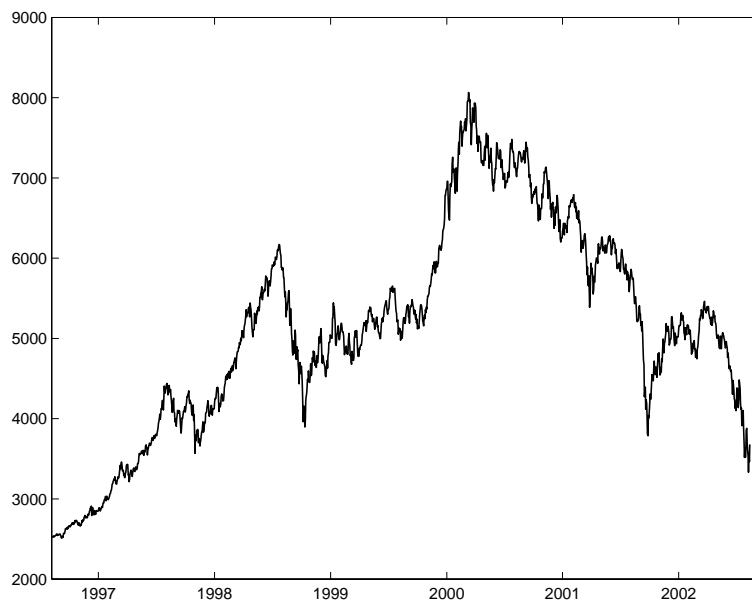


Diplomarbeit

# NICHTPARAMETRISCHE TREND- UND VOLATILITÄTSSCHÄTZER FÜR FINANZZEITREIHEN

Daniel Kopf  
5. Februar 2003



Betreuer: Prof. Dr. Jürgen Franke  
Fachbereich Mathematik  
Universität Kaiserslautern

# Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die auf unterschiedlichste Art und Weise zum Gelingen dieser Diplomarbeit beigetragen haben. Mein ganz besonderer Dank gilt Prof. Dr. Jürgen Franke, der schon früh mein Interesse an diesem Thema geweckt hat und mir immer mit viel Geduld und Engagement hilfreich zur Seite stand. Des Weiteren möchte ich meinem Bruder Andreas und all meinen Freunden danken, deren Ratschläge und Tipps mir immer sehr viel bedeutet haben.

Abschließend möchte ich mich noch ganz besonders bei meinen Eltern für die einzigartige Unterstützung bedanken.

# Erklärung

Hiermit versichere ich, dass ich diese Diplomarbeit selbständig und nur unter Verwendung der im Literaturverzeichnis angegebenen Hilfsmittel angefertigt habe.

Kaiserslautern, den 5. Februar 2003

Daniel Kopf

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Grundlagen der Regression</b>	<b>3</b>
2.1	Nichtparametrische Regression . . . . .	3
2.1.1	Kernschätzer für $m(x)$ . . . . .	3
2.1.2	Glättungsparameter $h$ . . . . .	5
2.1.3	Whitening by windowing principle . . . . .	6
2.2	Heteroskedastische Regression . . . . .	7
2.2.1	Glättungsparameter $\tilde{h}$ . . . . .	9
2.3	Nichtparametrische AR(1)-ARCH(1)-Modelle . . . . .	9
2.4	Kleinste-Quadrate-Schätzer . . . . .	11
2.4.1	KQ-Schätzer bei konstanter Varianz . . . . .	11
2.4.2	Gewichtete KQ-Schätzer . . . . .	13
<b>3</b>	<b>Theorie der gewichteten Kernschätzer</b>	<b>15</b>
3.1	Gewichtete Kernschätzer . . . . .	15
3.2	Asymptotisches Verhalten . . . . .	17
3.3	Algorithmus . . . . .	21
<b>4</b>	<b>Simulationsstudie mit Regressionsdaten</b>	<b>23</b>
4.1	Das Modell . . . . .	23
4.2	Nadaraya-Watson-Schätzer . . . . .	24
4.3	Gewichteter Kernschätzer . . . . .	25
4.4	Vergleich der beiden Schätzer . . . . .	27
4.5	Tabelle der Schätzfehler . . . . .	28
4.6	Varianzproblematik . . . . .	29
4.7	Fazit . . . . .	29
<b>5</b>	<b>Simulationsstudie mit Zeitreihendaten</b>	<b>31</b>
5.1	Das Modell . . . . .	31
5.2	Nadaraya-Watson-Schätzer . . . . .	32
5.3	Gewichteter Kernschätzer . . . . .	33
5.4	Vergleich der beiden Schätzer . . . . .	34
5.5	Tabelle der Schätzfehler . . . . .	34
5.6	Fazit . . . . .	35

<b>6</b>	<b>Betrachtung anhand realer Daten</b>	<b>36</b>
6.1	DAX-Kurse . . . . .	36
6.1.1	Die Daten . . . . .	36
6.1.2	Gewichtete Kernschätzer . . . . .	37
6.2	Dow Jones-Kurse . . . . .	38
6.2.1	Die Daten . . . . .	38
6.2.2	Gewichtete Kernschätzer . . . . .	39
<b>7</b>	<b>Resümee und Ausblick</b>	<b>40</b>

# Kapitel 1

## Einleitung

Schon seit Beginn des 20. Jahrhunderts beschäftigen sich viele Menschen mit den Aktienmärkten. Dabei bleibt es nicht aus, dass man versucht, mathematische Methoden anzuwenden, um die zukünftigen Aktienkurse vorherzusagen. Die ersten Schritte in dieser sogenannten Finanzzeitreihenanalyse wurden in den dreißiger Jahren von Working, Cowles und Jones getan. Obwohl sie sehr bald feststellten, dass eine Vorhersage der Kurse nicht möglich ist, blieb dieses ein Hauptthema der Forschung auf dem Gebiet der Finanzmärkte.

Lange Zeit beschäftigten sich die Forscher mit der Theorie der Irrfahrten, bis Box und Jenkins 1976 auf die Idee kamen, ein allgemeineres Modell, das sogenannte ARIMA-Modell (autoregressive integrated moving average model) zu verwenden. Darauf basierend entwickelte Engle 1982 das Modell des ARCH-Prozesses (autoregressive conditional heteroscedastic process), aus dem dann später der allgemeinere GARCH-Prozess entstand. Dieser kommt auch heutzutage noch in vielen Bereichen der Finanzmathematik zur Anwendung.

Jedoch haben die oben genannten Modelle den Nachteil, dass sie parametrisch sind und man somit schon zu Beginn eine strukturelle Annahme über das Modell treffen muss. Um dies zu vermeiden, kann man ein allgemeines nichtparametrisches Modell der Form

$$X_t = m(X_{t-1}, \dots, X_{t-p}) + \sigma(X_{t-1}, \dots, X_{t-q})\varepsilon_t \quad t = 1, 2, \dots$$

verwenden, wobei  $\varepsilon_t$  unabhängig identisch verteilte Zufallsvariablen mit Erwartungswert 1 und Varianz 0 sind. Die Berechnung des bedingten Erwartungswertes  $m(\underline{x}) = E[X_j | (X_{j-1}, \dots, X_{j-p}) = \underline{x}]$  und der bedingten Varianz  $\sigma^2(\underline{x}) = Var[X_j | (X_{j-1}, \dots, X_{j-q}) = \underline{x}]$  kann auf unterschiedlichste Arten erfolgen, wobei an dieser Stelle auf das Buch von Härdle und Tsybakov (1997) sowie die Arbeit von Franke, Kreiss und Mammen (1997) verwiesen wird. Eine andere Möglichkeit der Berechnung bieten die von Nadaraya und Watson schon 1964 entwickelten nichtparametrischen Kernschätzer.

Ein neuer Ansatz, der in dieser Arbeit untersucht wird, stellt die Verwendung von gewichteten Kernschätzern dar, wobei ich mich auf das obige Modell mit  $p = q = 1$  beschränke. Diese basieren auf den von Nadaraya und Watson entwickelten Kernschätzern, wobei eine Gewichtung in Anlehnung an die gewichteten Kleinste-Quadrate-Schätzer erfolgt.

In Kapitel 2 werden zunächst einmal die Grundlagen der nichtparametrischen Regression sowie die der Kleinste-Quadrate-Schätzer und die Entstehung des obigen Modells behan-

delt. Kapitel 3 führt dann in die Theorie der gewichteten Kernschätzer ein, wobei auch das asymptotische Verhalten genauer untersucht wird. Des Weiteren wird ein numerischer Algorithmus zur Berechnung der Kernschätzer angegeben.

Die Simulationsstudie der gewichteten Kernschätzer anhand von Regressionsdaten und Zeitreihendaten sowie die praktische Beurteilung erfolgen in Kapitel 4 und 5. Reale Zeitreihendaten bilden danach im sechsten Kapitel die Grundlage für die praktische Betrachtung der neuen Schätzer.

Im letzten Kapitel folgt dann ein Resümee und ein kleiner Ausblick auf die gewichteten Kernschätzer für allgemeinere Modelle.

# Kapitel 2

## Grundlagen der Regression

### 2.1 Nichtparametrische Regression

In diesem Abschnitt werden einige grundlegende Begriffe aus der Theorie der nichtparametrischen Regression eingeführt und im Speziellen die Methode der Kernschätzer erklärt. Die Problemstellung der Regression geht hierbei von Datenpaaren  $\{(X_j, Y_j)\}_{j=1}^N$  aus, die folgende Modellbeziehung erfüllen sollen.

**Modell 1:**

$$\begin{aligned} Y_j &= m(X_j) + \varepsilon_j & j &= 1, \dots, N \\ X_1, X_2, \dots, X_N && \text{u.i.v.} \\ \varepsilon_1, \varepsilon_2, \dots, \varepsilon_N && \text{u.i.v. mit } E[\varepsilon_j] = 0 \end{aligned}$$

Hierbei ist speziell die Funktion  $m(x)$  von Interesse. Im Gegensatz zur parametrischen Regression, bei der man eine subjektive Annahme über die Struktur der Funktion  $m(x)$  treffen muss, ist dies bei der nichtparametrischen Regression, die als eine lokale Durchschnittsbildung verstanden werden kann, nicht nötig.

#### 2.1.1 Kernschätzer für $m(x)$

Würde man einen allgemeinen Durchschnitt über alle Funktionswerte bilden, so erhielte man die waagrechte Ausgleichsgerade  $y = \frac{1}{N}Y_N$ . Die Bildung eines lokalen Durchschnitts um den Funktionswert  $x$  erreicht man dagegen, indem man die  $Y_j$  entsprechend der Entfernung von  $X_j$  zu  $x$  gewichtet. Hierbei erscheint zur Gewichtung die Verwendung eines Kerns  $K$  sinnvoll. Unter einem Kern versteht man eine stetige, begrenzte und symmetrische Funktion  $K$  mit  $\int K(u)du = 1$ . Zur besseren Steuerung der Gewichtung verwendet man einen mit dem Glättungsfaktor  $h$  skalierten Kern  $K_h$  mit

$$K_h(u) = \frac{1}{h} \int K\left(\frac{u}{h}\right)du . \quad (2.1)$$

Die Wahl des richtigen Kerns fällt schwer, da es viele unterschiedliche Möglichkeiten gibt. Der, für numerische Zwecke am Häufigsten verwendete, ist der Epanechnikov-Kern

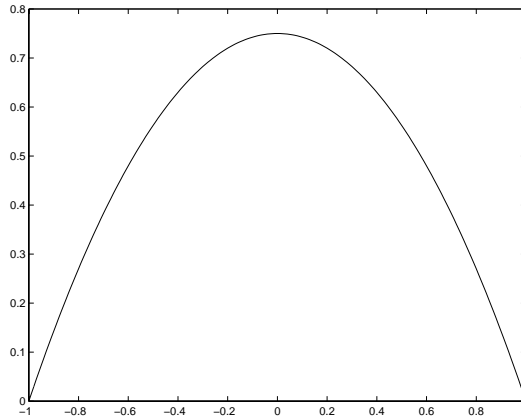


Abbildung 2.1: Der Epanechnikov Kern

[Epa69], da er einen kompakten Träger hat [Här93, Kap 4.5]. Er ist definiert über :

$$K(u) = \frac{3}{4} (1 - u^2) 1_{[-1,1]}(u) .$$

Um der Anforderung gerecht zu werden, dass sich die Gewichte zu eins aufsummieren, muss man sie noch durch die Dichte  $p$  der  $X_j$  dividieren. Da diese Dichte nicht explizit gegeben ist, verwendet man hierfür den Rosenblatt-Parzen-Schätzer  $\hat{p}(x, h)$  [Par62] [Ros56]

$$\hat{p}(x, h) = \frac{1}{N} \sum_{j=1}^N K_h(x - X_j) . \quad (2.2)$$

Der sich hieraus ergebende Schätzer für  $m(x)$  wurde 1964 von Nadaraya und Watson entwickelt und wird deswegen auch Nadaraya-Watson-Schätzer genannt [Nad64] [Wat64].

$$\hat{m}(x, h) = \frac{\sum_{j=1}^N K_h(x - X_j) Y_j}{\sum_{j=1}^N K_h(x - X_j)} \quad (2.3)$$

**Satz 1.** Sei  $K(u) \geq 0$  ,  $\lim_{|u| \rightarrow \infty} (u \cdot K(u)) = 0$  und  $E[Y_j^2] < \infty$ . Dann gilt für  $N \rightarrow \infty$  und  $h \rightarrow 0$  mit  $Nh \rightarrow \infty$

$$(a) \quad \hat{p}(x, h) \xrightarrow{p} p(x)$$

$$(b) \quad \hat{m}(x, h) \xrightarrow{p} m(x)$$

für alle  $x$ , an denen  $m(x)$  und  $p(x)$  stetig sind mit  $p(x) > 0$ .

Auf den Beweis des Satzes wird an dieser Stelle verzichtet und auf [Här93] verwiesen.



### 2.1.2 Glättungsparameter $h$

Wie man bisher gesehen hat, hängt der Nadaraya-Watson-Schätzer stark vom Glättungsparameter  $h$  ab. Aufgrund der Definition des Kerns ist leicht ersichtlich, dass für große  $h$  der lokale Durchschnitt an der Stelle  $x$  über viele Werte  $Y_j$  gebildet wird. Es leuchtet ein, dass die Kurve der Funktion  $\hat{m}(x, h)$  demnach für große  $h$  glatter ist als für kleine. Die Frage der nichtparametrischen Regression ist also auch eine Frage wie glatt die geschätzte Funktion sein soll. Da dies aber auch davon abhängt, wozu man den Schätzer verwenden will, ist diese Entscheidung von Fall zu Fall unterschiedlich zu treffen.

Trotzdem ist es möglich, eine Bandbreite  $h$  zu bestimmen, die im Allgemeinfall gute Dienste leistet. Hierzu geht man davon aus, dass der Parameter  $h$  den Average Squared Error, gegeben durch

$$ase(\hat{m})(h) = d_A(h) = \frac{1}{N} \sum_{j=1}^N (\hat{m}(X_j, h) - m(X_j))^2 \quad (2.4)$$

minimieren soll. Dies führt auf die optimale Bandbreite

$$\hat{h} = \arg \min_h \left[ \frac{1}{N} \sum_{j=1}^N (Y_j - \hat{m}(X_j, h))^2 \right],$$

die ohne Probleme numerisch berechnet werden kann. Das Problem an dieser Stelle ist, dass  $\hat{h}$  nicht erwartungstreu (engl. biased) ist. Zurückzuführen ist dies auf die Tatsache, dass zur Berechnung von  $\hat{m}(X_i, h)$  bereits der Wert  $Y_i$  verwendet wurde, was dazu führt, dass  $\hat{m}(X_i, h)$  näher an  $Y_i$  liegt als  $m(X_i)$ .

Um dem entgegen zu wirken, benutzt man die Methode der Kreuz-Validierung (engl. cross-validation), bei der man das Paar  $(X_i, Y_i)$  bei der Berechnung des Rosenblatt-Parzen- und des Nadaraya-Watson-Schätzers unberücksichtigt lässt. Damit hängt  $\hat{m}_{-i}(x, h)$  nicht mehr von  $(X_i, Y_i)$  ab und es ergibt sich

$$\hat{m}_{-i}(x, h) = \frac{1}{(N-1) \cdot \hat{p}_{-i}(x, h)} \sum_{\substack{j=1 \\ j \neq i}}^N K_h(x - X_j) Y_j \quad (2.5)$$

$$\hat{p}_{-i}(x, h) = \frac{1}{(N-1)} \sum_{\substack{j=1 \\ j \neq i}}^N K_h(x - X_j). \quad (2.6)$$

Diese mit Hilfe der Kreuz-Validierung erhaltene Schätzer sind erwartungstreu und werden nun dazu benutzt, den Average Squared Error (2.4) zu minimieren. Daraus ergibt sich dann folgende optimale Bandbreite  $\hat{h}_{CV}$ :

$$\begin{aligned} \hat{h}_{CV} &= \arg \min_h CV(h) \\ &= \arg \min_h \left[ \frac{1}{N} \sum_{j=1}^N (Y_j - \hat{m}_{-j}(X_j, h))^2 \right]. \end{aligned} \quad (2.7)$$

Härdle und Marron haben bewiesen, dass  $\hat{h}_{CV}$  asymptotisch optimal ist, das heißt für

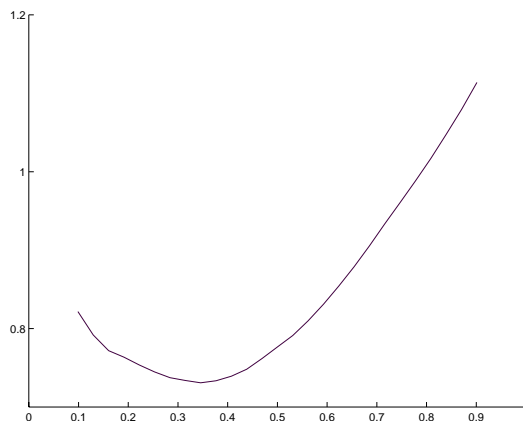


Abbildung 2.2: Eine Beispielfunktion  $CV(h)$

$N \rightarrow \infty$  und

$$H_N = \left[ \frac{1}{C \cdot N^{1-\delta}}, \frac{1}{C \cdot N^\delta} \right]$$

mit Konstanten  $C, \delta \in (0, 0.5)$  gilt:

$$\frac{d_A(\hat{h}_{CV})}{\inf_{h \in H_N} d_A(h)} \xrightarrow{a.s.} 1.$$

Auf den genauen Wortlaut des Satzes und den Beweis wird aufgrund der vielen technischen Voraussetzungen an dieser Stelle verzichtet. Er kann im Buch von Härdle [Här93] nachgelesen werden.

### 2.1.3 Whitening by windowing principle

Bisher wurden unabhängig identisch verteilte  $X_1, \dots, X_N$  betrachtet, was jedoch in der Praxis nur selten der Fall ist. Häufig ist eine Beobachtung  $Y_j$  in irgendeiner Weise von den vorhergehenden abhängig. Mit dieser Thematik hat sich J. D. Hart 1996 beschäftigt und herausgefunden, dass die in den vorherigen Abschnitten für unabhängige Beobachtungen genannten Schätztechniken auch für nicht unabhängige anwendbar sind. Diese Tatsache wird 'Whitening by windowing principle' genannt [Har96].

Veranschaulichen kann man dies am Beispiel eines Epanechnikov-Kerns, da hierbei aufgrund des kompakten Trägers  $[-h, h]$  nur die Werte in dem Intervall  $[x - h, x + h]$  von Bedeutung sind und keineswegs alle beobachteten Daten.

Bildet man an den beiden Stellen  $x_1$  und  $x_2$  beispielsweise die Kernschätzer, so sind für  $\hat{m}(x_1)$  nur die Wertepaare in  $M_1 = \{(X_i, Y_i) \mid X_i \in [x_1 - h, x_1 + h]\}$  von Interesse. Gleiches gilt für  $\hat{m}(x_2)$  und  $M_2 = \{(X_i, Y_i) \mid X_i \in [x_2 - h, x_2 + h]\}$ . Diese beiden Mengen  $M_1$  und  $M_2$  sind nun für beliebig kleines  $h$  disjunkt, und somit kann eine Abhängigkeit der Wertepaare untereinander vernachlässigt werden. Da der Parameter  $h$  in einer gewissen Relation zu dem Stichprobenumfang  $N$  steht, kann man weiter feststellen, dass die bisher für unabhängige Daten besprochenen Techniken für große  $N$  auch bei abhängigen Daten

verwendet werden können.

Eine typische Anwendung findet dieses Prinzip bei Zeitreihenanalysen, im Speziellen bei der später zu betrachtenden Analyse von Finanzzeitreihen. Hierfür werden neben der Stationarität der Finanzzeitreihe noch weitere Mischungs-Bedingungen benötigt, die an dieser Stelle aber nur erwähnt und aufgrund der Komplexität weggelassen werden.

## 2.2 Heteroskedastische Regression

In dem bisherigen Modell ging man davon aus, dass die  $\varepsilon_1, \dots, \varepsilon_N$  unter anderem identisch verteilt sind, was bedeutet, dass sie die gleiche Varianz haben. Nun soll ein Modell betrachtet werden, das eine Varianzveränderung zulässt.

**Modell 2:**

$$\begin{aligned} Y_j &= m(X_j) + v_j & j &= 1, \dots, N \\ X_1, X_2, \dots, X_N &\text{ u.i.v.} \\ v_1, v_2, \dots, v_N &\text{ mit } E[v_j] = 0 \end{aligned}$$

Geht man nun davon aus, dass die Varianz ebenso wie auch der Wert  $Y_j$  von  $X_j$  abhängt, kann man das Modell so umformulieren, dass man neben  $m(X_j)$  eine weitere Funktion  $\sigma(X_j)$  mit  $v_j = \sigma(X_j)\varepsilon_j$  hat, die die Varianzveränderung angibt. Um diese Varianzveränderung eindeutig angeben zu können, benötigt man als weitere Bedingung die Tatsache, dass die Varianz der  $\varepsilon_j$  gleich 1 ist.

**Modell 2':**

$$\begin{aligned} Y_j &= m(X_j) + \sigma(X_j)\varepsilon_j & j &= 1, \dots, N \\ X_1, X_2, \dots, X_N &\text{ u.i.v.} \\ \varepsilon_1, \varepsilon_2, \dots, \varepsilon_N &\text{ u.i.v. mit } E[\varepsilon_j] = 0, \text{Var}[\varepsilon_j] = 1 \end{aligned}$$

Als Schätzer für  $m(x)$  kann natürlich weiterhin der Nadaraya-Watson-Schätzer (2.3) verwendet werden, da sich nur die Varianz ändert, was aber für die Bildung lokaler Durchschnitte nicht von Interesse ist, solange  $E[v_j] = 0$  gilt und man annehmen kann, dass die Varianz  $\sigma(x)$  in einer kleinen Umgebung um  $x$  ungefähr konstant ist.

Für die Schätzung der Funktion  $\hat{\sigma}(x)$  gibt es zwei verschiedene Möglichkeiten, die im Folgenden kurz vorgestellt werden.

Die erste Idee basiert darauf, dass  $\sigma(X_j)$  im Modell 2' in anderer Form dargestellt werden kann. Aufgrund der Varianzannahme der  $\varepsilon_j$  gilt für die Varianz der  $Y_j$

$$\begin{aligned} \text{Var}[Y_j|X_j = x] &= \text{Var}[m(X_j) + \sigma(X_j)\varepsilon_j|X_j = x] \\ &= \text{Var}[m(x)] + \sigma^2(x) \cdot \text{Var}[\varepsilon_j] \\ &= \sigma^2(x) \end{aligned}$$

und somit kann man  $\sigma^2(x)$  schreiben als

$$\sigma^2(x) = \text{Var}[Y_j|X_j = x] = \underbrace{E[Y_j^2|X_j = x]}_{=: q(x)} - \underbrace{[E[Y_j|X_j = x]]^2}_{= m^2(x)}.$$

Da für die Funktion  $m(x)$  schon ein zuverlässiger Schätzer  $\hat{m}(x)$  existiert, fehlt zur Bestimmung von  $\hat{\sigma}(x)$  nur noch ein Schätzer für  $q(x)$ . Diesen kann man in Anlehnung an den Nadaraya-Watson-Schätzer (2.3) bestimmen durch

$$\hat{q}(x, \tilde{h}) = \frac{\sum_{j=1}^N K_{\tilde{h}}(x - X_j) Y_j^2}{\sum_{j=1}^N K_{\tilde{h}}(x - X_j)}. \quad (2.8)$$

Somit ergibt sich der Schätzer für die Varianz  $\hat{\sigma}^2(x, h, \tilde{h})$  durch

$$\hat{\sigma}^2(x, h, \tilde{h}) = \hat{q}(x, \tilde{h}) - \hat{m}^2(x, h). \quad (2.9)$$

Der zweiten Idee liegt das Modell 2 zu Grunde. Da

$$\begin{aligned} E[v_j^2 | X_j = x] &= E[\sigma^2(X_j) \varepsilon_j^2 | X_j = x] \\ &= \sigma^2(x) \cdot E[\varepsilon_j^2] \\ &= \sigma^2(x) \end{aligned}$$

gilt, kann man bei bekannten  $v_j$  die Varianz  $\sigma^2(x)$  wie folgt schätzen:

$$\hat{\sigma}^2(x, \tilde{h}) = \frac{\sum_{j=1}^N K_{\tilde{h}}(x - X_j) v_j^2}{\sum_{j=1}^N K_{\tilde{h}}(x - X_j)}.$$

Da in unserem Fall die Werte  $v_j$  nicht bekannt sind, bleibt nichts anderes übrig, als auch sie zu schätzen. Dies geschieht mittels des bekannten Schätzers  $\hat{m}(x)$  und der Gleichung aus dem Modell 2:

$$\hat{v}_j(h) = Y_j - \hat{m}(X_j, h). \quad (2.10)$$

Damit ergibt sich für den Schätzer der Varianz die Formel

$$\hat{\sigma}^2(x, h, \tilde{h}) = \frac{\sum_{j=1}^N K_{\tilde{h}}(x - X_j) \hat{v}_j^2(h)}{\sum_{j=1}^N K_{\tilde{h}}(x - X_j)}. \quad (2.11)$$

Im Folgenden wird der zweite Varianzschätzer (2.11) verwendet, da er gegenüber dem ersten entscheidende Vorteile aufweist. So kann es im ersten Modell im Speziellen bei Verwendung von verschiedenen Glättungsparametern  $h, \tilde{h}$  für  $\hat{m}$  und  $\hat{\sigma}$  aufgrund der Differenz in (2.9) dazu kommen, dass der Varianzschätzer negativ wird, was zu Schwierigkeiten in der numerischen Berechnung führen würde. Dies wird aufgrund der Berechnung in (2.11) vermieden. Fan und Yao haben sich 1998 dem Schätzer (2.11) angenommen und gezeigt, dass er für  $N \rightarrow \infty$  genauso gut ist, wie wenn die Funktion  $m(x)$  in (2.10) explizit bekannt wäre [FY98]. Des Weiteren haben die beiden betrachteten Schätzer zwar die selbe asymptotische Varianz, jedoch hat ersterer im Vergleich zum zweiten einen zusätzlichen Term in der Formel des Bias. Dies kann einen nachteiligen Effekt auf die Qualität des Schätzers haben.

### 2.2.1 Glättungsparameter $\tilde{h}$

Für den Volatilitätsschätzer benötigt man nun ebenfalls einen Glättungsparameter  $\tilde{h}$ . Es wäre zwar möglich, den gleichen Wert zu nehmen wie für die Schätzung der Trendfunktion, allerdings ist es offensichtlich, dass es sich hierbei um andere Datenpaare handelt. Deswegen muss man, um ein möglichst gutes Ergebnis zu erhalten, einen anderen Glättungsparameter wählen.

Das Prinzip entspricht dem aus Kapitel 2.1.2. Man versucht den Average Squared Error zu minimieren. Dieser ist nun gegeben durch:

$$ase(\hat{\sigma})(\tilde{h}) = d_B(\tilde{h}) = \frac{1}{N} \sum_{j=1}^N (\hat{\sigma}^2(X_j, \tilde{h}) - v_j^2)^2. \quad (2.12)$$

Da jedoch die  $v_j$  nicht explizit gegeben sind, muss man diese durch Verwendung des Schätzers  $\hat{m}(x)$  annähern. Es gilt:

$$\begin{aligned} \hat{v}_j^2 &= (Y_j - \hat{m}(X_j))^2 \\ &= Y_j^2 - 2Y_j\hat{m}(X_j) + \hat{m}^2(X_j). \end{aligned}$$

Hierbei verwendet man genauso wie schon bei dem Glättungsparameter  $h$  der Trendfunktion das Prinzip der Kreuz-Validierung. Dieses führt nun zu folgendem optimalen  $\tilde{h}_{CV}$ :

$$\begin{aligned} \tilde{h}_{CV} &= \arg \min_{\tilde{h}} d_B(\tilde{h}) \\ &= \arg \min_{\tilde{h}} \left[ \frac{1}{N} \sum_{j=1}^N (\hat{\sigma}_{-j}^2(X_j, \tilde{h}) - Y_j^2 + 2 \cdot Y_j \hat{m}_{-j}(X_j) - \hat{m}_{-j}^2(X_j))^2 \right]. \end{aligned} \quad (2.13)$$

## 2.3 Nichtparametrische AR(1)-ARCH(1)-Modelle

In den letzten Jahren wurden viele parametrische Modelle für die Modellierung von Finanzzeitreihen entwickelt, jedoch hatten sie alle das Problem, dass die Funktionsscharen fest vorgegeben waren und man nur die Parameter schätzen konnte. Mit Hilfe der nichtparametrischen Regression kann man Funktionen schätzen, ohne ihren Grad oder den genauen Aufbau zu kennen. Nichtsdestotrotz benötigt man eine Annahme über das Modell, das man schätzen möchte. Eines der wohl bekanntesten Modelle für die Bildung von Aktienpreisen, das schon Black und Scholes für ihre Optionspreisformel verwendeten, geht davon aus, dass die Preise  $S_t$  lognormal-verteilt sind [KK99]. Sie sind damit die Lösungen der stochastischen Differentialgleichung

$$dS_t = b(t)S_t dt + \sigma(t)S_t dW_t \quad (2.14)$$

wobei  $W_t$  eine eindimensionale Brownsche Bewegung ist. Die Funktion  $b(t)$  kann hierbei als mittlere Ertragsrate und  $\sigma(t)$  als Volatilitätsfunktion des Preisprozesses zum Zeitpunkt  $t$  interpretiert werden.

Geht man davon aus, dass die Ertragsrate und die Volatilität konstant sind, so kann man die Veränderung  $\Delta S$  im Aktienpreis über ein Zeitintervall  $\Delta t$  wie folgt beschreiben:

$$\begin{aligned}\frac{\Delta S}{S} &= b\Delta t + \sigma W_{\Delta t} \\ &= b\Delta t + \sigma\varepsilon\sqrt{\Delta t} .\end{aligned}$$

Für  $\varepsilon \sim \mathcal{N}(0, 1)$  folgt die letzte Gleichheit aufgrund der Stationaritätseigenschaft einer Brownschen Bewegung, woraus folgt, dass  $W_{\Delta t} \sim \mathcal{N}(0, \Delta t)$  verteilt ist.

Damit kann man nun eine diskrete Version der Aktienpreisformel aufstellen.

$$\begin{aligned}S_{t+\Delta t} &= S_t + S_t \frac{\Delta S}{S} \\ S_{t+1} &= S_t + bS_t + \sigma S_t \varepsilon_{t+1}\end{aligned}\tag{2.15}$$

Will man nun die Theorie der Kernschätzer auf die Aktienpreise anwenden, so wird man feststellen, dass die benötigte identische Verteilung der  $S_t$  nicht gegeben ist. Um dies zu erreichen, werden anstelle der Aktienpreise die sogenannten Returns  $R_t$  verwendet. Sie sind definiert über

$$R_{t+1} = \frac{S_{t+1} - S_t}{S_t}\tag{2.16}$$

und entsprechen der prozentualen Veränderung des Kurses zum Vortag. Es ist klar ersichtlich, dass diese nun eine identische Verteilung haben.

Die diskrete Black-Scholes-Formel für Aktienpreise lässt sich nun wie folgt umformulieren:

$$\begin{aligned}S_{t+1} &= S_t + bS_t + \sigma S_t \varepsilon_{t+1} \\ S_{t+1} - S_t &= bS_t + \sigma S_t \varepsilon_{t+1} \\ \frac{S_{t+1} - S_t}{S_t} &= b + \sigma \varepsilon_{t+1} \\ R_{t+1} &= b + \sigma \varepsilon_{t+1} .\end{aligned}$$

Entgegen der obigen Annahme, dass  $b$  und  $\sigma$  Konstanten sind, ist dies in der Praxis selten der Fall. Häufiger ist es vielmehr so, dass der Wert und die Volatilität des Returns  $R_{t+1}$  von dem letzten Return  $R_t$  abhängen. Daraus ergibt sich dann die diskrete Black-Scholes-Formel für die Returns

$$R_{t+1} = b(R_t) + \sigma(R_t)\varepsilon_{t+1} .\tag{2.17}$$

Für  $\sigma(R_t) \equiv 1$  entspricht das Modell (2.17) einem nichtlinearen autoregressiven Prozess der Ordnung 1 (kurz: AR(1)-Prozess) und für  $b(R_t) \equiv 0$  mit  $\sigma^2(R_t) = \omega + \alpha S_t^2$  einem autoregressiven bedingten heteroskedastischen Prozess der Ordnung 1 (kurz: ARCH(1)-Prozess).

Deswegen spricht man bei dem obigen Modell auch von einem AR(1)-ARCH(1)-Prozess, von einigen Autoren auch CHARN(1)-Prozess genannt.

Es fällt auf, dass die Funktion (2.17) für  $X_t = R_t$ ,  $Y_t = R_{t+1}$  und  $b(x) \equiv m(x)$  der aus unserem Modell 2' gleicht. Somit kann man, werden die Bedingungen des Modells 2' erfüllt, die Theorie der Kernschätzer auf die Finanzzeitreihe  $R_t$  anwenden.

Bis auf die Unabhängigkeit der  $R_t$  untereinander werden alle Bedingungen erfüllt, und

diese kann mit Hilfe des "Whitening by windowing principle" von J. D. Hart aus Kapitel 2.1.3 angenommen werden. Somit lassen sich die bisher erhaltenen Kenntnisse über Kernschätzer auch auf AR(1)-ARCH(1)-Finanzzeitreihen anwenden.

## 2.4 Kleinste-Quadrate-Schätzer

Die Methode der Kleinsten-Quadrate-Schätzer, die im Folgenden vorgestellt wird, stammt aus der parametrischen Regression. Betrachtet man zum Beispiel einmal die Größe  $Y_j$  und das Gewicht  $X_j$  von beliebigen Personen, so stellt man fest, dass das Gewicht bei gegebener Größe stark schwanken kann und umgekehrt. Trotzdem ist offensichtlich, dass bei zunehmender Größe auch das durchschnittliche Gewicht größer wird. Man könnte nun, um das erwartete Gewicht bei gegebener Größe zu schätzen, eine Regressionsgerade der Form

$$Y_j = a + bX_j + Z_j \quad (2.18)$$

durch die beobachteten Werte legen, wobei  $Z_j$  die individuelle Abweichung widerspiegelt. Eine solche mögliche Regressionsgerade ist in Abb. 2.3 einmal dargestellt.

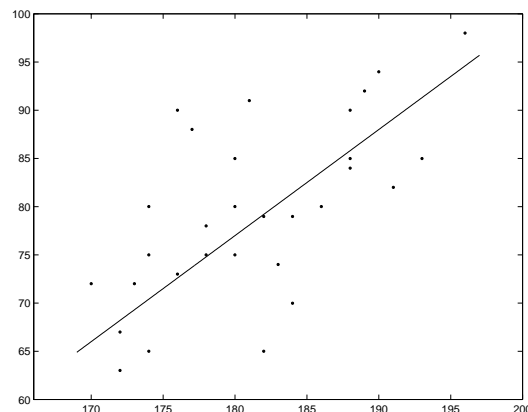


Abbildung 2.3: Größe und Gewichte von 40 Männern

Bei gegebenen Daten-Paaren  $(X_j, Y_j)$  kann man nun mit Hilfe der Methode der Kleinsten-Quadrate die Parameter  $a$  und  $b$  im Modell (2.18) schätzen. Hierbei wird die Gerade derart gewählt, dass die Summe der quadrierten Differenzen zwischen den tatsächlichen Gewichten und der Regressionsgeraden minimiert wird. Diese Technik wird im nächsten Abschnitt kurz vorgestellt, bevor dann auf die Besonderheit der Gewichtet-Kleinsten-Quadrate-Schätzer eingegangen wird.

### 2.4.1 KQ-Schätzer bei konstanter Varianz

Die Methode der Kleinste-Quadrate-Schätzer wurde unabhängig voneinander von Carl-Friedrich Gauss (1777-1855) und Adrien Marie Legendre (1752-1833) entwickelt. Hierbei

wird davon ausgegangen, dass die Wertepaare  $(X_j, Y_j)$  folgendes Modell erfüllen:

**Modell R1:**

$$Y_j = a + bX_j + Z_j \quad j = 1, \dots, N$$

$$E[Z_j] = 0 \text{ und } Var[Z_j] = \sigma_Z^2$$

Wie oben erwähnt, werden die Schätzer nun so gewählt, dass die Summe der quadrierten Residuen möglichst klein wird. Demnach fordert man nun für gegebene Paare  $(X_j, Y_j)$ ,  $j = 1, \dots, N$ , dass die Kleinste-Quadrate-Schätzer für  $a$  und  $b$  die Funktion

$$D(a, b) = \sum_{j=1}^N (Y_j - a - bX_j)^2 \quad (2.19)$$

minimieren. Damit müssen die Schätzer  $\hat{a}$  und  $\hat{b}$  folgende Gleichungen erfüllen:

$$\frac{\partial D}{\partial a}(\hat{a}, \hat{b}) = -2 \sum_{j=1}^N (Y_j - \hat{a} - \hat{b}X_j) = 0$$

$$\frac{\partial D}{\partial b}(\hat{a}, \hat{b}) = -2 \sum_{j=1}^N X_j (Y_j - \hat{a} - \hat{b}X_j) = 0 .$$

Löst man nun dieses Gleichungssystem, so erhalten wir die Kleinste-Quadrate-Schätzer  $\hat{a}$  und  $\hat{b}$ . Sie sind gegeben durch

$$\hat{b} = \frac{\sum_{j=1}^N (X_j - \bar{X}_N) Y_j}{\sum_{j=1}^N (X_j - \bar{X}_N)^2} \quad (2.20)$$

$$\hat{a} = \bar{Y}_N - \hat{b} \bar{X}_N \quad (2.21)$$

mit  $\bar{X}_N = \frac{1}{N} \sum_{j=1}^N X_j$  und  $\bar{Y}_N = \frac{1}{N} \sum_{j=1}^N Y_j$ . Somit ergibt sich als Regressionsfunktion die Gerade  $\hat{g}(x) = \hat{a} + \hat{b}x$ . Es bleibt zu erwähnen, dass das Paar  $(\bar{X}, \bar{Y})$  immer auf dieser Regressionsgerade liegt.

Eine etwas einfachere Möglichkeit der Darstellung bietet die Matrixschreibweise. Das Modell hat dann die Form

$$Y = X\beta + Z \quad (2.22)$$

mit

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix} ; \quad X = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_N \end{pmatrix}$$

$$\beta = \begin{pmatrix} a \\ b \end{pmatrix} ; \quad Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_N \end{pmatrix} .$$

Der KQ-Schätzer  $\hat{\beta} = (\hat{a} \ \hat{b})^T$ , der die Funktion  $D(a, b)$  (2.19) minimiert, ist nun gegeben durch

$$\hat{\beta} = (X^T X)^{-1} X^T Y . \quad (2.23)$$



Er ist eine lineare Funktion der Beobachtungen  $Y_1, \dots, Y_N$  und falls das Modell R1 korrekt ist, so ist er auch erwartungstreu, das heißt es gilt

$$E[\hat{a}] = a \quad \text{und} \quad E[\hat{b}] = b .$$

Eine Anwendung des Gauß-Markov-Theorems besagt, dass der Kleinste-Quadrate-Schätzer unter allen erwartungstreuen Schätzern die geringste Varianz aufweist [Bre73] [Wei80]. Dies alles gilt allerdings nur, wenn die Modellbedingungen erfüllt sind. Im Speziellen muss die Varianz der  $Z_j$  konstant sein. Ist dies nicht der Fall, so ist es sinnvoll die im Folgenden beschriebenen gewichteten Kleinste-Quadrate-Schätzer zu verwenden.

## 2.4.2 Gewichtete KQ-Schätzer

Ändert sich die Varianz der  $Z_j$  und somit auch die Varianz der  $Y_j$ , so lässt sich die gewöhnliche Methode der KQ-Schätzer nicht anwenden, da zu beachten ist, dass die Aussagekraft derjenigen Wertepaare  $(X_j, Y_j)$  mit einer großen Varianz geringer ist als solche mit einer kleinen Varianz. Voraussetzen muss man hierbei allerdings, dass die einzelnen Faktoren  $\sigma_j$  der Varianzveränderungen bekannt sind, wobei man dann

$$Z_j = \sigma_j \varepsilon_j \quad \text{mit} \quad \text{Var}[\varepsilon_j] = \sigma_\varepsilon^2$$

hat. Der Spezialfall, dass alle Varianzen  $\sigma_j$  explizit bekannt wären, würde dazu führen, dass  $\text{Var}[\varepsilon_j] = 1$  ist.

**Modell R2:**

$$\begin{aligned} Y_j &= a + bX_j + \sigma_j \varepsilon_j & j &= 1, \dots, N \\ & & \sigma_j & \text{bekannt} \\ & & E[\varepsilon_j] &= 0 \quad \text{und} \quad \text{Var}[\varepsilon_j] = \sigma_\varepsilon^2 \end{aligned}$$

Um für dieses Modell nun Schätzer zu finden, wird die Funktion zuerst einmal umformuliert. Dividiert man jeden Summanden der Modellfunktion durch  $\sigma_j$ , so lässt sie sich äquivalent schreiben als

$$\frac{Y_j}{\sigma_j} = a \frac{1}{\sigma_j} + b \frac{X_j}{\sigma_j} + \varepsilon_j . \quad (2.24)$$

Dieses Modell entspricht nun den Anforderungen des Modells R1 und somit kann man hier die KQ-Schätzer anwenden. Die zu minimierende Summe der quadrierten Residuen  $D(a, b)$  ist nun gegeben durch

$$\begin{aligned} D(a, b) &= \sum_{j=1}^N \left( \frac{Y_j}{\sigma_j} - a \frac{1}{\sigma_j} - b \frac{X_j}{\sigma_j} \right)^2 \\ &= \sum_{j=1}^N \frac{1}{\sigma_j^2} (Y_j - a - bX_j)^2 . \end{aligned} \quad (2.25)$$

Diese Darstellung verdeutlicht auch, warum man hierbei von gewichteten KQ-Schätzern spricht, denn die quadrierten Residuen werden mit der inversen Varianz gewichtet. Residuen, die aus einer größeren Varianz  $\sigma_j$  resultieren, sind demnach für die Berechnung des

zu minimierenden Fehlers weniger wichtig als andere.  
Der gewichtete KQ-Schätzer  $\hat{\beta}$  berechnet sich nun über

$$\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y \quad (2.26)$$

mit

$$\begin{aligned} \Sigma &= \text{Diag}(\sigma_1^2, \dots, \sigma_N^2) \\ &= \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{pmatrix}. \end{aligned}$$

# Kapitel 3

## Theorie der gewichteten Kernschätzer

In diesem Kapitel soll nun die Idee der Gewichtung der Beobachtungen entsprechend ihrer Volatilität in die Theorie der heteroskedastischen Kernschätzer eingebaut werden. Dazu wird zuerst die Theorie der gewichteten Kernschätzer eingeführt, bevor dann ihr asymptotisches Verhalten genauer untersucht und ein möglicher Algorithmus zur Berechnung gegeben wird.

### 3.1 Gewichtete Kernschätzer

Bei den Kleinste-Quadrate-Schätzern hat man gesehen, dass eine Verwendung von Gewichten durchaus Sinn macht, da Beobachtungen, die eine höhere Varianz aufweisen, nicht so aussagekräftig sind wie andere mit einer niedrigeren Varianz. Dieses Prinzip der Gewichtung soll nun auch auf den, in Kapitel 2 beschriebenen, nichtparametrischen Kernschätzer  $\hat{m}(x, h)$  angewendet werden.

Geht man zunächst von der Tatsache aus, dass die Varianzen  $\sigma_j^2$  bekannt sind, dann sind die gewichteten Kleinste-Quadrate-Schätzer diejenigen Schätzer, die die gewichtete Fehlerfunktion

$$D(a, b) = \sum_{j=1}^N \frac{1}{\sigma_j^2} (Y_j - a + bX_j)^2$$

minimieren.

Analog wird nun die Fehlerfunktion des Kernschätzers ebenfalls mit dem Faktor  $\frac{1}{\sigma_j^2}$  gewichtet. Dazu wird zuerst gezeigt, dass der gewöhnliche Kernschätzer von Nadaraya und Watson über die Fehlerfunktion

$$D(\mu) = \sum_{j=1}^N (Y_j - \mu)^2 K_h(x - X_j) \tag{3.1}$$

gegeben ist. Aufgrund der geforderten Minimierungseigenschaft muss der Schätzer nun die Gleichung

$$\frac{\partial D}{\partial \mu}(\hat{\mu}) = -2 \sum_{j=1}^N (Y_j - \hat{\mu}) K_h(x - X_j) = 0$$

erfüllen und somit folgt durch Umformung

$$\hat{\mu} = \frac{\sum_{j=1}^N K_h(x - X_j) Y_j}{\sum_{j=1}^N K_h(x - X_j)} = \hat{m}(x, h). \quad (3.2)$$

Damit hat man gezeigt, dass der Kernschätzer aus Kapitel 2 die Fehlerfunktion (3.1) minimiert. Entsprechend der gewichteten Fehlerfunktion der Kleinste-Quadrate-Schätzer lautet die gewichtete Fehlerfunktion der Kernschätzer  $\tilde{D}(\mu)$  somit

$$\tilde{D}(\mu) = \sum_{j=1}^N \frac{1}{\sigma^2(X_j)} (Y_j - \mu)^2 K_h(x - X_j). \quad (3.3)$$

Den gewichteten Kernschätzer erhält man, indem man die Ableitung der Fehlerfunktion gleich null setzt. Es gilt:

$$\begin{aligned} \frac{\partial \tilde{D}}{\partial \mu}(\hat{\mu}) &= \sum_{j=1}^N \frac{-2}{\sigma^2(X_j)} (Y_j - \hat{\mu}) K_h(x - X_j) = 0 \\ \Leftrightarrow \hat{\mu} \cdot \sum_{j=1}^N \frac{1}{\sigma^2(X_j)} K_h(x - X_j) &= \sum_{j=1}^N \frac{1}{\sigma^2(X_j)} Y_j K_h(x - X_j) \end{aligned}$$

und somit

$$\hat{\mu} = \frac{\sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\sigma^2(X_j)}}{\sum_{j=1}^N K_h(x - X_j) \frac{1}{\sigma^2(X_j)}}$$

Damit ergibt sich für den gewichteten Kernschätzer bei bekannter Varianz  $\sigma(X_j)$  die Formel

$$\hat{m}(x, h) = \frac{\sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\sigma^2(X_j)}}{\sum_{j=1}^N K_h(x - X_j) \frac{1}{\sigma^2(X_j)}}. \quad (3.4)$$

Betrachtet man nun im Folgenden das eigentliche heteroskedastische Problem mit dem dazugehörigen Modell 2'. Hierbei sind die Varianzen  $\sigma(X_j)$  nun nicht mehr explizit gegeben, sondern müssen ebenfalls geschätzt werden.

Die gewichteten Kernschätzer  $\hat{m}(x, h, \tilde{h})$  und  $\hat{\sigma}^2(x, h, \tilde{h})$  sind nun in Analogie zu den heteroskedastischen Schätzern aus Kapitel 2 gegeben durch

$$\hat{m}(x, h, \tilde{h}) = \frac{\sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\hat{\sigma}^2(X_j, h, \tilde{h})}}{\sum_{j=1}^N K_h(x - X_j) \frac{1}{\hat{\sigma}^2(X_j, h, \tilde{h})}} \quad (3.5)$$

und

$$\hat{\sigma}^2(x, h, \tilde{h}) = \frac{\sum_{j=1}^N K_{\tilde{h}}(x - X_j) (Y_j - \hat{m}(X_j, h, \tilde{h}))^2}{\sum_{j=1}^N K_{\tilde{h}}(x - X_j)}. \quad (3.6)$$

Wie man sieht, hängt der Schätzer  $\hat{m}(x, h, \tilde{h})$  von  $\hat{\sigma}(x, h, \tilde{h})$  und umgekehrt ab. Deswegen ist die Berechnung der beiden Schätzer nicht explizit möglich. Man kann nur durch Verwendung von numerischen Verfahren versuchen, die Lösungen anzunähern. Hierzu wird später ein iterativer Algorithmus vorgestellt.

Zuvor jedoch wird im nächsten Abschnitt das asymptotische Verhalten der neuen Schätzer betrachtet.

## 3.2 Asymptotisches Verhalten

Analog zum ersten Satz in Kapitel 2, kann man nun folgende Aussage über das Konvergenzverhalten des gewichteten Kernschätzers bei bekannten Varianzen  $\sigma(X_j)$  treffen.

**Satz 2.** Seien  $\sigma(X_j)$  für alle  $j = 1, \dots, N$  gegeben und sei  $K(u) \geq 0$ ,  $\lim_{|u| \rightarrow \infty} (u \cdot K(u)) = 0$  und  $E[Y_j^2] < \infty$ . Dann gilt für  $N \rightarrow \infty$  und  $h \rightarrow 0$  mit  $Nh \rightarrow \infty$

$$\hat{m}(x, h) = \frac{\sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\sigma^2(X_j)}}{\sum_{j=1}^N K_h(x - X_j) \frac{1}{\sigma^2(X_j)}} \xrightarrow{p} m(x) \quad (3.7)$$

für alle  $x$ , an denen  $m(x)$  stetig ist.

### Beweis:

Der Beweis des Satzes geschieht, indem zuerst einmal die folgenden vier Konvergenzen bewiesen werden, wobei  $m(x) = \frac{r(x)}{q(x)}$  und  $q(x) = \frac{p(x)}{\sigma^2(x)}$  gilt.

$$(1) \text{ bias}(\hat{q}(x, h)) := \text{bias}\left(\frac{1}{N} \sum_{j=1}^N K_h(x - X_j) \frac{1}{\sigma^2(X_j)}\right) \longrightarrow 0$$

$$(2) \text{ var}(\hat{q}(x, h)) := \text{var}\left(\frac{1}{N} \sum_{j=1}^N K_h(x - X_j) \frac{1}{\sigma^2(X_j)}\right) \longrightarrow 0$$

$$(3) \text{ bias}(\hat{r}(x, h)) := \text{bias}\left(\frac{1}{N} \sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\sigma^2(X_j)}\right) \longrightarrow 0$$

$$(4) \text{ var}(\hat{r}(x, h)) := \text{var}\left(\frac{1}{N} \sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\sigma^2(X_j)}\right) \longrightarrow 0$$

Gelten diese vier Aussagen, so gilt wegen (1)+(2)

$$\text{mse}(\hat{q}(x, h)) = E(\hat{q}(x, h) - q(x))^2 = \text{var}(\hat{q}(x, h)) + [\text{bias}(\hat{q}(x, h))]^2 \longrightarrow 0$$

und somit gilt

$$\hat{q}(x, h) \xrightarrow{p} q(x). \quad (3.8)$$

Aufgrund der Aussagen (3)+(4) gilt außerdem

$$\text{mse}(\hat{r}(x, h)) = \text{var}(\hat{r}(x, h)) + [\text{bias}(\hat{r}(x, h))]^2 \longrightarrow 0$$

und somit

$$\hat{r}(x, h) \xrightarrow{p} r(x). \quad (3.9)$$

Aus diesen beiden Konvergenzen in Wahrscheinlichkeit folgt nun

$$\hat{m}(x, h) = \frac{\hat{r}(x, h)}{\hat{q}(x, h)} \xrightarrow{p} m(x) = \frac{r(x)}{q(x)} \quad (3.10)$$

und somit der Beweis des Satzes.

Im Folgenden werden nun die Aussagen (1), (2), (3) und (4) gezeigt:

zu (1):

Da die Gleichung  $\text{bias}(\hat{q}(x, h)) = E(\hat{q}(x, h) - q(x))$  gilt, kann man nun folgende Betrachtung anstellen.

$$\begin{aligned} & |E(\hat{q}(x, h) - q(x))| \\ &= \left| \int K_h(x-u) \frac{p(u)}{\sigma^2(u)} du - \int K_h(u) q(x) du \right| \\ &= \left| \int K_h(u) \frac{p(x-u)}{\sigma^2(x-u)} du - \int K_h(u) \frac{p(x)}{\sigma^2(x)} du \right| \\ &= \left| \int K_h(u) \left[ \frac{p(x-u)}{\sigma^2(x-u)} - \frac{p(x)}{\sigma^2(x)} \right] du \right| \\ &\leq \left| \int_{|u| \geq \delta} u K_h(u) \frac{p(x-u)}{u \sigma^2(x-u)} du \right| + \frac{p(x)}{\sigma^2(x)} \int_{|u| \geq \delta} K_h(u) du + \\ &\quad \sup_{|u| < \delta} \left| \frac{p(x-u)}{\sigma^2(x-u)} - \frac{p(x)}{\sigma^2(x)} \right| \int_{|u| < \delta} K_h(u) du \\ &\leq \sup_{|u| \geq \delta} |u K_h(u)| \int_{|u| \geq \delta} \frac{p(x-u)}{u \sigma^2(x-u)} du + \frac{p(x)}{\sigma^2(x)} \int_{|v| \geq \frac{\delta}{h}} K(v) dv + \sup_{|u| < \delta} \left| \frac{p(x-u)}{\sigma^2(x-u)} - \frac{p(x)}{\sigma^2(x)} \right| \\ &\leq \sup_{|v| \geq \frac{\delta}{h}} |v K(v)| \frac{1}{\delta} \int_{|u| \geq \delta} \frac{p(x-u)}{\sigma^2(x-u)} du + \frac{p(x)}{\sigma^2(x)} \int_{|v| \geq \frac{\delta}{h}} K(v) dv + \sup_{|u| < \delta} \left| \frac{p(x-u)}{\sigma^2(x-u)} - \frac{p(x)}{\sigma^2(x)} \right| \end{aligned}$$

$\delta$  kann nun so klein gewählt werden, dass der dritte Term beliebig klein gemacht werden kann. Der zweite Term geht für  $h \rightarrow 0$  ebenfalls gegen 0, da damit  $\frac{\delta}{h}$  gegen unendlich geht. Das Integral im ersten Term ist für gegebenes  $x$  beschränkt, was dazu führt, dass auch der erste Term wegen  $h \rightarrow 0$  gegen 0 konvergiert. Deswegen geht der gesamte *bias* gegen 0 und somit gilt die Aussage (1).

zu (2):

$$\begin{aligned}
& \text{var}(\hat{q}(x, h)) \\
&= \frac{1}{N} \text{var}\left(K_h(x - X_1) \frac{1}{\sigma^2(X_1)}\right) \\
&= \frac{1}{N} E\left(K_h^2(x - X_1) \frac{1}{\sigma^4(X_1)}\right) - \frac{1}{N} \left[ E\left(K_h(x - X_1) \frac{1}{\sigma^2(X_1)}\right) \right]^2 \\
&= \frac{1}{N} \int K_h^2(x - u) \frac{p(u)}{\sigma^4(u)} du - \frac{1}{N} \left[ \int K_h(x - u) \frac{p(u)}{\sigma^2(u)} du \right]^2 \\
&= \frac{1}{Nh} \int \frac{1}{h} K^2\left(\frac{x - u}{h}\right) \frac{p(u)}{\sigma^4(u)} du - \frac{1}{N} \left[ \int K_h(x - u) \frac{p(u)}{\sigma^2(u)} du \right]^2 \\
&= \frac{1}{Nh} \cdot \frac{p(x)}{\sigma^4(x)} \int K^2(v) dv - \frac{1}{N} \left[ \int K_h(x - u) \frac{p(u)}{\sigma^2(u)} du \right]^2 \\
&\leq \frac{1}{N^\delta} \cdot \frac{p(x)}{\sigma^4(x)} \int K^2(v) dv - \frac{1}{N} \left[ \int K_h(x - u) \frac{p(u)}{\sigma^2(u)} du \right]^2
\end{aligned}$$

Die letzte Ungleichung folgt aus der Tatsache, dass  $h$  im Intervall  $[N^{-(1-\delta)}, N^{-\delta}]$  liegt, wobei  $\delta$  beliebig aber fest gewählt werden kann. Nach (1) gilt

$$\left[ \int K_h(x - u) \frac{p(u)}{\sigma^2(u)} du \right]^2 \xrightarrow{N \rightarrow \infty} q^2(x) .$$

Für  $N \rightarrow \infty$  gehen nun beide Terme gegen 0, da im Speziellen  $\frac{1}{N^\delta}$  und  $\frac{1}{N}$  gegen 0 konvergieren.

zu (3):

$$\begin{aligned}
& |E(\hat{r}(x, h) - r(x))| \\
&= |E(\hat{r}(x, h)) - r(x)| \\
&= \left| \frac{1}{N} \sum_{j=1}^N E\left(K_h(x - X_j) \frac{Y_j}{\sigma^2(X_j)}\right) - r(x) \right| \\
&= \left| E\left(K_h(x - X_1) \frac{Y_1}{\sigma^2(X_1)} \mid X_1\right) - r(x) \right| \\
&= \left| E\left(K_h(x - X_1) \frac{E(Y_1 \mid X_1)}{\sigma^2(X_1)}\right) - r(x) \right| \\
&= \left| E\left(K_h(x - X_1) \frac{m(X_1)}{\sigma^2(X_1)}\right) - r(x) \right| \\
&= \left| \int K_h(x - u) \frac{m(u)}{\sigma^2(u)} p(u) du - \int K_h(u) \frac{m(x)}{\sigma^2(x)} p(x) du \right| \\
&= \left| \int K_h(u) \frac{m(x - u)}{\sigma^2(x - u)} p(x - u) du - \int K_h(u) \frac{m(x)}{\sigma^2(x)} p(x) du \right| \\
&= \left| \int K_h(u) \left[ \frac{m(x - u)}{\sigma^2(x - u)} p(x - u) - \frac{m(x)}{\sigma^2(x)} p(x) \right] du \right|
\end{aligned}$$

Es kann nun analog zu dem Beweis der Aussage (1) gezeigt werden, dass dieser Term gegen 0 strebt. Daraus folgt wiederum, dass der Bias von  $\hat{r}(x, h)$  gegeben durch

$$\text{bias}(\hat{r}(x, h)) = E(\hat{r}(x, h) - r(x))$$

gegen 0 konvergiert.

zu (4):

$$\begin{aligned} \text{var}(\hat{r}(x, h)) &= \text{var}\left(\frac{1}{N} \sum_{j=1}^N K_h(x - X_j) \frac{Y_j}{\sigma^2(x_j)}\right) \\ &= \frac{1}{N} E\left(K_h^2(x - X_1) \frac{m^2(X_1)}{\sigma^4(X_1)}\right) - \frac{1}{N} \left[E\left(K_h(x - X_1) \frac{m(X_1)}{\sigma^2(X_1)}\right)\right]^2 \\ &= \frac{1}{N} \int K_h^2(x - u) \frac{m^2(u)}{\sigma^4(u)} p(u) du - \frac{1}{N} \left[\int K_h(x - u) \frac{m(u)}{\sigma^2(u)} p(u) du\right]^2 \\ &\leq \frac{1}{N^\delta} \cdot \frac{m^2(x)}{\sigma^4(x)} p(x) \int K^2(v) dv - \frac{1}{N} \left[\int K_h(x - u) \frac{m(u)}{\sigma^2(u)} p(u) du\right]^2 \end{aligned}$$

Die letzte Ungleichung folgt wiederum analog zu der im Beweis von (2). Für den zweiten Term gilt

$$\left[\int K_h(x - u) \frac{m(u)}{\sigma^2(u)} p(u) du\right]^2 \xrightarrow{N \rightarrow \infty} r^2(x).$$

Deswegen konvergieren die beiden Terme für  $N \rightarrow \infty$  gegen 0 und somit geht auch die Varianz von  $\hat{r}(x, h)$  gegen 0.

Damit sind nun alle 4 Aussagen bewiesen und es folgt somit der Beweis des Satzes.  $\square$

Der Beweis kann auch auf einfache Weise veranschaulicht werden. Denn aufgrund der Tatsache, dass  $N$  gegen unendlich strebt, geht die Bandbreite  $h$  gegen 0. Nun gehen aber die Varianzen  $\sigma^2(X_j)$  für  $X_j \in (x - h, x + h)$  bei kleinem  $h$  gegen den gleichen Wert, und somit konvergiert der gewichtete Kernschätzer gegen den Nadaraya-Watson-Schätzer. Der Beweis folgt damit aus Satz 1.

Mit Hilfe dieses Satzes weiß man somit, dass die gewichteten Kernschätzer bei gegebener Varianzfunktion  $\sigma^2(x)$  als Schätzer verwendet werden können.

Für die heteroskedastische Regression und somit den Schätzer (3.5) gilt nun das selbe asymptotische Verhalten, da der Varianzschätzer (3.6) bekanntermaßen gegen  $\sigma^2(x)$  konvergiert.



### 3.3 Algorithmus

Wie man gesehen hat, lassen sich die Schätzer (3.5) und (3.6) nicht explizit berechnen, sondern müssen über numerische Verfahren angenähert werden. An dieser Stelle soll nun ein Algorithmus vorgestellt werden, mit dessen Hilfe man die Schätzer näherungsweise bestimmen kann. Dieser Algorithmus basiert vom Prinzip her auf einem iterativen Verfahren.

---

*Algorithmus*

---

**(Schritt 1)** Abschnitt der äußeren 10% der Daten aus numerischen Gründen

Setze  $(\hat{\sigma}_{-i}^{(0)2}(X_j))_{i,j \in \{1, \dots, N\}} = 1$

Setze  $(\hat{\sigma}^{(0)2}(X_j))_{j \in \{1, \dots, N\}} = 1$

(Hinweis: oBdA kann zu Beginn die Varianz gleich 1 angenommen werden, da eine Konstante keine Rolle spielt.)

Berechnung des Intervalls  $H$  in dem die Bandbreite  $h$  liegt.

$$H = \left[ \frac{1}{N^{1-\delta}}, \frac{1}{N^\delta} \right]$$

Setze  $t=1$

**(Schritt 2)** Berechnung der optimalen Bandbreite  $h^{(t)}_{opt}$  über

$$h^{(t)}_{opt} = \arg \min_{h \in H} \frac{1}{N} \sum_{j=1}^N \left[ Y_j - \frac{\sum_{k=1, k \neq j}^N K_h(X_j - X_k) \frac{Y_k}{\hat{\sigma}_{-j}^{(t-1)2}(X_k)}}{\sum_{k=1, k \neq j}^N K_h(X_j - X_k) \frac{1}{\hat{\sigma}_{-j}^{(t-1)2}(X_k)}} \right]^2$$

**(Schritt 3)** Berechnung des Schätzers für  $m(x)$

$$\hat{m}^{(t)}(x) = \frac{\sum_{k=1}^N K_{h^{(t)}_{opt}}(x - X_k) \frac{Y_k}{\hat{\sigma}^{(t-1)2}(X_k)}}{\sum_{k=1}^N K_{h^{(t)}_{opt}}(x - X_k) \frac{1}{\hat{\sigma}^{(t-1)2}(X_k)}}$$

sowie von

$$\begin{aligned} \hat{m}^{(t)}(X_j) &= \frac{\sum_{k=1}^N K_{h^{(t)}_{opt}}(X_j - X_k) \frac{Y_k}{\hat{\sigma}^{(t-1)2}(X_k)}}{\sum_{k=1}^N K_{h^{(t)}_{opt}}(X_j - X_k) \frac{1}{\hat{\sigma}^{(t-1)2}(X_k)}} & \forall j = 1, \dots, N \\ \hat{m}^{(t)}_{-i}(X_j) &= \frac{\sum_{k=1, k \neq i}^N K_{h^{(t)}_{opt}}(X_j - X_k) \frac{Y_k}{\hat{\sigma}_{-i}^{(t-1)2}(X_k)}}{\sum_{k=1, k \neq i}^N K_{h^{(t)}_{opt}}(X_j - X_k) \frac{1}{\hat{\sigma}_{-i}^{(t-1)2}(X_k)}} & \forall i, j = 1, \dots, N \end{aligned}$$

(Schritt 4) Berechnung der optimalen Bandbreite  $\tilde{h}^{(t)}_{opt}$  über

$$\tilde{h}^{(t)}_{opt} = \arg \min_{\tilde{h} \in H} \frac{1}{N} \sum_{j=1}^N \left[ Y_j^2 - \hat{m}^{(t)}_{-j}(X_j) - \frac{\sum_{k=1, k \neq j}^N K_{\tilde{h}}(X_j - X_k) [Y_k - \hat{m}^{(t)}_{-j}(X_k)]^2}{\sum_{k=1, k \neq j}^N K_{\tilde{h}}(X_j - X_k)} \right]^2$$

(Schritt 5) Berechnung des Varianzschätzers  $\hat{\sigma}^2(x)$

$$\hat{\sigma}^{(t)2}(x) = \frac{\sum_{k=1}^N K_{\tilde{h}^{(t)}_{opt}}(x - X_k) [Y_k - \hat{m}^{(t)}(X_k)]^2}{\sum_{k=1}^N K_{\tilde{h}^{(t)}_{opt}}(x - X_k)}$$

sowie von

$$\hat{\sigma}^{(t)2}(X_j) = \frac{\sum_{k=1}^N K_{\tilde{h}^{(t)}_{opt}}(X_j - X_k) [Y_k - \hat{m}^{(t)}(X_k)]^2}{\sum_{k=1}^N K_{\tilde{h}^{(t)}_{opt}}(X_j - X_k)} \quad \forall j = 1, \dots, N$$

$$\hat{\sigma}^{(t)2}_{-i}(X_j) = \frac{\sum_{k=1, k \neq i}^N K_{\tilde{h}^{(t)}_{opt}}(X_j - X_k) [Y_k - \hat{m}^{(t)}_{-i}(X_k)]^2}{\sum_{k=1, k \neq i}^N K_{\tilde{h}^{(t)}_{opt}}(X_j - X_k)} \quad \forall i, j = 1, \dots, N$$

(Schritt 6) Falls das Abbruchkriterium (z.B. maximale Anzahl der Iterationen) erfüllt ist (STOP)

Ansonsten setze  $t=t+1$  und mache weiter mit (Schritt 2)

---

# Kapitel 4

## Simulationsstudie mit Regressionsdaten

In diesem Kapitel werden nun die gewichteten Kernschätzer auf erzeugte Daten angewendet, um so die Qualität des Schätzers zu überprüfen. Dazu wird zuerst das Modell definiert, dann werden im Folgenden die Nadaraya-Watson-Schätzer und die gewichteten Kernschätzer berechnet und anschließend werden diese beiden dann miteinander verglichen. Die Berechnung des Schätzers erfolgt dabei nach dem Algorithmus aus dem vorherigen Kapitel.

### 4.1 Das Modell

Das Modell, das hier näher betrachtet werden soll, ist das klassische heteroskedastische Modell, das auch schon in Kapitel 2.2 beschrieben wurde:

$$\begin{aligned} Y_j &= m(X_j) + \sigma(X_j)\varepsilon_j & j &= 1, \dots, N \\ X_1, X_2, \dots, X_N & \text{u.i.v.} \\ \varepsilon_1, \varepsilon_2, \dots, \varepsilon_N & \text{u.i.v. mit } E[\varepsilon_j] = 0, \text{Var}[\varepsilon_j] = 1 . \end{aligned}$$

Um Daten entsprechend dem obigen Modell zu erzeugen, werden zunächst die  $X_j$ ,  $j = 1, \dots, N$  gleichverteilt auf dem Intervall  $[-3,3]$  gewählt. Die Innovationen  $\varepsilon_1, \dots, \varepsilon_N$  wählt man  $\mathcal{N}(0, 1)$ -verteilt. Als Trendfunktion dient eine sogenannte "bump function", anhand derer man die Eigenschaften des Schätzers gut erkennen kann. Die Volatilitätsfunktion wird periodisch gewählt:

$$\begin{aligned} m(x) &= x + \frac{10}{\sqrt{\pi}} \cdot e^{-2x^2} \\ \sigma^2(x) &= 0.2 + \sin^2(x) . \end{aligned}$$

Anhand dieses Modells werden nun zufällige Datenpaare  $(X_j, Y_j)$  erzeugt, die im Folgenden zur Berechnung der verschiedenen Schätzer verwendet werden.

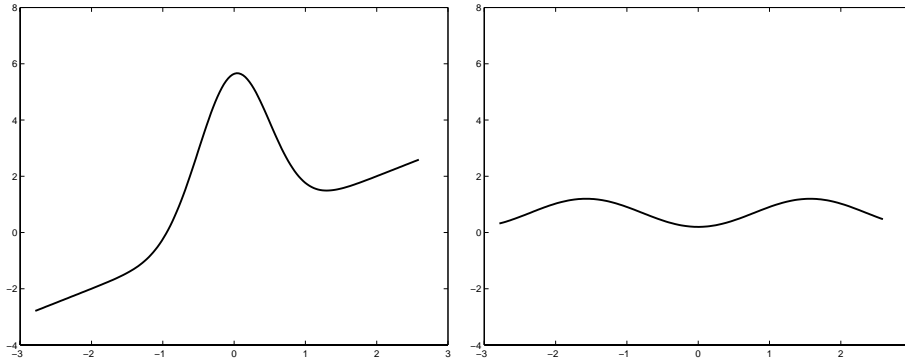


Abbildung 4.1: Die Trendfunktion  $m(x)$  und die Volatilitätsfunktion  $\sigma^2(x)$

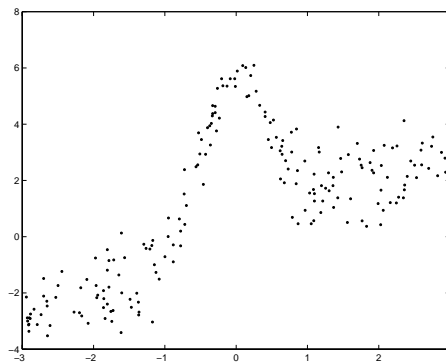


Abbildung 4.2: Nach dem Modell erzeugte Daten

## 4.2 Nadaraya-Watson-Schätzer

Berechnet man nun den Nadaraya-Watson-Schätzer (2.3), der sich als Ergebnis der ersten Iteration aus dem vorgestellten Algorithmus ergibt, so erhält man die in Abb.4.3 dargestellten Trend- und Volatilitätsschätzer:

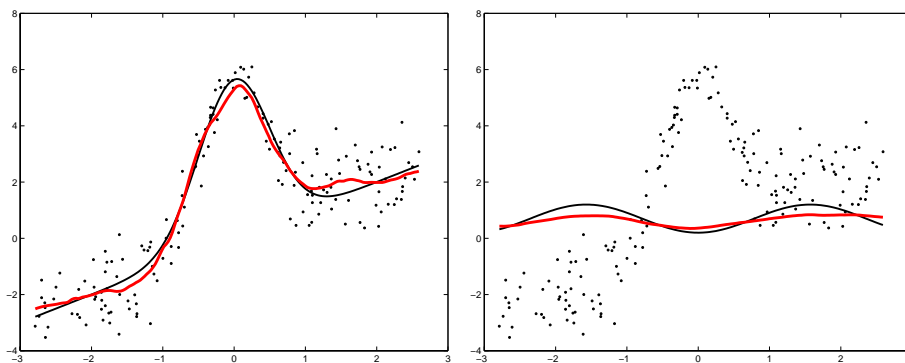


Abbildung 4.3: Der Trendschätzer  $\hat{m}(x)$  und der Volatilitätsschätzer  $\hat{\sigma}^2(x)$

Um die Qualität des Schätzers beurteilen zu können, betrachtet man das Fehlerschaubild (Abb.4.4). Es gibt für jedes  $X_j$  den Abstand zwischen dem geschätzten Funktionswert

$\hat{m}(X_j)$  und dem wahren Funktionswert  $m(X_j)$  wieder.

Ingesamt fällt bei der Betrachtung der letzten Abbildungen auf, dass der Nadaraya-Watson-Schätzer speziell in der Höckerspitze der bump-function und im Übergang zwischen dem Höcker und der Geraden Probleme hat, die wahre Trendfunktion zu treffen.

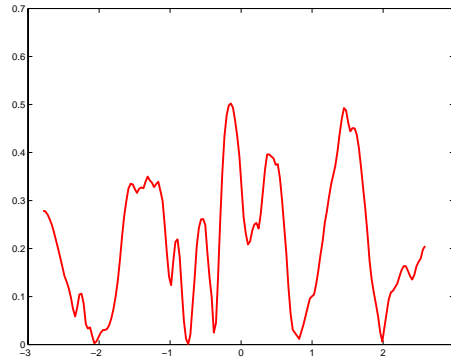


Abbildung 4.4: Fehlerschaubild des Nadaraya-Watson-Schätzers  $\hat{m}(x)$

### 4.3 Gewichteter Kernschätzer

Obwohl der iterative Algorithmus den gewichteten Kernschätzer nur näherungsweise liefert, stellt man durch Betrachtung der Abbildungen (4.6)-(4.9) fest, dass sich der Schätzer unter normalen Umständen nach wenigen Iterationen nicht mehr ändert. Insofern kann man davon ausgehen, dass bereits nach 5 Iterationen ein ausreichend zufriedenstellendes Ergebnis erreicht wird. Die Abbildungen auf der nächsten Seite zeigen die Trend- und Volatilitätsschätzer der zweiten bis fünften Iteration.

Das folgende Fehlerschaubild gibt, analog zu obigem Fehlerschaubild, die Abstände zwischen dem Schätzer der 5. Iteration und der wahren Trendfunktion wieder.

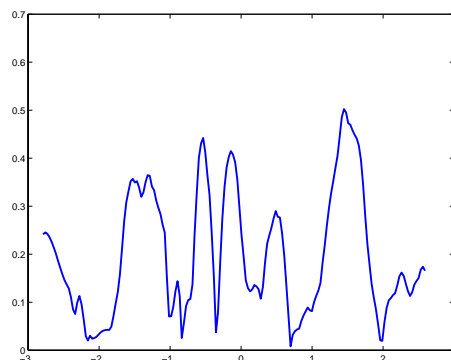


Abbildung 4.5: Fehlerschaubild der 5. Iteration

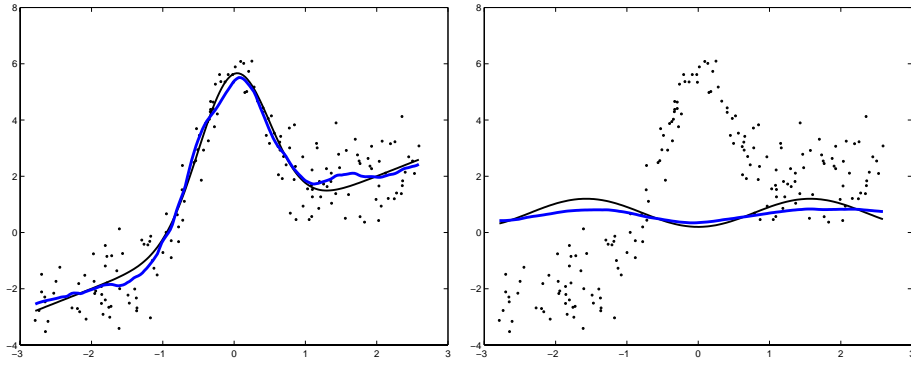


Abbildung 4.6: Gewichtete Kernschätzer der 2. Iteration

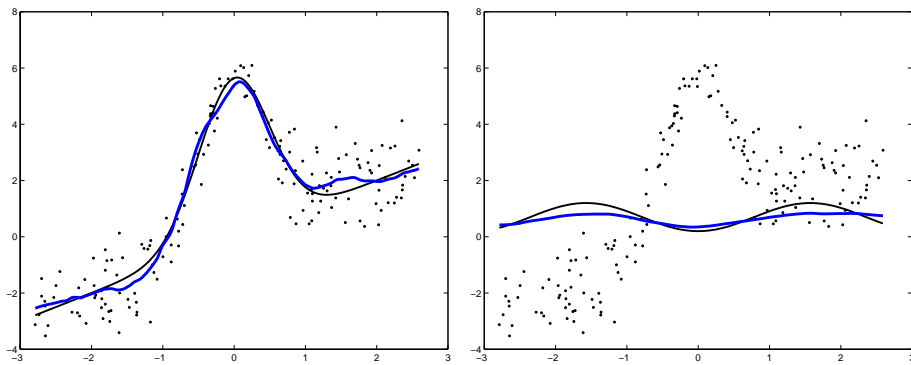


Abbildung 4.7: Gewichtete Kernschätzer der 3. Iteration

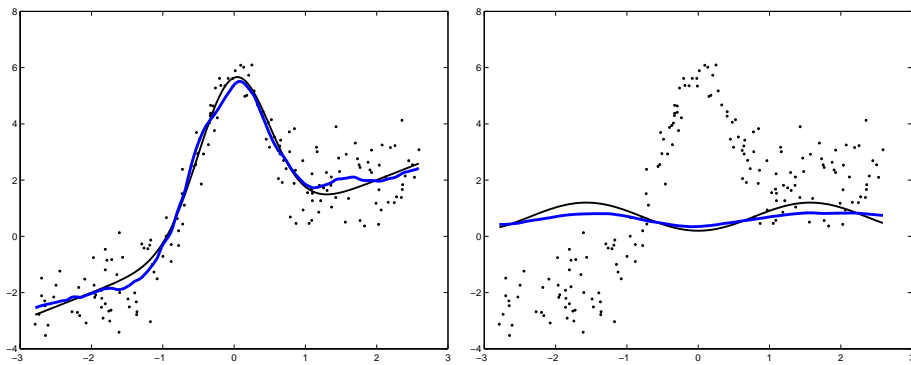


Abbildung 4.8: Gewichtete Kernschätzer der 4. Iteration

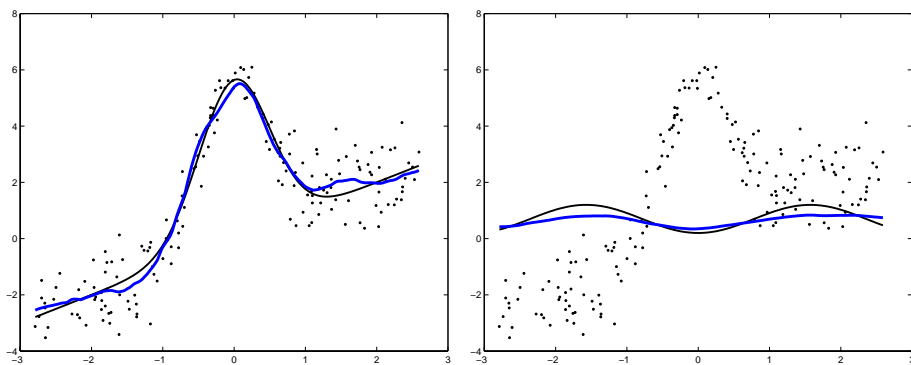


Abbildung 4.9: Gewichtete Kernschätzer der 5. Iteration

## 4.4 Vergleich der beiden Schätzer

Um die beiden Schätzer untereinander vergleichen zu können, werden zunächst die Fehler im Vergleich zueinander betrachtet. Die folgende Abbildung zeigt die beiden Fehlerschaubilder übereinandergelegt.

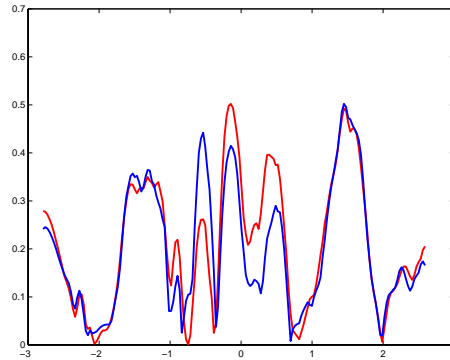


Abbildung 4.10: Fehlerschaubild der 1. (rot) und 5. Iteration (blau)

Bei der Betrachtung stellt man fest, dass der gewichtete Kernschätzer der Trendfunktion im Vergleich zum Nadaraya-Watson-Schätzer speziell in diesem Fall im Bereich des Höckers der bump-function vorteilhafter ist. Dies hat damit zu tun, dass aufgrund der Wahl der Varianzfunktion dort eine Gewichtung zugunsten des Höckers stattfindet. Andererseits sind die Seiten des Höckers dafür wegen des starken Gefälles etwas anfälliger und es kommt in diesem Beispiel zu einer Erhöhung des Fehlers. In den Bereichen der Geraden, also in den äußeren Bereichen verhält sich der gewichtete Kernschätzer ziemlich ähnlich zu dem Nadaraya-Watson-Schätzer.

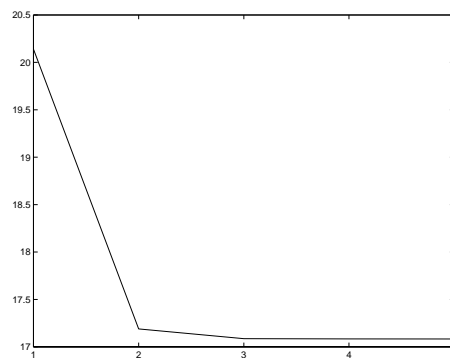


Abbildung 4.11: Gewichteter Gesamtfehler der jeweiligen Iteration

Betrachten man den gewichteten Gesamtfehler (Abb.4.11), der sich über die Formel

$$Fehler = \frac{1}{N} \sum_{j=1}^N \left( \frac{\hat{m}(X_j) - m(X_j)}{\hat{\sigma}(X_j)} \right)^2 \quad (4.1)$$

ergibt, so kann man feststellen, dass er in diesem Fall durch Verwendung des gewichteten Kernschätzers deutlich gesenkt werden konnte.

## 4.5 Tabelle der Schätzfehler

Dass die im Speziellen gezeigten Ergebnisse des vorherigen Abschnitts auch allgemein gelten, soll nun anhand einer Schätzfehler-tabelle gezeigt werden. Hierzu wurden für verschiedene Stichprobenumfänge jeweils 100 verschiedene Datenmengen anhand des Modells generiert und damit die Schätzfehler der ersten 5 Iterationen berechnet. Die folgende Tabelle zeigt die Fehler auf:

Stichproben- umfang	Nadaraya-W.- Schätzer	2. Iteration	3. Iteration	4. Iteration	5. Iteration	Verhältnis
N=50	19.8095	20.7082	20.9618	20.9166	20.9112	47 : 53
N=100	20.4509	20.6641	20.3808	20.2764	20.2669	53 : 47
N=200	20.9603	19.6157	19.4590	19.4372	19.4354	56 : 44
N=500	23.0325	22.3455	22.1998	22.2059	22.1812	59 : 41

Das Verhältnis gibt an, in wie vielen Fällen der Schätzfehler geringer beziehungsweise höher wurde. Liegt das Verhältnis bei 56:44, so heißt das, dass in 56 Fällen der gewichtete Kernschätzer besser war als der Nadaraya-Watson-Schätzer und dass es in 44 Fällen umgekehrt war.

Es fällt auf, dass der gewichtete Kernschätzer für größere  $N$  in den meisten Fällen besser ist, was jedoch nicht für kleinere Stichprobenumfänge gilt. Diese Problematik hat mit dem Varianzschätzer zu tun und wird im nächsten Abschnitt gesondert erläutert.

Vorher wird aber noch einmal ein durchschnittliches Fehlerschaubild von 50 verschiedenen Datensätzen bei  $N=200$  betrachtet, um so das asymptotische Verhalten besser beurteilen zu können.

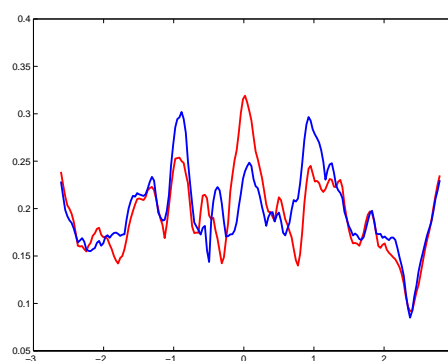


Abbildung 4.12: Fehlerschaubild von 50 Durchläufen bei  $N=200$

Man stellt fest, dass hier genauso wie in dem einzelnen Beispiel vorher, der Fehler im Bereich der Höckerspitze geringer geworden ist, während er an den Seiten des Höckers tendenziell eher gestiegen ist.



## 4.6 Varianzproblematik

Wie in der Fehlertabelle im vorherigen Abschnitt gesehen, ist der gewichtete Kernschätzer für kleinere Stichprobenumfänge  $N$  nicht besonders gut geeignet. Dies hat damit zu tun, dass für die Berechnung von  $\hat{m}(x)$  der Varianzschätzer  $\hat{\sigma}^2(x)$  benötigt wird. Dies setzt aber voraus, dass man für die Varianz einen geeigneten Schätzer zur Verfügung hat. Das Problem ist nun, dass für kleine  $N$  der Varianzschätzer nicht immer optimale Ergebnisse liefert. Dies hat unter anderem damit zu tun, dass oftmals eine zu geringe Bandbreite  $h$  zu einem stark schwankenden Varianzschätzer, wie zum Beispiel in Abbildung (4.13) dargestellt, führt. Dieser sorgt seinerseits für eine extreme Gewichtung, die bei geringem  $N$  auch eine starke Auswirkung auf den Trendschätzer haben kann. Im Folgenden haben wir einmal ein Beispiel einer solchen Problematik dargestellt, in der der Fehler (4.1) von 33.4012 für den Nadaraya-Watson-Schätzer auf 53.0302 in der 5. Iteration gestiegen ist.

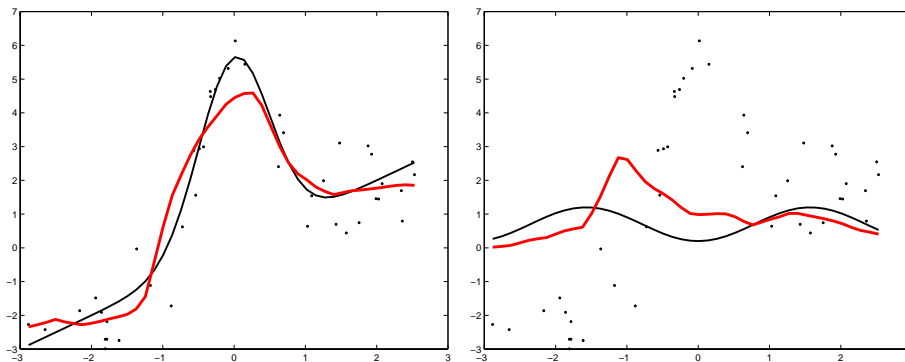


Abbildung 4.13: Trend- und Volatilitätsschätzer (N.-W.-Schätzer)

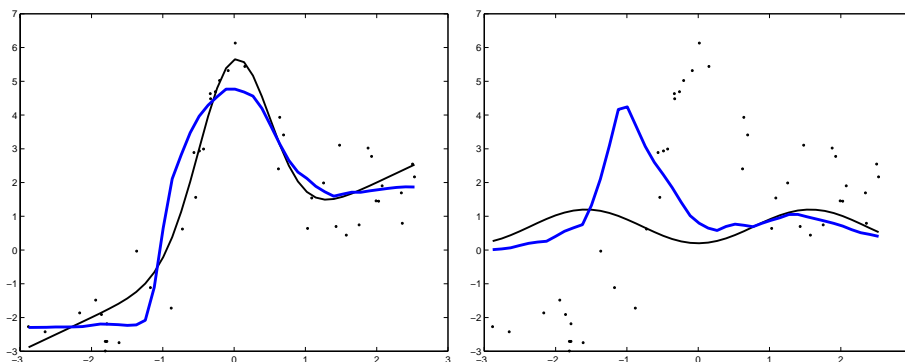


Abbildung 4.14: Gewichteter Trend- und Volatilitätsschätzer (5. Iteration)

## 4.7 Fazit

Man stellt fest, dass der gewichtete Kernschätzer nicht unbedingt in allen Bereichen besser als der gewöhnliche Nadaraya-Watson-Schätzer ist. Und vorallem für kleine Stichprobenumfänge  $N$  muss man auf die Qualität des Varianzschätzers achten, um in Problemfällen nicht die falsche Entscheidung zu treffen.

Nichtsdestotrotz kann man anhand der Fehlertabelle sehen, dass der gewichtete Kernschätzer nicht nur in gewissen Situationen sondern vielmehr auf dem ganzen Regressionsbereich einen kleineren Fehler aufweist. Auch wenn der Rechenaufwand höher ist als beim Nadaraya-Watson-Schätzer, so macht ihn das trotzdem für die Anwendung interessant.

# Kapitel 5

## Simulationsstudie mit Zeitreihendaten

Da nun im vorherigen Abschnitt die gewichteten Kernschätzer anhand von Regressionsdaten qualitativ überprüft wurden, soll dies in diesem Kapitel mit Hilfe von Zeitreihendaten erfolgen. Da der Schätzer letztendlich auf Aktienkurse angewendet werden soll, betrachtet man an dieser Stelle Finanzzeitreihen anstatt der allgemeinen Zeitreihen.

### 5.1 Das Modell

Zu diesem Zweck generiert man sich zuerst einmal Zeitreihendaten, um danach die gewichteten Kernschätzer damit zu berechnen. Dazu verwendet man das in Kapitel 2.3 kennengelernte AR(1)-ARCH(1)-Modell

$$X_j = m(X_{j-1}) + \sigma(X_{j-1})\varepsilon_j \quad j = 1, \dots, N$$
$$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N \text{ u.i.v. mit } E[\varepsilon_j] = 0, \text{Var}[\varepsilon_j] = 1 .$$

In diesem Fall wählt man die Trend- und Volatilitätsfunktionen wie folgt, wobei die Innovationen  $\varepsilon_1, \dots, \varepsilon_N$  wie schon im vorherigen Kapitel  $\mathcal{N}(0, 1)$ -verteilt sind:

$$m(x) = \frac{1}{10}x^2$$
$$\sigma^2(x) = 1 + \frac{1}{9}x^2 + \frac{1}{9}x^{-2} .$$

Diese sind für Finanzzeitreihen durchaus realistisch (Abb 5.1) und erzeugen zum Beispiel die Finanzzeitreihe aus Abbildung 5.2. Hierbei zeigt das erste Bild die Finanzzeitreihe in ihrem zeitlichen Verlauf und das zweite die zur Berechnung des Schätzers verwendeten Datenpaare  $(X_j, X_{j+1})$   $j = 1, \dots, N$ .

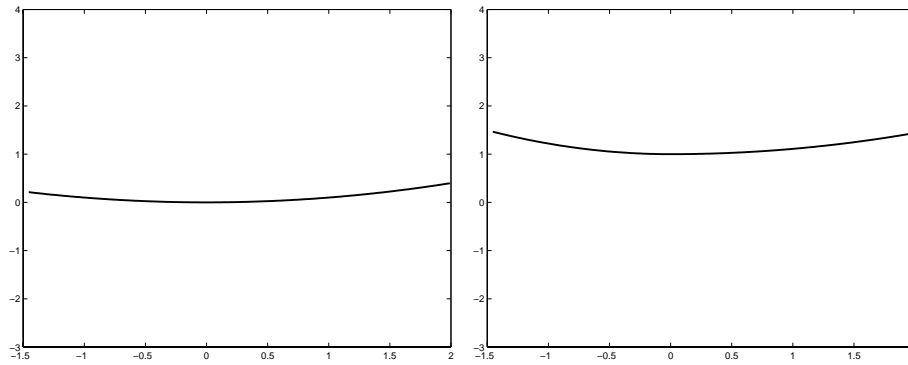


Abbildung 5.1: Die Trendfunktion  $m(x)$  und die Volatilitätsfunktion  $\sigma^2(x)$

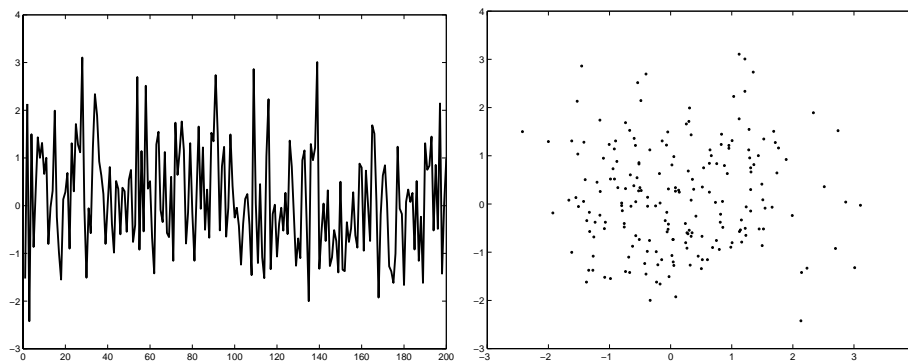


Abbildung 5.2: Nach dem Modell erzeugte Finanzzeitreihe

## 5.2 Nadaraya-Watson-Schätzer

Berechnet man anhand dieser Daten nun den Nadaraya-Watson-Schätzer (2.3), so erhält man die nachfolgenden Schätzer der Trend- und Volatilitätsfunktion.

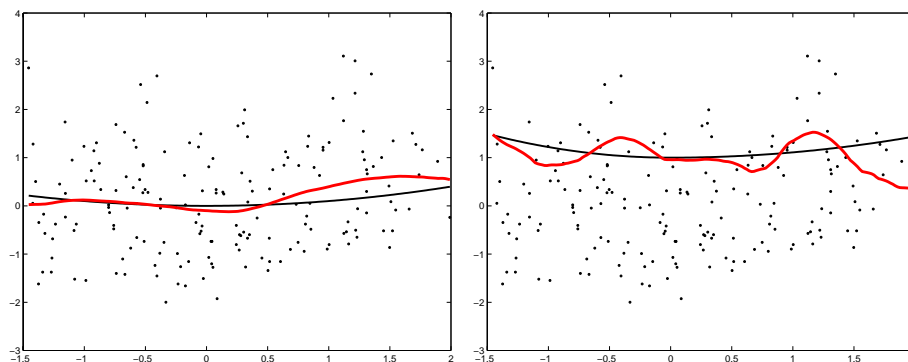


Abbildung 5.3: Nadaraya-Watson-Schätzer  $\hat{m}(x)$  und  $\hat{\sigma}^2(x)$

Um die Qualität des Schätzers genauer beurteilen zu können, schaut man sich, wie im vorherigen Kapitel, das Fehlerschaubild des Trendschätzers an. Hierbei stellt man fest, dass kein markantes Fehlerbild erkennbar ist.

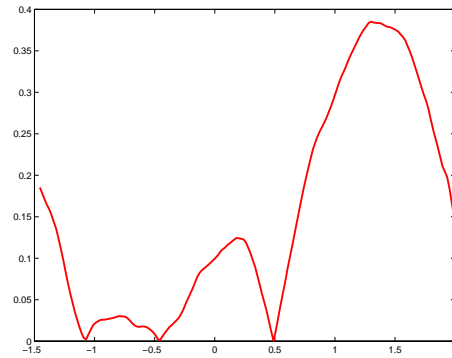


Abbildung 5.4: Fehlerschaubild des Nadaraya-Watson-Schätzers

### 5.3 Gewichteter Kernschätzer

Auch bei den Zeitreihendaten kann man feststellen, dass der Algorithmus schon nach wenigen Iterationen keine ersichtlichen Unterschiede in den Schätzern liefert. Im Gegensatz zum vorherigen Kapitel wird allerdings auf eine Abbildung aller Iterationen verzichtet, sondern man betrachtet gleich die gewichteten Kernschätzer und das Fehlerschaubild der 5. Iteration.

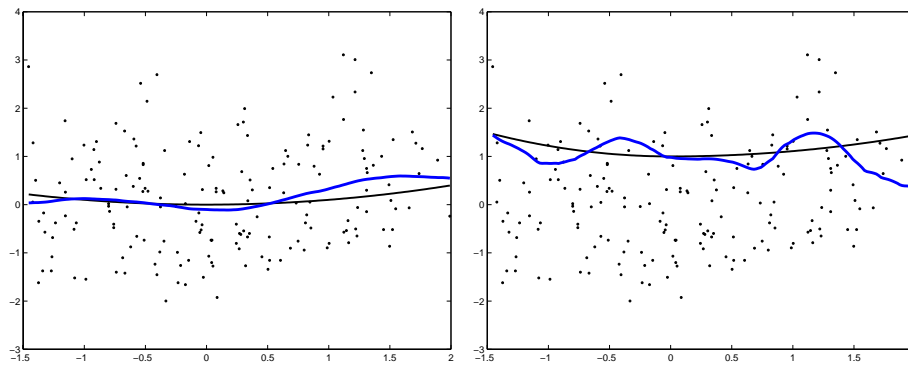


Abbildung 5.5: Gewichtete Kernschätzer der 5. Iteration

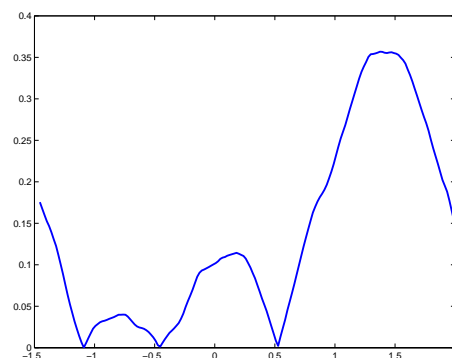


Abbildung 5.6: Fehlerschaubild des gewichteten Trendschätzers

## 5.4 Vergleich der beiden Schätzer

Zum Vergleich der beiden Schätzer betrachtet man erneut die Fehlerschaubilder. Es wird sowohl aus dem Fehlerschaubild als auch aus der Abbildung der Gesamtfehler sofort ersichtlich, dass der Fehler des gewichteten Trendschätzers der 5. Iteration gegenüber dem des Nadaraya-Watson-Schätzers gesenkt werden konnte.

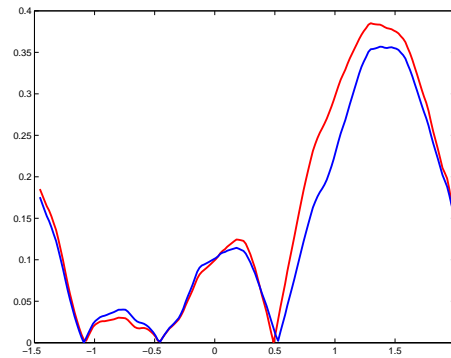


Abbildung 5.7: Fehlerschaubild der 1. (rot) und 5. Iteration (blau)

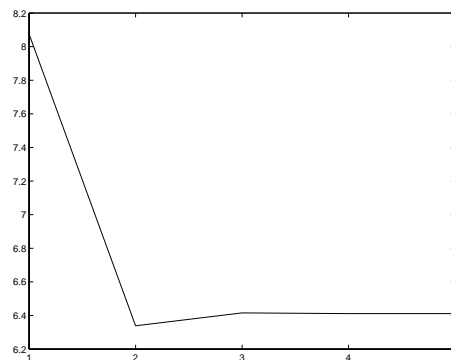


Abbildung 5.8: Fehler der jeweiligen Iteration

## 5.5 Tabelle der Schätzfehler

Man betrachtet nun wieder die Tabelle der Schätzfehler um die Ergebnisse nicht nur auf spezielle Beispiele zu stützen. Die folgende Tabelle zeigt die Durchschnittsfehler bei 100 verschiedenen Simulationen des gewählten Modells.

Stichproben- umfang	Nadaraya-W.- Schätzer	2. Iteration	3. Iteration	4. Iteration	5. Iteration	Verhältnis
N=50	18.3490	19.1576	19.3859	19.3451	19.3403	48 : 52
N=100	20.4059	20.6641	20.3808	20.2764	20.2669	53 : 47
N=200	4.5816	4.4747	4.4695	4.4756	4.4699	65 : 35
N=500	16.5527	16.0201	16.1157	16.0987	16.0914	58 : 42

Es wird deutlich, dass der Fehler zwar nicht bedeutend verringert werden kann, dass jedoch trotzdem in einem Großteil der Fälle der Fehler gesenkt werden konnte. Für kleinere  $N$  taucht auch an dieser Stelle wieder die Problematik des Varianzschätzers, wie sie im vorherigen Kapitel beschrieben wurde, auf. Gerade bei einer flachen Trendfunktion wie es hier und auch meistens bei Finanzzeitreihen der Fall ist, wirkt sich dieses Problem eines extrem schlechten Varianzschätzers besonders stark aus.

Das folgende Bild zeigt ein durchschnittliches Fehlerschaubild von 50 Durchläufen.

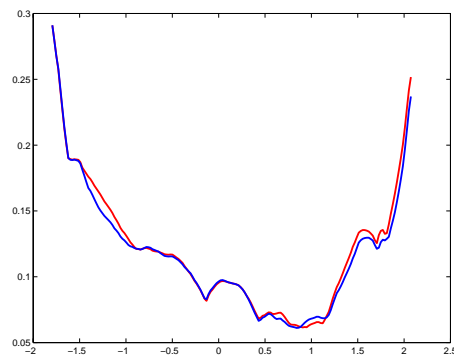


Abbildung 5.9: Fehlerschaubild von 50 Durchläufen bei  $N=200$

## 5.6 Fazit

Trotz der Varianzproblematik kann man den gewichteten Kernschätzer auch sehr gut auf Zeitreihen anwenden, da auch hier in den meisten Fällen der Fehler gesenkt werden konnte. Jedoch empfiehlt es sich auch hier einen Blick auf den Varianzschätzer und dessen Aussagekraft zu werfen, um erhöhte Fehler zu vermeiden.

# Kapitel 6

## Betrachtung anhand realer Daten

Da in den vorherigen beiden Kapiteln veranschaulicht wurde, dass eine Verwendung der gewichteten Kernschätzer Sinn macht und sie im Vergleich zu den gewöhnlichen Nadaraya-Watson-Schätzern oft einen geringeren Fehler aufweisen, sollen diese nun auf reale Daten aus der Finanzwelt angewendet werden.

### 6.1 DAX-Kurse

In diesem Abschnitt werden sie zuerst einmal für den Deutschen Aktienindex DAX berechnet. Dazu geht man davon aus, dass die Kurse des DAX dem AR(1)-ARCH(1)-Modell aus Kapitel 2.3 genügen.

#### 6.1.1 Die Daten

Als Grundlage dienen in diesem Fall die Tagesendstände  $X_t$  des DAX im Zeitraum vom 09.08.1996 bis 08.08.2002. Die Zeitreihe der Renditen  $R_t$  ergibt sich damit aus den prozentualen Veränderungen zum Vortagesstand.

$$R_t = \frac{X_t - X_{t-1}}{X_{t-1}} \quad (6.1)$$

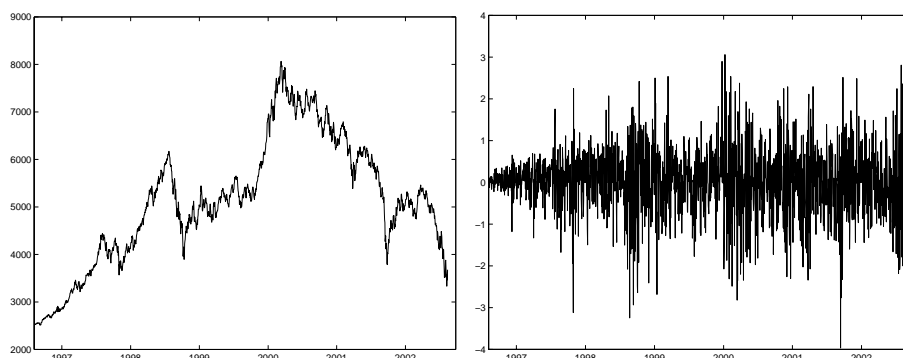


Abbildung 6.1: Aktienkursentwicklung und Zeitreihe der prozentualen Tagesveränderungen



Zur Berechnung der Trend- und Volatilitätsschätzer werden die Datenpaare  $(R_{t-1}, R_t)$ , die wie folgt aussehen, verwendet.

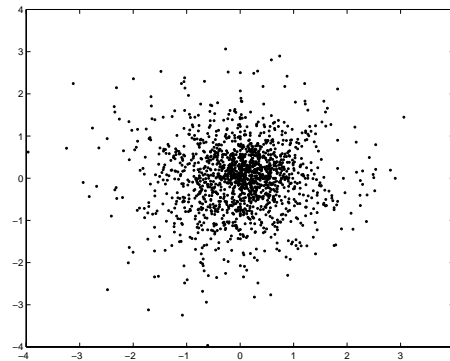


Abbildung 6.2: Die Datenpaare  $(R_{t-1}, R_t)$

### 6.1.2 Gewichtete Kernschätzer

Wendet man nun den Algorithmus aus Kapitel 3 auf diese Datenpaare an, so erhält man nach der 3. Iteration folgende Schätzer für den Trend und die Volatilität. Die jeweils ersten Schaubilder zeigen hierbei die zur Wahl der Glättungsparameter relevanten Funktionen.

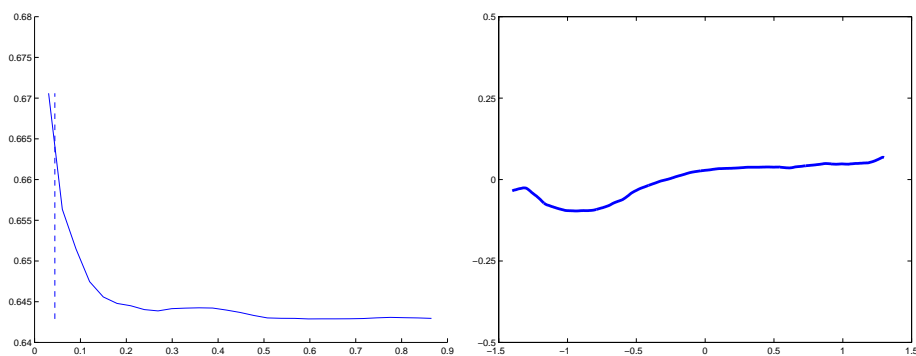


Abbildung 6.3: Trendschätzer des DAX

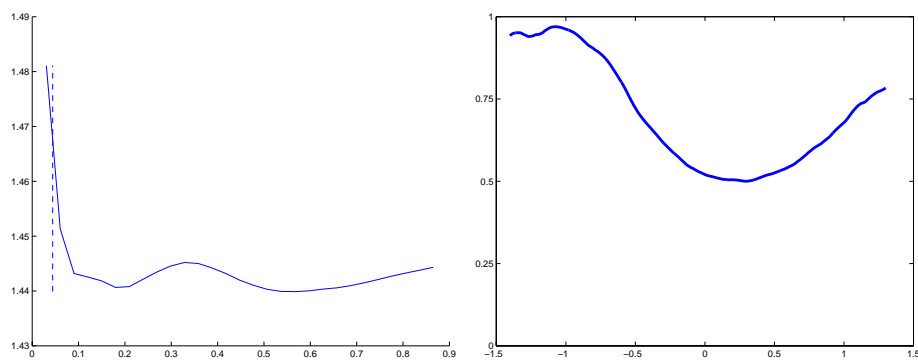


Abbildung 6.4: Volatilitätsschätzer des DAX

Bei Betrachtung des Trendschätzers wird ersichtlich, dass dieser für Vortagesänderungen zwischen  $-1.5$  und  $-0.5$  Prozent deutlich negativ ist, was auch den Erwartungen an den Schätzer entspricht. Ist die Veränderung heute dagegen größer null, so ist der Trend für morgen auch positiv.

Auch der Volatilitätsschätzer entspricht den Erwartungen, da die bedingte Varianz für eine betragsmäßig kleine Vortagesänderung am geringsten ist. In den anderen Fällen kommt es aufgrund von Panikreaktionen oftmals zu großen Veränderungen, ob positiv oder negativ. Dies führt somit zu einer größeren Varianz.

## 6.2 Dow Jones-Kurse

Als zweites sollen die gewichteten Kernschätzer nun noch anhand der Kurse des amerikanischen Aktienindex Dow Jones betrachtet werden.

### 6.2.1 Die Daten

Hierzu verwendet man die Tagesendkurse vom 08.08.1998 bis zum 08.08.2002. Der Kursverlauf und die zugehörige Zeitreihe  $R_t$ , die sich entsprechend dem vorherigen Abschnitt berechnet, sind in den folgenden Grafiken dargestellt.

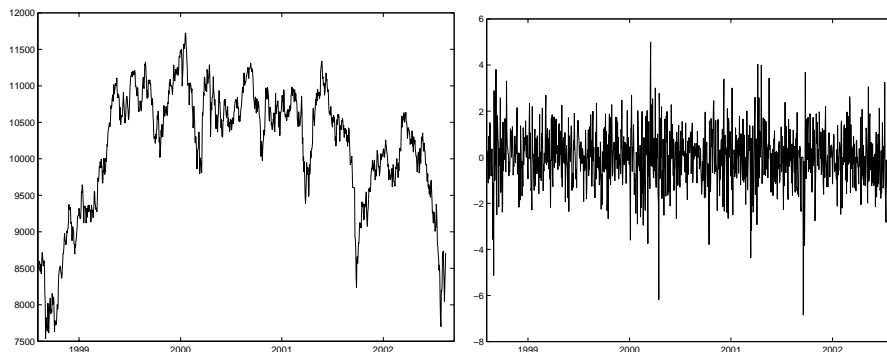


Abbildung 6.5: Grafik über den Kursverlauf und die zugehörige Zeitreihe

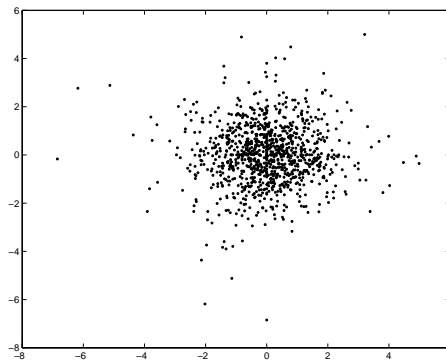


Abbildung 6.6: Die zur Berechnung verwendeten Datenpaare

Mit Hilfe dieser Datenpaare wird nun der gewichtete Kernschätzer für die Trend- und Volatilitätsfunktion berechnet.

### 6.2.2 Gewichtete Kernschätzer

Die sich daraus ergebenden Schätzer sehen wie folgt aus, wobei die ersten Abbildungen jeweils die zugehörigen Funktionen (2.7) und (2.13) zur Wahl des optimalen Glättungsparameters  $h$  und  $\tilde{h}$  zeigen.

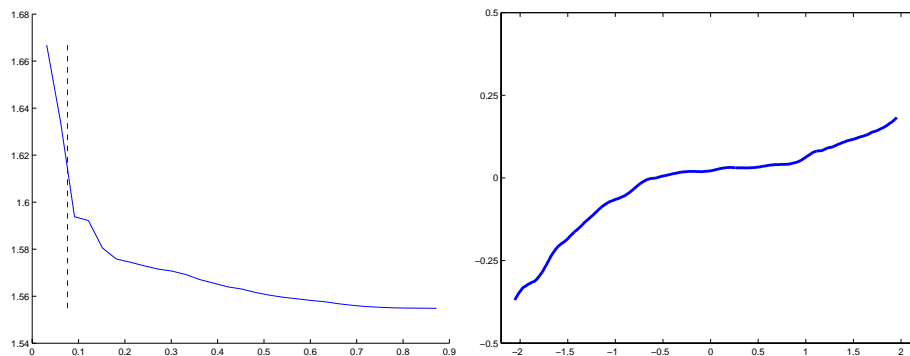


Abbildung 6.7: Trendschätzer des Dow Jones

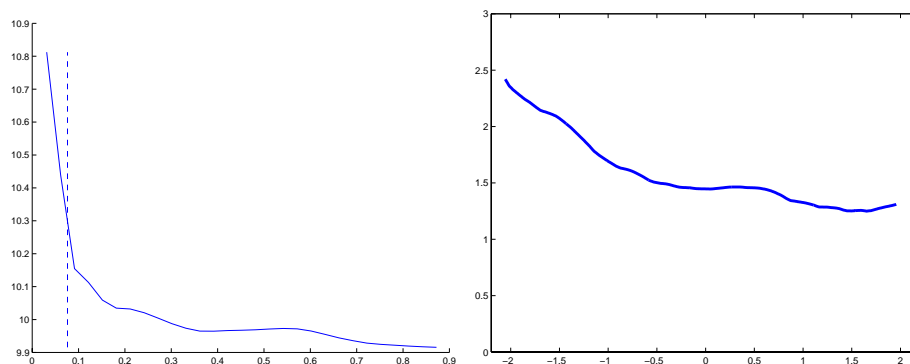


Abbildung 6.8: Volatilitätsschätzer des Dow Jones

Diese sind nicht unbedingt identisch mit denen des DAX, was aber auch nicht zu erwarten war, da erstens die Kurshistorie nicht so weit zurückgeht wie beim DAX und da es ja doch unterschiedliche wirtschaftliche Entwicklungen in den beiden Ländern gab. Dennoch lassen sich auch hier deutliche Merkmale des Aktienmarktes wiederfinden, wie zum Beispiel die signifikant höhere Volatilität bei extrem negativen Vortagesveränderungen und die Tendenz den Vortagestrend fortzusetzen.

# Kapitel 7

## Resümee und Ausblick

Zusammenfassend lassen sich die Ergebnisse in kurzer Form darstellen. Die Methode der gewichteten Kernschätzer hat gegenüber den gewöhnlichen Nadaraya-Watson-Schätzern keine allzu gravierenden Nachteile. Der Rechenaufwand ist höher, da die erste Iteration des kennengelernten Algorithmus bereits den Nadaraya-Watson-Schätzer liefert. Jedoch ergibt sich schon nach wenigen Iterationen ein zufriedenstellend genauer gewichteter Kernschätzer. Auch die festgestellte Varianzproblematik kann durch eine genauere Betrachtung des Varianzschätzers vernachlässigt werden.

Hinzu kommt der wohl entscheidende Vorteil, dass in den meisten Fällen der Fehler zwischen dem Schätzer und der zu schätzenden Funktion verkleinert werden kann.

Alles in allem lässt sich somit sagen, dass die Methode der gewichteten Kernschätzer eine durchaus effiziente Alternative zu den gewöhnlichen Nadaraya-Watson-Schätzern darstellt. Insbesondere in den Fällen, in denen der Rechenaufwand gegenüber der Qualität des Schätzers keine so große Rolle spielt.

Vorausblickend gibt es noch eine Reihe von Punkten, die eine genauere Betrachtung verdienen. So habe ich mich hier in dieser Arbeit zum Beispiel auf den speziellen Fall beschränkt, dass der Wert  $Y_j$  nur von  $X_j$  abhängt. Bei Finanzzeitreihen ist es jedoch vielmehr so, dass eine Abhängigkeit von mehreren vorherigen Werten besteht.

Demnach müßte man das viel allgemeinere Modell

$$X_t = m(X_{t-1}, \dots, X_{t-p}) + \sigma(X_{t-1}, \dots, X_{t-p})\varepsilon_t \quad t = 1, 2, \dots$$

wählen.

Die entsprechenden mehrdimensionalen gewichteten Kernschätzer, die aber auch noch genaueren Betrachtungen und Untersuchungen unterzogen werden müssten, würden dann wie folgt aussehen [Här93] [AT90]:

$$\hat{m}(\underline{x}, \underline{h}, \tilde{\underline{h}}) = \frac{\sum_{j=p}^N \left[ \prod_{i=1}^p K_{h_i}(x_i - X_{j+1-i}) \right] \frac{Y_j}{\hat{\sigma}^2(X_j, \dots, X_{j-p+1}, \underline{h}, \tilde{\underline{h}})}}{\sum_{j=p}^N \left[ \prod_{i=1}^p K_{h_i}(x_i - X_{j+1-i}) \right] \frac{1}{\hat{\sigma}^2(X_j, \dots, X_{j-p+1}, \underline{h}, \tilde{\underline{h}})}} \quad \underline{x}, \underline{h}, \tilde{\underline{h}} \in \mathbb{R}^p$$

$$\hat{\sigma}^2(\underline{x}, \underline{h}, \tilde{\underline{h}}) = \frac{\sum_{j=d}^N \left[ \prod_{i=1}^p K_{\tilde{h}_i}(x_i - X_{j+1-i}) \right] \left[ Y_j - \hat{m}(X_j, \dots, X_{j-d+1}, \underline{h}, \tilde{\underline{h}}) \right]^2}{\sum_{j=d}^N \left[ \prod_{i=1}^p K_{\tilde{h}_i}(x_i - X_{j+1-i}) \right]} \quad \underline{x}, \underline{h}, \tilde{\underline{h}} \in \mathbb{R}^p .$$

# Literaturverzeichnis

- [AT90] B. Auestad und D. Tjøstheim. Identification of nonlinear time series. In *Beimetrika*, Jgg. 77, Seiten 669–687. 1990.
- [BJ76] G. E. P. Box und G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Fransisco: Holden Day, 1976.
- [Bre73] L. Breiman. *Statistics - With a View toward Applications*. Houghton Mifflin Company, 1973.
- [CJ37] A. Cowles und H. E. Jones. Some a posteriori probabilities in stock market action. In *Econometrica*, Jgg. 5, Seiten 280–294. 1937.
- [Cow33] A. Cowles. Can Stock Market Forecasters Forecast? In *Econometrica*, Jgg. 1, Seiten 309–324. 1933.
- [Cow44] A. Cowles. Stock Market Forecasting. In *Econometrica*, Jgg. 12, Seiten 206–214. 1944.
- [DS81] N. R. Draper und H. Smith. *Applied Regression Analysis*. John Wiley and Sons, 1981.
- [Eli00] M. Elifnew. A Nonparametric Black-Scholes Model in Discrete Time. Diplomarbeit, Universität Kaiserslautern, 2000.
- [Eng82] R. F. Engle. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. In *Econometrica*, Jgg. 50, Seiten 987–1008. 1982.
- [Epa69] V. Epanechnikov. Nonparametric estimates of a multivariate probability density. In *Theory of Probability and its Applications*, Jgg. 14, Seiten 153–158. 1969.
- [FHH01] J. Franke, W. Härdle und C. Hafner. *Einführung in die Statistik der Finanzmärkte*. Springer Verlag, 2001.
- [FKM97] J. Franke, J. Kreiss und E. Mammen. Bootstrap of Kernel Smoothing in nonlinear Time Series. In *Technical Report*. Department of Mathematics, University of Kaiserslautern, 1997.
- [Fra98] J. Franke. Nonlinear and Nonparametric Methods for Analyzing Financial Time Series. In *Report in Wirtschaftsmathematik*, Jgg. 44. Universität Kaiserslautern, 1998.

- [FY98] J. Fan und Q. Yao. Efficient estimation of conditional variance functions in stochastic regression. In *Biometrika*, Jgg. 85, Seiten 645–660. 1998.
- [Har96] J. D. Hart. Some automated methods of smoothing time-dependent data. In *Journal of Nonparametric Statistics*, Jgg. 6, Seiten 115–142. 1996.
- [Här93] W. Härdle. *Applied Nonparametric Regression*. Cambridge Press, 1993.
- [HT97] W. Härdle und A. Tsybakov. Local Polynomial Estimation of the Volatility Function. In *Journal of Econometrics*, Jgg. 81, Seiten 223–242. 1997.
- [KK99] R. Korn und E. Korn. *Optionsbewertung und Portfolio-Optimierung*. Vieweg Verlag, 1999.
- [Mil93] T. C. Mills. *The econometric Modelling of Financial Time Series*. Cambridge University Press, 1993.
- [Nad64] E. A. Nadaraya. On estimating regression. In *Theory of Probability and their Applications*, Jgg. 10, Seiten 186–190. 1964.
- [Par62] E. Parzen. On estimation of a probability density and mode. In *Annals of Statistics*, Jgg. 35, Seiten 1065–1076. 1962.
- [PTT01] D. Pena, G.C. Tiao und R.S. Tsay. *A course in Time Series Analysis*, Kapitel 12. John Wiley and Sons, 2001.
- [Ros56] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. In *Annals of Statistics*, Jgg. 27, Seiten 642–669. 1956.
- [Wat64] G. S. Watson. Smooth Regression Analysis. In *Sankhya, Series A*, Jgg. 26, Seiten 359–372. 1964.
- [Wei80] S. Weisberg. *Applied Linear Regression*. John Wiley and Sons, 1980.
- [Wor34] H. Working. A random-difference series for use in the analysis of time series. In *Journal of the American Statistical Association*, Jgg. 29, Seiten 11–24. 1934.
- [YHN99] L. Yang, W. Härdle und J. P. Nielsen. Nonparametric Autoregression with Multiplicative Volatility and Additive Mean. In *Journal of Time Series Analysis*, Jgg. 20, Seiten 579–604. 1999.