

Härdle, Wolfgang Karl; Chen, Ying; Schulz, Rainer

**Working Paper**

## Prognose mit nichtparametrischen Verfahren

Papers, No. 2004,07

**Provided in Cooperation with:**

CASE - Center for Applied Statistics and Economics, Humboldt University Berlin

*Suggested Citation:* Härdle, Wolfgang Karl; Chen, Ying; Schulz, Rainer (2004) : Prognose mit nichtparametrischen Verfahren, Papers, No. 2004,07, Humboldt-Universität zu Berlin, Center for Applied Statistics and Economics (CASE), Berlin

This Version is available at:

<https://hdl.handle.net/10419/22181>

**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.*

# Prognose mit nichtparametrischen Verfahren

Ying Chen

Wolfgang Härdle

Rainer Schulz

CASE - Center for Applied Statistics and Economics

Humboldt-Universität zu Berlin


## 1 Einleitung

Statistische Prognosen basieren auf der Annahme, dass ein funktionaler Zusammenhang zwischen der zu prognostizierenden Variable  $y$  und anderen  $J$  dimensionalen beobachtbaren Variablen  $x = (x_1, \dots, x_J)^T \in \mathfrak{R}^J$  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 bivariaten Fall ist das lineare Modell  $y = \beta_0 + \beta_1 x$ . Sind die beiden unbekannt Parameter  $\beta_0$  und  $\beta_1$  mit Hilfe 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 in Folge zu schlechten Prognosen führen.

Nichtparametrische Verfahren gehen ebenfalls von einem funktionalen Zusammenhang aus, geben aber kein festes parametrisches Modell vor. Sie sind deshalb hervorragend geeignet, 1) um Daten explorativ darzustellen, 2) um parametrische Modelle zu überprüfen und 3) um selbst als Schätzer für den funktionalen Zusammenhang zu dienen, Cleveland und Devlin (1988). Nichtparametrische Verfahren können daher problemlos auch zur Prognose eingesetzt werden.

Abschnitt 2 stellt nichtparametrische Verfahren vor und erläutert deren grundsätzliche Struktur. Der Schwerpunkt liegt auf bivariaten Modellen und auf der Motivation der verwendeten Verfahren. Abschnitt 3 präsentiert eine Anwendung für eine Zeitreihe von Wechselkursvarianz. Es werden Prognosen mit nichtparametrischen Verfahren berechnet und deren Güte mit den Prognosen eines AR(1) Zeitreihenmodells verglichen. Für die gewählten

Daten zeigt sich, dass diese durch das parametrische Modell sehr gut erfasst werden und das nichtparametrische Modell keine bessere Prognosegüte liefert. Zugleich veranschaulichen die praktischen Beispiele, wie parametrische Verfahren für Prognosen eingesetzt werden können. Alle Beispiele, welche mit dem Symbol  versehen sind, lassen sich interaktiv über [www.xplore-stat.de](http://www.xplore-stat.de) ausführen.

## 2 Nichtparametrische Verfahren

### 2.1 Einführung

Ein nichtparametrisches Modell geht von dem allgemeinen funktionalen Zusammenhang aus

$$y = m(x) , \quad (1)$$

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

$$y = m(x) + \varepsilon \quad (2)$$

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 Erwartungswertsfunktion  $m(x)$  zu schätzen. Dies geschieht durch lokale Mittelwertbildung über die  $i = 1, \dots, n$  Beobachtungen  $y_i$  mit

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

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

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

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

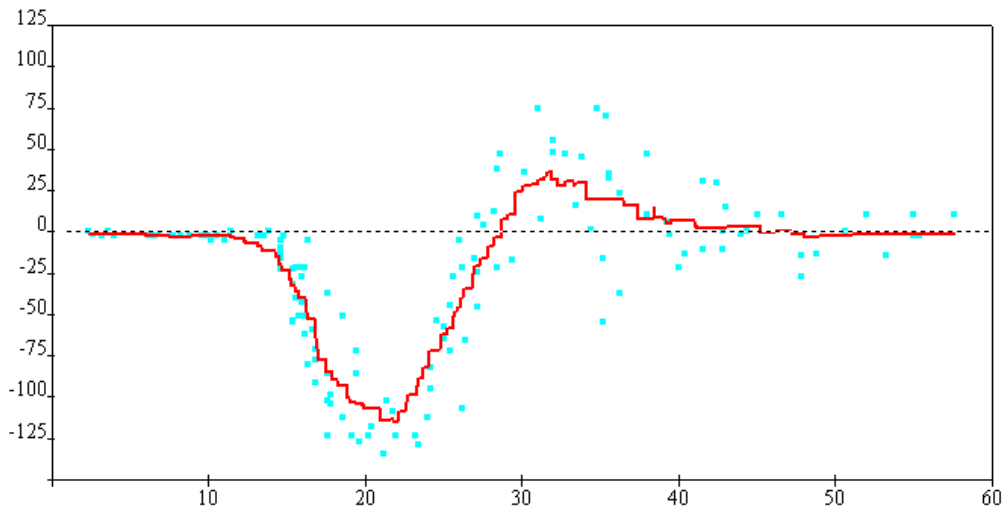
wobei

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

die Menge der  $k$  Beobachtungen ist, die am nächsten an  $x$  liegen. Die Nähe kann etwa mit der euklidischen Distanz  $\sqrt{(x_i - x)^\top (x_i - x)}$  bestimmt werden;  $I(\cdot)$  ist die Indikatorfunktion. Wird etwa  $k = 12$  gewählt, so ist

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

und für jedes  $x$  wird das arithmetische Mittel aus 12 Beobachtungen berechnet. Es ist offensichtlich, dass  $\hat{m}(x)$  dem arithmetischen Mittel aller Beobachtungen entspricht, wenn  $k = n$  gewählt wird. Abbildung 1 zeigt die geschätzte Funktion für den bivariaten Fall:  $(x, y) \in \mathfrak{R}^2$ .



**Abbildung 1:** Nichtparametrische Regression mit kNN Schätzer mit  $k = 12$ , uniformer Gewichtung und Mittelwertbildung für alle einbezogenen Beobachtungen. Quadrate geben die Beobachtungen an. Es handelt sich um den *Motorcycle* Datensatz mit  $n = 133$  Beobachtungen aus Table 2, Appendix 2 in Härdle (1990).

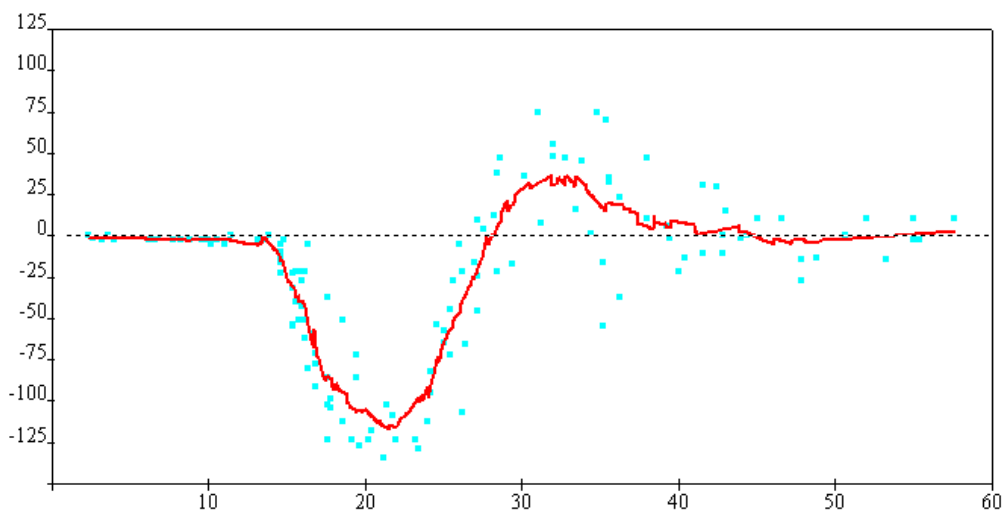


Man erkennt, dass  $\hat{m}(x)$  eine Treppenfunktion ist, die für alle  $x$  konstant ist, die über die gleiche Indexmenge verfügen. Der frei zu wählende Parameter  $k$  bestimmt die Glattheit der geschätzten Kurve. Variiert man  $k$ , so entsteht eine Familie von Regressionskurven, die alle durch lokale Mittelung entstanden sind.

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, so dass

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

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



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



## 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 ein genereller Ansatz für die nichtparametrische Regression vorgestellt, der die Beispiele aus der Einleitung als Sonderfälle umfasst. Da die Gewichte bei der lokal gewichteten Regression nicht direkt auf die Beobachtung  $y_i$  angewendet werden – wie in (3) –, sondern auf die quadrierten Abweichungen

$y_i - m(x_i)$ , werden die Gewichtungsfunktionen mit  $K(\cdot)$  bezeichnet. Man findet sie in der Literatur unter dem Namen Kernfunktionen. Hastie, Tibshirani und Friedman (2001) geben einen guten und knappen Überblick. Das effektive Datengewicht der lokalen Regression hat wieder die Form (3), so dass  $\hat{m}(x)$  auch hier ein gewichteter Durchschnitt der Beobachtungen  $y_i$  ist. Die Gewichte  $G(\cdot)$  sind dabei im Allgemeinen komplizierte Ausdrücke der Differenzen  $x_i - x$  für  $i = 1, \dots, n$  und den Kernfunktionen  $K(\cdot)$ .

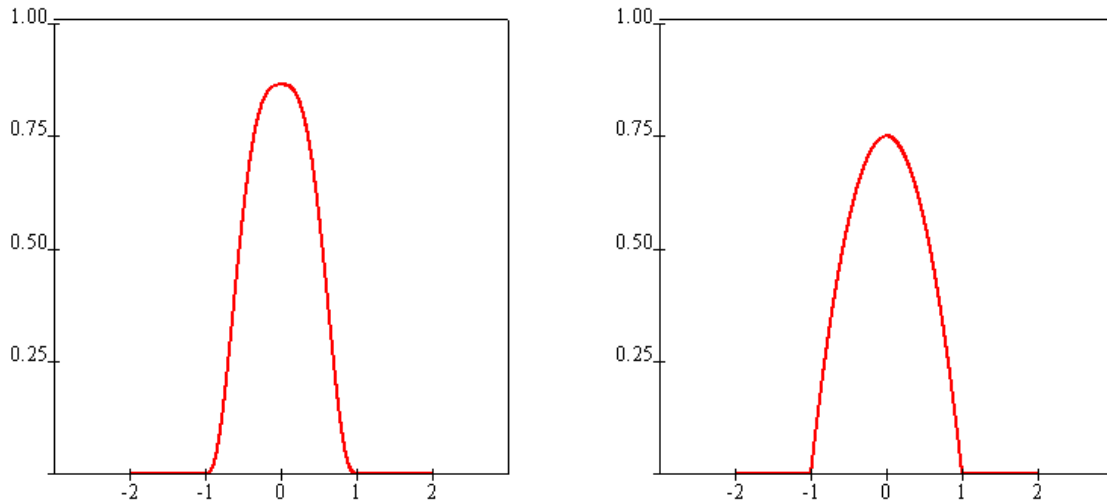
Die Tricube Funktion

$$K(u_i) = \frac{70}{81}(1 - |u_i|^3)^3 I(|u_i| \leq 1) \quad (5)$$

und die Epanechnikov Funktion

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

sind zwei häufig angewandte Kernfunktionen.  $u_i$  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:** Linkes Diagramm zeigt die Tricube Kernfunktion (5) und rechtes Diagramm zeigt die Epanechnikov Kernfunktion (6).



Falls immer  $k$  Beobachtungen einbezogen werden sollen, ist im bivariaten Fall

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

wobei

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

die absolute Distanz der am weitesten von  $x$  entfernten Beobachtungen in  $M(x, k)$  ist. 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  $K > 0$ . Für  $x_i = x$  nehmen die Kernfunktion (5) und (6) mit 1 beziehungsweise 0,8875 den maximalen Wert an. Für die am weitesten entfernte Beobachtung in  $M(x, k)$  gilt  $|u_i| = 0$  und entsprechend  $K = 0$ . Da  $|u_i| > 1$  für alle Beobachtungen gilt, die kein Element von  $M(x, k)$  sind, ist in diesen Fällen  $K = 0$ . Abbildung 3 zeigt beide Kernfunktionen.

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

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

wobei  $h > 0$  die Bandweite angibt. 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.

Bei der lokal gewichteten linearen Regression wird im bivariaten Fall für jedes  $x$  (beziehungsweise für ein hinreichend feines Gitter von  $x$  Werten) folgendes lokales Minimierungsproblem gelöst

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

Als Schätzer für  $\beta = (\beta_0, \beta_1)^T$  erhält man

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

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 Gewichte  $K(u_i)$  und die  $(n \times 1)$  Matrix  $Y$  die Beobachtungen  $y_i$ . Der Schätzer  $\hat{\beta}(x)$  ähnelt dem gewichteten Kleinstquadrat Schätzer, Johnston (1984). Der große Unterschied besteht jedoch daran, 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

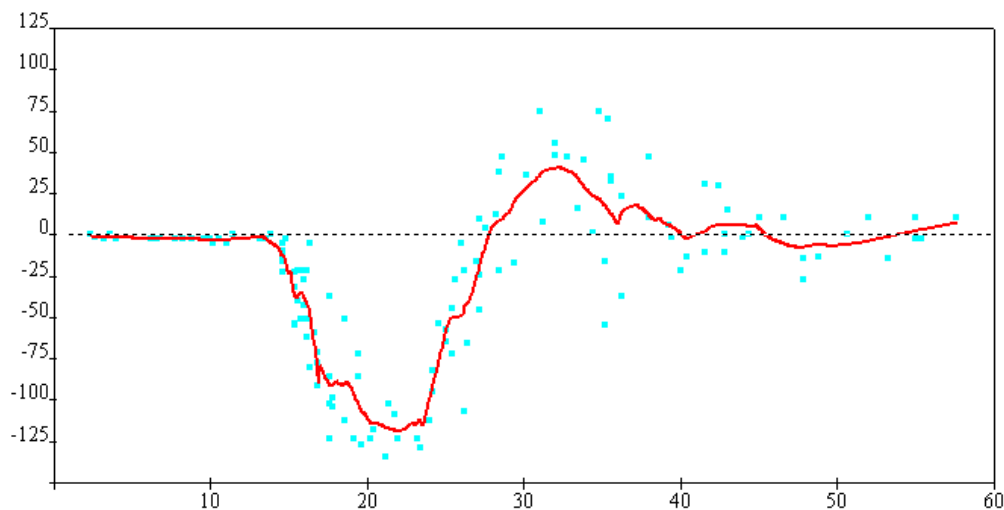
$$m(x_i) = m(x) + m'(x)(x_i - x) \quad (10)$$


approximiert. Definiert man  $\beta_0 = m(x)$  und  $\beta_1 = m'(x)$ , so entspricht der Term in der geschweiften Klammer in (8) approximativ  $y_i - m(x_i)$ . Bei der Schätzung der unbekannt Parameter  $\beta$  werden Beobachtungen nahe an  $x$  stärker gewichtet als Beobachtungen, die fern von  $x$  liegen. Es folgt aus (10), dass der Schätzer für  $m(x)$  durch den Schätzer für  $\beta_0$  gegeben ist, so dass mit  $e_1 = (1, 0, \dots, 0)^\top$

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

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 auch in der Form (3) schreiben. Damit ist  $\hat{m}(x)$  der gewichtete Mittelwert der  $y$  Beobachtungen, wobei die Gewichte  $G_i(x)$  von den exogenen Variablen  $x_i - x$ ,  $i = 1, \dots, n$  und der Kernfunktion  $K$  abhängen.

Abbildung 4 zeigt die lokale lineare Regression für  $k = 12$ , wobei der Tricube Kernel (5) und (7) verwendet wurde.



**Abbildung 4:** Lokale lineare Regression mit  $k = 12$  und Tricube Gewichtung für alle einbezogenen Beobachtungen. Quadrate geben die Beobachtungen an. 



Falls statt der linearen Regressionsfunktion in (8) lediglich die lokale Konstante  $\beta_0$  angesetzt wird, erhält man für (9) den Nadaraya-Watson Schätzer

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

Auch dies ist ein gewichteter Durchschnitt der  $y_i$  Beobachtungen. 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.

Für die empirische Anwendung muss die Kernfunktion  $K(\cdot)$  und der Glättungsparameter  $k$  beziehungsweise  $h$  gewählt werden. Es ist offensichtlich, dass die geschätzte Funktion  $\hat{m}(x)$  desto glatter wird, je mehr Beobachtungen einbezogen werden, desto größer also  $k$  und  $h$  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 die totale Variation  $\{\hat{m}(x) - m(x)\}^2$  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 mit Hilfe von Approximationen in Ausdrücke transformieren, die durch die Wahl des Glättungsparameters minimiert werden können. Alternativ können Kreuzvalidierungs-Techniken 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 – zumindestens asymptotisch – keine Rolle, solange der Glättungsparameter entsprechend angepasst wird. Härdle et al. (2004) 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. Und selbstverständlich lässt sich die lokale Regression auch auf multivariate Daten anwenden. Diese Erweiterungen, asymptotische Eigenschaften, Konstruktion von Konfidenzbändern und Testverfahren werden ausführlich in Härdle (1990), Fan und Gijbels (1996) und Härdle et al. (2004) diskutiert.

## 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, dann ist die Prognose für die nächste Periode mit

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

gegeben. Dies entspricht dem bedingten Erwartungswert

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

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

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

wobei im Zähler von  $u_t$  die Abweichungen  $y_t - y_T$  steht. Der Ansatz lässt sich leicht für Prognosen mit Horizont  $T > 1$  erweitern. Überblicke zur nichtparametrischen Zeitreihenanalyse geben Härdle, Lütkepohl und Chen (1997) und Heiler (2001).

### 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 Preisen von Optionen oder für VaR Kalkulationen benötigt werden. Einen Überblick zum Stand der Forschung geben Poon und Granger (2003). Allerdings werden dort keine nichtparametrischen Modelle diskutieren, da deren Vorhersagekraft nach einigen Studien nicht besser ist als diejenige von parametrischen Modellen, siehe etwa Diebold und Nason (1990). Natürlich hängt dieses Ergebnis von der gewählten Anwendung ab und es ist immer von Vorteil, parametrische und nichtparametrische Modellen gleichzeitig zu verwenden.

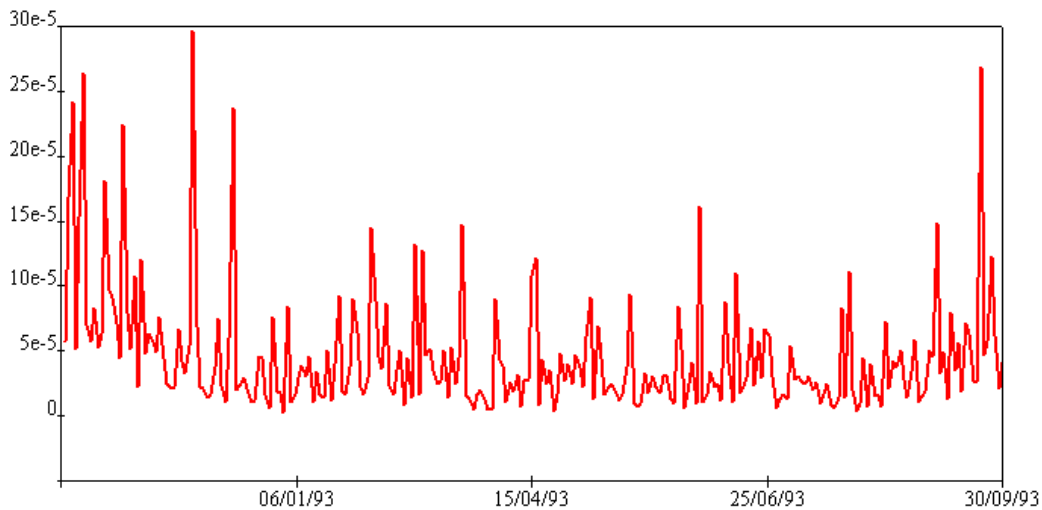
Die quadrierte Volatilität wird für den Wechselkurs von Deutscher Mark (DEM) und US Dollar (USD) für 258 Tage vom 1 Oktober 1992 bis zum 30 September 1993 mit Hilfe von logarithmierten Intra-Tageskursen berechnet. Die Ausgangsdaten umfassen 25434 Beobachtungen (siehe Tabelle 1), und sind online über die Datenbank MD\*Base ([www.quantlet.org/mdbase](http://www.quantlet.org/mdbase)) abrufbar (demusd.dat). Die tägliche quadrierte Volatilität wird mit

$$\sigma_t^2 = 0,361 * Range_t^2 \quad (13)$$

berechnet, wobei  $Range_t$  die Differenz zwischen höchstem und niedrigstem Log-Kurs pro Tag  $t$  ist. Parkinson (1980) und Brandt und Diebold (2003) haben gezeigt, dass der range-based Schätzer ein zuverlässiges Maß für die quadrierte Volatilität ist. Abbildung 5 zeigt die 258 berechneten täglichen quadrierten Volatilitäten (siehe auch Tabelle 1).

Data	Mean	Std. Error	Min.	Max.	Observations
DEM/USD	0,48	0,04	0,33	0,56	25434
Variance	$4,57 * 10^{-5}$	$4,65 * 10^{-5}$	$0,20 * 10^{-5}$	$29,60 * 10^{-5}$	258

**Tabelle 1:** Intra-Tageskurs von Deutscher Mark (DEM) und US Dollar (USD) und tägliche quadrierte Volatilität des DEM/USD Wechselkurses von 1. Oktober 1992 bis zum 30. September 1993.



**Abbildung 5:** Tägliche quadrierte Volatilität des DEM/USD Wechselkurses von 1. Oktober 1992 bis zum 30. September 1993. Werte wurden mit dem range-based Schätzer (12) berechnet.

Um die Ein-Schritt-Prognosen zu berechnen, werden jeweils die  $T = 180$  vorhergehenden Beobachtungen genutzt, was einer Jahreshälfte entspricht. Insgesamt werden  $H = 78$  Prognosen berechnet. Der Glättungsparameter  $k$ , der die lokale Umgebung definiert, wird durch die Kreuzvalidierungs-Technik (gestützt 258 Beobachtungen) gewählt. Genauer wird die verallgemeinerte Kreuzvalidierung Technik verwendet: Eine Penalizing-Funktion  $\Xi$  bestraft zu kleine Glättungsparameter, siehe Härdle, Müller, Sperlich und Werwatz (2004),

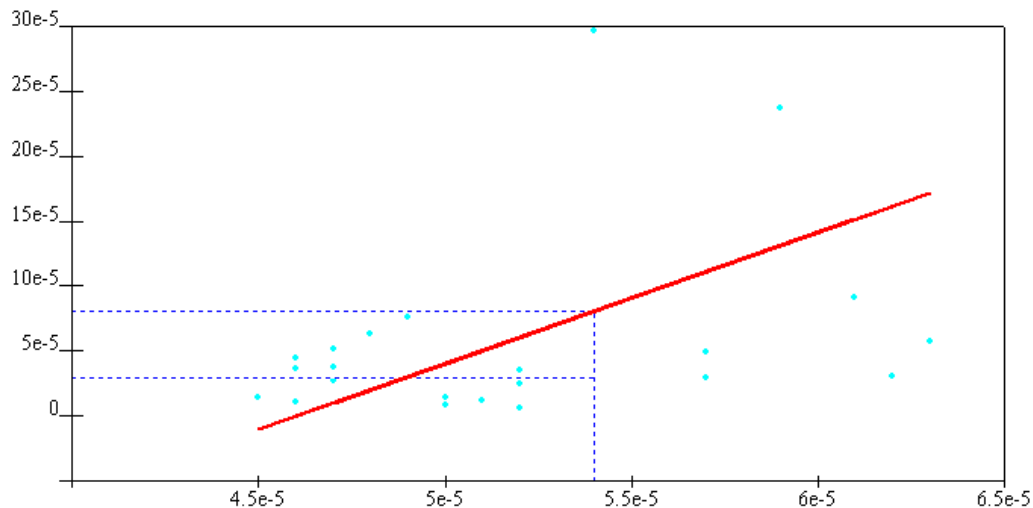
$$\Xi_{GCV}(u) = (1 - u)^{-2},$$

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

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

Der erhaltene optimale Glättungsparameter ist  $k = 22$ . Abbildung 6 illustriert den LWR-Vorhersage-Prozess für die Varianz am 21. April 1993. Die Punkte sind die 22 Nachbarn, mit denen die lokalen Parameter berechnet werden. Man sieht, dass die Punkte eine lineare Kurve mit positiver Drift implizieren. Man sieht zugleich, dass negative Prognosen nicht grundsätzlich ausgeschlossen sind. Lügen etwa die Punkte in der rechten oberen Ecke weiter oben, so würde die Regressionsgerade steiler verlaufen 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.



**Abbildung 6:** LWR Prognose der quadrierten Volatilität für den 21. April 1993 mit  $k=22$  benachbarten und zeitlich vorhergehenden Beobachtungen.



Alle Berechnungen werden in XploRe durchgeführt, Härdle, Klinke und Müller (2000), Härdle und Tschernig (2000). Abbildung 7 zeigt die 78 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 22 lokalen Nachbarn aus den 180 zeitlich vorausgehenden Beobachtungen geschätzt. Das AR(1) Modell wird jeweils für alle 180 vorausgehende Beobachtungen angepasst. Die Prognosefehler des LWR Modells und des AR(1) Modells werden in der Abbildung 8 gezeigt.

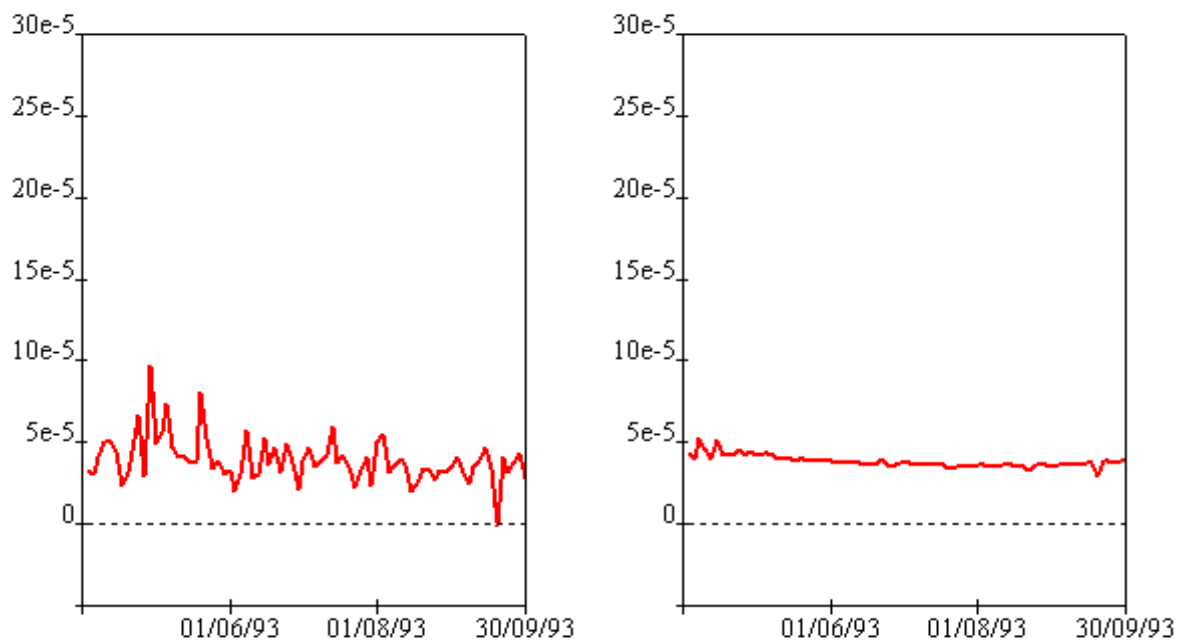
Tabelle 2 zeigt Statistiken für die durchschnittlichen Prognosefehler, wobei der mittlere quadratische

$$MQPF = \frac{1}{H} \sum_{h=1}^H (\hat{Y}_{T+h} - Y_{T+h})^2$$

und der mittlere absolute

$$MAPF = \frac{1}{H} \sum_{h=1}^H |\hat{Y}_{T+h} - Y_{T+h}|$$

angegeben wird. Die Prognosen eines Verfahrens sind desto besser, je geringer diese Fehler sind. Tabelle 2 zeigt die Prognosefehler der AR(1) und LWR Modelle. Das nichtparametrische Modell hat einen leicht größeren Prognosefehler im Durchschnitt 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, Herwatz und Spokoiny (2003). Zudem ist die nichtparametrische LWR für die gewählte Anwendung hilfreich, da sie offenbart, dass ein parametrisches Modell für eine statistische Modellierung ausreichend ist.

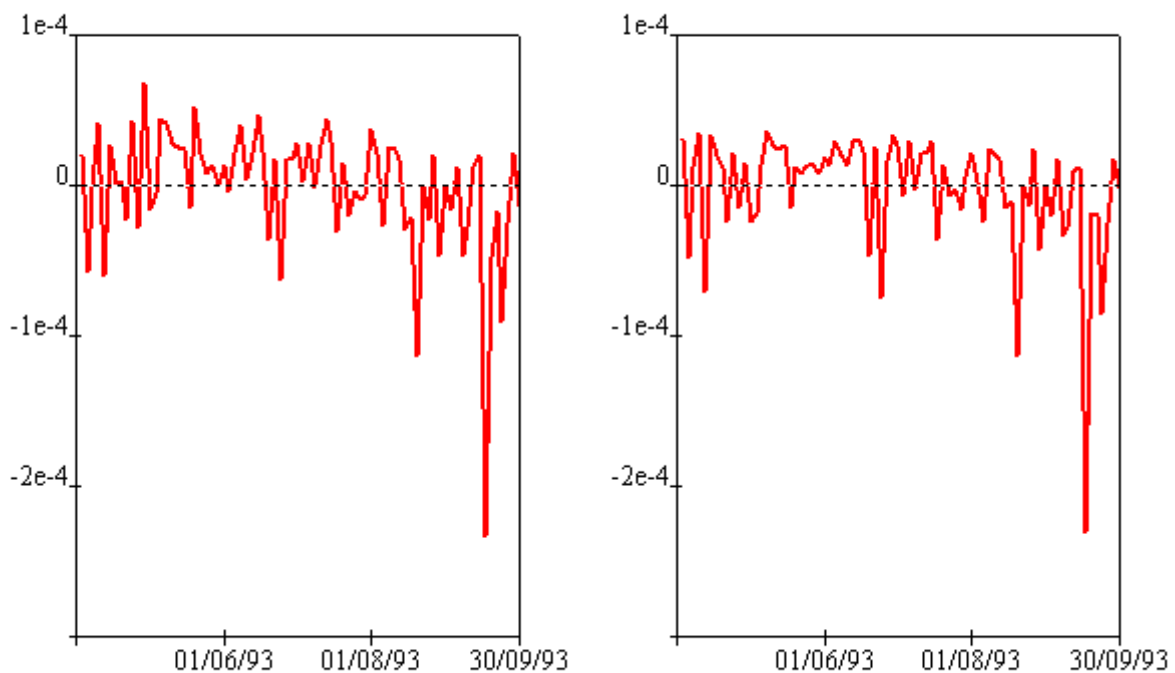



**Abbildung 7:**  $H = 78$  lokale gewichtete (links) und AR (1) (rechts) Ein-Schritt-Prognosen, wobei  $k = 22$  lokale vorhergehenden Beobachtungen verwendet wurden.



	MQPF ( $10^{-9}$ )	MAPF ( $10^{-5}$ )	Rolling Window Size
AR (1)	1,54	2,57	180
LWR( $k = 22$ )	1,76	2,85	180

**Tabelle 2:** Die Ergebnisse der Vorhersagen mittlere MQPF und MAPF.



**Abbildung 8:** Prognosefehler des LWR Modells (links) und AR(1) Modells (rechts)  
mit  $T = 180$  und  $k = 22$ . 

**Danksagung:** Für finanzielle Unterstützung bedanken wir uns bei der Deutschen Forschungsgemeinschaft, SFB 373 *Quantifikation und Simulation Ökonomischer Prozesse*.

## Literatur

Brandt, M.W. und Diebold, F.X. (2003), A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations, *Journal of Business*, forthcoming.

Cleveland, W.S. (1979): Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* **74**, 829-836.

Cleveland, W.S. und Devlin, S.J. (1988): Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, *Journal of the American Statistical Association* **83**, 596-610.

- Diebold, F.X. und Nason, J.A. (1990): Nonparametric Exchange Rate Prediction?, *Journal of International Economics* **28**, 315-332.
- Fan, J. und Gijbels, I. (1996): *Local Polynomial Modeling and its Applications*. London: Chapman and Hall.
- Härdle, W. (1990): *Applied Nonparametric Regression*. Econometric Society Monographs No. 19. Cambridge: Cambridge University Press.
- Härdle, W., Herwartz, H. und Spokoiny, V. (2003): Time Inhomogeneous Multiple Volatility Modeling, *Journal of Financial Econometrics* **1**, 55-95.
- Härdle, W. und Tschernig, R. (2000): Flexible Time Series Analysis. In Härdle, W., Hlávka, Z. und Klinke, S. (Hrsg.): *Xplore – Application Guide*. Berlin: Springer, 397-457.
- Härdle, W., Klinke, S. und Müller, M., Hrsg. (2000): *XploRe – Learning Guide*. Berlin: Springer.
- Härdle, W., Lütkepohl, H. und Chen, R. (1997): A Review of Nonparametric Time Series Analysis, *International Statistical Review* **65**, 49-72.
- Härdle, W., Müller, M., Sperlich, S. und Werwatz, A. (2004): *Introduction to Non- and Semiparametric Modeling*. Berlin: Springer Verlag.
- Hastie, T., Tibshirani, R. und Friedman, J. (2001): *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer Series in Statistics. Berlin: Springer Verlag.
- Heiler, S. (2001): 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: John Wiley, 308-347.
- Johnston, J. (1984): *Econometric Methods*. Dritte Auflage. New York: McGraw-Hill.
- Parkinson, M. (1980): The Extreme Value Method for Estimating the Variance of the Rate of Return, *Journal of Business* **53**, 61-65.
- Poon, S.-H. und Granger, C.W.J. (2003): Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature* **41**, 478-539.