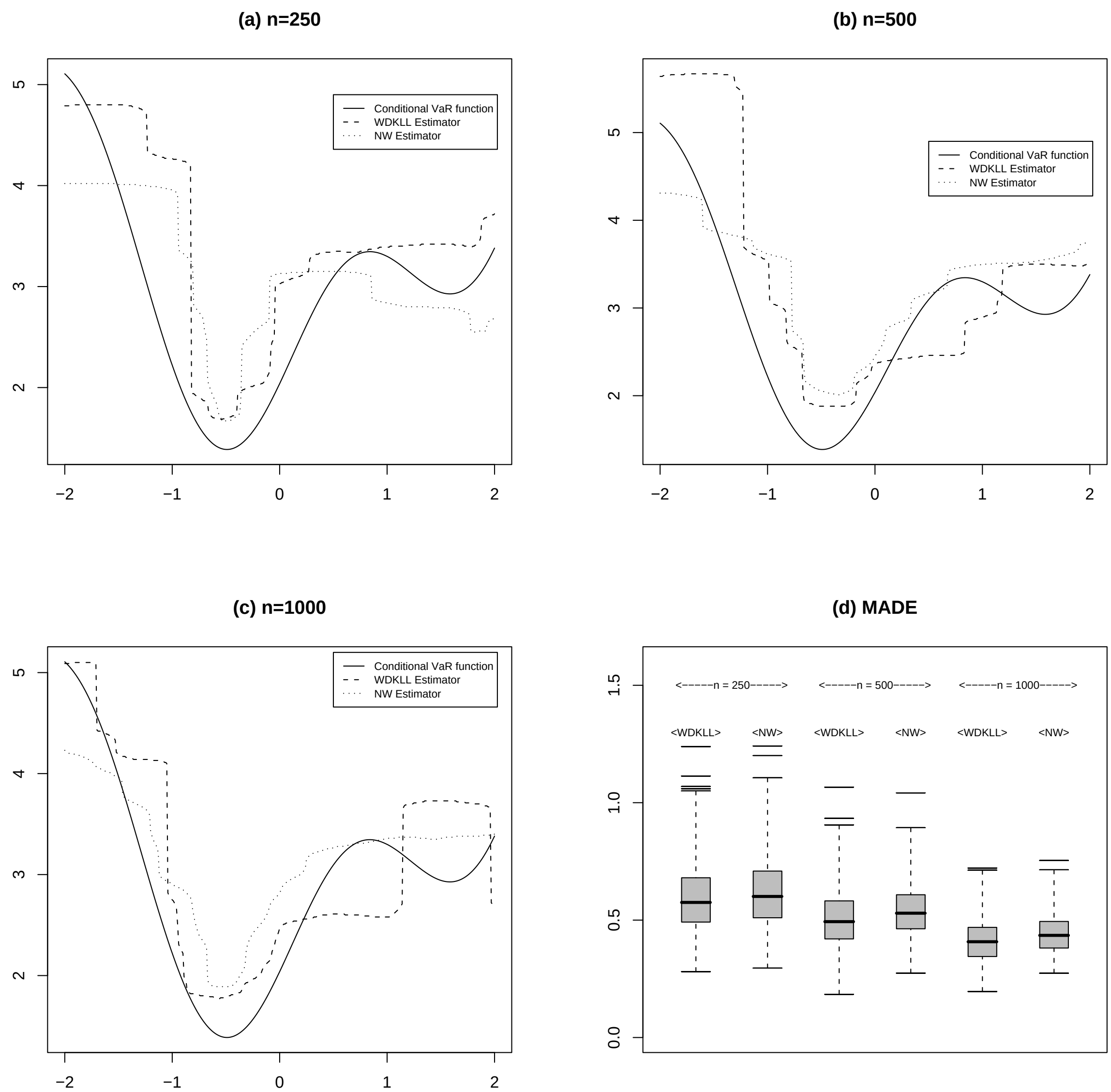

Figure 3: Simulation results for Example 1 when $p = 0.01$. Displayed in (a) - (c) are the true conditional VaR functions (in solid lines), the estimated WDKLL conditional VaR functions (in dashed lines), and the estimated NW conditional VaR functions (in dotted lines) for $n = 250$, 500 and 1000, respectively. Boxplots of the 500 MADE values for both WDKLL and NW estimation of the conditional VaR are plotted in (d).

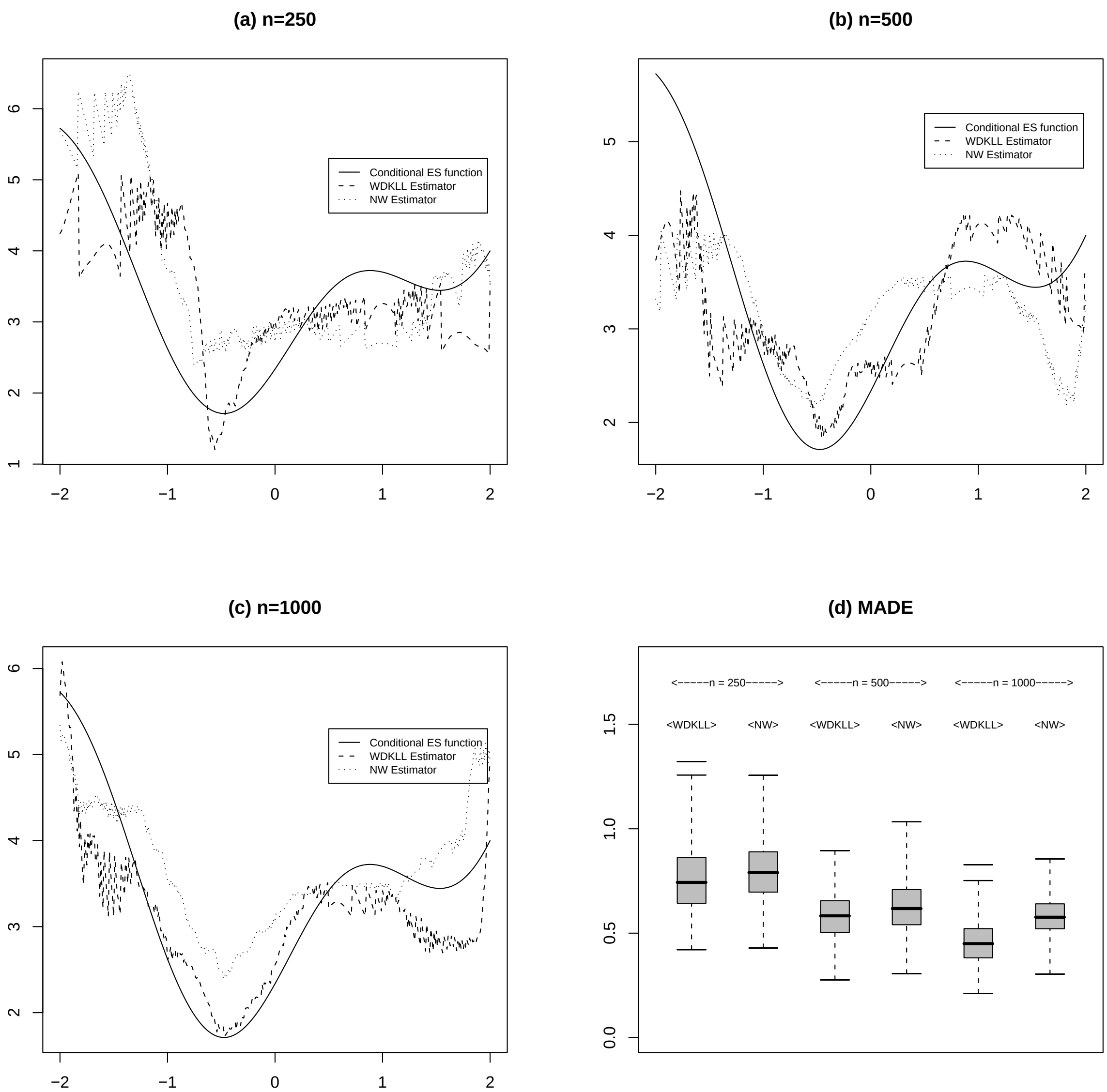

Figure 4: Simulation results for Example 1 when $p = 0.01$. Displayed in (a) - (c) are the true conditional ES functions (in solid lines), the estimated WDKLL conditional ES functions (in dashed lines), and the estimated NW conditional ES functions (in dotted lines) for $n = 250$, 500 and 1000, respectively. Boxplots of the 500 MADE values for both WDKLL and NW estimation of the conditional VaR are plotted in (d).

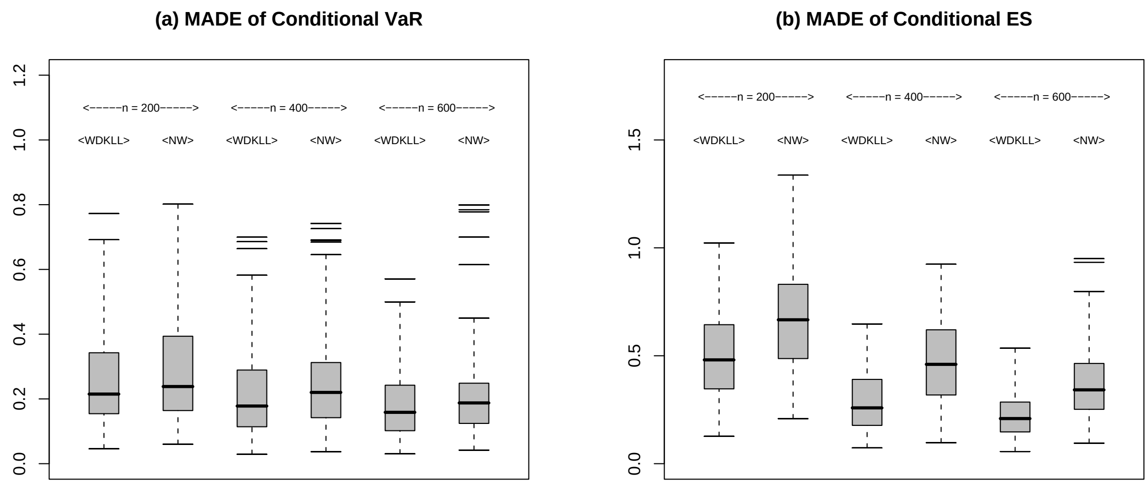

Figure 5: Simulation results for Example 2 when $p = 0.05$. (a) Boxplots of MADE for both WDKLL and NW conditional VaR estimates. (b) Boxplots of MADE for Both WDKLL and NW conditional ES estimates.

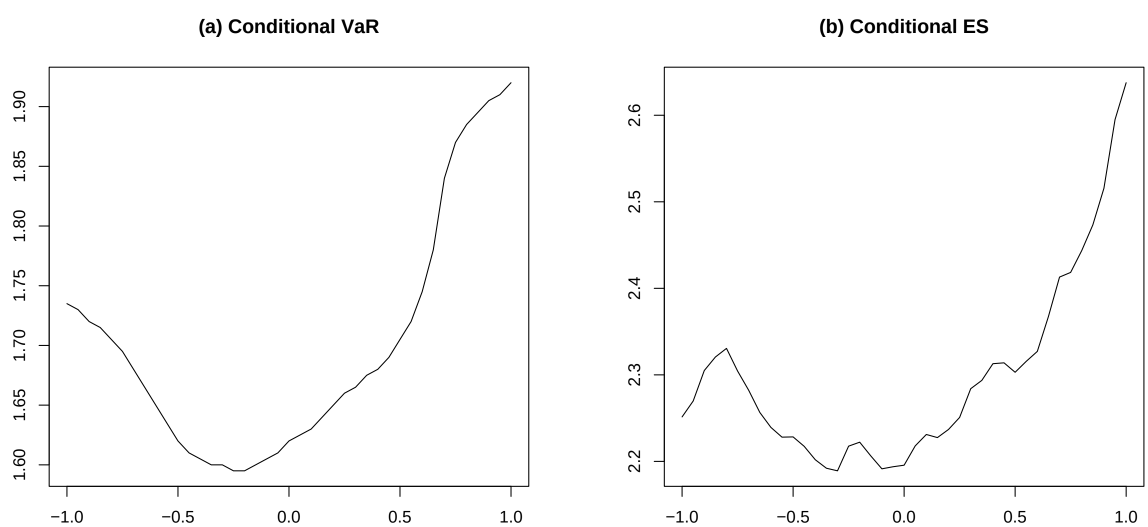

Figure 6: (a) 5% conditional VaR estimate for DJI index. (b) 5% conditional ES estimate for DJI index.

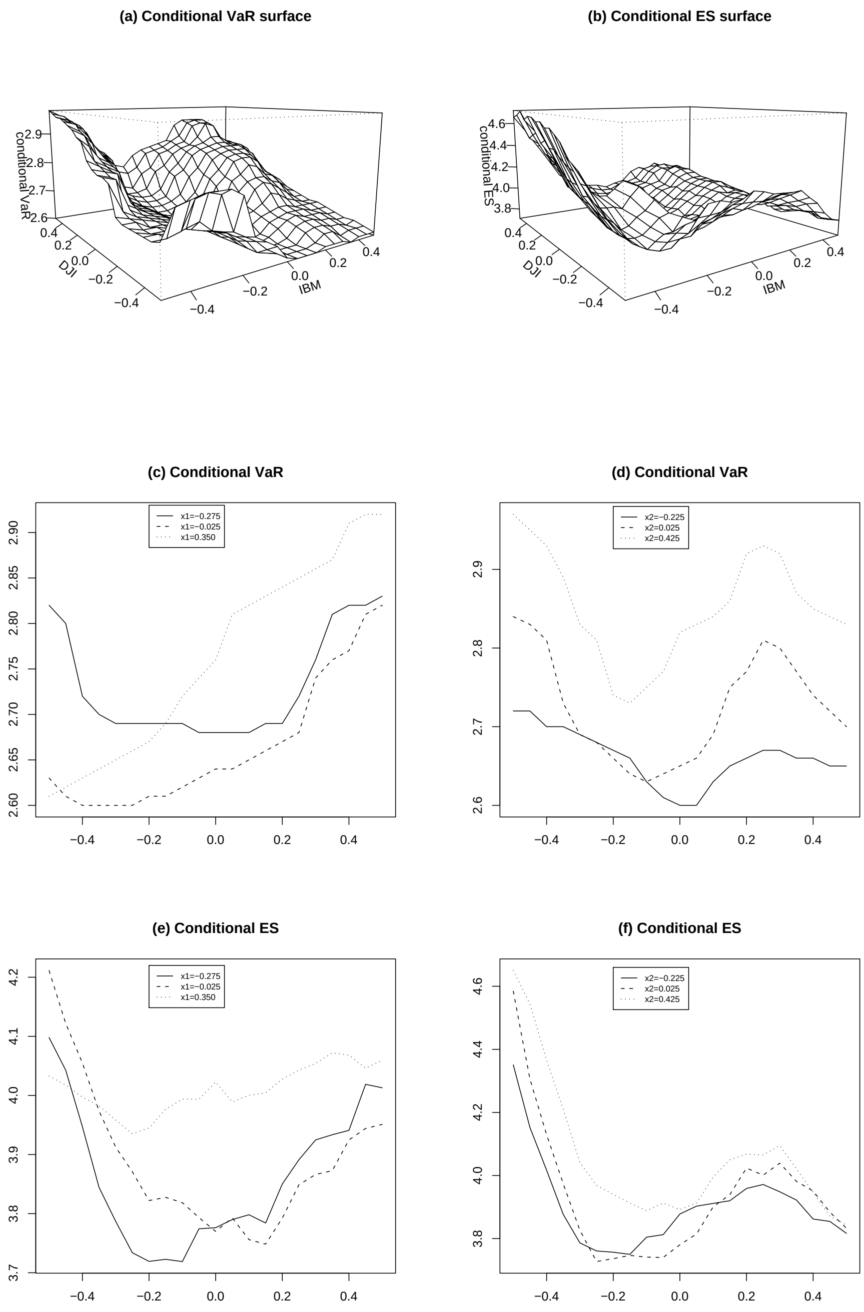

Figure 7: (a) 5% conditional VaR estimates for IBM stock returns. (b) 5% conditional ES estimates for IBM stock returns index. (c) 5% conditional VaR estimates for three different values of lagged negative IBM returns (-0.275, -0.025, 0.325). (d) 5% conditional VaR estimates for three different values of lagged negative DJI returns (-0.225, 0.025, 0.425). (e) 5% conditional ES estimates for three different values of lagged negative IBM returns (-0.275, -0.025, 0.325). (f) 5% conditional ES estimates for three different values of lagged negative DJI returns (-0.225, 0.025, 0.425).