Econometric Theory, **29**, 2013, 673–698.
doi:10.1017/S0266466612000667

NONPARAMETRIC INFERENCE FOR CONDITIONAL QUANTILES OF TIME SERIES

KE-LI XU
Texas A&M University

This paper considers model-free hypothesis testing and confidence interval construction for conditional quantiles of time series. A new method, which is based on inversion of the smoothed empirical likelihood of the conditional distribution function around the local polynomial estimator, is proposed. The associated inferential procedures do not require variance estimation, and the confidence intervals are automatically shaped by data. We also construct the bias-corrected empirical likelihood, which does not require undersmoothing. Limit theories are developed for mixing time series. Simulations show that the proposed methods work well in finite samples and outperform the normal confidence interval. An empirical application to inference of the conditional value-at-risk of stock returns is also provided.

1. INTRODUCTION

Quantile regression (Koenker and Bassett, 1978), with median regression as a special case, has become popular among practitioners as an alternative modeling approach to the classical mean regression. It characterizes distributional impacts of the covariates on the dependent variable beyond the traditional average effects. The estimation of the quantile regression is relatively insensitive to outlier observations, thereby providing a robust measure of location; see Buchinsky (1998) and Koenker (2005) for recent surveys on this topic.

Parametric quantile regression assumes a priori a functional form (linear or nonlinear) of the conditional quantile function, and the valid estimation and inference rely on correct model specification. The nonparametric model, which is the focus of the current paper, provides a flexible alternative. While nonparametric point estimation has attracted substantial recent attention in the literature (see Chaudhuri, 1991; Koenker, Portnoy, and Ng, 1992; Yu and Jones, 1998; Cai, 2002; Li and Racine, 2008; and references therein), the issues on inferences remain much less studied despite their practical relevance. For example, the well-known

The author thanks the co-editor, two anonymous referees, and the participants of the third CIREQ time series conference at Montreal for helpful comments, and University of Alberta School of Business for partial financial support through a Canadian Utilities faculty award and Killam research grant. Address correspondence to Ke-Li Xu, 3063 Allen, 4228 TAMU, Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA; e-mail: keli.xu@econmail.tamu.edu.

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

measure of the market risk in financial risk management, value-at-risk (VaR) (see, e.g., Jorion, 2000), can be computed by evaluating a lower-tail quantile of the conditional distribution of portfolio returns, and it characterizes the maximum potential loss for a given probability and time horizon. In most cases, a portfolio manager is not only interested in a single estimate but also a safe margin to understand the estimation risk (Christoffersen and Gonçalves, 2005). The existing works on inference of the VaR, like that of Christoffersen and Gonçalves, as well as those of Chan, Deng, Peng, and Xia (2007) and Gao and Song (2008), focus on the zero mean parametric framework. The associated interval estimation procedures hinge on a decomposition of the volatilities and the innovations and also the specification of the volatility process, e.g., a linear autoregressive conditional heteroskedasticity (ARCH)-type model. As another example in which inference is relevant, parents may be interested in whether the weights of their children are higher than the 90% quantile in the growth chart to obtain an early warning of potential medical problems caused by overweight.

In this paper we consider fully nonparametric quantile regression models that do not assume any representation of the dependent variable of interest. The model-free conditional quantile estimates and the associated confidence intervals are derived directly from the nonparametrically estimated conditional cumulative distribution function (CDF). In such models, the conventional inferential method (e.g., Wald test or Wald confidence interval) is based on the asymptotic normality of the conditional quantile estimate. Although this method is computationally easy to implement, it suffers from several practical deficiencies. As a key step of the asymptotic normality-based inference, the estimation of the standard error involves the conditional density estimate, which could be very inaccurate and imprecise, for example at low or high quantiles due to the sparseness of the data on the extremal observations. As a result, the associated confidence interval could be seriously size-distorted and fluctuate widely in lengths. In this paper we introduce an alternative computationally inexpensive approach to construct confidence intervals for conditional quantiles that do not require variance estimation. We also drop the symmetry assumption of the normal confidence intervals.

The proposed method, which is readily extendible to the case of multiple covariates, is based on inversion of the empirical likelihood (EL, Owen, 1988, 1990) confidence interval (CI) for the conditional CDF. The idea is illustrated in Figure 1. The solid curve is the nonparametrically estimated conditional CDF of Intel stock returns when the observed lagged log return is zero; see Section 4.4 for more details. Suppose we are interested in 5% conditional quantile q. The point estimate $\widehat{q}$ is obtained as the log return at which the estimated conditional distribution is 0.05, which is displayed by the solid arrow in the figure. The 90% EL CI of the conditional distribution function at the estimated 5% quantile (cf. Section 2) is constructed with two endpoints, A and B. Then the EL CI of the 5% conditional quantile is just the collection of points at which the estimated CDFs fall in the interval $[A, B]$. The obtained confidence interval is displayed in the figure between two dashed arrows. For comparison, the Wald CI, which is

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

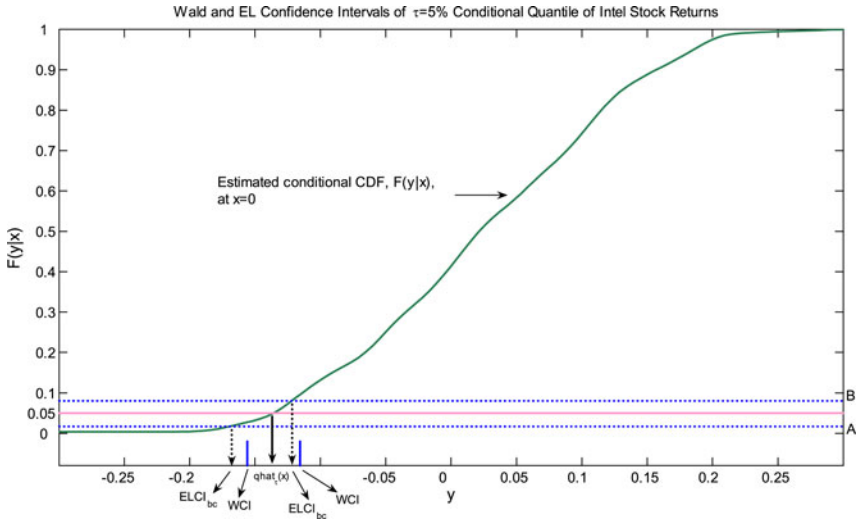


FIGURE 1. Comparison of the 90% empirical likelihood confidence interval (ELCI) and the Wald confidence interval (WCI) of the 5% conditional quantile of the Intel stock returns, i.e., $q_\tau(x)$, where $\tau = 0.05$ and $x = 0$. Numerical values: $\widehat{q}_\tau(x) = -0.1357$, $WCI = [-0.1560, -0.1155]$, $ELCI_{bc} = [-0.1681, -0.1196]$. The $ELCI_{bc}$ for the conditional CDF at $\widehat{q}_\tau(x)$ is $[0.0171, 0.0802]$.

predetermined to be symmetric around $\widehat{q}$, is also plotted by two vertical bars on the x-axis.

The empirical likelihood, as a nonparametric likelihood, has been proven as an efficient and robust way to construct confidence intervals for parametric and nonparametric models and is gaining popularity (e.g., Owen, 2001; Kitamura, 2006). This paper can be viewed as an extension of Chen and Hall (1993), who considered the EL CIs for the nonparametric unconditional quantiles in an independent and identically distributed (i.i.d.) setting, to the empirically more relevant nonparametric conditional quantile model given the existence of covariates. We also establish the validity of EL under weaker assumption on dependence of observations allowing for stationary mixing time series. Relatively large data sets are suitable for our applications, especially when tail quantiles are of interest and conditioned variables take extreme values due to the nonparametric nature of the proposed methods. Other recent applications of EL to conditional quantile models include Whang (2006) and Otsu (2008), who focused on the correctly specified parametric models.

The remainder of the paper is organized as follows. Section 2 introduces the double-smoothing local polynomial estimation of the conditional distribution function. The empirical likelihood for the conditional CDF, which we expect to be of independent interest, is derived, and an asymptotic distribution theory

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

is developed allowing for dynamically dependent data. We also introduce the bias-corrected empirical likelihood to relax the conventional undersmoothing assumption. The empirical likelihoods are inverted to form confidence intervals for conditional quantiles in Section 3. The implementation issues are discussed, and the finite sample performance is evaluated in Section 4 via simulations and an empirical application to the conditional value-at-risk of stock returns. Comparisons between empirical likelihood methods and Wald confidence intervals are also provided. Section 5 discusses the extension to the case of multidimensional covariates. The proofs of the main results are contained in the Appendix.

2. EMPIRICAL LIKELIHOOD FOR CONDITIONAL CDF FUNCTION

2.1. Local Polynomial Estimation

Let $\{X_t, Y_t : t = 1, \ldots, T\}$ be a two-dimensional stationary random process, where Y_t is the dependent variable of the main interest and X_t is the covariate. For clarity of exposition, we state our main results for a scalar covariate. The extension to the case of multiple covariates is discussed in Section 5. Denote $F(y|x)$ as the conditional CDF of Y_t at y given $X_t = x$, which is the focus of this section. Let $f(y|x)$ be the corresponding conditional density function and $\mu(\cdot)$ be the marginal density of X_t.

Estimation and inference of conditional CDF are important in many applications. In finance, popular risk measures like VaR and expected shortfall are essentially the estimates of functionals of conditional CDF. Consider a simple linear Gaussian AR-ARCH model for financial returns: $r_t = \varphi_0 + \varphi_1 r_{t-1} + (\varphi_2 + \varphi_3 r_{t-1}^2)^{1/2}\varepsilon_t$, where $\varepsilon_t \sim \mathcal{N}(0,1)$. Then the CDF of r_t conditional on r_{t-1} is $\mathcal{N}(\varphi_0 + \varphi_1 r_{t-1}, \varphi_2 + \varphi_3 r_{t-1}^2)$. The next-period VaR is obtained as a quantile of the normal distribution, and the expected shortfall is calculated as an expectation conditional on the loss exceeding the VaR, given the value on today's return. In this paper we focus on an alternative approach to estimating the conditional CDF that does not rely on any parametric specification (like linearity above) on mean, volatility, or other distributional aspects.

Suppose that G is a user-chosen CDF. The kernel-weighted local polynomial (Fan and Yao, 2003) double-smoothing estimator of $F(y|x)$ is defined as $\widehat{F}(y|x) = \widehat{\beta}_0$, where

$$\{\widehat{\beta}_j, j = 0, 1, \ldots, p\} = \arg\min_{\beta_j} \sum_{t=1}^{T} \left(G((y - Y_t)/h_0) - \textstyle\sum_{j=0}^{p} \beta_j (X_t - x)^j \right)^2 K_h(X_t - x), \qquad (1)$$

with $K_h(\cdot) = h^{-1}K(\cdot/h)$, for a kernel function $K(\cdot)$ and smoothing bandwidths h and h_0. When $h_0 = 0$, $G((y - Y_t)/h_0)$ reduces to $\mathbb{I}(Y_t \le y)$, where $\mathbb{I}(\cdot)$ is the indicator function, and $\widehat{F}(y|x)$ is the conventional single-smoothing estimator (Hall, Wolff, and Yao, 1999). By smoothing over both directions of conditioned

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

and conditioning variables, the resulting estimator not only appears closer to a distribution function, but also has more flexibility to reduce the mean squared error when the optimal bandwidths are used (Fan, Yao, and Tong, 1996; Yu and Jones, 1998; Hansen, 2004; Li and Racine, 2008; Gourieroux and Jasiak, 2010). The estimator $\widehat{F}(y|x)$ can be expressed in closed form. Define

$$\widehat{S}_k = T^{-1}\sum_{t=1}^{T}[(X_t - x)/h]^k K_h(X_t - x),$$

where $k = 0, 1, \ldots, 2p$ and $\widehat{S}$ is the $(p+1)$ by $(p+1)$ matrix with the (i, j)th element $\widehat{S}_{i+j-2}$. Let e_1 be the $(p+1)$-dimension vector $(1, 0, \ldots, 0)'$. Then $\widehat{F}(y|x)$ can be written as

$$\widehat{F}(y|x) = \frac{1}{Th}\sum_{t=1}^{T} W_p((X_t - x)/h)G((y - Y_t)/h_0), \tag{2}$$

where $W_p(u) = e_1' H^{-1}\widehat{S}^{-1}H^{-1}(1, uh, \ldots, (uh)^p)'K(u)$ with $H = \mathrm{diag}(1, h, \ldots, h^p)$ satisfying

$$\frac{1}{Th}\sum_{t=1}^{T} W_p\big((X_t - x)/h\big) = 1. \tag{3}$$

For example, when $p = 0$, $\widehat{F}(y|x)$ is the Nadaraya-Watson or local constant estimator and $W_p^{NW}((X_t - x)/h) = \big[(Th)^{-1}\sum_{t=1}^{T} K\big((X_t - x)/h\big)\big]^{-1} \cdot K((X_t - x)/h)$. When $p = 1$, $\widehat{F}(y|x)$ is the local linear estimator and $W_p^{LL}((X_t - x)/h) = \big[(Th)^{-1}\sum_{t=1}^{T} K((X_t - x)/h)\big(M_{T,2} - ((X_t - x)/h)M_{T,1}\big)\big]^{-1} \cdot K((X_t - x)/h)$ $[M_{T,2} - ((X_t - x)/h)M_{T,1}]$, where $M_{T,j} = (Th)^{-1}\sum_{t=1}^{T}[(X_t - x)/h]^j K((X_t - x)/h)$ for $j = 1, 2$.

The following conditions are assumed throughout the paper. Let C be a generic finite constant.

Assumption A1. The kernel function K is bounded with a bounded support.

Assumption A2. Let $f_{X_1,X_l}(x_1, x_l)$ be the joint density of X_1 and X_l for any $l \geq 2$. Assume that $0 < \mu(x_1) \leq C$, $0 < f_{X_1,X_l}(x_1, x_l) \leq C$.

Assumption A3. The process $\{X_t, Y_t\}$ is stationary mixing.[1] For ρ-mixing processes with the mixing coefficients $\rho(l)$, $\sum_{l=1}^{\infty}\rho(l) < \infty$. For α-mixing processes with the mixing coefficients $\alpha(l)$, $\sum_{l=1}^{\infty} l^a[\alpha(l)]^{1-2/\delta} < \infty$, where $\delta > 2$ and $a > 1 - 2/\delta$.

Assumption A4. For ρ-mixing and α-mixing processes, we assume, respectively, that there is a sequence of positive integers s_T such that $s_T \to \infty$, $s_T = o\big(\sqrt{Th}\big)$ and

$$\sqrt{T/h}\rho(s_T) \to 0, \qquad \sqrt{T/h}\alpha(s_T) \to 0, \qquad \text{as } T \to \infty.$$

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

Assumption A5. Assume that $F(y|\cdot)$ and $\mu(\cdot)$ are continuous at x, for any y.

Assumption A6. $Th \to \infty$ and $Th^{2p+3} = O(1)$, as $T \to \infty$.

Assumption A7. $Thh_0^4 = O(1)$, as $T \to \infty$.

Assumptions A1–A5 were used in Masry and Fan (1997). They are standard for the local polynomial estimator in the time series context. The mixing assumptions as in Assumption A3 restrict serial dependence of time series. They have been verified to be reasonably weak, allowing for a variety of widely used time series models including linear processes (like autoregressive moving average (ARMA) models), Markov processes (like ARCH models), generalized autoregressive conditional heteroskedasticity (GARCH) models, and stochastic volatility models. Carrasco and Chen (2002) provided conditions to ensure β-mixing (which implies α-mixing) for various GARCH models (excluding the IGARCH model), stochastic volatility models, and autoregressive conditional duration models. For a ρ-mixing process, which is necessarily α-mixing, we impose the less restrictive coefficient-decaying assumption. Assumption A4 relates convergence rates of mixing coefficients to bandwidth h. The smaller h is, the stronger the condition on the mixing coefficient that is imposed. This reflects the manner in which serial dependence is weakened by spatial smoothing; see Hart (1996) and Fan and Yao (2003, Sec. 5.3). Assumption A4 is not needed if the mixing coefficients decay exponentially. Assumptions A6 and A7 restrict the convergence rates of h and h_0, and they imply that $h, h_0 \to 0$ as $T \to \infty$.

Denote S as the $(p+1)$ by $(p+1)$ matrix with the (i,j)th element $\int u^{i+j-2}K(u)\mathrm{d}u$, S^* as the $(p+1)$ by $(p+1)$ matrix with the (i,j)th element $\int u^{i+j-2}K^2(u)\mathrm{d}u$ and c_p as the $p+1$ by 1 vector with the ith element $\int u^{p+2-i}K(u)\mathrm{d}u$. The following result gives the limit distribution of $\widehat{F}(y|x)$.

LEMMA 1. *Under Assumptions A1–A7, as* $T \to \infty$,

$$\sqrt{Th}\left[\widehat{F}(y|x) - F(y|x) - B(x,y)\right] \xrightarrow{d} \mathcal{N}(0, V_F(x,y)), \tag{4}$$

where

$$B(x,y) = \frac{h_0^2}{2}\frac{\mathrm{d}^2F(y|x)}{\mathrm{d}y^2}G_2 + \frac{h^{p+1}}{(p+1)!}\frac{\mathrm{d}^{p+1}F(y|x)}{\mathrm{d}x^{p+1}}e_1'S^{-1}c_p,$$

$$V_F(x,y) = F(y|x)[1-F(y|x)]e_1'S^{-1}S^*S^{-1}e_1/\mu(x),$$

and $G_2 = \int u^2 g(u)du$ *with* $g(\cdot) = G'(\cdot)$.

Note that higher-order local polynomial smoothing reduces the bias incurred by smoothing in the X_t direction (with a smaller order of h) but not the one in the Y_t direction (of the order h_0^2). The lemma can be compared with Yu and Jones (1998), Hall et al. (1999), and Cai (2002), who considered local linear estimation (or its variant) of the conditional CDF. Yu and Jones were interested in an i.i.d.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

setting, and Hall et al. (1999) only considered single-smoothing estimators. Li and Racine (2008) considered the case with multivariate X_t allowing for continuous and categorical covariates in an i.i.d. setting, and they focused on a local constant estimator.

2.2. Empirical Likelihood

The empirical likelihood for the conditional CDF can be built upon the estimating equation

$$\frac{1}{T}\sum_{t=1}^{T} W_p((X_t - x)/h)\left(G\big((y - Y_t)/h_0\big) - \widehat{F}(y|x)\right) = 0,$$

which is obtained by rearranging (2). Let $g_t(\theta) = W_p((X_t - x)/h)\left(G((y - Y_t)/h_0) - \theta\right)$. Define the EL ratio function for a candidate value θ as

$$\widetilde{R}(\theta) = \max_{(p_1,\ldots,p_T)} \left\{ \prod_{t=1}^{T} T p_t \Big| \sum_{t=1}^{T} p_t g_t(\theta) = 0,\ p_t \geq 0,\ \sum_{t=1}^{T} p_t = 1 \right\}. \tag{5}$$

The largest EL ratio is achieved at $\theta = \widehat{F}(y|x)$ with $p_t = 1/T$; i.e., $\widetilde{R}(\widehat{F}(y|x)) = 1$. The idea of an EL confidence interval, which will be defined below, is to collect the values of θ around $\widehat{F}(y|x)$ in the sense that the values have EL ratios close to unity. Define $R(\theta) = -2\log \widetilde{R}(\theta)$. The empirical likelihood construction is related to Chen and Qin (2000), which is concerned about nonparametric inference of a regression function. For the single-smoothing local linear estimator (i.e., $h_0 = 0$ and $p = 1$), (5) reduces to their EL in view of the expectation representation of a CDF.

The constrained optimization (5) can be solved via the method of Lagrange multiplier at $p_t = [T(1 + \lambda(\theta)g_t(\theta))]^{-1}$, where $\lambda(\theta)$ satisfies

$$\sum_{t=1}^{T} \frac{g_t(\theta)}{1 + \lambda(\theta)g_t(\theta)} = 0. \tag{6}$$

Thus

$$R(\theta) = 2\sum_{t=1}^{T} \log\left(1 + \lambda(\theta)g_t(\theta)\right). \tag{7}$$

To study the asymptotic behavior of the EL ratio, we strengthen Assumptions A6 and A7.

Assumption A6′. $Th \to \infty$ and $Th^{2p+3} \to 0$, as $T \to \infty$.

Assumption A7′. $Thh_0^4 \to 0$, as $T \to \infty$.

The undersmoothing conditions in A6′ and A7′ are used to eliminate the biases induced by smoothing X_t and Y_t. Assumption A6′ will be relaxed in Section 2.3.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

The following result provides the limit distribution of $R(\theta)$ when evaluated at $F(y|x)$.

THEOREM 1. *Under Assumptions A1–A5, A6′, and A7′,* $R(F(y|x)) \xrightarrow{d} \chi^2(1)$, *as* $T \to \infty$.

The empirical likelihood introduces a likelihood function in our purely nonparametric framework. Thus the confidence interval for $F(y|x)$ can be constructed just as the traditional parametric likelihood; i.e., the $100(1-\gamma)\%$ EL CI for $F(y|x)$ is $ELCI = \{\theta : R(\theta) < \chi^2_{1-\gamma}(1)\}$, where $\chi^2_{1-\gamma}(1)$ is the $(1-\gamma)$-quantile of $\chi^2(1)$ distribution.

2.3. Bias-Corrected Empirical Likelihood

The estimating function $g_t(\theta)$ has zero expectation at the true value $F(y|x)$ plus the biases, instead of at $F(y|x)$. The undersmoothing Assumptions A6′ and A7′ presume that the biases are negligible. Assumption A6′ is particularly restrictive, as it excludes the optimal bandwidth; see Section 4.1. We now introduce an empirical likelihood that implicitly corrects the bias in the X_t direction. Let $\widehat{g}_t(\theta) = W_p((X_t - x)/h)\left(G((y - Y_t)/h_0) - \theta - \left[\widehat{F}(y|X_t) - \widehat{F}(y|x)\right]\right)$. Define $\widetilde{R}_{bc}(\theta)$ and $R_{bc}(\theta)$ similarly to $\widetilde{R}(\theta)$ and $R(\theta)$, respectively, except that $\widehat{g}_t(\theta)$ is used in place of $g_t(\theta)$. The following result shows that Theorem 1 applies to $R_{bc}(\theta)$ without requiring Assumption A6′.

THEOREM 2. *Under Assumptions A1–A6 and A7′,* $R_{bc}(F(y|x)) \xrightarrow{d} \chi^2(1)$, *as* $T \to \infty$.

The confidence interval for $F(y|x)$ based on the bias-corrected empirical likelihood $R_{bc}(\theta)$ is then constructed as $ELCI_{bc} = \{\theta : R_{bc}(\theta) < \chi^2_{1-\gamma}(1)\}$.

3. EL CONFIDENCE INTERVALS FOR CONDITIONAL QUANTILES

The $\tau-$quantile of the conditional CDF $F(y|x)$ is defined as $q_\tau(x) = F^{-1}(\tau|x) = \inf\{y : F(y|x) \geq \tau\}$. We follow Yu and Jones (1998) and Li and Racine (2008) and estimate $q_\tau(x)$ by $\widehat{q}_\tau(x) = \widehat{F}^{-1}(\tau|x)$. The empirical likelihood confidence interval for $q_\tau(x)$ is thus obtained by inverting the EL confidence interval (*ELCI* or $ELCI_{bc}$) for $F(\tau|x)$. The steps of implementation are given below.

Algorithm 1 (EL CI for $q_\tau(x)$)

1. Evaluate $\widehat{F}(y|x)$ for a sufficiently fine partition of the range of Y, denoted as $\{y_b, b = 1, \ldots, B\}$.
2. Obtain the estimate of the $\tau-$quantile of $F(y|x)$as $\widehat{q}_\tau(x) = \min_{1 \leq b \leq B}\{y_b : \widehat{F}(y_b|x) \geq \tau\}$.
3. Form the $100(1-\gamma)\%$ EL confidence interval (*ELCI* or $ELCI_{bc}$) around $\widehat{F}(\widehat{q}_\tau(x)|x)$ as $\left[\widehat{F}^L, \widehat{F}^U\right]$.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

4. Obtain the estimates of the $\widehat{F}^L-$quantile and $\widehat{F}^U-$quantile of $F(y|x)$ as $\widehat{q}_\tau^L(x) = \min_{1\le b\le B}\{y_b : \widehat{F}(y_b|x) \ge \widehat{F}^L\}$ and $\widehat{q}_\tau^U(x) = \min_{1\le b\le B}\{y_b : \widehat{F}(y_b|x) \ge \widehat{F}^U\}$, respectively, where $\{\widehat{F}(y_b|x), b = 1,\dots,B\}$ were calculated in Step 1.
5. The EL confidence interval of $q_\tau(x)$ is formed as $\left[\widehat{q}_\tau^L(x), \widehat{q}_\tau^U(x)\right]$.

The following result follows directly from Theorems 1 and 2.

THEOREM 3. *The EL confidence interval described above contains the true value of* $q_\tau(x)$ *with probability* $100(1-\gamma)\%$, *as* $T \to \infty$.

Alternatively, EL CI for conditional quantile $q_\tau(x)$ can be constructed directly from the EL ratio function [$\widetilde{R}(\theta)$ or $\widetilde{R}_{bc}(\theta)$] for conditional CDF. Consider $\widetilde{R}(\theta)$, for example. Rewrite $\widetilde{R}(\theta)$ in (5) as $\widetilde{R}(\theta, x, y)$. Similarly to the proof of Theorem 1, we can show that $-2\log\widetilde{R}(\tau, x, y) \xrightarrow{d} \chi^2(1)$, as $T \to \infty$. Then the EL CI for $q_\tau(x)$ is obtained as $\{y : -2\log\widetilde{R}(\tau, x, y) < \chi^2_{1-\gamma}(1)\}$. This method does not require the point estimate of $q_\tau(x)$ and is an extension of Chen and Hall (1993) to the conditional model. We considered such constructed confidence intervals in simulations and found little difference with the ones formed by inversion of CIs for conditional CDF.

4. IMPLEMENTATION AND NUMERICAL EXAMPLES

4.1. Bandwidth Selection

Important issues in practical implementation include the selection of smoothing parameters h and h_0. For the most popular member of (1), the local linear estimator (viz. $p = 1$), Assumptions A6 and A7′ assure that h and h_0 should satisfy $h \sim T^{\xi}$, where $\xi \in (-1, -1/5]$, and $h_0 \sim T^{\xi_0}$, where $\xi_0 \in (-\infty, -(1+\xi)/4)$, for the bias-corrected EL to be valid. The usual optimal data-driven bandwidth selectors when estimating conditional quantiles can be used to obtain confidence intervals. Yu and Jones (1998) considered the local linear conditional quantile estimator in an i.i.d. setting and derived the integrated mean squared error (IMSE). They showed that the optimal choices of h_0 and h satisfy $h_0 \sim T^{-2/5}$ and $h \sim T^{-1/5}$; see also Li and Racine (2007, p. 186; 2008). The optimal rates satisfy our Assumptions A6 and A7′. Li and Racine (2008) considered the conditional quantile estimator via inversion of the conditional CDF estimator, as in our paper, and adopted the bandwidth selector of Hall, Racine, and Li (2004) to jointly select h and h_0. Their least-squares-cross-validation-(LSCV)-based procedure is more computationally involved than one in estimating the conditional mean since the integration of mean squared error (MSE) is over both dimensions of X and Y. The resulting bandwidths are not of the optimal rates, and they proposed adjustments.

Another useful method leads to the ad hoc rule-of-thumb (ROT) bandwidths $h_0^{ROT} = \widehat{\sigma}(Y)T^{-2/5}$ and $h^{ROT} = \widehat{\sigma}(X)T^{-1/5}$, where $\widehat{\sigma}(\cdot)$ means sample standard

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

deviation. They are much simpler to compute and are found to perform well in finite samples in both estimation and inference. In the simulations of Li and Racine (2008), under a nonlinear multiple quantile regression design, ad hoc bandwidths for continuous variables have comparable performance to the LSCV bandwidths, especially under heteroskedasticity. LSCV tends to select bandwidths that vary widely (over realizations), and the resulting estimator has relatively large variance. Here we emphasize that the usual IMSE-minimizing bandwidths, as those mentioned above, are designed to optimize estimation performance and may not be optimal for inference purposes. Optimal bandwidths for inference involve balancing of size and power, and a thorough treatment in a general context, deserve further research.

4.2. Wald Confidence Intervals for Conditional Quantiles

The conventional Wald confidence interval can be constructed if an estimate of the asymptotic variance is available. Under the assumptions of Theorem 1, it follows that by an application of the delta method to (4),

$$\sqrt{Th}[\widehat{q}_\tau(x) - q_\tau(x)] \xrightarrow{d} \mathcal{N}(0, V_\tau(x)),$$

where $V_\tau(x) = \tau(1-\tau)e_1' S^{-1} S^* S^{-1} e_1 / [f^2(q_\tau(x)|x)\mu(x)]$, noting that $\mathrm{d}q_\tau(x)/\mathrm{d}x = [f(q_\tau(x)|x)]^{-1}$ and $\tau = F(q_\tau(x)|x)$. The asymptotic variance $V_\tau(x)$ involves several unknown quantities that need to be separately estimated. The estimated asymptotic variance is

$$\widehat{V}_\tau(x) = \frac{\tau(1-\tau)e_1' S^{-1} S^* S^{-1} e_1}{\widehat{f}^2(\widehat{q}_\tau(x)|x) \cdot \widehat{\mu}(x)},$$

where $\widehat{f}(\widehat{q}_\tau(x)|x) = \left(\sum_{t=1}^T K_h(X_t - x)\right)^{-1} \sum_{t=1}^T K_h(X_t - x) g_{h_0}(Y_t - \widehat{q}_\tau(x))$ and $\widehat{\mu}(x) = T^{-1} \sum_{t=1}^T K_h(X_t - x)$ with $g(\cdot) = G'(\cdot)$ and $g_{h_0}(\cdot) = g(\cdot/h_0)/h_0$. Note that h_0 has to be chosen large enough so that $\widehat{f}(\widehat{q}_\tau(x)|x)$ is not zero, especially when τ is small. The Wald confidence interval for $q_\tau(x)$ at $100(1-\gamma)\%$ significance level is formed as $[\widehat{q}_\tau(x) - z_{1-\gamma/2}\sqrt{\widehat{V}_\tau(x)}, \widehat{q}_\tau(x) + z_{1-\gamma/2}\sqrt{\widehat{V}_\tau(x)}]$, where $z_{1-\gamma/2}$ is the $(1-\gamma/2)-$quantile of the standard normal distribution.

The Wald confidence intervals for conditional quantiles are subject to at least two criticisms. First, they are predetermined to be symmetric around the point estimates, and thus the shapes of the intervals are not data-driven. Second, their qualities rely strongly on how well the variances are estimated. The denominator of $\widehat{V}_\tau(x)$ is usually poorly estimated when τ is small (as in the VaR application) and/or when x is far from the majority of data due to the sparseness of local observations. Also, in these cases large bandwidths h and h_0 (probably larger than those used in the point estimate $\widehat{q}_\tau(x)$) should be used to avoid overly long confidence intervals. We will see numerical examples of this point in the next sections.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

4.3. Simulations

We present simulation evidence to illustrate the finite sample performance of the proposed methods and compare with the conventional Wald approach. Samples of $T+201$ observations are generated from two data generating processes (DGPs):

$$\begin{aligned}
&\text{DGP 1}: r_t = 0.0174 + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0,1),\\
&\qquad \sigma_t^2 = 0.0126 + 0.3526(r_{t-1} - 0.0174)^2;\\
&\text{DGP 2}: r_t = 0.0221 + \sigma_t \varepsilon_t, \quad \varepsilon_t \sim t(6),\\
&\qquad \sigma_t^2 = 0.0134 + 0.2492(r_{t-1} - 0.0221)^2,
\end{aligned}$$

with the initial observation set as zero. For each sample we discard the first 200 observations to eliminate the start-up effects. We set $T = 371$ to match the sample size of the empirical application analyzed in Section 4.4, and the coefficients are obtained by fitting the linear ARCH models to the data (Tsay, 2005, Sec. 3.4.4). The ARCH effect is reduced when the heavy-tailed innovation distribution is used.[2] Let $Y_t = r_t$ and $X_t = r_{t-1}$, and we are interested in the $\tau \in \{3\%, 5\%\}$ conditional quantile of the distribution of Y_t given $X_t = x$, where $x \in \{0, 0.08\}$.

We only consider the local linear estimation ($p = 1$), and in this case the term $e_1' S^{-1} S^* S^{-1} e_1$ in $\widehat{V}_\tau(x)$ reduces to $\int K^2(u)\mathrm{d}u$. The kernels used are the standard normal distribution function G and the Epanechnikov kernel $K(u) = 0.75(1-u^2)\mathbb{I}_{[-1,1]}$. For illustration, we consider the bandwidths $h_0 = h_0^{ROT}$ and $h = h(j) = 1.5^{j-2} h^{ROT}$, where $j \in \{1,2,3\}$ and h_0^{ROT}, h^{ROT} are as defined above, and the same bandwidths are used to construct Wald CIs and EL CIs for fair comparison. The number of replications is 1,000, and the nominal size of the confidence interval is 90%.

4.3.1. Simulation Results. The results are summarized in Tables 1, 2, and 3 for two DGPs. The point estimates of conditional quantiles are quite close to the true values with acceptable standard errors. Standard errors are larger when the risk level is more extreme ($\tau = 3\%$), the conditioned variable takes value farther away from the majority of the data ($x = 0.08$), or the data have heavier tails (DGP 2). Not surprisingly, in these cases the CIs are generally longer using two types of methods analyzed below. The well-known smoothing bandwidth effects, the shorter bandwidth the larger the estimation standard error and thus the longer the CI, are also clearly identified in our results.

Wald confidence intervals are generally under-covered. The under-coverage is more severe for 3% VaR than for 5% VaR; taking DGP 1, for example, the nominal coverage rates are between 63% and 78% for 3% VaR, and between 71% and 83% for 5% VaR. We find that the estimated variances that are used in Wald CIs tend to be smaller than the actual standard errors of the point estimates (see the last rows of the tables, where we report the medians of the squared roots of variance estimates over replications), which partly explains the under-coverage of Wald CIs.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

TABLE 1. DGP 1: Point estimates and 90% Wald CIs and EL CIs for conditional quantile $q_\tau(x)$ where $\tau = 0.05$

	$x = 0.00$			$x = 0.08$		
h	$h(1)$	$h(2)$	$h(3)$	$h(1)$	$h(2)$	$h(3)$
		$\tau = 5\%$				
$q_\tau(x)$	−.168	−.168	−.168	−.177	−.177	−.177
Mean $\widehat{q}_\tau(x)$	−.166	−.168	−.170	−.177	−.177	−.180
Std. Dev.	.030	.025	.021	.034	.028	.023
EL CI Cov.%	83.2%	83.1%	85.4%	81.1%	84.9%	87.9%
EL_{bc} CI Cov.%	89.5%	89.4%	92.2%	87.8%	91.9%	92.9%
WCI Cov.%	76.1%	78.5%	81.8%	71.6%	77.6%	82.9%
EL Len. Mean	.095	.077	.066	.101	.086	.074
EL Len. Med.	.088	.073	.063	.091	.081	.069
EL Len. 75%	.118	.095	.080	.125	.105	.088
EL Len. 90%	.155	.121	.096	.167	.134	.110
EL_{bc} Len. Mean	.107	.085	.071	.116	.097	.078
EL_{bc} Len. Med.	.095	.079	.068	.102	.088	.073
EL_{bc} Len. 75%	.129	.103	.086	.139	.114	.093
EL_{bc} Len. 90%	.170	.130	.104	.191	.149	.117
WCI Len. Mean	.209	.285	.110	.224	.200	.120
WCI Len. Med.	.075	.069	.062	.074	.075	.065
WCI Len. 75%	.129	.112	.096	.121	.123	.106
WCI Len. 90%	.274	.212	.173	.266	.242	.176
WCI Len. Inf.%	7.5%	4.7%	1.0%	10.3%	3.7%	1.4%
Var. Est. Med.	.023	.021	.019	.023	.023	.020

Notes: For point estimates, we report the mean and standard deviation over 1,000 replications. For confidence intervals, we report coverage rates (Cov.) and the mean, median, 75% quantiles, and 90% quantiles of the lengths (Len.). The second-to-last row reports the percentage of Wald CIs with infinite length. The last row reports the median of estimated variances that are used in Wald CIs.

Empirical likelihood CIs are more accurate than Wald CIs with milder undercoverage. After bias correction, they are even more accurate, with the coverage rates very close to the nominal level. In Figure 2 we plot all confidence intervals for 5% VaR in 1,000 realizations generated by Wald method [Panel (a)] and empirical likelihood (bias corrected) [Panel (b)] when $x = 0$ and $h = h(2)$. They are sorted in the ascending order of the lower bounds of CIs. We can clearly see the cases of rejections when the true value is smaller than the lower bounds of CIs (at the right ends of the plots) and when it is larger than the upper bounds of CIs. The coverage rates are actually 78.5% and 89.4% for Wald CIs and EL CIs, respectively. We find that the improvements of EL CIs over Wald CIs appear to be robust to the smoothing bandwidths used, risk levels, the values of the conditional variable, and DGPs.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

TABLE 2. DGP 1: Point estimates and 90% Wald CIs and EL CIs for conditional quantile $q_\tau(x)$ where $\tau = 0.03$

	$x = 0.00$			$x = 0.08$		
h	$h(1)$	$h(2)$	$h(3)$	$h(1)$	$h(2)$	$h(3)$
		$\tau = 3\%$				
$q_\tau(x)$	−.195	−.195	−.195	−.205	−.205	−.205
Mean $\widehat{q}_\tau(x)$	−.194	−.195	−.196	−.205	−.205	−.208
Std. Dev.	.034	.029	.027	.041	.035	.029
EL CI Cov.%	82.7%	85.0%	81.6%	78.5%	81.6%	84.8%
EL_{bc} CI Cov.%	87.3%	91.0%	90.6%	85.8%	86.0%	90.2%
WCI Cov.%	69.2%	75.0%	77.8%	63.5%	71.0%	75.8%
EL Len. Mean	.100	.092	.076	.105	.099	.085
EL Len. Med.	.092	.085	.071	.096	.093	.079
EL Len. 75%	.129	.114	.097	.139	.130	.104
EL Len. 90%	.165	.149	.119	.183	.169	.141
EL_{bc} Len. Mean	.132	.105	.091	.139	.121	.099
EL_{bc} Len. Med.	.105	.093	.079	.115	.103	.088
EL_{bc} Len.75%	.155	.127	.110	.173	.149	.118
EL_{bc} Len. 90%	.254	.166	.136	.266	.200	.156
WCI Len. Mean	.206	.157	.143	.397	.150	.246
WCI Len. Med.	.075	.073	.070	.069	.070	.073
WCI Len. 75%	.155	.114	.107	.114	.112	.116
WCI Len. 90%	.347	.234	.191	.246	.221	.225
WCI Len. Inf.%	18.1%	10.0%	3.0%	22.7%	12.2%	7.6%
Var. Est. Med.	.023	.022	.021	.021	.021	.022

Notes: For point estimates, we report the mean and standard deviation over 1,000 replications. For confidence intervals, we report coverage rates (Cov.) and the mean, median, 75% quantiles, and 90% quantiles of the lengths (Len.). The second-to-last row reports the percentage of Wald CIs with infinite length. The last row reports the median of estimated variances that are used in Wald CIs.

We also check the lengths of various confidence intervals. Tables 1, 2, and 3 report the average, median, 75% quantile, and 90% quantile of the lengths of Wald and EL CIs over realizations. Although in median, Wald CIs are shorter than two types of EL CIs, they can be absurd. We find very long Wald CIs and ones with infinite lengths in simulations. For a particular case when $x = 0$ and $h = h(2)$, this literally can be seen in Figure 2, in which the first few sorted Wald CIs are very long and EL CIs exhibit much less fluctuation in lengths. Eleven too-wide Wald CIs are not displayed in the figure. The percentages of realizations that give infinitely long Wald CIs are reported in the second-to-last rows of the tables. Wald CIs with infinite length are more likely to happen when the bandwidth h or h_0 is small so that the conditional density estimate $\widehat{f}(\widehat{q}_\tau(x)|x)$ is close to zero. When $h = h(1)$ is used, about 7%–10% infinitely long Wald CIs would emerge

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

TABLE 3. DGP 2: Point estimates and 90% Wald CIs and EL CIs for conditional quantile $q_\tau(x)$ where $\tau = 0.05$

	$x = 0.00$			$x = 0.08$		
h	$h(1)$	$h(2)$	$h(3)$	$h(1)$	$h(2)$	$h(3)$
		$\tau = 5\%$				
$q_\tau(x)$	−.204	−.204	−.204	−.210	−.210	−.210
Mean $\widehat{q}_\tau(x)$	−.207	−.208	−.209	−.212	−.212	−.214
Std. Dev.	.047	.037	.031	.046	.042	.034
EL CI Cov.%	85.0%	83.1%	86.7%	83.3%	82.6%	85.7%
EL_{bc} CI Cov.%	90.2%	89.4%	91.4%	89.7%	90.0%	91.1%
WCI Cov.%	78.2%	78.5%	82.1%	69.8%	76.1%	81.7%
EL Len. Mean	.141	.116	.098	.150	.126	.104
EL Len. Med.	.122	.106	.093	.129	.115	.096
EL Len. 75%	.177	.141	.117	.185	.154	.124
EL Len. 90%	.240	.180	.147	.265	.206	.160
EL_{bc} Len. Mean	.163	.124	.104	.170	.139	.114
EL_{bc} Len. Med.	.137	.111	.098	.141	.123	.104
EL_{bc} Len. 75%	.192	.146	.124	.204	.164	.134
EL_{bc} Len. 90%	.283	.196	.153	.288	.226	.178
WCI Len. Mean	2.025	.187	.167	.372	.251	.318
WCI Len. Med.	.100	.094	.088	.098	.100	.091
WCI Len. 75%	.161	.154	.145	.156	.167	.143
WCI Len. 90%	.355	.280	.229	.311	.296	.266
WCI Len. Inf.%	8.7%	4.8%	2.1%	8.5%	4.6%	3.2%
Var. Est. Med.	.030	.029	.027	.030	.030	.028

Notes: For point estimates, we report the mean and standard deviation over 1,000 replications. For confidence intervals, we report coverage rates (Cov.) and the mean, median, 75% quantiles, and 90% quantiles of the lengths (Len.). The second-to-last row reports the percentage of Wald CIs with infinite length. The last row reports the median of estimated variances that are used in Wald CIs.

after 1,000 realizations with finite Wald CIs. Even when $h = h(3)$, there are still about 1%–3% infinitely long Wald CIs. The proportion of infinite Wald CIs is more remarkable for 3% VaR. In contrast, the EL CIs circumvent this difficulty by selecting the variance estimates implicitly via the internal mechanism of nonparametric likelihood. Bias-corrected EL CIs are usually longer than uncorrected ones due to more variation introduced by the bias-correction term.

4.4. Intel Stock Returns

In this section we consider an empirical application to stock returns of Intel Corporation. The data of monthly log stock returns from January 1973 to December 2003 are obtained from Tsay (2005). The first-order lagged log return is used

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

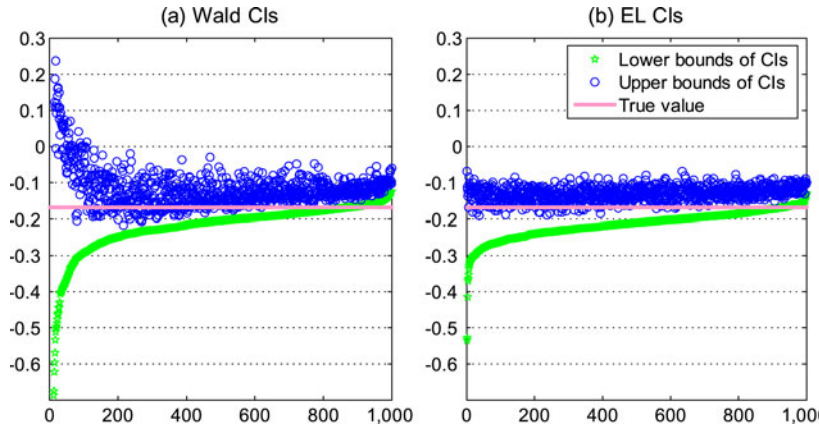


FIGURE 2. 1,000 realizations of Wald CIs (Panel a) and empirical likelihood CIs (bias corrected) (Panel b) for 5% VaR when $x = 0$. In each figure the CIs are sorted in ascending order of the lower bounds of CIs. The coverage rates are 78.5% and 89.4% for Wald CIs and EL CIs, respectively. The bandwidths $h_0 = h_0^{ROT}$ and $h = h^{ROT}$ are used. Eleven Wald CIs with too-small lower bounds are not displayed.

as the conditioned variable X_t. Implementation issues summarized in Sections 4.1 and 4.3 also apply here. In particular, $h_0 = h_0^{ROT}$ and $h = h^{ROT}$ are used. Suppose we are interested in the τ-quantile of the distribution of the log returns, which corresponds to (the negative of) the τ-level VaR, conditional on a value of the observed lagged log return.

Table 4 gives the values of the augmented Dickey-Fuller (ADF) and Elliott-Rothenberg-Stock (1996) DF-GLS unit root test statistics for the return and the squared return series.[3] It provides strong evidence for stationarity of the series. We plot the autocorrelation functions (ACFs) for two series in Figure 3. From these results we verify the weak dependence of the return series (the squared returns are more serially dependent but nevertheless not persistent), and thus the assumptions maintained in the paper are satisfied.

TABLE 4. Test statistics of augmented Dickey-Fuller (ADF) and Elliott-Rothenberg-Stock (1996) (ERS) DF-GLS unit root tests of Intel stock returns

	ADF unit root tests			ERS unit root tests	
	None	Constant	Trend	Constant	Trend
r_t	−12.8	−13.5	−13.5	−8.3	−8.5
r_t^2	−9.3	−11.5	−11.6	−4.8	−5.4
1% critical value	−2.58	−3.44	−3.98	−2.57	−3.48

Note: All are significant at the 1% level.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

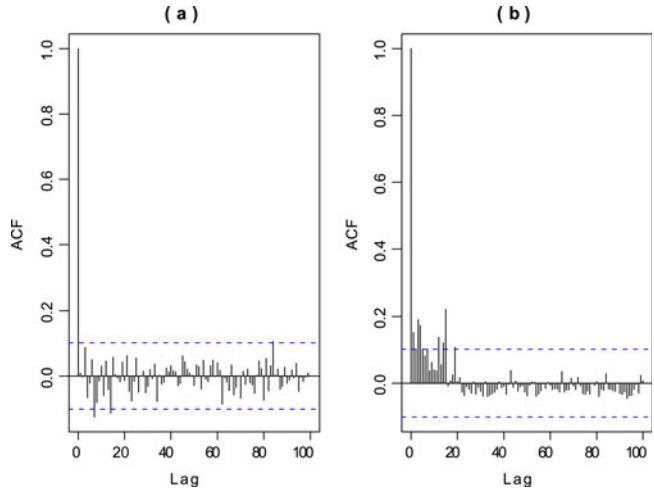


FIGURE 3. Autocorrelation functions (ACFs) of Intel stock returns (Panel a) and squared returns (Panel b).

We consider the confidence intervals for $\tau = 5\%$ quantiles at various values of the covariate. The results are shown in Figures 4 and 5. Four cases are drawn to illustrate different patterns of empirical likelihood CIs compared with Wald CIs where: (i) EL CIs are left-skewed ($x = 0$); (ii) EL CIs are right-skewed ($x = -0.04$); (iii) EL CIs are wider than Wald CIs ($x = -0.06$); or (iv) EL CIs are narrower than Wald CIs ($x = -0.07$). Automatic shaping of EL CIs can lead to meaningfully different consequences for financial decision-makers from those based on Wald CIs. For example, suppose we observe zero lagged log return and we then estimate the next-day 5% VaR around -0.136. The EL CI tells the risk

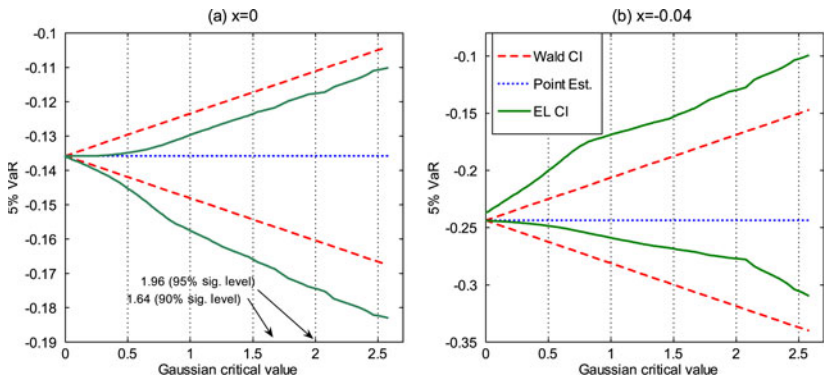


FIGURE 4. Confidence curve (Cook and Weisberg, 1990) for 5% value at risk (VaR) when (a) $x = 0$ and (b) $x = -0.04$. The (unconditional) VaR estimate by historical simulation is -0.167.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

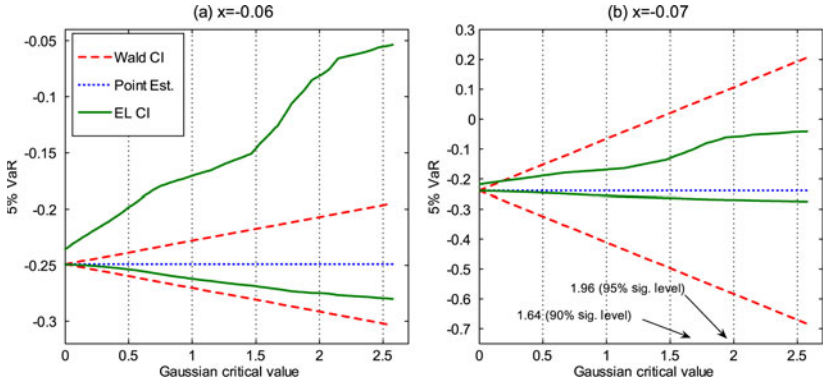


FIGURE 5. Confidence curve (Cook and Weisberg, 1990) for 5% value at risk (VaR) when (a) $x = -0.06$ and (b) $x = -0.07$. The (unconditional) VaR estimate by historical simulation is -0.167.

manager that the loss would more likely be above 0.136 (times the amount of the financial position) than below it, i.e., she is likely to suffer more losses, the information that the Wald CI fails to provide.

We illustrate the construction of EL CI further by focusing on the case when $x = 0$. As mentioned in the Introduction, Figure 1 shows how the EL CI for the conditional quantile is obtained by inverting the EL CI for the conditional CDF. Figure 6 plots the Lagrange multipliers $\lambda(\theta)$ and $\lambda_{bc}(\theta)$ and EL ratio statistics $R(\theta)$ and $R_{bc}(\theta)$ when computing the EL CI for the conditional CDF.

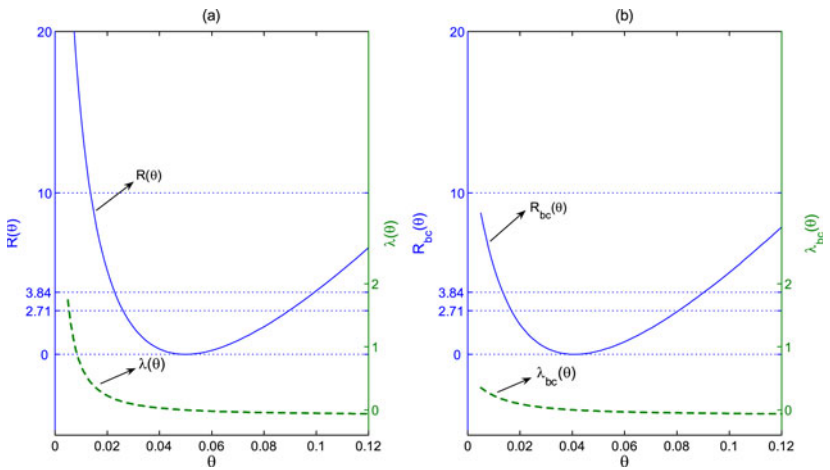


FIGURE 6. The double y-axes plot of the Lagrange multiplier λ and EL ratio R for conditional CDF $F(\widehat{q}_{0.05}(0)|0)$ of Intel stock returns. (a) EL; (b) bias-corrected EL.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

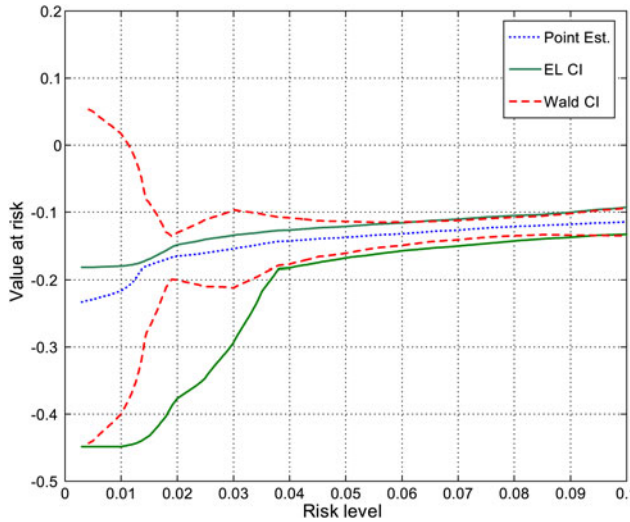


FIGURE 7. Point estimates and 90% confidence intervals for value at risk (VaR) at different risk levels when $x = 0$.

The point estimates of 5% VaRs are between -0.25 and -0.13 for the four representative values of the covariate considered. These can be compared with the unconditional quantile estimate by historical simulation, which is -0.167. In contrast to historical simulation, in which equal weights are assigned to all the historical observations, the conditional method we adopted here puts more weight on observations that bear more resemblance (in terms of the values of covariates) to the forecasting target, the return distribution next day.

We also consider VaR estimates at different risk levels τ ranging from 0.3% to 10%, together with 90% Wald CIs and EL CIs, as shown in Figure 7. The Wald CIs are of infinite length when the risk level is equal to or lower than 0.4%, and thus are not displayed in the figure. For instance, at the 0.4% level, the VaR is estimated as -0.217 with the 90% EL CI $[-0.449, -0.184]$, while the denominator of the Wald standard error is zero, and the Wald method does not generate a confidence interval of practical value. Figure 7 shows that at low risk levels, the left-skewness of EL CI is even more remarkable. Figure 7 also indicates that the most useful inference works for risk levels higher than around 4% in the current application, and a larger data set should be entertained to understand VaRs at lower tails if the model-free method is used.

5. EXTENSION TO THE CASE WITH MULTIPLE COVARIATES

The inference methods proposed in the paper extend to the case when there is more than one covariate. Multidimensionality of covariates does not change the number of model restrictions in the empirical likelihood (5), and the only

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

complication is that the model restrictions involve smoothing in higher dimension. To illustrate this, we adopt Li and Racine's (2008) framework, which allows for mixed continuous and discrete covariates, and we only consider the local linear estimation of the conditional CDF. Suppose $X_t = (X^c_{1t}, \ldots, X^c_{q_1 t}, X^d_{1t}, \ldots, X^d_{q_2 t})'$ is a $q-$dimensional vector of covariates with the first q_1 variables being continuous and the last q_2 variables being discrete,[4] where $q = q_1 + q_2$, and $x = (x^c_1, \ldots, x^c_{q_1}, x^d_1, \ldots, x^d_{q_2})'$. Let $\Phi_{h,\lambda}(X_t, x) = \Pi_{j=1}^{q_1} K_{h_j}(X^c_{jt} - x^c_j) \cdot \Pi_{j=1}^{q_2} L_{\lambda_j}(X^d_{jt}, x^d_j)$ be a generalized product kernel for mixed data, where $L_{\lambda_j}(X^d_{jt}, x^d_j) = \mathbb{I}(X^d_{jt} = x^d_j) + \lambda_j \mathbb{I}(X^d_{jt} \neq x^d_j)$ and λ_j's are smoothing bandwidths for discrete covariates. The local linear estimator of $F(y|x)$ is defined as $\widehat{F}(y|x) = \widehat{\beta}_0$, where

$$\left\{\widehat{\beta}_0, \widehat{\beta}_1\right\} = \min_{\beta_0, \beta_1} \sum_{t=1}^{T} \left(G((y - Y_t)/h_0) - \beta_0 - \beta_1'(X_t - x)\right)^2 \Phi_{h,\lambda}(X_t, x).$$

Rewrite $\widehat{F}(y|x)$ in an explicit form as

$$\widehat{F}(y|x) = \sum_{t=1}^{T} W_p^{LL}(X_t, x) \cdot G((y - Y_t)/h_0),$$

where

$$W_p^{LL}(X_t, x) = \left(\sum_{t=1}^{T} \Phi_{h,\lambda}(X_t, x) - M'_{T,1} M_{T,2}^{-1} \left(\sum_{t=1}^{T} (X_t - x)\Phi_{h,\lambda}(X_t, x)\right)\right)^{-1} \cdot \left(\Phi_{h,\lambda}(X_t, x) - M'_{T,1} M_{T,2}^{-1} (X_t - x)\Phi_{h,\lambda}(X_t, x)\right), \quad \textbf{(8)}$$

with $M_{T,1} = \sum_{t=1}^{T} (X_t - x)\Phi_{h,\lambda}(X_t, x)$ and $M_{T,2} = \sum_{t=1}^{T} (X_t - x)(X_t - x)' \Phi_{h,\lambda}(X_t, x)$. Li and Racine (2008) developed the asymptotic theory for the local constant (or Nadaraya-Watson) estimator of $F(y|x)$ for i.i.d. data, and the local linear estimator $\widehat{F}(y|x)$ derived here is expected to further reduce the bias of Li and Racine's (2008) estimator. Then the empirical likelihood in (5) (and also the bias-corrected empirical likelihood) can be computed using the multivariate local linear weights in (8), and confidence intervals of the conditional quantiles of the dependent variable given values of multivariate covariates therefore can be obtained. We also incorporate bias correction in a similar way as in Section 2.3. The smoothing bandwidths h_0 and h_j for a particular covariate are expected to be larger than the unicovariate case when no other covariate is included in the regression. Take, for instance, the case of no discrete covariates ($q = q_1$). Following the lines of Sections 2 and 3, if h_j's ($j = 1, \ldots, q$) are of the same order, they should satisfy $Th_1^q \to \infty$, $Th_1^{q+4} = O(1)$ and $Th_1^q h_0^4 \to 0$ for the bias-corrected EL to be valid. These conditions are met by the optimal bandwidths (Li and Racine, 2008) that minimize IMSE, $h_0 = O(T^{-2/(q+4)})$ and $h_1 = O(T^{-1/(q+4)})$. The presence of discrete covariates causes added complications, and further investigation of related issues is left for future research.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

NOTES

1. See, e.g., Davidson (1994) for the definitions of various mixing processes.
2. It can be verified that the simple ARCH models in DGPs 1 and 2 are strictly stationary (Nelson, 1990), $\alpha-$mixing with exponentially decaying coefficients (Basrak, Davis, and Mikosch, 2002) and have finite fourth moments, so the assumptions are satisfied.
3. For both unit root tests, we use Bayesian Information Criterion to select the numbers of lags.
4. For example, Tsay (2005) and Li and Racine (2008) considered several categorical variables and checked whether they are useful to forecast the VaR of stock returns next day. These categorical variables were generated to capture, e.g., the year-end effects and the possibility of panic selling on the previous trading day.

REFERENCES

Basrak, B., R.A. Davis, & T. Mikosch (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Applications* 99, 95–115.

Buchinsky, M. (1998) Recent advances in quantile regression models. *Journal of Human Resources* 33, 88–126.

Cai, Z. (2002) Regression quantiles for time series data. *Econometric Theory* 18, 169–192.

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

Chan, N.H., S.J. Deng, L. Peng, & Z. Xia (2007) Interval estimation for the conditional value-at-risk based on GARCH models with heavy tailed innovations. *Journal of Econometrics* 137, 556–576.

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

Chen, S.X. & P. Hall (1993) Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics* 21, 1166–1181.

Chen, S.X. & Y.S. Qin (2000) Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* 87, 946-953.

Christoffersen, P. & S. Gonçalves (2005) Estimation risk in financial risk management. *Journal of Risk* 7, 1–28.

Cook, R.D. & S. Weisberg (1990) Confidence curves in nonlinear regression. *Journal of American Statistical Association* 85, 544–551.

Davidson, J. (1994) *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press.

Elliott, G., T.J. Rothenberg & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Fan, J., Q. Yao, & H. Tong (1996) Estimation of conditional densities and sensitivity measures. *Biometrika* 83, 189–206.

Gao, F. & F. Song (2008) Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* 24, 1404–1428.

Gourieroux, C. & J. Jasiak (2010) Local likelihood density estimation and value at risk. *Journal of Probability and Statistics* 2010, article 754251.

Hall, P., J. Racine, & Q. Li (2004) Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association* 99, 1015–1026.

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

Hansen, B. (2004) Nonparametric Estimation of Smooth Conditional Distributions. Technical report, University of Wisconsin.

Hart, J.D. (1996) Some automated methods of smoothing time-dependent data. *Journal of Nonparametric Statistics* 6, 115–142.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

Jorion, P. (2000) *Value-at-Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill.
Kitamura, Y. (2006) Empirical likelihood methods in econometrics: Theory and practice. In R. Blundell, P. Torsten & W.K. Newey (eds.), *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*. Cambridge University Press.
Koenker, R. (2005) *Quantile Regression*. Cambridge University Press.
Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.
Koenker, R., S. Portnoy, & P. Ng (1992) Nonparametric estimation of conditional quantile functions. In Y. Dodge (ed.), *L1-Statistical Analysis and Related Methods,* pp. 217–229. Elsevier.
Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Li, Q. & J.S. Racine (2008) Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics* 26, 423–434.
Masry, E. & J. Fan (1997) Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24, 165–179.
Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.
Otsu, T. (2008) Conditional empirical likelihood estimation and inference for quantile regression models. *Journal of Econometrics* 142, 508–538.
Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.
Owen, A. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.
Owen, A. (2001) *Empirical Likelihood*. Chapman and Hall/CRC.
Tsay, R.S. (2005) *Analysis of Financial Time Series, 2nd Ed*. Wiley-Interscience.
Whang, Y.-J. (2006) Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory* 22, 173–205.
Yu, K. & M.C. Jones (1998) Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.

APPENDIX: Proofs

In this section we provide the proofs of the main results. We write law of iterated expectations, law of large numbers, and central limit theorem simply as LIE, LLN, and CLT respectively.

Proof of Lemma 1. By (2), the sampling error of $\widehat{F}(y|x)$ can be written as

$$\begin{aligned}\sqrt{Th}[\widehat{F}(y|x)-F(y|x)] &= \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} W_p\left(\tfrac{X_t-x}{h}\right)\left(G\left(\tfrac{y-Y_t}{h_0}\right)-F(y|x)\right)\\ &= \frac{1}{\sqrt{Th}}\sum_{t=1}^{T} W_p\left(\tfrac{X_t-x}{h}\right)(\mathbb{I}(Y_t\le y)-F(y|x))\\ &\quad+\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} W_p\left(\tfrac{X_t-x}{h}\right)\left(G\left(\tfrac{y-Y_t}{h_0}\right)-\mathbb{I}(Y_t\le y)\right)\\ &:= A_T+B_T. \end{aligned} \tag{A.1}$$

First we will show that

$$B_T=\sqrt{Th}h_0^2\frac{\mathrm{d}^2F(y|x)}{\mathrm{d}y^2}G_2/2+o_p\left(\sqrt{Th}h_0^2\right). \tag{A.2}$$

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

Note that $W_p(u) = e_1' \widehat{S}^{-1}(1, u, \ldots, u^p)' K(u)$ and $\widehat{S} \xrightarrow{p} \mu(x)S$. We can write

$$W_p(u) = \mu(x)^{-1}[1 + o_p(1)] \cdot K^*(u), \tag{A.3}$$

where $K^*(u) = e_1' S^{-1}(1, u, \ldots, u^p)' K(u)$ and the $o_p(1)$ part in (A.3) does not depend on u. The equivalent kernel $K^*(u)$ satisfies (Fan and Yao, 2003, pp. 237–238; hereafter FY)

$$\int K^*(u)\mathrm{d}u = 1 \tag{A.4}$$

and

$$\int K^*(u)^2 \mathrm{d}u = e_1' S^{-1} S^* S^{-1} e_1. \tag{A.5}$$

Let $Q_t = h^{-1} K^*\left(\frac{X_t - x}{h}\right)\left(G\left(\frac{y - Y_t}{h_0}\right) - \mathbb{I}(Y_t \le y)\right)$. Then $B_T = \sqrt{Th}\mu(x)^{-1}$ $[1 + o_p(1)]\left(T^{-1}\sum_{t=1}^T Q_t\right)$. So (A.2) holds if we show that

$$T^{-1}\sum_{t=1}^T Q_t = \mu(x)h_0^2 \frac{\mathrm{d}^2 F(y|x)}{\mathrm{d}y^2} G_2/2 + o_p\left(h_0^2\right). \tag{A.6}$$

We will use the facts

$$\mathrm{E}\left[G\left((y - Y_t)/h_0\right)|X_t\right] = F(y|X_t) + h_0^2 \cdot \int u^2 g(u)\mathrm{d}u \cdot \frac{\mathrm{d}^2 F(y|X_t)}{\mathrm{d}y^2}/2 + o_p\left(h_0^2\right), \tag{A.7}$$

$$\mathrm{E}\left[G^2((y - Y_t)/h_0)|X_t\right] = F(y|X_t) - h_0 \cdot 2\int uG(u)g(u)\mathrm{d}u \cdot \frac{\mathrm{d}F(y|X_t)}{\mathrm{d}y} + o_p\left(h_0^2\right), \tag{A.8}$$

$$\mathrm{E}[G((y - Y_t)/h_0)\mathbb{I}(Y_t \le y)|X_t] = h_0 F_0(y|X_t)\left[G(y)y - \int_{-\infty}^y \upsilon g(\upsilon)\mathrm{d}\upsilon\right] + O_p\left(h_0^2\right). \tag{A.9}$$

Among them, (A.7) and (A.8) have been shown by Li and Racine (2008). Now we show (A.9). The left-hand side of (A.9) equals $\int G((y - y_t)/h_0)\mathbb{I}(y_t \le y) f(y_t|X_t)\mathrm{d}y_t$, which can be rewritten as, after a change of variables $\upsilon = (y - y_t)/h_0$,

$$\begin{aligned}
&= -h_0 \int_{-\infty}^y G(\upsilon) f(y - \upsilon h_0|X_t)\mathrm{d}\upsilon = -\int_{-\infty}^y G(\upsilon)\mathrm{d}F(y - \upsilon h_0|X_t) \\
&= -G(y)F(y - yh_0|X_t) + \int_{-\infty}^y F(y - \upsilon h_0|X_t)g(\upsilon)\mathrm{d}\upsilon \\
&= -G(y)\left[F(y|X_t) - h_0 F_0(y|X_t)y + h_0^2 F_{00}(y|X_t)y^2/2\right] \\
&\quad + \int_{-\infty}^y \left[F(y|X_t) - h_0 F_0(y|X_t)\upsilon + h_0^2 F_{00}(y|X_t)\upsilon^2/2\right] g(\upsilon)\mathrm{d}\upsilon + o_p\left(h_0^2\right) \\
&= -G(y)\left[F(y|X_t) - h_0 F_0(y|X_t)y + h_0^2 F_{00}(y|X_t)y^2/2\right] \\
&\quad + \left[G(y)F(y|X_t) - h_0 F_0(y|X_t)\int_{-\infty}^y \upsilon g(\upsilon)\mathrm{d}\upsilon + h_0^2 F_{00}(y|X_t)\int_{-\infty}^y \upsilon^2 g(\upsilon)\mathrm{d}\upsilon/2\right] + o_p\left(h_0^2\right) \\
&= h_0 F_0(y|X_t)\left[G(y)y - \int_{-\infty}^y \upsilon g(\upsilon)\mathrm{d}\upsilon\right] + O_p\left(h_0^2\right).
\end{aligned}$$

Thus (A.9) holds.

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

By stationarity of $\{X_t, Y_t\}$,

$$\begin{aligned}\mathrm{E}\left(T^{-1}\textstyle\sum_{t=1}^{T} Q_t\right) = \mathrm{E}Q_t &\overset{\text{LIE}}{=} \mathrm{E}\left(h^{-1}K^*\left(\tfrac{X_t-x}{h}\right)\mathrm{E}\left[\left(G\left(\tfrac{y-Y_t}{h_0}\right)-\mathbb{I}(Y_t\le y)\right)|X_t\right]\right)\\ &\overset{\text{(A.7)}}{=} \mathrm{E}\left(h^{-1}K^*\left(\tfrac{X_t-x}{h}\right)\left[h_0^2\frac{\mathrm{d}^2F(y|X_t)}{\mathrm{d}y^2}G_2/2+o_p\left(h_0^2\right)\right]\right)\\ &\overset{\text{(A.4)}}{=} h_0^2\mu(x)\frac{\mathrm{d}^2F(y|x)}{\mathrm{d}y^2}G_2/2+o\left(h_0^2\right). \qquad \textbf{(A.10)}\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathrm{Var}(Q_t) &\overset{\text{(A.10)}}{=} \mathrm{E}Q_t^2+O\left(h_0^4\right)\\ &= \mathrm{E}h^{-2}K^{*2}\left(\tfrac{X_t-x}{h}\right)\left(G\left(\tfrac{y-Y_t}{h_0}\right)-\mathbb{I}(Y_t\le y)\right)^2+O\left(h_0^4\right)\\ &\overset{\text{LIE, (A.8) and (A.9)}}{=} \mathrm{E}\left[h^{-2}K^{*2}\left(\tfrac{X_t-x}{h}\right)(Ch_0+o_p(h_0))\right]+O\left(h_0^4\right)\\ &\overset{\text{(A.5)}}{=} O(h_0/h).\end{aligned}$$

By the stationarity of $\{X_t, Y_t\}$,

$$\mathrm{Var}\left(T^{-1}\textstyle\sum_{t=1}^{T} Q_t\right) = T^{-1}\mathrm{Var}(Q_1)+2T^{-1}\displaystyle\sum_{l=1}^{T-1}(1-l/T)\mathrm{Cov}(Q_1,Q_{l+1}).$$

Let $d_T\to\infty$ be a sequence of integers such that $d_T h\to 0$. Let

$$J_1=\sum_{l=1}^{d_T-1}|\mathrm{Cov}(Q_1,Q_{l+1})|,\ J_2=\sum_{l=d_T}^{T}|\mathrm{Cov}(Q_1,Q_{l+1})|.$$

Following the steps in FY (pp. 262–263; see also Masry and Fan, 1997), under Assumptions A1–A7 we can show that $J_1, J_2 = o(1/h)$. Thus

$$\mathrm{Var}\left(T^{-1}\textstyle\sum_{t=1}^{T} Q_t\right)\le O(h_0/(Th))+O(1/(Th))=O(h_0/(Th))\to 0. \qquad \textbf{(A.11)}$$

Then (A.6) follows from (A.10) and (A.11). So (A.2) holds.

The asymptotic normality of A_T in (A.1) follows from Theorem 6.3 of FY (p. 238; see also Masry and Fan, 1997) by noting that $\mathrm{Var}(\mathbb{I}(Y_t\le y)|X=x)=F(y|x)[1-F(y|x)]$. Thus the proof of the lemma is complete. ■

Proof of Theorem 1. The proof contains the following steps:

(i) First we establish some preliminary results. According to Lemma 1, by undersmoothing (Assumptions A6′ and A7′), we have

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T} g_t(F(y|x)) \overset{(3)}{=} \sqrt{Th}\left[\widehat{F}(y|x)-F(y|x)\right]\xrightarrow{d}\mathcal{N}(0,V_F(x,y)). \qquad \textbf{(A.12)}$$

Next we show that

$$\frac{1}{Th}\sum_{t=1}^{T} g_t^2(F(y|x))\xrightarrow{p} V_F(x,y). \qquad \textbf{(A.13)}$$

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

Note that

$$\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^{T} g_t^2(F(y|x)) \\
&\quad = \frac{1}{Th}\sum_{t=1}^{T} W_p^2\left(\tfrac{X_t-x}{h}\right)\left(G\left(\tfrac{y-Y_t}{h_0}\right)-F(y|x)\right)^2 \\
&\quad \overset{\text{LLN}}{=} \mathrm{E}h^{-1}W_p^2\left(\tfrac{X_t-x}{h}\right)\left(G\left(\tfrac{y-Y_t}{h_0}\right)-F(y|x)\right)^2+o_p(1) \\
&\quad \overset{\text{LIE}}{=} \mathrm{E}\left(h^{-1}W_p^2\left(\tfrac{X_t-x}{h}\right)\cdot\mathrm{E}\left[\left(G\left(\tfrac{y-Y_t}{h_0}\right)-F(y|x)\right)^2|X_t\right]\right)+o_p(1) \\
&\quad \overset{\text{(A.7),(A.8)}}{=} \mathrm{E}\left(h^{-1}W_p^2\left(\tfrac{X_t-x}{h}\right)\cdot\left[F(y|X_t)(1-F(y|X_t))+O(h_0)\right]\right)+o_p(1) \\
&\quad \overset{\text{(A.3)}}{=} \mathrm{E}\left(h^{-1}\left[\mu(x)^{-1}(1+o_p(1))\cdot K^*\left(\frac{X_t-x}{h}\right)\right]^2\right. \\
&\qquad \left.\cdot\left[F(y|X_t)(1-F(y|X_t))+O(h_0)\right]\right)+o_p(1) \\
&\quad = \int h^{-1}\left[\mu(x)^{-1}(1+o_p(1))\cdot K^*\left(\frac{u-x}{h}\right)\right]^2 \\
&\qquad \cdot[F(y|u)(1-F(y|u))+O(h_0)]\mu(u)\mathrm{d}u+o_p(1) \\
&\quad \overset{z=(u-x)/h}{=} \int\left[f(x)^{-1}(1+o_p(1))\cdot K^*(z)\right]^2 \\
&\qquad \cdot[F(y|x+zh)(1-F(y|x+zh))+O(h_0)]f(x+zh)\mathrm{d}z+o_p(1) \\
&\quad \xrightarrow{p} V_F(x,y),
\end{aligned}$$

where we have used the LLN for mixingales and the fact that the function of the mixing process is still mixing, and the last probability limit is by (A.5). Then (A.13) holds.

(ii) Now we study the Lagrange multiplier $\lambda(F(y|x))$. By the proof of (A.13), we have $\mathrm{Var}[h^{-1/2}g_t(F(y|x))] \to V_F(x,y)$, as $T\to\infty$. So there exists C such that

$$\mathrm{Var}\left[h^{-1/2}g_t(F(y|x))\right] \le C < \infty \qquad \textbf{(A.14)}$$

for all t. Following the proof of Lemma 11.2 in Owen (2001, p. 218), $\max_{1\le t\le T}|h^{-1/2}g_t(F(y|x))| = o_p\left(T^{1/2}\right)$ or

$$\max_{1\le t\le T}|g_t(F(y|x))| = o_p\left(\sqrt{Th}\right). \qquad \textbf{(A.15)}$$

By (6), for any finite θ, $0 = |\sum_{t=1}^T\left(\frac{\lambda(\theta)g_t^2(\theta)}{1+\lambda(\theta)g_t(\theta)}-g_t(\theta)\right)| \ge \sum_{t=1}^T\frac{|\lambda(\theta)|g_t^2(\theta)}{|1+\lambda(\theta)g_t(\theta)|} - |\sum_{t=1}^T g_t(\theta))| \ge \sum_{t=1}^T\frac{|\lambda(\theta)|g_t^2(\theta)}{1+|\lambda(\theta)|\max_{1\le t\le T}|g_t(\theta)|} - \left|\sum_{t=1}^T g_t(\theta)\right|$, or

$$|\lambda(\theta)|\left(\sum_{t=1}^T g_t^2(\theta)-\max_{1\le t\le T}|g_t(\theta)|\cdot\left|\sum_{t=1}^T g_t(\theta)\right|\right) \le \left|\sum_{t=1}^T g_t(\theta)\right|.$$

Thus by (A.12), (A.13), and (A.15)

$$|\lambda(F(y|x))| = O_p\left((Th)^{-1/2}\right). \qquad \textbf{(A.16)}$$

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

By (6) and using the equality $(1-a)^{-1} = 1 + a + a^2/(1-a)$,

$$
\begin{aligned}
0 &= (Th)^{-1/2} \sum_{t=1}^{T} g_t(F(y|x)) \Big(1 - \lambda(F(y|x)) g_t(F(y|x)) + \frac{\lambda^2(F(y|x)) g_t^2(F(y|x))}{1+\lambda(F(y|x)) g_t(F(y|x))}\Big) \\
&= (Th)^{-1/2} \sum_{t=1}^{T} g_t(F(y|x)) - (Th)^{-1/2} \lambda(F(y|x)) \sum_{t=1}^{T} g_t^2(F(y|x)) \\
&\quad + (Th)^{-1/2} \lambda^2(F(y|x)) \sum_{t=1}^{T} \frac{g_t^3(F(y|x))}{1+\lambda(F(y|x)) g_t(F(y|x))} \\
&= (Th)^{-1/2} \sum_{t=1}^{T} g_t(F(y|x)) - (Th)^{-1/2} \lambda(F(y|x)) \sum_{t=1}^{T} g_t^2(F(y|x)) + o_p(1)
\end{aligned}
\tag{A.17}
$$

since

$$
\begin{aligned}
&\Big|(Th)^{-1/2} \lambda^2(F(y|x)) \textstyle\sum_{t=1}^{T} \frac{g_t^3(F(y|x))}{1+\lambda(F(y|x)) g_t(F(y|x))}\Big| \\
&\quad \le \lambda^2(F(y|x))(Th)^{-1/2} \sum_{t=1}^{T} \frac{|g_t^3(F(y|x))|}{|1+\lambda(F(y|x)) g_t(F(y|x))|} \\
&\quad \le C[1+o_p(1)] \lambda^2(F(y|x))(Th)^{-1/2} \sum_{t=1}^{T} |g_t^3(F(y|x))| = o_p(1),
\end{aligned}
$$

where the last inequality uses the fact $\max_{1\le t\le T} |\lambda(F(y|x)) g_t(F(y|x))| = o_p(1)$ [which follows from (A.15) and (A.16)] and so for all t, $|1 + \lambda(F(y|x)) g_t(F(y|x))|^{-1} = O_p(1)$. The last equality uses (A.16) and the fact that $(Th)^{-1/2} \sum_{t=1}^{T} |g_t^3(F(y|x))| = o_p(Th)$ [which follows from the proof of Lemma 11.3 in Owen (2001, p. 218) and (A.14)]. So, by (A.17),

$$
\lambda(F(y|x)) = \big(\textstyle\sum_{t=1}^{T} g_t^2(F(y|x))\big)^{-1} \big(\sum_{t=1}^{T} g_t(F(y|x))\big) + o_p(1). \tag{A.18}
$$

(iii) Finally, we look at the log EL ratio statistic $R(F(y|x))$. By Taylor expansion of $\log(1+\lambda(F(y|x)) g_t(F(y|x)))$ with respect to λ at zero,

$$
\begin{aligned}
\log(1+\lambda(F(y|x)) g_t(F(y|x))) &= \lambda(F(y|x)) g_t(F(y|x)) \\
&\quad - \frac{1}{2} \lambda^2(F(y|x)) g_t^2(F(y|x)) + \xi_t,
\end{aligned}
\tag{A.19}
$$

where $\xi_t = g_t^3(F(y|x)) \lambda^3(F(y|x))[1 + \dot{\lambda} g_t(F(y|x))]^{-3}/3$, with $\dot{\lambda}$ being a point between $\lambda(F(y|x))$ and 0. Following the similar arguments used in (ii), we have $\sum_{t=1}^{T} \xi_t = o_p(1)$. So by (7) and (A.19),

$$
\begin{aligned}
R(F(y|x)) &= 2 \sum_{t=1}^{T} \log\big(1 + \lambda(\theta) g_t(\theta)\big) \\
&= \big(\textstyle\sum_{t=1}^{T} g_t^2(F(y|x))\big)^{-1} \big(\sum_{t=1}^{T} g_t(F(y|x))\big)^2 + o_p(1) \\
&\xrightarrow{d} \chi^2(1),
\end{aligned}
$$

where the weak convergence follows from (A.12) and (A.13). The proof of the theorem is complete. ■

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press

Proof of Theorem 2. It follows from the proof of Theorem 1 if we show (A.12) and (A.13) for $\widehat{g}_t(F(y|x))$ instead of $g_t(F(y|x))$ without the undersmoothing condition A6′. We first show (A.12). We write

$$\begin{aligned}
&\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}\widehat{g}_t(F(y|x)) \\
&\quad=\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}W_p((X_t-x)/h)\cdot\Big(G((y-Y_t)/h_0)-F(y|x)-\Big[\widehat{F}(y|X_t)-\widehat{F}(y|x)\Big]\Big) \\
&\quad=\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}g_t(F(y|x))-\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}W_p((X_t-x)/h)\cdot\big[\widehat{F}(y|X_t)-\widehat{F}(y|x)\big]. \quad \textbf{(A.20)}
\end{aligned}$$

By Lemma 1, under Assumptions A1–A6 and A7′,

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}g_t(F(y|x))\xrightarrow{d}\mathcal{N}\big(\widetilde{B}_X(x,y),V_F(x,y)\big),$$

where $\widetilde{B}_X(x,y)=\lim_{T\to\infty}\sqrt{Th}\frac{h^{p+1}}{(p+1)!}\frac{\mathrm{d}^{p+1}F(y|x)}{\mathrm{d}x^{p+1}}e_1'S^{-1}c_p$. We shall show that the second term in (A.20) converges in probability to $\widetilde{B}_X(x,y)$. Standard results (e.g. FY, Thm. 6.3) give that $\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}W_p((X_t-x)/h)\cdot[F(y|X_t)-F(y|x)]\xrightarrow{p}\widetilde{B}_X(x,y)$. So we only need to show that

$$\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}W_p((X_t-x)/h)\cdot\Big(\Big[\widehat{F}(y|X_t)-\widehat{F}(y|x)\Big]-[F(y|X_t)-F(y|x)]\Big)=o_p(1). \quad \textbf{(A.21)}$$

By Taylor expansion of $\widehat{F}(y|X_t)$ and $F(y|X_t)$ with respect to X_t at x, $[\widehat{F}(y|X_t)-\widehat{F}(y|x)]-[F(y|X_t)-F(y|x)]=[\widehat{F}_x'(y|\overline{x})-F_x'(y|\overline{\overline{x}})](X_t-x)$, where $\overline{x}$ and $\overline{\overline{x}}$ are two points between X_t and x. Note that $\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}W_p((X_t-x)/h)(X_t-x)\overset{(A.3)}{=}\mu(x)^{-1}[1+o_p(1)]h\cdot\frac{1}{\sqrt{Th}}\sum_{t=1}^{T}K^*((X_t-x)/h)(X_t-x)/h\overset{\mathrm{CLT}}{=}O_p(h)$. By the convergence rate of the derivative estimator, $\widehat{F}_x'(y|\overline{x})-F_x'(y|\overline{\overline{x}})=O_p(1/(h\sqrt{Th}))$. So the left-hand side of (A.21) is $O_p(1/\sqrt{Th})$ and thus (A.21) holds. Similarly, we can show (A.13). The proof of Theorem 2 is complete. ■

https://doi.org/10.1017/S0266466612000667 Published online by Cambridge University Press