

SFB 649 Discussion Paper 2010-041

Prognose mit nichtparametrischen Verfahren

Wolfgang Karl Härdle*
Rainer Schulz**
Weining Wang***



*Humboldt-Universität zu Berlin, Germany & National Central
University, Taipei, Taiwan

**University of Aberdeen Business School, United Kingdom

***Humboldt- Universität zu Berlin, Germany

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Prognose mit nichtparametrischen Verfahren

Wolfgang Karl Härdle, Rainer Schulz und Weining Wang*

Keywords: time series, semiparametric model, k-NN estimation, local polynomial regression, volatility forecasting,

AMS 2000 subject classification: 62G08, 62G20, 62M10

JEL classification: C14, C32, G12

*Härdle: Professor at Humboldt-Universität zu Berlin and Director of C.A.S.E. - Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Berlin, Germany & Department of Finance, National Central University, Taipei, Taiwan, R.O.C. Schulz: University of Aberdeen Business School, Aberdeen, United Kingdom. Wang: Ladislaus von Bortkiewicz Chair of Statistics, School of Business and Economics Humboldt-Universität zu Berlin, Berlin, Germany. Für finanzielle Unterstützung bedanken wir uns bei der Deutschen Forschungsgemeinschaft, SFB 649 „Ökonomisches Risiko“.

9 Prognose mit nichtparametrischen Verfahren


Wolfgang Karl Härdle, Rainer Schulz und Weining Wang

9.1 Einleitung

Statistische Prognosen basieren auf der Annahme, dass ein funktionaler Zusammenhang zwischen der zu prognostizierenden Variable y und anderen J -dimensional beobachtbaren Variablen $x = (x_1, \dots, x_J) \in \mathbb{R}^J$ besteht. Kann der funktionale Zusammenhang geschätzt werden, so kann im Prinzip für jedes x der zugehörige y Wert prognostiziert werden.

Bei den meisten Anwendungen wird angenommen, dass der funktionale Zusammenhang einem niedrigdimensionalen parametrischen Modell entspricht oder durch dieses zumindest gut wiedergegeben wird. Ein Beispiel im univariaten Fall ist das lineare Modell $y = \beta_0 + \beta_1 x$. Sind die beiden unbekannt Parameter β_0 und β_1 mithilfe historischer Daten geschätzt, so lässt sich für jedes gegebene x sofort der zugehörige y Wert prognostizieren. Allerdings besteht hierbei die Gefahr, dass der wirkliche funktionale Zusammenhang nicht dem gewählten Modell entspricht. Dies kann infolge zu schlechten Prognosen führen.

Nichtparametrische Verfahren gehen ebenfalls von einem funktionalen Zusammenhang aus, geben aber kein festes parametrisches Modell vor und zwingen die Daten damit in kein Prokrustes Bett. Sie sind deshalb hervorragend geeignet, um 1) Daten explorativ darzustellen, 2) parametrische Modelle zu überprüfen und 3) selbst als Schätzer für den funktionalen Zusammenhang zu dienen (Cleveland [2], Cleveland und Devlin [3]). Nichtparametrische Verfahren können daher problemlos auch zur Prognose eingesetzt werden.

Dieses Kapitel ist wie folgt strukturiert. Abschnitt 9.2 stellt nichtparametrische Verfahren vor und erläutert deren grundsätzliche Struktur. Der Schwerpunkt liegt auf dem univariaten Regressionsmodell und auf der Motivation der vorgestellten Verfahren. Abschnitt 9.3 präsentiert eine praktische Anwendung für eine Zeitreihe von Wechselkursvolatilitäten. Es werden Prognosen mit nichtparametrischen Verfahren berechnet und deren Güte mit den Prognosen eines AR(1)-Zeitreihenmodells verglichen, vgl. auch Kapitel 14 dieses Buches. Es zeigt sich für die gewählte Anwendung, dass das parametrische Modell die Daten sehr gut erfasst. Das nichtparametrische Modell liefert in dieser Anwendung keine bessere Prognosegüte. Zugleich veranschaulicht die Anwendung, wie nichtparametrische Verfahren für die Modelvalidierung eingesetzt werden können. Und natürlich zeigt es auch, wie solche Verfahren für Prognosen eingesetzt werden können. Abschnitt 9.4 präsentiert die Literatur, die für weitere Lektüre herangezogen werden kann. Alle praktischen Beispiele im Text, welche mit dem Symbol  versehen sind, lassen sich von der Adresse www.quantlet.de herunterladen.

9.2 Nichtparametrische Verfahren

9.2.1 Einführung

Ein nichtparametrisches Modell geht von dem allgemeinen funktionalen Zusammenhang

$$y = m(x) \tag{1}$$

aus, wobei $m(\cdot)$ eine glatte Funktion in den erklärenden Variablen x ist. Selbstverständlich lässt sich auch das lineare Regressionsmodell mit $m(x) = x^T \beta$ in obiger Form darstellen. Nichtparametrische Verfahren lassen jedoch die glatte Funktion $m(\cdot)$ für die Schätzung unspezifiziert. Mit dem Störterm ε wird (1) in das statistische Modell

$$y = m(x) + \varepsilon \tag{2}$$

überführt. Der bedingte Erwartungswert des Störterms in (2) ist null und es folgt

$$E[y|x] = m(x).$$

Es ist das Ziel nichtparametrischer Verfahren, die bedingte Erwartungswertfunktion $m(x)$ zu schätzen. Dies geschieht durch lokale Mittelwertbildung über die $i = 1, \dots, n$ Beobachtungen y_i mit

$$\hat{m}(x) = \sum_{i=1}^n G_i(x) y_i. \tag{3}$$

Der Wert des Gewichtes $G_i(x)$ hängt davon ab, wie nahe die zu y_i gehörenden Ausprägungen der erklärenden Variablen x_i an x liegen.

Der *k-nearest neighbor* (kNN) Schätzer mit gleichen Gewichten ist ein anschauliches Beispiel für (3). Hierbei werden die k nächsten Nachbarn zur Berechnung des Durchschnitts verwendet, wobei $n \geq k \geq 1$. In diesem Fall gilt für die Gewichte in (3)

$$G_i(x) = \frac{I\{i \in M(x,k)\}}{k},$$

wobei

$$M(x, k) = \{i: x_i \text{ ist eine der } k \text{ nächsten Beobachtungen von } x\}$$

die Indexmenge der k Beobachtungen ist, die am nächsten an x liegen. $I(\cdot)$ ist die Indikatorfunktion, welche den Wert 1 annimmt, falls i ein Element von $M(x, k)$ ist, sodass $G_i(x) = 1/k$. Entsprechend nimmt die Indikatorfunktion den Wert 0 an, falls i kein Element von $M(x, k)$ ist, sodass $G_i(x) = 0$. Die Nähe zu x kann etwa mit der euklidischen Distanz $\sqrt{(x_i - x)^T (x_i - x)}$ berechnet werden. Wählen wir zum Beispiel $k = 12$, so ist

$$\hat{m}(x) = \frac{1}{12} \sum_{i \in M(x,12)} y_i \tag{4}$$

und der kNN-Schätzer entspricht dem arithmetischen Mittel aus den 12 Beobachtungen der abhängigen Variablen, deren zugehörigen erklärenden Variablen am nächsten an x liegen. Weiterhin ist offensichtlich, dass $\hat{m}(x)$ dem arithmetischen

Mittel aller Beobachtungen der abhängigen Variablen entspricht, sobald $k = n$ gewählt wird.

Abbildung 1 zeigt die geschätzte univariate Funktion $\hat{m}(x)$ für einen Beispieldatensatz mit einer erklärenden Variable, $(y, x) \in \mathbb{R}^2$. Bei den Daten handelt es sich um den Motorcycle Datensatz mit 133 Beobachtungen aus Härdle [6] (Table 2, Appendix 2). Man erkennt, dass $\hat{m}(x)$ eine Treppenfunktion ist, die für alle x konstant ist, welche über die gleiche Indexmenge verfügen. Der frei zu wählende Parameter k bestimmt die Glattheit der geschätzten Kurve. Variiert man k , so entsteht eine Familie von Regressionskurven, die alle durch lokale Mittelung entstanden sind.

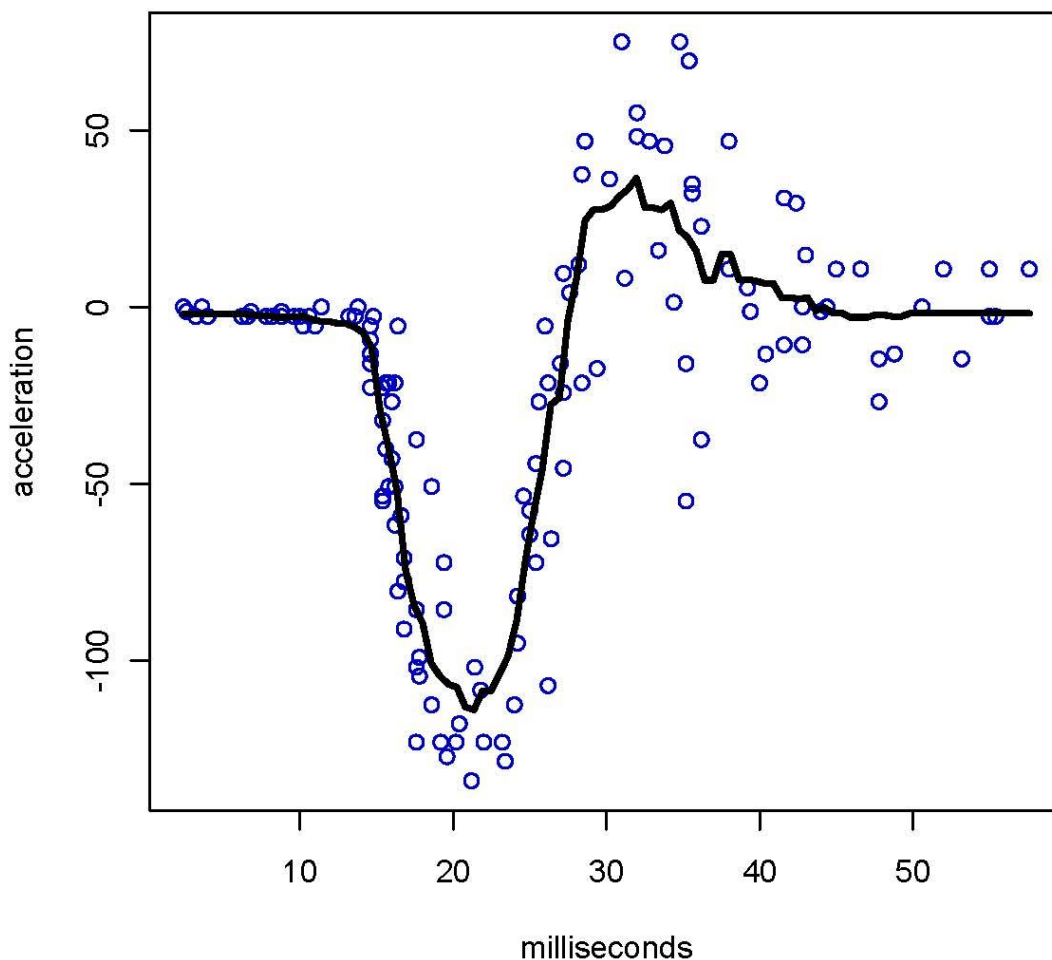



Abb. 1: Nichtparametrische Regression mit kNN Schätzer mit $k = 12$, uniformer Gewichtung und Mittelwertbildung für alle einbezogenen Beobachtungen. Quadrate geben die Beobachtungen an. 

Statt das arithmetische Mittel aller Beobachtungen der Indexmenge $M(x, k)$ zu nehmen, kann auch eine lokale lineare Regression für die Beobachtungen in $M(x, k)$ angepasst werden, sodass gilt

$$\hat{m}(x) = \hat{\beta}_0(x) + \hat{\beta}_1(x)x.$$

Abbildung 2 zeigt die geschätzte Funktion, die aufgrund der uniformen Gewichtung (4) unstetig ist.

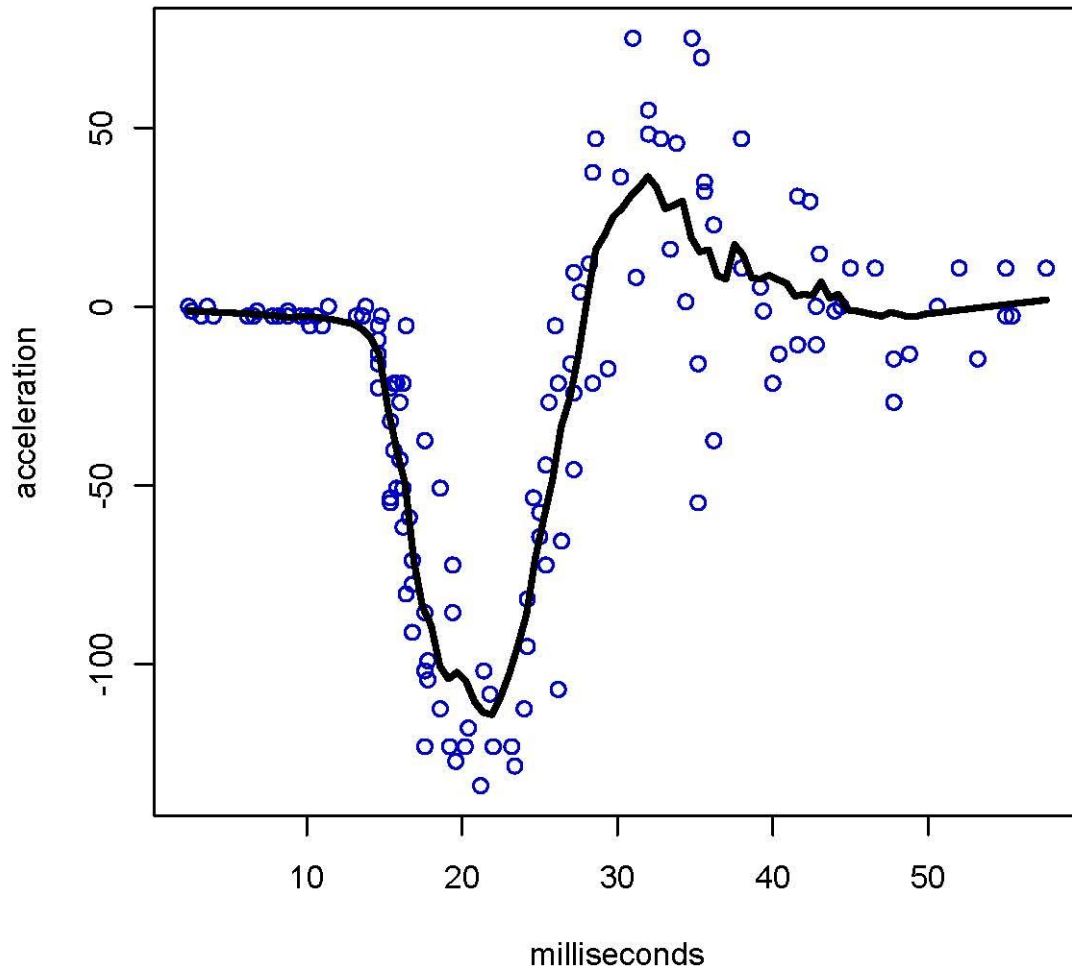



Abb. 2: Nichtparametrische Regression mit kNN Schätzer mit $k = 12$, uniformer Gewichtung und linearer Regression für alle einbezogenen Beobachtungen. Quadrate geben die Beobachtungen an. 

9.2.2 Lokal gewichtete lineare Regression

Um eine glatte Funktion $\hat{m}(x)$ zu erhalten, dürfen nicht alle Beobachtungen für die Schätzung gleich gewichtet werden, sondern Beobachtungen nahe x müssen stark und entfernte Beobachtungen schwach gewichtet werden. Eine stetige Gewichtungsfunktion reduziert den anfänglichen Einfluss von Beobachtungen, die neu in die Indexmenge kommen oder diese verlassen. Zwei stetige Gewichtungsfunktionen werden im Folgenden genauer vorgestellt.

Zugleich wird mit der lokal gewichteten Regression (LWR) ein genereller Ansatz für die nichtparametrische Regression vorgestellt, der die Beispiele aus der Einleitung als Sonderfälle umfasst. Diese Generalisierung erfordert zugleich, dass wir die Notation erweitern und sogenannte Kernfunktionen $K(\cdot)$ einführen. Hastie, Tibshirani und Friedman [12] geben einen guten und knappen Überblick zu solchen Kernfunktionen.

Kernfunktionen werden bei der lokal gewichteten Regression auf die quadrierten Abweichungen $y_i - m(x_i)$ angewandt, siehe weiter unten. Das effektive Datengewicht des lokalen Regressionsschätzers hat dann wieder die Form (3), so dass $\hat{m}(x)$ auch hier ein gewichteter Durchschnitt der Beobachtungen y_i ist. Die Gewichte $G_i(\cdot)$ sind dabei im Allgemeinen komplizierte Ausdrücke der Kernfunktionen $K(\cdot)$ und der Differenzen $x_i - x$ für $i = 1, \dots, n$. Nur in Spezialfällen erhalten wir solch einen einfachen Schätzer wie in Gleichung (4).

Zwei Kernfunktionen, die zu einer glatten Funktion $\hat{m}(x)$ führen, sind die Tricube-Funktion

$$K(u_i) = \frac{70}{63} (1 - |u_i|^3)^3 I(|u_i| \leq 1) \quad (5)$$

und die Epanechnikov-Funktion

$$K(u_i) = \frac{3}{4} (1 - u_i^2) I(|u_i| \leq 1) \quad (6)$$

Beide Kernfunktionen werden oft in empirischen Arbeiten angewandt. u_i misst den Abstand von x_i zu x und die Kernfunktionen geben den Beobachtungen ein desto höheres Gewicht, je näher x_i an x liegt. Abbildung 3 zeigt beide Kernfunktionen. Für $x_i = x$ nehmen die Kernfunktionen (5) und (6) mit 0,86 bzw. 0,75 den jeweils maximalen Wert an.

Falls immer k Beobachtungen in die Schätzung einbezogen werden sollen, setzt man im univariaten Fall mit einer erklärenden Variable

$$u_i = \frac{x_i - x}{d(x, k)} \quad (7)$$

und gewichtet dann weiterhin jede Beobachtung mit der gewählten Kernfunktion. Hierbei ist

$$d(x, k) = \max\{|x_i - x| : x_i \in M(x, k)\}$$

die absolute Distanz der am weitesten von x entfernten Beobachtung in $M(x, k)$. Für alle Beobachtungen aus $M(x, k)$, die näher als $d(x, k)$ an x liegen, gilt $|u_i| < 1$ und für (5) und (6) folgt $K(u_i) > 0$. Für die am weitesten entfernte Beobachtung in $M(x, k)$ gilt $|u_i| = 1$ und entsprechend $K(u_i) = 0$. Da $|u_i| > 1$ für alle Beobachtungen gilt, die kein Element von $M(x, k)$ sind, ist in diesen Fällen $K(u_i) = 0$.

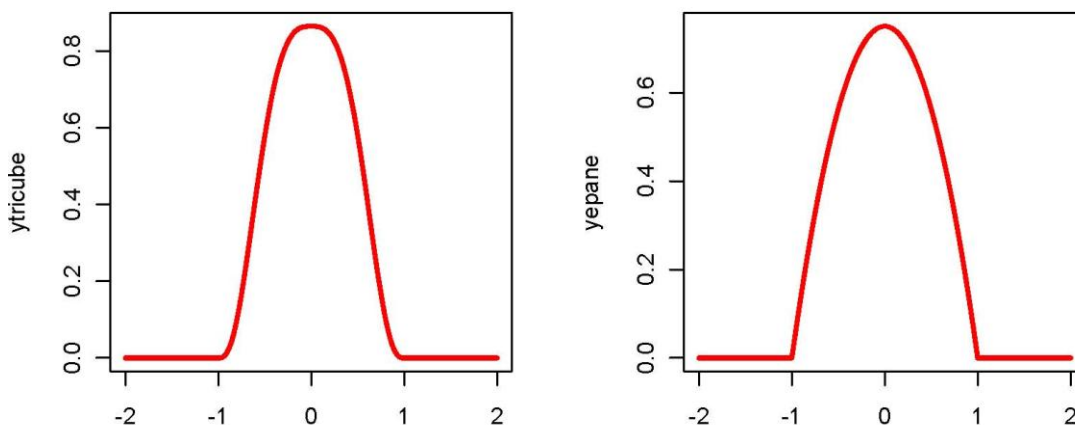



Abb. 3: Linkes Diagramm zeigt die Tricube-Kernfunktion (5) und rechtes Diagramm zeigt die Epanechnikov-Kernfunktion (6). 

Statt mit einer festen Anzahl an Beobachtungen können die Gewichte auch auf einem Intervall mit fester Länge definiert sein. In diesem Fall ist

$$u_i = \frac{x_i - x}{h},$$

wobei $h > 0$ die sogenannte Bandweite ist. Alle Beobachtungen, die im Intervall $(x - h, x + h)$ liegen, erhalten mit (5) und (6) strikt positive Gewichte, während Beobachtungen außerhalb des Intervalls ein Gewicht von 0 erhalten. Mit festen Bandweiten kann die Anzahl der Beobachtungen mit positiven Gewichten mit x variieren.

Wir zeigen nun für den univariaten Fall, wie man den Schätzer für $\hat{m}(x)$ mit einer lokal gewichteten linearen Regression erhält. Hierzu wird folgendes lokale Minimierungsproblem gelöst

$$\min_{\beta} \sum_{i=1}^n \{y_i - \beta_0 - \beta_1(x_i - x)\}^2 K(u_i). \tag{8}$$

Als Schätzer für $\beta = (\beta_0, \beta_1)^T$ erhalten wir

$$\hat{\beta}(x) = (X^T W X)^{-1} X^T W Y, \tag{9}$$

wobei die erste Spalte der $(n \times 2)$ Matrix X aus Einsen besteht und die zweite Spalte die Beobachtungen x_i enthält. Die $(n \times n)$ Diagonalmatrix W enthält die Kerngewichte $K(u_i)$ und die $(n \times 1)$ Matrix Y die Beobachtungen y_i . Der Schätzer $\hat{\beta}(x)$ ähnelt dem gewichteten Kleinstquadrat-Schätzer, siehe etwa [14]. Ein kleines Gedankenexperiment zeigt uns, dass für konstantes $K(u_i)$ Formel (8) äquivalent zum allgemeinen Kleinstquadrat Ansatz ist. Der Fall konstanter Kerngewichte tritt ein, wenn $h \rightarrow \infty$ und damit $u_i \rightarrow 0$ für $i = 1, \dots, n$. Im Zusammenspiel mit der Normalverteilung kann man also (8) als lokalen log-likelihood interpretieren. Damit eröffnet sich auch für nichtparametrische Prognoseverfahren die gesamte Klasse der robusten Statistik und der generalisierten linearen Modelle, genauer die Klasse aller (Quasi) Likelihood basierten Prognoseverfahren.

Der einzige Unterschied zum klassischen Zugang besteht darin, dass (8) nur lokal für den Punkt x minimiert wird. Bezogen auf das Ausgangsmodell (1) mit der glatten Funktion $m(x)$ wird diese für jede Beobachtung durch eine Taylor-Entwicklung erster Ordnung

$$m(x_i) = m(x) + m'(x)(x_i - x) \tag{10}$$

approximiert. Definiert man $\beta_0 \stackrel{\text{def}}{=} m(x)$ und $\beta_1 \stackrel{\text{def}}{=} m'(x)$, so entspricht der Term in der geschweiften Klammer in (8) approximativ $y_i - m(x_i)$. Bei der Schätzung der Parameter β werden Beobachtungen mit x_i nahe an x stärker gewichtet als Beobachtungen, die fern von x liegen, wie aus (8) ersichtlich. Es folgt aus (10) und obigen Definitionen, dass der Schätzer für $m(x)$ durch den Schätzer für β_0 gegeben ist, sodass mit $e_1 = (1, 0, \dots, 0)^T$ gilt:

$$\hat{m}(x) = \hat{\beta}_0 = e_1^T (X^T W X)^{-1} X^T W Y.$$

Wie die rechte Seite zeigt, wird der $(n \times 1)$ Vektor Y mit einem $(1 \times n)$ Vektor multipliziert, der nicht von den endogenen Variablen y_i abhängt. Dies lässt sich folglich auch in der Form (3) schreiben. Wie eingangs erwähnt, ist $\hat{m}(x)$ damit der gewichtete Mittelwert der y Beobachtungen. Die Gewichte $G_i(x)$ hängen dabei von der Kernfunktion $K(\cdot)$ und den exogenen Variablen $x_i - x$, $i = 1, \dots, n$ ab.

Abbildung 4 zeigt die lokal-gewichtete lineare Regression für $k = 12$, wobei der Tricube-Kernel (5) und (7) verwendet wurden.

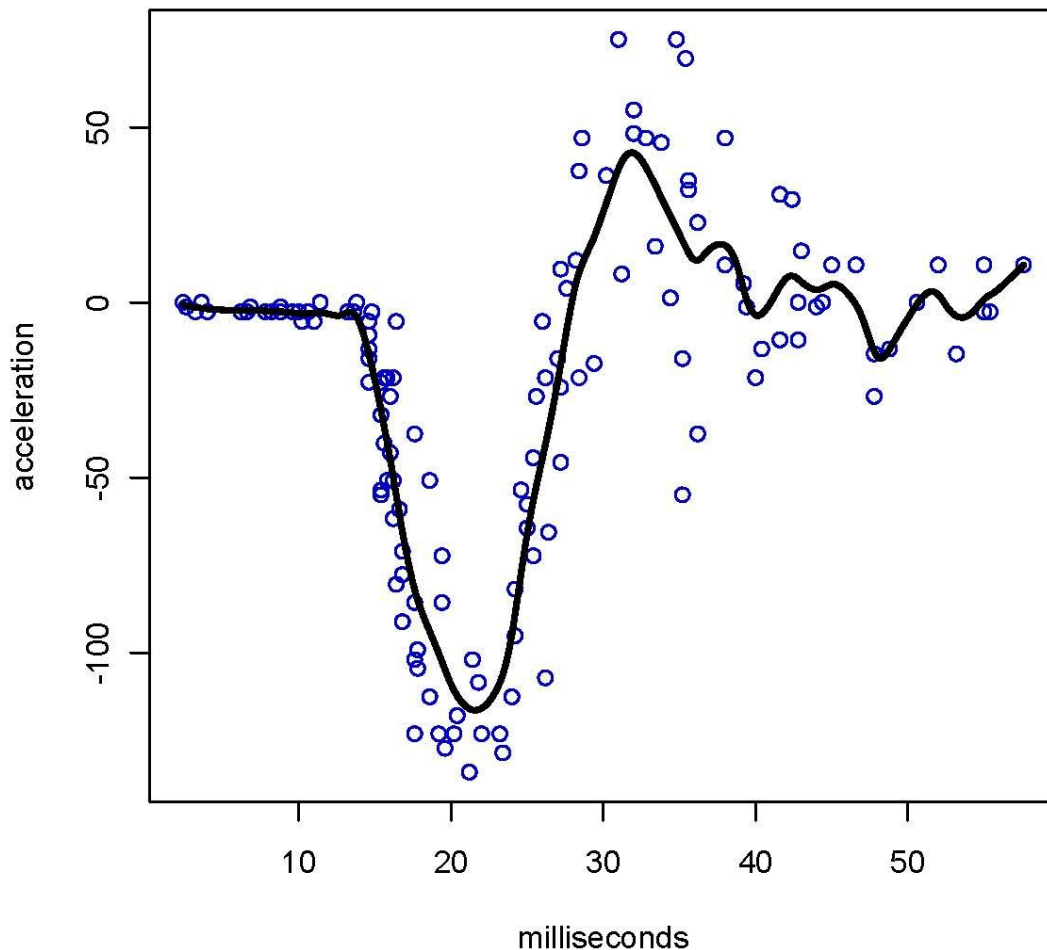



Abb. 4: Lokale lineare Regression mit $h = 1$ und Tricube-Gewichtung für alle einbezogenen Beobachtungen. Quadrate geben die Beobachtungen an. 

Falls statt der linearen Regressionsfunktion in (8) lediglich die lokale Konstante β_0 angesetzt wird, erhält man aus (9) den Nadaraya-Watson-Schätzer

$$\hat{m}(x) = \frac{\sum_{i=1}^n K(u_i) y_i}{\sum_{j=1}^n K(u_j)}. \quad (11)$$

Es ist offensichtlich, dass dieser Schätzer ein gewichteter Durchschnitt der y_i Beobachtungen ist. Für die uniforme Kernfunktion

$$K(u_i) = \begin{cases} 1/k & \text{wenn } |u_i| \leq 1 \\ 0 & \text{sonst} \end{cases}$$

erhält man für (11) sofort das arithmetische Mittel aller k Beobachtung in $M(x, k)$. Man kann ebenfalls zeigen, dass mit der uniformen Kernfunktion und einem lokal linearen Modell der Schätzer (9) der lokalen linearen Regression auf $M(x, k)$ entspricht. Damit sind dies, wie oben erwähnt, Spezialfälle des allgemeineren Ansatzes.

Für empirische Anwendungen müssen die Kernfunktion $K(\cdot)$ und der Glättungsparameter k beziehungsweise h gewählt werden. Es ist offensichtlich, dass die geschätzte Funktion $\hat{m}(x)$ umso glatter wird, je mehr Beobachtungen einbezogen werden, desto größer also k beziehungsweise h gewählt werden. Mit dem Nadaraya-Watson-Schätzer (11) prüft man leicht, dass für $h \rightarrow \infty$ und damit $u_i \rightarrow 0$ für alle i der Schätzer $\hat{m}(x)$ für alle x innerhalb des Spektrums der Beobachtungen gegen \bar{y} konvergiert. Daraus ist offensichtlich, dass Glattheit alleine kein Kriterium für die Wahl des Glättungsparameters sein kann. Der optimale Glättungsparameter sollte eher – im Sinne der oben diskutierten log-likelihood Interpretation der Gaußschen Fehler – die Abweichungen $\{\hat{m}(x) - m(x)\}^2$ über alle x minimieren. Da $m(x)$ nicht bekannt ist, kann diese Minimierung nicht direkt durchgeführt werden. Allerdings lassen sich erwartete mittlere Abweichungen theoretisch berechnen und mithilfe von Approximationen in Ausdrücke transformieren, die durch die Wahl des Glättungsparameters minimiert werden können. Alternativ können Kreuzvalidierungstechniken verwendet werden, wobei $m(x_i)$ durch y_i ersetzt wird und Beobachtung i nicht für die Schätzung von $\hat{m}(x_i)$ verwendet wird. Durch die Wahl des Glättungsparameters wird danach die Summe der quadrierten Abweichungen minimiert. Die Wahl der Kernfunktion spielt – zumindest asymptotisch – keine Rolle, solange der Glättungsparameter entsprechend angepasst wird. Härdle u.a. [11] geben einen detaillierten Überblick.

Eine natürliche Erweiterung der vorgestellten lokal gewichteten Regression besteht darin, höhere Ordnungen für die Taylor-Approximation zu verwenden und $m(x)$ mit einem Polynom der Ordnung $p > 1$ zu modellieren. Selbstverständlich lässt sich die lokale Regression auch auf multivariate Daten anwenden. Diese Erweiterungen, asymptotische Eigenschaften, Konstruktion von Konfidenzbändern und Testverfahren werden ausführlich in Härdle [6], Fan und Gijbels [5] und Härdle u.a. [11] diskutiert.

9.2.3 Prognose

Eine Punktprognose besteht darin, den Wert der endogenen Variable y an einer bestimmten Stelle x_p zu schätzen. Die obigen Ausführungen haben deutlich gemacht, dass mit nichtparametrischer Regression eine gesamte Funktion geschätzt werden kann. Für eine Prognose muss nun natürlich nicht die gesamte Funktion geschätzt werden, sondern lediglich $\hat{m}(x_p)$.

Die nichtparametrische Regressionstechnik lässt sich leicht auf stationäre Zeitreihenmodelle anwenden. Ein einfaches Beispiel ist ein autoregressives Modell erster Ordnung

$$y_t = m(y_{t-1}) + \varepsilon_t,$$

was der Form des allgemeinen Modells (1) entspricht. Für Modelle höherer Ordnung gehen weitere verzögerte Variablen ein. Falls die aktuellen Beobachtungen bis Periode T reichen, ist die Prognose für die nächste Periode mit

$$\hat{y}_{t+1} = \hat{m}(y_T)$$

gegeben. Dies entspricht dem bedingten Erwartungswert

$$E[y_{T+1}|y_T] = m(y_T),$$

Für die Prognose wird dann die unbekannte Funktion $m(y_T)$ durch den nichtparametrischen Schätzer ersetzt. Für eine gewichtete Regression mit lokaler Konstante – dem Nadaraya-Watson-Schätzer (11) – erhält man etwa als Prognose

$$\hat{m}(y_T) = \frac{\sum_{t=T-1}^{T-1} K(u_t) y_{t+1}}{\sum_{t=T-1}^{T-1} K(u_t)}, \quad (12)$$

wobei im Zähler von u_t die Abweichungen $y_t - y_T$ stehen. Der Ansatz lässt sich leicht für Prognosen mit Horizont $T > 1$ erweitern. Überblicke zur nichtparametrischen Zeitreihenanalyse geben Härdle, Lütkepohl und Chen [10] und Heiler [13].

9.3 Anwendung auf Volatilitäten

Zur Illustration wird ein nichtparametrisches Modell für die Vorhersage von Wechselkursvolatilitäten genutzt. Die Ergebnisse werden mit den Vorhersagen eines linearen Modells verglichen. Die Vorhersage von Volatilitäten ist ein sehr aktiver Bereich der Finanzmarktökonomie, da Volatilitätsvorhersagen für die Preise von Optionen, Variance Swaps oder für Value-at-Risk (VaR) Kalkulationen benötigt werden. Einen Überblick zum Stand der Forschung geben Poon und Granger [16], Diebold und Nason [4].

Die quadrierte Volatilität wird für den Wechselkurs e_t von EURO (EUR) und US Dollar (USD) für die Periode vom 1. Januar 2002 bis zum 26. Mai 2003 mithilfe von Tageskursen berechnet. Die Ausgangsdaten sind von Bloomberg data bases. Die tägliche quadrierte Volatilität wird mit

$$\sigma_t^2 = \ln(\ln e_t^{\max} - \ln e_t^{\min})^2 \quad (13)$$

berechnet, wobei die Differenz mit der höchsten und niedrigsten Kursnotierung für den Tag t berechnet wird. Parkinson [15] und Brandt und Diebold [1] haben gezeigt, dass der „range-based“-Schätzer ein zuverlässiger Schätzer für die quadrierte Volatilität ist. Abbildung 5 zeigt die 365 berechneten täglichen quadrierten Volatilitäten (siehe auch Tabelle 1).

Zeitreihe	Mittelwert	Standardfehler	Minimum	Maximum	Beobachtungen
e_t	1,53	0,07	1,41	1,66	366
$\sigma_t^2 \times 10^{-5}$	1,95	3,13	0,00	21,68	365

Tab. 1: Deskriptive Statistiken für den EUR/USD Tageswechselkurs und dessen tägliche quadrierte Volatilität für den Zeitraum vom 1. Januar 2002 bis zum 26. Mai 2003.

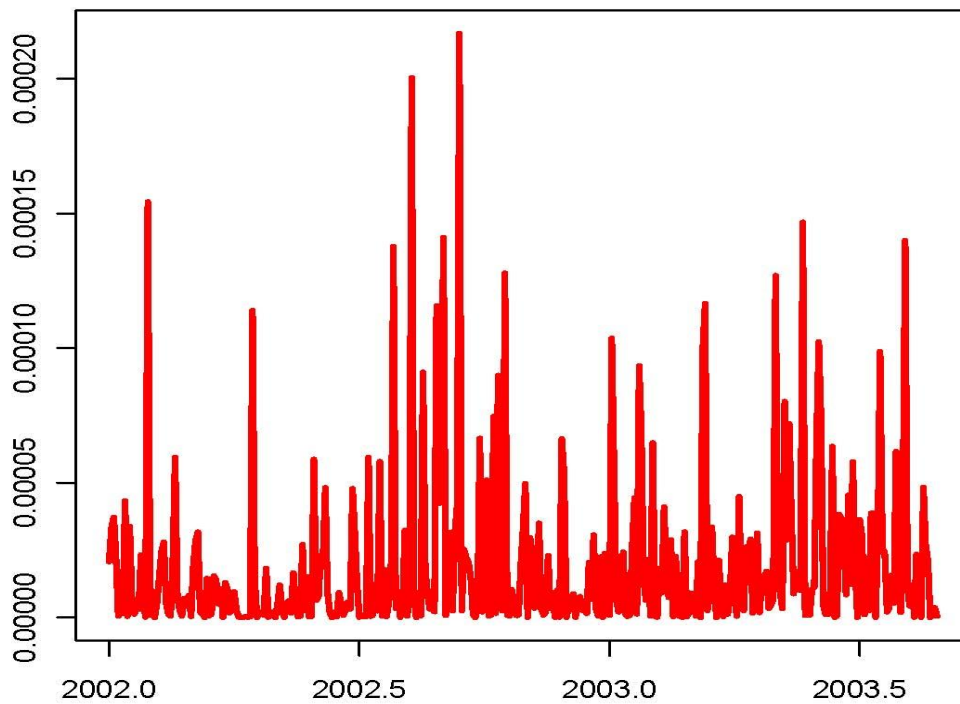



Abb. 5: Tägliche quadrierte Volatilität des EUR/USD Wechselkurses von 1. Januar 2002 bis zum 26. Mai 2003. 

Um die Ein-Schritt-Prognosen zu berechnen, werden jeweils die $T = 250$ vorhergehenden Beobachtungen genutzt, was einer Jahreshälfte entspricht. Insgesamt werden $H = 65$ Prognosen berechnet. Der Glättungsparameter k , der die lokale Umgebung definiert, wird durch die Kreuzvalidierungs-Technik (gestützt durch 258 Beobachtungen) gewählt. Genauer wird die verallgemeinerte Kreuzvalidierung-Technik verwendet: Eine Penalizing-Funktion \mathcal{E} bestraft zu kleine Glättungsparameter, siehe Härdle u.a. [11],

$$\mathcal{E}_{GCV}(u) = (1 - u)^{-2},$$

wobei $u = 1/k$. Der Glättungsparameter minimiert das Produkt der Summe der quadrierten Abweichungen von $y_i - \hat{m}(x_i)$ und der verallgemeinerten Kreuzvalidierung-Penalizing-Funktion

$$k = \arg \min \frac{1}{n} \sum_{i=1}^n \{y_i - \hat{m}(x_i)\}^2 \mathcal{E}_{GCV}(1/k).$$

Der erhaltene optimale Glättungsparameter ist $k = 44$. Abbildung 6 illustriert den LWR-Vorhersage-Prozess für die Varianz am 30. Januar 2003. Die Punkte sind die 44 Nachbarn, mit denen die lokalen Parameter berechnet werden.

Man sieht, dass die Punkte eine lineare Kurve mit negativer Drift implizieren. Man sieht zugleich, dass positive Prognosen nicht grundsätzlich ausgeschlossen sind. Lügen etwa die Punkte in der rechten oberen Ecke weiter oben, so würde die Regressionsgerade steiler ansteigen und sie könnte für kleine Beobachtungswerte im negativen Bereich verlaufen. Neben den lokal gewichteten Regressionen wird auch ein parametrisches Modell angepasst. Da für den gesamten Zeitraum das lineare AR(1)-Modell mit 258 Beobachtungen die beste Anpassung bringt, wird dieses auch für die Prognosen verwendet.

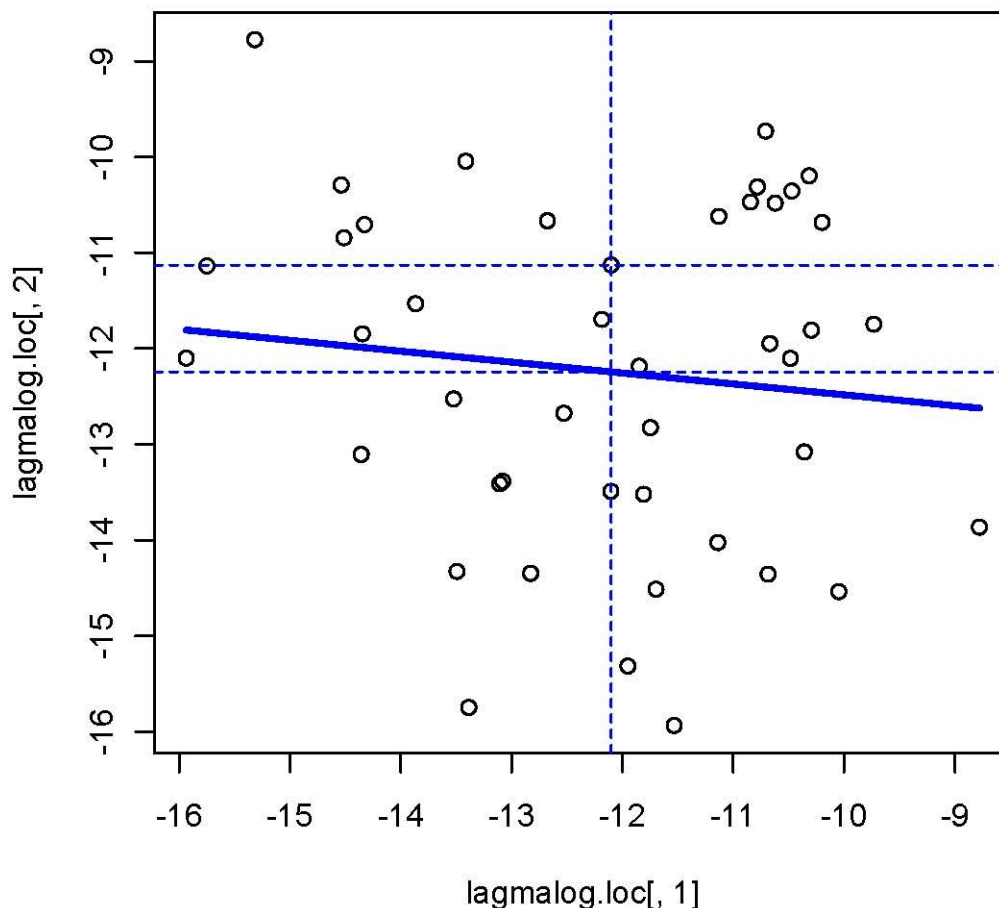


Abb. 6: LWR-Prognose der quadrierten Volatilität für den 30. Januar 2003 mit $k = 44$ benachbarten und zeitlich vorhergehenden Beobachtungen. . [Q](#)

Alle Berechnungen werden in *R* durchgeführt, Härdle u.a. [8], Härdle u.a. [9]. Abbildung 7 zeigt die 65 Ein-Schritt-Prognosen. Das linke Panel ist für die LWR-Prognosen und das rechte Panel für die Prognosen mit dem parametrischen AR(1)-Modell. Für die LWR wurde der Tricube Kernel (5) und (7) verwendet. Für jede LWR-Prognose werden die Parameter mit den 44 lokalen Nachbarn aus den 250 zeitlich vorausgehenden Beobachtungen geschätzt. Das AR(1)-Modell wird jeweils für alle 250 vorausgehende Beobachtungen angepasst. Die Prognosefehler des LWR-Modells und des AR(1)-Modells werden in der Abbildung 8 gezeigt.

Tabelle 2 zeigt Statistiken für die durchschnittlichen Prognosefehler, wobei der mittlere quadratische Fehler

$$MQF = \frac{1}{H} \sum_{h=1}^H (\hat{y}_{T+h} - y_{T+h})^2$$

und der mittlere absolute Fehler

$$MAF = \frac{1}{H} \sum_{h=1}^H |\hat{y}_{T+h} - y_{T+h}|$$

angegeben werden, vgl. auch Kapitel 19 dieses Buches. Die Prognosen eines Verfahrens sind umso besser, je geringer diese Fehler sind. Tabelle 2 zeigt die Prognosefehler der AR(1)- und LWR-Modelle. Das nichtparametrische Modell hat im Durchschnitt einen leicht größeren Prognosefehler als parametrische Verfahren. Dies ist zunächst einmal eine schwächere Prognose-Leistung, die jedoch nicht überinterpretiert werden sollte, da die nichtparametrischen Verfahren gerade bei wechselnden Regimen und nichtstationären Phänomenen eine wesentlich bessere Leistung erzielen, Härdle u.a. [7]. Zudem ist die nichtparametrische LWR für die gewählte Anwendung hilfreich, da sie offenbart, dass ein parametrisches Modell für eine statistische Modellierung ausreichend ist. Dies könnte auch dadurch überprüft werden, indem man den Glättungsparameter k immer weiter vergrößert. In der Tat würden sich für dieses Gedankenexperiment die linke und die rechte Seite des Panels in Abb. 7 angleichen.

	$MQF \times 10^{-9}$	$MAF \times 10^{-5}$	Schätzfenster
AR(1)	1,364	2,168	250
LWR ($k = 44$)	1,366	2,170	250

Tab. 2: Die Ergebnisse der Vorhersagen beider Modelle anhand des mittlere MSE und MAE.

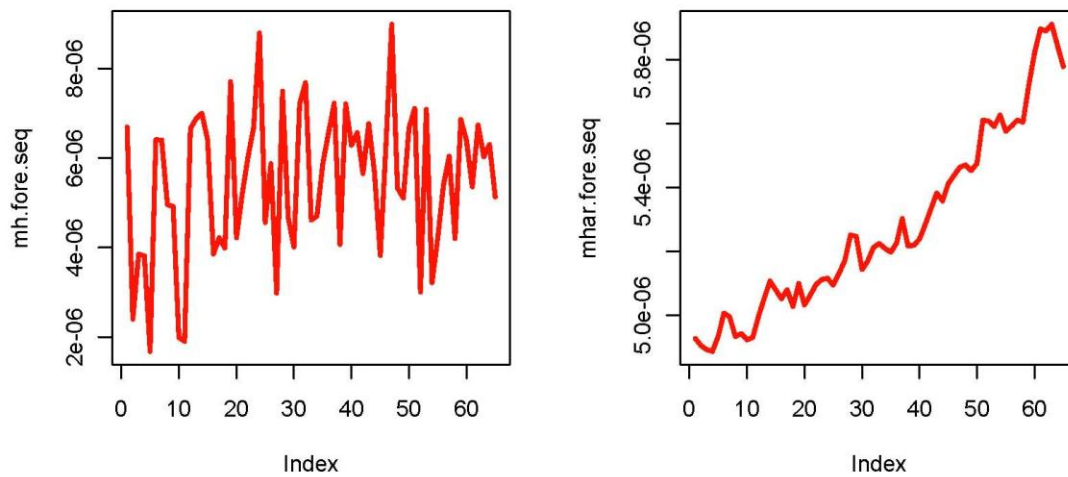


Abb. 7: $H = 65$ lokale gewichtete (links) und AR(1) (rechts) Ein-Schritt-Prognosen, wobei $k = 44$ lokale vorhergehenden Beobachtungen verwendet wurden. 

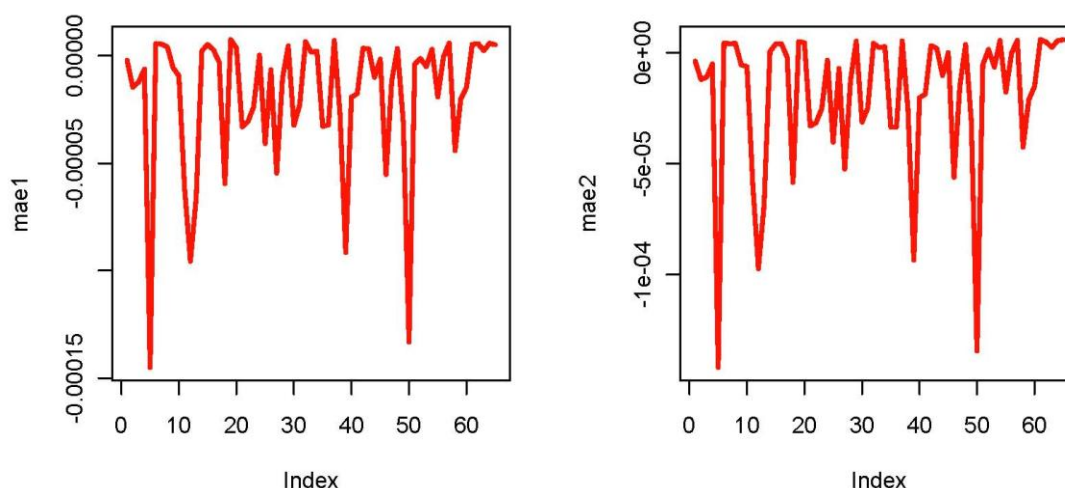



Abb. 8: Prognosefehler des LWR Modells (links) und AR(1)-Modells (rechts) mit $T = 250$ und $k = 44$. 

Danksagung: Für finanzielle Unterstützung bedanken wir uns bei der Deutschen Forschungsgemeinschaft, SFB 649 „Ökonomisches Risiko“.

9.4 Literatur

- [1] Brandt, M.W. und Diebold, F.X., A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations, *Journal of Business*, vol. 78, no. 4 (2005).
- [2] Cleveland, W.S., Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* 74 (1979), S. 829 ff.
- [3] Cleveland, W.S. und Devlin, S.J., Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association* 83 (1988), S. 596 ff.
- [4] Diebold, F.X. und Nason, J.A., Nonparametric Exchange Rate Prediction?, *Journal of International Economics* 28 (1990), S. 315 ff.
- [5] Fan, J. und Gijbels, I., *Local Polynomial Modeling and its Applications*. London 1996.
- [6] Härdle, W., *Applied Nonparametric Regression*. Econometric Society Monographs No. 19, Cambridge 1990.
- [7] Härdle, W., Herwartz, H. und Spokoiny, V., Time Inhomogeneous Multiple Volatility Modeling, *Journal of Financial Econometrics* 1 (2003), S. 55 ff.
- [8] Härdle, W., Okhrin, O. und Okhrin, Y., *Basis Elements of Computational Statistics*, Springer Verlag, Heidelberg 2011.
- [9] Härdle, W. und Simar, L., *Applied Multivariate Statistical Analysis*, Springer Verlag, Heidelberg 2011.
- [10] Härdle, W., Lütkepohl, H. und Chen, R., A Review of Nonparametric Time Series Analysis, *International Statistical Review* 65 (1997), S. 49 ff.
- [11] Härdle, W., Müller, M., Sperlich, S. und Werwatz, A., *Introduction to Non- and Semiparametric Modeling*, Berlin 2004.
- [12] Hastie, T., Tibshirani, R. und Friedman, J., *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Springer Series in Statistics, Berlin 2001.
- [13] Heiler, S., Nonparametric Time Series Analysis: Nonparametric Regression, Locally Weighted Regression, Autoregression, and Quantile Regression, in: Peña, D., Tiao, G.C. und Tsay R.S. (Hrsg.), *A Course in Time Series Analysis*, Wiley Series in Probability and Statistics, New York 2001, S. 308 ff.

- [14] Greene, W. H., *Econometric Analysis*, 6. Aufl., Upper Saddle River, NJ 2008.
- [15] Parkinson, M., The Extreme Value Method for Estimating the Variance of the Rate of Return, *Journal of Business* 53 (1980), S. 61 ff.
- [16] Poon, S.-H. und Granger, C.W.J., Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature* 41 (2003), S. 478 ff.

SFB 649 Discussion Paper Series 2010

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Volatility Investing with Variance Swaps" by Wolfgang Karl Härdle and Elena Silyakova, January 2010.
- 002 "Partial Linear Quantile Regression and Bootstrap Confidence Bands" by Wolfgang Karl Härdle, Ya'acov Ritov and Song Song, January 2010.
- 003 "Uniform confidence bands for pricing kernels" by Wolfgang Karl Härdle, Yarema Okhrin and Weining Wang, January 2010.
- 004 "Bayesian Inference in a Stochastic Volatility Nelson-Siegel Model" by Nikolaus Hautsch and Fuyu Yang, January 2010.
- 005 "The Impact of Macroeconomic News on Quote Adjustments, Noise, and Informational Volatility" by Nikolaus Hautsch, Dieter Hess and David Veredas, January 2010.
- 006 "Bayesian Estimation and Model Selection in the Generalised Stochastic Unit Root Model" by Fuyu Yang and Roberto Leon-Gonzalez, January 2010.
- 007 "Two-sided Certification: The market for Rating Agencies" by Erik R. Fasten and Dirk Hofmann, January 2010.
- 008 "Characterising Equilibrium Selection in Global Games with Strategic Complementarities" by Christian Basteck, Tijmen R. Daniels and Frank Heinemann, January 2010.
- 009 "Predicting extreme VaR: Nonparametric quantile regression with refinements from extreme value theory" by Julia Schaumburg, February 2010.
- 010 "On Securitization, Market Completion and Equilibrium Risk Transfer" by Ulrich Horst, Traian A. Pirvu and Gonçalo Dos Reis, February 2010.
- 011 "Illiquidity and Derivative Valuation" by Ulrich Horst and Felix Naujokat, February 2010.
- 012 "Dynamic Systems of Social Interactions" by Ulrich Horst, February 2010.
- 013 "The dynamics of hourly electricity prices" by Wolfgang Karl Härdle and Stefan Trück, February 2010.
- 014 "Crisis? What Crisis? Currency vs. Banking in the Financial Crisis of 1931" by Albrecht Ritschl and Samad Sarferaz, February 2010.
- 015 "Estimation of the characteristics of a Lévy process observed at arbitrary frequency" by Johanna Kappusl and Markus Reiß, February 2010.
- 016 "Honey, I'll Be Working Late Tonight. The Effect of Individual Work Routines on Leisure Time Synchronization of Couples" by Juliane Scheffel, February 2010.
- 017 "The Impact of ICT Investments on the Relative Demand for High-Medium-, and Low-Skilled Workers: Industry versus Country Analysis" by Dorothee Schneider, February 2010.
- 018 "Time varying Hierarchical Archimedean Copulae" by Wolfgang Karl Härdle, Ostap Okhrin and Yarema Okhrin, February 2010.
- 019 "Monetary Transmission Right from the Start: The (Dis)Connection Between the Money Market and the ECB's Main Refinancing Rates" by Puriya Abbassi and Dieter Nautz, March 2010.
- 020 "Aggregate Hazard Function in Price-Setting: A Bayesian Analysis Using Macro Data" by Fang Yao, March 2010.
- 021 "Nonparametric Estimation of Risk-Neutral Densities" by Maria Grith, Wolfgang Karl Härdle and Melanie Schienle, March 2010.

SFB 649 Discussion Paper Series 2010

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 022 "Fitting high-dimensional Copulae to Data" by Ostap Okhrin, April 2010.
- 023 "The (In)stability of Money Demand in the Euro Area: Lessons from a Cross-Country Analysis" by Dieter Nautz and Ulrike Rondorf, April 2010.
- 024 "The optimal industry structure in a vertically related market" by Raffaele Fiocco, April 2010.
- 025 "Herding of Institutional Traders" by Stephanie Kremer, April 2010.
- 026 "Non-Gaussian Component Analysis: New Ideas, New Proofs, New Applications" by Vladimir Panov, May 2010.
- 027 "Liquidity and Capital Requirements and the Probability of Bank Failure" by Philipp Johann König, May 2010.
- 028 "Social Relationships and Trust" by Christine Binzel and Dietmar Fehr, May 2010.
- 029 "Adaptive Interest Rate Modelling" by Mengmeng Guo and Wolfgang Karl Härdle, May 2010.
- 030 "Can the New Keynesian Phillips Curve Explain Inflation Gap Persistence?" by Fang Yao, June 2010.
- 031 "Modeling Asset Prices" by James E. Gentle and Wolfgang Karl Härdle, June 2010.
- 032 "Learning Machines Supporting Bankruptcy Prediction" by Wolfgang Karl Härdle, Rouslan Moro and Linda Hoffmann, June 2010.
- 033 "Sensitivity of risk measures with respect to the normal approximation of total claim distributions" by Volker Krätschmer and Henryk Zähle, June 2010.
- 034 "Sociodemographic, Economic, and Psychological Drivers of the Demand for Life Insurance: Evidence from the German Retirement Income Act" by Carolin Hecht and Katja Hanewald, July 2010.
- 035 "Efficiency and Equilibria in Games of Optimal Derivative Design" by Ulrich Horst and Santiago Moreno-Bromberg, July 2010.
- 036 "Why Do Financial Market Experts Misperceive Future Monetary Policy Decisions?" by Sandra Schmidt and Dieter Nautz, July 2010.
- 037 "Dynamical systems forced by shot noise as a new paradigm in the interest rate modeling" by Alexander L. Baranovski, July 2010.
- 038 "Pre-Averaging Based Estimation of Quadratic Variation in the Presence of Noise and Jumps: Theory, Implementation, and Empirical Evidence" by Nikolaus Hautsch and Mark Podolskij, July 2010.
- 039 "High Dimensional Nonstationary Time Series Modelling with Generalized Dynamic Semiparametric Factor Model" by Song Song, Wolfgang K. Härdle, and Ya'acov Ritov, July 2010.
- 040 "Stochastic Mortality, Subjective Survival Expectations, and Individual Saving Behavior" by Thomas Post and Katja Hanewald, July 2010.
- 041 "Prognose mit nichtparametrischen Verfahren" by Wolfgang Karl Härdle, Rainer Schulz, and Weining Wang, August 2010.