Econometric Theory, **18**, 2002, 169–192. Printed in the United States of America.

REGRESSION QUANTILES FOR TIME SERIES

ZONGWU CAI
University of North Carolina at Charlotte

In this paper we study nonparametric estimation of regression quantiles for time series data by inverting a weighted Nadaraya–Watson (WNW) estimator of conditional distribution function, which was first used by Hall, Wolff, and Yao (1999, *Journal of the American Statistical Association* 94, 154–163). First, under some regularity conditions, we establish the asymptotic normality and weak consistency of the WNW conditional distribution estimator for α-mixing time series at both boundary and interior points, and we show that the WNW conditional distribution estimator not only preserves the bias, variance, and, more important, automatic good boundary behavior properties of local linear "double-kernel" estimators introduced by Yu and Jones (1998, *Journal of the American Statistical Association* 93, 228–237), but also has the additional advantage of always being a distribution itself. Second, it is shown that under some regularity conditions, the WNW conditional quantile estimator is weakly consistent and normally distributed and that it inherits all good properties from the WNW conditional distribution estimator. A small simulation study is carried out to illustrate the performance of the estimates, and a real example is also used to demonstrate the methodology.

1. INTRODUCTION

In nonparametric estimation of regression function, most investigations are concerned with the regression function $m(\mathbf{x})$, the conditional mean of Y given value $\mathbf{x}$ of a predictor $\mathbf{X}$. See, for example, the books by Wahba (1990), Wand and Jones (1995), and Fan and Gijbels (1996) for good introductions and interesting applications to the general subject areas and Härdle (1990), Engle and Granger (1991), and Granger and Teräsvirta (1993) for applications to econometrics in particular. However, new insights about the underlying structures can be gained by considering the other aspects of the conditional distribution $F(y|\mathbf{x})$ of Y given $\mathbf{X} = \mathbf{x}$ other than the mean function $m(\mathbf{x})$. Regression (conditional) quantiles $q_p(\mathbf{x})$, $0 < p < 1$, of Y given $\mathbf{X} = \mathbf{x}$, the topic of this paper, coupled with the conditional distribution $F(y|\mathbf{x})$, are key aspects of inference in various statistical problems and are of particular interest in econometrics.

This research was supported in part by National Science Foundation grant DMS 0072400 and funds provided by the University of North Carolina at Charlotte. The author is indebted to the coeditor Benedikt M. Pötscher and two referees for their constructive suggestions and comments that helped to improve an earlier version of this paper. Address correspondence to: Zongwu Cai, Department of Mathematics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA; e-mail: zcai@uncc.edu.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

Although some individual quantiles, such as the conditional median, are sometimes of interest in practice, more often one wishes to obtain a collection of conditional quantiles that can characterize the entire conditional distribution. More important, another application of conditional quantiles is construction of prediction intervals for the next value given a small section of the recent past values in a stationary time series $\{Y_1, \ldots, Y_n\}$, which motivates this work.

Estimation of conditional quantiles has gained particular attention during the recent three decades because of their useful applications in various fields such as econometrics and finance and also the related fields. Although Hogg (1975) uses the idea of regression quantile technique to study salary data, Koenker and Bassett (1978) first introduce regression quantiles and provide detailed background and motivation from econometrics. Koenker and Bassett (1982) use their proposed techniques to test heteroskedasticity, and Powell (1986) applies the ideas to consider censored data in econometrics. Fan and Gijbels (1996, p. 229) use them to study the change pattern of family income during the Reagan administration (1981–1988). For more applications, see Cole (1988), Cole and Green (1992), Yu and Jones (1998), and Hall, Wolff, and Yao (1999). Of particular interest is the median function $q_{1/2}(\mathbf{x})$ for asymmetric distribution, which can provide a useful alternative to the ordinary regression based on the mean. Regression quantiles can also be useful for the estimation of predictive intervals. For example, in predicting the response from a given covariate $\mathbf{X} = \mathbf{x}$, estimates of $q_{\alpha/2}(\mathbf{x})$ and $q_{1-\alpha/2}(\mathbf{x})$ can be used to obtain a $(1 - \alpha)100\%$ nonparametric predictive interval. We refer to the papers by Koenker (1994) and Zhou and Portnoy (1996) for detailed discussions on the advantages of the direct use of regression quantiles to construct confidence and prediction intervals and other aspects.

For cases when $\mathbf{X}$ and Y satisfy a linear model with independent and identically distributed (i.i.d.) errors, Hogg (1975) and Koenker and Bassett (1978) develop innovative procedures, called regression quantiles, for inference about $q_p(\mathbf{x})$; Koenker and Bassett (1982) use them to test heteroskedasticity; Powell (1986) extends them to censored data; Koenker (1994) and Zhou and Portnoy (1996) discuss how to use them to construct confidence sets. In a nonparametric setting, several authors study the asymptotic properties of nonparametric estimation of conditional quantiles, such as kernel and nearest neighbor, including Stone (1977), Lejeune and Sarda (1988), Troung (1989), Samanta (1989), Bhattacharya and Gangopadhyay (1990), and Chaudhuri (1991) for i.i.d. errors; Roussas (1969) and Roussas (1991) for Markovian processes; and Troung and Stone (1992) and Boente and Fraiman (1995) for mixing dependence. It is well known that 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 modification. To attenuate these drawbacks, recently some new methods of estimating conditional quantiles have been proposed. The first one, a more direct approach using the "check" function such as a robustified local linear smoother, is provided by Fan, Hu, and Troung (1994) and further extended by Yu and Jones (1997, 1998). An

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

alternative procedure is first to estimate the conditional distribution function using the "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, which is called the Yu and Jones estimator. See Yu and Jones (1997, 1998) for details. A detailed comparison of these two methods can be found in Yu and Jones (1997, 1998). According to Yu and Jones (1998), a particular preference is the Yu and Jones type estimator.

As pointed out by Hall et al. (1999), although local linear estimators of the Yu and Jones type have some attractive properties such as no boundary effects, design adaptation, and mathematical efficiency (see, e.g., Fan and Gijbels, 1996), they have the disadvantage of producing conditional distribution function estimators that are not constrained either to lie between zero and one or to be monotone increasing although some modifications in implementation have been addressed by Yu and Jones (1998). In both these respects, the Nadaraya–Watson (NW) methods are superior, despite their rather large bias and boundary effects. The properties of positivity and monotonicity are particularly advantageous if the method of inverting the conditional distribution estimator is applied to produce an estimator of a conditional quantile. To overcome these difficulties, Hall et al. (1999) propose a weighted version of the NW (WNW) estimator, which is designed to possess the superior properties of local linear methods such as bias reduction and no boundary effect and to preserve the property that the NW estimator is always a distribution function. Also, Hall et al. discuss the asymptotic normality of the WNW estimator for β-mixing under some stronger assumptions (see Remark 1, which follows). However, they do not provide rigorous theoretical justification, and they do not discuss the boundary behavior. Furthermore, Hall et al. (1999) conduct an empirical study on the comparison of WNW with other methods such as NW, local logistic, and local linear. The basic techniques are not novel to this paper because the WNW method was first used by Hall et al., but many details and insights are.

The goal of this paper is twofold. First, we establish asymptotic normality and weak consistency for the WNW estimator of conditional distribution for α-mixing under a set of weaker conditions at both boundary and interior points. It is therefore shown, to the first order, that the WNW method enjoys the same convergence rates as those of the local linear "double-kernel" procedure of Yu and Jones (1998). An important consequence of this study is that the WNW estimator has desired sampling properties at both boundary and interior points of the support of the design density, which seems to be seminal. Second, we derive the WNW estimator of the conditional quantile by inverting the WNW conditional distribution estimator. We show that the WNW quantile estimator always exists as a result of the WNW distribution being a distribution function itself and that it inherits all advantages from the WNW estimator of conditional distribution, as we describe later.

Although our interest in conditional quantile estimation is motivated by forecasting from time series data, we introduce our methods in a more general set-

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

ting (α-mixing) that includes time series modeling as a special case. Our theoretical results are derived under an α-mixing assumption.

For reference convenience, we first introduce the mixing coefficient. Let $\mathcal{F}_a^b$ be the σ-algebra generated by $\{(\mathbf{X}_t, Y_t)\}_{t=a}^b$. Define

$$\alpha(t) = \sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_t^\infty\}.$$

This is called the strong mixing coefficient of the stationary process $\{(\mathbf{X}_t, Y_t)\}_{-\infty}^{\infty}$. If $\alpha(t) \to 0$ as $t \to \infty$, the process is called strongly mixing or α-mixing.

Among various mixing conditions used in the literature, α-mixing is reasonably weak and is known to be fulfilled for many time series models. Gorodetskii (1977) and Withers (1981) derive the conditions under which a linear process is α-mixing. In fact, under very mild assumptions linear autoregressive and more generally bilinear time series models are α-mixing with mixing coefficients decaying exponentially. Auestad and Tjøstheim (1990) provide illuminating discussions on the role of α-mixing (including geometric ergodicity) for model identification in nonlinear time series analysis. Chen and Tsay (1993) show that the functional autoregressive process is geometrically ergodic under certain conditions. Furthermore, Masry and Tjøstheim (1995, 1997) demonstrate that under some mild conditions, both autoregressive conditional heteroskedastic (ARCH) processes and nonlinear additive autoregressive models with exogenous variables, particularly popular in finance and econometrics, are stationary and α-mixing.

The plan of the paper is as follows. In Section 2, we concentrate on the WNW estimator of conditional distribution. In Section 3, we discuss the WNW estimator of conditional quantiles. In both sections, the asymptotic normality and weak consistency of the estimators at both boundary and interior points are stated, and an ad hoc estimator of the asymptotic variance is also presented. In Section 4, a small simulation study is carried out to illustrate the estimates, and the methodology is also applied to a real example. All technical proofs are given in the Appendix.

2. CONDITIONAL DISTRIBUTION ESTIMATE

2.1. Weighted Nadaraya–Watson Estimate

Let $p_t(\mathbf{x})$, for $1 \le t \le n$, denote the weight functions of the data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and the design point $\mathbf{x}$ with the property that each $p_t(\mathbf{x}) \ge 0$, $\sum_{t=1}^n p_t(\mathbf{x}) = 1$, and

$$\sum_{t=1}^{n} (\mathbf{X}_t - \mathbf{x}) p_t(\mathbf{x}) K_h(\mathbf{x} - \mathbf{X}_t) = 0, \tag{1}$$

where $K(\cdot)$ is a kernel function, $K_h(\cdot) = K(\cdot/h)/h$, and $h = h_n > 0$ is the bandwidth. Motivated by the property of the local linear estimator, constraint (1) can be regarded as a discrete moment condition (see Fan and Gijbels, 1996,

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

(3.12), p. 63). Of course, $\{p_t(\mathbf{x})\}$ satisfying these conditions are not uniquely defined, and we specify them by maximizing $\prod_{t=1}^n p_t(\mathbf{x})$ subject to the constraints. The weighted version of the NW estimator of the conditional distribution $F(y|\mathbf{x})$ of Y_t given $\mathbf{X}_t = \mathbf{x}$ is defined as

$$\hat{F}(y|\mathbf{x}) = \frac{\sum_{t=1}^n p_t(\mathbf{x}) K_h(\mathbf{x} - \mathbf{X}_t) I(Y_t \le y)}{\sum_{t=1}^n p_t(\mathbf{x}) K_h(\mathbf{x} - \mathbf{X}_t)}.$$

Note that $0 \le \hat{F}(y|\mathbf{x}) \le 1$ and it is monotone in y. We show in Theorem 1, which follows, that $\hat{F}(y|\mathbf{x})$ is first-order equivalent to a local linear estimator, which does not enjoy either of these properties, and, more important, in Theorem 2 that $\hat{F}(y|\mathbf{x})$ has automatic good behavior at boundaries.

A question naturally arises regarding how to choose the weights. The idea is from empirical likelihood. Namely, by maximizing $\sum_{t=1}^n \log\{p_t(\mathbf{x})\}$ subject to the constraints $\sum_{t=1}^n p_t(\mathbf{x}) = 1$ and (1) through the Lagrange multiplier, the $\{p_t(\mathbf{x})\}$ are simplified to

$$p_t(\mathbf{x}) = n^{-1}\{1 + \lambda(\mathbf{X}_t - \mathbf{x}) K_h(\mathbf{x} - \mathbf{X}_t)\}^{-1},$$

where λ, a function of data and $\mathbf{x}$, is uniquely defined by (1), which ensures that $\sum_{t=1}^n p_t(\mathbf{x}) = 1$. Equivalently, λ is chosen to maximize

$$L_n(\lambda) = \frac{1}{nh} \sum_{t=1}^n \log\{1 + \lambda(\mathbf{X}_t - \mathbf{x}) K_h(\mathbf{x} - \mathbf{X}_t)\}. \tag{2}$$

In implementation, the Newton–Raphson scheme is recommended to find the root of the equation $L_n'(\lambda) = 0$.

2.2. Sampling Properties

In this section, we establish weak consistency with a rate and asymptotic normality for the WNW estimator $\hat{F}(y|\mathbf{x})$ under α-mixing although Hall et al. (1999) gave asymptotic normality for β-mixing without detailed proofs under a set of stronger conditions. For expositional purposes, we consider only the special case that $\mathbf{X}$ is a scalar. We first introduce some notation. Let $g(\cdot)$ denote the marginal density of X_t. Define $\mu_2 = \int u^2 K(u)\, du$ and $\nu_0 = \int K^2(u)\, du$. Let $F^{(i)}(y|x) = (\partial/\partial x)^i F(y|x)$. We now impose the following regularity conditions.

B1. For fixed y and x, $g(x) > 0$, $0 < F(y|x) < 1$, $g(\cdot)$ is continuous at x, and $F(y|x)$ has continuous second-order derivative with respect to x.

B2. The kernel $K(\cdot)$ is a symmetric, bounded, and compactly supported density.

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

B3. The process $\{(X_t, Y_t)\}$ is α-mixing with the mixing coefficient satisfying $\alpha(t) = O(t^{-(2+\delta)})$ for some $\delta > 0$.

B4. As $n \to \infty$, $h \to 0$ and $nh \to \infty$.

B5. 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.

B6. $nh^{1+2/\delta} \to \infty$.

Remark 1. Note that condition B3 is weaker than that in Hall et al. (1999) for β-mixing, which is stronger than α-mixing. Because B6 is satisfied by bandwidths of optimal size (i.e., $h \approx n^{-1/5}$) if $\delta > \frac{1}{2}$, we do not concern ourselves with such refinements.

THEOREM 1. *Suppose that conditions B1–B5 hold. Then, as* $n \to \infty$,

$$\hat{F}(y|x) - F(y|x) = \frac{1}{2} h^2 \mu_2 F^{(2)}(y|x) + o_p(h^2) + O_p((nh)^{-1/2}), \tag{3}$$

and in addition, if B6 holds,

$$\sqrt{nh}[\hat{F}(y|x) - F(y|x) - B(y|x) + o_p(h^2)] \xrightarrow{D} N(0, \sigma^2(y|x)), \tag{4}$$

where the bias and variance are given, respectively, by

$$B(y|x) = \frac{1}{2} h^2 \mu_2 F^{(2)}(y|x)$$

and

$$\sigma^2(y|x) = \nu_0 F(y|x)[1 - F(y|x)]/g(x). \tag{5}$$

Remark 2. It may be seen from Theorem 1 that first, the WNW estimator $\hat{F}(y|x) \to F(y|x)$ in probability with a rate, which, of course, implies that $\hat{F}(y|x)$ is consistent. Also, to the first order, the WNW method enjoys the same convergence rates as those of the local linear "double-kernel" procedure of Yu and Jones (1998), under similar regularity conditions. However, Yu and Jones (1998) treat only the case of independent data.

As for the boundary behavior of the WNW estimator, we offer Theorem 2, which follows. Without loss of generality, we consider the left boundary point $x = ch$, $0 < c < 1$. From Fan et al. (1994), we take $K(\cdot)$ to have support $[-1,1]$ and $g(\cdot)$ to have support $[0,1]$. First, we introduce the following notation. Let

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

and λ_c be the root of equation $L_c(\lambda) = 0$, namely, $L_c(\lambda_c) = 0$. For example, $\lambda_c \approx 1.8$ for $c = 0.5$ and $\lambda_c \approx 1.1$ for $c = 0.6$. Figure 1 depicts the solutions of $L_c(\lambda) = 0$ for c taking values from 0.5 to 1 with increment 0.1.

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

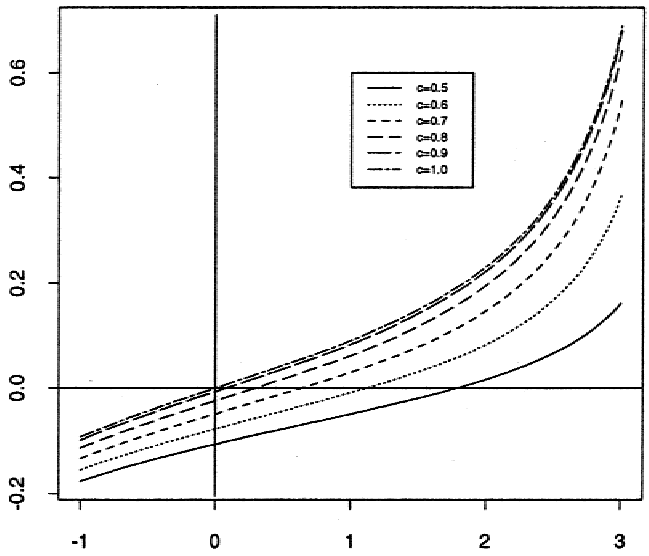


FIGURE 1. Plot of $L_c(\lambda)$ versus λ for c taking values from 0.5 to 1 with increment 0.1.

THEOREM 2. *Suppose that the conditions of Theorem 1 hold. Then,*

$$\sqrt{nh}[\hat{F}(y|ch) - F(y|ch) - B_c(y) + o_p(h^2)] \xrightarrow{D} N(0, \sigma_c^2(y)), \tag{6}$$

where the bias term is given by

$$B_c(y) = \frac{h^2 \beta_0(c) F^{(2)}(y|0+)}{2\beta_1(c)}$$

and the variance is

$$\sigma_c^2(y) = \frac{\beta_2(0) F(y|0+)[1 - F(y|0+)]}{\beta_1^2(c) g(0+)}$$

with

$$\beta_0(c) = \int_{-1}^{c} \frac{u^2 K(u)}{1 - \lambda_c u K(u)}\, du \quad \textit{and} \quad \beta_j(c) = \int_{-1}^{c} \frac{K^j(u)}{\{1 - \lambda_c u K(u)\}^j}\, du,$$

$j = 1$ *and* 2.

Of course, $g(0+) = \lim_{z \downarrow 0} g(z)$.

Theorems 1 and 2 reflect two of the major advantages of the WNW estimator: (a) no dependence of the asymptotic bias on the design density $g(\cdot)$ and, indeed, its dependence on the simple conditional distribution curvature $F^{(2)}(\cdot|\cdot)$; and (b) automatic good behavior at boundaries, at least with regard to orders of

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

magnitude, without the need of boundary correction. Also, we remark that a similar result (6) holds for the right boundary point $x = 1 - ch$. If the point 0 were an interior point, then the expression (6) would hold with $c = 1$ and $\lambda_c = 0$. Furthermore, note that because the proofs of Theorems 1 and 2 are similar, we present only the detailed proof of Theorem 1 and give a brief outline of the proof of Theorem 2 in the Appendix.

The explicit expression of the asymptotic variance of $\hat{F}(y|x)$ given in (5) gives a more direct and simpler way to construct the estimate of $\sigma^2(y|x)$, as follows:

$$\hat{\sigma}^2(y|x) = \nu_0 \hat{F}(y|x)[1 - \hat{F}(y|x)]/\hat{g}(x), \tag{7}$$

where $\hat{g}(x)$ is any consistent density estimator of $g(x)$ that might be obtained by a local linear procedure.

3. QUANTILE ESTIMATE

Our interest here is to estimate the pth conditional quantile function $q_p(x)$ of Y_t given $X_t = x$ for any $0 < p < 1$, defined by

$$q_p(x) = \inf\{y \in \Re : F(y|x) \geq p\},$$

which is assumed to be unique. As demonstrated in Section 2, the WNW estimator of the conditional distribution function $\hat{F}(y|x)$ possesses the following advantages: design adaptation and no boundary effects, which are the same as the local linear counterpart; and more important, being between zero and one and monotone, two properties that the local linear estimator does not enjoy. We thus define the WNW type conditional quantile estimator $\hat{q}_p(x)$ in principle to satisfy $\hat{F}(\hat{q}_p(x)|x) = p$ so that

$$\hat{q}_p(x) = \inf\{y \in \Re : \hat{F}(y|x) \geq p\} \equiv \hat{F}^{-1}(p|x). \tag{8}$$

We remark that $\hat{q}_p(x)$ always exists because $\hat{F}(y|x)$ is between zero and one and monotone in y, and it involves only one bandwidth, so that it makes practical implementation more appealing. In contrast, the "double-kernel" estimator of Yu and Jones (1998) has some difficulty inverting the conditional distribution estimator because of lack of monotonicity, and it requires choosing two bandwidths, although the second bandwidth should not be very sensitive. Furthermore, we show in Theorems 3 and 4, which follow, that the WNW estimator $\hat{q}_p(x)$ maintains the aforementioned advantages of $\hat{F}(\cdot|x)$. To this end, we need the following additional conditions.

C1. Assume that $F(y|x)$ has a conditional density $f(y|x)$ and $f(y|x)$ is continuous at x.

C2. $f(q_p(x)|x) > 0$.

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

THEOREM 3. *Suppose that conditions B1–B5 hold. Then, as* $n \to \infty$,

$$\hat{q}_p(x) \to q_p(x) \quad \textit{in probability.} \tag{9}$$

In addition, if conditions C1 and C2 are satisfied, then,

$$\sqrt{nh}[\hat{q}_p(x) - q_p(x) - B_p(x) + o_p(h^2)] \xrightarrow{D} N(0, \sigma_p^2(x)), \tag{10}$$

where the bias and variance are given, respectively, by

$$B_p(x) = -\frac{B(q_p(x)|x)}{f(q_p(x)|x)} \quad \textit{and} \quad \sigma_p^2(x) = \frac{\sigma^2(q_p(x)|x)}{f^2(q_p(x)|x)} = \frac{\nu_0 p[1-p]}{f^2(q_p(x)|x)g(q_p(x))}.$$

An important way of assessing the performance of $\hat{q}_p(x)$ is by its asymptotic mean squared error (AMSE). As an application of Theorem 3, the AMSE of $\hat{q}_p(x)$ is given by

$$\text{AMSE}(\hat{q}_p(x)) = \frac{h^4}{4}\left\{\frac{\mu_2 F^{(2)}(q_p(x)|x)}{f(q_p(x)|x)}\right\}^2 + \frac{1}{nh}\frac{\nu_0 p(1-p)}{f^2(q_p(x)|x)g(q_p(x))}. \tag{11}$$

Comparing (11) with Theorem 1 in Yu and Jones (1998) for the "double-kernel" estimator reveals that (11) does not have the extra two terms from the vertical smoothing "in the y direction." Minimizing AMSE in (11), therefore, yields the optimal bandwidth

$$h_{opt} = \left[\frac{\mu_2^2\{F^{(2)}(q_p(x)|x)\}^2}{\nu_0 p(1-p)/g(q_p(x))}\right]^{1/5} n^{-1/5}.$$

In the same manner as in (7), the consistent estimate of $\sigma_p^2(x)$ is

$$\hat{\sigma}_p^2(x) = \frac{\nu_0 p[1-p]}{\hat{f}^2(q_p(x)|x)\hat{g}(q_p(x))},$$

where $\hat{f}(y|x)$ can be obtained by using the local linear "double-kernel" method of Fan et al. (1996).

Similar to Theorem 2, we consider the boundary behavior of the WNW estimator $\hat{q}_p(x)$ in the following theorem. The proof of Theorem 4 is omitted because it is similar to that of Theorem 3, which can be found in the Appendix.

THEOREM 4. *Suppose that the conditions of Theorem 3 hold. Then, as* $n \to \infty$,

$$\sqrt{nh}[\hat{q}_p(ch) - q_p(ch) - B_{p,c} + o_p(h^2)] \xrightarrow{D} N(0, \sigma_{p,c}^2),$$

where the bias and variance are given, respectively, by

$$B_{p,c} = -\frac{B_c(q_p(0+))}{f(q_p(0+)|0+)} = -\frac{h^2\beta_2(c)F^{(2)}(q_p(0+)|0+)}{2\beta_1(c)f(q_p(0+)|0+)}$$

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

and

$$\sigma^2_{p,c} = \frac{\sigma^2_c(q_p(0+))}{f^2(q_p(0+)|0+)} = \frac{\beta_0(0)p[1-p]}{\beta_1^2(c)f^2(q_p(0+)|0+)g(q_p(0+))}.$$

Similarly, we can derive the AMSE of $\hat{q}_p(ch)$

$$\text{AMSE}(\hat{q}_p(ch)) = \frac{h^4}{4}\left\{\frac{\beta_2(c)F^{(2)}(q_p(0+)|0+)}{\beta_1(c)f(q_p(0+)|0+)}\right\}^2 + \frac{1}{nh}\frac{\beta_0(0)p(1-p)}{\beta_1^2(c)f^2(q_p(0+)|0+)g(q_p(0+))},$$

and the corresponding optimal bandwidth is

$$h_{c,opt} = \left[\frac{\beta_2^2(c)\{F^{(2)}(q_p(0+)|0+)\}^2}{\beta_0(c)p(1-p)/g(q_p(0+))}\right]^{1/5} n^{-1/5}.$$

Remark 3. As mentioned earlier, the WNW method involves only one bandwidth in estimating the conditional distribution and quantile. As for the bandwidth selection issue, because the WNW estimate is a linear smoother, we recommend using the optimal bandwidth selector, called the nonparametric Akaike information criterion, proposed by Cai and Tiwari (2000), which is used in our implementation in Section 4.

Remark 4. From the foregoing discussions, it appears clear that one of the advantages of the conditional distribution/quantile technique is to cope with heteroskedasticity automatically, which is particularly relevant for prediction in ARCH modeling. Also, it is convenient to use the conditional quantile for detecting conditional heteroskedasticity. To this end, we assume that Y_t is related to X_t through the model

$$Y_t = \mu(X_t) + \sigma(X_t)\varepsilon_t, \tag{12}$$

where $\mu(\cdot)$ is the mean function, $\sigma^2(\cdot)$ is the variance function, and X_t and ε_t are independent. The conditional distribution and quantile of Y_t given $X_t = x$ are

$$F(y|x) = F_\varepsilon\{(y-\mu(x))/\sigma(x)\} \quad \text{and} \quad q_p(x) = \mu(x) + \sigma(x)q_\varepsilon(p),$$

where $F_\varepsilon(\cdot)$ is the distribution of ε_t and $q_\varepsilon(p)$ is the pth quantile. An informal way to test conditional heteroskedasticity is to use a graph. If the curves of $F(y|x)$ or $q_p(x)$ for different values of x are parallel, this indicates that $\sigma(\cdot)$ should be a constant. We hope to report in greater detail on a formal test for heteroskedasticity using a regression quantile technique for time series in future work.

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

4. EXAMPLES

We now illustrate the methods discussed earlier with a small simulation and a real data set for time series. The purpose of the examples is to demonstrate that the foregoing methods work reasonably well for time series data in terms of predictive utility and detecting heteroskedasticity graphically. Throughout this section, the Epanechnikov kernel $K(u) = 0.75(1 - u^2)_+$ is used.

4.1. A Simulated Example

We begin the illustration with a simulated example of the ARCH model in (12) with $X_t = Y_{t-1}$:

$$Y_t = 0.6Y_{t-1} + \sigma(Y_{t-1})\varepsilon_t,$$

where $\sigma(x) = \sqrt{0.5(1 + x^2)}$ and $\{\varepsilon_t\}$ are i.i.d. $N(0,1)$. Here $n + 5$ data points are generated. The first n observations are used for estimation of the conditional distribution and quantile, and the last 5 observations are left for construction of predictive intervals (PI). Sample sizes are $n = 100$, 300, and 500. The 95% prediction intervals $[\hat{q}_{0.025}(Y_{n+j-1}), \hat{q}_{0.975}(Y_{n+j-1})]$ for Y_{n+j} $(1 \le j \le 5)$ are computed. Finally, for each n, 500 replications are performed, and the ratio of the length of PI versus the range of data and the coverage frequencies are recorded. In Table 1, we present the median of the 500 values for ratios and coverage frequencies for each n and Y_{n+j} with its standard deviation in parentheses. The ratio of the length of PI versus the range of data decreases when n increases, and the coverage frequency increases and is close to 95% when n increases. Figures 2a and 2b display the WNW estimates of conditional distributions $F(y|x) = \Phi((y - 0.6x)/\sigma(x))$ (solid line in Figure 2a) of Y_t given $X_t = x$ and their quantiles $q_x(p) = 0.6x + \sigma(x)\Phi^{-1}(p)$ (solid line in Figure 2b), dotted line $(x = -1.76)$ and dashed line $(x = -0.75)$, based on a random sample with $n = 500$, where $\Phi(\cdot)$ is the standard normal distribution. Clearly, Figure 2 shows that the two curves of conditional distribution/quantile for two different values of X_t are not parallel, which implies that the conditional variance is not a constant. This example shows that the performance of the WNW estimate is reasonably good.

TABLE 1. The postsample prediction for the ARCH model

	n	Y_{n+1}	Y_{n+2}	Y_{n+3}	Y_{n+4}	Y_{n+5}
Ratio of length	100	0.46 (0.19)	0.49 (0.18)	0.42 (0.18)	0.45 (0.18)	0.44 (0.17)
of PI vs. range	300	0.29 (0.15)	0.30 (0.14)	0.30 (0.15)	0.30 (0.16)	0.28 (0.13)
of data	500	0.23 (0.13)	0.25 (0.12)	0.24 (0.14)	0.24 (0.12)	0.23 (0.10)
Coverage	100	0.90 (0.30)	0.90 (0.30)	0.88 (0.32)	0.88 (0.33)	0.91 (0.29)
Frequencies	300	0.94 (0.23)	0.93 (0.26)	0.95 (0.22)	0.91 (0.29)	0.92 (0.27)
	500	0.94 (0.23)	0.96 (0.21)	0.95 (0.22)	0.95 (0.22)	0.95 (0.23)

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

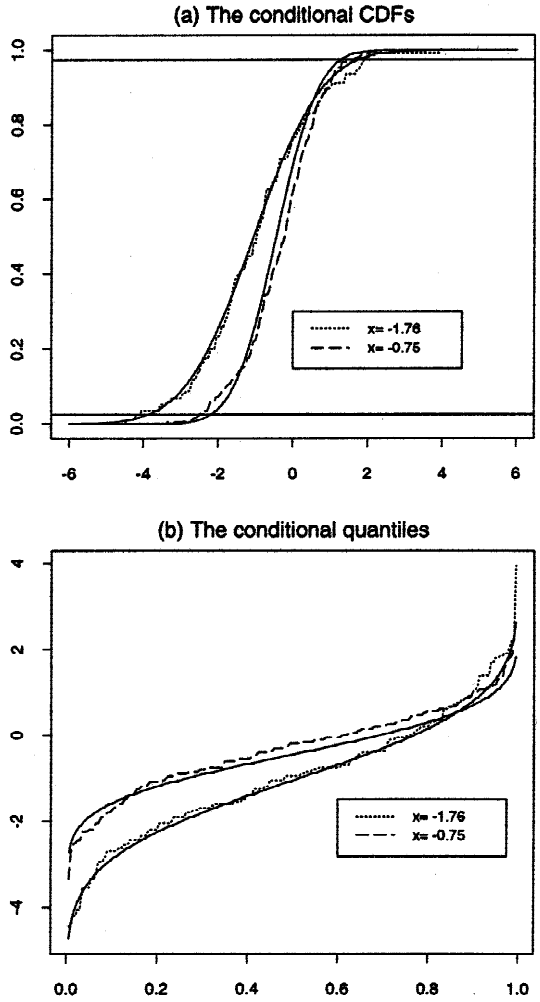


FIGURE 2. (a) Conditional distributions and their WNW estimates. (b) Conditional quantiles and their WNW estimates. Solid line, true function; dotted line ($X_t = -1.76$), and dashed line ($X_t = -0.75$).

4.2. A Real Example

Finally, we consider the International Airline Passenger Data $\{U_t, t = 1, \ldots, 144\}$ of Box and Jenkins (1976, p. 531) (monthly totals [in thousands] of passengers from January 1949 to December 1960). This data set has been analyzed extensively in time series literature (see, e.g., Brockwell and Davis, 1991; Faraway and Chatfield, 1998). The previous work focuses mainly on linear or nonlinear

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

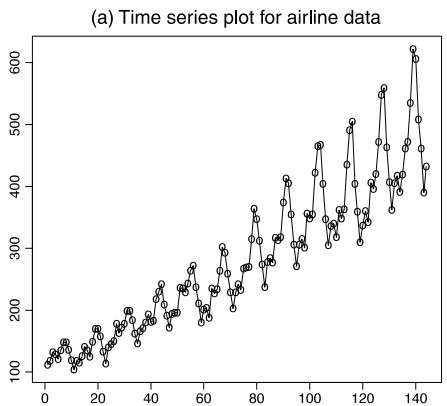
autoregressive moving average (ARMA) modeling. The main purpose of this analysis is to use the regression quantiles technique to construct prediction intervals and to check heteroskedasticity graphically without making any model assumptions. Figure 3a shows clearly that variability increases as U_t increases (called multiplicative seasonality) and also shows tread. To eliminate the tread and seasonality, following the convention in the literature, we consider transformations. We first use the logarithmic transformation because $\{U_t\}$ is a series whose standard deviation increases linearly with the mean. The transformed series $V_t = \log(U_t)$, shown in Figure 3b, does not display increase in variability with V_t but clearly shows a linear tread. Second, by following the analysis in Brockwell and Davis (1991, pp. 284–287), we apply the difference operator $(1 - B)(1 - B^{12})$ to $\{V_t\}$ to obtain the new series $Y_t = (1 - B)(1 - B^{12})V_t$, shown in Figure 3c, which does not display any apparent deviations from stationarity. The first 125 transformed observations are used for estimation, and the last 6 observations are left for prediction. The WNW estimates of conditional distribution of Y_t given $X_t = Y_{t-1}$ are depicted in Figure 3d for two different values of X_t (0.0086, −0.0501), and Figure 3e gives the WNW estimates of two conditional quantiles. Figure 3f presents the estimated surface of the conditional quantile. Both Figures 3d and 3e indicate that two curves are almost parallel, which implies that there is no clear indication that there exists heteroskedasticity of $\{Y_t\}$ given X_t. Now we consider the forecasting for the last 6 observations based on both the WNW and NW estimators, which are computed by using the same bandwidth for each case. The 95% prediction intervals are reported in Table 2, which shows that the WNW method outperforms the NW approach, although all predictive intervals based on both methods contain the corresponding true values. The average lengths of the intervals for WNW and NW are 0.166 and 0.179, respectively, which are 58.7% and 63.6% of the range of the data. Therefore, we can conclude that without making any complex model assumptions, the prediction intervals based on the WNW method for time series data work reasonably well.

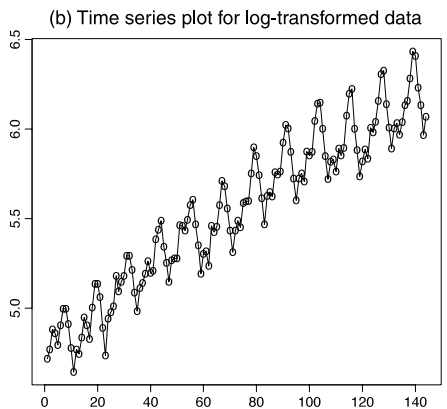
TABLE 2. The postsample predictive intervals for airline data

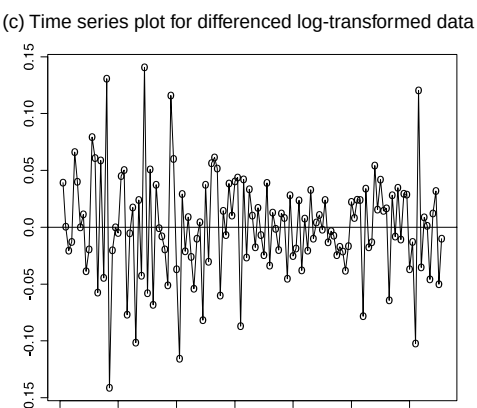
Observation	True value	P.I. for WNW	P.I. for NW
Y_{126}	0.001	[−0.101, 0.045]	[−0.101, 0.045]
Y_{127}	−0.046	[−0.101, 0.055]	[−0.101, 0.067]
Y_{128}	0.012	[−0.102, 0.131]	[−0.102, 0.131]
Y_{129}	0.032	[−0.101, 0.042]	[−0.101, 0.042]
Y_{130}	−0.050	[−0.086, 0.044]	[−0.086, 0.044]
Y_{131}	−0.010	[−0.054, 0.131]	[−0.115, 0.141]

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press







https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

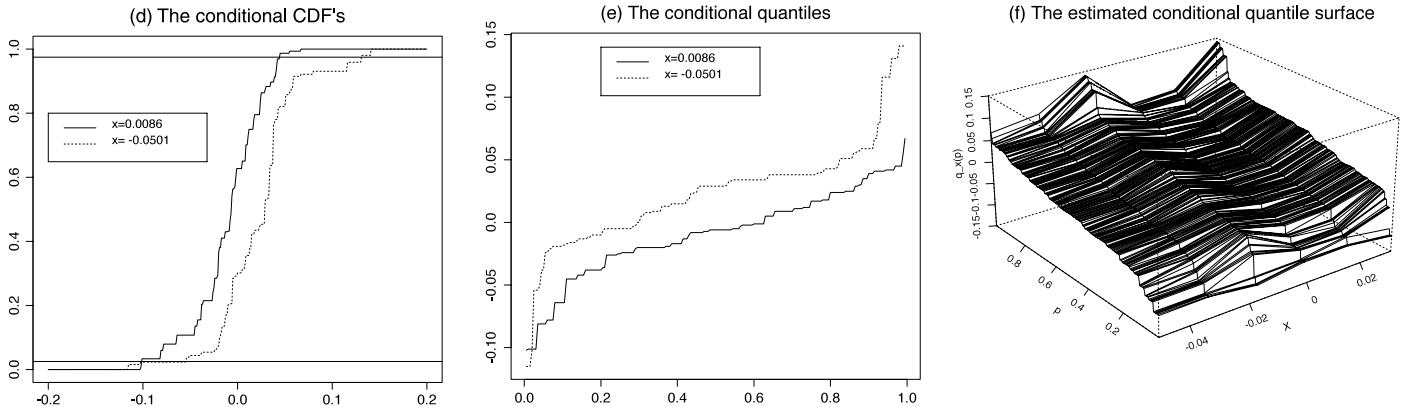


FIGURE 3. (a) Time series plot of Airline data. (b) Time series plot of log-transformed data. (c) Time series plot of twice-differenced log-transformed data. (d) WNW estimates of conditional CDF's for two different values of covariate X_t (0.0086, −0.0501). (e) WNW estimates of conditional quantiles for two different values of covariate X_t (0.0086, −0.0501). (f) Estimated conditional quantile $q_x(p)$ of Y_t given $Y_{t-1} = x$.

REFERENCES

Auestad, B. & D. Tjøstheim (1990) Identification of nonlinear time series: First order characterization and order determination. *Biometrika* 77, 669–687.

Bhattacharya, P.K. & A.K. Gangopadhyay (1990) Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics* 18, 1400–1415.

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

Box, G.E.P. & G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

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

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

Chen, R. & R.S. Tsay (1993) Functional-coefficient autoregressive models. *Journal of the American Statistical Association* 88, 298–308.

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

Cole, T.J. (1988) Fitting smoothed centile curves to reference data. *Journal of the Royal Statistical Society, Series A* 151, 385–418.

Cole, T.J. & P.J. Green (1992) Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine* 11, 1305–1319.

Engle, R.F. & Granger, C.W.J. (1991) *Long-Run Economic Relationships*. Oxford, UK: Oxford University Press.

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

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

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

Faraway, J. & C. Chatfield (1998) Time series forecasting with neural networks: A comparative study using the airline data. *Applied Statistics* 47, 231–250.

Gorodetskii, V.V. (1977) On the strong mixing property for linear sequences. *Theory of Probability and Its Applications* 22, 411–413.

Granger, C.W.J. & T. Teräsvirta (1993) *Modeling Nonlinear Economic Relationships*. Oxford, UK: Oxford University Press.

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.

Härdle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

Hogg, R.V. (1975) Estimates of percentile regression lines using salary data. *Journal of the American Statistical Association* 70, 56–59.

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

Koenker, R. (1994) Confidence intervals for regression quantiles. In P. Mandl and M. Huskova (eds.), *Proceedings of the Fifth Prague Symposium on Asymptotic Statistics*, pp. 349–359. Heidelberg: Physica.

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

Koenker, R. & G.W. Bassett (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50, 43–61.

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

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory* 11, 258–289.
Masry, E. & D. Tjøstheim (1997) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–252.
Powell, J.L. (1986) Censored regression quantiles. *Journal of Econometrics* 32, 143–155.
Roussas, G.G. (1969) Nonparametric estimation of the transition distribution function of a Markov process. *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.
Shao, Q. & H. Yu (1996) Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* 24, 2098–2127.
Stone C.J. (1977) Consistent nonparametric regression (with discussion). *Annals of Statistics* 5, 595–645.
Troung, Y.K. (1989) Asymptotic properties of kernel estimators based on local median. *Annals of Statistics* 17, 606–617.
Troung, Y.K. & C.J. Stone (1992) Nonparametric function estimation involving time series. *Annals of Statistics* 20, 77–97.
Tucker, H. (1967) *A Graduate Course in Probability*. New York: Academic Press.
Volkonskii, V.A. & Yu.A. Rozanov (1959) Some limit theorems for random functions. I. *Theory of Probability and Its Applications* 4, 178–197.
Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: SIAM.
Wand, M.P. & M.C. Jones (1995) *Kernel Smoothing*. London: Chapman and Hall.
Withers, C.S. (1981) Conditions for linear processes to be strong mixing. *Zeitschrift für Wahrscheinlichkeitstheorie verwandte Gebiete* 57, 477–480.
Yu, K. & 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. & M.C. Jones (1998) Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.
Zhou, K.Q. & Portnoy, S.L. (1996). Direct use of regression quantiles to construct confidence sets in linear models. *Annals of Statistics* 24, 287–306.

APPENDIX

Note that we use the same notation as in Sections 2 and 3. Let $\varepsilon_t = I(Y_t \leq y) - F(y|X_t)$ and

$$b_t(x) = \left[1 - \frac{h\mu_2 g'(x)}{2\nu_2 g(x)}\,(X_t - x)K_h(x - X_t)\right]^{-1}, \tag{A.1}$$

where $\nu_2 = \int u^2 K^2(u)\,du$. Then,

$$J_1 \equiv \sqrt{\frac{h}{n}} \sum_{t=1}^{n} b_t(x)\varepsilon_t K_h(x - X_t) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \zeta_t, \tag{A.2}$$

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

where $\zeta_t = \sqrt{h} b_t(x)\varepsilon_t K_h(x - X_t)$. Let C denote a generic constant that might take a different value at different places.

LEMMA 1. *Under assumptions B1–B5, we have*

$$\mathrm{var}(J_1) \to \nu_0 F(y|x)[1 - F(y|x)]g(x) = \sigma^2(y|x)g^2(x) \equiv \theta^2(y|x).$$

Proof. It is easy to see that $E[\zeta_t] = 0$ because $E[\varepsilon_t | X_t] = 0$ and

$$\mathrm{var}(J_1) = E[\zeta_t^2] + \sum_{t=2}^{n} \left(1 - \frac{t-1}{n}\right) \mathrm{cov}(\zeta_1, \zeta_t). \tag{A.3}$$

A straightforward manipulation yields

$$E[\zeta_t^2] = hE[b_t^2(x)\varepsilon_t^2 K_h^2(x - X_t)] = \sigma^2(y|x)g^2(x) + o(1).$$

Choose $d_n = O(h^{-1/(1+\delta/2)})$ and decompose the second term on the right-hand side of (A.3) into two terms as follows:

$$\sum_{t=2}^{n} = \sum_{t=2}^{d_n} + \sum_{t=d_n+1}^{n} \equiv J_{11} + J_{12}.$$

For J_{11}, it follows by condition B5 that $|\mathrm{cov}(\zeta_1, \zeta_t)| \le Ch$, so that $J_{11} = O(d_n h) = o(1)$. For J_{12}, condition B2 implies that $K_h(\cdot) \le Ch^{-1}$ and $|(X_t - x)K_h(x - X_t)| \le C$, so that $|\zeta_t| < Ch^{-1/2}$. Then, it follows from Theorem 17.2.1 of Ibragimov and Linnik (1971, p. 306) that

$$|\mathrm{cov}(\zeta_1, \zeta_t)| \le Ch^{-1}\alpha(t-1),$$

which implies that

$$J_{12} \le Ch^{-1} \sum_{t \ge d_n} \alpha(t) \le Ch^{-1} d_n^{-(1+\delta)} = o(1).$$

This completes the proof of the lemma. ■

LEMMA 2. *Under assumptions B1–B5, we have*

$$\lambda = -\frac{h\mu_2 g'(x)}{2\nu_2 g(x)}\{1 + o_p(1)\} \quad \text{and} \quad p_t(x) = b_t(x)\{1 + o_p(1)\}.$$

Proof. Define, for $1 \le j \le 3$,

$$A_j = \frac{1}{n} \sum_{t=1}^{n} (X_t - x)^j K_h^j(x - X_t).$$

Using the same arguments as those in Lemma 1, we have

$$A_1 = -\frac{1}{2}\mu_2 h^2 g'(x) + o_p(h^2), \qquad A_2 = h\nu_2 g(x) + o_p(h^2), \qquad \text{and} \qquad A_3 = O_p(h^2). \tag{A.4}$$

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

By (6.4) in Chen and Hall (1993),

$$|\lambda| \le \frac{|A_1|}{A_2 - C_1|A_1|} = O_p(h).$$

By a Taylor expansion,

$$0 = A_1 - \lambda A_2 + \lambda^2 A_3 - \lambda^3 A_4 + \cdots,$$

so that

$$\lambda = \frac{A_1}{A_2} + \lambda^2 \frac{A_3}{A_2} - \lambda^3 \frac{A_4}{A_2} + \cdots.$$

Therefore, substituting (A.4) into the preceding equation, we prove the lemma. ■

Proof of Theorem 1. It follows from Lemma 2 that

$$\hat{F}(y|x) - F(y|x) = \frac{\sum_{t=1}^{n} [I(Y_t \le y) - F(y|x)] p_t(x) K_h(x - X_t)}{\sum_{t=1}^{n} p_t(x) K_h(x - X_t)}$$

$$\equiv \{(nh)^{-1/2} J_1 + J_2\} J_3^{-1} \{1 + o_p(1)\},$$

where

$$J_2 = \frac{1}{n} \sum_{t=1}^{n} [F(y|X_t) - F(y|x)] p_t(x) K_h(x - X_t)$$

and

$$J_3 = \frac{1}{n} \sum_{t=1}^{n} b_t(x) K_h(x - X_t).$$

By condition B1 and (1) and also by the Taylor expansion, we have

$$J_2 = \frac{1}{2n} \sum_{t=1}^{n} F^{(2)}(y|x)(X_t - x)^2 b_t(x) K_h(x - X_t) + o_p(h^2) = B(y|x) g(x) + o_p(h^2)$$

by following the same line as that used in the proof of Lemma 1. Similarly,

$$J_3 = g(x) + o_p(1). \tag{A.5}$$

Therefore,

$$\sqrt{nh}[\hat{F}(y|x) - F(y|x) - B(y|x) + o_p(h^2)] = g^{-1}(x) J_1 + o_p(1). \tag{A.6}$$

This, in conjunction with Lemma 1, implies (3). To prove (4), it suffices to establish the asymptotic normality of J_1 by (A.6). To this end, we employ Doob's small-block and large-block technique (see, e.g., Ibragimov and Linnik, 1971, p. 316). Namely, partition

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

$\{1,\ldots,n\}$ into $2q_n + 1$ subsets with large block of size $r = r_n$ and small block of size $s = s_n$. Set

$$q = q_n = \left\lfloor \frac{n}{r_n + s_n} \right\rfloor, \tag{A.7}$$

where $\lfloor x \rfloor$ denotes the integer part of x. Define the random variables, for $0 \leq j \leq q-1$,

$$\eta_j = \sum_{i=j(r+s)}^{j(r+s)+r-1} \zeta_i, \qquad \xi_j = \sum_{i=j(r+s)+r}^{(j+1)(r+s)} \zeta_i, \qquad \text{and} \quad \eta_q = \sum_{i=q(r+s)}^{n-1} \zeta_i.$$

Then,

$$J_1 = \frac{1}{\sqrt{n}} \left\{ \sum_{j=0}^{q-1} \eta_j + \sum_{j=0}^{q-1} \xi_j + \eta_q \right\} \equiv \frac{1}{\sqrt{n}} \{Q_{n,1} + Q_{n,2} + Q_{n,3}\}.$$

We show that, as $n \to \infty$,

$$\frac{1}{n} E[Q_{n,2}]^2 \to 0, \qquad \frac{1}{n} E[Q_{n,3}]^2 \to 0, \tag{A.8}$$

$$\left| E[\exp(itQ_{n,1})] - \prod_{j=0}^{q-1} E[\exp(it\eta_j)] \right| \to 0, \tag{A.9}$$

$$\frac{1}{n} \sum_{j=0}^{q-1} E(\eta_j^2) \to \theta^2(y|x), \tag{A.10}$$

where $\theta^2(y|x)$ is defined in Lemma 1 and

$$\frac{1}{n} \sum_{j=0}^{q-1} E[\eta_j^2 I\{|\eta_j| \geq \varepsilon\theta(y|x)\sqrt{n}\}] \to 0 \tag{A.11}$$

for every $\varepsilon > 0$. Expression (A.8) implies that $Q_{n,2}$ and $Q_{n,3}$ are asymptotically negligible in probability. Expression (A.9) shows that the summands η_j in $Q_{n,1}$ are asymptotically independent. Expressions (A.10) and (A.11) are the standard Lindeberg–Feller conditions for asymptotic normality of $Q_{n,1}$ for the independent setup.

Let us first establish (A.8). To this effect, we define the large-block size r_n by $r_n = \lfloor (nh_n)^{1/2} \rfloor$ and the small-block size $s_n = \lfloor (nh_n)^{1/2}/\log n \rfloor$. Then, as $n \to \infty$,

$$s_n/r_n \to 0 \quad \text{and} \quad (n/r_n)\alpha(s_n) \to 0. \tag{A.12}$$

Observe that

$$E[Q_{n,2}]^2 = \sum_{j=0}^{q-1} \operatorname{var}(\xi_j) + 2 \sum_{0 \leq i < j \leq q-1} \operatorname{cov}(\xi_i, \xi_j) \equiv F_1 + F_2. \tag{A.13}$$

It follows from stationarity and Lemma 1 that

$$F_1 = q_n \operatorname{var}(\xi_1) = q_n \operatorname{var}\left(\sum_{j=1}^{s_n} \zeta_j \right) = q_n s_n [\theta^2(y|x) + o(1)]. \tag{A.14}$$

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

Next consider the second term F_2 on the right-hand side of (A.13). Let $r_j^* = j(r_n + s_n)$; then $r_j^* - r_i^* \geq r_n$ for all $j > i$. We therefore have

$$|F_2| \leq 2 \sum_{0 \leq i < j \leq q-1} \sum_{j_1=1}^{s_n} \sum_{j_2=1}^{s_n} |\operatorname{cov}(\zeta_{r_i^*+r_n+j_1}, \zeta_{r_j^*+r_n+j_2})| \leq 2 \sum_{j_1=1}^{n-r_n} \sum_{j_2=j_1+r_n}^{n} |\operatorname{cov}(\zeta_{j_1}, \zeta_{j_2})|.$$

By stationarity and Lemma 1, one obtains

$$|F_2| \leq 2n \sum_{j=r_n+1}^{n} |\operatorname{cov}(\zeta_1, \zeta_j)| = o(n). \tag{A.15}$$

Hence, by (A.12)–(A.15), we have

$$\frac{1}{n} E[Q_{n,2}]^2 = O(q_n \quad s_n \quad n^{-1}) + o(1) = o(1). \tag{A.16}$$

It follows from stationarity, (A.12), and Lemma 1 that

$$\operatorname{var}[Q_{n,3}] = \operatorname{var}\left(\sum_{j=1}^{n-q_n(r_n+s_n)} \zeta_j \right) = O(n - q_n(r_n + s_n)) = o(n). \tag{A.17}$$

Combining (A.12), (A.16), and (A.17), we establish (A.8). As for (A.10), by stationarity, (A.12), and Lemma 1, it is easily seen that

$$\frac{1}{n} \sum_{j=0}^{q_n-1} E(\eta_j^2) = \frac{q_n}{n} E(\eta_1^2) = \frac{q_n r_n}{n} \cdot \frac{1}{r_n} \operatorname{var}\left(\sum_{j=1}^{r_n} \zeta_j \right) \to \theta^2(y|x).$$

To establish (A.9), we make use of Lemma 1.1 of Volkonskii and Rozanov (1959) to obtain

$$\left| E[\exp(itQ_{n,1})] - \prod_{j=0}^{q_n-1} E[\exp(it\eta_j)] \right| \leq 16(n/r_n)\alpha(s_n)$$

tending to zero by (A.12).

It remains to establish (A.11). To this end, we employ Theorem 4.1 of Shao and Yu (1996) and condition B4 to obtain

$$\begin{aligned} E[\eta_1^2 I\{|\eta_1| \geq \varepsilon\theta(y|x)\sqrt{n}\}] &\leq Cn^{-\delta/2} E(|\eta_1|^{2+\delta}) \\ &\leq Cn^{-\delta/2} r_n^{(2+\delta)/2} \{E(|\zeta_1|^{2(2+\delta)})\}^{1/2}. \end{aligned} \tag{A.18}$$

It is easy to see that

$$E(|\zeta_1|^{2(2+\delta)}) \leq Ch_n^{-(1+\delta)}. \tag{A.19}$$

Therefore, by (A.18) and (A.19),

$$E[\eta_1^2 I\{|\eta_1| \geq \varepsilon\theta(y|x)\sqrt{n}\}] \leq Cn^{-\delta/2} r_n^{(2+\delta)/2} h_n^{-(1+\delta)/2}.$$

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

Thus, by (A.7) and the definition of r_n, and using condition B6, we obtain

$$\frac{1}{n}\sum_{j=0}^{q-1} E[\eta_j^2 I\{|\eta_j| \geq \varepsilon\theta(y|x)\sqrt{n}\}] \leq C r_n^{\delta/2} n^{-\delta/2} h_n^{-(1+\delta)/2} \leq C(nh^{1+2/\delta})^{-\delta/4} \to 0.$$

This completes the proof of the theorem. ■

To prove Theorem 2, we need the following lemma.

LEMMA 3. *Under assumptions B1–B5, we have*

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

where

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

Proof. Let $\hat{\lambda} = \text{argmax}_\lambda L_n(\lambda)$ so that $L_n'(\hat{\lambda}) = 0$, where $L_n(\cdot)$ is defined in (2). It suffices to show that $\hat{\lambda} \to \lambda_c$ in probability. To this end, denote by S_ε the interval $\lambda_c \pm \varepsilon$. We show that for any sufficiently small ε, the probability

$$\sup_{\lambda \in S_\varepsilon} L_n(\lambda) \leq L_n(\lambda_c)$$

tends to one. By the Taylor expansion,

$$L_n(\lambda) - L_n(\lambda_c) = L_n'(\lambda_c)(\lambda - \lambda_c) + \frac{1}{2} L_n''(\lambda_c)(\lambda - \lambda_c)^2 + \frac{1}{6} L_n'''(\lambda^*)(\lambda - \lambda_c)^3$$

with λ^* lying between λ and λ_c. It is easy to show that

$$L_n'(\lambda_c) = o_p(1), \qquad L_n''(\lambda_c) = -\beta_3(c)g(0+) + o_p(1), \qquad \text{and} \quad L_n'''(\lambda^*) = O_p(1),$$

where

$$\beta_3(c) = \int_{-1}^{c} \frac{u^2 K^2(u)}{[1 - \lambda_c u K(u)]^2}\, du.$$

This concludes with probability tending to one that when ε is small enough, for all $\lambda \in S_\varepsilon$,

$$L_n(\lambda) - L_n(\lambda_c) \leq 0,$$

which completes the proof of the lemma. ■

Proof of Theorem 2. By replacing $b_t(x)$ in ζ_t by $b_t^c(ch)$ and following the same arguments as those used in the proof of Theorem 1, we can prove the theorem via Lemma 3. ■

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

LEMMA 4. *Under assumptions B1–B5 and C1, we have, for any* $\delta_n \to 0$,

$$\hat{F}(y+\delta_n|x) - \hat{F}(y|x) = f(y|x)\delta_n + o_p(\delta_n) + o_p((nh)^{-1/2}).$$

Proof. Let

$$J_4 = \frac{1}{n}\sum_{t=1}^{n} b_t(x)[I(Y_t \le y+\delta_n) - I(Y_t \le y)]K_h(x - X_t).$$

Then,

$$\hat{F}(y+\delta_n|x) - \hat{F}(y|x) = J_4 J_3^{-1}\{1 + o_p(1)\}. \tag{A.20}$$

It is easy to see that

$$E(J_4) = E[\{F(y+\delta_n|x) - F(y|x)\}b_t(x)K_h(x - X_t)] = f(y|x)g(x)\delta_n + o(\delta_n).$$

Similarly,

$$\mathrm{var}(J_4) = O(\delta_n(nh)^{-1}).$$

Therefore,

$$J_4 = f(y|x)g(x)\delta_n + o_p(\delta_n) + o_p((nh)^{-1/2}).$$

This, coupled with (A.5) and (A.20), proves the lemma. ■

We now embark on the proof of Theorem 3.

Proof of Theorem 3. First, we prove (9). To this end, by (3), we have, for all x and y,

$$\hat{F}(y|x) \to F(y|x) \quad \text{in probability.}$$

It follows by Theorem 1 of Tucker (1967, pp. 127–128) that

$$\sup_{y\in\Re} |F(y|x) - F(y|x)| \to 0 \quad \text{in probability} \tag{A.21}$$

because $F(y|x)$ is a distribution function. The assumption that $q_p(x)$ is unique implies that, for any fixed x, there is an $\varepsilon = \varepsilon(x) > 0$ such that

$$\delta = \delta(\varepsilon) = \min\{p - F(q_p(x) - \varepsilon|x), F(q_p(x) + \varepsilon|x) - p\} > 0.$$

It is easy to see that the following inequalities hold:

$$P\{|\hat{q}_p(x) - q_p(x)| > \varepsilon\} \le P\{|F(\hat{q}_p(x)|x) - p| > \delta\}$$

$$\le P\left\{\sup_y |\hat{F}(y|x) - F(y|x)| > \delta\right\},$$

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press

which tends to zero by (A.21). Therefore, (9) holds. We now prove (10). For any u, let $\delta_n = B_p(x) + (nh)^{-1/2}\sigma_p(x)u$. Then,

$$\begin{aligned} Q_n(u) &\equiv P[\sqrt{nh}\sigma_p^{-1}(x)\{\hat{q}_p(x) - q_p(x) - B_p(x) + o_p(h^2)\} \le u] \\ &\approx P\{\hat{q}_p(x) \le q_p(x) + \delta_n\} = P\{\hat{F}(q_p(x) + \delta_n|x) \ge p\} \\ &\approx P\{\hat{F}(q_p(x)|x) \ge -f(q_p(x)|x)\delta_n + p\} \end{aligned}$$

by Lemma 4. Therefore,

$$Q_n(u) \approx P[\sqrt{nh}\sigma^{-1}(q_p(x)|x)\{\hat{F}(q_p(x)|x) - p - B(q_p(x)|x)\} \ge -u] \approx \Phi(u)$$

by Theorem 1, where $\Phi(\cdot)$ is the standard normal distribution. Therefore, we finish the proof of the theorem. ■

https://doi.org/10.1017/S0266466602181096 Published online by Cambridge University Press