Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

# NONLINEAR QUANTILE REGRESSION UNDER DEPENDENCE AND HETEROGENEITY

**Walter Oberhofer and Harry Haupt**
*University of Regensburg*

Regensburger Diskussionsbeiträge zur Wirtschaftswissenschaft 388
University of Regensburg Discussion Papers in Economics 388

UNIVERSITÄT REGENSBURG
Wirtschaftswissenschaftliche Fakultät

# NONLINEAR QUANTILE REGRESSION UNDER DEPENDENCE AND HETEROGENEITY

Walter Oberhofer and Harry Haupt

Address correspondence to:
Harry Haupt, Department of Economics and Econometrics, University of Regensburg, Universitaetsstr. 31, 93053 Regensburg, Germany, e-mail: harald.haupt@wiwi.uni-r.de.

Abstract:
This paper derives the asymptotic normality of the nonlinear quantile regression estimator with dependent errors. The required assumptions are weak, and it is neither assumed that the error process is stationary nor that it is mixing. In fact, the notion of weak dependence introduced in this paper, can be considered as a quantile specific local variant of known concepts. The connection of the derived asymptotic results to corresponding results of least squares estimation is obvious.

Kurzfassung:
In dieser Arbeit wird die asymptotische Normalität des nichtlinearen Quantilsregressionsschätzers bei abhängigen Fehlertermen bewiesen. Die Annahmen die dabei zu Grunde liegen sind sehr schwach, wobei gezeigt wird, dass weder die Stationarität noch eine Mixing-Eigenschaft des Fehlerprozesses erforderlich sind. Von besonderer Bedeutung ist die in diesem Papier eingeführte quantilsspezifische Form von schwacher Abhängigkeit, die als lokale Variante existierender Konzepte interpretiert werden kann. Zudem zeigt sich, dass die Asymptotik starke Parallelen zum Fall der Minimumquadratschätzung aufweist.

JEL classification: C22.

# 1 Introduction

THE CONCEPT OF QUANTILE REGRESSION, introduced in the seminal paper of Koenker and Bassett (1978), has become a widely used and accepted technique in many areas of theoretical and applied econometrics. Many of the numerous research frontiers in this fast evolving field have been reviewed and summarized in recent survey articles (see inter alia Buchinsky, 1998, and Yu et al., 2003). In addition to the more methodological literature, there are also important, non-technical attempts to bring the key concepts and especially the applicability of quantile estimation to a wider audience outside the statistical profession (see for example Koenker and Hallock, 2001).

In this paper we consider the case where the dependent variable $y$ and covariates $x_1, \ldots, x_K$ satisfy a nonlinear model with additive errors. Often, the error process is assumed to be independent and identically distributed (i.i.d.). This assumption has been challenged in different ways in the literature. Koenker and Bassett (1982) first investigated the case of heteroscedasticity based on regression quantiles, other authors discussed this case for the most prominent quantile, the median (see for example Knight, 1999, Zhao, 2001, and the literature cited there). Nonlinear quantile regression models have been discussed in Oberhofer (1982), Weiss (1991), Koenker and Park (1994), and Mukherjee (2000), with the first two papers considering least absolute deviations (LAD), and the second paper making a weak dependence assumption in the form of strongly mixing errors. Quantile regression with dependent errors have been discussed for LAD estimation by Phillips (1991), for unconditional quantiles in a parametric context by Oberhofer and Haupt (2005), and in a nonparametric context by De Gooijer and Zerom (2003) and Ioannides (2004). In the context of pure time series models, the nonparametric estimation of regression quantiles under dependence has been discussed recently by Cai (2002), who also provides a survey of the preceding literature in this context.

In this paper we allow the error process to be heteroscedastic and weakly dependent. It is well known that in quantile regression no moments of the error process are needed and that the density of the error process enters instead of the variance. In the case of weak dependence, however, it is usually assumed that the random processes considered are strong mixing (and stationary). Existing asymptotic results in this contexts often rely — at least in an econometric approach — on concepts established in the various works (mainly on the general class of optimization estimators) of Halbert White and co-authors (see inter alia Gallant and White, 1988, Wooldridge and White, 1988, and White, 1994).

We prove the asymptotic normality of the regression quantiles under weak assumptions, utilizing specific properties of regression quantiles under weak dependence. It becomes evident, that the weak dependence condition is a generic consequence of the quantile specific modelling process and is the only condition required for an indicator process depending on the sign of the error process. Similar to the case of i.i.d. disturbances, the asymptotic distribution of the estimator is also strongly connected to the corresponding distribution appearing in least squares estimation.

The following Section 2 introduces the model framework and derives the loss function of conditional quantile estimation. The remainder of the paper is organized as follows: In Section 3 we establish and discuss the required assumptions for asymptotic normality and we derive the limiting distribution of regression quantiles under dependence for the special case of a linear regression function. This discussion is a convenient starting point for the discussion of the general nonlinear case which follows in Section 4. In the appendix we prove some necessary preliminary results.

# 2 Regression quantiles

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let $\{y_t\}_{t\in\mathbb{N}}$ be an $\mathcal{F}$-measurable random sequence with right continuous distribution function $F_t(y)$. Then, consider the regression model

$$(2.1) \qquad y_t - g(x_t, \beta_0) = u_t, \quad 1 \leq t \leq T,$$

where $\beta_0 \in D_\beta \subset \mathbb{R}^K$ is a vector of unknown parameters, the $1 \times L$ vectors $x_t$ are deterministic and given, the dependent variables $y_t$ are observable, and $g(x, \beta_0)$ is in general a nonlinear function defined for $x \in D_x$ and $\beta \in D_\beta$ from $D_x \times D_\beta \to \mathbb{R}$, where $x_t \in D_x$ for all $t$.

We assume that $g(x, \beta)$ has the following Taylor expansion (with remainder) for all $x \in D_x$ and $\beta$ in the neighborhood of $\beta_0$:

$$g(x,\beta) = g(x,\beta_0) + \frac{\partial g(x,\beta)}{\partial \beta'}|_{\beta=\beta_0}(\beta - \beta_0) + (\beta-\beta_0)'\left(\frac{1}{2}\frac{\partial^2 g(x,\beta)}{\partial\beta\partial\beta'}|_{\beta=\beta^*}\right)(\beta-\beta_0),$$

where $\beta^* = \beta_0 + \xi(\beta - \beta_0)$ and $0 < \xi < 1$. For ease of notation we introduce the row vector

$$(2.2) \qquad w_t \equiv \frac{\partial g(x_t,\beta)}{\partial \beta'}|_{\beta=\beta_0}$$

and the $K \times K$ matrix

$$(2.3) \qquad W_t(\beta_t^*) \equiv \left( \frac{1}{2} \frac{\partial^2 g(x_t, \beta)}{\partial \beta \partial \beta'} |_{\beta=\beta_t^*} \right).$$

According to the growth of the components of $w_t$ depending on $t$, the components of the estimator have to be normalized. Therefore, we introduce the $K \times K$ diagonal matrix $D_T = \text{diag}(d_{1T}, \ldots, d_{KT})$, where

$$(2.4) \qquad d_{iT} \equiv \sqrt{\frac{1}{T} \sum_{t=1}^{T} w_{it}^2}, \quad 1 \leq i \leq K.$$

It is assumed that $d_{iT}$ is positive for $1 \leq i \leq K$ and for large enough $T$. For later convenience we define

$$(2.5) \qquad h_t \equiv g(x_t, \beta) - g(x_t, \beta_0) = w_t(\beta - \beta_0) + (\beta - \beta_0)' W_t(\beta_t^*)(\beta - \beta_0),$$

the $1 \times K$ vectors

$$(2.6) \qquad z_{tT} \equiv w_t D_T^{-1},$$

constituting the rows of the $T \times K$ matrix $Z_T$, and the $K \times 1$ vectors

$$(2.7) \qquad v \equiv \sqrt{T} D_T (\beta - \beta_0).$$

Note that due to (2.7), $\beta_t^*$ can be rewritten as $\beta_t^* = \beta_0 + \xi T^{-1/2} D_T^{-1} v$. Then, due to the definitions (2.4)-(2.7), the left hand side of the estimating equation based on (2.1) can be transformed in the following way:

$$\begin{aligned} (2.8) \; y_t - g(x_t, \beta) &= g(x_t, \beta_0) + u_t - g(x_t, \beta) \\ &= u_t - h_t = u_t - \frac{1}{\sqrt{T}} z_{tT} v - \frac{1}{T} v' D_T^{-1} W_t(\beta_t^*) D_T^{-1} v, \quad 1 \leq t \leq T. \end{aligned}$$

Note that $h_t$ depends on $\beta$ in (2.5). However, due to the substitution of $\beta$ by $v$ in (2.7), $h_t$ in (2.8) should be properly denoted as $h_{tT}(v)$. We bear this in mind, but abstain from the latter notation for the sake of notational simplicity.

Our aim is to derive the asymptotic distribution of the $\vartheta$-quantile regression estimator $\hat{\beta}_T$, i.e. $\beta = \hat{\beta}_T$ minimizing the objective function

$$(2.9) \qquad \sum_{t=1}^{T} \left[ \vartheta |y_t - g(x_t, \beta)|_+ + (1 - \vartheta) |y_t - g(x_t, \beta)|_- \right],$$

where $0 < \vartheta < 1$, and for $w \in \mathbb{R}$, we define

$$|w|_+ = \begin{cases} w & \text{if} \quad w > 0, \\ 0 & \text{if} \quad w \leq 0, \end{cases} \qquad \text{and} \quad |w|_- = \begin{cases} 0 & \text{if} \quad w > 0, \\ -w & \text{if} \quad w \leq 0. \end{cases}$$

For technical reasons – i.e. to avoid the need of using $E(u_t)$ – the objective function (2.9) will be rewritten as

$$\sum_{t=1}^{T} \Big[ \vartheta |y_t - g(x_t,\beta)|_+ + (1-\vartheta)|y_t - g(x_t,\beta)|_- - \vartheta|u_t|_+ - (1-\vartheta)|u_t|_- \Big],$$

or, equivalently, under consideration of the derivation of (2.8)

$$\sum_{t=1}^{T} \Big[ \vartheta|u_t - h_t|_+ + (1-\vartheta)|u_t - h_t|_- - \vartheta|u_t|_+ - (1-\vartheta)|u_t|_- \Big]. \tag{2.10}$$

This sum will be considered as a function of $v$, and will be denoted as $A_T(v)$. It defines a scalar random variable depending on $v \in \mathbb{R}^K$, such that if $\hat{v}_T$ is a minimand of $A_T(v)$, the estimator $\hat{\beta}_T$ of the parameter vector $\beta_0$ of the $\vartheta$th regression quantile is, due to (2.7), given by

$$\hat{\beta}_T = \beta_0 + \frac{1}{\sqrt{T}} D_T^{-1} \hat{v}_T.$$

# 3 Linear regression

In this section we discuss the special case of a linear regression function. Thus, instead of (2.1) we consider

$$y_t - x_t\beta_0 = u_t, \quad 1 \leq t \leq T,$$

where $L = K$. Then, by setting $g(x_t, \beta) = x_t\beta$, the definitions (2.2)-(2.5) are rendered to

$$w_t = x_t, \quad W_t = 0, \quad \text{and} \quad h_t = x_t(\beta - \beta_0),$$

and

$$d_{iT} \equiv \sqrt{\frac{1}{T}\sum_{t=1}^{T} x_{it}^2}, \quad 1 \leq i \leq L,$$

which can be interpreted as a measure of the growth of $x_{it}$ and $z_{tT} \equiv x_t D_T^{-1}$ is the normalized regressor vector.

## 3.1 Discussion of the assumptions

By $F_t(z)$ we denote the distribution of $u_t$ and by $F_{s,t}(z,w)$ the common distribution of $(u_s, u_t)$ for $s \neq t$. In this framework, the existence of a measurable estimator $\hat{\beta}_T$ usually is ensured by Theorem 3.10 of Pfanzagl (1969), which, if $\beta_0$ is an inner point of a compact set, is valid under the assumptions stated below. The following assumptions are needed:

(A1) *The* $1 \times L$ *vectors* $x_t$ *are deterministic and known,* $t = 1, 2, \ldots$

(A2) *For some real number* $M$, $|z_{tT}v| \leq M < \infty$ *for* $1 \leq t \leq T$, *and all* $T$, *where* $v \in \mathcal{C}$, *and* $\mathcal{C}$ *is any compact subset of* $\mathbb{R}^K$.

(A3) *The density* $f_t(z)$ *of* $F_t(z)$ *exists in the near of zero, is continuous at* $z = 0$ *uniformly in* $t$, *and* $\lim_{T\to\infty} T^{-3/2} \sum_{t=1}^{T} f_t(0) = 0$.

(A4) *The density* $f_{s,t}(z,w)$ *of* $F_{s,t}(z,w)$ *exists for* $s \neq t$ *in the neighborhood of* $(0,0)$, *is continuous at* $(z,w) = (0,0)$ *uniformly in* $s$ *and* $t$, *and* $\lim_{T\to\infty} T^{-1} \sum_{|k|=1}^{T} \alpha_0(k|\, u) = 0$, *where* $\alpha_0(k|\, u) = \sup |\, f_{t,t+k}(0,0) - f_t(0) f_{t+k}(0)|$, *and the supremum is taken over* $t, t+k \in \mathbb{N}$.

(A5) $F_t(0) = P(u_t \leq 0) = \vartheta$, $0 < \vartheta < 1$, *for all* $t$.

(A6) $T^{-1} Z_T' \Omega_T Z_T$ *converges for* $T \to \infty$ *to a* $K \times K$ *matrix* $\Sigma$, *where* $\Omega_T$ *is a* $T \times T$ *matrix with generic element*

$$\omega_{s,t} = \begin{cases} F_{s,t}(0,0) - \vartheta^2 & \text{for} \quad s \neq t, \\ \vartheta - \vartheta^2 & \text{for} \quad s = t. \end{cases}$$

(A7) $T^{-1} Z_T' \Phi_T Z_T$ *converges for* $T \to \infty$ *to a non-singular* $K \times K$ *matrix* $V$, *where* $\Phi_T$ *is a* $T \times T$ *diagonal matrix with diagonal elements* $\varphi_t = f_t(0)$, $1 \leq t \leq T$.

(A8) *The Bernoulli process*

$$\gamma_t = \begin{cases} -\vartheta & \text{if} \quad u_t > 0, \\ 1 - \vartheta & \text{if} \quad u_t \leq 0. \end{cases}$$

*satisfies the conditions of a CLT (e.g., White, 1994, Theorem A.3.7).*

It is not necessary for the regressors to be deterministic as postulated in assumption (A1). Similar behavior can be expected of random regressors $\{x_t\}$ independent of the disturbances $\{u_t\}$. For example, let $\{x_t\}$ be a stationary sequence with $E(x_t x_t')$ finite and non-singular. Then, by choosing $D_T$ as the identity matrix, almost all realizations would have the necessary limiting properties. Note that in assumption (A2) the boundedness of the expression $z_{tT}v$ is equivalent to that of $x_{iT}/d_{iT}$. The counterpart of (A2) in least squares estimation is $\max_{1\leq t\leq T} T^{-1/2}|x_{it}| \to 0 \quad \text{for} \quad T \to \infty$. For $x_{it} = 1$, assumption (A2) leads to $d_{iT} = 1$ and $z_{itT} = 1$. In the case of a polynomial trend, for example $x_t\beta = \beta_1 + \beta_2 t + \beta_3 t^2$, $d_{1T}$ is of order $O(1)$, $d_{2T}$ is of order $O(T)$, and $d_{3T}$ is of order $O(T^2)$.

If the disturbances are i.i.d., assumptions (A3) and (A4) are implied by the existence of $f(z)$ in the neighborhood of $z = 0$, and the continuity of $f(z)$ at $z = 0$. Usually, in the case of independence, the stronger assumption is made, in which $f_t(0)$ is uniformly bounded, implying (A3). Assumption (A5) is a common normalization in quantile regression.

Assumptions (A4), (A6), and (A8) restrict the dependence structures imposed on the quantile regression model. Generally, the properties of the Bernoulli process $\{\gamma_t\}$ defined in (A8), and the behavior of the distribution functions and densities in the near of $z = 0$, and $(z, w) = (0, 0)$, respectively, are vital for weak dependence concepts in the quantile estimation framework.

By virtue of assumption (A4), a too strong dependence of the errors is excluded. This assumption can be considered as an infinitesimal weak dependence condition and it can be interpreted as a quantile specific variant of the "dependence index sequence" introduced by Castellana and Leadbetter (1986). In the case of independence, the sum in assumption (A4) is equal to zero for all $T$.

Assumption (A6) ensures the existence of the covariance matrix in the limit, and at the same time it reflects the dependence structure and heterogeneity of the error process. Note, that a too strong dependence hinders convergence in (A6). If, for example, all regressors are growing with the same order, we can assume without loss of generality that $D_T$ is equal to the identity matrix and if in addition $x_{1t} = 1$, then, according to (A6), $T^{-1}\sum_{s,t=1}^{T}\omega_{s,t}$ must converge. Further, it is important to note explicitly, that it is not necessary to assume a mixing property that requires the whole $\sigma$-algebras $\sigma(u_t\big|t \leq m)$ and $\sigma(u_t\big|t \geq m+k)$, for all $m = 1, 2, \ldots$, respectively. A peculiarity of quantile regression lies in the fact, that the only thing that matters is a local mixing condition for the point $(0, 0)$. In this sense, for $s \neq t$, we can view $\omega_{s,t} = F_{s,t}(0, 0) - F_s(0)F_t(0)$ as a local measure of dependence (or a local mixing coefficient), and $f_{s,t}(0, 0) - f_s(0)f_t(0)$ as an analogous infinitesimal measure. Here, an interesting peculiarity of quantile estimation arises, which can be seen from the fact that in the assumptions no moments of the error process are required. As we will see, $(\vartheta - \vartheta^2)/f_t(0)^2$ corresponds to variances and $(F_{s,t}(0,0) - \vartheta^2)/f_s(0)f_t(0)$ to covariances. Obviously, independence of the two events $\{\omega|\, \omega \in \Omega, u_s(\omega) \leq 0\}$ and $\{\omega|\, \omega \in \Omega, u_t(\omega) \leq 0\}$ implies $\omega_{s,t} = 0$. If the limit of the matrix $Z_T'\Omega_T Z_T$ is singular, then the limiting distribution of $\sqrt{T}D_t^{-1}(\hat{\beta} - \beta_0)$ is singular, too. In least squares estimation the matrix $Z_T'\Omega_T Z_T$ corresponds to $T^{-1}X_T'E(uu')X_T/\sigma^2$, and $Z_T'\Phi_T Z_T$ in assumption (A7), which controls the form of heteroscedasticity, corresponds to $T^{-1}\sigma^2 X'X$, respectively.

It is clear from the discussion of assumption (A6), that it is neither necessary to assume that the error process is strongly mixing, nor that is is near epoch dependent. Thus, assumption (A8) can be formulated in different ways. It is not necessary to assume that $\{\gamma_t\}$ is strongly mixing (e.g., Oberhofer and Haupt, 2005), since it is possible to employ simple and less abstract moment conditions as in Oberhofer (2005), which in turn are a simple, special case of Doukhan's and Louhichi's (1999) notion of weak dependence. As has been shown by Nze and Doukhan (2004), the latter notion is implied by near epoch dependence. Thus, the use of a CLT based on near epoch dependence on an underlying mixing process (see White, 1994) seems to be quite restrictive in the context of quantile estimation.

## 3.2 Asymptotic normality

The proof of our central theorem requires three preliminary Lemmata, which will be stated and proven in Appendix A. For the derivation and discussion of the asymptotic normality result it is convenient to introduce some further definitions. A typical element of $A_T(v)$ defined in (2.10), is denoted by $a_{tT}(v)$, leading to $A_T(v) = \sum_{t=1}^{T} a_{tT}(v)$, where

$$a_{tT}(v) = \begin{cases} -\vartheta h_t & \text{if} \quad u_t > \max(0, h_t), \\ u_t - \vartheta h_t & \text{if} \quad h_t < u_t \leq 0, \\ -u_t + (1-\vartheta)h_t & \text{if} \quad 0 < u_t \leq h_t, \\ (1-\vartheta)h_t & \text{if} \quad u_t \leq \min(0, h_t). \end{cases} \tag{3.1}$$

Then split up $a_{tT}(v)$ in

$$a_{tT}(v) = b_{tT}(v) + h_t\gamma_t, \tag{3.2}$$

where

$$h_t\gamma_t = \begin{cases} -\vartheta h_t & \text{for} \quad u_t > 0, \\ (1-\vartheta)h_t & \text{for} \quad u_t \leq 0, \end{cases} \tag{3.3}$$

and, consequently

$$b_{tT}(v) = \begin{cases} 0 & \text{if} \quad u_t > \max(0, h_t), \\ u_t - h_t & \text{if} \quad h_t < u_t \leq 0, \\ -u_t + h_t & \text{if} \quad 0 < u_t \leq h_t, \\ 0 & \text{if} \quad u_t \leq \min(0, h_t), \end{cases} \tag{3.4}$$

for $1 \leq t \leq T$, and we define $B_T(v) = \sum_{t=1}^{T} b_{tT}(v)$. The expression $h_t\gamma_t$ defined in (3.3) has an interesting interpretation. It contains the component $h_t$, which arises from the deviation between the regression function and its true value, and a second component, which arises from the error defined in equation (2.1). The decomposition of $a_{tT}(v)$ in (3.2) allows us to study its asymptotic behavior by studying separately that of $b_{tT}(v)$ and $h_t\gamma_t$ in three preliminary Lemmata (given in the Appendix).

Firstly, in Lemma 1 it will be shown that $E[B_T(v)]$ converges to

$$\lim_{T\to\infty} \frac{1}{2}\frac{1}{T} v' Z_T' \Phi_T Z_T v = \frac{1}{2} v' V v.$$

Secondly, in Lemma 2 we prove that

$$\lim_{T\to\infty} \mathrm{Var}[B_T(v)] = 0.$$

Finally, in the proof of Lemma 3 it will be shown that $C_T(v) = \sum_{t=1}^{T} h_t\gamma_t$ converges in distribution to $Cv$, where the $1 \times K$ random vector $C$ is normally distributed with mean zero and covariance matrix

$$\lim_{T\to\infty} \frac{1}{T} Z_T' \Omega_T Z_T.$$

As a consequence, $A_T(v)$ converges in distribution to $A(v) = \frac{1}{2}v'Vv + Cv$, with the minimizing value $\hat{v} = -V^{-1}C'$, and, for $T \to \infty$ the limiting distribution of $\hat{v}_T = T^{1/2}D_T(\hat{\beta}_T - \beta_0)$ will be normal with mean zero and covariance matrix $V^{-1} \lim_{T\to\infty} \left[\frac{1}{T} Z_T'\Omega_T Z_T\right] V^{-1}$. $A(v)$ can be interpreted as the limit of a second-order Taylor approximation of $A_T(v)$. It is interesting, however, that – in the linear case – we do not need a Taylor approximation in the proof. The analogy to the corresponding covariance matrix of ordinary least squares with serially correlated disturbances is obvious.

Theorem 1: *The minimizing value $\hat{v}_T = \sqrt{T}D_T(\hat{\beta}_T - \beta_0)$ of $A_T(v)$ converges in distribution to a normal distribution with mean zero and covariance matrix*

$$V^{-1} \lim_{T\to\infty} \left[\frac{1}{T} Z_T' \Omega_T Z_T\right] V^{-1}. \tag{3.5}$$

PROOF: According to Lemmata 1-3, $A_T(v)$ converges for $T \to \infty$ in distribution to $A(v) = \frac{1}{2}v'Vv + Cv$, where $C$ is normally distributed with mean zero and covariance matrix $\lim_{T\to\infty} T^{-1}Z_T'\Omega_T Z_T$. For the convergence in distribution of the minimizing value $\hat{v}_T$ to $\hat{v}$, it is required, that the function $A_T(v)$ is convex and that $\sum_{t=1}^{T} b_{tT}(v)$ converges uniformly for $v \in \mathcal{C}$, where $\mathcal{C}$ is any compact subset of $\mathbb{R}^K$. That the former requirement is fulfilled has been shown by Pollard (1991) and Geyer (1996), the latter has been shown in Lemma 1.

*Q.E.D.*

# 4 Nonlinear regression function

In the nonlinear case (2.1), in addition to assumptions (A1), (A3)-(A8) we have to assume:

(A2') *There exist real numbers* $M_1 < \infty$ *and* $M_2 < \infty$ *such that* $|w_t D_T^{-1} v| \leq M_1$ *and* $|v' D_T^{-1} W_t(\beta) D_T^{-1} v| \leq M_2$ *in the neighborhood of* $\beta = \beta_0$ *for all* $t$, $T$ *and all* $v \in \mathcal{L}$, *where* $\mathcal{L}$ *is a compact subset of* $\mathbb{R}^K$, *and* $W_t(\beta)$ *is continuous in the near of* $\beta = \beta_0$.

(A9) $\lim_{T\to\infty} T^{-2} \sum_{s,t=1}^{T} |\omega_{s,t}| = 0$.

Note that from defining the coefficient $\lambda_0(k) = \sup_t |\omega_{t,t+k}|$, and the requirement $\lambda_0(k) = o(1)$, follows assumption (A9).

As first step we show that the three preliminary Lemmata 1-3 remain valid in the case of a nonlinear regression function if we replace assumption (A2) by assumptions (A2') and (A9). From (A2') we obtain according to (2.8)

$$h_t = \frac{1}{\sqrt{T}} \left[ z_{tT} v + O(\frac{1}{\sqrt{T}}) \right], \tag{4.1}$$

where $T^{-1/2} z_{tT} v = O(T^{-1/2})$. Consequently, from the definition of $\beta_t^*$ and (2.7), if $T^{-1/2} D_T^{-1}$ converges to zero, we get $\lim_{T\to\infty} \beta_t^* = \beta_0$, and

$$\sum_{t=1}^{T} h_t^2 - \frac{1}{T} \sum_{t=1}^{T} (z_{tT} v)^2 \to 0, \tag{4.2}$$

$$T \sum_{\substack{s,t=1 \\ s \neq t}}^{T} h_s^2 h_t^2 - \frac{2}{T} \sum_{\substack{s,t=1 \\ s \neq t}}^{T} (z_{sT} v)^2 (z_{tT} v)^2 \to 0, \tag{4.3}$$

(4.4) $$|h_t|^3 = O(T^{3/2}),$$

and, due to (A9),

(4.5) $$\sum_{s,t=1}^{T} h_s h_t \omega_{s,t} - \frac{1}{T}\sum_{s,t=1}^{T} [(z_{sT}v)(z_{tT}v)\omega_{s,t}] \to 0.$$

Thus, it is straightforward to verify that the following assertions remain valid, respectively: Lemma 1 due to (4.2), Lemma 2 due to (4.3) and (4.4), and Lemma 3 due to (4.5).

Due to the fact that in the nonlinear case the loss function $A_T(v)$ is not convex in general, unfortunately the proof of Theorem 1 can not be extended in such a simple manner. In Lemma 1 we consider the matrix $\frac{1}{T}Z_T'\Phi_T Z_T$ with the limit $V$. According to assumption (A7), the matrix $V$ is nonsingular and we define

$$\tilde{v}_T \equiv -V^{-1}\Gamma_T,$$

where $\Gamma_T$ is implicitly used in Lemma 3, since

$$C_T(v) = \sum_{t=1}^{T} h_t\gamma_t = v'\sum_{t=1}^{T}\left[\frac{1}{\sqrt{T}}z_{tT}' + \frac{1}{T}D_T^{-1}W_t(\beta_t^*)D_T^{-1}v\right]\gamma_t \equiv v'\Gamma_T.$$

By virtue of Lemma 3 and assumptions (A2') and (A9), $\Gamma_T$ converges in distribution to $C$, and $C$ is normally distributed. Therefore, in order to prove Theorem 1 for the nonlinear case, we have to show $\operatorname{plim}(\hat{v}_T - \tilde{v}_T) = 0$, where $\hat{v}_T$ is defined as $T^{1/2}D_T(\hat{\beta}_T - \beta_0)$. Due to Lemma 1 and Lemma 2, the loss function can be written as

$$A_T(v) = B_T(v) + C_T(v) = \lim_{T\to\infty} E[B_T(v)] + v'\Gamma_T + R_T(v) = \frac{1}{2}v'Vv + v'\Gamma_T + R_T(v).$$

From Lemma 1 and Lemma 2 follows

(4.6) $$\operatorname*{plim}_{T\to\infty} R_T(v) = 0,$$

for $v \in \mathcal{L}$, where $\mathcal{L}$ is defined in (A2'). Due to the definition of $\tilde{v}_T$ we obtain

(4.7) $$A_T(v) - A_T(\tilde{v}_T) = \frac{1}{2}(v_T - \tilde{v}_T)'V(v_T - \tilde{v}_T) + R_T(v) - R_T(\tilde{v}_T).$$

Due to (4.7) and the positive definiteness of $V$, for $v = \hat{v}_T$ – the minimizing value of $A_T(v)$ – for every $\epsilon > 0$, $\eta > 0$ there exists a $T_0$ such that for $T > T_0$

(4.8) $$P\left((\hat{v}_T - \tilde{v}_T)'(\hat{v}_T - \tilde{v}_T) > \epsilon\right) \leq \eta.$$

This implies $\operatorname{plim}(\hat{v}_T - \tilde{v}_T) = 0$, and the proof is complete.

# Appendix: Proofs of the preliminary Lemmata

LEMMA 1: $E\left[\sum_{t=1}^{T} b_{tT}(v)\right]$ *converges for* $T \to \infty$ *to*

$$\frac{1}{2}v'\left[\lim_{T\to\infty}\frac{1}{T}Z_T'\Phi_T Z_T\right]v = \frac{1}{2}v'Vv.$$

*The convergence is uniform for* $v \in \mathcal{C} \subset \mathbb{R}^K$*, where* $\mathcal{C}$ *is any compact set.*

PROOF: By the definitions of $h_t$ and $b_{tT}(v)$ given above, we get

$$E\big[b_{tT}(v)\big] = \begin{cases} \int_0^{h_t}(h_t - z)f_t(z)dz, & \text{if} \quad h_t > 0, \\ \int_{h_t}^{0}(z - h_t)f_t(z)dz, & \text{if} \quad h_t < 0, \end{cases} \tag{A.1}$$

under consideration of $E[b_{tT}(v)] = 0$ for $h_t = 0$. Then, for $h_t > 0$, and under (A3)

$$\frac{1}{2}h_t^2 \inf_{0\le z\le h_t} f_t(z) \le E\big[b_{tT}(v)\big] \le \frac{1}{2}h_t^2 \sup_{0\le z\le h_t} f_t(z). \tag{A.2}$$

The argumentation is analogous for the case $h_t < 0$ and is left to the reader. Due to (A2),

$$\lim_{T\to\infty} h_t = \lim_{T\to\infty}\frac{1}{\sqrt{T}}z_{tT}v = 0.$$

Thus, due to assumptions (A3), (A7), and the definition of $h_t$, from (A.2) follows the proof of the assertion.

*Q.E.D.*

LEMMA 2: *For every* $v \in \mathcal{C} \subset \mathbb{R}^K$*, where* $\mathcal{C}$ *is any compact set,*

$$\lim_{T\to\infty} \mathrm{Var}\left[\sum_{t=1}^{T} b_{tT}(v)\right] = 0.$$

*The convergence is uniform in* $\mathcal{C}$*.*

PROOF: By definition

$$\mathrm{Var}\left[\sum_{t=1}^{T} b_{tT}(v)\right] = \sum_{s,t=1}^{T}\{E[b_s(v)b_t(v)] - E[b_s(v)]E[b_t(v)]\}.$$

Then, for $s \neq t$, $h_s > 0$, $h_t > 0$, and by the definition of $b_{tT}(v)$

$$\begin{aligned} E[b_s(v)b_t(v)] - E[b_s(v)]E[b_t(v)] &= \int\limits_{z\le h_s}\int\limits_{w\le h_t}(h_s - z)(h_t - w)[f_{s,t}(z,w) - f_s(z)f_t(w)]dzdw \\ &\le \frac{1}{4}h_s^2 h_t^2 \sup_{\substack{z\le h_s \\ w\le h_t}} |\, f_{s,t}(z,w) - f_s(z)f_t(w)|. \end{aligned}$$

By analogous argumentation for the remaining cases ($h_s > 0$, $h_t < 0$ etc.), we finally get

$$\text{(A.3)} \qquad \sum_{\substack{s,t=1\\ s\neq t}}^{T} \operatorname{Cov}[b_{sT}(v), b_{tT}(v)] \leq \frac{1}{4} \sum_{\substack{s,t=1\\ s\neq t}}^{T} h_s^2 h_t^2 \sup_{\substack{0\leq |z|\leq |h_s|\\ 0\leq |w|\leq |h_t|}} | f_{s,t}(z,w) - f_s(z) f_t(w)|,$$

where the expression on the right hand side of (A.3) is bounded from above by

$$\frac{1}{4}\frac{1}{T} \sum_{|k|=0}^{T-1} \sup_{t,t+k\in\mathbb{N}} \sup_{\substack{0\leq |z|\leq |h_t|\\ 0\leq |w|\leq |h_{t+k}|}} | f_{t,t+k}(z,w) - f_t(z) f_{t+k}(w)| \frac{1}{T} \sum_{s=1}^{T-k} (z_{sT}v)^2 (z_{s+kT}v)^2.$$

Analogously, for $s = t$, we get

$$\sum_{t=1}^{T} \operatorname{Var}[b_{tT}(v)] \leq \sum_{t=1}^{T} \left[ \frac{1}{3} |h_t|^3 \sup_{0\leq |z|\leq |h_t|} | f_t(z)| \right]$$

instead of (A.3). Then, from $(\mathsf{A2})$ and $(\mathsf{A4})$ follows the assertion.

*Q.E.D.*

LEMMA 3: *$C_T(v) = \sum h_t\gamma_t$ converges for $T \to \infty$ in distribution to $Cv$, where the $1 \times K$ random vector $C$ is normally distributed with mean zero and covariance matrix $\lim_{T\to\infty} T^{-1} Z_T' \Omega_T Z_T$.*

PROOF: Obviously, from the definitions of $h_t$ and $\gamma_t$, we have $E(h_t\gamma_t) = 0$ and $\operatorname{Var}(h_t\gamma_t) = \vartheta(1-\vartheta)h_t^2$, and for $s \neq t$, the covariance between $h_s\gamma_s$ and $h_t\gamma_t$ is given by $h_s h_t \big[\vartheta^2 P(u_s > 0, u_t > 0) + (1-\vartheta)^2 P(u_s \leq 0, u_t \leq 0) - \vartheta(1-\vartheta)P(u_s > 0, u_t \leq 0) - \vartheta(1-\vartheta)P(u_s \leq 0, u_t > 0)\big] = h_s h_t \{\vartheta^2[1 - F_{s,t}(\infty, 0) - F_{s,t}(0,\infty) + F_{s,t}(0,0)] + (1-\vartheta)^2 F_{s,t}(0,0) - \vartheta(1-\vartheta)[F_{s,t}(\infty,0) - F_{s,t}(0,0) + F_{s,t}(0,\infty) - F_{s,t}(0,0)]\}$, where we define $F_{t,t}(0,0) = F_t(0) = \vartheta$. Thus,

$$Cov[h_s\gamma_s, h_t\gamma_t] = h_s h_t \big[F_{s,t}(0,0) - \vartheta^2\big],$$

and, finally

$$\operatorname{Var}[C_T(v)] = \frac{1}{T} v' Z_T' \Omega_T Z_T v.$$

Then, due to assumption $(\mathsf{A8})$, the proof of the assertion follows from the CLT given in White (1994, Theorem A.3.7) and upon application of the Cramér-Wold device.

*Q.E.D.*

## References


[1] Buchinsky, M. (1998): "Recent Advances in Quantile Regression," *Journal of Human Resources*, 33, 88-126.

[2] Cai, Z. (2002): "Regression Quantiles for Time Series," *Econometric Theory*, 18, 169-192.

[3] Castellana, J.V., and Leadbetter, M.R. (1986): "On Smoothed Probability Density Estimation for Stationary Processes," *Stochastic Processes and their Applications*, 21, 179-193.

[4] De Gooijer, J.G., and Zerom, D. (2003): "On Additive Conditional Quantiles with High-Dimensional Covariates," *Journal of the American Statistical Association*, 98, 135-146.

[5] Doukhan, P., and Louhichi, S. (1999): "A New Weak Dependence Condition and Applications to Moment Inequalities," *Stochastic Processes and their Applications*, 84, 313-342.

[6] Doukhan, P. (1994): *Mixing.* Springer Verlag, New York.

[7] Gallant, A.R., and White, H. (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Basil Blackwell, Oxford.

[8] Geyer, C.J. (1996): "On the Asymptotics of Convex Stochastic Optimization," unpublished manuscript.

[9] Ioannides, D.A. (2004): "Fixed Design Regression Quantiles for Time Series," *Statistics and Probability Letters*, 68, 235-245.

[10] Knight, K. (1999): "Asymptotics for $L_1$-Estimators of Regression Parameters under Heteroscadasticity," *Canadian Journal of Statistics*, 27, 497-507.

[11] Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33-50.

[12] Koenker, R., and G. Bassett (1982): "Robust Tests for Heteroscedasticity based on Regression Quantiles," *Econometrica*, 50, 43-61.

[13] Koenker, R., and K. F. Hallock (2001): "Quantile Regression," *Journal of Economic Perspectives*, 15, 143-156.

[14] Koenker, R., and Park, B. (1994): "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265-283.

[15] Mukherjee, K. (2000): "Linearization of Randomly Weighted Empiricals under Long Range Dependence with Applications to Nonlinear Regression Quantiles," *Econometric Theory*, 16, 301-323.

[16] Nze, P.A., and Doukhan, P. (2004): "Weak Dependence: Models and Applications to Econometrics," *Econometric Theory*, 20, 169-192.

[17] Oberhofer, W. (1982): "The Consistency of Nonlinear Regression Minimizing the $L_1$ Norm," *Annals of Statistics*, 10, 316-319.

[18] Oberhofer, W. (2005): "Moment Conditions for Central Limit Theorems under Dependence," unpublished manuscript, University of Regensburg.

[19] Oberhofer, W., and Haupt, H. (2005): "The Asymptotic Distribution of the Unconditional Quantile Estimator under Dependence," *Statistics and Probability Letters*, fortcoming.

[20] Pfanzagl, J. (1969): "On the Measurability and Consistency of Minimum Contrast Estimator," *Metrika*, 14, 249-272.

[21] Phillips, P.C.B. (1991): "A Shortcut to LAD Estimator Asymptotics," *Econometric Theory*, 7, 450-463.

[22] Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186-199.

[23] Weiss, A.A. (1991): "Estimating Nonlinear Dynamic Models using Least Absolute Error Estimation," *Econometric Theory*, 7, 46-68.

[24] White, H. (1994): *Estimation, Inference and Specification Analysis.* Econometric Society Monographs No. 22.

[25] Wooldridge, J.M., White, H. (1988): "Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes," *Econometric Theory*, 4, 210-230.

[26] Yu, K., Z. Lu, and J. Stander (2003): "Quantile Regression: Applications and Current Research Areas," *The Statistician*, 52, 331-350.

[27] Zhao, Q. (2001): "Asymptotic Efficient Median Regression in the Presence of Heteroscedasticity of Unknown Form," *Econometric Theory*, 17, 765-784.