

Abberger, Klaus

**Working Paper**

## Kreuzvalidierung in der nichtparametrischen Quantilsregression

Diskussionsbeiträge - Serie II, No. 254

**Provided in Cooperation with:**

Department of Economics, University of Konstanz

*Suggested Citation:* Abberger, Klaus (1995) : Kreuzvalidierung in der nichtparametrischen Quantilsregression, Diskussionsbeiträge - Serie II, No. 254, Universität Konstanz, Sonderforschungsbereich 178 - Internationalisierung der Wirtschaft, Konstanz

This Version is available at:

<https://hdl.handle.net/10419/101524>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

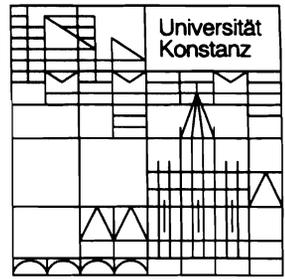
*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

**Sonderforschungsbereich 178  
„Internationalisierung der Wirtschaft“**

Diskussionsbeiträge



Juristische  
Fakultät

Fakultät für Wirtschafts-  
wissenschaften und Statistik

---

Klaus Abberger

**Kreuzvalidierung in der  
nichtparametrischen Quantilsregression**

22. JUNI 1995 Weltwirtschaft  
K.-G.

W 113 (254) mi br sig gla

---

**Kreuzvalidierung in der  
nichtparametrischen Quantilsregression**

611396

**Klaus Abberger**

W 113 (254)



Serie II - Nr. 254

April 1995

# Kreuzvalidierung in der nichtparametrischen Quantilsregression

Klaus Abberger

SFB 178, Universität Konstanz

## **Zusammenfassung:**

Nichtparametrische Verfahren zur Dichte- und Regressionsschätzung setzen die Wahl eines Glättungsparameters voraus. Ein oft verwendetes Verfahren zu dessen Bestimmung ist die Kreuzvalidierung. Die Übertragung dieser Methode auf die Quantilsregression ist Gegenstand der vorliegenden Arbeit. Es werden ein Kreuzvalidierungsverfahren definiert und einige Eigenschaften dieser Vorgehensweise aufgezeigt. Simulationen für verschiedene Verteilungsannahmen sollen weitere Einblicke in die Anwendbarkeit der vorliegenden Prozedur geben.

# 1 Einleitung

Die populärsten nichtparametrischen Schätzverfahren sind die Kernschätzung, die Nächste-Nachbarn-Schätzung und die Splineglättung. Hauptanwendungsbereiche dieser Methoden sind die Dichteschätzung und die Regressionsanalyse, wobei für die letztere als weiteres nichtparametrisches Verfahren die lokal gewichtete Regression zur Verfügung steht. Diesen nichtparametrischen Verfahren ist gemeinsam, daß sie die Wahl eines Glättungsparameters voraussetzen. Ein sehr häufig benutztes Verfahren zur Bestimmung eines Glättungsparameters ist die Kreuzvalidierung. Ziel dieses Aufsatzes ist es, die Anwendbarkeit der Kreuzvalidierung für die nichtparametrische Quantilsregression zu untersuchen. Dazu wird im nächsten Abschnitt ein Kernschätzer für bedingte Quantile definiert und die Kreuzvalidierung zur Ermittlung von geeigneten Bandweiten diskutiert. Im dritten Abschnitt dieser Arbeit wird der Mechanismus für die Bandweitenwahl in Simulationen studiert. Eine Zusammenfassung der Ergebnisse bildet den Abschluß dieses Aufsatzes.

## 2 Nichtparametrische Quantilsregression und Bandweitenwahl

Der Quantilsregression liegt folgendes Modell zugrunde: Es sei  $\{Y, X\}$  eine bivariate Zufallsvariable und  $F(y | x) = F_x(y)$  bezeichne die bedingte Verteilung von  $Y$  gegeben  $X = x$ . Das bedingte  $\alpha$ -Quantil  $q_\alpha(x)$  ist eine Funktion von  $x$ , wobei  $q_\alpha(x) = \inf\{y \in \mathbb{R} | F_x(y) \geq \alpha\}$ ,  $0 < \alpha < 1$ . Gegeben die Beobachtungen  $\{Y_i, X_i\}_{i=1}^n$ , lautet das Modell

$$Y_i = q_\alpha(X_i) + U_i, \quad (1)$$

mit  $P_x(U_i \leq 0) = \alpha$ .

Untersucht werden die Eigenschaften der Kernschätzung bedingter Quantile. Mit einer geeigneten Kernfunktion  $K(\cdot)$  und Bandweiten  $h_n$ , ist ein Kernschätzer für die bedingte

Verteilungsfunktion bestimmt durch

$$F_{n,x}(y) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \mathbb{I}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}. \quad (2)$$

Einen Schätzer für die bedingten  $\alpha$ -Quantile erhält man über die Definition der bedingten  $\alpha$ -Quantile. Er lautet

$$q_{n,\alpha}(x) = \inf\{y \in \mathbb{R} \mid F_{n,x}(y) \geq \alpha\}, 0 < \alpha < 1. \quad (3)$$

Für die folgenden Untersuchungen muß an verschiedenen Stellen der Modellrahmen noch weiter konkretisiert werden. Es sei  $X$  deterministisch und die  $\{Y_i\}_{i=1}^n$  seien unabhängig.

Für die Bandweiten und die Kernfunktion seien folgende Bedingungen erfüllt:

$$h_n \rightarrow 0 \quad , \text{ für } n \rightarrow \infty; \quad (B. 1)$$

$$nh_n \rightarrow \infty \quad , \text{ für } n \rightarrow \infty; \quad (B. 2)$$

$$K \text{ besitze einen kompakten Träger;} \quad (K. 1)$$

$$K \text{ sei Lipschitz-stetig;} \quad (K. 2)$$

$$\int K(u)du = 1; \quad (K. 3)$$

$$K \text{ symmetrisch.} \quad (K. 4)$$

Es seien  $f_x(y) = \frac{\delta F_x(y)}{\delta y}$ ,  $F_x^{(10)}(y) = \frac{\delta F_x(y)}{\delta x}$  und  $F_x^{(20)}(y) = \frac{\delta^2 F_x(y)}{\delta x^2}$  die Bezeichnungen für die Ableitungen der bedingten Verteilungsfunktion. Dann seien folgende Bedingungen erfüllt:

$$\begin{aligned} &\text{die Ableitungen } f_x(y), F_x^{(10)}(y) \text{ und } F_x^{(20)}(y) \\ &\text{existieren in einer Umgebung von } (y, x); \end{aligned} \tag{S. 1}$$

$$X_i = \frac{i - \frac{1}{2}}{n}. \tag{S. 2}$$

Wie soll nun die Bandweite  $h$  für verschiedene  $\alpha$  gewählt werden? Ausgangspunkt ist die Verlustfunktion

$$E \left[ \sum_{i=1}^n \varphi(Y_i - g(X_i)) \right], \tag{4}$$

mit der Quantilsfunktion nach Koenker und Basset  $\varphi(u) = \alpha \mathbb{I}_{\{u \geq 0\}} \cdot u + (\alpha - 1) \mathbb{I}_{\{u < 0\}} \cdot u$ .

Es sei  $E[Y_i | X_i] = \mu_Y(X_i)$ . Dann ist

$$E[\varphi(Y_i - g(X_i))] = \alpha(\mu_Y(X_i) - g(X_i)) + \int_{-\infty}^{g(X_i)} F(y | X_i) dy. \tag{5}$$

Durch Differenzieren erhält man

$$\frac{\delta E[\varphi(Y_i - g(X_i))]}{\delta g(X_i)} = F(g(X_i) | X_i) - \alpha. \tag{6}$$

Der Erwartungswert (4) wird daher minimal für

$$g(X_i) = q_\alpha(X_i), \quad \forall i = 1, \dots, n. \tag{7}$$

und hat im Optimum den Wert

$$E \left[ \sum_{i=1}^n \varphi(Y_i - q_\alpha(X_i)) \right]. \tag{8}$$

Der Ausdruck (4) wird zur Kreuzvalidierung verwendet. Er wird geschätzt durch

$$\sum_i \varphi(Y_i - \hat{g}(X_i)). \quad (9)$$

Setzt man für  $\hat{g}(X_i)$  den Kernschätzer  $q_{n,\alpha}(X_i)$  ein und minimiert über die Bandweite  $h$ , so resultiert stets eine optimale Bandweite von Null, d.h.  $\hat{g}(X_i) = Y_i$ . Daher wird statt des gewöhnlichen Kernschätzers für  $\hat{g}(X_i)$  der leave-one-out Schätzer der bedingten Quantile eingesetzt. Der leave-one-out Schätzer läßt an der Stelle  $X_i$  die zugehörige Beobachtung  $Y_i$  unberücksichtigt. Er ist definiert durch

$$F_{n,X_i}^{(-i)}(y) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h_n}\right) \mathbb{I}_{\{Y_j \leq y\}}}{\sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h_n}\right)} \quad (10)$$

und

$$q_{n,\alpha}^{(-i)}(X_i) : \inf \left\{ y \in \mathbb{R} \mid F_{n,X_i}^{(-i)}(y) \geq \alpha \right\}, \quad 0 < \alpha < 1. \quad (11)$$

Kreuzvalidierung bedeutet, daß der leave-one-out Schätzer (11) in (9) eingesetzt wird und die Bandweite ausgewählt wird, die diesen Ausdruck minimiert. Es ist also

$$h_{KV} = \min_h \left\{ \sum_{i=1}^n \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i)) \right\}. \quad (12)$$

Um einen tieferen Einblick in die Arbeitsweise dieses Verfahrens zu erhalten, wird der Erwartungswert der Summe (9) untersucht. Es gilt

$$E \left[ \sum_i \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i)) \right] = \sum_i E \left[ \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i)) \right] \quad (13)$$

und für unabhängige  $\{Y_i\}$  ist

$$E \left[ \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i)) \right] = \alpha \mu_Y(X_i) - \alpha E \left[ q_{n,\alpha}^{(-i)}(X_i) \right] + \int_{-\infty}^{E \left[ q_{n,\alpha}^{(-i)}(X_i) \right]} F(y \mid X_i) dy$$

$$\begin{aligned}
&= \alpha \mu_Y(X_i) - \alpha q_\alpha(X_i) + \alpha q_\alpha(X_i) - \alpha E [q_{n,\alpha}^{(-i)}(X_i)] \\
&\quad + \int_{-\infty}^{q_\alpha(X_i)} F(y | X_i) dy + \int_{q_\alpha(X_i)}^{E[q_{n,\alpha}^{(-i)}(X_i)]} F(y | X_i) dy \\
&= E [\varphi(Y_i - q_\alpha(X_i))] + \alpha [q_\alpha(X_i) - E [q_{n,\alpha}^{(-i)}(X_i)]] \\
&\quad + \int_{q_\alpha(X_i)}^{E[q_{n,\alpha}^{(-i)}(X_i)]} F(y | X_i) dy \\
&= E [\varphi(Y_i - q_\alpha(X_i))] + c_i | q_\alpha(X_i) - E [q_{n,\alpha}^{(-i)}(X_i)] | . \quad (14)
\end{aligned}$$

Die Werte  $c_i$  ergeben sich aus dem Mittelwertsatz der Integralrechnung. Es folgt für den Erwartungswert der Summe

$$\begin{aligned}
E \left[ \sum_{i=1}^n \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i)) \right] &= E \left[ \sum_{i=1}^n \varphi(Y_i - q_\alpha(X_i)) \right] \\
&\quad + \sum_{i=1}^n c_i | q_\alpha(X_i) - E [q_{n,\alpha}^{(-i)}(X_i)] | . \quad (15)
\end{aligned}$$

Es tritt ein Bias in Form der gewichteten Summe  $\sum_i c_i | q_\alpha(X_i) - E [q_{n,\alpha}^{(-i)}(X_i)] |$  auf. Die Gewichtung hängt von der Steilheit der bedingten Verteilungsfunktion  $F(y | X_i)$  zwischen den Stellen  $q_\alpha(X_i)$  und  $E [q_{n,\alpha}^{(-i)}(X_i)]$  ab. Je größer die absolute Differenz  $| \alpha - F(q_{n,\alpha}^{(-i)}(X_i) | X_i) |$  ist, desto stärker ist die Gewichtung.

Die gewählte Bandweite  $h_{KV}$  soll die auftretende Verzerrung minimieren. Diese Bandweite wird dann in der Schätzung  $q_{n,\alpha}(X_i)$  verwendet. Daher wird der Bias in (15) weiter untersucht. Es ist

$$\begin{aligned}
\sum_i c_i | q_\alpha(X_i) - E [q_{n,\alpha}^{(-i)}(X_i)] | &= \sum_i c_i | q_\alpha(X_i) - E [q_{n,\alpha}(X_i)] \\
&\quad + E [q_{n,\alpha}(X_i)] - E [q_{n,\alpha}^{(-i)}(X_i)] | .
\end{aligned}$$

Es seien  $F_{n,X_i}(q_{n,\alpha}(X_i) | X_i) = F_{n,X_i}^{(-i)}(q_{n,\alpha}^{(-i)}(X_i) | X_i) = F_{X_i}(q_\alpha(X_i) | X_i) = \alpha$  eindeutig und es seien die Bedingungen (K.1) - (K.4), (S.1) und (S.2) erfüllt. Dann gilt mit  $x \in$

$(h_n, 1 - h_n)$  für die Erwartungswerte der Schätzer:

$$\begin{aligned}
F_{n,x}(q_{n,\alpha}(x)) &= \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \mathbb{I}_{\{Y_i \leq q_{n,\alpha}(x)\}}}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} = \alpha \\
0 &= \frac{1}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) [\alpha - \mathbb{I}_{\{Y_i \leq q_{n,\alpha}(x)\}}] ; \\
E[0] &= \frac{1}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) [\alpha - F(q_{n,\alpha}(x) | X_i)] = 0. \quad (16)
\end{aligned}$$

Taylorreihenentwicklung der bedingten Verteilungsfunktion um den Punkt  $(x, q_\alpha(x))$  und die Substitution  $U_i = \frac{x-X_i}{h_n}$  ergeben

$$\begin{aligned}
F(q_{n,\alpha}(x) | X_i) &\approx \alpha + f(q_\alpha(x) | x)(q_{n,\alpha}(x) - q_\alpha(x)) - h_n U_i F^{(10)}(q_\alpha(x) | x) \\
&\quad + \frac{1}{2} h_n^2 U_i^2 F^{(20)}(q_\alpha(x) | x).
\end{aligned}$$

Durch Einsetzen und Umstellen resultiert aus (16)

$$E[q_{n,\alpha}(x)] \approx q_\alpha(x) + \frac{1}{2} h_n^2 \left[ \frac{F^{(20)}(q_\alpha(x) | x)}{f(q_\alpha(x) | x)} \right] \cdot \frac{\sum_i U_i^2 K(U_i)}{\sum_i K(U_i)}. \quad (17)$$

Mit der gleichen Vorgehensweise erhält man

$$E[q_{n,\alpha}^{(-i)}(x)] \approx q_\alpha(x) + \frac{1}{2} h_n^2 \left[ \frac{F^{(20)}(q_\alpha(x) | x)}{f(q_\alpha(x) | x)} \right] \cdot \frac{\sum_i U_i^2 K(U_i)}{\sum_i K(U_i) - K(0)}. \quad (18)$$

Die Differenz der beiden Erwartungswerte ist

$$E[q_{n,\alpha}(x)] - E[q_{n,\alpha}^{(-i)}(x)] \approx \frac{1}{2} h_n^2 \left[ \frac{F^{(20)}(q_\alpha(x) | x)}{f(q_\alpha(x) | x)} \right] \left[ \frac{K(0) \cdot \sum_i U_i^2 K(U_i)}{K(0) \sum_i K(U_i) - \left[ \sum_i K(U_i) \right]^2} \right]. \quad (19)$$

Mit der Integralapproximation aus dem Anhang gilt

$$E[q_{n,\alpha}(x)] - E[q_{n,\alpha}^{(-i)}(x)] \approx \frac{1}{2} h_n^2 \left[ \frac{F^{(20)}(q_\alpha(x) | x)}{f(q_\alpha(x) | x)} \right] \left[ \frac{K(0) \int u^2 K(u) du}{K(0) - n h_n} \right]$$

$$\begin{aligned}
&\approx \frac{1}{2}h_n^2 \left[ \frac{F^{(20)}(q_\alpha(x) | x)}{f(q_\alpha(x) | x)} \right] \left[ \frac{K(0) \int u^2 K(u) du}{nh_n \left[ \frac{K(0)}{nh_n} - 1 \right]} \right] \\
&= o(h_n^2)
\end{aligned} \tag{20}$$

Der leave-one out Schätzer konvergiert gegen den gewöhnlichen Kernschätzer. Diese Konvergenz ist schneller, als die Konvergenz von  $E[q_{n,\alpha}(X_i)]$  gegen  $q_\alpha(X_i)$ , da letztere, wie aus Gleichung (17) ersichtlich ist, von der Ordnung  $O(h_n^2)$  ist.

Da die Berechnungen nur für die Werte  $x = X_i$  gelten, die nicht im Randbereich liegen, ist es sinnvoll, bei Anwendung der Kreuzvalidierung eine Gewichtsfunktion  $\omega(X_i)$  in Form einer Indikatorfunktion einzuführen, die den Werten am Rand ein Gewicht von Null gibt. Die Summe (9) wird modifiziert zu

$$\sum_i \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i)) \cdot \omega(X_i). \tag{21}$$

Die obigen Betrachtungen geben einen Einblick in die Wirkungsweise der Bandweitenwahl durch Kreuzvalidierung für die nichtparametrische Quantilsregression. Es wurde nicht gezeigt, ob die diskutierte Art der Bandweitenwahl Optimalitätskriterien genügt. Alternative Verfahren zur Bandweitenwahl für die Quantilsregression sind uns bisher nicht bekannt.

### 3 Simulation

Die Schätzung bedingter Quantile mit automatisierter Bandweitenwahl durch Kreuzvalidierung wird nun in verschiedenen Simulationen auf ihre Tauglichkeit hin überprüft. Im folgenden sei  $X_i = i$ ,  $\{Y_i\}$  stets unabhängig und der Stichprobenumfang jeweils  $n = 600$ . Die  $\{Y_i\}$  sind jeweils aus der gleichen Verteilungsfamilie, die Streuung der Verteilung hängt von der Stelle  $i$  ab. Die Abbildungen 1, 2 und 3 zeigen drei Beispiele. Abbildung 1 enthält eine Realisierung bei Ziehung aus Normalverteilungen. Es sind weiter die tatsächlichen 0,25- und 0,75-Quantile und die geschätzten Quantile dargestellt. Abbildung 2 zeigt ein

Beispiel für Cauchy-Verteilungen. Dieser Fall ist besonders interessant, da die Cauchy-Verteilung so breite Schwänze besitzt, daß weder Erwartungswert noch Varianz existieren. Man erkennt, daß der extreme Wert von über 3000 nicht auf die Schätzung durchschlägt. Die Schätzung ist robust. In Abbildung 3 stammen die Daten aus Lognormalverteilungen. Die Schiefe dieser Verteilung spiegelt sich auch in der Schätzung wider. Diese Beispiele zeigen, daß die verwendete Schätzmethodik in der Lage ist, die den Daten zugrundeliegende Struktur zu erkennen. Aussagekräftiger als die Auswertung von Schätzungen für jeweils einmalige Ziehungen von 600 Werten ist die Betrachtung von Simulationen mit einer größeren Anzahl von Wiederholungen. Die Ergebnisse von drei solchen Simulationen sind in den Abbildungen 4, 5 und 6 dargestellt. Es wurden 100 mal 600 Daten gezogen und 100 mal die Quantilsfunktionen geschätzt. Um die Ergebnisse interpretierbar darzustellen sind in den Graphiken die tatsächlichen Quantile sowie für verschiedene Punkte auf der Abszisse die Box-Plots der an diesen Stellen errechneten Schätzungen eingetragen. D.h. die zu bestimmten Werten  $i$  vorhandenen 100 Schätzwerte für  $q_\alpha(i)$  werden durch einen Box-Plot repräsentiert. In Abbildung 4 ist die zugrundeliegende Verteilung eine t-Verteilung mit 6 Freiheitsgraden. Die verwendete Schätzprozedur ist in der Lage, die Struktur in den Daten zu erkennen, wengleich ein leichter, für die Kernglättung üblicher Bias auftritt. Ebenfalls aus t-Verteilungen mit 6 Freiheitsgraden stammen die Beobachtungen in Abbildung 5, doch sind die vorgegebenen Verläufe der bedingten Quantile hier "welliger". Die Kreuzvalidierung reagiert auf diesen Umstand und erkennt die höhere Frequenz der zugrundeliegenden Funktionen. Abbildung 6 zeigt die Simulationsergebnisse für eine schiefe Verteilung, die Lognormalverteilung. Das Schätzverfahren erkennt die Schiefe der Verteilung. Unproblematisch ist die Schätzung der 0,25-Quantile. Die geschätzten Werte liegen alle sehr eng um die tatsächliche Funktion. Diffuser ist das Bild für die 0,75-Quantile. In den Spitzen der Funktion sind die Schätzungen verzerrt und weisen eine hohe Variabilität auf. Dagegen sind in den Tälern der Funktion Verzerrung und Variabilität der Schätzung gering. Hier spielt die in (15) aufgezeigte Gewichtung der absoluten Abstände zwischen tatsächlichen Quantilen und den Erwartungswerten der geschätzten Quantile eine Rolle.

Abbildung 1: Tatsächliche und geschätzte bedingte Quantile für normalverteilte Daten ( $n = 600$ )

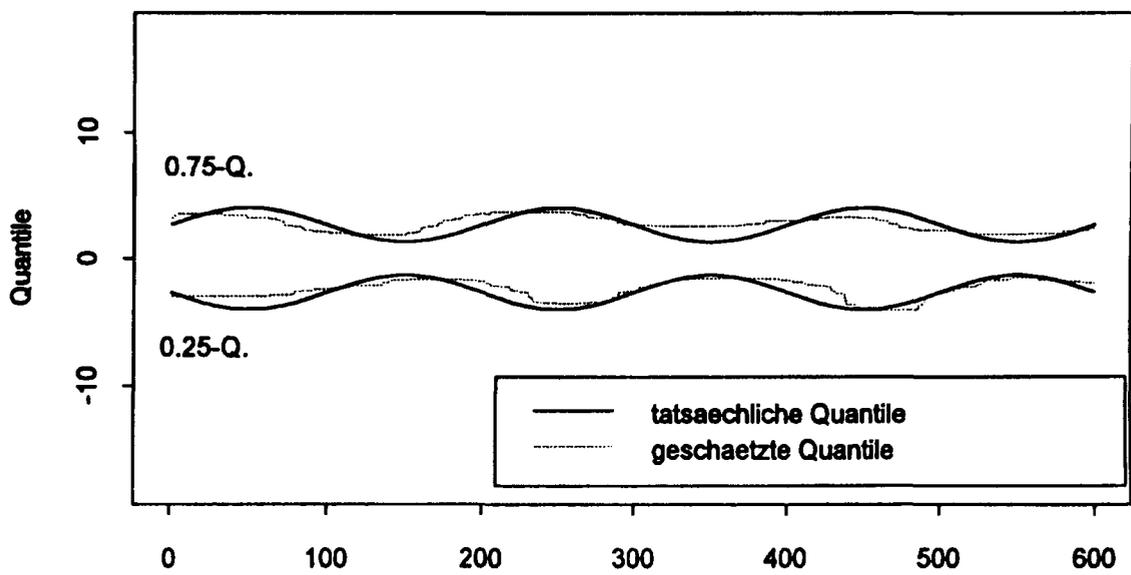
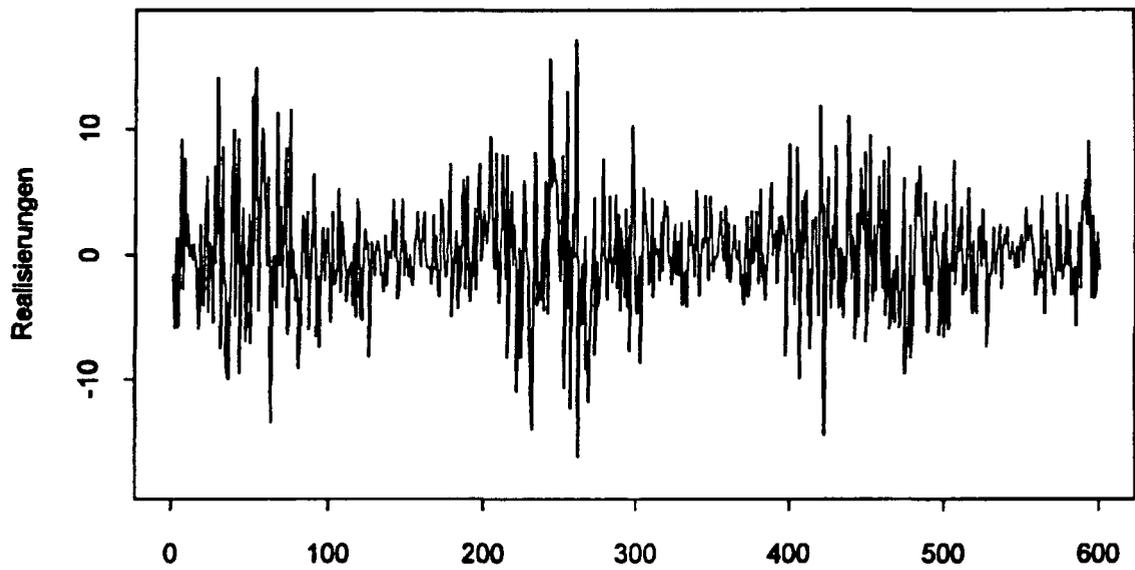


Abbildung 2: Tatsächliche und geschätzte bedingte Quantile für Cauchy-verteilte Daten ( $n = 600$ )

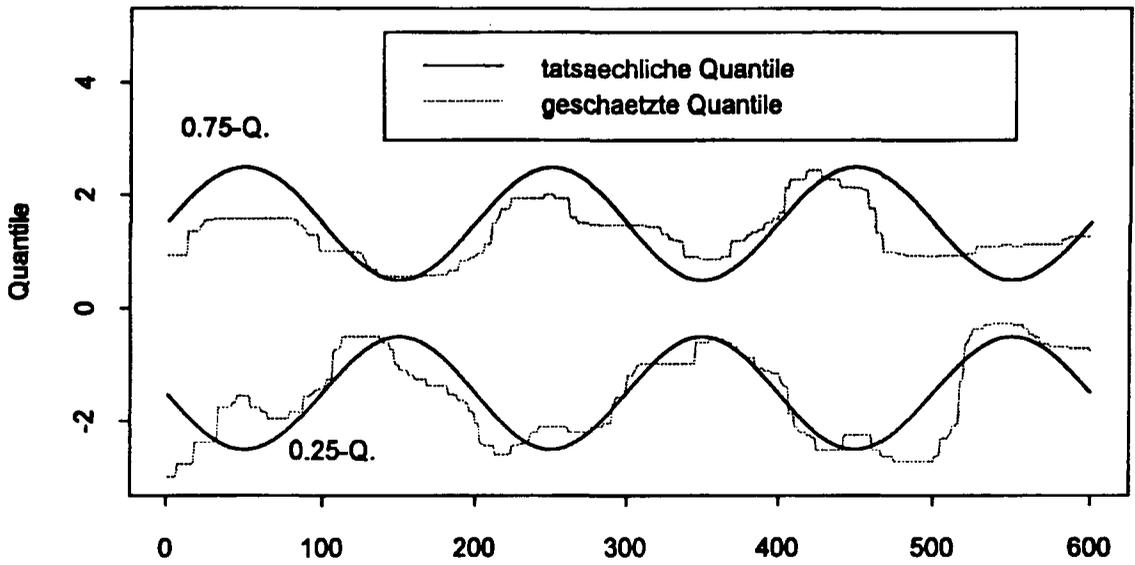
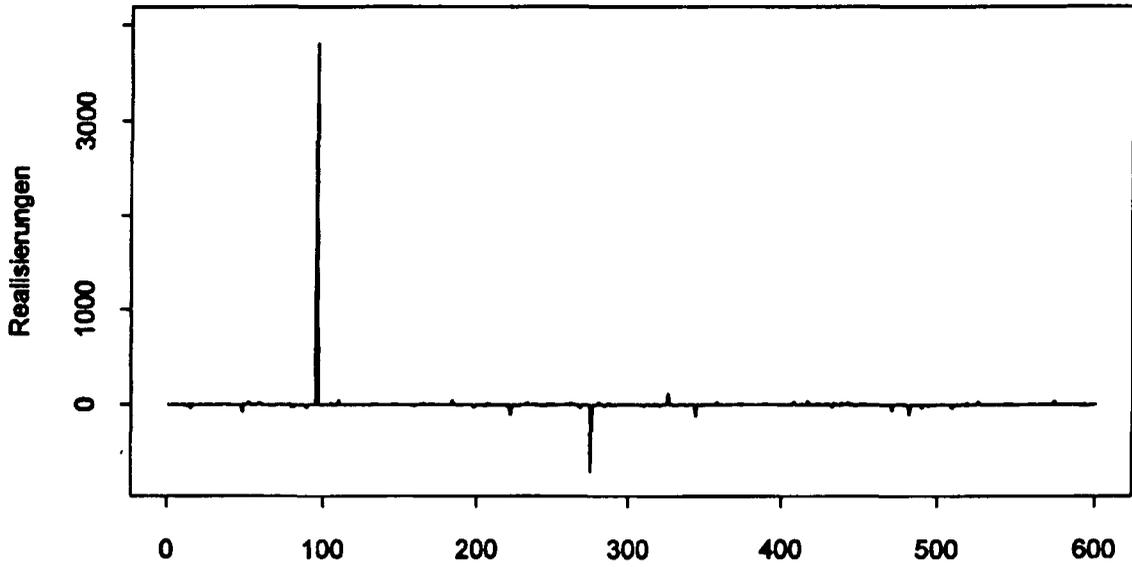


Abbildung 3: Tatsächliche und geschätzte bedingte Quantile für lognormalverteilte Daten  
( $n = 600$ )

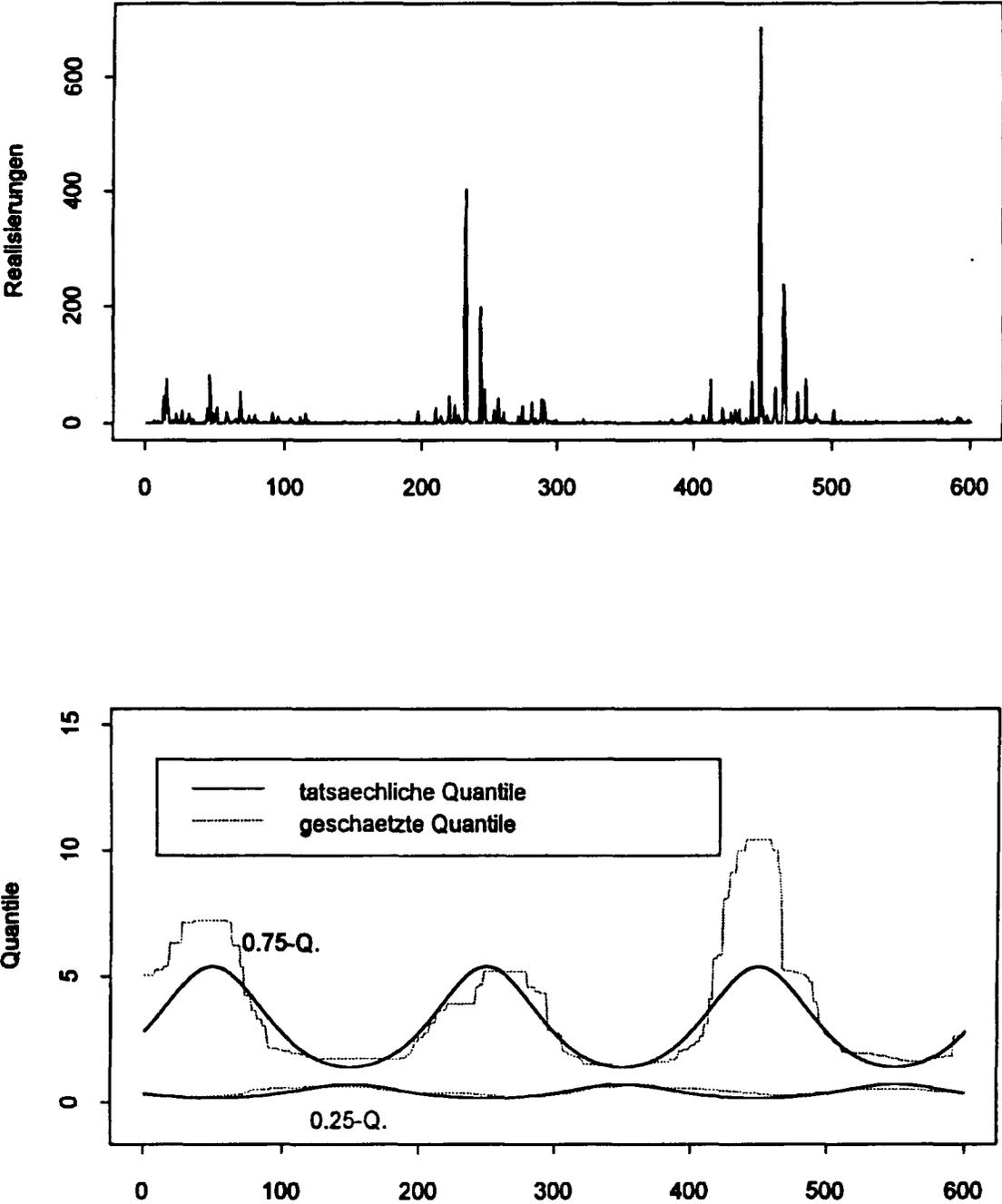
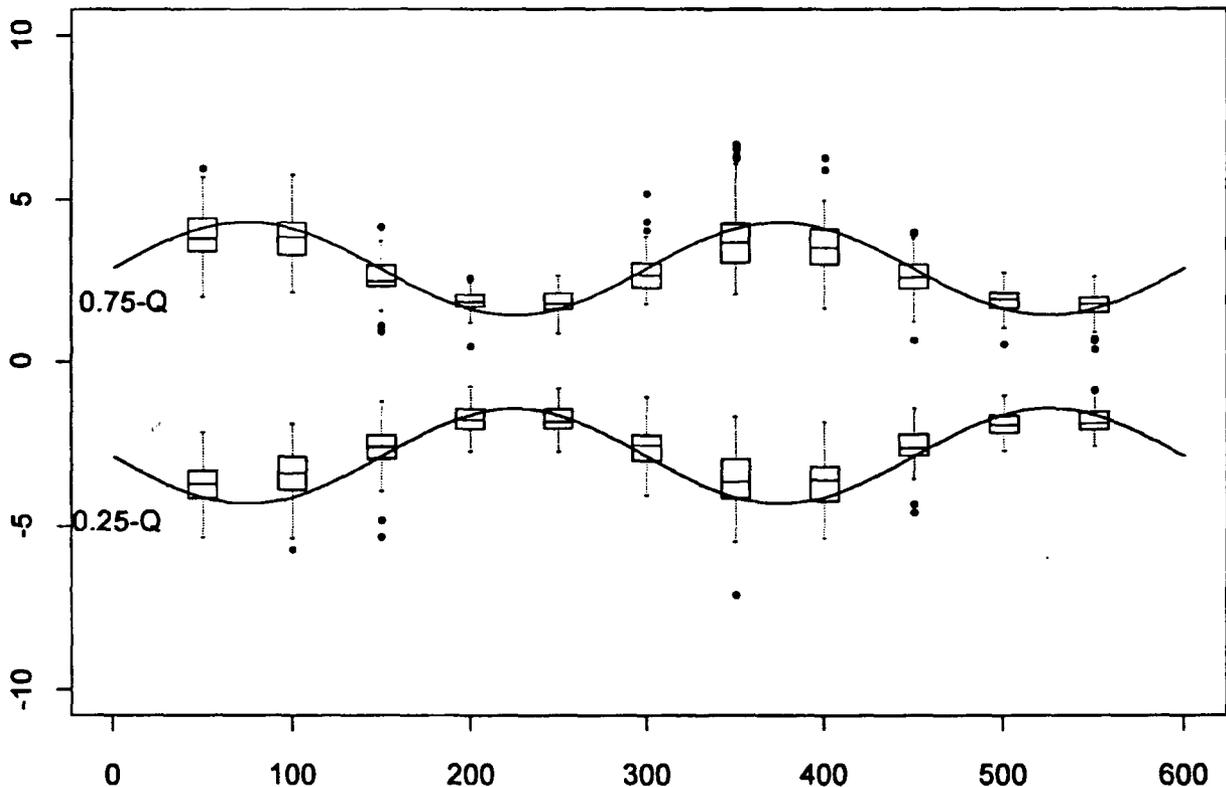
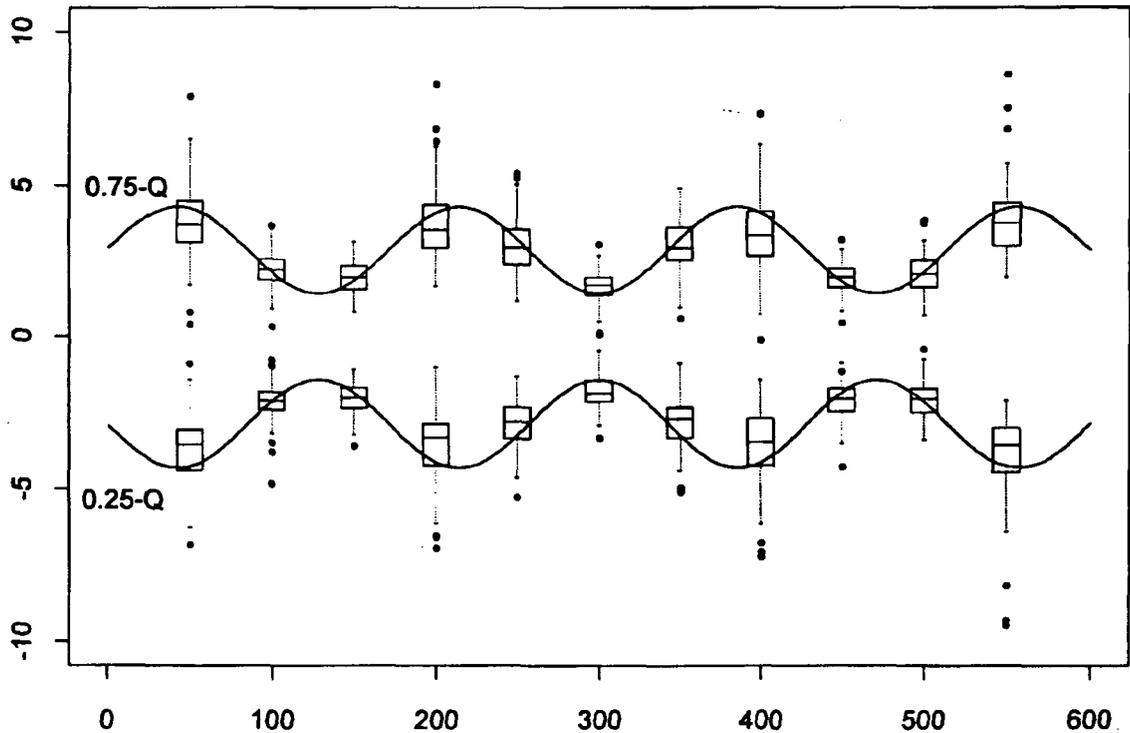


Abbildung 4: Tatsächliche bedingte Quantile und Simulationsergebnisse für t-verteilte Daten mit 6 Freiheitsgraden ( $n = 600$ , Anzahl Wiederholungen = 100)



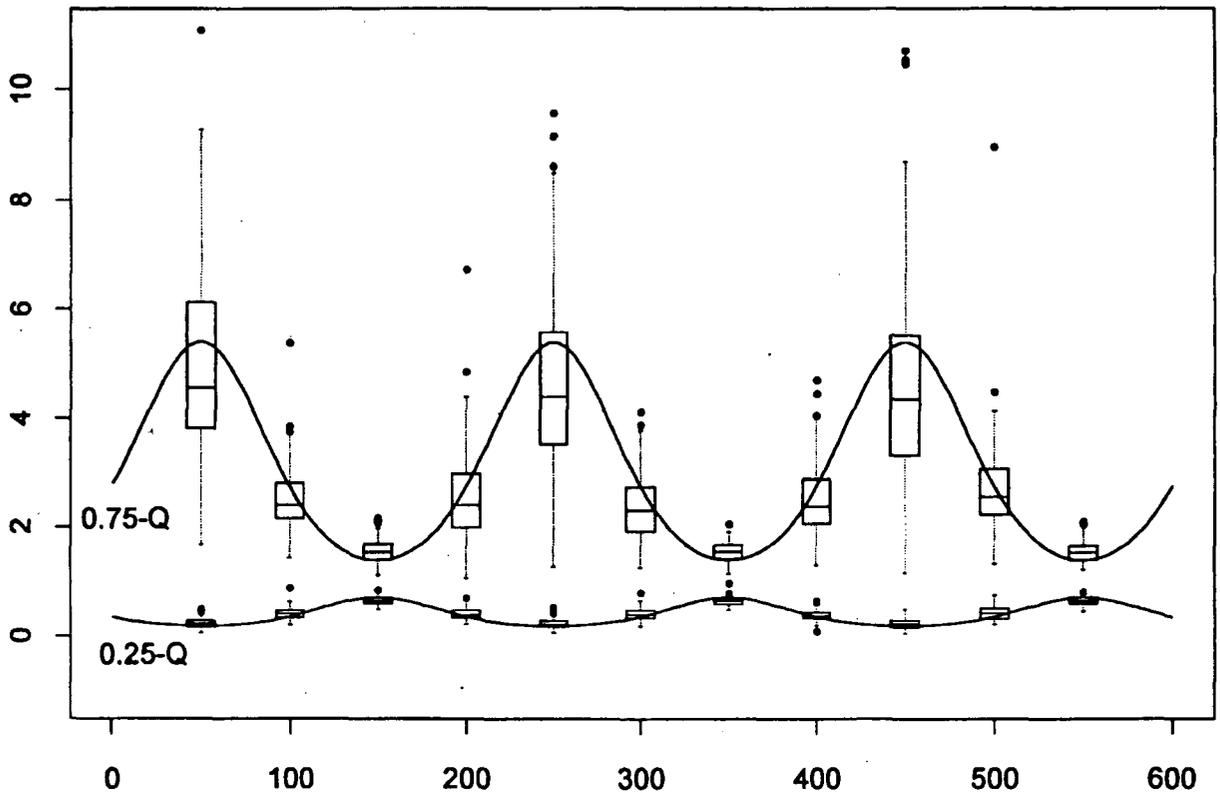
Eine absolute Abweichung wird um so stärker gewichtet, je steiler die bedingte Verteilungsfunktion in diesem Bereich ist. Für die Lognormalverteilung variiert diese Steilheit in der Umgebung der dargestellten 0,75-Quantile sehr stark. In den Minimas der Funktion ist die bedingte Verteilung recht steil. In den Spitzen dagegen ist die bedingte Verteilung relativ flach. Man erkennt diese Unterschiede auch in den Daten aus Abbildung 3. Die Schätzprozedur nimmt durch die Gewichtung Fehler bei der Schätzung in den Spitzen zugunsten einer guten Anpassung in den Minimas in Kauf. Die Steilheit der bedingten Verteilungsfunktion beeinflusst nicht nur die Kreuzvalidierung, sondern auch den Erwartungswert des Kernschätzers. In Gleichung (17) wird der approximative Bias des Kernschätzers für bedingte Quantile angegeben. Die Verzerrung hängt durch  $F^{(20)}(q_\alpha(x) | x)$  von der Glattheit der Quantilsfunktion ab. Wegen der Division durch

Abbildung 5: Tatsächliche bedingte Quantile und Simulationsergebnisse für t-verteilte Daten mit 6 Freiheitsgraden ( $n = 600$ , Anzahl Wiederholungen = 100)



die bedingte Dichte spielt aber auch die Steilheit der bedingten Verteilungsfunktion an der Stelle  $x$  eine Rolle. Je steiler die bedingte Verteilungsfunktion in  $q_\alpha(x)$  ist, desto geringer ist der Bias der Schätzung  $q_{n,\alpha}(x)$ . Je flacher die bedingte Verteilungsfunktion, desto größer der Bias. Die Kreuzvalidierung zur Bandweitenbestimmung gewichtet diese Fehler in Bereichen flacher Verteilungsfunktionen geringer als die Fehler in Umgebungen mit steilen Verteilungsfunktionen.

Abbildung 6: Tatsächliche bedingte Quantile und Simulationsergebnisse für lognormalverteilte Daten mit 6 Freiheitsgraden ( $n = 600$ , Anzahl Wiederholungen = 100)



## 4 Schlußbetrachtung

In diesem Aufsatz wird die Kreuzvalidierung als automatisches Verfahren der Bandweitenwahl in der nichtparametrischen Quantilsregression diskutiert. Die Kreuzvalidierung wird über die Verlustfunktion  $\sum_i \varphi(Y_i - q_{n,\alpha}^{(-i)}(X_i))\omega(X_i)$  mit  $\varphi(u) = \alpha \mathbb{I}_{\{u \geq 0\}} \cdot u + (\alpha - 1) \mathbb{I}_{\{u < 0\}} \cdot u$  definiert. Einige Eigenschaften dieser Vorgehensweise werden diskutiert. Die Erwartungswertbildung für diese Summe zeigt, daß ein Bias in Form der gewichteten Summe der absoluten Abstände zwischen tatsächlichen Quantilen und den Erwartungswerten der leave-one-out Schätzungen der Quantile auftritt. Die Gewichtung hängt von der Steilheit der bedingten Verteilungsfunktion zwischen diesen beiden Werten ab. Simulationen für verschiedene Verteilungsfamilien zeigen, daß die Schätzprozedur in der Lage ist, die

den Daten zugrundeliegende Struktur zu erkennen. Die Beispiele beinhalten Ziehungen aus Normalverteilungen und aus t-Verteilungen, welche breitere Schwänze besitzen als Normalverteilungen. Sowie Ziehungen aus Cauchy-Verteilungen, für die erste und höhere Momente nicht existieren, und Lognormalverteilungen, die schief sind. Insbesondere die Betrachtung der Simulationsergebnisse für lognormalverteilte Daten zeigt die Rolle der Steilheit der bedingten Verteilungsfunktion. Es resultiert eine relativ gute Anpassung in Bereichen, in denen die bedingte Verteilung steil ist. Schlechter ist die Schätzung in Zonen, in denen die bedingte Verteilung flach ist. In den anderen Beispielen, in denen die Veränderung der Steilheit nicht so dramatisch ist wie bei der Lognormalverteilung, spielt dieser Effekt nur eine untergeordnete Rolle.

## 5 Anhang

**Lemma 1: Integralapproximation der Summe über die Kernfunktion.**

Mit (K.1), (K.2) und (S.2) folgt:

$$\begin{aligned}
 & \left| \frac{1}{nh_n} \sum_{i=1}^n K(U_i) - \int K(u) du \right| \\
 & \leq \frac{1}{nh_n} \sum_{i \in J} \left| K(U_i) - nh_n \int_{u_i}^{u_{i-1}} K(z) dz \right| \\
 & = \frac{1}{nh_n} \sum_{i \in J} \left| K(U_i) - nh_n (U_{i-1} - U_i) K(\xi_i) \right|, \quad U_i \leq \xi_i < U_{i-1}, \\
 & \leq \frac{1}{nh_n} \sum_{i \in J} |K(U_i) - K(\xi_i)| \\
 & \leq \frac{1}{nh_n} \sum_{i \in J} L |U_i - \xi_i| \leq \frac{1}{nh_n} \sum_{i \in J} O\left(\frac{1}{nh_n}\right) = O\left(\frac{1}{nh_n}\right).
 \end{aligned}$$

Wobei  $J$  die Indexmenge der Beobachtungen mit einem Gewicht größer Null ist. Es ist  $|J| = O\left(\frac{1}{nh_n}\right)$ , mit  $|J|$  gleich der Mächtigkeit der Menge  $J$ . Auf die gleiche Weise lassen sich die Ausdrücke  $\sum_i K^2(U_i)$ ,  $\sum_i U_i K(U_i)$  und  $\sum_i U_i^2 K^2(U_i)$  durch Integrale approximieren.

## 6 Literatur

- Abberger K. (1994):** Nichtparametrische Schätzung bedingter Quantile in Finanzmarktdaten, Diskussionsbeitrag Nr. 225 Ser. II, Sonderforschungsbereich 178, Universität Konstanz.
- Härdle W. (1989):** Applied nonparametric regression, Cambridge University Press, Cambridge.
- Heiler S., Michels P. (1994):** Deskriptive und Explorative Datenanalyse, Oldenbourg Verlag, München.
- Jones M.C., Hall P. (1990):** Mean Squared Error Properties of Kernel Estimates of Regression Quantiles, *Statistics & Probability Letters* 10, 283-289.
- Koenker R., Basset G. (1978):** Regression Quantiles, *Econometrica* 46, 33-50.
- Koenker R., Portnoy S., Ng P. (1992):** Nonparametric Estimation of Conditional Quantile Functions, in: *L<sub>1</sub>-Statistical Analysis and Related Methods* (Ed.: Y. Dodge), North-Holland, New York.