

Asymptotischer Vergleich höherer Ordnung adaptiver linearer Schätzer in Regressionsmodellen

Dissertation



zur Erlangung des akademischen Grades
Dr. rer. nat.
im Fach Mathematik

eingereicht an der

Mathematisch–Naturwissenschaftlichen Fakultät II
der Humboldt–Universität zu Berlin

von

Dipl. Math. **Pierre Emmanuel ILOUGA**

geb. am 22. November 1968 in Yaoundé (Kamerun)

Präsident der Humboldt–Universität zu Berlin

Prof. Dr. Dr. h. c. Hans Meyer

Dekan der Mathematisch–Naturwissenschaftlichen Fakultät II

Prof. Dr. Bodo Krause

Gutachter:

1. Prof. Dr. habil. Olaf Bunke, Humboldt–Universität zu Berlin
2. Prof. Dr. Enno Mammen, Universität Heidelberg
3. Prof. Dr. Michael H. Neumann, Universität Köln

Tag der Einreichung

Berlin, den 22.08.2000

Tag der mündlichen Prüfung

Berlin, den 20.04.2001

Danksagung

An dieser Stelle möchte ich mich bei meinem Betreuer, *Prof. Dr. Olaf Bunke* für die nötige Unterstützung, die wertvollen Kommentare und Hinweise bedanken, die dazu geführt haben, daß diese Arbeit in der jetzigen Form zustande kam.

Ich bedanke mich ebenfalls bei *Dr. Rolf Thrum* für die Bereitschaft, die er mehrmals zeigte, interessante Diskussionen mit mir zu führen, bei *Dr. Bernd Droge* für Hinweise bei den graphischen Darstellungen und bei *Dipl. Math. Daniela Syrbe* für viele Vorschläge bei der exakten Formulierung von Aussagen.

Ein besonderes Dankeschön geht an meine Familie in Kamerun, die mich während meines ganzen Studiums unterstützt hat.

Inhaltsverzeichnis

1	Einführung	1
2	Das Modell und die Voraussetzungen	3
2.1	Das Modell	3
2.2	Die Voraussetzungen	4
3	Die Schätzer der Regressionsfunktion	12
3.1	Glättung mit der optimalen Bandbreite	12
3.2	Die adaptiven Schätzungen der Regressionsfunktion	14
3.2.1	Der Plug-In Schätzer	14
3.2.2	Die anderen adaptiven Schätzer	15
4	Asymptotischer Vergleich der adaptiven Schätzer	18
4.1	Asymptotische Resultate 2. Ordnung	18
4.2	Asymptotische Resultate 3. Ordnung	19
5	Der Fall $Pf = 0$	26
5.1	Adaptive Schätzungen	27
5.2	Eine k -stufige Plug-In Schätzung	30
6	Die Stauchung des Mittelwertvektors	33
6.1	Schätzung mit dem optimalen Parameter	34
6.2	Die adaptiven Schätzungen	34
7	Verbesserung des Risikos zweiter Ordnung	39
8	Numerische Resultate	42
8.1	Der Fall glatter Funktionen	42
8.2	Der Fall unglatter Funktionen	53
9	Anwendung auf die Regression mit wachsender Anzahl von Versuchspunkten	65

10	Hilfssätze und Beweise	73
10.1	Hilfssätze	73
10.2	Beweise	77

Kapitel 1

Einführung

Die nichtparametrische Schätzung der Regressionsfunktion hat sich seit vielen Jahren empfohlen, wenn es darum geht, den Zusammenhang zwischen einer Einflußgröße und einer abhängigen Variable zu beschreiben. Wird dieser Zusammenhang durch eine unbekannte Funktion f dargestellt, so hängt die Schätzung von f von einem Parameter h (Bandbreite oder Modelldimension) ab. Er wird im allgemeinen als Glättungsparameter bezeichnet. In der nichtparametrischen Regressionsfunktionsschätzung kontrolliert dieser Parameter die Glattheit der Schätzung. Eine große Bandbreite führt zu einer Überglättung von f . Man wird eine Schätzung mit großer Verzerrung bekommen. Dagegen liefert eine zu kleine Bandbreite eine Schätzung mit großer Variabilität. Aus diesem Grund spielt die Wahl des Glättungsparameters eine wichtige Rolle (siehe *Wand und Jones (1995)*, *Härdle (1990)* und *Silverman (1986)*). Für praktische Probleme ist es wünschenswert, die Bandbreite adaptiv, d.h. datenabhängig, zu wählen. Zu dieser Problematik sind bereits eine ganze Reihe von Arbeiten erschienen. *Ruppert, Sheather und Wand (1995)* konstruieren eine Plug-In Schätzung von h für die lokal lineare Regression mit der Rate $n^{-2/7}$, wobei n der Stichprobenumfang ist. *Hall, Sheather, Jones und Marron (1991)* entwickelten eine $n^{-1/2}$ -konsistente Bandbreite für die Dichteschätzung, indem sie den Bias höherer Ordnung der Dichteschätzung aufschreiben und dazu eine $n^{-1/2}$ -konvergente Schätzung für das Integral über das Quadrat der zweiten Ableitung der Dichtefunktion konstruieren. Eine Zusammenfassung der Methoden zur Wahl der Bandbreite bei der Dichteschätzung findet man in *Jones, Marron und Sheather (1996)*. In der Arbeit von *Härdle, Hall und Marron (1988)* wird für die nichtparametrische Regression bewiesen, daß die aus automatischen Methoden resultierenden Bandbreiten die gleiche Konvergenzrate von $n^{-1/10}$ besitzen. Eine ähnliche asymptotische Äquivalenz im Sinne der Gleichheit der Risiken zweiter Ordnung der adaptiven Schätzungen der Regressionsfunktion beweist *Neumann (1992)* für eine feste Anzahl von

Versuchspunkten und groß werdende Anzahl von Wiederholungen an jedem Versuchspunkt bzw. klein werdende Varianz der Beobachtungen.

Da die adaptiven Schätzungen der Regressionsfunktion nicht mehr linear in den Beobachtungen sind, wird in dieser Arbeit ihre Güte dadurch verglichen, daß ihre Risiken verglichen werden. Für große Stichprobenumfänge sind die Risiken zweiter Ordnung der adaptiven Schätzungen gleich den Risiken zweiter Ordnung der Plug-In Schätzung. Um einen Unterschied der adaptiven Schätzer quantifizieren zu können, ist es nötig, ihre Risiken höherer Ordnung zu vergleichen. Die Berechnung der Risiken höherer Ordnung ist zum Beispiel in *Huang (1999)* ein Ausweg, um eine Antwort auf das „Even-Odd“-Phänomen bei der Schätzung des Medians einer Stichprobe zu geben.

In dieser Dissertation werden (unter Annahme von normalverteilten Beobachtungen) verschiedene Kriterien zur Wahl der Bandbreite miteinander verglichen, indem der Vergleich der Risiken höherer Ordnung der entsprechenden adaptiven Schätzungen der Regressionsfunktionen durchgeführt wird. Dies wird bei fester Anzahl von Versuchspunkten und wachsender Anzahl von Wiederholungen an jedem Versuchspunkt gemacht. Ein neues Kriterium (Full Cross-validation genannt, siehe *Droge (1996)* und *Bunke, Droge und Polzehl (1999)*), das gute Eigenschaften bei der Problematik der Modellwahl in der linearen und nichtlinearen Regression aufweist (*Droge (1999)*), wird auch hier zur Wahl der Bandbreite in Betracht gezogen.

Diese Arbeit ist wie folgt gegliedert: zunächst wird das Modell beschrieben und es werden alle Voraussetzungen aufgeführt, die für die asymptotischen Entwicklungen notwendig sind. Danach werden die Schätzungen der unbekanntes Regressionsfunktion sowie die Kriterien zur Wahl der Bandbreite vorgestellt. Im zentralen Teil der Arbeit geht es um den asymptotischen Vergleich der verschiedenen adaptiven Schätzer der Regressionsfunktion. Dazu werden die asymptotischen Resultate zweiter Ordnung, die in der Literatur zu finden sind angegeben. Es wird auch den Fall behandelt, daß die zu schätzende Funktion „sehr glatt“ ist. Anschließend wird ein Schätzer der Regressionsfunktion angegeben, dessen Risiko zweiter Ordnung kleiner als das Risiko zweiter Ordnung des Plug-In Schätzers ist. In einer Simulationsstudie werden die Risiken der adaptiven Schätzer verglichen. Es folgt ein Kapitel, das zur Untersuchung im Falle einer wachsenden Anzahl von Versuchspunkten dient. Am Ende der Arbeit sind Hilfssätze und Beweise enthalten.

Kapitel 2

Das Modell und die Voraussetzungen

2.1 Das Modell

Wir betrachten ein Regressionsmodell der Form

$$y_{ij} = f(x_i) + \varepsilon_{ij} \quad i = 1, \dots, m \quad j = 1, \dots, n_i. \quad (2.1)$$

Die n_i Beobachtungen (Wiederholungen) y_{ij} ($j = 1, \dots, n_i$) sind Realisierungen von Zufallsgrößen Y_{ij} . Die Anzahl m der Versuchspunkte ist fest. Die Versuchspunkte x_i ($i = 1, \dots, m$) sind nichtzufällige Punkte aus \mathbb{R}^p , $p \geq 1$.

Die Fehler ε_{ij} haben den Erwartungswert 0 und homogene Varianz σ^2 und seien paarweise unkorreliert.

Im Regressionsmodell (2.1) sind sowohl die Funktion f als auch die Varianz σ^2 unbekannt. Solche Modelle kommen beispielsweise in der Analyse von Feldexperimenten vor. Auf Seite 10 wird der Faktorplan von *Green, Jennison and Seheult (1985)* als Beispiel angegeben.

Das Ziel ist, den Vektor $f := (f_1, \dots, f_m)^T := (f(x_1), \dots, f(x_m))^T$ zu schätzen. Die erste Schätzung, die man vorschlagen kann, ist die kleinste Quadratschätzung

$$\bar{Y} := (\bar{Y}_1, \dots, \bar{Y}_m)^T \quad \text{mit} \quad \bar{Y}_i := n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}.$$

In dieser Arbeit werden wir lineare Schätzungen (linear in \bar{Y}) von f der Gestalt

$$\hat{f}_h = K(n, h)\bar{Y} \quad (2.2)$$

betrachten. Die Glättungsmatrix $K(n, h)$ ist eine $m \times m$ -Matrix, die von n und einem Parameter $h \in H := [0, h_{max}]$ ($h_{max} < \infty$) abhängt, wobei $n := \sum_{i=1}^m n_i$ die Gesamtzahl der Beobachtungen ist. Die Schätzung (2.2) ist nicht mehr erwartungstreu. Dagegen gibt es, wie wir später sehen werden, eine Menge $H_0 \subset H$ so, daß der mittlere quadratische Fehler von \hat{f}_h kleiner ist als der von \bar{Y} für alle $h \in H_0$.

Es wird für die Konstruktion adaptiver (datenabhängiger) Schätzungen von f eine Varianzschätzung $\hat{\sigma}^2$ verwendet, die auf

$$s^2 := (n - m)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (2.3)$$

beruht. Die Güte einer Schätzung der Form (2.2) wird beurteilt durch ihren mittleren quadratischen Fehler

$$M(h) := MSE(\hat{f}_h) = E\|\hat{f}_h - f\|^2, \quad (2.4)$$

wobei $\|x\|^2 := \sum_{i=1}^m x_i^2$, $x \in \mathbb{R}^m$.

Für die Schätzung (2.2) ist

$$M(h) = \|[I_m - K(n, h)]f\|^2 + \sigma^2 \text{tr}[K(n, h)^T K(n, h)M_n], \quad (2.5)$$

wobei $M_n = \text{Diag}(n_1^{-1}, \dots, n_m^{-1})$ die $m \times m$ -Matrix mit den Elementen n_i^{-1} auf der Hauptdiagonale und Nullen sonst ist.

2.2 Die Voraussetzungen

Um eine asymptotische Entwicklung des Glättungsparameters und der mittleren quadratischen Fehler der Schätzungen von f herleiten zu können, müssen einige Voraussetzungen an die Glättungsmatrix und die wachsenden Stichprobenumfänge $n_i = n_i(n)$ gemacht werden.

V1 $K(n, h)$ ist stetig in $h \in H$ für jedes $n > \tilde{n}$ ($\tilde{n} \geq m$).

V2 $K(n, 0) = I_m$

V3 Es existiert ein $h_d < h_{max}$ so, daß $K(n, h)$ fünfmal stetig differenzierbar auf $[0, h_d]$ ist. Wir führen die folgenden Bezeichnungen ein:

$$P_n := -\frac{\partial K(n, h)}{\partial h} \Big|_{h=0} \quad (2.6)$$

$$Q_n := (2!)^{-1} \frac{\partial^2 K(n, h)}{\partial h^2} \Big|_{h=0} \quad (2.7)$$

$$R_n := -(3!)^{-1} \frac{\partial^3 K(n, h)}{\partial h^3} \Big|_{h=0} \quad (2.8)$$

$$T_n := (4!)^{-1} \frac{\partial^4 K(n, h)}{\partial h^4} \Big|_{h=0} \quad (2.9)$$

V4 Es existieren Matrizen P, Q, R und T so, daß

$$\lim_{n \rightarrow \infty} \|P_n - P\|^2 = 0 \quad (2.10)$$

$$\lim_{n \rightarrow \infty} \|Q_n - Q\|^2 = 0 \quad (2.11)$$

$$\lim_{n \rightarrow \infty} \|R_n - R\|^2 = 0 \quad (2.12)$$

$$\lim_{n \rightarrow \infty} \|T_n - T\|^2 = 0. \quad (2.13)$$

(Für eine Matrix A ist $\|A\| := \sqrt{\lambda_{max}(A^T A)}$; $\lambda_{max}(A^T A)$ ist der maximale Eigenwert von $A^T A$.)

V5 Die Diagonalelemente $P_{n,ii}$ sind größer oder gleich 0 und es existiert mindestens ein $j \in \{1, \dots, m\}$ so, daß $P_{jj} > 0$.

V6 $\|P f\| > 0$. ($\Rightarrow \exists n_0 > 0 : \|P_n f\| > 0 \forall n \geq n_0$).

V7 $\lim_{n \rightarrow \infty} n_i n^{-1} =: c_i > 0 \quad i = 1, \dots, m$.

V8 $\varepsilon_{ij} \sim N(0, \sigma^2) \quad i = 1, \dots, m; j = 1, \dots, n_i$.

Deutung der Voraussetzungen:

Nach der ersten Voraussetzung braucht die Glättungsmatrix $K(n, h)$ in h nur ab einem gewissen Stichprobenumfang n stetig zu sein. Sie ist schwächer als die Forderung, daß die Matrizen $K(n, h)$ für jedes $n \geq m$ stetig sein sollen. Die zweite Voraussetzung ist eine natürliche Voraussetzung, die besagt, daß nicht geglättet wird, wenn die Bandbreite 0 ist. Für $h = 0$ ist die Schätzung von f im Punkt x_i gleich dem Mittelwert der Wiederholungen an dem Punkt. Die Voraussetzung **V3**, zusammen mit den Voraussetzungen **V1** und **V2**, ermöglicht es, die Glättungsmatrix $K(n, h)$ durch ein Polynom in h zu approximieren (Taylorentwicklung im Punkt $h = 0$). Die 4. Voraussetzung **V4** gibt die Grenzwerte der Matrizen P_n, Q_n, R_n und T_n an wenn n sehr groß wird. Die fünfte Voraussetzung **V5** sichert, wie wir später sehen werden, daß der optimale Glättungsparameter echt positiv ist. Die Tatsache, daß der optimale Glättungsparameter h_{opt} echt positiv ist, ist notwendig dafür, daß die Schätzung $\hat{f}_{h_{opt}}$ besser ist als \bar{Y} . Durch die Voraussetzung **V6** werden solche Funktionen geschätzt, die nicht zum Nullraum von P bzw. P_n für große n gehören. Der Nullraum von P_n steht für Regressionsfunktionen, die sehr glatt sind. Eine Funktion f ist sehr glatt, falls $K(n, h)f = f \quad \forall h \in H$, d. h. eine Glättung von ihr mit einer beliebigen Bandbreite nichts ändert. Aus $K(n, h)f = f \quad \forall h \in H$ folgt, daß $P_n f = 0$. Wenn die Glättungsmatrix die Gestalt $K(n, h) = (I_m + hP_n)^{-1}$ hat, (siehe die Beispiele 2, 3 und 5 auf Seite 8 bis 10), dann entspricht der Nullraum von P_n genau den sehr glatten Funktionen. Allgemein werden wir unter einer glatten Funktion diejenige verstehen, die der Ungleichung $\|[I_m - K(n, h)]f\|^2 \leq c$, ($c > 0$ klein vorgegeben) genügt. Die in Voraussetzung **V7** erwähnten Stichprobenumfänge n_i wachsen so an, daß für große n_i der Quotient n_i/n gegen eine bestimmte, echt positive Konstante konvergiert. Im Spezialfall $n_1 = \dots = n_m$ ist $c_1 = \dots = c_m = 1/m$. Mit der Voraussetzung **V8** erhalten die asymptotischen Entwicklungen des mittleren quadratischen Fehlers der adaptiven Schätzungen eine etwas einfachere Form. Damit wird ein Vergleich unter ihnen ermöglicht.

Beispiel 1 Normalkernelndichtematrix (vgl. Neumann (1992))

Wir betrachten eindimensionale äquidistante Versuchspunkte $x_i = x_{i-1} + \Delta$ $i = 2, \dots, m$ ($x_1 = 0, \Delta > 0$) und wir schätzen f mit der Nadaraya-Watson Kernschätzung mit der Bandbreite b ($b \in (0, \infty)$). Die Kernfunktion sei hierbei die Dichte der Standardnormalverteilung. Mit der Transformation

$h = \exp(-\Delta^2/2b^2)$, wobei jetzt $h \in (0, 1)$, hat die Kernglättungsmatrix die Form

$$K(n, h) := \left[\left(\frac{h^{|i-j|^2}}{\sum_{k=1}^m h^{|j-k|^2}} \right) \right]_{i,j=1,\dots,m} \quad h \in [0, 1].$$

Daraus ergibt sich

$$\begin{aligned} P_n &:= -\frac{\partial K(n, h)}{\partial h} \Big|_{h=0} \\ &= \begin{pmatrix} 1 & -1 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 2 & -1 & \cdot & \cdot & \cdot & \cdot \\ 0 & -1 & 2 & -1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & -1 & 2 & -1 \\ 0 & \cdot & \cdot & \cdot & 0 & -1 & 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} Q_n &:= (2!)^{-1} \frac{\partial^2 K(n, h)}{\partial h^2} \Big|_{h=0} \\ &= \begin{pmatrix} 1 & -2 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 4 & -2 & \cdot & \cdot & \cdot & \cdot \\ 0 & -2 & 4 & -2 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & -2 & 4 & -2 \\ \cdot & \cdot & \cdot & \cdot & -2 & 4 & -1 \\ 0 & \cdot & \cdot & \cdot & 0 & -2 & 1 \end{pmatrix} \end{aligned}$$

$$R_n := -(3!)^{-1} \frac{\partial^3 K(n, h)}{\partial h^3} \Big|_{h=0}$$

$$= \frac{1}{6} \begin{pmatrix} 1 & -4 & 0 & \cdot & \cdot & \cdot & 0 \\ -1 & 8 & -4 & 0 & \cdot & \cdot & \cdot \\ 0 & -4 & 8 & -4 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & -4 & 8 & -4 \\ \cdot & \cdot & \cdot & \cdot & -4 & 8 & -1 \\ 0 & \cdot & \cdot & \cdot & 0 & -4 & 1 \end{pmatrix}$$

Beispiel 2 Gewichtete kleinste Quadratschätzung mit Straffterm (vgl. Green and Silverman (1994))

Hierbei schätzt man die Regressionsfunktion f durch Minimierung (bezüglich f) der Größe

$$S(f) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} |Y_{ij} - f_i|^2 + hf^T Af ,$$

wobei A eine symmetrische und positiv semidefinite Matrix ist, die die Unglattheit von f bestraft. Die Lösung \hat{f}_h von $\left. \frac{\partial S(f)}{\partial f} \right|_{f=\hat{f}_h} = 0$ ist

$$\hat{f}_h = (I + hA)^{-1} \bar{Y} .$$

Somit ist

$$K(n, h) = (I + hA)^{-1}$$

$$P_n = - \left. \frac{\partial K(n, h)}{\partial h} \right|_{h=0} = A$$

$$Q_n = (2!)^{-1} \left. \frac{\partial^2 K(n, h)}{\partial h^2} \right|_{h=0} = A^2$$

$$R_n = -(3!)^{-1} \left. \frac{\partial^3 K(n, h)}{\partial h^3} \right|_{h=0} = A^3 .$$

Ist der Straffterm durch das Quadrat der euklidischen Norm des $(m-1)$ -dimensionalen Differenzenvektors $f^{(1)} := (f_2 - f_1, \dots, f_m - f_{m-1})^T$ gegeben, so entspricht die Matrix A der Matrix P_n von Beispiel 1.

Beispiel 3 Fortsetzung von Beispiel 2

Wenn diesmal der Straffterm durch das Quadrat der euklidischen Norm des $(m - 2)$ -dimensionalen Vektors der Differenzen zweiter Ordnung $f^{(2)} := (f_2^{(1)} - f_1^{(1)}, \dots, f_{m-1}^{(1)} - f_{m-2}^{(1)})^T$ gegeben ist, so hat die Glättungsmatrix $K(n, h)$ immer noch dieselbe Gestalt $K(n, h) = (I_m + hB)^{-1}$. Hierbei sind

$$P_n = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ -2 & 5 & -4 & 1 & 0 & \cdot & \cdot & \cdot \\ 1 & -4 & 6 & -4 & 1 & \cdot & \cdot & \cdot \\ 0 & \cdot \\ \cdot & \cdot \\ \cdot & 0 \\ \cdot & \cdot & \cdot & 1 & -4 & 6 & -4 & 1 \\ \cdot & \cdot & \cdot & \cdot & 1 & -4 & 5 & -2 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 & -2 & 1 \end{pmatrix} = B,$$

$$Q_n = B^2, \quad R_n = B^3.$$

Beispiel 4 Ridge Schätzung mit Projektionsmatrix (vgl. Bunke (1999))

Es sei \mathbf{P} eine $m \times m$ -Projektionsmatrix. Wir schätzen f mit

$$\hat{f}_c = c\mathbf{P}\bar{Y} + (I_m - \mathbf{P})\bar{Y}, \quad 0 < c \leq 1. \quad (2.14)$$

Diese Schätzung hat die Gestalt einer Ridge Schätzung mit Parameter $\alpha := (1 - c)/c$ (falls $c > 1/2$), denn

$$\begin{aligned} (I_m + \alpha\mathbf{P})^{-1} &= I_m + \sum_{k=1}^{\infty} (-1)^k \alpha^k \mathbf{P}^k \\ &= I_m + \mathbf{P} \sum_{k=1}^{\infty} (-1)^k \alpha^k \\ &= I_m - \frac{\alpha}{1 + \alpha} \mathbf{P} \\ &= (I_m - \mathbf{P}) + \left(1 - \frac{\alpha}{1 + \alpha}\right) \mathbf{P} \\ &= (I_m - \mathbf{P}) + \frac{1}{1 + \alpha} \mathbf{P} \\ &= (I_m - \mathbf{P}) + c\mathbf{P} \quad \text{für } \alpha := (1 - c)/c. \end{aligned}$$

Die Schätzung (2.14) kann umgeschrieben werden in die folgende Gleichungen:

$$\hat{f}_c = \bar{Y} + (c - 1)\mathbf{P}\bar{Y} \quad (2.15)$$

$$= [\mathbf{I}_m - (1 - c)\mathbf{P}]\bar{Y} \quad (2.16)$$

Hierbei sind $P_n = \mathbf{P}$, $Q_n = R_n = T_n = 0$.

Beispiel 5 Analyse von Feldexperimenten (vgl. Green, Jennison and Seheult (1985))

In diesen Feldexperimenten geht es um die Schätzung der Effekte verschiedener Düngemittel (m insgesamt) und verschiedener Feldtypen (l insgesamt) auf den Ertrag. Es seien $n=lm$ die Gesamtzahl der Felder,

$y \dots$	$n \times 1$	Vektor der Erträge
$D \dots$	$n \times m$	Designmatrix für die Düngemittelleffekte
$\tau \dots$	$m \times 1$	Vektor der Düngemittelleffekte
$\xi \dots$	$n \times 1$	Vektor, der die Feldeffekte enthält
$\eta \dots$	$n \times 1$	Fehlervektor.

Green, Jennison und Seheult gehen von folgendem Modell aus:

$$y = D\tau + \xi + \eta .$$

Die wichtigste Annahme in ihrer Betrachtung ist, daß die Funktion ξ glatt ist. Mit anderen Worten soll der Unterschied der Feldeffekte benachbarter Felder des gleichen Typs klein sein. Das wird in folgenden Gleichungen ausgedrückt:

$$\xi_{i+1} - 2\xi_i + \xi_{i-1} \approx 0 . \quad (2.17)$$

oder in der Matrixschreibweise $\Delta\xi \approx 0$, wobei Δ eine $[m(l-2)] \times n$ Tridiagonalmatrix ist.

Sie definieren Schätzungen für τ und ξ durch Minimierung (bezüglich τ und ξ) von

$$\|y - D\tau - \xi\|^2 + \lambda(\Delta\xi)^T(\Delta\xi) , \quad \lambda > 0 .$$

Der Term $(\Delta\xi)^T(\Delta\xi)$ ist der Ausdruck, der die Unglattheit von ξ (Gleichung (2.17)) bestraft. Je größer λ ist, desto glatter ist die Schätzung von ξ . Unter gegebenen Effekte τ schätzt man ξ durch

$$\hat{\xi} = (\mathbf{I}_n + \lambda\Delta^T\Delta)^{-1}(y - D\tau) .$$

Die Matrix $S_\lambda := (\mathbf{I}_n + \lambda\Delta^T\Delta)^{-1}$ erfüllt die Voraussetzungen **V1** bis **V5** vom Abschnitt 2.2.

Bemerkung 2.2.1

Beispiele für glatte Regressionsfunktionen im Beispiel 1, d. h. für die $P_n f \approx 0$ gilt, sind die konstanten Funktionen und die Geraden mit kleinem Anstieg; denn ist

$f_i = ax_i + b$, so gilt für $2 \leq i \leq m - 1$

$$\begin{aligned} [P_n f]_i &= 2f_i - f_{i-1} - f_{i+1} \\ &= 2(ax_i + b) - (ax_{i-1} + b) - (ax_{i+1} + b) \\ &= 2a(x_{i-1} + \Delta) - ax_{i-1} - a(x_{i-1} + 2\Delta) \\ &= 0 \end{aligned}$$

und für $i = 1$ oder $i = m$

$$[P_n f]_i = a\Delta.$$

Also ist $P_n f = 0$ für $a = 0$ und für sehr kleines a ist $\|P_n f\| \approx 0$.

Kapitel 3

Die Schätzer der Regressionsfunktion

3.1 Glättung mit der optimalen Bandbreite

Wir führen weitere Bezeichnungen ein:

$$C := \text{Diag}(c_1^{-1}, \dots, c_m^{-1}) = \lim_{n \rightarrow \infty} n M_n \quad (3.1)$$

$$\lambda := \text{tr}(P C) = \lim_{n \rightarrow \infty} n \text{tr}(P_n M_n) \quad \lambda_n := \text{tr}(P_n M_n) \quad (3.2)$$

$$\delta := \text{tr}(P^T P C) = \lim_{n \rightarrow \infty} n \text{tr}(P_n^T P_n M_n) \quad \delta_n := \text{tr}(P_n^T P_n M_n) \quad (3.3)$$

$$\rho := \text{tr}([P^T P + 2Q] C) = \lim_{n \rightarrow \infty} n \text{tr}([P_n^T P_n + 2Q_n] M_n) \quad (3.4)$$

$$\rho_n := \text{tr}([P_n^T P_n + 2Q_n] M_n) \quad (3.5)$$

$$\omega := f^T P^T P f = \lim_{n \rightarrow \infty} f^T P_n^T P_n f = \lim_{n \rightarrow \infty} \omega_n \quad (3.6)$$

$$\nu := f^T P^T Q f = \lim_{n \rightarrow \infty} f^T P_n^T Q_n f = \lim_{n \rightarrow \infty} \nu_n \quad (3.7)$$

Es sei $h_{opt} = h_{opt}(f, \sigma^2, n) = \arg \min_{h \in H} M(h)$, wobei $M(h)$ durch (2.5) gegeben ist. Da f und σ^2 unbekannt sind, ist $M(h)$ und demzufolge auch h_{opt} unbekannt. Der folgende Satz gibt die asymptotischen Entwicklungen der optimalen nichtzufälligen Bandbreite und des mittleren quadratischen Fehlers der Schätzung mit dieser Bandbreite an.

Satz 3.1.1 (Ilouga(1994))

Unter den Voraussetzungen **V1** bis **V7** und den Bezeichnungen (3.1) bis (3.7) gilt:

$$1) \quad h_{opt} = \frac{\sigma^2 \lambda}{n \omega} - \frac{\sigma^4 \lambda \rho}{n^2 \omega^2} + 3 \frac{\sigma^4 \lambda^2 \nu}{n^2 \omega^3} + o(\sigma^4 n^{-2}) \quad (3.8)$$

$$2) \quad M(h_{opt}) = \sigma^2 \text{tr}(M_n) - \frac{\sigma^4 \lambda^2}{n^2 \omega} + \frac{\sigma^6 \lambda^2 \rho}{n^3 \omega^2} - 2 \frac{\sigma^6 \lambda^3 \nu}{n^3 \omega^3} + o(\sigma^6 n^{-3}). \quad (3.9)$$

Der Beweis dieses Satzes wurde in *Ilouga(1994)* erbracht. Er geht aus von der Gleichung $M'(h_{opt}) = 0$ und basiert auf einem iterativen Verfahren zur Nullstellenbestimmung von Polynomen vom Grad größer oder gleich 3. Das Verfahren wurde zum Beispiel in *Barndorff-Nielsen(1988)*, Seite 44, für die Berechnung von Maximum-Likelihood-Schätzungen in parametrischen statistischen Modellen verwendet. Eine ähnliche Methode wird benutzt, um die Cornisch-Fischer Entwicklung von Quantilen stetiger Verteilungen in Abhängigkeit von den entsprechenden Quantilen der Standardnormalverteilung und umgekehrt zu bekommen (vgl. *Patel und Read (1996)*, Seiten 172-173).

Bemerkung 3.1.1 Mit den Bezeichnungen

$$M_1(h_{opt}) := \sigma^2 \text{tr}(M_n) \quad (3.10)$$

$$M_2(h_{opt}) := M_1(h_{opt}) - \frac{\sigma^4 \lambda^2}{n^2 \omega} \quad (3.11)$$

und

$$M_3(h_{opt}) := M_2(h_{opt}) + \frac{\sigma^6 \lambda^2 \rho}{n^3 \omega^2} - 2 \frac{\sigma^6 \lambda^3 \nu}{n^3 \omega^3}$$

können wir das optimale Risiko $M(h_{opt})$ in der folgende Form darstellen:

$$\begin{aligned} M_\infty(h_{opt}) &:= M(h_{opt}) \\ &= M_1(h_{opt}) + o(\sigma^2 n^{-1}) \\ &= M_2(h_{opt}) + o(\sigma^4 n^{-2}) \\ &= M_3(h_{opt}) + o(\sigma^6 n^{-3}). \end{aligned}$$

Damit stellen $M_1(h_{opt})$, $M_2(h_{opt})$ und $M_3(h_{opt})$ Approximationen erster, zweiter und dritter Ordnung des optimalen Risikos dar.

Bemerkung 3.1.2

Zwei Arten von Asymptotiken sind an dieser Stelle denkbar. Zum Einen könnte die Anzahl n_i der Wiederholungen im Versuchspunkt x_i groß sein. Zum Anderen, wenn die Anzahl der Wiederholungen im Versuchspunkt x_i nicht groß ist, so könnte σ^2 klein sein. Eine Asymptotik für $\sigma^2 \rightarrow 0$ könnte in diesem Fall betrieben werden. Im Satz 3.1.1 liegt eine Asymptotik für $v^2 := \sigma^2/n \rightarrow 0$ vor, die für beide o.g. Fälle gültig ist. In dieser Arbeit wird es um die Asymptotik für große n_i (bzw. n) gehen.

3.2 Die adaptiven Schätzungen der Regressionsfunktion

3.2.1 Der Plug-In Schätzer

In diesem Abschnitt wird h_{opt} geschätzt. In der Gleichung (3.8) schätzen wir die Approximation zweiter Ordnung der optimalen Bandbreite (d. h. die ersten drei Glieder auf der rechten Seite von (3.8)), indem wir die Unbekannten f und σ^2 durch erwartungstreue Schätzer $\hat{f} := \bar{Y}$ und $\hat{\sigma}^2 := s^2$ ersetzen. Dadurch erhalten wir die sogenannte Plug-In Schätzung für die Bandbreite

$$\hat{h}_{p-i} := \min \left\{ \frac{s^2 \lambda_n}{\hat{\omega}_n} - \frac{s^4 \lambda_n \rho_n}{\hat{\omega}_n^2} + 3 \frac{s^4 \lambda_n^2 \hat{\nu}_n}{\hat{\omega}_n^3}, h_{max} \right\}, \quad (3.12)$$

wobei $\hat{\omega}_n := \bar{Y}^T P_n^T P_n \bar{Y}$ und $\hat{\nu}_n := \bar{Y}^T P_n^T Q_n \bar{Y}$. Der Plug-In Schätzer für die Regressionsfunktion lautet dann

$$\hat{f}_{p-i} := K(n, \hat{h}_{p-i}) \bar{Y}. \quad (3.13)$$

Bemerkung 3.2.1

Es gelte $E(\varepsilon_{i1}^4) < \infty$, $i = 1, \dots, m$ (**V8** ist nicht notwendigerweise erfüllt) und es sei $\bar{\varepsilon} := \bar{Y} - f$. Dann ist:

1. Nach Anwendung des Satzes von Abschnitt 27.2 in Cramér (1946), Seite 345-346

$\hat{\omega}_n = \omega_n + O_p(n^{-1/2})$ sowie $\hat{\nu}_n = \nu_n + O_p(n^{-1/2})$, denn

$$\begin{aligned} \text{Var}(\hat{\omega}_n) &= E(2\bar{\varepsilon}^T P_n^T P_n f + \bar{\varepsilon}^T P_n^T P_n \bar{\varepsilon} - \sigma^2 \delta_n)^2 \\ &= 4\sigma^2 f^T P_n^T P_n M_n P_n^T P_n f + 4E(\bar{\varepsilon}^T P_n^T P_n f \cdot \bar{\varepsilon}^T P_n^T P_n \bar{\varepsilon}) \\ &\quad + E(\bar{\varepsilon}^T P_n^T P_n \bar{\varepsilon})^2 + O(n^{-2}) \end{aligned}$$

$$\begin{aligned}
&\leq 4\sigma^2 f^T P_n^T P_n M_n P_n^T P_n f + 4\|f\| \|P_n\|^4 E \|\bar{\varepsilon}\|^3 \\
&\quad + \|P_n\|^4 E \|\bar{\varepsilon}\|^4 + O(n^{-2}) \\
&= O(n^{-1}) .
\end{aligned}$$

Ebenso ist $\text{Var}(\hat{\nu}_n) = O(n^{-1})$.

$$2. \quad s^2 = \sigma^2 + O_p(n^{-1/2}) \quad \text{und} \quad s^4 = \sigma^4 + O_p(n^{-1/2}) .$$

Somit ist, falls $\hat{h}_{p-i} < h_{max}$ (dies gilt übrigens mit einer Wahrscheinlichkeit, die für große n gegen 1 konvergiert),

$$\begin{aligned}
\hat{h}_{p-i} &= \frac{(\sigma^2 + O_p(n^{-1/2}))\lambda_n}{\omega_n + O_p(n^{-1/2})} - \frac{(\sigma^4 + O_p(n^{-1/2}))\lambda_n \rho_n}{(\omega_n + O_p(n^{-1/2}))^2} \\
&\quad + 3 \frac{(\sigma^4 + O_p(n^{-1/2}))\lambda_n^2 (\nu_n + O_p(n^{-1/2}))}{(\omega_n + O_p(n^{-1/2}))^3} \\
&= \frac{\sigma^2 \lambda_n}{\omega_n} - \frac{\sigma^4 \lambda_n \rho_n}{\omega_n^2} + 3 \frac{\sigma^4 \lambda_n^2 \nu_n}{\omega_n^3} + O_p(n^{-3/2}) .
\end{aligned}$$

Demzufolge ist wegen (3.8)

$$\frac{\hat{h}_{p-i} - h_{opt}}{h_{opt}} = O_p(n^{-1/2}) . \quad (3.14)$$

3.2.2 Die anderen adaptiven Schätzer

Bei den folgenden Adaptationsverfahren wird so vorgegangen, daß man die Funktion $M(h)$ der Gleichung (2.5) schätzt und diese Schätzung minimiert. Ist also $\hat{M}(h)$ ein Schätzer für $M(h)$ und ist

$$\hat{h} \in \arg \min_{h \in H} \hat{M}(h) = \{\bar{h} : \hat{M}(\bar{h}) = \min \hat{M}(\cdot)\} ,$$

so definiert man die entsprechende adaptive Schätzung durch

$$\hat{f}_{\hat{h}} := K(n, \hat{h}) \bar{Y} . \quad (3.15)$$

Die Menge $\arg \min_{h \in H} \hat{M}(h)$ ist nicht leer und es kann nach dem Lemma 2 von Jennrich (1969) $\hat{h}_n : \mathbb{R}^n \rightarrow [0, h_{max}]$ als meßbare Funktion gewählt werden, denn mit $y := (y_{11}, \dots, y_{1n_1}, \dots, y_{m1}, \dots, y_{mn_m})^T \in \mathbb{R}^n$ haben wir die Eigenschaften:

1. $\hat{M}(h) = \hat{M}(h, y)$ ist eine reellwertige Funktion auf $[0, h_{max}] \times \mathbb{R}^n$;
2. $\forall h \in [0, h_{max}]$ ist $\hat{M}(h, y)$ eine meßbare Funktion in y ;
3. $\forall y \in \mathbb{R}^n$ ist $\hat{M}(h, y)$ stetig in h .

In dieser Arbeit werden wir die folgenden Schätzer für $M(h)$ betrachten:

1- C_p -Schätzer (vgl. *Mallows (1973)*)

$$\hat{M}_{cp}(h) := \|[I_m - K(n, h)]\bar{Y}\|^2 + s^2 tr \{[2K(n, h) - I_m]M_n\} \quad (3.16)$$

Unter der Annahme normalverteilter Fehler ist $\hat{M}_{cp}(h)$ die beste erwartungstreue Schätzung für $M(h)$ (vgl. *Bunke und Droge (1984)*).

2- Cross-validation Schätzer (vgl. *Efron und Tibshirani (1993)*)

$$\hat{M}_{cv}(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{hi}^{-ij})^2, \quad (3.17)$$

wobei \hat{f}_{hi}^{-ij} das i -te Element des m -dimensionalen Vektors $\hat{f}_h^{-ij} := K(n-1, h)\bar{Y}^{-ij}$ ist und

$$\bar{Y}_k^{-ij} = \begin{cases} \bar{Y}_k & \text{falls } k \neq i \\ (n_i - 1)^{-1} \sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il} & \text{falls } k = i \end{cases} \quad (3.18)$$

3- Full Cross-validation Schätzer (1. Version)

$$\hat{M}_{fcv1}(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{hi}^{+ij})^2, \quad (3.19)$$

wobei \hat{f}_{hi}^{+ij} das i -te Element des m -dimensionalen Vektors $\hat{f}_h^{+ij} := K(n, h)\bar{Y}^{+ij}$ ist und für $k = 1, \dots, m$

$$\bar{Y}_k^{+ij} = \begin{cases} \bar{Y}_k & \text{falls } k \neq i \\ n_i^{-1} (\sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il} + \bar{Y}_i) & \text{falls } k = i \end{cases} \quad (3.20)$$

4- Full Cross-validation Schätzer (2. Version)

$$\hat{M}_{fcv2}(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{hi}^{+ij})^2,$$

wobei \hat{f}_{hi}^{+ij} wie oben, und für $k = 1, \dots, m$ ist

$$\bar{Y}_k^{+ij} = \begin{cases} \bar{Y}_k & \text{falls } k \neq i \\ n_i^{-1} \left(\sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il} + [K(n, h)\bar{Y}]_i \right) & \text{falls } k = i \end{cases} \quad (3.21)$$

Bemerkung 3.2.2

Statt, wie bei Cross-validation eine Beobachtung y_{ij} wegzulassen und die weggelassene Beobachtung „vorherzusagen“, ersetzt man bei Full Cross-validation die weggelassene Beobachtung y_{ij} durch den Mittelwert aller Beobachtungen im Punkt x_i (1. Version) oder durch die Schätzung der Regressionsfunktion f im Punkt x_i (2. Version) (vgl. Bunke, Droge und Polzehl (1999)).

Bemerkung 3.2.3

Wir verwenden hier eine gewichtete Modifizierung vom ursprünglichen Full Cross-validation von Bunke, Droge und Polzehl (1999) aus folgenden Gründen:

1. In der Situation, in der man Wiederholungen hat, bildet man die Mittelwerte der Quadrate der „Full Cross-validierten“ Residuen $Y_{ij} - \hat{f}_{hi}^{+ij}$ an jedem Versuchspunkt x_i . Wenn man die so gebildeten Mittelwerte summiert, führt das dazu, daß die Residuen an jedem Versuchspunkt die gleiche Größenordnung im Full Cross-validation Score (3.19) haben, unabhängig davon, ob sich die Stichprobenumfänge n_i möglicherweise stark voneinander unterscheiden können. Falls $n_i = n_j$ für alle i und j , ergibt (3.19) den Full Cross-validation Score von Bunke, Droge und Polzehl (1999).
2. Die Größen $(Y_i - \hat{f}_h^{+i}(x_i))^2$, $i = 1, \dots, m$, sind Schätzungen des quadratischen Vorhersagefehlers SEP_i (square error of prediction) einer Regressionsfunktion f im Versuchspunkt x_i , wenn keine Wiederholungen vorliegen. Hat man aber n_i Wiederholungen y_{ij} ($j = 1, \dots, n_i$) im Punkt x_i , so wird man SEP_i mit

$$\widehat{SEP}_i = n_i^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{hi}^{+ij})^2$$

schätzen.

Kapitel 4

Asymptotischer Vergleich der adaptiven Schätzer

4.1 Asymptotische Resultate 2. Ordnung

Dieser Abschnitt dient dazu, einen Überblick über die Ergebnisse 2. Ordnung zu geben, die in der Literatur zu finden sind. In *Neumann (1992)* wurde bewiesen, daß die Risikenapproximationen 2. Ordnung der adaptiven Schätzung (3.15), (wobei \hat{h} durch Minimierung von (3.16) entsteht) und der Plug-In Schätzung gleich sind. Der folgende Satz gibt die Gestalt der optimalen Bandbreite bei der Schätzung von f sowie die Risikoapproximation 2. Ordnung der Schätzung mit der optimalen Bandbreite an.

Satz 4.1.1 (Neumann (1992))

Unter den Voraussetzungen V1, V2, V5, V6, V7, den Gleichungen (2.6), (2.10), (3.1), (3.2) und (3.6) gilt :

$$1) \quad h_{opt} = \frac{\sigma^2 \lambda}{n \omega} + o(\sigma^2 n^{-1}) \quad (4.1)$$

$$2) \quad M(h_{opt}) = \sigma^2 \text{tr}(M_n) - \frac{\sigma^4 \lambda^2}{n^2 \omega} + o(\sigma^4 n^{-2}) . \quad (4.2)$$

Somit definieren wir die Plug-In Schätzung der Regressionsfunktion durch

$$\hat{f}_{p-i} := K(n, \hat{h}_{p-i}) \bar{Y} \quad , \quad (4.3)$$

wobei

$$\hat{h}_{p-i} := \min \left\{ \frac{s^2 \lambda_n}{\hat{\omega}_n}, h_{max} \right\} . \quad (4.4)$$

Im Satz 4.1.2 werden Approximationen 2. Ordnung des Risikos der Plug-In Schätzung von f angegeben. Es sei $\kappa := f^T P^T P C P f$.

Satz 4.1.2 (Neumann (1992))

Unter den Voraussetzungen **V1**, **V2**, **V5**, **V6**, **V7**, **V8** den Gleichungen (2.6), (2.10), (3.1), (3.2) und (3.6) gilt :

$$E\|\hat{f}_{p-i} - f\|^2 = \sigma^2 \text{tr}(M_n) - \frac{\sigma^4 \lambda^2}{n^2 \omega} + 4 \frac{\sigma^4 \lambda^2 \kappa}{n^2 \omega^2} + o(n^{-2}) \quad (4.5)$$

$$M_2(\hat{h}_{p-i}) = \sigma^2 \text{tr}(M_n) - \frac{\sigma^4 \lambda^2}{n^2 \omega} + 4 \frac{\sigma^4 \lambda^2 \kappa}{n^2 \omega^2} . \quad (4.6)$$

Das Glied $4\sigma^4 \lambda^2 \kappa n^{-2} \omega^{-2}$ ist der Preis, den man zahlen muß, wenn man f mit der Plug-In Bandbreite \hat{h}_{p-i} statt mit der optimalen Bandbreite h_{opt} schätzt. Für die Beispiele 1, 2, 3, 4 und 5 ist $\kappa \geq 0$. Also sind für diese Beispiele die Risiken 2. Ordnung der Plug-In Schätzung größer als die Risiken zweiter Ordnung der Schätzung mit der optimalen Bandbreite.

Es seien \hat{h}_{cp} , \hat{h}_{cv} , \hat{h}_{fcv1} , \hat{h}_{fcv2} die Bandbreiten, die man durch Minimierung der Schätzer $\hat{M}_{cp}(h)$, $\hat{M}_{cv}(h)$, $\hat{M}_{fcv1}(h)$, $\hat{M}_{fcv2}(h)$ als meßbare Funktion in den Beobachtungsvektor erhält.

Es seien außerdem

$$\begin{aligned} \hat{f}_{cp} &:= K(n, \hat{h}_{cp}) \bar{Y} \\ \hat{f}_{cv} &:= K(n, \hat{h}_{cv}) \bar{Y} \\ \hat{f}_{fcv1} &:= K(n, \hat{h}_{fcv1}) \bar{Y} \quad \text{und} \\ \hat{f}_{fcv2} &:= K(n, \hat{h}_{fcv2}) \bar{Y} \end{aligned}$$

die entsprechenden adaptiven Schätzungen von f . Es gilt der folgende

Satz 4.1.3

Unter den Voraussetzungen von Satz 4.1.2 und den Gleichungen (3.16) bis (3.21) gilt für $\hat{f} \in \{\hat{f}_{cp}, \hat{f}_{cv}, \hat{f}_{fcv1}, \hat{f}_{fcv2}\}$

$$E\|\hat{f} - f\|^2 = E\|\hat{f}_{p-i} - f\|^2 + o(n^{-2}) . \quad (4.7)$$

Für $\hat{f} = \hat{f}_{cp}$ ist (4.7) bereits von Neumann (1992) bewiesen worden. Die Beweise dieser Aussage für die anderen adaptiven Schätzer sowie alle anderen Beweise findet man im Kapitel 10.2.

4.2 Asymptotische Resultate 3. Ordnung

Die Aussagen des Abschnitts 4.1 stellen eine gewisse „asymptotische Äquivalenz“ der Adaptationsverfahren für große Stichprobenumfänge dar. Wenn

aber die Stichprobenumfangänge nicht so groß sind, können sich die Risiken der adaptiven Schätzungen unterscheiden.

Um die Unterschiede zwischen den Risiken der verschiedenen adaptiven Schätzer quantifizieren zu können (und somit einen noch präziseren Vergleich der Adaptationen vollführen zu können) ist es notwendig, die Risiken 3. Ordnung jeder der adaptiven Schätzungen zu berechnen. Wir werden den Vergleich der adaptiven Schätzungen so durchführen, daß wir ihre Risiken 3. Ordnung mit den Risiken 3. Ordnung der Plug-In Schätzung vergleichen. Die explizite Darstellung der Formel für die Risiken 3. Ordnung der Plug-In Schätzung wird später im Satz 4.2.5 angegeben.

Der Satz 4.2.1 vergleicht die Risiken 3. Ordnung der Plug-In Schätzung mit den Risiken 3. Ordnung der C_p -Schätzung von f .

Satz 4.2.1 (Ilouga (1994))

Unter den Voraussetzungen V1 bis V8, den Bezeichnungen (3.1) bis (3.7), den Gleichungen (3.12), (3.13) und (3.16) gilt:

$$E\|\hat{f}_{cp} - f\|^2 = E\|\hat{f}_{p-i} - f\|^2 + \frac{8\sigma^6\lambda\delta\kappa}{n^3\omega^3} + o(n^{-3}) \quad (4.8)$$

$$M_3(\hat{h}_{cp}) = M_3(\hat{h}_{p-i}) + \frac{8\sigma^6\lambda\delta\kappa}{n^3\omega^3}. \quad (4.9)$$

Bemerkung 4.2.1 *Eine Aussage, welche die bessere dieser beiden Adaptationsmethoden ist, hängt vom Vorzeichen von κ ab. Ist $K(n, h)$ die Glättungsmatrix der Nadaraya-Watson Kernschätzung mit Normalkern, so ist $\kappa \geq 0$, denn alle Eigenwerte der Matrix P_n in Beispiel 1 sind in diesem Fall größer oder gleich 0. Diese Behauptung gilt ebenfalls für die gewichtete kleinste Quadratschätzung mit Straffunktional, falls der Strafterm durch das Quadrat der euklidischen Norm des $(m-1)$ -dimensionalen Vektors der Differenzen 1. Ordnung $f^{(1)} := (f_2 - f_1, \dots, f_m - f_{m-1})^T$ oder des $(m-2)$ -dimensionalen Vektors der Differenzen 2. Ordnung $f^{(2)} := (f_2^{(1)} - f_1^{(1)}, \dots, f_{m-1}^{(1)} - f_{m-2}^{(1)})^T$ gegeben ist (vgl. Beispiele 2 und 3). Für die Beispiele 4 und 5 gilt ebenfalls $\kappa \geq 0$. Folglich ist die Plug-In Adaptation besser als die Adaptation durch Minimierung von \hat{M}_{cp} für diese genannten Beispiele.*

Der nachfolgende Satz vergleicht die Risiken von \hat{f}_{p-i} und \hat{f}_{cv} . Wir definieren zunächst folgende Größen:

$$\begin{aligned} \kappa_n &:= f^T P_n^T P_n M_n P_n f \\ \gamma_{1n} &:= 2 \left\{ \sum_{i=1}^m P_{n,ii}^2 n_i^{-2} (n_i - 1)^{-1} - (n - m)^{-1} \left(\sum_{i=1}^m P_{n,ii} n_i^{-1} \right)^2 \right\} \end{aligned}$$

$$\gamma_{2n} := 4\kappa_n \sum_{i=1}^m P_{n,ii} n_i^{-2}$$

$$\gamma_1 := 2 \left\{ \sum_{i=1}^m P_{ii}^2 c_i^{-3} - \lambda^2 \right\}$$

$$\gamma_2 := 4\kappa \sum_{i=1}^m P_{ii} c_i^{-2} .$$

Satz 4.2.2

Unter den Voraussetzungen **V1** bis **V8**, den Bezeichnungen (3.1) bis (3.7), den Gleichungen (3.16), (3.17) und (3.18) gilt:

$$E\|\hat{f}_{cv} - f\|^2 = E\|\hat{f}_{cp} - f\|^2 + \frac{\sigma^4}{n^3} \left(\frac{\gamma_1}{\omega} + \frac{\gamma_2}{\omega^2} \right) + o(n^{-3}) \quad (4.10)$$

$$M_3(\hat{h}_{cv}) = M_3(\hat{h}_{cp}) + \frac{\sigma^4}{n^3} \left(\frac{\gamma_1}{\omega} + \frac{\gamma_2}{\omega^2} \right) . \quad (4.11)$$

Korollar 4.2.1

Unter den Voraussetzungen der Sätze 4.2.1 und 4.2.2 gilt:

$$E\|\hat{f}_{cv} - f\|^2 = E\|\hat{f}_{p-i} - f\|^2 + \frac{8\sigma^6 \lambda \delta \kappa}{n^3 \omega^3} + \frac{\sigma^4}{n^3} \left(\frac{\gamma_1}{\omega} + \frac{\gamma_2}{\omega^2} \right) + o(n^{-3}) \quad (4.12)$$

$$M_3(\hat{h}_{cv}) = M_3(\hat{h}_{p-i}) + \frac{8\sigma^6 \lambda \delta \kappa}{n^3 \omega^3} + \frac{\sigma^4}{n^3} \left(\frac{\gamma_1}{\omega} + \frac{\gamma_2}{\omega^2} \right) . \quad (4.13)$$

Bemerkung 4.2.2

Aus Lemma 10.1.3 folgt, daß γ_1 und γ_2 größer oder gleich 0 sind. Aus diesem Grund ist die Adaptation mit Cross-validation in dritter Ordnung schlechter als die Plug-In Adaptation (Gleichung (4.13)) und schlechter als die Adaptation mit dem C_p -Kriterium (Gleichung (4.11)), falls κ positiv ist.

Der folgende Satz vergleicht die Adaptation durch Cross-validation mit der Adaptation durch Full Cross-validation (1. Version).

Satz 4.2.3

Unter den Voraussetzungen **V1** bis **V8**, den Bezeichnungen (3.1) bis (3.7), den Gleichungen (3.17) bis (3.20) gilt:

$$E\|\hat{f}_{fcv1} - f\|^2 = E\|\hat{f}_{cv} - f\|^2 - \frac{\sigma^4\gamma_2}{n^3\omega^2} + o(n^{-3}) \quad (4.14)$$

$$M_3(\hat{h}_{fcv1}) = M_3(\hat{h}_{cv}) - \frac{\sigma^4\gamma_2}{n^3\omega^2}. \quad (4.15)$$

Korollar 4.2.2

Unter den Voraussetzungen der Sätze 4.2.2 und 4.2.3 gilt:

$$E\|\hat{f}_{fcv1} - f\|^2 = E\|\hat{f}_{cp} - f\|^2 + \frac{\sigma^4\gamma_1}{n^3\omega} + o(n^{-3}) \quad (4.16)$$

$$M_3(\hat{h}_{fcv1}) = M_3(\hat{h}_{cp}) + \frac{\sigma^4\gamma_1}{n^3\omega}. \quad (4.17)$$

Aus der Gleichung (4.15) ist ersichtlich, daß die Adaptation mit Cross-validation (in dritter Ordnung) schlechter ist als die Adaptation mit Full Cross-validation (1. Version), die ihrerseits wiederum schlechter ist als die Adaptation mit dem C_p -Kriterium (Gleichung (4.17)), falls es ein Paar (i, j) gibt $(i, j \in \{1, \dots, m\})$ mit $P_{n,ii} \neq P_{n,jj}$ oder $n_i \neq n_j$. Sind alle Diagonalelemente von P_n gleich und $n_i = n_j \forall (i, j)$, so ist $\gamma_{1n} = 0$ und die Risiken 3.Ordnung der adaptiven Schätzung mit Full Cross-validation (1.Version) würden gleich den Risiken 3. Ordnung der adaptiven Schätzung mit dem C_p -Kriterium sein. Eine Situation, in der $\gamma_{1n} = 0$ ist, kommt beispielsweise vor, wenn die Regressionsfunktion f mit Schätzungen der Form $\hat{f}_c := c\bar{Y}$, $c \in (0, 1]$ geschätzt wird. Dabei ist P_n gleich der m -dimensionale Einheitsmatrix I_m . Im Kapitel 6 werden wir ausführlich dieses Beispiel behandeln.

Wodurch sich die Risiken 3. Ordnung der beiden Full Cross-validation Varianten unterscheiden, zeigt der

Satz 4.2.4

Unter den Voraussetzungen **V1** bis **V8**, den Bezeichnungen (3.1) bis (3.7) und den Gleichungen (3.19) bis (3.21) gilt:

$$E\|\hat{f}_{fcv2} - f\|^2 = E\|\hat{f}_{fcv1} - f\|^2 + T_{fcv} + o(n^{-3}) \quad (4.18)$$

$$M_3(\hat{h}_{fcv2}) = M_3(\hat{h}_{fcv1}) + T_{fcv}, \quad (4.19)$$

wobei

$$T_{fcv} = -\frac{8\sigma^4\lambda}{n^3\omega^2} \left(\frac{2f^T P^T P C P f f^T P^T C P f}{\omega} - f^T P^T C P C P f \right). \quad (4.20)$$

Bemerkung 4.2.3

Hat man $n_1 = \dots = n_m =: n$, so sind $C := mI_m$, $\lambda = m \operatorname{tr}(P)$,
 $\kappa = m f^T P^T P P f$. Damit ist

$$T_{fcv} = -\frac{8\sigma^4 \lambda \kappa}{n^3 \omega^2} = -\frac{8\sigma^4 \operatorname{tr}(P) f^T P^T P P f}{n^3 \omega^2} \leq 0$$

für die Beispiele 1, 2, 3, 4 und 5. Demzufolge ist für diese Beispiele die Adaptation mit Full Cross-validation (2. Version) in dritter Ordnung besser als die Adaptation mit der ersten Version.

Beim Vergleich der Adaptation durch Full Cross-validation (2. Version) mit der Adaptation durch dem C_p -Kriterium kann keine allgemeingültige Aussage getroffen werden wie bei den anderen Vergleichen. Es gilt nämlich

$$M_3(\hat{h}_{fcv2}) - M_3(\hat{h}_{cp}) = T_{fcv} + \frac{\sigma^4 \gamma_1}{n^3 \omega}.$$

Aber es liefern einige spezielle Situationen noch spezielle Ergebnisse.

Korollar 4.2.3

Es sei $n_1 = \dots = n_m = n$. Unter den Bedingungen von Korollar 4.2.2 und Satz 4.2.4 gilt:

$$E\|\hat{f}_{fcv2} - f\|^2 = E\|\hat{f}_{cp} - f\|^2 - \frac{8\sigma^4 \lambda \kappa}{n^3 \omega^2} + \frac{\sigma^4 \gamma_1}{n^3 \omega} + o(n^{-3}) \quad (4.21)$$

$$M_3(\hat{h}_{fcv2}) = M_3(\hat{h}_{cp}) - \frac{8\sigma^4 \lambda \kappa}{n^3 \omega^2} + \frac{\sigma^4 \gamma_1}{n^3 \omega}, \quad (4.22)$$

wobei γ_1 die Gestalt

$$\gamma_1 = 2 \left(\sum_{i=1}^m P_{ii}^2 - m^{-1} \left(\sum_{i=1}^m P_{ii} \right)^2 \right)$$

annimmt und λ bzw κ die spezielle Form wie in der Bemerkung 4.2.3 haben.

Aus Gleichung (4.22) sieht man, daß das Vorzeichen von $-\frac{8\sigma^4 \lambda \kappa}{n^3 \omega^2} + \frac{\sigma^4 \gamma_1}{n^3 \omega}$ darüber entscheidet, welche der beiden Adaptationen in dritter Ordnung besser ist. Sind alle Diagonalelemente von P_n und alle Stichprobenumfänge n_i gleich, dann ist $\gamma_1 = 0$. In diesem Fall ist die Full Cross-validation Adaptation (2. Version) besser als die C_p -Adaptation. Wie schon bemerkt, werden wir im 6. Abschnitt ein Beispiel mit $\gamma_1 = 0$ betrachten.

Korollar 4.2.4

Es sei $n_1 = \dots = n_m = n$. Unter den Bedingungen von Satz 4.2.1 und Korollar 4.2.3 gilt:

$$E\|\hat{f}_{fcv2} - f\|^2 = E\|\hat{f}_{p-i} - f\|^2 + \frac{8\sigma^4\lambda\kappa}{n^3\omega^2} \left(\frac{\sigma^2\delta}{\omega} - 1 \right) + \frac{\sigma^4\gamma_1}{n^3\omega} + o(n^{-3}) \quad (4.23)$$

$$M_3(\hat{h}_{fcv2}) = M_3(\hat{h}_{p-i}) + \frac{8\sigma^4\lambda\kappa}{n^3\omega^2} \left(\frac{\sigma^2\delta}{\omega} - 1 \right) + \frac{\sigma^4\gamma_1}{n^3\omega}. \quad (4.24)$$

Das Korollar 4.2.4 gibt an, welchen Preis man zu zahlen hat, wenn man anstelle der Plug-In Schätzung (deren Bandbreite die Konvergenzgeschwindigkeit $n^{-1/2}$ hat) die Full Cross-validation Adaptation (2. Version) für die Schätzung von f verwendet. Damit die Schätzung mit der Full Cross-validation Adaptation (2. Version) besser wird als die Plug-In Schätzung von f , ist es notwendig (aber nicht hinreichend), daß $\sigma^2 < \omega/\delta$ (falls κ positiv ist).

An dieser Stelle wird der Satz angegeben, in dem die Risiken 3. Ordnung der Plug-In Schätzung von f berechnet werden.

Wir bezeichnen die $(m_1 m_2 \times n_1 n_2)$ -Matrix $A \otimes B := (a_{ij} B)$ als das Tensorprodukt der $(m_1 \times n_1)$ -Matrix $A = (a_{ij})$ und der $(m_2 \times n_2)$ -Matrix $B = (b_{kl})$ und mit Ψ_0 die Matrix der vierten Momente einer m -dimensionalen standardnormalverteilte Zufallsgröße Z .

$$(Z \sim N_m(0, I_m) \quad \Psi_0 := E(ZZ^T \otimes ZZ^T)).$$

Es seien weiterhin:

$$\begin{aligned} \|Z\|_C^2 &:= Z^T C Z \\ \Omega &:= (C \otimes C)^{\frac{1}{2}} \Psi_0 (C \otimes C)^{\frac{1}{2}} \\ T_1 &:= \text{tr} [(P^T P \otimes P) \Omega] \\ T_{2f} &:= \text{tr} [(P^T P f f^T P^T P \otimes P) \Omega] \\ T_{3f} &:= \text{tr} [(P^T P f f^T P^T \otimes P^T P) \Omega] \\ T_{4f} &:= \text{tr} [(P^T P f f^T P^T P \otimes P^T P f f^T P) \Omega]. \end{aligned}$$

Dann gilt:

Satz 4.2.5

Es seien die Voraussetzungen **V1** bis **V8** erfüllt. Dann gilt für die Schätzung (3.13)

$$\begin{aligned} E\|\hat{f}_{p-i} - f\|^2 &= \sigma^2 \text{tr}(M_n) - \frac{\sigma^4 \lambda^2}{n^2 \omega} + 4 \frac{\sigma^4 \lambda \kappa}{n^2 \omega^2} - \frac{\sigma^6 \lambda^2 \delta}{n^3 \omega^2} \\ &+ \frac{2\sigma^4}{n^2(n-m)} \frac{\lambda^2}{\omega} + \frac{4\sigma^6 \lambda^2 \|P^T P f\|_C^2}{n^3 \omega^3} - \frac{8\sigma^6 \lambda \rho \kappa}{n^3 \omega^3} \\ &+ \frac{2\sigma^6 \lambda}{n^3 \omega^2} T_1 - \frac{8\sigma^6 \lambda}{n^3 \omega^3} (T_{2f} + T_{3f}) + \frac{16\sigma^6 \lambda}{n^3 \omega^4} T_{4f} \\ &- \frac{2\sigma^6 \lambda^3 \nu}{n^3 \omega^3} + \frac{2\sigma^6 \lambda^2 \mu}{n^3 \omega^2} - \frac{4\sigma^6 \lambda^2}{n^3 \omega^3} f^T P^T (3CP^T + 2PC) Q f + O(n^{-7/2}) \quad . \end{aligned}$$

Kapitel 5

Der Fall $Pf = 0$

In vorangegangenen Kapiteln haben wir stets angenommen, daß die Regressionsfunktion f nicht zum Nullraum von P_n gehört. Für solche Funktionen galten die asymptotischen Entwicklungen, die in den Sätzen der Kapitel 3 und 4 gegeben sind. Für die „sehr glatten“ Regressionsfunktionen f (d. h. für die $Pf = 0$ gilt) sind die Risikoapproximationen der Schätzer in den Kapiteln 3 und 4 nicht mehr geeignet, denn die Voraussetzung **V6** wird verletzt. Somit ist der Nenner der in den Formeln vorkommenden Ausdrücke gleich 0. Die Risikoapproximationen sind ebenfalls ungeeignet, wenn die quadratischen Formen $f^T P^T P f$ sehr klein sind.

In diesem Abschnitt wird ein Schätzer angegeben, mit dem man diese Schwierigkeit umgeht. Wir nehmen an, daß die Stichprobenumfänge alle gleich sind, d. h. $n_1 = \dots = n_m = n$.

Es seien $\mathcal{R}(P)$ der Bildraum und \mathcal{N}_0 der Nullraum von P . Wir bezeichnen mit \mathbb{P} die Projektionsmatrix auf $\mathcal{R}(P)$, d. h. $\mathbb{P}^2 = \mathbb{P}$, $\mathbb{P}f = f \quad \forall f \in \mathcal{R}(P)$ und $\mathbb{P}f := \arg \inf_{g \in \mathcal{R}(P)} \{\|g - f\|\}$. Hier ist die Projektion bezüglich der euklidischen Norm in \mathbb{R}^m zu verstehen. Wenn man sicher wüßte, daß die wahre Regressionsfunktion zum Nullraum von P gehört, dann würde man f beispielsweise mit der kleinsten Quadratschätzung unter linearer Restriktion

$$\hat{f}_0 = (I_m + M_n P^T P)^{-1} \bar{Y} \quad (5.1)$$

schätzen oder einen anderen Schätzer verwenden, der unter der Restriktion $Pf = 0$ gute Eigenschaften aufweist. In der Praxis weiß man anhand der Beobachtungen y_{ij} jedoch nicht, ob die zu schätzende Funktion f zum Raum \mathcal{N}_0 gehört oder nicht, wobei aber eine Vermutung vorliegen könnte, daß die Regressionsfunktion sehr glatt ist, also $f \in \mathcal{N}_0$ gilt. Für eine derartige Situation wird in *Bunke (1999)* der folgende Schätzer für f vorgeschlagen:

$$\hat{f}_c = c \mathbb{P} \bar{Y} + (I_m - \mathbb{P}) \bar{Y} \quad , \quad 0 \leq c \leq 1 . \quad (5.2)$$

Mit $h := 1 - c$ ist

$$\hat{f}_h = (I_m - h\mathbb{P})\bar{Y} \quad , \quad h \in [0, 1]. \quad (5.3)$$

Sei $l := \text{tr}(\mathbb{P}) = m - \dim(\mathcal{N}_0)$ und $h_{opt} = \arg \min_{h \in [0, 1]} M(h)$, wobei $M(h)$ das quadratische Risiko der Schätzung (5.3) ist. Es gilt der

Satz 5.0.6

Für die Schätzung (5.3) gilt:

$$h_{opt} = [1 + n\|\mathbb{P}f\|^2/(\sigma^2 l)]^{-1} \quad (5.4)$$

$$M(h_{opt}) = \frac{\sigma^2 m}{n} - \frac{\sigma^4}{n^2} \frac{l^2}{(\|\mathbb{P}f\|^2 + n^{-1}\sigma^2 l)} \quad . \quad (5.5)$$

Falls $f \in \mathcal{N}_0$, so sind

$$h_{opt} = 1 \quad \text{und} \quad M(h_{opt}) = \frac{\sigma^2}{n}(m - l) \quad \text{unabhängig von } f.$$

5.1 Adaptive Schätzungen

Selbst unter der Vermutung, daß die unbekannte Regressionsfunktion in \mathcal{N}_0 liegt, müßte man die Bandbreite datenabhängig wählen. Es werden hier zwei adaptive Schätzungen gebildet.

Durch Schätzung des Minimums h_{opt} von $M(h)$ in (5.4) erhält man die „Plug-In“ Bandbreite

$$\hat{h}_{p-i} = [1 + n\|\mathbb{P}\bar{Y}\|^2/(s^2 l)]^{-1} \quad , \quad (5.6)$$

die echt kleiner als 1 ist. Die entsprechende adaptive Schätzung ist

$$\hat{f}_{p-i} = (I_m - \hat{h}_{p-i}\mathbb{P})\bar{Y} \quad . \quad (5.7)$$

Andererseits erhält man durch Minimierung der Schätzung (3.16) von $M(h)$ die Schätzung der Bandbreite

$$\hat{h} = \frac{s^2 l}{n\|\mathbb{P}\bar{Y}\|^2} \quad , \quad (5.8)$$

die größer als 1 sein kann. Man wählt deshalb

$$\hat{h}_{cp} = \min(1, \hat{h}) \quad (5.9)$$

und

$$\hat{f}_{cp} = (I_m - \hat{h}_{cp}\mathbb{P})\bar{Y} \quad . \quad (5.10)$$

Bemerkung 5.1.1

Wenn man bei der Bildung der Plug-In Schätzung von (5.4) die erwartungstreue Schätzung $\|\widehat{\mathbb{P}f}\|^2 = \|\mathbb{P}\bar{Y}\|^2 - n^{-1}s^2l$ für $\|\mathbb{P}f\|^2$ einsetzt, erhält man $\hat{h}_{p-i} = \hat{h}$.

Bemerkung 5.1.2

Unter der Voraussetzung **V8** gilt:

$$Y_1 := m(n-1) \frac{s^2}{\sigma^2} \sim \chi_{m(n-1)}^2 \quad (5.11)$$

$$Y_2 := \frac{n}{\sigma^2} \|\mathbb{P}\bar{Y}\|^2 \sim \chi_{(l, n\sigma^{-2}f^T \mathbb{P}f)}^2, \quad (5.12)$$

wobei χ_k^2 und $\chi_{(k, \lambda)}^2$ die zentrale Chi-Quadratverteilung mit k Freiheitsgraden bzw. die nichtzentrale Chi-Quadratverteilung mit k Freiheitsgraden und Nichtzentralitätsparameter λ bezeichnen. Y_1 und Y_2 sind stochastisch unabhängig und es gilt

$$\begin{aligned} \hat{h} &= \frac{s^2 l}{n \|\mathbb{P}\bar{Y}\|^2} \\ &= \frac{\sigma^2 l}{m(n-1)} \frac{m(n-1) s^2 / \sigma^2}{n^{-1} \sigma^2 n \|\mathbb{P}\bar{Y}\|^2 / (n^{-1} \sigma^2)} \\ &= \frac{Y_1 / m(n-1)}{Y_2 / l} \sim F_{m(n-1), l, (0, n\sigma^{-2}f^T \mathbb{P}f)}, \end{aligned}$$

wobei $F_{k_1, k_2, (\lambda_1, \lambda_2)}$ die doppeltnichtzentrale F -Verteilung mit k_1 und k_2 Freiheitsgraden und Nichtzentralitätsparametern λ_1 und λ_2 ist. Es gilt für $l \geq 3$

$$E(\hat{h}) = l / (l - 2) > 1, \quad \text{falls } \|\mathbb{P}f\|^2 = 0, \quad (5.13)$$

und $E(\hat{h})$ hängt nicht von n ab.

Für beliebige Werte von $\|\mathbb{P}f\|^2$ erhalten wir (Thrum (1999))

$$\begin{aligned} E(\hat{h}) &= E\left(\frac{l}{Y_2}\right) \\ &= \sum_{k=0}^{\infty} p_k(\tau^2/2) \frac{l}{(l-2) + 2k}, \end{aligned} \quad (5.14)$$

wobei $\tau^2 := n\sigma^{-2}\|\mathbb{P}f\|^2$ ist, und $p_k(\lambda) := \frac{\lambda^k}{k!}e^{-\lambda}$ die Wahrscheinlichkeit für k bei einer Poisson Verteilung mit Parameter λ ist. Die Gleichung (5.13) folgt aus der Gleichung (5.14), wenn der Nichtzentralitätsparameter $n\sigma^{-2}\|\mathbb{P}f\|^2$ gleich 0 gesetzt wird.

Die Verteilungsfunktion F von $\hat{h}_{cp} := \min(\hat{h}, 1)$ ist gegeben durch

$$\begin{aligned} F(x) &= 0, \quad \text{falls } x \leq 0 \\ F(x) &= 1, \quad \text{falls } x > 1. \end{aligned}$$

und

$$\begin{aligned} F(x) &= P(\hat{h}_{cp} < x) \\ &= P(\hat{h}_{cp} < x, \hat{h} < 1) + P(\hat{h}_{cp} < x, \hat{h} > 1) \\ &= P(\hat{h} < x) \\ &= P\left(\frac{1}{\hat{h}} > \frac{1}{x}\right) \\ &= 1 - P\left(\frac{1}{\hat{h}} < \frac{1}{x}\right) \\ &= 1 - F_{l, m(n-1), (n\sigma^{-2}f^T\mathbb{P}f, 0)}(x^{-1}), \end{aligned}$$

für $0 < x \leq 1$.

Wir bezeichnen die Zufallsgröße $L(h) := \|\hat{f}_h - f\|^2$ als den quadratischen Verlust der Schätzung (5.3) von f .

Satz 5.1.1

Es sei $\mathbb{P}f = 0$. Dann gilt für alle Beobachtungsvektoren y

$$L(\hat{h}_{p-i}) > L(\hat{h}_{cp}). \quad (5.15)$$

Bemerkung 5.1.3 (mit Beweis)

Im Fall $\|\mathbb{P}f\| = 0$ ist die Risikofunktion der Schätzung (5.3) monoton fallend auf $[0, 1]$ (denn ihre Ableitung $M'(h) = -2\sigma^2 l(1-h)/n$ ist negativ) und es gilt $M'(1) = 0$ (Satz 5.0.6).

Aus den Gleichungen (5.6) und (5.8) folgt, daß jede Realisierung der Zufallsgröße \hat{h}_{p-i} nicht nur kleiner als 1 ist, sondern auch kleiner als die entsprechende Realisierung von \hat{h} ist. Daraus folgt, daß $\hat{h}_{p-i} < \hat{h}_{cp}$. Somit ist $M(\hat{h}_{cp}) < M(\hat{h}_{p-i})$ für alle Stichprobenumfänge n . Dies gilt dann auch für alle Funktionen einer gewissen ϵ -Umgebung ($\epsilon = \epsilon(n, \sigma) > 0$ genügend klein) $\mathcal{N}_\epsilon := \{f \in \mathbb{R}^m : \|\mathbb{P}f\| < \epsilon\}$ des Nullraumes von \mathbb{P} .

Aus Satz 5.1.1 ist schlußzufolgern, daß die adaptive Schätzung von f durch Minimierung des Kriteriums (3.16) zu verwenden ist, falls vermutet wird, daß die unbekannte Regressionsfunktion glatt ist. Die Plug-In Schätzung, die für den Fall unglatter Funktionen am besten war wird hier schlechter als der C_p -Schätzer.

5.2 Eine k -stufige Plug-In Schätzung

In diesem Abschnitt werden wir eine k -stufige Plug-In Schätzung der optimalen Bandbreite h_{opt} (Gleichung (5.4)) konstruieren.

Im Abschnitt 5.1 hatten wir die unbekannte Regressionsfunktion f der Gleichung (5.4) mit \bar{Y} und die Varianz σ^2 mit s^2 geschätzt, um die Plug-In Schätzung (5.6) zu definieren. Wir führen die Bezeichnungen

$$\hat{h}_{p-i}^{(0)} = 0 \quad \hat{f}_{p-i}^{(0)} = \bar{Y} \quad (5.16)$$

so ein, daß die Plug-In Schätzung der Bandbreite in der Gleichung (5.6) geschrieben werden kann als

$$\hat{h}_{p-i}^{(1)} = s^2 l / \left(n \|\mathbb{P} \hat{f}_{p-i}^{(0)}\|^2 + s^2 l \right) \quad (5.17)$$

$$= s^2 l / \left[n \|\mathbb{P} \bar{Y}\|^2 + s^2 l \right] . \quad (5.18)$$

Die entsprechende adaptive Schätzung von f wird

$$\hat{f}_{p-i}^{(1)} = \left(I_m - \hat{h}_{p-i}^{(1)} \mathbb{P} \right) \bar{Y} . \quad (5.19)$$

Die 2-stufige Plug-In Schätzung von f wird dadurch gebildet, daß man das unbekannte f in der Gleichung (5.4) mit $\hat{f}_{p-i}^{(1)}$ schätzt. Man erhält demnach

$$\hat{h}_{p-i}^{(2)} = s^2 l / \left(n \|\mathbb{P} \hat{f}_{p-i}^{(1)}\|^2 + s^2 l \right) \quad (5.20)$$

$$= s^2 l / \left[n \|\mathbb{P} \left(I_m - \hat{h}_{p-i}^{(1)} \mathbb{P} \right) \bar{Y}\|^2 + s^2 l \right] \quad (5.21)$$

$$= s^2 l / \left[n \left(1 - \hat{h}_{p-i}^{(1)} \right)^2 \|\mathbb{P} \bar{Y}\|^2 + s^2 l \right] \quad (5.22)$$

und

$$\hat{f}_{p-i}^{(2)} = \left(I_m - \hat{h}_{p-i}^{(2)} \mathbb{P} \right) \bar{Y} . \quad (5.23)$$

Aus Gleichung (5.22) folgt

$$\hat{h}_{p-i}^{(2)} > s^2 l / \left[n \|\mathbb{P}\bar{Y}\|^2 + s^2 l \right] = \hat{h}_{p-i}^{(1)} , \quad (5.24)$$

denn $0 < \hat{h}_{p-i}^{(1)} < 1$ und demzufolge $0 < (1 - \hat{h}_{p-i}^{(1)})^2 < 1$ gilt. Es gilt ebenfalls aus Gleichung (5.22), daß $\hat{h}_{p-i}^{(2)} < 1$.

Wir können also rekursiv eine Folge $\{\hat{h}_{p-i}^{(k)}\}_{k \geq 1}$ von Plug-In Bandbreiten wie folgt definieren:

$$\hat{h}_{p-i}^{(0)} = 0. \quad \text{Für } k \geq 1 \text{ ist}$$

$$\hat{f}_{p-i}^{(k-1)} = \left(I_m - \hat{h}_{p-i}^{(k-1)} \mathbb{P} \right) \bar{Y}$$

$$\hat{h}_{p-i}^{(k)} = s^2 l / \left(n \left\| \mathbb{P} \hat{f}_{p-i}^{(k-1)} \right\|^2 + s^2 l \right) \quad (5.25)$$

$$= s^2 l / \left[n \left(1 - \hat{h}_{p-i}^{(k-1)} \right)^2 \left\| \mathbb{P} \bar{Y} \right\|^2 + s^2 l \right] , \quad (5.26)$$

denn es gilt

$$\begin{aligned} \left\| \mathbb{P} \hat{f}_{p-i}^{(k-1)} \right\|^2 &= \left\| \mathbb{P} \left(I_m - \hat{h}_{p-i}^{(k-1)} \mathbb{P} \right) \bar{Y} \right\|^2 \\ &= \left(1 - \hat{h}_{p-i}^{(k-1)} \right)^2 \left\| \mathbb{P} \bar{Y} \right\|^2 . \end{aligned}$$

Diese Folge hat folgende Eigenschaften:

- (i) $\{\hat{h}_{p-i}^{(k)}\}_{k \geq 1}$ ist monoton wachsend. Dies folgt aus den Gleichungen (5.24), (5.26) und der Konstruktionsvorschrift.
- (ii) $\{\hat{h}_{p-i}^{(k)}\}_{k \geq 1}$ ist nach oben beschränkt: $\hat{h}_{p-i}^{(k)} < 1 \quad \forall k \geq 1$ (Gleichung (5.26)).
- (iii) 1 ist die kleinste obere Schranke von $\{\hat{h}_{p-i}^{(k)}\}_{k \geq 1}$, falls $\|\mathbb{P}f\| = 0$.

Beweis: angenommen sei, $\exists c \in (0, 1) : \forall k \geq 1 \quad \hat{h}_{p-i}^{(k)} < c$.

Es sei $\delta = 1 - c > 0$. Aus (5.6) und (5.8) gilt erstmal die Darstellung

$$\hat{h}_{p-i} = \hat{h} / (1 + \hat{h}) .$$

$$\begin{aligned}
\hat{h}_{p-i}^{(k)} < c &\Rightarrow \hat{h}_{p-i} < c, \quad (\text{denn } \{\hat{h}_{p-i}^{(k)}\}_{k \geq 1} \text{ ist monoton wachsend}) \\
&\Rightarrow \hat{h}/(1 + \hat{h}) < c \\
&\Rightarrow \hat{h} < c/(1 - c) = 1/\delta - 1
\end{aligned}$$

d.h. $\hat{h} \in (0, 1/\delta - 1)$. Das steht in Widerspruch mit der Tatsache, daß \hat{h} eine zentrale F -Verteilung mit $(l, m(n - 1))$ Freiheitsgraden besitzt, falls $\|Pf\| = 0$. Die Realisierungen einer solchen Zufallsgröße liegen nämlich auf der ganzen positiven reellen Achse $(0, \infty)$.

Aus (i) bis (iii) folgt, daß

$$p - \lim_{k \rightarrow \infty} \hat{h}_{p-i}^{(k)} = 1 = h_{opt}. \quad (5.27)$$

Es folgt aus dem Beweis von Satz 5.1.1 ($L(h)$ ist monoton fallend, falls $Pf = 0$), daß

$$L(\hat{h}_{p-i}^{(k)}) < L(\hat{h}_{p-i}^{(k-1)}) \quad f.s. \quad \forall k \geq 1 \quad (5.28)$$

und wegen der Stetigkeit von $L(h)$ in h , daß

$$p - \lim_{k \rightarrow \infty} L(\hat{h}_{p-i}^{(k)}) = L(1). \quad (5.29)$$

Aus der Gleichung (5.28) folgt, daß ab einer gewissen Anzahl k_0 von Plug-In Stufen die Schätzung $\hat{f}_{p-i}^{(k_0)}$ besser wird als die Schätzung \hat{f}_{cp} , falls die Regressionsfunktion sehr glatt ist. Mit einem Argument, daß auf dieselbe Überlegung basiert existiert eine Anzahl k_ϵ von Plug-In Stufen, so daß die Schätzung $\hat{f}_{p-i}^{(k_\epsilon)}$ besser ist als die C_p Schätzung, falls $f \in \mathcal{N}_\epsilon$ ist ($\epsilon = \epsilon(n, \sigma)$ sehr klein).

Kapitel 6

Die Stauchung des Mittelwertvektors

Wir leiten dieses Kapitel mit dem Beispiel der Messung einer Naturkonstante ein. Es sei eine unbekannte Größe μ , die zu schätzen ist. Hat man dafür eine Stichprobe (Messungen) $\{Y_i\}$ vom Umfang n und setzt man das Modell

$$Y_i = \mu + \varepsilon_i \quad i = 1, \dots, n \quad , \quad E \varepsilon_i = 0 \quad Var \varepsilon_i = \sigma^2$$

voraus, dann erhält man mit der kleinsten Quadrat Methode die Schätzung

$$\hat{\mu} = \bar{Y} := n^{-1} \sum_{i=1}^n Y_i \tag{6.1}$$

für μ .

Wenn aber der unbekannte Parameter μ in einem Intervall der Form $B_a = \left\{ \left| \frac{\mu}{\sigma} \right|^2 \leq a \right\}$ liegt, dann ist die Schätzung

$$\hat{\mu}_a := \frac{an}{1+an} \bar{Y} = c \bar{Y} \tag{6.2}$$

für μ minimax in der Klasse B_a und linear in \bar{Y} . Die Konstruktion solcher minimax linearer Schätzungen in multivariaten Regressionsmodellen unter Parameterrestriktionen in Form von Ellipsoiden wird in *Bunke (1999)* ausführlich vorgestellt und diskutiert.

Das Ziel dieses Kapitels ist die datenabhängige Wahl der Konstante bei einer Stauchung.

Wir betrachten jetzt im Regressionsmodell (2.1) die Stauchungen des Mittelwertvektors der Gestalt:

$$\hat{f}_c := c \bar{Y} \quad , \quad \text{wobei} \quad 0 < c \leq 1 \quad . \tag{6.3}$$

6.1 Schätzung mit dem optimalen Parameter

Es sei c_{opt} das (wahre nicht zufällige und unbekannte) $c \in (0, 1]$ mit $c_{opt} = \arg \min_{c \in (0, 1]} M(c)$, wobei

$$M(c) := E \|\hat{f}_c - f\|^2, \quad (6.4)$$

und es sei $t_n := \text{tr}(M_n)$. Die folgende Aussage gibt die Form von c_{opt} und $M(c_{opt})$ an.

Aussage 6.1.1

$$1) \quad c_{opt} = 1 - \sigma^2 t_n (f^T f + \sigma^2 t_n)^{-1} \quad (6.5)$$

$$2) \quad M(c_{opt}) = \sigma^2 t_n \left[1 - \frac{\sigma^2}{n} \frac{\text{tr}(C) f^T f}{(f^T f + \frac{\sigma^2}{n} \text{tr}(C))^2} - \frac{\sigma^4}{n^2} \frac{\text{tr}(C)^2}{(f^T f + \frac{\sigma^2}{n} \text{tr}(C))^2} \right] \quad (6.6)$$

Wir sehen, daß der Ausdruck in eckigen Klammern kleiner oder gleich 1 für grosse n ist. Demzufolge verbessert $\hat{f}_{c_{opt}}$ die kleinste Quadratschätzung \bar{Y} . Damit wir die Bezeichnungen möglichst vereinheitlichen, setzen wir:

$$h := 1 - c, \quad h \in [0, 1)$$

$$h_{opt} := 1 - c_{opt} = \frac{\sigma^2 t_n}{f^T f + \sigma^2 t_n} \quad (6.7)$$

$$\hat{f}_h := (I_m - h I_m) \bar{Y}. \quad (6.8)$$

h wird weiterhin als Glättungsparameter bezeichnet und die Matrix $I_m - h I_m$ als Glättungsmatrix.

6.2 Die adaptiven Schätzungen

1. Der Plug-In Schätzer

Wir schätzen die optimale Bandbreite h_{opt} in (6.7) dadurch, daß wir f durch \bar{Y} ersetzen. Die Varianz wird mit

$$\hat{\sigma}^2 := \alpha s^2 + \gamma \bar{Y}^T \Gamma_n \bar{Y} \quad \alpha \in (0, 1] \quad \text{und} \quad \gamma \geq 0 \quad (6.9)$$

geschätzt, wobei Γ_n eine $m \times m$ -Matrix bezeichnet, die symmetrisch und positiv semidefinit ist. Die minimax-Eigenschaften solcher Varianzschätzungen in der Klasse aller quadratischen bedingt erwartungstreuen Schätzungen von σ^2 wurden in *Bunke (1999)* diskutiert. Ist nämlich die Regressionsfunktion sehr glatt im Sinne von $\Gamma_n f = 0$, so ist die Schätzung (6.9) besser als s^2 für $\alpha \in (0, 1)$ (besser im Sinne des mittleren quadratischen Fehlers). Wir definieren jetzt

$$\hat{h}_{p-i} := \frac{\hat{\sigma}^2 t_n}{\|\bar{Y}\|^2 + \hat{\sigma}^2 t_n} \quad (6.10)$$

$$\hat{f}_{p-i} := [I_m - \hat{h}_{p-i} I_m] \bar{Y} . \quad (6.11)$$

Analog zum Abschnitt 3.2.2 werden folgende Schätzer $\hat{M}(h)$ betrachtet:

2. C_p -Schätzer

$$\hat{M}_{cp}(h) = h^2 \|\bar{Y}\|^2 + \hat{\sigma}^2 (1 - 2h) t_n \quad (6.12)$$

3. Cross-validation Schätzer

$$\hat{M}_{cv}(h) = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{hi}^{-ij})^2 \quad (6.13)$$

wobei \hat{f}_{hi}^{-ij} das i -te Element des m -dimensionalen Vektors $\hat{f}_h^{-ij} = (1 - h) \bar{Y}^{-ij}$ ist und

$$\bar{Y}_k^{-ij} = \begin{cases} \bar{Y}_k & \text{falls } k \neq i \\ (n_i - 1)^{-1} \sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il} & \text{falls } k = i \end{cases} \quad (6.14)$$

4. Full Cross-validation Schätzer (1. Version)

$$\hat{M}_{fcv1}(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{f}_{hi}^{+ij})^2 , \quad (6.15)$$

wobei \hat{f}_{hi}^{+ij} das i -te Element des m -dimensionalen Vektors $\hat{f}_h^{+ij} = (1-h)\bar{Y}^{+ij}$ ist und für $k = 1, \dots, m$

$$\bar{Y}_k^{+ij} = \begin{cases} \bar{Y}_k & \text{falls } k \neq i \\ n_i^{-1} \left(\sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il} + \bar{Y}_i \right) & \text{falls } k = i \end{cases} \quad (6.16)$$

5. Full Cross-validation Schätzer (2. Version)

Wie die 1. Version aber diesmal

$$\bar{Y}_k^{+ij} = \begin{cases} \bar{Y}_k & \text{falls } k \neq i \\ n_i^{-1} \left[\sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il} + (1-h)\bar{Y}_i \right] & \text{falls } k = i \end{cases} \quad (6.17)$$

Die folgenden Aussagen sind Schlußfolgerungen aus den Sätzen von Kapitel 4, denn die Glättungsmatrix $I_m - hI_m$ ist ein Spezialfall der Matrix $K(n, h)$ aus Kapitel 2.2. Die Bezeichnungen haben hierbei eine spezielle Gestalt:

$$\begin{aligned} \omega_n &= \omega = f^T f \quad \kappa = f^T C f \\ \rho &= \lambda = t := \lim_{n \rightarrow \infty} n t_n \\ \gamma_1 &= 2\{tr(C^3) - t^2\} =: \tau_1 \geq 0 \quad \gamma_2 = 4tr(C^2)f^T C f =: \tau_2 \geq 0. \end{aligned}$$

In der Aussage 6.2.1 werden die Risiken der Plug-In und der C_p -Schätzungen verglichen. Sei

$$\alpha_0 := \left[1 - 4 \frac{f^T C f}{f^T f tr(C)} \right].$$

Aussage 6.2.1

Es sei $f \neq 0$. Unter den Voraussetzungen **V7**, **V8** und den Gleichungen (6.9), (6.10), (6.11), (6.12) gilt:

- 1) $M_3(\hat{h}_{p-i}) = M_3(\hat{h}_{cp}) - \frac{2\alpha^2 \sigma^6 t^2}{n^3 (f^T f)^2} \left[\frac{4f^T C f}{f^T f} - (1-\alpha)tr(C) \right]$
- 2) Ist $\frac{\max_{1 \leq i \leq m} \{n_i\}}{\min_{1 \leq i \leq m} \{n_i\}} < m/4$ (gilt beispielsweise, wenn die Stichproben-

umfänge n_i alle gleich sind und $m \geq 5$ ist) so gilt :

$$M_3(\hat{h}_{p-i}) \leq M_3(\hat{h}_{cp}) \quad \forall \alpha \in I_{\alpha_0} := [\alpha_0, 1].$$

$$3) \quad \arg \max_{\alpha \in I_{\alpha_0}} \{M_3(\hat{h}_{cp}) - M_3(\hat{h}_{p-i})\} = 1 \quad \text{und für } \alpha = 1 \quad \text{ist}$$

$$M_3(\hat{h}_{cp}) = M_3(\hat{h}_{p-i}) + \frac{8\sigma^6 t^2 f^T C f}{n^3 \|f\|^6}.$$

Die nächsten Aussagen werden demzufolge für $\hat{\sigma}^2 = s^2$ angegeben. In (6.9) setzen wir also $\alpha = 1$ und $\gamma = 0$, und erhalten damit Terme, deren Deutung einfacher ist.

Aussage 6.2.2

Es sei $f \neq 0$. Unter den Voraussetzungen **V7**, **V8** und den Gleichungen (6.8) und (6.10) bis (6.17) sind:

$$\begin{aligned} 1) \quad M_3(\hat{h}_{cv}) &= M_3(\hat{h}_{p-i}) + \frac{8\sigma^6 t^2 f^T C f}{n^3 (f^T f)^3} + \frac{\sigma^4}{n^3} \left(\frac{\tau_1}{f^T f} + \frac{\tau_2}{(f^T f)^2} \right) \\ &= M_3(\hat{h}_{cp}) + \frac{\sigma^4}{n^3} \left(\frac{\tau_1}{f^T f} + \frac{\tau_2}{(f^T f)^2} \right) \end{aligned}$$

$$\begin{aligned} 2) \quad M_3(\hat{h}_{fcv1}) &= M_3(\hat{h}_{cv}) - \frac{\sigma^4 \tau_2}{n^3 (f^T f)^2} \\ &= M_3(\hat{h}_{cp}) + \frac{\sigma^4 \tau_1}{n^3 f^T f} \end{aligned}$$

$$3) \quad M_3(\hat{h}_{fcv2}) = M_3(\hat{h}_{fcv1}) - \frac{8\sigma^4 t}{n^3 (f^T f)^2} \left(\frac{2(f^T C f)^2}{f^T f} - f^T C^2 f \right)$$

$$4) \quad M_3(\hat{h}_{fcv2}) = M_3(\hat{h}_{cp}) + \frac{\sigma^4 \tau_1}{n^3 f^T f} - \frac{8\sigma^4 t}{n^3 (f^T f)^2} \left(\frac{2(f^T C f)^2}{f^T f} - f^T C^2 f \right)$$

$$5) \quad M_3(\hat{h}_{fcv2}) = M_3(\hat{h}_{p-i}) + \frac{8\sigma^4 \text{tr}(C) f^T C f}{n^3 (f^T f)^2} \left(\frac{\sigma^2 \text{tr}(C)}{f^T f} - 1 \right).$$

Bemerkung 6.2.1

Falls $n_1 = \dots = n_m =: n$ so sind $C = mI_m$, $\tau_1 = 0$, $t = m^2$ und $\tau_2 = 4m^4 f^T f$. Damit haben wir

$$\begin{aligned}
 1)' \quad M_3(\hat{h}_{cv}) &= M_3(\hat{h}_{cp}) + \frac{4m\sigma^4}{n^3 f^T f} \\
 2)' \quad M_3(\hat{h}_{fcv1}) &= M_3(\hat{h}_{cp}) \\
 3)' \quad M_3(\hat{h}_{fcv2}) &= M_3(\hat{h}_{fcv1}) - \frac{8m\sigma^4}{n^3 f^T f} \\
 4)' \quad M_3(\hat{h}_{fcv2}) &= M_3(\hat{h}_{cp}) - \frac{8m\sigma^4}{n^3 f^T f} \\
 5)' \quad M_3(\hat{h}_{fcv2}) &= M_3(\hat{h}_{p-i}) + \frac{8\sigma^4}{n^3 (f^T f)^2} \left(\frac{\sigma^2 m^2}{f^T f} - 1 \right).
 \end{aligned}$$

In dieser Situation sind die Risiken 3. Ordnung der Schätzung mit der Full Cross-validation Adaptation (2. Version) kleiner als die Risiken 3. Ordnung des C_p -Schätzers, und diese sind gleich den Risiken 3. Ordnung der Schätzung mit der Full Cross-validation Adaptation (1. Version). Zur Gleichung 5)' ist folgendes zu bemerken:

Ist $\sigma^2 < m^{-2} f^T f$, so ist die Full Cross-validation Adaptation (2. Version) besser als die Plug-In Adaptation. Ist dagegen die Varianz der Beobachtungen „groß“ ($\sigma^2 > m^{-2} f^T f$), so ist die Plug-In Adaptation besser.

Analog zum Satz 4.2.5 seien

$$\begin{aligned}
 T_1 &:= \text{tr}[\Omega] \\
 T_{2f} &:= \text{tr}[(f f^T \otimes I_m)\Omega] \\
 T_{3f} &:= \text{tr}[(f f^T \otimes f f^T)\Omega].
 \end{aligned}$$

Aussage 6.2.3

Sei $f \neq 0$. Unter den Voraussetzungen **V7**, **V8** und der Gleichung (6.10) gilt für die Schätzung (6.11):

$$\begin{aligned}
 M_3(\hat{h}_{p-i}) &= \sigma^2 \text{tr}(M_n) - \frac{\sigma^4 t^2}{n^2 f^T f} + 4 \frac{\sigma^4 t f^T C f}{n^2 (f^T f)^2} - \frac{\sigma^6 t^3}{n^3 (f^T f)^2} \\
 &+ \frac{2\sigma^4}{n^2(n-m)} \frac{t^2}{f^T f} - 4 \frac{\sigma^6 t^2 f^T C f}{n^3 (f^T f)^3} + 2 \frac{\sigma^6 t}{n^3 (f^T f)^2} T_1 \\
 &- 16 \frac{\sigma^6 t}{n^3 (f^T f)^3} T_{2f} + 16 \frac{\sigma^6 t}{n^3 (f^T f)^4} T_{3f}.
 \end{aligned}$$

Kapitel 7

Verbesserung des Risikos zweiter Ordnung

In diesem Kapitel werden wir eine adaptive Schätzung für f angeben, welche die Approximation zweiter Ordnung des Risikos der bisher definierten adaptiven Schätzungen unter normalverteilten Beobachtungsfehlern verbessert.

Im Abschnitt 4.1, Satz 4.1.3 wurde bereits erwähnt, daß die Risiken 2. Ordnung der Schätzung \hat{f}_{p-i} und die Risiken 2. Ordnung der Schätzung \hat{f}_{cp} gleich sind.

Der neue Plug-In Schätzer

Um den Plug-In Schätzer der Regressionsfunktion zu konstruieren, setzen wir in (4.4) eine Varianzschätzung der Form (6.9) mit $\gamma = 0$ ein:

$$\hat{\sigma}^2 = \alpha s^2. \quad (7.1)$$

Diese Varianzschätzung ist zwar verfälscht, aber ihr mittlerer quadratischer Fehler ist (für gewisse Werte α) kleiner als die Varianz von s^2 (vgl. *Bunke (1999)*). Es seien

$$\tilde{h}_\alpha := \min\{\hat{\omega}_n^{-1}\alpha s^2\lambda_n, h_{\max}\} \quad (7.2)$$

$$\tilde{f}_\alpha := K(n, \tilde{h}_\alpha)\bar{Y} \quad (7.3)$$

Die Berechnung des Risikos von \tilde{f}_α ergibt:

$$E\|\tilde{f}_\alpha - f\|^2 = \sigma^2 \text{tr}(M_n) + \frac{\alpha^2 \sigma^4 \lambda^2}{n^2 \omega} - \frac{2\alpha \sigma^4 \lambda^2}{n^2 \omega} + \frac{4\alpha \sigma^4 \lambda \kappa}{n^2 \omega^2} + o(n^{-2}) \quad (7.4)$$

$$M_2(\tilde{h}_\alpha) = \sigma^2 \text{tr}(M_n) + \frac{\alpha^2 \sigma^4 \lambda^2}{n^2 \omega} - \frac{2\alpha \sigma^4 \lambda^2}{n^2 \omega} + \frac{4\alpha \sigma^4 \lambda \kappa}{n^2 \omega^2}. \quad (7.5)$$

Dadurch sieht man Folgendes:

1. Ist α zu klein (d. h. liegt α in der Nähe von 0), so gibt (7.5) das Risiko von \bar{Y} an. Dieses Risiko ist größer als das der adaptiven Schätzung.
2. Ist α zu groß (d. h. liegt α in der Nähe von 1), so gibt (7.5) das Risiko zweiter Ordnung (Gleichung (4.6)) an, das gerade mit Hilfe von α verkleinert werden soll.

Wir nehmen also das „asymptotisch optimale“ α , das (7.5) minimiert,

$$\alpha_{opt} = 1 - \frac{2\kappa}{\lambda\omega} \quad (7.6)$$

und schätzen α durch

$$\tilde{\alpha}_n := 1 - \frac{2\bar{Y}^T P_n^T P_n M_n P_n \bar{Y}}{\lambda_n \|P_n \bar{Y}\|^2}. \quad (7.7)$$

Wir definieren nun

$$\tilde{h}_{\tilde{\alpha}_n} := \min\{\hat{\omega}_n^{-1} \tilde{\alpha}_n s^2 \lambda_n, h_{\max}\} \quad (7.8)$$

und

$$\tilde{f}_{\tilde{\alpha}_n} := K(n, \tilde{h}_{\tilde{\alpha}_n}). \quad (7.9)$$

Bemerkung 7.0.2

Für $n_1 = \dots = n_m =: n$ ist

$$\tilde{\alpha}_n := 1 - \frac{2\bar{Y}^T P_n^T P_n P_n \bar{Y}}{\text{tr}(P_n) \|P_n \bar{Y}\|^2}.$$

Ist jetzt P_n die Matrix von Beispiel 1, so folgt aus dem Kreisesatz von Geršgorin (vgl. Kielbasinski und Schwetlick (1988), Seite 39), daß die Eigenwerte λ_i , $i = 1, \dots, m$, von P_n auf einem der Kreise

$$\mathcal{K}_1 := \{\lambda : |\lambda - 1| \leq 1\} \quad \text{oder}$$

$$\mathcal{K}_2 := \{\lambda : |\lambda - 2| \leq 2\}$$

liegen, d. h. $\lambda_{\max}(P_n) \leq 4$. Somit ist

$$\begin{aligned} \sup_{\bar{Y} : P_n \bar{Y} \neq 0} \frac{2\bar{Y}^T P_n^T P_n P_n \bar{Y}}{\text{tr}(P_n) \|P_n \bar{Y}\|^2} &= \sup_{\bar{Y} : P_n \bar{Y} \neq 0} \frac{2\bar{Y}^T P_n^T P_n P_n \bar{Y}}{2(m-1) \|P_n \bar{Y}\|^2} \\ &= \frac{4}{m-1} \leq 1 \quad \text{genau dann, wenn } m \geq 5. \end{aligned}$$

In diesem Fall liegen die Realisierungen von $\tilde{\alpha}_n$ zwischen 0 und 1.

Für die adaptive Schätzung $\tilde{f}_{\tilde{\alpha}_n}$ gilt der

Satz 7.0.1

*Es sei $2\kappa < \lambda\omega$. Unter den Voraussetzungen **V1**, **V2**, **V5** bis **V8** und den Gleichungen (2.6) und (2.10) ist:*

$$M_2(\tilde{h}_{\tilde{\alpha}_n}) < M_2(\hat{h}_{p-i}). \quad (7.10)$$

Bemerkung 7.0.3

Für $m \geq 5$ erfüllen die Nadaraya-Watson Kernschätzung und die gewichtete kleinste Quadratschätzung mit Straffunktional die Bedingung $2\kappa < \lambda\omega$ (siehe Bemerkung 7.0.2). Nach dem Satz 4.1.3 gilt für große n , daß $M_2(\hat{h}_{p-i}) = M_2(\hat{h}_{crit})$, wobei \hat{h}_{crit} eins der Kriterien (3.16) bis (3.21) minimiert. Somit gilt auch $M_2(\tilde{h}_{\tilde{\alpha}_n}) < M_2(\hat{h}_{crit})$. Für ausreichend große Stichprobenumfänge (so groß, daß $M_2(\hat{h}_{p-i}) \approx M_3(\hat{h}_{p-i})$), wird somit auch $M_2(\tilde{h}_{\tilde{\alpha}_n}) < M_3(\hat{h}_{p-i})$ sein.

Kapitel 8

Numerische Resultate

8.1 Der Fall glatter Funktionen

Hier werden wir untersuchen, wie sich der Plug-In, der C_p , die k -stufigen Plug-In ($k = 2, 3$) als adaptive Schätzungen und im Vergleich auch der nicht-geglätete Projektionsschätzer einer glatten Regressionsfunktion f für kleine bis mittlere Stichprobenumfänge verhalten. Wir untersuchen zugleich wie sich diese Schätzer verhalten, wenn der Abstand der zu schätzende Regressionsfunktion zur Menge der sehr glatten Funktionen langsam vergrößert (Abbildungen 8.4 bis 8.16).

Dafür betrachten wir die Matrix P von Beispiel 1 und bezeichnen mit \mathbb{P} die Projektionsmatrix auf dem Bildraum von P . Aus der Bemerkung 2.2.1 wissen wir, daß sehr glatte Funktionen bezüglich P die konstanten Funktionen sind, und die glatten Funktionen sind Geraden $f(x) = ax + b$, wobei a klein und b eine beliebige reelle Zahl sind. Demzufolge sind die Funktionen f , für die $\mathbb{P}f = 0$ ist, ebenfalls konstante Funktionen.

Die Anzahl der Versuchspunkte ist $m = 10$. Diese 10 Versuchspunkte sind äquidistant im Intervall $[0, 1]$. Die Varianz σ^2 ist gleich 1. Die Stichprobenumfänge $n_1 = \dots = n_{10} = n$ sind von $n = 20$ bis $n = 130$ mit der Schrittweite 10. Die Regressionsfunktionen f sind Geraden $f(x) = ax + 1$, wobei der Anstieg a langsam erhöht wird, um zu sehen, wie sich die entsprechenden Schätzungen verhalten. Die Erhöhung des Anstiegs a bedeutet die Vergrößerung der Umgebung \mathcal{N}_ϵ der Menge \mathcal{N}_0 der sehr glatten Funktionen. Es werden die Regressionsfunktionen mit $a = 0, 1/8, 2/8, 3/8, 4/8, 5/8, 6/8, 8/8, 12/8$ und $16/8$ betrachtet.

Das Ziel ist die Schätzung der Größe $D_n := E \|\tilde{f} - f\|^2 / M(h_{opt})$ für verschiedene n , wobei \tilde{f} jeweils für die Schätzungen $\hat{f}_{p-i}, \hat{f}_{cp}$, die Projek-

tionsschätzung \hat{f} , die Schätzungen $\hat{f}_{p-i}^{(k)}$ ($k = 2, 3$) steht (vgl. Kapitel 5). Die Berechnung dieser Größe D_n hat als Ziele, die adaptiven Schätzungen untereinander zu vergleichen und gleichzeitig der Vergleich der adaptiven Schätzung mit der optimalen Schätzung zu vollführen. Die Anzahl der Simulationen für die Berechnung von D_n ist $N_{sim} = 5000$. Die Ergebnisse der Simulation sind in den Abbildungen 8.3 bis 8.16 dargestellt. Dabei kennzeichnen pi , cp *proj*, $pi2s$ und $pi3s$ die Größe D_n für die Plug-In Schätzung, die C_p Schätzung, den Projektionsschätzer, die 2-stufige bzw. 3-stufige Plug-In Schätzung von f . Die Größe $glat$ in den Abbildungen ist $\|\mathbb{P}f\|^2$. Ab der Abbildung 8.7 ($a = 1/2$) ist die Kurve für die Projektionsschätzung nicht mehr aufgetragen, da die Werte D_n dabei sehr groß werden. In den Tabellen 8.1 und 8.2 stehen bei $a = 0$ bzw. $a = 1/8$ die Werte von D_n , wenn \tilde{f} die kleinste Quadrat Schätzung \bar{Y} (lse), der Projektionsschätzer (proj) oder \tilde{f} einer der oben erwähnten adaptiven Schätzer ist.

Die folgenden Bemerkungen sind zu machen:

1. Für $a = 0$ ist der Projektionsschätzer gleichmäßig besser als die adaptiven Schätzer. Der C_p -Schätzer ist seinerseits besser für alle n als der Plug-In und 2-stufige Plug-In Schätzer, aber schlechter als der 3-stufige Plug-In Schätzer (Abbildung 8.3). Für $a = 1/8$ bleibt der Projektionsschätzer immer noch gut und die 3-stufige Plug-In adaptive Schätzung wird etwas besser (Abbildung 8.4).
2. Ab $a = 1/4$ ist der Projektionsschätzer nicht mehr geeignet. Er wird sogar schlechter für größer werdende n (Abbildungen 8.5 und 8.6). Die Schätzung $\hat{f}_{p-i}^{(3)}$ wird zunehmend schlechter als \hat{f}_{p-i} und $\hat{f}_{p-i}^{(2)}$ (ab der Abbildung 8.6).
3. Je größer a wird, d. h. je unglatter die Regressionsfunktion wird, desto besser wird die Schätzung \hat{f}_{p-i} im Vergleich zur Schätzung \hat{f}_{cp} bei wachsenden n (Abbildungen 8.6 und 8.7) und desto schlechter werden die k -stufigen Plug-In Schätzungen ($k > 1$). Ab $a = 5/8$ (f unglatt) ist der Plug-In Schätzer gleichmäßig besser als der C_p -Schätzer (Abbildungen 8.8 bis 8.12).
4. Bei wachsender Unglattheit der Regressionsfunktion werden die C_p und Plug-In Kriterien besser (d.h. die Werte von D_n nähern sich an die 1). Die Schätzung \hat{f}_{p-i} wird dabei besonders geeigneter, sobald die Unglattheit der Regressionsfunktion zunimmt (Abbildungen 8.13 und 8.14).

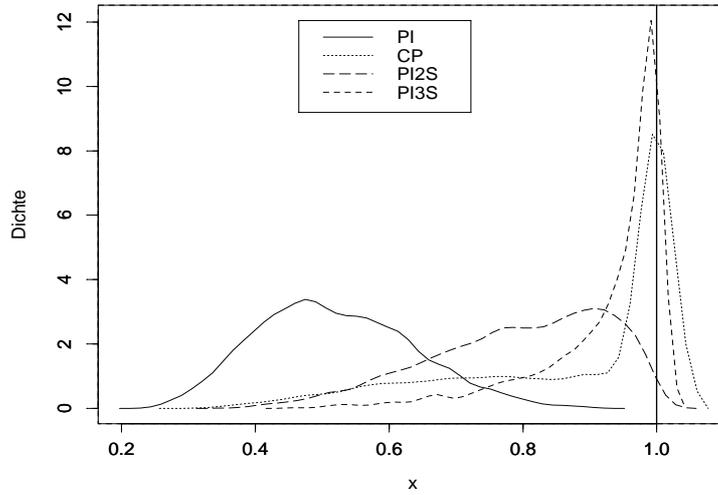


Abbildung 8.1: Kern-dichteschätzungen auf Grundlage von 1500 simulierten Bandbreiten \hat{h}_{p-i} , \hat{h}_{cp} , $\hat{h}_{p-i}^{(2)}$ und $\hat{h}_{p-i}^{(3)}$ bei $a = 0$ und $n = 30$. Die Bandbreiten dieser Dichteschätzungen mit Epanechnikovkern wurden mit der rule-of-thumb Methode mit Normalkern als Referenzdichte gewählt. Die vertikale Linie zeigt die Position der optimalen Bandbreite h_{opt} an.

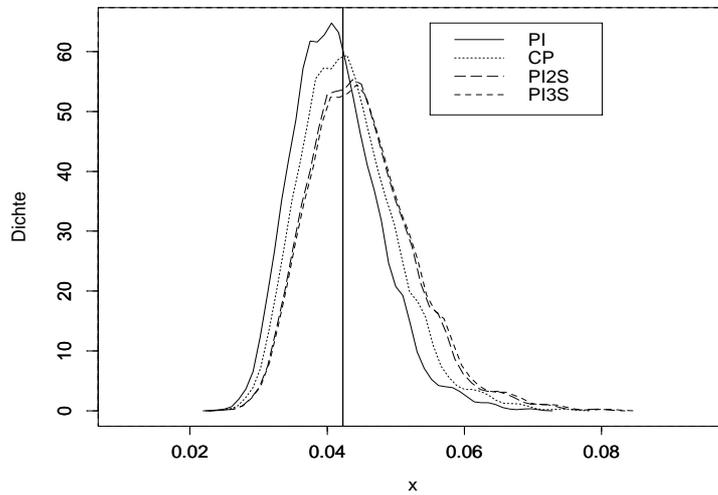


Abbildung 8.2: Kern-dichteschätzungen auf Grundlage von 1500 simulierten Bandbreiten \hat{h}_{p-i} , \hat{h}_{cp} , $\hat{h}_{p-i}^{(2)}$, und $\hat{h}_{p-i}^{(3)}$ bei $a = 2$ und $n = 30$. Die Bandbreiten dieser Dichteschätzungen mit Epanechnikovkern wurden mit der rule-of-thumb Methode mit Normalkern als Referenzdichte gewählt. Die vertikale Linie zeigt die Position der optimalen Bandbreite h_{opt} an.

Krit.	n=20	n=30	n=40	n=50	n=60	n=70	n=80	n= 100	n=120	n=130
lse	9.930	9.943	9.916	9.998	9.935	10.023	9.994	9.993	10.059	10.007
proj	1.001	0.993	0.992	0.999	1.003	1.025	0.993	0.998	0.997	0.999
pi	3.567	3.568	3.543	3.596	3.553	3.623	3.501	3.574	3.634	3.585
cp	1.659	1.655	1.619	1.656	1.618	1.679	1.644	1.617	1.689	1.641
pi2s	1.781	1.776	1.751	1.787	1.746	1.811	1.776	1.752	1.808	1.765
pi3s	1.242	1.231	1.214	1.241	1.195	1.261	1.229	1.202	1.242	1.207

Tabelle 8.1: Werte von D_n für die entsprechenden Schätzungen bei $a = 0$.

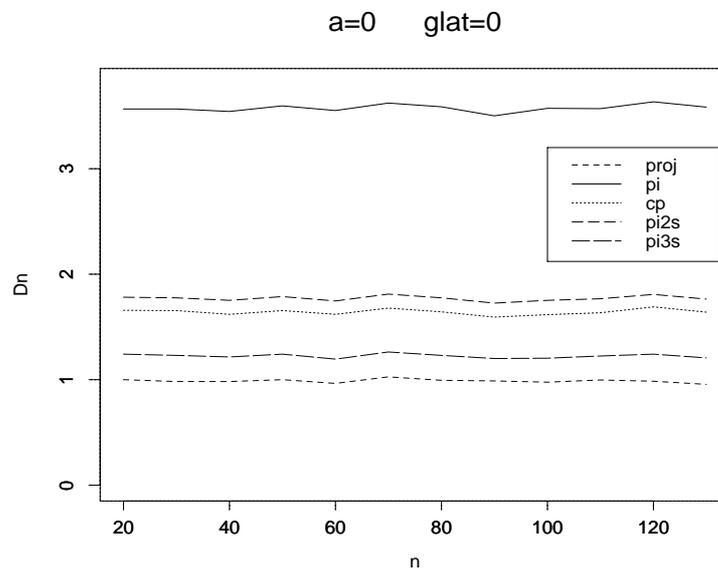


Abbildung 8.3: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

Krit.	n=20	n=30	n=40	n=50	n=60	n=70	n=80	n= 100	n=120	n=130
lse	7.665	6.922	6.257	5.781	5.396	4.980	4.718	4.226	3.893	3.724
proj	0.999	1.032	0.998	1.041	1.057	1.045	1.061	1.093	1.109	1.144
pi	2.879	2.661	2.423	2.299	2.190	2.031	1.961	1.812	1.726	1.669
cp	1.506	1.461	1.374	1.373	1.355	1.295	1.290	1.264	1.265	1.243
pi2s	1.558	1.495	1.388	1.374	1.347	1.274	1.260	1.219	1.209	1.186
pi3s	1.170	1.164	1.101	1.127	1.126	1.089	1.093	1.094	1.110	1.104

Tabelle 8.2: Werte von D_n für die entsprechenden Schätzungen bei $a = 1/8$.

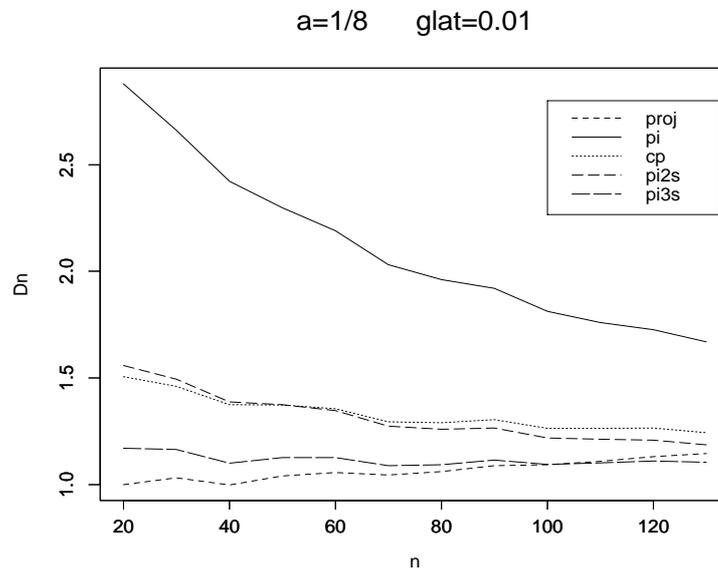


Abbildung 8.4: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

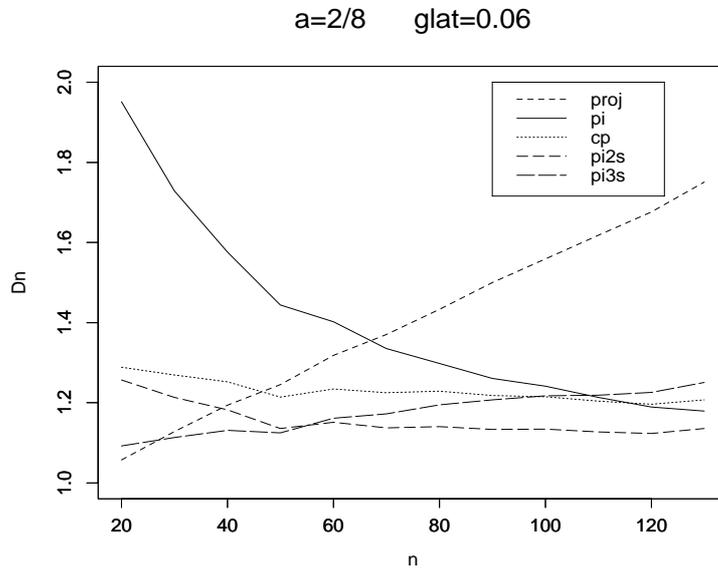


Abbildung 8.5: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

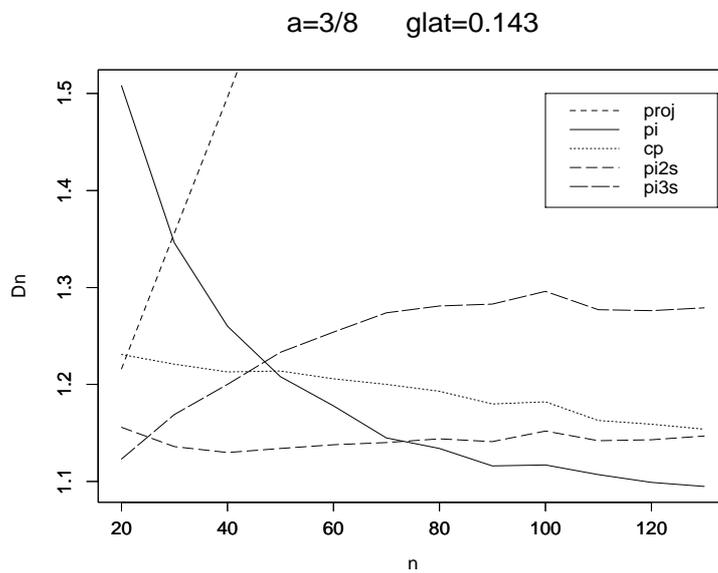


Abbildung 8.6: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

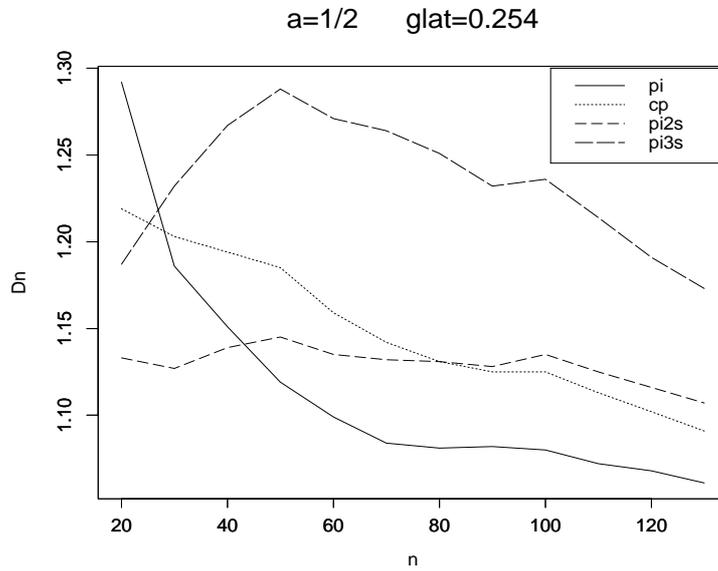


Abbildung 8.7: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

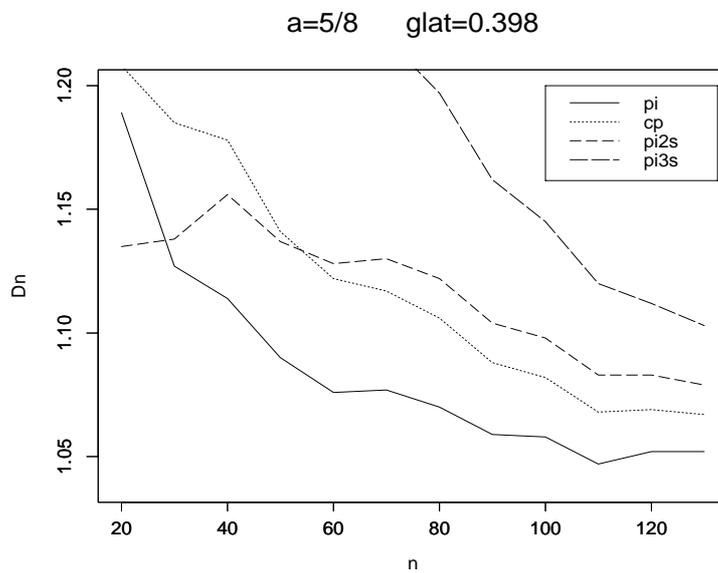


Abbildung 8.8: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

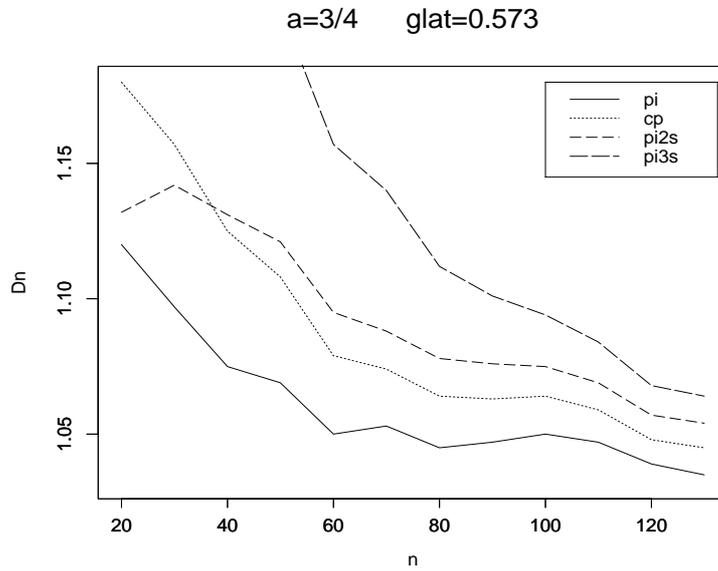


Abbildung 8.9: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

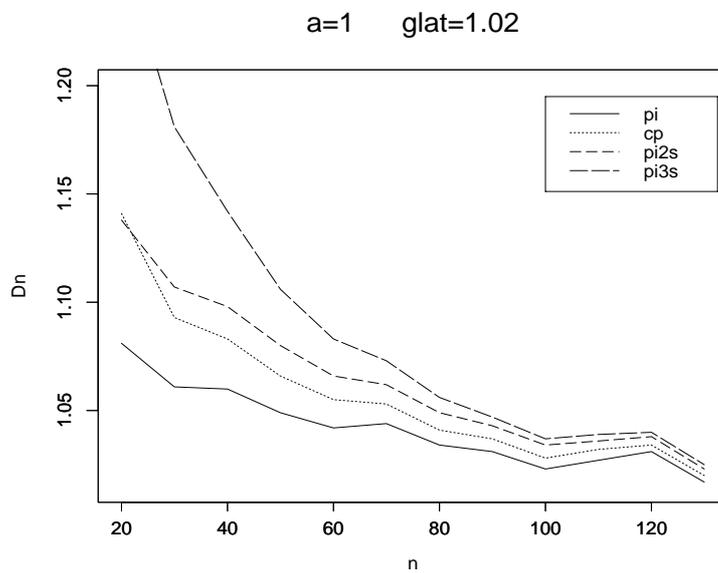


Abbildung 8.10: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

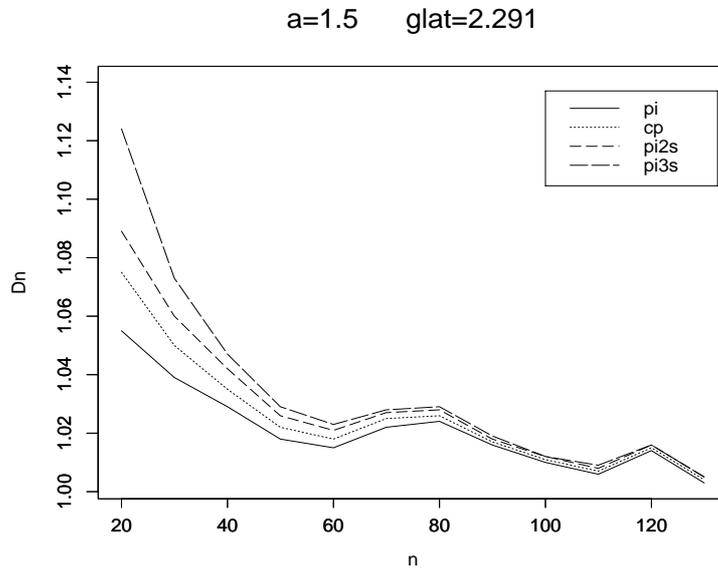


Abbildung 8.11: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

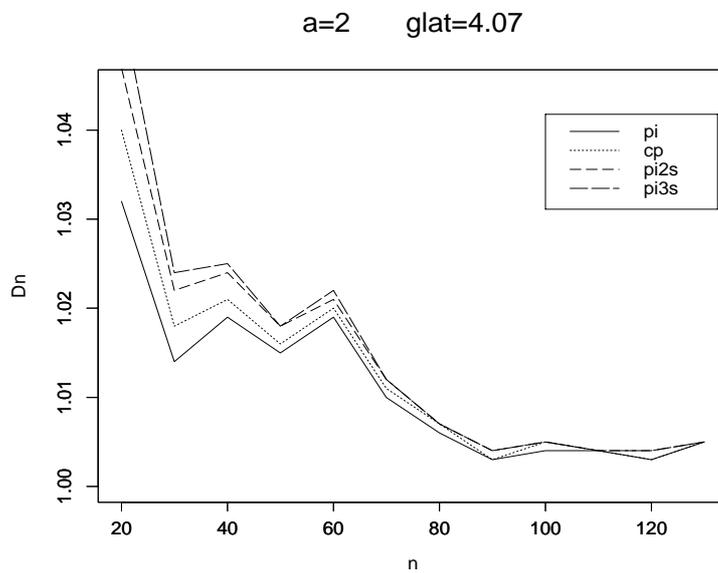


Abbildung 8.12: Quotient der Risiken der adaptiven Schätzungen und der optimalen Schätzung in Abhängigkeit von n .

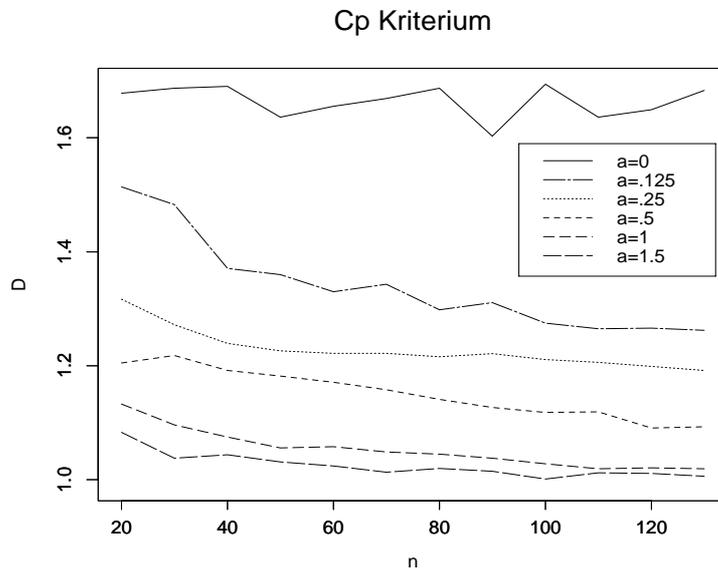


Abbildung 8.13: Verhalten des C_p Kriteriums bei wachsendem a in Abhängigkeit von n .

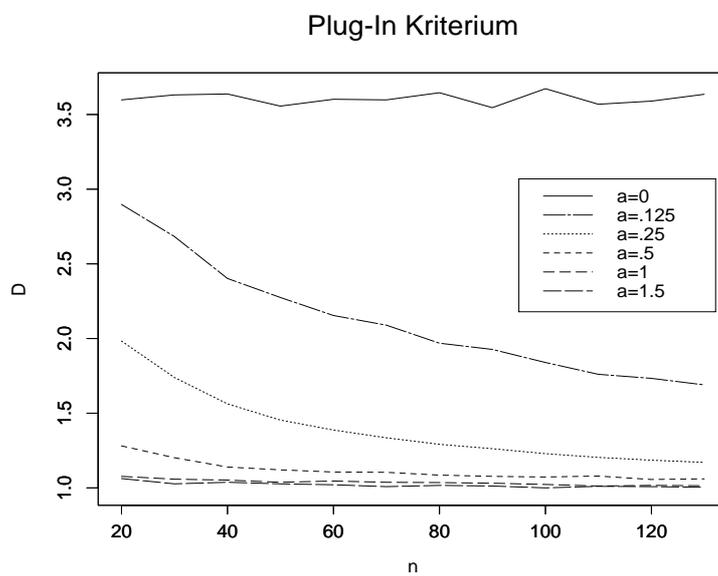


Abbildung 8.14: Verhalten des Plug-In Kriteriums bei wachsendem a in Abhängigkeit von n .

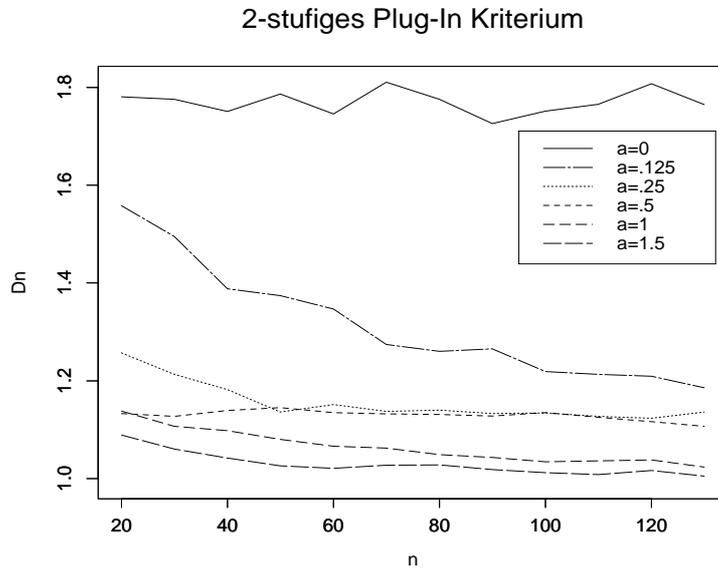


Abbildung 8.15: Verhalten des 2-stufigen Plug-In Kriteriums bei wachsendem a in Abhängigkeit von n .

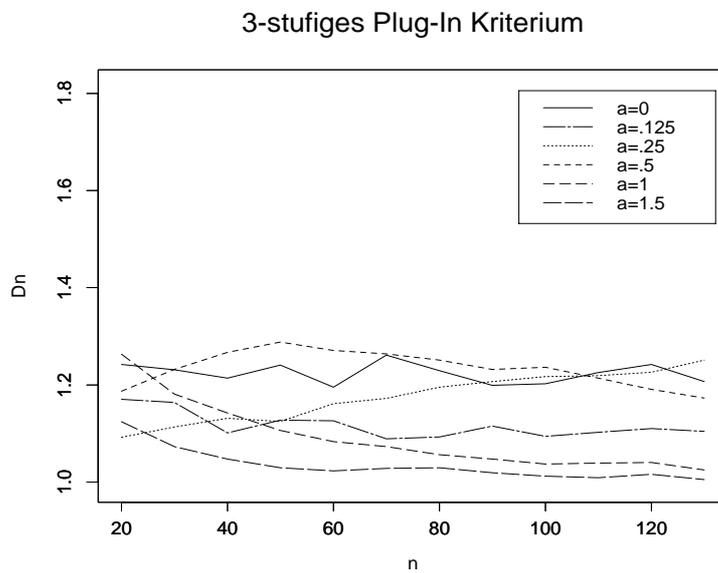


Abbildung 8.16: Verhalten des 3-stufigen Plug-In Kriteriums bei wachsendem a in Abhängigkeit von n .

8.2 Der Fall unglatter Funktionen

In diesem Abschnitt untersuchen wir, wie sich die Risiken der verschiedenen adaptiven Schätzungen von den Risiken der optimalen Schätzung für kleine bis mittlere Stichprobenumfänge unterscheiden. Wir betrachten dabei 2 verschiedene Regressionsfunktionen f im Modell:

$$y_{ij} = f(x_i) + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, 1), \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

$n_1 = \dots = n_m = n$; $x_i = \frac{i-1}{m-1}$. Der Vektor $f := (f_1, \dots, f_m)^T$ ist durch relativ unglatte nichtlineare Regressionsfunktionen gegeben, die in *Hurvich et al. (1998)* für Simulationsvergleiche nichtparametrischer Schätzungen betrachtet wurden.

Beispiel 1:

$$f(x) = (1 - 48x + 218x^2 - 315x^3 + 145x^4)/2$$

Beispiel 2:

$$f(x) = e^{x-1/3} \mathbb{1}_{[0,1/3)}(x) + e^{-2(x-1/3)} \mathbb{1}_{[1/3,1]}(x).$$

Es wurden $m = 5$, $m = 10$ und $m = 15$ Versuchspunkte festgelegt. Die Stichprobenumfänge sind $n = 15, 20, 25, 30, 40, 50$. Die Glättungsmatrix ist $K(n, h) = I_m - hI_m$ (die Schätzungen entsprechen dabei eine Stauchung der kleinsten Quadratschätzung). Da f und σ^2 gegeben sind, kann man die exakten Werte für h_{opt} , $M(h_{opt})$ und die Risiken 3. Ordnung der Schätzung mit der optimalen Bandbreite $M_3(h_{opt})$ berechnen (Gleichungen (6.5) und (6.6)). Die zwei Größen

$$R = \frac{|R_k - M(h_{opt})|}{M(h_{opt})} \quad \text{und} \quad \tilde{R} = \frac{|R_k - M_3(h_{opt})|}{M(h_{opt})}$$

wurden ebenfalls berechnet. Dabei steht R_k für folgende Risiken:

1. Risiko der kleinsten Quadratschätzung \bar{Y} (bezeichnet mit *lse* in den Tabellen)
2. Risiko 2. Ordnung der Schätzung von f mit der Plug-In Adaptation (Gleichung (4.6)) (bezeichnet mit *pi1*).
3. Risiko 3. Ordnung der Schätzung von f mit der Plug-In Adaptation (Satz 4.2.5) (bezeichnet mit *pi2*).
4. Risiko 2. Ordnung der Schätzung von f mit der „ α -Plug-In“ Adaptation (Gleichung (7.5)) (bezeichnet mit *alp*).

5. Risiko 3. Ordnung der Schätzung von f mit der C_p Adaptation (Satz 4.2.1) (bezeichnet mit cp).
6. Risiko 3. Ordnung der Schätzung von f mit der Full Cross-validation Adaptation (2. version) (Korollar 4.2.4) (bezeichnet mit $fcv2$).
7. Risiko 3. Ordnung der Schätzung von f mit der Cross-validation Adaptation (Korollar 4.2.1) (bezeichnet mit cv).

Es ist zu bemerken, daß in dieser Situation die C_p Adaptation und die Full Cross-validation Adaptation (1. version) dasselbe ergeben (denn $\gamma_1 = 0$). Die Werte in den Tabellen stehen für $100R$ bzw. $100\tilde{R}$.

Die Abbildungen 8.17 und 8.18 zeigen die Funktionen der Beispiele 1 und 2. Alle adaptiven Schätzungen sind erstmal besser als die kleinste Quadratschätzung. Im den Beispielen 1 und 2 kann die kleinste Quadratschätzung die feinen Strukturen der wahren Regressionsfunktion nicht widerspiegeln. Bei festem n bleibt sie fast genauso schlecht für $m = 5$, $m = 10$ und $m = 15$, während die adaptiven Schätzungen besser werden mit wachsendem m . Aus den Abbildungen 8.21 und 8.24 ist zu sehen, daß für $m = 15$ die Plug-In Adaptationen eindeutig besser als die anderen Adaptationen sind und dies besonders für kleinere n . Aus diesen Beispielen könnte man sagen: je mehr feine Strukturen die Funktion hat, desto schlechter bleibt die kleinste Quadratschätzung und desto geeigneter werden die Plug-In Adaptationen im Vergleich zu den automatischen Kriterien (C_p , Cross-validation und Full Cross-validation Kriterien) bei kleinen n . Aus den Werten der sechs Tabellen sieht man, daß der Mittelwertvektor aus dem Beispiel 1 etwas schwieriger zu schätzen ist als der Vektor aus dem 2. Beispiel. Das Gemeinsame an den zwei Beispielen ist:

1. Die Risiken 2. Ordnung der „ α -Plug-In“ Schätzung sind nicht nur kleiner als die Risiken 2. Ordnung der Plug-In Schätzung, sondern auch kleiner als die Risiken dritter Ordnung der anderen adaptiven Schätzungen.
2. Die kleinsten Risiken sind immer entweder die Risiken 3. Ordnung der Plug-In Schätzung oder die Risiken 2. Ordnung der „ α Plug-In“ Schätzung.
3. Von allen automatischen Kriterien liefert immer die Full Cross-validation Adaptation (2. Version) die kleinsten Risiken.

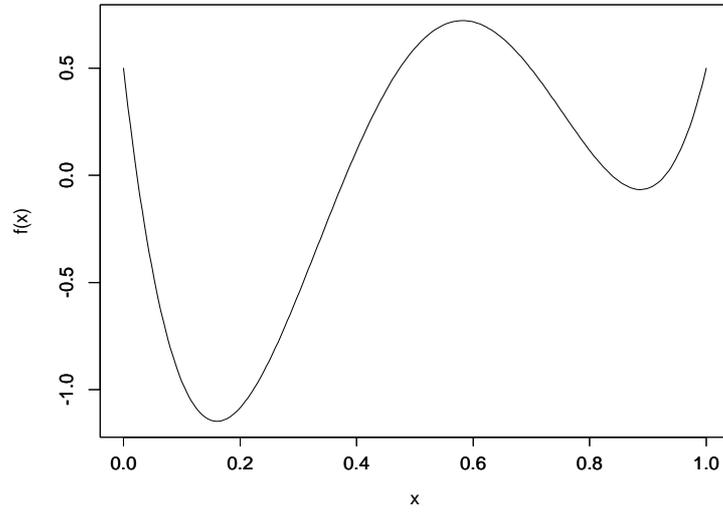


Abbildung 8.17: $f(x) = (1 - 48x + 218x^2 - 315x^3 + 145x^4)/2$

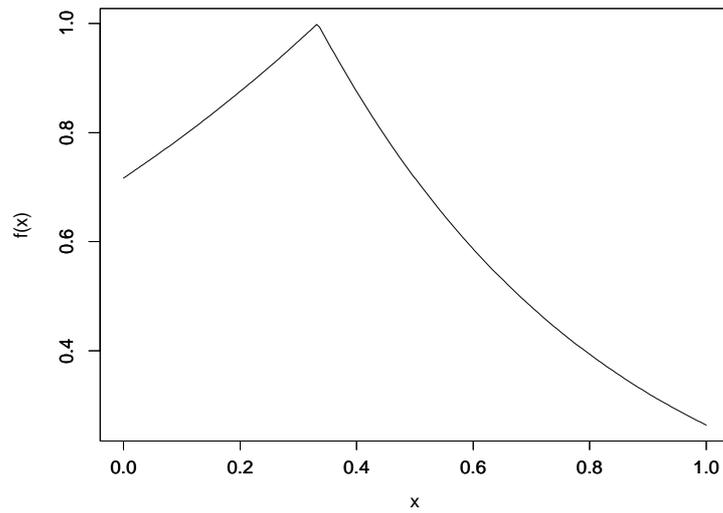


Abbildung 8.18: $f(x) = e^{(x-1/3)} \mathbb{1}_{[0,1/3)}(x) + e^{-2(x-1/3)} \mathbb{1}_{[1/3,1]}(x)$

Krit.	n= 15		n= 20		n= 25		n= 30		n= 50	
	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}
lse	19.663	18.903	14.748	14.427	11.798	11.634	9.832	9.737	5.899	5.878
pi1	14.957	14.197	11.363	11.042	9.160	8.996	7.672	7.577	4.650	4.629
alp	11.193	10.432	8.655	8.335	7.050	6.885	5.944	5.849	3.650	3.630
pi2	7.916	7.156	7.559	7.238	6.786	6.621	6.051	5.956	4.086	4.066
cp	15.319	14.559	11.552	11.231	9.275	9.111	7.750	7.655	4.676	4.655
fcv1	15.319	14.559	11.552	11.231	9.275	9.111	7.750	7.655	4.676	4.655
fcv2	12.809	12.049	10.198	9.877	8.431	8.267	7.174	7.079	4.476	4.455
cv	16.574	15.814	12.229	11.908	9.698	9.533	8.038	7.943	4.776	4.755

Tabelle 8.3: Beispiel 1 **m=5**

Krit.	n= 15		n= 20		n= 25		n= 30		n= 50	
	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}
lse	18.006	17.422	13.505	13.258	10.804	10.678	9.003	8.93	5.402	5.386
pi1	5.257	4.673	4.308	4.061	3.621	3.495	3.115	3.042	1.986	1.970
alp	4.407	3.823	3.694	3.448	3.142	3.016	2.722	2.649	1.758	1.742
pi2	3.901	3.318	3.571	3.325	3.160	3.034	2.799	2.726	1.875	1.860
cp	6.962	6.378	5.227	4.981	4.195	4.069	3.506	3.433	2.122	2.106
fcv1	6.962	6.378	5.227	4.981	4.195	4.069	3.506	3.433	2.122	2.106
fcv2	5.829	5.245	4.614	4.368	3.812	3.685	3.245	3.172	2.030	2.015
cv	7.529	6.945	5.534	5.287	4.386	4.260	3.637	3.564	2.167	2.151

Tabelle 8.4: Beispiel 1 **m=10**

Krit.	n= 15		n= 20		n= 25		n= 30		n= 50	
	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}
lse	18.448	17.821	13.836	13.571	11.069	10.933	9.224	9.146	5.535	5.518
pi1	2.424	1.796	2.286	2.021	2.053	1.918	1.836	1.757	1.251	1.234
alp	2.035	1.407	2.006	1.741	1.835	1.699	1.657	1.578	1.147	1.130
pi2	2.543	1.915	2.348	2.083	2.091	1.956	1.862	1.783	1.260	1.243
cp	4.693	4.065	3.510	3.246	2.817	2.682	2.357	2.279	1.432	1.415
fcv1	4.693	4.065	3.510	3.246	2.817	2.682	2.357	2.279	1.432	1.415
fcv2	3.916	3.288	3.090	2.826	2.555	2.419	2.178	2.100	1.370	1.353
cv	5.081	4.453	3.720	3.456	2.948	2.813	2.447	2.368	1.464	1.447

Tabelle 8.5: Beispiel 1 $m=15$

Krit.	n= 15		n= 20		n= 25		n= 30		n= 50	
	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}
lse	15.637	15.255	11.728	11.567	9.382	9.300	7.819	7.771	4.691	4.681
pi1	12.021	11.638	9.107	8.946	7.330	7.247	6.133	6.085	3.709	3.699
alp	9.128	8.745	7.011	6.849	5.688	5.605	4.784	4.736	2.923	2.913
pi2	8.311	7.928	7.086	6.925	6.062	5.979	5.264	5.216	3.405	3.394
cp	12.835	12.452	9.545	9.383	7.602	7.520	6.318	6.270	3.773	3.763
fcv1	12.835	12.452	9.545	9.383	7.602	7.520	6.318	6.270	3.773	3.763
fcv2	10.906	10.524	8.496	8.335	6.945	6.863	5.869	5.821	3.616	3.606
cv	13.799	13.417	10.069	9.908	7.931	7.848	6.543	6.495	3.852	3.841

Tabelle 8.6: Beispiel 2 $m=5$

Krit.	n= 15		n= 20		n= 25		n= 30		n= 50	
	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}
lse	14.436	14.135	10.827	10.70	8.662	8.597	7.218	7.180	4.331	4.323
pi1	4.524	4.223	3.627	3.501	3.014	2.950	2.575	2.537	1.620	1.612
alp	3.863	3.562	3.147	3.021	2.638	2.573	2.265	2.227	1.439	1.431
pi2	3.773	3.473	3.216	3.089	2.756	2.691	2.397	2.359	1.557	1.549
cp	5.681	5.380	4.256	4.129	3.408	3.343	2.844	2.806	1.714	1.706
fcv1	5.681	5.380	4.256	4.129	3.408	3.343	2.844	2.806	1.714	1.706
fcv2	4.800	4.499	3.776	3.649	3.107	3.042	2.637	2.600	1.642	1.633
cv	6.122	5.821	4.496	4.369	3.558	3.493	2.947	2.909	1.750	1.742

Tabelle 8.7: Beispiel 2 **m=10**

Krit.	n= 15		n= 20		n= 25		n= 30		n= 50	
	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}	R	\tilde{R}
lse	14.421	14.121	10.816	10.689	8.653	8.588	7.211	7.173	4.326	4.318
pi1	2.321	2.021	2.026	1.900	1.758	1.694	1.542	1.504	1.016	1.008
alp	2.027	1.727	1.813	1.687	1.591	1.526	1.404	1.367	0.936	0.928
pi2	2.543	2.244	2.146	2.020	1.833	1.768	1.592	1.555	1.034	1.026
cp	3.813	3.513	2.838	2.711	2.267	2.202	1.890	1.852	1.138	1.130
fcv1	3.813	3.513	2.838	2.711	2.267	2.202	1.890	1.852	1.138	1.130
fcv2	3.226	2.926	2.518	2.392	2.066	2.002	1.752	1.715	1.090	1.082
cv	4.106	3.806	2.997	2.871	2.367	2.302	1.958	1.921	1.162	1.154

Tabelle 8.8: Beispiel 2 **m=15**

Beispiel 1 $m=5$

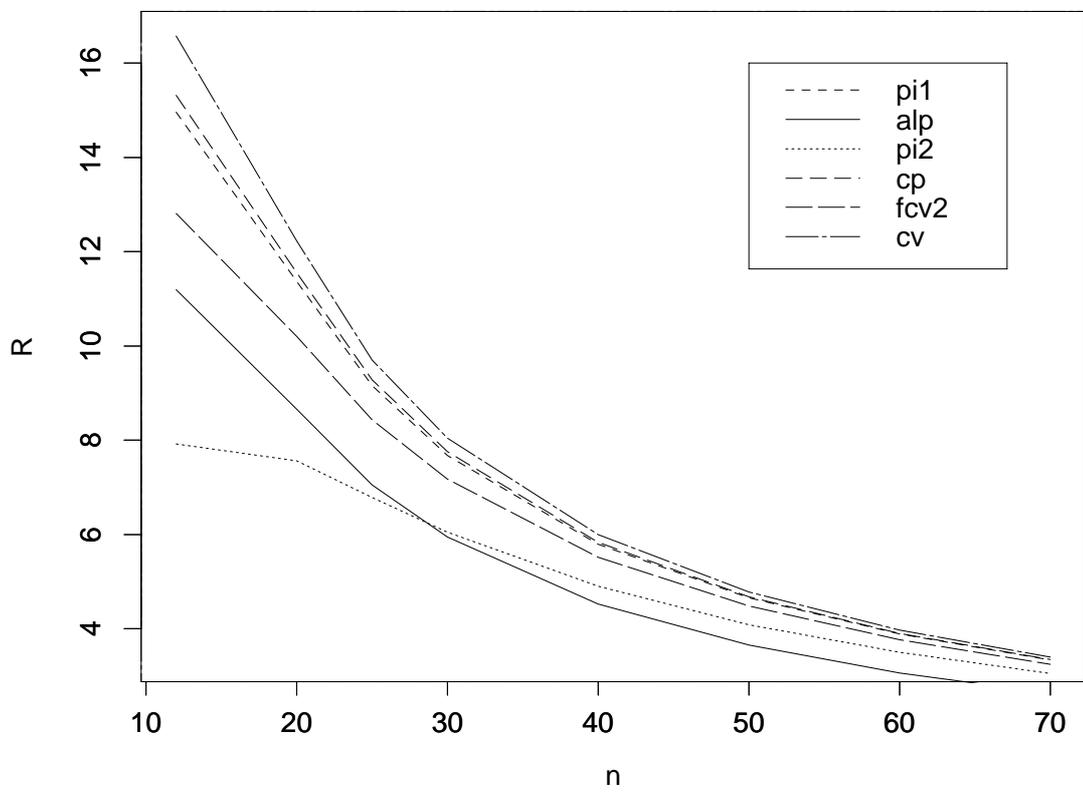


Abbildung 8.19: Die normierten Risiken der adaptiven Schätzungen in Abhängigkeit von n

Beispiel 1 $m=10$

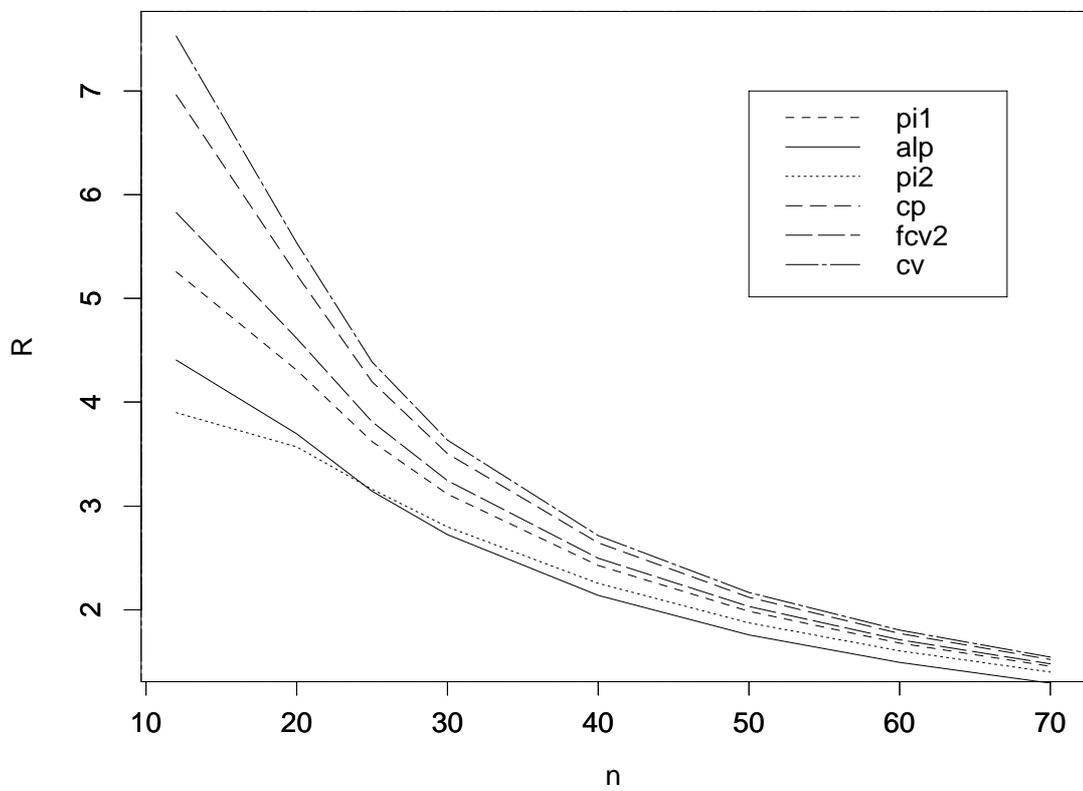


Abbildung 8.20: Die normierten Risiken der adaptiven Schätzungen in Abhängigkeit von n

Beispiel 1 $m=15$

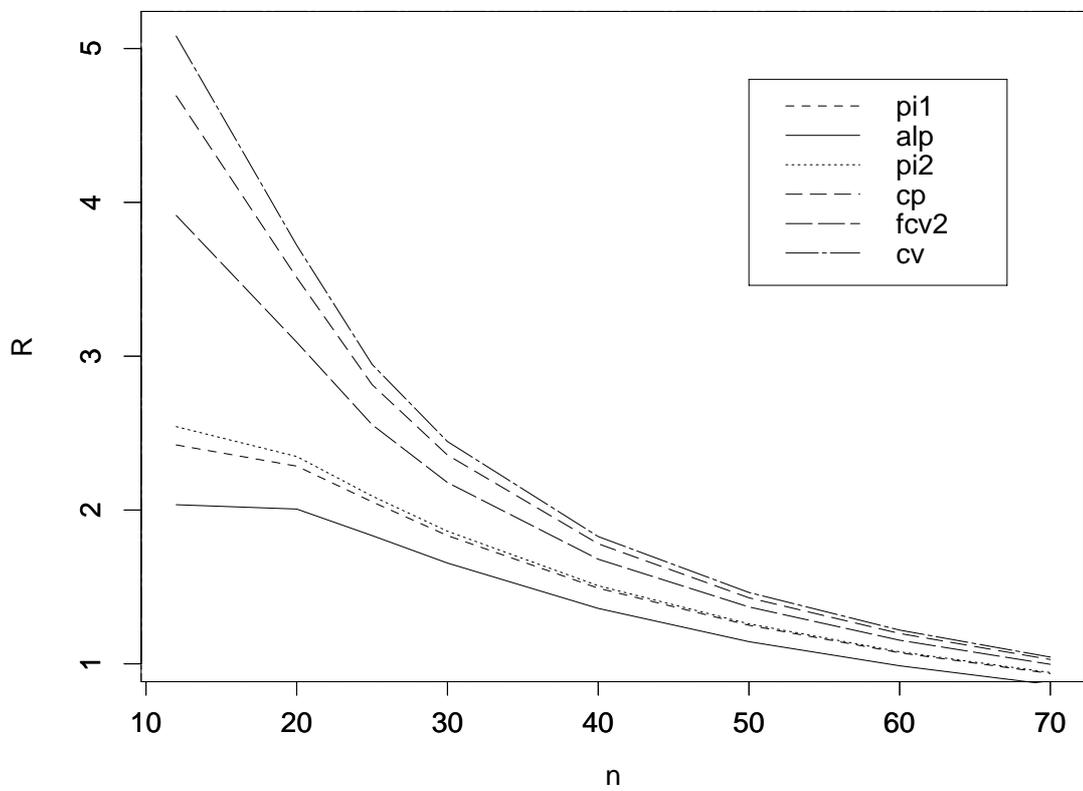


Abbildung 8.21: Die normierten Risiken der adaptiven Schätzungen in Abhängigkeit von n

Beispiel 2 $m=5$

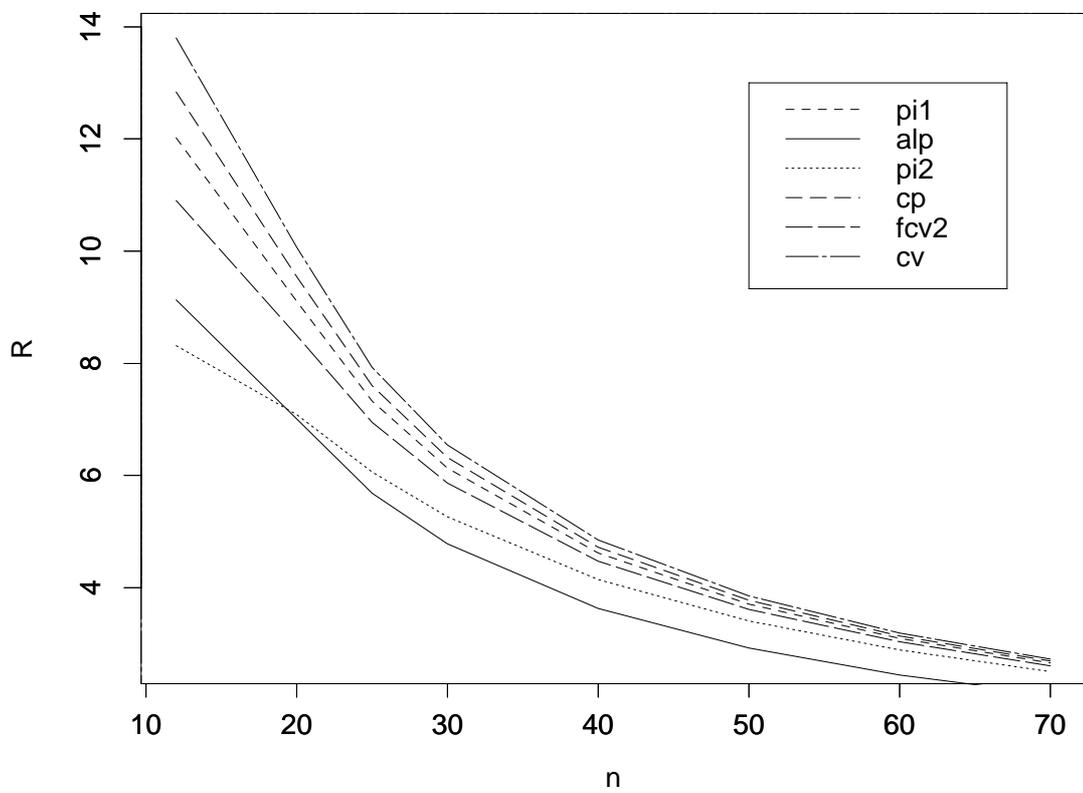


Abbildung 8.22: Die normierten Risiken der adaptiven Schätzungen in Abhängigkeit von n

Beispiel 2 $m=10$

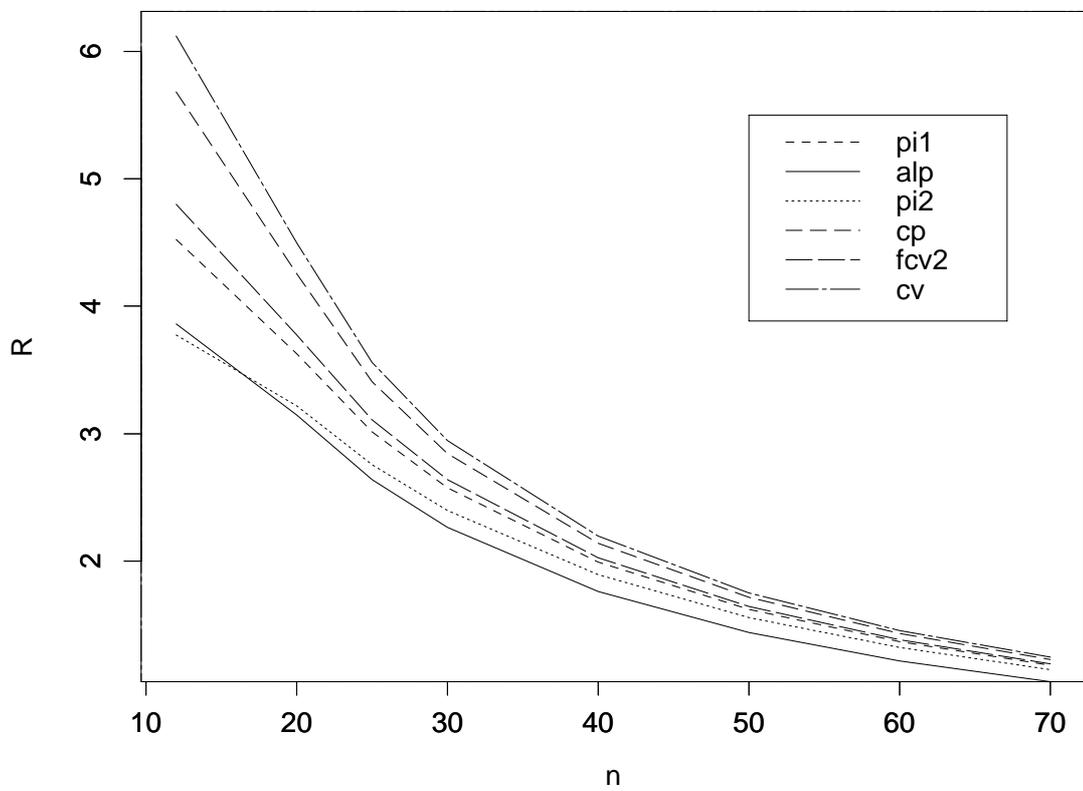


Abbildung 8.23: Die normierten Risiken der adaptiven Schätzungen in Abhängigkeit von n

Beispiel 2 $m=15$

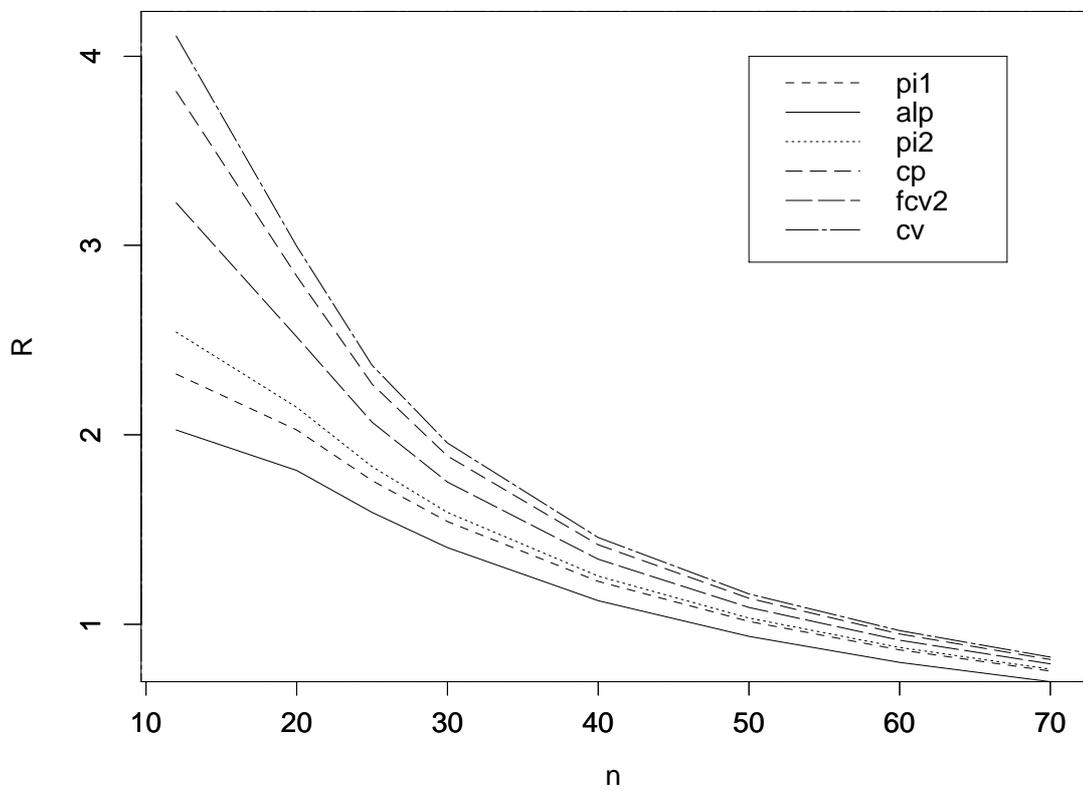


Abbildung 8.24: Die normierten Risiken der adaptiven Schätzungen in Abhängigkeit von n

Kapitel 9

Anwendung auf die Regression mit wachsender Anzahl von Versuchspunkten

Bisher haben wir feste Versuchspunkte x_i betrachtet. In diesem Kapitel sind die Versuchspunkte zufällig und ihre Verteilung ist durch eine Dichte beschrieben. Wir untersuchen jetzt eine Asymptotik, bei der die Anzahl n der Versuchspunkte gegen unendlich geht und es keine Wiederholungen in den Versuchspunkten gibt. Es wird das Analogon zum Satz 3.1.1 angegeben, sowie die Konvergenzgeschwindigkeit der Approximation zweiter Ordnung der optimalen Bandbreite. Die ersten Resultate sind in dieser Richtung von *Fan und Huang (1997)* erzielt worden, wobei sie keine explizite Form wie in der Gleichung (3.8) angegeben haben. Dadurch erhielten sie (statt eine etwas schnellere $O(n^{-4/5})$, wie in dieser Dissertation) nur die $O(n^{-2/5})$ Konvergenzrate der Approximation erster Ordnung der optimalen Bandbreite.

Um eine Verwechslung mit der Bezeichnung der Dichte zu vermeiden, werden wir in diesem Kapitel die Regressionsfunktion mit m und die Dichte der Zufallsgröße X mit f bezeichnen.

Wir betrachten unabhängig, identisch verteilte Zufallsgrößen $(X_1, Y_1), \dots, (X_n, Y_n)$. Die X_i sind die erklärenden Variablen mit einer positiven Dichte f und nehmen Werte in einem abgeschlossenen und endlichen Intervall \mathcal{X} an. Die Variablen Y_i sind eindimensional und sind die abhängigen Variablen.

Das Ziel ist die Schätzung der bedingten Erwartung $m(x) = E(Y|X = x)$ für jeden Punkt $x \in \mathcal{X}$. Wir können dieses Problem mit Hilfe des Modells

$$Y_i = m(X_i) + \sigma^2(X_i)^{1/2} \varepsilon_i, \quad i = 1, \dots, n \quad (9.1)$$

ausdrücken, wobei $\sigma^2(x) = \text{Var}(Y|X = x)$ die bedingte Varianz von Y unter $X = x$ ist. Die Zufallsgrößen $X_1, \dots, X_n, \varepsilon_1, \dots, \varepsilon_n$ sind dabei unabhängig,

wobei X_i die Dichte f hat und die Fehler ε_i identisch verteilt sind mit

$$E(\varepsilon_1) = 0 \quad \text{Var}(\varepsilon_1) = 1 .$$

Die Funktionen

$$m : \mathcal{X} \longrightarrow \mathbb{R} \quad \text{und} \quad \sigma^2 : \mathcal{X} \longrightarrow \mathbb{R}$$

sind die unbekanntes Regressions- bzw. Varianzfunktion.

Sei $K(\cdot)$ eine Kernfunktion 2. Ordnung d. h.

$$\int xK(x)dx = 0 \quad \text{und} \quad \int x^2K(x)dx \neq 0 .$$

Die Schätzung von $m(x)$ ist hier die Nadaraya-Watson Kernschätzung mit der Bandbreite h , die durch

$$\hat{m}_h^{NW}(x) = \frac{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)Y_i}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)} \quad (9.2)$$

definiert ist. Ein Untersuchungsziel ist asymptotische Entwicklungen der Formen (3.8) und (3.9) anzugeben.

Für die Beurteilung der Güte einer Schätzung \hat{m}_h^{NW} werden wir das „aktuelle Risiko“ also den bedingten mittleren integrierten quadratischen Fehler $M(h)$ verwenden, in Anlehnung an *Fan und Huang (1997)*.

$$\begin{aligned} M(h) &:= E \left(\int [\hat{m}_h^{NW}(x) - m(x)]^2 f(x) dx | X_1, \dots, X_n \right) \quad (9.3) \\ &= \int (E[\hat{m}_h^{NW}(x) | X_1, \dots, X_n] - m(x))^2 f(x) dx \\ &\quad + \int E \left\{ (\hat{m}_h^{NW}(x) - E[\hat{m}_h^{NW}(x) | X_1, \dots, X_n])^2 \middle| X_1, \dots, X_n \right\} f(x) dx \\ &= IBias^2(\hat{m}_h^{NW}(x) | X_1, \dots, X_n) \\ &\quad + IVar(\hat{m}_h^{NW}(x) | X_1, \dots, X_n) , \end{aligned}$$

wobei $IBias^2(\hat{m}_h^{NW}(x) | X_1, \dots, X_n)$ und $IVar(\hat{m}_h^{NW}(x) | X_1, \dots, X_n)$ für das Integral des Quadrates des bedingten Bias und das Integral der bedingten Varianz der Schätzung (9.2) im Punkt x stehen. Die „aktuelle“ optimale Bandbreite ist durch $h_{opt} = \arg \min_h M(h)$. Damit man eine Entwicklung höherer

Ordnung von h_{opt} herleiten kann, ist es wichtig, $Bias(\hat{m}_h^{NW}(x)|X_1, \dots, X_n)$ und $Var(\hat{m}_h^{NW}(x)|X_1, \dots, X_n)$ in der Form

$$Bias(\hat{m}_h^{NW}(x)|X_1, \dots, X_n) \approx h^2 b_1(x) - h^4 b_2(x) \quad (9.4)$$

und

$$\begin{aligned} Var(\hat{m}_h^{NW}(x)|X_1, \dots, X_n) &\approx n^{-1} h^{-1} v_1(x) + n^{-1} v_0(x) \\ &- n^{-1} h v_2(x) \end{aligned} \quad (9.5)$$

sowie

$$\begin{aligned} M(h) &\approx h^4 B_1 - h^6 B_2 + n^{-1} h^{-1} V_1 + n^{-1} V_0 - n^{-1} h V_2 \\ &= h^4 \int b_1(x)^2 f(x) dx - h^6 \int b_1(x) b_2(x) f(x) dx \\ &\quad + n^{-1} h^{-1} \int v_1(x) f(x) dx + n^{-1} \int v_0(x) f(x) dx \\ &\quad - n^{-1} h \int v_2(x) f(x) dx \end{aligned} \quad (9.6)$$

darzustellen. Man braucht deshalb einige Voraussetzungen über die Regressionsfunktion, die Dichte f , die Varianzfunktion und die Kernfunktion.

- v1** Die Regressionsfunktion m sei viermal stetig differenzierbar auf \mathcal{X} .
- v2** Die Dichte f sei dreimal stetig differenzierbar und nehme nur positive Werte auf \mathcal{X} an.
- v3** Die Varianzfunktion σ^2 sei zweimal stetig differenzierbar auf \mathcal{X} .
- v4** K sei ein Kern 2. Ordnung und eine Wahrscheinlichkeitsdichte, die symmetrisch um 0 ist.
- v5** $h = h(n)$ und für $n \rightarrow \infty$ konvergieren $h \rightarrow 0$, $nh \rightarrow \infty$, $nh^7 \rightarrow 0$.

Es seien

$$A_1 := \frac{V_1}{4B_1} \quad A_2 := \frac{V_2}{4B_1} \quad A_3 := \frac{6B_2}{4B_1} . \quad (9.7)$$

Satz 9.0.1

Unter den Voraussetzungen **v1** bis **v5** gilt:

$$1) \quad h_{opt} = (A_1 n^{-1})^{1/5} + \left(\frac{A_2}{5A_1} + \frac{A_3}{5} \right) (A_1 n^{-1})^{3/5} + O(n^{-1}) \quad (9.8)$$

$$2) \quad \frac{h_{opt} - h_0}{h_{opt}} = O(n^{-4/5}) \quad (9.9)$$

wobei

$$h_0 := (A_1 n^{-1})^{1/5} + \left(\frac{A_2}{5A_1} + \frac{A_3}{5} \right) (A_1 n^{-1})^{3/5}. \quad (9.10)$$

Im Satz 2 von *Fan und Huang (1997)* wird die Rate $O(n^{-2/5})$ in der Gleichung (9.9) erreicht. Der Grund ist, daß sie statt die Entwicklung 2. Ordnung (9.8) die Entwicklungen 1. Ordnung $h_{opt} = (A_1 n^{-1})^{1/5} + O_p(n^{-3/5})$ sowie $h_0 := (A_1 n^{-1})^{1/5}$ verwendet haben.

Wir werden jetzt für die Nadaraya-Watson Kernschätzung die Entwicklungen für den Bias und die Varianz in einem Punkt x angeben, mit denen man die Approximationen (9.8) und (9.9) herleitet. Für den lokal linearen Schätzer sind diese Ausdrücke in *Fan und Huang (1997)* angegeben. Wir bezeichnen mit:

$$\begin{aligned} \mu_i &:= \int z^i K(z) dz & \nu_i &:= \int z^i K^2(z) dz \quad i = 0, 1, 2, \dots \text{ und} \\ g(x) &:= f(x)\sigma^2(x). \end{aligned}$$

Satz 9.0.2

Unter den Voraussetzungen **v1** bis **v5** gilt:

$$\begin{aligned} 1) \quad & \text{Bias}(\hat{m}_h^{NW}(x)|X_1, \dots, X_n) = h^2 \mu_2 \left(\frac{f'(x)m'(x)}{f(x)} + \frac{1}{2} m''(x) \right) \\ & - h^4 \frac{\mu_2^2}{2} \frac{f''(x)}{f(x)^2} \left(f'(x)m'(x) + \frac{1}{2} m''(x)f(x) \right) \\ & + h^4 \mu_4 \left(\frac{m'(x)f'''(x)}{3!f(x)} + \frac{m''(x)f''(x)}{4} \right) + o(h^4) + O_p(n^{-1/2}h^{1/2}) \\ 2) \quad & \text{Var}(\hat{m}_h^{NW}(x)|X_1, \dots, X_n) = (nh)^{-1} \frac{\nu_0 \sigma^2(x)}{f(x)} + n^{-1} \frac{\nu_1 g'(x)}{f(x)^2} \\ & + n^{-1} h \left(\frac{\nu_2 g''(x)}{2} - \mu_2 \nu_0 f''(x) \sigma^2(x) \right) f(x)^{-2} + o(n^{-1}h) + O_p((nh)^{-3/2}) \end{aligned}$$

Mit Hilfe dieses Satzes erhält man die Größen A_1 , A_2 und A_3 , die im Satz 9.0.1 vorkommen. Aus (9.6) sind nämlich:

$$B_1 = \mu_2^2 \int \left(\frac{f'(x)m'(x)}{f(x)} + \frac{1}{2}m''(x) \right)^2 f(x) dx$$

$$\begin{aligned} B_2 &= \mu_2^3 \int \left(\frac{f'(x)m'(x)}{f(x)} + \frac{1}{2}m''(x) \right)^2 f''(x) dx \\ &- 2\mu_2\mu_4 \int \left(\frac{f'(x)m'(x)}{f(x)} + \frac{1}{2}m''(x) \right) \times \\ &\quad \times \left(\frac{m'(x)f'''(x)}{3!} + \frac{m''(x)f''(x)f(x)}{4} \right) dx \end{aligned}$$

$$V_1 = \int \nu_0 \sigma^2(x) dx$$

$$V_2 = - \int \left(\frac{\nu_2 g''(x)}{2} - \mu_2 \nu_0 f''(x) \sigma^2(x) \right) f(x)^{-1} dx$$

Diese Konstanten A_1 , A_2 und A_3 lauten:

$$A_1 = \frac{\int \nu_0 \sigma^2(x) dx}{4\mu_2^2 \int [f'(x)m'(x)/f(x) + m''(x)/2]^2 f(x) dx} \quad (9.11)$$

$$A_2 = \frac{\int [\mu_2 \nu_0 f''(x) \sigma^2(x) - g''(x) \nu_2 / 2] f(x)^{-1} dx}{4\mu_2^2 \int [f'(x)m'(x)/f(x) + m''(x)/2]^2 f(x) dx} \quad (9.12)$$

und

$$\begin{aligned} A_3 &= \frac{3\mu_2 \int [f'(x)m'(x)/f(x) + m''(x)/2]^2 f''(x) dx}{2 \int [f'(x)m'(x)/f(x) + m''(x)/2]^2 f(x) dx} \\ &- \frac{3\mu_4 \int [f'(x)m'(x)/f(x) + m''(x)/2] [f'''(x)m'(x)/3!] dx}{\mu_2 \int [f'(x)m'(x)/f(x) + m''(x)/2]^2 f(x) dx} \\ &- \frac{3\mu_4 \int [f'(x)m'(x)/f(x) + m''(x)/2] [m''(x)f''(x)f(x)/4] dx}{\mu_2 \int [f'(x)m'(x)/f(x) + m''(x)/2]^2 f(x) dx} \end{aligned} \quad (9.13)$$

Beispiel

Die Varianzfunktion sei konstant, d. h. $\sigma^2(x) = \sigma^2$ und die Versuchspunkte seien gleichverteilt auf $[0, 1]$. Die Ausdrücke für A_1 , A_2 und A_3 vereinfachen sich und sind:

$$A_1 = \frac{\nu_0 \sigma^2}{\mu_2^2 \int_0^1 [m''(x)]^2 dx}, \quad A_2 = 0 \quad \text{und} \quad A_3 = 0.$$

Das heißt, bei gleichverteilten Versuchspunkten und konstanter Varianz bringt für mittlere bis große n die Entwicklung 2. Ordnung, also das Glied der Ordnung $O(n^{-3/5})$ in der asymptotischen Entwicklung (9.8) keinerlei Verbesserungen gegenüber der Entwicklung 1. Ordnung.

Für den lokal linearen Schätzer (vgl. *Wand and Jones (1995)*) erhält man (siehe *Fan und Huang (1997)*):

$$A_1 = \frac{\nu_0 \sigma^2}{\mu_2^2 \int_0^1 [m''(x)]^2 dx}$$
$$A_2 = 0 \quad \text{und} \quad A_3 = -\frac{\mu_4 \int_0^1 m''(x) m^{(4)}(x)}{4\mu_2 \int_0^1 [m''(x)]^2 dx}$$

und die optimale Bandbreite ist

$$h_{opt} = \left(\frac{\nu_0 \sigma^2}{n \mu_2^2 \int_0^1 [m''(x)]^2 dx} \right)^{1/5} - \left(\frac{\mu_4 \int_0^1 m''(x) m^{(4)}(x)}{20\mu_2 \int_0^1 [m''(x)]^2 dx} \right) \times$$
$$\times \left(\frac{\nu_0 \sigma^2}{n \mu_2^2 \int_0^1 [m''(x)]^2 dx} \right)^{3/5} + O(n^{-1}).$$

Bemerkung 9.0.1

Die von Ruppert, Sheather und Wand (1995), Seite 1266 bemerkte Tatsache, wonach die Hinzunahme des Gliedes der Ordnung $n^{-3/5}$ bei der Konstruktion einer Plug-In Schätzung der Dichte (Hall et al. (1991)) in einer Simulationsstudie keine Verbesserung gebracht hat, wird auch für die Schätzung der Regressionsfunktion in dieser letzten Formel ersichtlich. Um nämlich die schnellere Konvergenzrate $n^{-1/2}$ bei der Plug-In Schätzung von (9.10) zu erzielen, muß das zweite Glied der Ordnung $n^{-3/5}$ gut geschätzt werden. Das bedeutet, die Größe $\int_0^1 [m''(x)]^2 dx$ muß mit der Rate $n^{-1/2}$ geschätzt werden und die Größe $\int_0^1 m''(x)m^{(4)}(x)dx$ darf mit der schlechteren Rate $n^{-1/10}$ geschätzt werden. Andernfalls ist die Konvergenzrate der Summe der Plug-In Schätzungen der ersten zwei Glieder von h_{opt} dieselbe wie die von der Plug-In Schätzung des $O(n^{-1/5})$ -Gliedes von h_{opt} .

Zusammenfassung:

Mit asymptotischen Entwicklungen höherer Ordnung haben wir verschiedene adaptive Schätzungen der Regressionsfunktion in einem nichtparametrischen Modell mit Wiederholungen an den Versuchspunkten untereinander verglichen. Da die Risiken zweiter Ordnung aller adaptiven Schätzungen gleich sind, ist es notwendig gewesen, ihre Risiken dritter Ordnung zu berechnen um die Adaptationsverfahren der Güte nach anzuordnen. Die expliziten Ausdrücke für die Differenz der Risiken dritter Ordnung von den Risiken zweiter Ordnung der adaptiven Schätzungen konnten wir berechnen. Dadurch sahen wir, daß für die Nadaraya-Watson Kernschätzung mit Normalkern die Adaptation mit Hilfe des Cross-validation Kriteriums am schlechtesten war. Für Schätzungen von f der Form $c\bar{Y}$ ist die beste Adaptation entweder die Plug-In oder die Full Cross-validation (2. Version) (Bemerkung 6.2.1). Außerdem haben wir einen Schätzer der Regressionsfunktion durch Unterschätzen der Varianz konstruiert, dessen Risiken zweiter Ordnung kleiner sind als die Risiken zweiter Ordnung der adaptiven Schätzungen.

Im letzten Teil der Arbeit haben wir die Methodik auf die Situation mit wachsender Anzahl von zufälligen Versuchspunkten angewandt. Es ist aber zunächst nicht gelungen, die Risiken dritter Ordnung der adaptiven Schätzungen in diesem Fall zu berechnen und eine Risikoapproximation zweiter Ordnung für Kernschätzungen abzuleiten. Ähnlich wie bei der in den früheren Kapiteln untersuchten Asymptotik für adaptive Schätzer wäre nun der nächste Schritt das bisjetzt in dieser Dissertation noch offen gelassene, aber durchaus noch schwierigere Problem einer Asymptotik höherer Ordnung für adaptive Schätzungen und ihr Vergleich mit der aktuell optimalen Schätzung $\hat{m}_{h_{opt}}$ zu untersuchen. Für die Risikoapproximationen für adaptive Schätzungen ist es nämlich wichtig, explizite Ausdrücke für die Entwicklung zweiter Ordnung der Bandbreiten zu erarbeiten, welche die Kriterien minimieren, um sie letztendlich in einer Entwicklung höherer Ordnung für den Bias und die Varianz einzusetzen. Die Vermutung, daß für eine wachsende Anzahl von Versuchspunkten die Full Cross-validation Adaptation möglicherweise besser ist als die Cross-validation Adaptation für kleine bis mittlere Stichprobenumfänge, bedarf ebenfalls weitere Untersuchungen.

Kapitel 10

Hilfssätze und Beweise

10.1 Hilfssätze

Es seien

$$Y_{(n)} := (Y_{11}, \dots, Y_{1n_1}, \dots, Y_{m1}, \dots, Y_{mn_m})^T \quad (n \times 1 \text{ Vektor})$$

$$\varepsilon_{(n)} := (\varepsilon_{11}, \dots, \varepsilon_{1n_1}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{mn_m})^T \quad (n \times 1 \text{ Vektor})$$

$$\Phi_{\varepsilon_{(n)}} := E(\varepsilon_{(n)} \otimes \varepsilon_{(n)} \varepsilon_{(n)}^T)$$

$$\Psi_{\varepsilon_{(n)}} := E(\varepsilon_{(n)} \varepsilon_{(n)}^T \otimes \varepsilon_{(n)} \varepsilon_{(n)}^T).$$

Lemma 10.1.1 (*Humak Band 3, Seite 323*)

Für beliebige $n \times n$ Matrizen A und B und beliebigen Vektor $a \in \mathbb{R}^n$ gilt:

$$\begin{aligned} E(\varepsilon_{(n)}^T a \varepsilon_{(n)}^T A \varepsilon_{(n)}) &= \text{tr} \left[(a^T \otimes A) \Phi_{\varepsilon_{(n)}} \right] \\ E(\varepsilon_{(n)}^T A \varepsilon_{(n)} \varepsilon_{(n)}^T B \varepsilon_{(n)}) &= \text{tr} \left[(A \otimes B) \Psi_{\varepsilon_{(n)}} \right] \end{aligned}$$

Lemma 10.1.2

Für eine bedingt (unter der Bedingung $\Gamma_n f = 0$) erwartungstreue Varianzschätzung $\hat{\sigma}^2 = \alpha s^2 + \gamma \bar{Y}^T \Gamma_n \bar{Y}$ der Form (6.9) (wobei die Zufallsvariablen Y_{ij} nicht notwendigerweise normalverteilt sind aber endliche vierte Momente besitzen) gilt:

$$s_i^2 - \hat{\sigma}^2 = O_p(n^{-1/2}), \quad (10.1)$$

wobei $s_i^2 := (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$ ist.

Beweis:

Zu zeigen ist, daß

$$\begin{aligned} E(s_i^2 - \hat{\sigma}^2)^2 &= O(n^{-1}), \quad \text{da für } \delta > 0 \\ P(n^{1/2}(s_i^2 - \hat{\sigma}^2) > \delta) &\leq \frac{n}{\delta^2} E(s_i^2 - \hat{\sigma}^2)^2 \end{aligned}$$

durch Anwendung der Tschebycheffsche Ungleichung.

$$\begin{aligned} E(s_i^2 - \hat{\sigma}^2)^2 &= E(s_i^2 - \sigma^2 - (\hat{\sigma}^2 - \sigma^2))^2 \\ &= E(s_i^2 - \sigma^2)^2 + \text{Var}(\hat{\sigma}^2) - 2E[(s_i^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)] \\ &= O(n_i^{-1}) + \text{Var}(\hat{\sigma}^2) - 2E[(s_i^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)] \end{aligned}$$

Wir haben also jetzt zu zeigen, daß $\text{Var}(\hat{\sigma}^2) = O(n^{-1})$. Danach wird der gemischte Term

$$\begin{aligned} 2E[(s_i^2 - \sigma^2)(\hat{\sigma}^2 - \sigma^2)] &\leq 2\sqrt{E(s_i^2 - \sigma^2)^2}\sqrt{E(\hat{\sigma}^2 - \sigma^2)^2} \\ &= O(n_i^{-1/2})O(n^{-1/2}) = O(n^{-1}) \end{aligned}$$

durch Anwendung der Cauchy-Schwarzschen Ungleichung. Mit den Bezeichnung

$$G_n := \text{Diag}(\mathbb{1}_{n_1}, \dots, \mathbb{1}_{n_m}) \quad (n \times m \text{ matrix}),$$

wobei $\mathbb{1}_{n_i} := (1, \dots, 1)^T$ ein n_i -dimensionaler Vektor ist, haben wir $\bar{Y} = M_n G_n^T Y_{(n)}$. Somit ist

$$\begin{aligned} \gamma \bar{Y}^T \Gamma_n \bar{Y} &= \gamma Y_{(n)}^T G_n M_n \Gamma_n M_n G_n^T Y_{(n)} \\ &= Y_{(n)}^T \Omega_{2,n} Y_{(n)}, \quad \text{wobei } \Omega_{2,n} := \gamma G_n M_n \Gamma_n M_n G_n^T \end{aligned}$$

ist.

Mit $\Omega_n := \Omega_{1,n} + \Omega_{2,n}$, wo

$\Omega_{1,n} := \frac{\alpha}{n-m} (I_n - \text{Diag}[n_1^{-1} \mathbb{1}_{n_1} \mathbb{1}_{n_1}^T, \dots, n_m^{-1} \mathbb{1}_{n_m} \mathbb{1}_{n_m}^T])$ ist, können wir

$\hat{\sigma}^2 = Y_{(n)}^T \Omega_n Y_{(n)}$ schreiben. Es gelten die folgenden Beziehungen:

$$\begin{aligned} G_n^T G_n &= M_n^{-1} \\ \|G_n\| &= \sqrt{\lambda_{\max}(G_n^T G_n)} \\ &= \sqrt{\max_{i=1, \dots, m} (n_i)} \leq \sqrt{\sum_{i=1}^m n_i} = \sqrt{n} \end{aligned}$$

$$\begin{aligned}
\|G_n^T\| &= \sqrt{\lambda_{\max}(G_n G_n^T)} \\
&= \sqrt{\lambda_{\max}(\text{Diag}_{1 \leq i \leq m}[\mathbb{1}_{n_i} \mathbb{1}_{n_i}^T])} = \sqrt{\max_{i=1, \dots, m} (n_i)}.
\end{aligned}$$

Da $(n-m)\Omega_{1,n}/\alpha$ eine Projektionsmatrix ist, ist $\|\Omega_{1,n}\| = O(n^{-1})$.

$$\begin{aligned}
\|\Omega_{2,n}\| &\leq \gamma \|M_n\|^2 \|\Gamma_n\| \|G_n\| \|G_n^T\| \\
&= O(n^{-1})^2 O(1) \sqrt{n} \sqrt{\max_{i=1, \dots, m} (n_i)} = O(n^{-1}).
\end{aligned}$$

Also ist $\|\Omega_n\| = O(n^{-1})$.

Es ist

$$\begin{aligned}
\text{Var}(\hat{\sigma}^2) &= \text{Var}(2\varepsilon_{(n)}^T \Omega_n G_n f + \varepsilon_{(n)}^T \Omega_n \varepsilon_{(n)}) \\
&= 2\sigma^2 f^T G_n^T \Omega_n^T \Omega_n G_n f + \text{Var}(\varepsilon_{(n)}^T \Omega_n \varepsilon_{(n)}) \\
&\quad + 2\text{cov}(2\varepsilon_{(n)}^T \Omega_n G_n f, \varepsilon_{(n)}^T \Omega_n \varepsilon_{(n)}) \\
&= (1) + (2) + (3).
\end{aligned}$$

$$\begin{aligned}
(1) &\leq 2\sigma^2 \|\Omega_n\|^2 \|G_n f\|^2 \\
&= 2\sigma^2 \|\Omega_n\|^2 \sum_{i=1}^m n_i f_i^2 \\
&\leq O(n^{-2}) \max_{i=1, \dots, m} \{f_i^2\} \sum_{i=1}^m n_i = O(n^{-1}).
\end{aligned}$$

$$\begin{aligned}
(2) &\leq \|\Omega_n\|^2 \text{Var} \left(\sum_{i=1}^m \sum_{j=1}^{n_i} \varepsilon_{ij}^2 \right) \\
&= O(n^{-2}) O(n) = O(n^{-1}).
\end{aligned}$$

$$\begin{aligned}
(3) &= 4E \{ \varepsilon_{(n)}^T \Omega_n G_n f \cdot [\varepsilon_{(n)}^T \Omega_n \varepsilon_{(n)} - \sigma^2 \text{tr}(\Omega_n)] \} \\
&= 4E(\varepsilon_{(n)}^T \Omega_n G_n f \cdot \varepsilon_{(n)}^T \Omega_n \varepsilon_{(n)}) \\
&\leq 4\|\Omega_n\|^2 E(\varepsilon_{(n)}^T G_n f \cdot \varepsilon_{(n)}^T \varepsilon_{(n)}) \\
&= 4\|\Omega_n\|^2 \text{tr} \left[((G_n f)^T \otimes I_n) \Phi_{\varepsilon_{(n)}} \right] = (*).
\end{aligned}$$

10.2 Beweise

Beweis von Satz 4.2.2

Aus den Voraussetzungen **V1**, **V2** und **V3** gilt gleichmäßig auf $H_d := [0, h_d]$ die folgende Entwicklung für die Glättungsmatrix:

$$K(n, h) = I_m - hP_n + h^2Q_n - h^3R_n + h^4T_n + O(h^5). \quad (10.3)$$

(vgl. *Hairer and Wanner, S. 252*). Der Term $O(h^5)$ steht hier für $\frac{h^5}{5!} \frac{\partial^5 K(n, h)}{\partial h^5} \Big|_{h=\bar{h}}$, wobei \bar{h} zwischen 0 und h liegt. Es sei

$$\bar{Y}^{-ij} := \left(\bar{Y}_1, \dots, \bar{Y}_{i-1}, (n_i - 1)^{-1} \sum_{\substack{l=1 \\ l \neq j}}^{n_i} Y_{il}, \bar{Y}_{i+1}, \dots, \bar{Y}_m \right)^T. \quad (10.4)$$

Wir definieren

$$CV(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n-1, h)\bar{Y}^{-ij}]_i)^2. \quad (10.5)$$

Wir werden jetzt $CV(h)$ in eine etwas einfachere Gestalt bringen, mit der wir nachher $\arg \min_{h \in H} CV(h)$ berechnen werden.

$$\begin{aligned} CV(h) &:= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n-1, h)\bar{Y}]_i + [K(n-1, h)\bar{Y}]_i \\ &\quad - [K(n-1, h)\bar{Y}^{-ij}]_i)^2 \\ &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n-1, h)\bar{Y}]_i)^2 \\ &\quad + 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n-1, h)\bar{Y}]_i) \times \\ &\quad \times ([K(n-1, h)\bar{Y}]_i - [K(n-1, h)\bar{Y}^{-ij}]_i) \\ &\quad + \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} ([K(n-1, h)\bar{Y}]_i - [K(n-1, h)\bar{Y}^{-ij}]_i)^2 \\ &= S_1 + S_2 + S_3. \end{aligned}$$

Seien $K^{(i)}(n-1, h)$ die i -te Zeile und $K(n-1, h)_{ii}$ das i -te Diagonalelement von $K(n-1, h)$.

$$\begin{aligned} S_3 &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [K^{(i)}(n-1, h) (\bar{Y} - \bar{Y}^{-ij})]^2 \\ &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [K(n-1, h)_{ii}]^2 [\bar{Y}_i - (\bar{Y}^{-ij})_i]^2, \quad \text{denn} \end{aligned}$$

$$\bar{Y} - \bar{Y}^{-ij} = (0, \dots, 0, \bar{Y}_i - (\bar{Y}^{-ij})_i, 0, \dots, 0)^T.$$

Es gilt die folgende Nebenrechnung:

$$\begin{aligned} \bar{Y}_i - (\bar{Y}^{-ij})_i &= \frac{1}{n_i} \sum_{k=1}^{n_i} Y_{ik} - \frac{1}{n_i - 1} \sum_{\substack{k=1 \\ k \neq j}}^{n_i} Y_{ik} \\ &= \frac{1}{n_i} Y_{ij} + \frac{1}{n_i} T_{-j} - \frac{1}{n_i - 1} T_{-j} \quad \text{wobei} \quad T_{-j} := \sum_{\substack{k=1 \\ k \neq j}}^{n_i} Y_{ik} \\ &= \frac{1}{n_i - 1} \left(\frac{n_i - 1}{n_i} Y_{ij} + \frac{n_i - 1}{n_i} T_{-j} - T_{-j} \right) \\ &= \frac{1}{n_i - 1} \left(Y_{ij} - \frac{1}{n_i} Y_{ij} - \frac{1}{n_i} T_{-j} \right) \\ &= \frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i). \end{aligned}$$

Also ist

$$\begin{aligned} S_3 &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [K(n-1, h)_{ii}]^2 \left[\frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i) \right]^2 \\ &= \sum_{i=1}^m [K(n-1, h)_{ii}]^2 \frac{1}{n_i} \frac{1}{n_i - 1} \left\{ \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right\} \\ &= \sum_{i=1}^m [K(n-1, h)_{ii}]^2 \frac{s_i^2}{n_i(n_i - 1)} \quad \text{mit} \quad s_i^2 := \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2. \end{aligned}$$

$$S_2 = 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n-1, h)\bar{Y}]_i) K(n-1, h)_{ii} \frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i)$$

$$\begin{aligned}
&= 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - \bar{Y}_i + ([I_m - K(n-1, h)] \bar{Y})_i\} \times \\
&\quad \times K(n-1, h)_{ii} \frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i) \\
&= 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i)^2 K(n-1, h)_{ii} \\
&\quad + 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} ([I_m - K(n-1, h)] \bar{Y})_i \times \\
&\quad \times K(n-1, h)_{ii} \frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i) \\
&= 2 \sum_{i=1}^m K(n-1, h)_{ii} \frac{s_i^2}{n_i} \quad , \quad \text{denn} \quad \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i) = 0 .
\end{aligned}$$

$$\begin{aligned}
S_1 &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n-1, h)\bar{Y}]_i)^2 \\
&= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i + \bar{Y}_i - [K(n-1, h)\bar{Y}]_i)^2 \\
&= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (\bar{Y}_i - [K(n-1, h)\bar{Y}]_i)^2 \\
&\quad + 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i) (\bar{Y}_i - [K(n-1, h)\bar{Y}]_i) \\
&= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m \frac{1}{n_i} n_i ([I_m - K(n-1, h)] \bar{Y})_i^2 \\
&\quad + 2 \sum_{i=1}^m \frac{1}{n_i} (\bar{Y}_i - [K(n-1, h)\bar{Y}]_i) \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i) \\
&= \sum_{i=1}^m \frac{n_i - 1}{n_i} s_i^2 + \|[I_m - K(n-1, h)] \bar{Y}\|^2 .
\end{aligned}$$

Wegen $\lim_{n \rightarrow \infty} P_n = P$, $\lim_{n \rightarrow \infty} Q_n = Q$, $\lim_{n \rightarrow \infty} R_n = R$ und $\lim_{n \rightarrow \infty} T_n = T$ kann die Norm der Differenz $K(n, h) - K(n-1, h)$ für große n beliebig klein gemacht

werden. Das heißt,

$$\forall \eta_0 > 0, \exists N = N(\eta_0) : \|K(n, h) - K(n-1, h)\| < \eta_0 \quad \forall n \geq N.$$

Da der Term $\sum_{i=1}^m n_i^{-1}(n_i - 1)s_i^2$ von h nicht abhängt, lautet das Kriterium, das zu minimieren ist

$$\begin{aligned} CV(h) &= \|[I_m - K(n, h)]\bar{Y}\|^2 + 2 \sum_{i=1}^m \frac{s_i^2}{n_i} K(n, h)_{ii} \\ &\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i - 1)} [K(n, h)_{ii}]^2. \end{aligned} \quad (10.6)$$

Es seien

$$\begin{aligned} \hat{\omega}_n &:= \bar{Y}^T P_n^T P_n \bar{Y} \quad \hat{\nu}_n := \bar{Y}^T P_n^T Q_n \bar{Y} \\ P_n &:= (P_{n,ii}) \quad Q_n := (Q_{n,ii}) \quad R_n := (R_{n,ii}). \end{aligned}$$

Die Gleichungen (10.6) und (10.3) ergeben

$$\begin{aligned} CV(h) &= h^2 \hat{\omega}_n - 2h^3 \hat{\nu}_n + h^4 \bar{Y}^T (Q_n^T Q_n + 2P_n^T R_n) \bar{Y} \\ &\quad + 2 \sum_{i=1}^m \frac{s_i^2}{n_i} [1 - P_{n,ii}h + Q_{n,ii}h^2 - R_{n,ii}h^3 + O(h^4)] \quad (10.7) \\ &\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i - 1)} [1 - P_{n,ii}h + Q_{n,ii}h^2 - R_{n,ii}h^3 + O(h^4)]^2 \\ &\quad + O_p(h^5) \end{aligned}$$

$$\begin{aligned} CV'(h) &= 2h \hat{\omega}_n - 6h^2 \hat{\nu}_n + 4h^3 \bar{Y}^T (Q_n^T Q_n + 2P_n^T R_n) \bar{Y} \\ &\quad + 2 \sum_{i=1}^m \frac{s_i^2}{n_i} [-P_{n,ii} + 2Q_{n,ii}h - 3R_{n,ii}h^2 + O(h^3)] \quad (10.8) \\ &\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i - 1)} 2 \{ [1 - P_{n,ii}h + Q_{n,ii}h^2 - R_{n,ii}h^3 + O(h^4)] \times \\ &\quad \times [-P_{n,ii} + 2Q_{n,ii}h - 3R_{n,ii}h^2 + O(h^3)] \} + O_p(h^4) \end{aligned}$$

$$\begin{aligned}
CV''(h) &= 2\hat{\omega}_n - 12h\hat{\nu}_n + 12h^2\bar{Y}^T(Q_n^T Q_n + 2P_n^T R_n)\bar{Y} \\
&\quad + 2\sum_{i=1}^m \frac{s_i^2}{n_i} [2Q_{n,ii} - 6R_{n,ii}h + O(h^2)] \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 2 [-P_{n,ii}h + 2Q_{n,ii}h^2 - 3R_{n,ii}h^3 + O(h^4)]^2 \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 2 \{ [1 - P_{n,ii}h + Q_{n,ii}h^2 - R_{n,ii}h^3 + O(h^4)] \times \\
&\quad \times [2Q_{n,ii}h - 6R_{n,ii}h + O(h^2)] \} + O_p(h^3)
\end{aligned}$$

$$\begin{aligned}
CV'''(h) &= -12\hat{\nu}_n + 24h\bar{Y}^T(Q_n^T Q_n + 2P_n^T R_n)\bar{Y} + 2\sum_{i=1}^m \frac{s_i^2}{n_i} [-6R_{n,ii} + O(h)] \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 6 \{ [-P_{n,ii} + 2Q_{n,ii}h - 3R_{n,ii}h^2 + O(h^3)] \times \\
&\quad \times [2Q_{n,ii}h - 6R_{n,ii}h + O(h^2)] \} \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 2 \{ [1 - P_{n,ii}h + Q_{n,ii}h^2 - R_{n,ii}h^3 + O(h^4)] \times \\
&\quad \times [-6R_{n,ii} + O(h)] \} + O_p(h^2)
\end{aligned}$$

$$\begin{aligned}
CV^{(4)}(h) &= 24\bar{Y}^T(Q_n^T Q_n + 2P_n^T R_n)\bar{Y} + \sum_{i=1}^m \frac{s_i^2}{n_i} O(1) \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 6 [2Q_{n,ii} - 6R_{n,ii}h + O(h^2)]^2 \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 8 \{ [-P_{n,ii} + 2Q_{n,ii}h - 3R_{n,ii}h^2 + O(h^3)] \times \\
&\quad \times [-6R_{n,ii} + O(h)] \} \\
&\quad + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} 2 [1 - P_{n,ii}h + Q_{n,ii}h^2 - R_{n,ii}h^3 + O(h^4)] O(1) \\
&\quad + O_p(h) .
\end{aligned}$$

Um die Nullstellen von $CV'(h)$ zu approximieren, wenden wir eine Methode, die in *Lawley (1956)* oder in *Barndorff-Nielsen (1988)* zur approximativen

Berechnung von wahren Parametern in statistischen Modellen angewandt wurde.

Es sei nämlich $\hat{h}_{cv} \in \arg \min_{h \in H} CV(h) = \{\tilde{h} : CV(\tilde{h}) = \min CV(h)\}$.

Man schreibt $CV'(\hat{h}_{cv}) = 0$ und macht eine Taylorentwicklung im Punkt $h = 0$.

$$\begin{aligned} 0 &= CV'(\hat{h}_{cv}) \\ &= CV'(0) + CV''(0) \hat{h}_{cv} + 1/2 CV'''(0) \hat{h}_{cv}^2 \\ &\quad + 1/6 CV^{(4)}(0) \hat{h}_{cv}^3 + O_p(\bar{h}^4) \end{aligned}$$

wobei \bar{h} zwischen 0 und \hat{h}_{cv} liegt.

Es folgt daraus

$$\begin{aligned} \hat{h}_{cv} &= -[CV''(0)]^{-1} \left\{ CV'(0) + 1/2 \hat{h}_{cv}^2 CV'''(0) \right. \\ &\quad \left. + 1/6 \hat{h}_{cv}^3 CV^{(4)}(0) + O_p(\bar{h}^4) \right\}. \end{aligned} \quad (10.9)$$

Ein sukzessives Einsetzen der Formel (10.9) in sich hat als Folge

$$\hat{h}_{cv} = A_0 + A_1 + A_2 + \dots \quad (10.10)$$

wobei

$$\begin{aligned} A_0 &= -[CV''(0)]^{-1} CV'(0) \\ A_1 &= -1/2 [CV''(0)]^{-3} CV'''(0) [CV'(0)]^2 \\ A_2 &= 1/6 [CV''(0)]^{-4} CV^{(4)}(0) [CV'(0)]^3 \end{aligned}$$

Es gilt (vgl. *Barndorff-Nielsen (1988), Seite 44*)

$$A_\nu = o_p(n^{-(\nu+1)/2}) \quad \nu \geq 3$$

und demzufolge ist $A_3 + A_4 + \dots = o_p(n^{-2})$. Wir haben

$$\begin{aligned} CV'(0) &= 2 \sum_{i=1}^m \frac{s_i^2}{n_i} (-P_{n,ii}) + 2 \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (-P_{n,ii}) \\ CV''(0) &= 2\hat{\omega}_n + 2 \sum_{i=1}^m \frac{s_i^2}{n_i} (2Q_{n,ii}) + 2 \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (P_{n,ii}^2 + 2Q_{n,ii}) \\ CV'''(0) &= -12\hat{\nu}_n + 2 \sum_{i=1}^m \frac{s_i^2}{n_i} (6R_{n,ii}) \\ &\quad + 2 \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (-6P_{n,ii}Q_{n,ii} - 6R_{n,ii}) \end{aligned}$$

$$CV^{(4)}(0) = O_p(1).$$

Zur Berechnung von A_0 , A_1 und A_2 wenden wir die folgende Entwicklung an:

$$\begin{aligned} \frac{1}{x+a} &= \frac{1}{x} + \left(\frac{1}{x+a} - \frac{1}{x} \right) \\ &= \frac{1}{x} - \frac{a}{x(x+a)} \\ &= \frac{1}{x} - \frac{a}{x^2} + \frac{a^2}{x^2(x+a)}. \end{aligned}$$

$$\begin{aligned} A_0 &= - \frac{2 \sum_{i=1}^m \frac{s_i^2}{n_i} (-P_{n,ii}) + 2 \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (-P_{n,ii})}{2\hat{\omega}_n + 2 \sum_{i=1}^m \frac{s_i^2}{n_i} (2Q_{n,ii}) + 2 \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (P_{n,ii}^2 + 2Q_{n,ii})} \\ &= \frac{- \sum_{i=1}^m \frac{s_i^2}{n_i} (-P_{n,ii})}{\hat{\omega}_n + \sum_{i=1}^m \frac{s_i^2}{n_i} (2Q_{n,ii}) + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (P_{n,ii}^2 + 2Q_{n,ii})} \\ &\quad + \frac{- \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (-P_{n,ii})}{\hat{\omega}_n + \sum_{i=1}^m \frac{s_i^2}{n_i} (2Q_{n,ii}) + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (P_{n,ii}^2 + 2Q_{n,ii})} \\ &= \frac{\sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii}}{\hat{\omega}_n} - \frac{\left(\sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii} \right) \left\{ \sum_{i=1}^m \frac{s_i^2}{n_i} 2Q_{n,ii} + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (P_{n,ii}^2 + 2Q_{n,ii}) \right\}}{\hat{\omega}_n^2} \\ &\quad + \frac{\sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} P_{n,ii}}{\hat{\omega}_n} + O_p \left(\max_{1 \leq i \leq m} \left\{ \frac{s_i^2}{n_i^3 \hat{\omega}_n^2} \right\} \right) \\ &= \frac{\sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii}}{\hat{\omega}_n} - \frac{\left\{ \sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii} \right\} \left(\sum_{i=1}^m \frac{s_i^2}{n_i} 2Q_{n,ii} \right)}{\hat{\omega}_n^2} \\ &\quad + \frac{\sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} P_{n,ii}}{\hat{\omega}_n} + O_p \left(\max_{1 \leq i \leq m} \left\{ \frac{s_i^2}{n_i^3 \hat{\omega}_n^2} \right\} \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{s^2 \sum_{i=1}^m \frac{P_{n,ii}}{n_i} + \sum_{i=1}^m \frac{(s_i^2 - s^2)}{n_i} P_{n,ii}}{\hat{\omega}_n} - \frac{\left(\sum_{i=1}^m \frac{(s^2 + s_i^2 - s^2)}{n_i} P_{n,ii} \right) \left(\sum_{i=1}^m \frac{(s^2 + s_i^2 - s^2)}{n_i} 2Q_{n,ii} \right)}{\hat{\omega}_n^2} \\
&\quad + \frac{\sum_{i=1}^m \frac{(s^2 + s_i^2 - s^2)}{n_i(n_i-1)} P_{n,ii}}{\hat{\omega}_n} + O_p \left(\max_{1 \leq i \leq m} \left\{ \frac{s_i^2}{n_i^3 \hat{\omega}_n^2} \right\} \right) \\
&= \frac{s^2 \lambda_n}{\hat{\omega}_n} + \frac{\sum_{i=1}^m \frac{(s_i^2 - s^2)}{n_i} P_{n,ii}}{\hat{\omega}_n} - \frac{2s^4 \lambda_n \mu_n}{\hat{\omega}_n^2} + \frac{s^2 \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2}}{\hat{\omega}_n} \\
&\quad + O_p \left(n^{-5/2} + \max_{1 \leq i \leq m} \left\{ \frac{s_i^2}{n_i^3 \hat{\omega}_n^2} \right\} \right),
\end{aligned}$$

wobei $\mu_n = \text{tr}(Q_n M_n)$

Wir haben in dieser letzten Gleichung Lemma 10.1.2 mit $\gamma = 0$ und $\alpha = 1$ angewandt. Das Lemma 10.1.2 werden wir auch in der nächsten Berechnung anwenden.

$$\begin{aligned}
A_1 &= \frac{1}{2} \frac{\left[-12\hat{\nu}_n + O_p \left(\max_{1 \leq i \leq m} \left\{ \frac{s_i^2}{n_i} \right\} \right) \right]}{8 \left[\hat{\omega}_n + \sum_{i=1}^m \frac{s_i^2}{n_i} (2Q_{n,ii}) + \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (P_{n,ii}^2 + 2Q_{n,ii}) \right]^3} \times \\
&\quad \times \left[2 \sum_{i=1}^m \frac{s_i^2}{n_i} (-P_{n,ii}) + 2 \sum_{i=1}^m \frac{s_i^2}{n_i(n_i-1)} (-P_{n,ii}) \right]^2 \\
&= \frac{3\hat{\nu}_n}{\hat{\omega}_n^3} \left[\sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii} \right]^2 + O_p \left(\max_{1 \leq i \leq m} \left\{ \frac{s_i^4}{n_i^3 \hat{\omega}_n^3} \right\} \right) \\
&= \frac{3\hat{\nu}_n}{\hat{\omega}_n^3} \left[\sum_{i=1}^m \frac{s^2 + s_i^2 - s^2}{n_i} P_{n,ii} \right]^2 + O_p \left(\max_{1 \leq i \leq m} \left\{ \frac{s_i^4}{n_i^3 \hat{\omega}_n^3} \right\} \right) \\
&= \frac{3s^4 \lambda_n^2 \hat{\nu}_n}{\hat{\omega}_n^3} + O_p \left(n^{-5/2} + \max_{1 \leq i \leq m} \left\{ \frac{s_i^4}{n_i^3 \hat{\omega}_n^3} \right\} \right)
\end{aligned}$$

$$\begin{aligned}
A_2 &= 1/6 \frac{O_p(1) \cdot O_p\left(\max_{1 \leq i \leq m} \{s_i^6 n_i^{-3}\}\right)}{\hat{\omega}_n^4} \\
&= O_p\left(\max_{1 \leq i \leq m} \left\{\frac{s_i^6}{n_i^3 \hat{\omega}_n^4}\right\}\right).
\end{aligned}$$

Auf Seite 90 werden wir zeigen, daß $\hat{h}_{cp} \in \arg \min_{h \in H} \hat{M}_{cp}$ die asymptotische Beziehung

$$\hat{h}_{cp} = \frac{s^2 \lambda_n}{\hat{\omega}_n} - 2 \frac{s^4 \lambda_n \mu_n}{\hat{\omega}_n^2} + \frac{3s^4 \lambda_n^2 \hat{\nu}_n}{\hat{\omega}_n^3} + O_p(s^6 n^{-3}) + o_p(s^4 n^{-2}) \quad (10.11)$$

genügt. Aus (10.10) und (10.11) erhält man:

$$\begin{aligned}
\hat{h}_{cv} &= \hat{h}_{cp} + \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] + \frac{s^2}{\hat{\omega}_n} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + o_p(n^{-2}) \quad (10.12) \\
&= \hat{h}_{cp} + a_1 + a_2 + o_p(n^{-2}).
\end{aligned}$$

$$\begin{aligned}
E \|\hat{f}_{cv} - f\|^2 &= E \|K(n, \hat{h}_{cv}) - f\|^2 \\
&= E \left\| \left[I_m - \hat{h}_{cv} P_n + \hat{h}_{cv}^2 Q_n + O_p(\hat{h}_{cv}^3) \right] \bar{Y} - f \right\|^2 \\
&= E \left\| \left[I_m - \hat{h}_{cp} P_n + \hat{h}_{cp}^2 Q_n + O_p(\hat{h}_{cp}^3) \right] \bar{Y} - f \right\|^2 \\
&\quad + E \left\| \left[-(a_1 + a_2 + o_p(n^{-2})) P_n + 2(a_1 + a_2 + o_p(n^{-2})) \hat{h}_{cp} Q_n + \right. \right. \\
&\quad \quad \left. \left. + (a_1 + a_2 + o_p(n^{-2}))^2 Q_n \right] \bar{Y} \right\|^2 + \\
&\quad + 2E \left\{ \left[\bar{\varepsilon} - \hat{h}_{cp} P_n \bar{Y} + \hat{h}_{cp}^2 Q_n \bar{Y} + O_p(\hat{h}_{cp}^3 \bar{Y}) \right]^T \times \right. \\
&\quad \quad \times \left[-(a_1 + a_2 + o_p(n^{-2})) P_n + 2(a_1 + a_2 + o_p(n^{-2})) \hat{h}_{cp} Q_n + \right. \\
&\quad \quad \left. \left. + (a_1 + a_2 + o_p(n^{-2}))^2 Q_n \right] \bar{Y} \right\} \\
&= V_1 + V_2 + V_3.
\end{aligned}$$

Es gilt:

$$\begin{aligned}
V_1 &= E \left\| \left[I_m - \hat{h}_{cp} P_n + \hat{h}_{cp}^2 Q_n + O_p(\hat{h}_{cv}^3) \right] \bar{Y} - f \right\|^2 \\
&= E \|\hat{f}_{cp} - f\|^2,
\end{aligned}$$

denn aus (10.12) gilt $\hat{h}_{cv} = \hat{h}_{cp} + O_p(n^{-3/2})$ und demzufolge ist $O_p(\hat{h}_{cv}) = O_p(\hat{h}_{cp})$.

$$\begin{aligned}
V_2 &= E \left\| \left[-(a_1 + a_2 + o_p(n^{-2})) P_n + 2(a_1 + a_2 + o_p(n^{-2})) \hat{h}_{cp} Q_n + \right. \right. \\
&\quad \left. \left. + (a_1 + a_2 + o_p(n^{-2}))^2 Q_n \right] \bar{Y} \right\|^2 \\
&= E \{ a_1^2 \|P_n \bar{Y}\|^2 \} + O(n^{-4}) \\
&= E \left\{ \frac{1}{\hat{\omega}_n^2} \left[\sum_{i=1}^m \frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right]^2 \|P_n \bar{Y}\|^2 \right\} + O(n^{-4}) \\
&= E \left\{ \frac{1}{\hat{\omega}_n} \left[\sum_{i=1}^m \frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right]^2 \right\} + O(n^{-4}),
\end{aligned}$$

denn die anderen Terme sind von kleinerer Ordnung.

a

$$\begin{aligned}
E \left\{ \left[\sum_{i=1}^m \frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right]^2 \right\} &= \sum_{i=1}^m \frac{P_{n,ii}^2}{n_i^2} E(s_i^2 - s^2)^2 \\
&\quad + \sum_{\substack{i,j=1 \\ i \neq j}}^m \frac{P_{n,ii} P_{n,jj}}{n_i n_j} E(s_i^2 - s^2)(s_j^2 - s^2).
\end{aligned}$$

Dabei ist, da s_i^2 für $i \neq k$ unabhängig von s_k^2 ist

$$\begin{aligned}
E(s_i^2 - s^2)^2 &= E(s_i^2 - \sigma^2)^2 + E(s^2 - \sigma^2)^2 - 2E(s_i^2 - \sigma^2)(s^2 - \sigma^2) \\
&= \frac{2\sigma^4}{n_i - 1} + \frac{2\sigma^4}{n - m} - 2E \left[(s_i^2 - \sigma^2) \frac{1}{n - m} \sum_{j=1}^m (n_j - 1) s_j^2 \right], \\
&= \frac{2\sigma^4}{n_i - 1} + \frac{2\sigma^4}{n - m} - 2 \frac{n_i - 1}{n - m} E \left[(s_i^2 - \sigma^2) s_i^2 \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{2\sigma^4}{n_i - 1} + \frac{2\sigma^4}{n - m} - 2\frac{n_i - 1}{n - m}E(s_i^4 - \sigma^4) \\
&= \frac{2\sigma^4}{n_i - 1} - \frac{2\sigma^4}{n - m}.
\end{aligned}$$

$$\begin{aligned}
E(s_i^2 - s^2)(s_j^2 - s^2) &= E(s_i^2 - \sigma^2)(s_j^2 - \sigma^2) - E(s_i^2 - \sigma^2)(s^2 - \sigma^2) \\
&\quad - E(s_j^2 - \sigma^2)(s^2 - \sigma^2) + E(s^2 - \sigma^2)^2 \\
&= E(s_i^2 - \sigma^2)E(s_j^2 - \sigma^2) - E\left[(s_i^2 - \sigma^2)\frac{1}{n - m}\sum_{k=1}^m(n_k - 1)s_k^2\right] \\
&\quad - E\left[(s_j^2 - \sigma^2)\frac{1}{n - m}\sum_{k=1}^m(n_k - 1)s_k^2\right] + Var(s^2) \\
&= 0 - \frac{n_i - 1}{n - m}E[(s_i^2 - \sigma^2)s_i^2] - \frac{n_j - 1}{n - m}E[(s_j^2 - \sigma^2)s_j^2] + Var(s^2) \\
&= 0 - \frac{2\sigma^4}{n - m} - \frac{2\sigma^4}{n - m} + \frac{2\sigma^4}{n - m} \\
&= -\frac{2\sigma^4}{n - m}.
\end{aligned}$$

Deshalb ist

$$\begin{aligned}
E\left\{\left[\sum_{i=1}^m \frac{(s_i^2 - s^2)}{n_i} P_{n,ii}\right]^2\right\} &= \sum_{i=1}^m \frac{P_{n,ii}^2}{n_i^2} \left(\frac{2\sigma^4}{n_i - 1} - \frac{2\sigma^4}{n - m}\right) \\
&\quad - \sum_{\substack{i,j=1 \\ i \neq j}}^m \frac{P_{n,ii}P_{n,jj}}{n_i n_j} \frac{2\sigma^4}{n - m}.
\end{aligned}$$

b

$$E(\hat{\omega}_n) = \omega_n + \sigma^2 \delta_n.$$

Insgesamt ist

$$V_2 = \frac{\sum_{i=1}^m \frac{P_{n,ii}^2}{n_i^2} \left(\frac{2\sigma^4}{n_i - 1} - \frac{2\sigma^4}{n - m}\right) - \sum_{\substack{i,j=1 \\ i \neq j}}^m \frac{P_{n,ii}P_{n,jj}}{n_i n_j} \frac{2\sigma^4}{n - m}}{\omega_n} + O(n^{-4})$$

$$\begin{aligned}
&= \frac{\sum_{i=1}^m \frac{P_{n,ii}^2}{n_i^2} \frac{2\sigma^4}{n_i-1} + O(n^{-4}) - \frac{2\sigma^4}{n-m} \left(\sum_{i=1}^m \frac{P_{n,ii}}{n_i} \right)^2}{\omega_n} \\
&= \frac{2\sigma^4}{n^3\omega} \left\{ \sum_{i=1}^m P_{ii}^2 c_i^{-3} - \lambda^2 \right\} + O(n^{-4}).
\end{aligned}$$

Für die Berechnung von V_3 ist zu beachten (*Patel und Read (1996), Seite 85*), daß :

- $\bar{\varepsilon}_i$ und s_i^2 sind unabhängig.
- $\bar{\varepsilon}_i$ und s_j^2 sind unabhängig $i \neq j$.
- $\bar{\varepsilon}$ und s_i^2 sind unabhängig.
- $\bar{\varepsilon}$ und s^2 sind unabhängig.

$$\begin{aligned}
V_3 &= 2E \left\{ -\bar{\varepsilon}^T (a_1 + a_2) P_n \bar{Y} + \bar{\varepsilon}^T 2(a_1 + a_2) \hat{h}_{cp} Q_n \bar{Y} \right. \\
&\quad \left. + \hat{h}_{cp} (P_n \bar{Y})^T (a_1 + a_2) P_n \bar{Y} + o(n^{-3}) \right\} \\
&= 2E \left\{ -\bar{\varepsilon}^T \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] P_n \bar{Y} \right\} \\
&\quad + 2E \left\{ -\bar{\varepsilon}^T \frac{s^2}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{P_{n,ii}}{n_i^2} \right] P_n \bar{Y} \right\} \\
&\quad + 2E \left\{ \hat{h}_{cp} (P_n \bar{Y})^T \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] P_n \bar{Y} \right\} \\
&\quad + 2E \left\{ \hat{h}_{cp} (P_n \bar{Y})^T \left[\frac{s^2}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{P_{n,ii}}{n_i^2} \right] P_n \bar{Y} \right] \right\} + o(n^{-3}) \\
&= Z_1 + \dots + Z_4 + o(n^{-3}).
\end{aligned}$$

$$\begin{aligned}
Z_1 &= -2E \left\{ \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] \bar{\varepsilon}^T P_n \bar{Y} \right\} \\
&= -2E \left\{ \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{s_i^2}{n_i} P_{n,ii} \right] \bar{\varepsilon}^T P_n \bar{Y} \right\} + 2E \left\{ \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{s^2}{n_i} P_{n,ii} \right] \bar{\varepsilon}^T P_n \bar{Y} \right\} \\
&= 0,
\end{aligned}$$

denn die Funktionen $g_1(s_i^2, \bar{\varepsilon}^T P_n \bar{Y}, \hat{\omega}_n) := \frac{s_i^2 \bar{\varepsilon}^T P_n \bar{Y}}{\hat{\omega}_n}$ und $g_2(s^2, \bar{\varepsilon}^T P_n \bar{Y}, \hat{\omega}_n) := \frac{s^2 \bar{\varepsilon}^T P_n \bar{Y}}{\hat{\omega}_n}$ haben dieselbe Taylorentwicklung im Punkt $(\sigma^2, \sigma^2 \lambda_n, \omega_n + \sigma^2 \delta_n)$.

$$\begin{aligned} Z_2 &= -2E \left\{ \frac{s^2}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{P_{n,ii}}{n_i^2} \right] \bar{\varepsilon}^T P_n \bar{Y} \right\} \\ &= -2 \left\{ E \left[\frac{s^2 \bar{\varepsilon}^T P_n f}{\hat{\omega}_n} \right] + E \left[\frac{s^2 \bar{\varepsilon}^T P_n \bar{\varepsilon}}{\hat{\omega}_n} \right] \right\} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} \\ &= -2 \{(1) + (2)\} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2}. \end{aligned}$$

$$(1) = E \left\{ -\frac{\sigma^2}{(\omega_n + \sigma^2 \delta_n)^2} (\bar{\varepsilon}^T P_n f - 0)(2\bar{\varepsilon}^T P_n^T P_n f + \bar{\varepsilon}^T P_n^T P_n \bar{\varepsilon} - \sigma^2 \delta_n) + O_p(n^{-3/2}) \right\}$$

$$= -\frac{2\sigma^4 f^T P_n^T P_n M_n P_n f}{(\omega_n + \sigma^2 \delta_n)^2} + O(n^{-3/2}).$$

$$(2) = \frac{\sigma^4 \lambda_n}{(\omega_n + \sigma^2 \delta_n)} + O(n^{-3/2}).$$

Insgesamt ist

$$Z_2 = 2 \left\{ \frac{2\sigma^4 f^T P_n^T P_n M_n P_n f}{\omega_n^2} - \frac{\sigma^4 \lambda_n}{\omega_n} \right\} \cdot \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + O(n^{-7/2}).$$

$$\begin{aligned} Z_3 &= 2E \left\{ \hat{h}_{cp}(P_n \bar{Y})^T \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] P_n \bar{Y} \right\} \\ &= 0. \end{aligned}$$

$$\begin{aligned} Z_4 &= 2E \left\{ \hat{h}_{cp} \frac{s^2}{\hat{\omega}_n} \|P_n \bar{Y}\|^2 \right\} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} \\ &= 2E \left\{ \frac{s^2 \lambda_n s^2}{\hat{\omega}_n} \right\} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + O(n^{-4}) \\ &= 2 \frac{\sigma^4 \lambda_n}{\omega_n} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + O(n^{-4}). \end{aligned}$$

Also ist

$$\begin{aligned} V_3 &= \frac{4\sigma^4 f^T P_n^T P_n M_n P_n f}{\omega_n^2} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + o(n^{-3}) \\ &= \frac{4\sigma^4 f^T P^T P C P f}{n^3 \omega^2} \sum_{i=1}^m P_{ii} c_i^{-2} + o(n^{-3}) \end{aligned}$$

Aus Lemma 10.1.3 folgt $V_2 \geq 0$. Zusammen hat man:

$$E\|\hat{f}_{cv} - f\|^2 = E\|\hat{f}_{cp} - f\|^2 + V_2 + V_3 + o(n^{-3}). \quad (10.13)$$

Das Korollar 4.2.1 folgt aus den Sätzen 4.2.1 und 4.2.2.

Beweis von (10.11)

Aus der Gleichung (3.16) haben wir

$$\begin{aligned} \hat{M}_{cp}(h) &= \|[I_m - K(n, h)]\bar{Y}\|^2 + s^2 \text{tr} \{ [2K(n, h) - I_m] M_n \} \\ &= h^2 \hat{\omega}_n - 2h^3 \hat{\nu}_n + s^2 \text{tr}(M_n) - 2s^2 h \lambda_n \\ &\quad + 2s^2 h^2 \mu_n + O_p(s^2 h^3 \|M_n\| + h^4 \|\bar{Y}\|^2). \end{aligned}$$

Die Ableitungen von \hat{M}_{cp} im Punkt 0 sind:

$$\begin{aligned} \hat{M}'_{cp}(0) &= [2h\hat{\omega}_n - 6h^2\hat{\nu}_n - 2s^2\lambda_n + 4s^2h\mu_n + O_p(s^2h^2\|M_n\| + h^3\|\bar{Y}\|^2)]_{|h=0} \\ &= -2s^2\lambda_n \end{aligned}$$

$$\begin{aligned} \hat{M}''_{cp}(0) &= [2\hat{\omega}_n - 12h\hat{\nu}_n + 4s^2\mu_n + O_p(s^2h\|M_n\| + h^2\|\bar{Y}\|^2)]_{|h=0} \\ &= 2\hat{\omega}_n + 4s^2\mu_n \end{aligned}$$

$$\begin{aligned} \hat{M}'''_{cp}(0) &= [-12\hat{\nu}_n + 4s^2\mu_n + O_p(s^2\|M_n\| + h\|\bar{Y}\|^2)]_{|h=0} \\ &= -12\hat{\nu}_n + O_p(n^{-1}s^2) \end{aligned}$$

$$\hat{M}^{(4)}_{cp}(0) = O_p(\|\bar{Y}\|^2).$$

$$\Rightarrow \hat{h}_{cp} = -\frac{\hat{M}'_{cp}(0)}{\hat{M}''_{cp}(0)} - \frac{1}{2} \frac{\hat{M}'''_{cp}(0)[\hat{M}'_{cp}(0)]^2}{[\hat{M}''_{cp}(0)]^3} + o_p(n^{-2}s^4)$$

$$\begin{aligned}
&= \frac{s^2 \lambda_n}{\hat{\omega}_n + 2s^2 \mu_n} + \frac{3s^4 \lambda_n^2 [\hat{\nu}_n + O_p(n^{-1} s^2)]}{(\hat{\omega}_n + 2s^2 \mu_n)^3} + o_p(n^{-2} s^4) \\
&= \frac{s^2 \lambda_n}{\hat{\omega}_n} - \frac{2s^4 \lambda_n \mu_n}{\hat{\omega}_n^2} + \frac{3s^4 \lambda_n^2 \hat{\nu}_n}{\hat{\omega}_n^3} + O_p(n^{-3} s^6) + o_p(n^{-2} s^4) .
\end{aligned}$$

Beweis von Satz 4.1.3

Der Beweis dieses Satzes folgt aus dem Beweis des Satzes 4.2.2. Alle Bandbreiten haben dieselbe asymptotische Entwicklung, nämlich

$$\begin{aligned}
\hat{h}_{p-i} &= \frac{s^2 \lambda_n}{\hat{\omega}_n} + o_p(n^{-1}) \\
\hat{h}_{cp} &= \hat{h}_{p-i} + o_p(n^{-1}) \\
\hat{h}_{cv} &= \hat{h}_{cp} + o_p(n^{-1}) \\
\hat{h}_{fcv1} &= \hat{h}_{cv} + o_p(n^{-1}) \\
\hat{h}_{fcv2} &= \hat{h}_{fcv1} + o_p(n^{-1}) .
\end{aligned}$$

Analog wie auf der Seite 85 erhält man die Gleichheit der entsprechenden Risiken.

Beweis von Satz 4.2.3

Die Idee zum Beweis von Satz 4.2.3 ist dieselbe wie beim Beweis von Satz 4.2.2. Es sei

$$FCV1(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n, h) \bar{Y}^{+ij}]_i)^2 , \quad (10.14)$$

wobei

$$\bar{Y}^{+ij} := \left(\bar{Y}_1, \dots, \bar{Y}_{i-1}, \hat{Y}_{+ij}, \bar{Y}_{i+1}, \dots, \bar{Y}_m \right)^T , \quad i = 1, \dots, m ; j = 1, \dots, n_i ,$$

mit

$$\hat{Y}_{+ij} := \frac{1}{n_i} \left(\sum_{\substack{k=1 \\ k \neq j}}^{n_i} Y_{ik} + \bar{Y}_i \right) .$$

$$\begin{aligned}
FCV1(h) &:= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n, h)\bar{Y}]_i + [K(n, h)\bar{Y}]_i - [K(n, h)\bar{Y}^{+ij}]_i)^2 \\
&= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n, h)\bar{Y}]_i)^2 \\
&\quad + 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n, h)\bar{Y}]_i) ([K(n, h)\bar{Y}]_i - [K(n, h)\bar{Y}^{+ij}]_i) \\
&\quad + \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} ([K(n, h)\bar{Y}]_i - [K(n, h)\bar{Y}^{+ij}]_i)^2 \\
&= S_1 + S_2 + S_3 .
\end{aligned}$$

Es sei $K^{(i)}(n, h)$ die i -te Zeile von $K(n, h)$.

$$\begin{aligned}
S_3 &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [K^{(i)}(n, h) (\bar{Y} - \bar{Y}^{+ij})]^2 \\
&= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [K(n, h)_{ii}]^2 [\bar{Y}_i - \hat{Y}^{+ij}]^2 ,
\end{aligned}$$

denn $\bar{Y} - \bar{Y}^{+ij} = (0, \dots, 0, \bar{Y}_i - \hat{Y}^{+ij}, 0, \dots, 0)^T$. Es gilt:

$$\begin{aligned}
\bar{Y}_i - \hat{Y}^{+ij} &= \frac{1}{n_i} \sum_{k=1}^{n_i} Y_{ik} - \frac{1}{n_i} \left(\sum_{\substack{k=1 \\ k \neq j}}^{n_i} Y_{ik} + \bar{Y}_i \right) \\
&= \frac{1}{n_i} Y_{ij} - \frac{1}{n_i^2} \sum_{k=1}^{n_i} Y_{ik} \\
&= \frac{1}{n_i} (Y_{ij} - \bar{Y}_i) .
\end{aligned}$$

$$\begin{aligned}
\Rightarrow S_3 &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [K(n, h)_{ii}]^2 \left\{ \frac{1}{n_i} (Y_{ij} - \bar{Y}_i) \right\}^2 \\
&= \sum_{i=1}^m \frac{1}{n_i^3} [K(n, h)_{ii}]^2 \sum_{j=1}^{n_i} (n_i - 1) \left\{ \frac{1}{n_i - 1} (Y_{ij} - \bar{Y}_i) \right\}^2 \\
&= \sum_{i=1}^m [K(n, h)_{ii}]^2 \frac{n_i - 1}{n_i^3} s_i^2
\end{aligned}$$

$$\begin{aligned}
S_2 &= 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - [K(n, h)\bar{Y}]_i) K(n, h)_{ii} \left\{ \frac{1}{n_i} (Y_{ij} - \bar{Y}_i) \right\} \\
&= 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i + \bar{Y}_i - [K(n, h)\bar{Y}]_i) K(n, h)_{ii} \left\{ \frac{1}{n_i} (Y_{ij} - \bar{Y}_i) \right\} \\
&= 2 \sum_{i=1}^m \frac{1}{n_i} K(n, h)_{ii} \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + 0 \\
&= 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} K(n, h)_{ii} s_i^2 .
\end{aligned}$$

$$S_1 = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \|[I_m - K(n, h)]\bar{Y}\|^2 .$$

Das zu minimierende Kriterium lautet:

$$\begin{aligned}
FCV1(h) &= \|[I_m - K(n, h)]\bar{Y}\|^2 + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} K(n, h)_{ii} s_i^2 \\
&\quad + \sum_{i=1}^m [K(n, h)_{ii}]^2 \frac{n_i - 1}{n_i^3} s_i^2 . \tag{10.15}
\end{aligned}$$

Es gilt:

$$\begin{aligned}
FCV1'(0) &= 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} s_i^2 (-P_{n,ii}) + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^3} s_i^2 (-P_{n,ii}) \\
FCV1''(0) &= 2\hat{\omega}_n + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} s_i^2 (2Q_{n,ii}) + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^3} s_i^2 (P_{n,ii}^2 + 2Q_{n,ii}) \\
FCV1'''(0) &= -12\hat{\nu}_n + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} s_i^2 (-6R_{n,ii}) \\
&\quad + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^3} s_i^2 (-P_{n,ii}Q_{n,ii} - 6R_{n,ii}) \\
FCV1^{(4)}(0) &= O_p(1) .
\end{aligned}$$

Analog zum vorigen Beweis läßt sich $\hat{h}_{fcv1} \in \arg \min_{h \in H} FCV1(h)$ in der Form

$$\hat{h}_{fcv1} = A_0 + A_1 + A_2 + o_p(n^{-2}) \quad \text{darstellen, wobei}$$

$$\begin{aligned} A_0 &= -\frac{FCV1'(0)}{FCV1''(0)} \\ &= \frac{\sum_{i=1}^m \frac{n_i-1}{n_i^2} s_i^2 P_{n,ii}}{\hat{\omega}_n} - \frac{\left(\sum_{i=1}^m \frac{n_i-1}{n_i^2} s_i^2 P_{n,ii}\right) \left(\sum_{i=1}^m \frac{n_i-1}{n_i^2} s_i^2 (2Q_{n,ii})\right)}{\hat{\omega}_n^2} \\ &\quad + \frac{\sum_{i=1}^m \frac{n_i-1}{n_i^3} s_i^2 P_{n,ii}}{\hat{\omega}_n} + O_p(\max_{1 \leq i \leq m} n_i^{-3}) \\ &= \frac{\sum_{i=1}^m \frac{s_i^2}{n_i^2} P_{n,ii}}{\hat{\omega}_n} - \frac{\left(\sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii}\right) \left(\sum_{i=1}^m \frac{s_i^2}{n_i} 2Q_{n,ii}\right)}{\hat{\omega}_n^2} + O_p(\max_{1 \leq i \leq m} n_i^{-3}) \\ &= \frac{s^2 \lambda_n}{\hat{\omega}_n} + \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] - \frac{2s^4 \lambda_n \mu_n}{\hat{\omega}_n^2} + O_p\left(n^{-5/2} + \max_{1 \leq i \leq m} n_i^{-3}\right) \end{aligned}$$

$$\begin{aligned} A_1 &= -\frac{1}{2} \frac{FCV1'''(0)[FCV1'(0)]^2}{[FCV1''(0)]^3} \\ &= \frac{3\hat{\nu}_n}{\hat{\omega}_n^3} \left(\sum_{i=1}^m \frac{n_i-1}{n_i^2} s_i^2 P_{n,ii}\right)^2 + O_p(\max_{1 \leq i \leq m} n_i^{-3}) \\ &= \frac{3\hat{\nu}_n}{\hat{\omega}_n^3} \left(\sum_{i=1}^m \frac{s_i^2}{n_i} P_{n,ii}\right)^2 + O_p(\max_{1 \leq i \leq m} n_i^{-3}) \\ &= \frac{3s^4 \lambda_n^2 \hat{\nu}_n}{\hat{\omega}_n^3} + O_p\left(n^{-5/2} + \max_{1 \leq i \leq m} n_i^{-3}\right) \end{aligned}$$

$$\begin{aligned} A_2 &= -\frac{1}{6} \frac{FCV1^{(4)}(0)[FCV1'(0)]^3}{[FCV1''(0)]^4} \\ &= O_p(\max_{1 \leq i \leq m} n_i^{-3}). \end{aligned}$$

$$\begin{aligned} \Rightarrow \hat{h}_{fcv1} &= \frac{s^2 \lambda_n}{\hat{\omega}_n} + \frac{1}{\hat{\omega}_n} \sum_{i=1}^m \left[\frac{(s_i^2 - s^2)}{n_i} P_{n,ii} \right] - \frac{2s^4 \lambda_n \mu_n}{\hat{\omega}_n^2} \\ &\quad + \frac{3s^4 \lambda_n^2 \hat{\nu}_n}{\hat{\omega}_n^3} + O_p\left(n^{-5/2} + \max_{1 \leq i \leq m} n_i^{-3}\right). \end{aligned} \quad (10.16)$$

Mit den Gleichungen (10.11) und (10.12) erhalten wir

$$\begin{aligned}\hat{h}_{fcv1} &= \hat{h}_{cv} - \frac{s^2}{\hat{\omega}_n} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + o_p(n^{-2}) \\ &= \hat{h}_{cv} - b_{cv} + o_p(n^{-2}),\end{aligned}\tag{10.17}$$

wobei $b_{cv} := \frac{s^2}{\hat{\omega}_n} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2}$ ist.

$$\begin{aligned}E\|\hat{f}_{fcv1} - f\|^2 &= E\left\| \left[I_m - \hat{h}_{fcv1} P_n + \hat{h}_{fcv1}^2 Q_n + O_p(\hat{h}_{fcv1}^3) \right] \bar{Y} - f \right\|^2 \\ &= E\left\| \left[I_m - \hat{h}_{cv} P_n + \hat{h}_{cv}^2 Q_n + O_p(\hat{h}_{cv}^3) \right] \bar{Y} - f \right\|^2 \\ &\quad + E\left\| \left[b_{cv} P_n - 2b_{cv} \hat{h}_{cv} Q_n - b_{cv}^2 Q_n \right] \bar{Y} \right\|^2 \\ &\quad + 2E\left\{ \left[\bar{\varepsilon} - \hat{h}_{cv} P_n \bar{Y} + \hat{h}_{cv}^2 Q_n \bar{Y} + O_p(\hat{h}_{cv}^3 \bar{Y}) \right]^T \times \right. \\ &\quad \left. \times \left[b_{cv} P_n - 2b_{cv} \hat{h}_{cv} Q_n - b_{cv}^2 Q_n \right] \bar{Y} \right\} \\ &= U_1 + U_2 + U_3.\end{aligned}$$

$$U_1 = E\|\hat{f}_{cv} - f\|^2 + o(n^{-3}).$$

$$\begin{aligned}U_2 &= E(b_{cv}^2 \|P_n \bar{Y}\|^2) + O(n^{-4}) \\ &= O(n^{-4}).\end{aligned}$$

$$\begin{aligned}U_3 &= 2E\left[\bar{\varepsilon}^T b_{cv} P_n \bar{Y} - 2\bar{\varepsilon}^T b_{cv} \hat{h}_{cv} Q_n \bar{Y} - (P_n \bar{Y})^T \hat{h}_{cv} b_{cv} P_n \bar{Y} \right] + O(n^{-4}) \\ &= Z_1 + Z_2 + Z_3 + O(n^{-4}).\end{aligned}$$

$$\begin{aligned}Z_1 &= 2E(\bar{\varepsilon}^T b_{cv} P_n \bar{Y}) \\ &= 2E\left(\frac{s^2 \bar{\varepsilon}^T P_n f}{\hat{\omega}_n} + \frac{s^2 \bar{\varepsilon}^T P_n \bar{\varepsilon}^T}{\hat{\omega}_n} \right) \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} \\ &= \left(\frac{-4\sigma^4 f^T P_n^T P_n M_n P_n f}{\omega_n^2} + \frac{2\sigma^4 \lambda_n}{\omega_n} \right) \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + o(n^{-3}).\end{aligned}$$

$$\begin{aligned}
Z_2 &= -4E \left(\bar{\varepsilon}^T \frac{s^2}{\hat{\omega}_n} \frac{s^2 \lambda_n}{\hat{\omega}_n} Q_n \bar{Y} \right) \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} \\
&= -4E \left(\frac{s^4}{\hat{\omega}_n^2} \bar{\varepsilon}^T Q_n f \right) \sum_{i=1}^m \lambda_n \frac{P_{n,ii}}{n_i^2} + O(n^{-4}) \\
&= O(n^{-4}) .
\end{aligned}$$

$$\begin{aligned}
Z_3 &= -2E \left((P_n \bar{Y})^T \frac{s^2 \lambda_n}{\hat{\omega}_n} \frac{s^2}{\hat{\omega}_n} P_n \bar{Y} \right) \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} \\
&= -2\lambda_n E \left(\frac{s^4}{\hat{\omega}_n} \right) \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} \\
&= -2 \frac{\sigma^4 \lambda_n}{\omega_n} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + O(n^{-4}) .
\end{aligned}$$

Somit ist

$$U_3 = \frac{-4\sigma^4 f^T P_n^T P_n M_n P_n f}{\omega_n^2} \sum_{i=1}^m \frac{P_{n,ii}}{n_i^2} + O(n^{-4})$$

und

$$E \|\hat{f}_{fcv1} - f\|^2 = E \|\hat{f}_{cv} - f\|^2 - \frac{4\sigma^4 f^T P^T P C P f}{n^3 \omega^2} \sum_{i=1}^m P_{ii} c_i^{-2} + o(n^{-3}) .$$

Das Korollar 4.2.2 folgt unmittelbar aus den Sätzen 4.2.2 und 4.2.3.

Beweis von Satz 4.2.4

$$FCV2(h) := \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \left(Y_{ij} - [K(n, h) \tilde{Y}^{+ij}]_i \right)^2$$

mit

$$\begin{aligned}
\tilde{Y}^{+ij} &:= (\bar{Y}_1, \dots, \bar{Y}_{i-1}, \tilde{Y}_{+ij}, Y_{i+1}, \dots, \bar{Y}_m)^T \quad \text{und} \\
\tilde{Y}_{+ij} &:= n_i^{-1} \left(\sum_{\substack{k=1 \\ k \neq j}}^{n_i} Y_{ik} + [K(n, h) \bar{Y}]_i \right) ,
\end{aligned}$$

wobei $[K(n, h) \bar{Y}]_i$ die Schätzung der Regressionsfunktion im Punkt x_i ist.

$$\begin{aligned}
FCV2(h) &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - [K(n, h)\bar{Y}]_i\}^2 \\
&+ 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - [K(n, h)\bar{Y}]_i\} \times \\
&\quad \times \{[K(n, h)\bar{Y}]_i - [K(n, h)\bar{Y}^{+ij}]_i\} \\
&+ \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{[K(n, h)\bar{Y}]_i - [K(n, h)\tilde{Y}^{+ij}]_i\}^2 \\
&= S_1 + S_2 + S_3 .
\end{aligned}$$

Es gilt

$$\begin{aligned}
\bar{Y}_i - \tilde{Y}_{ij} &= n_i^{-1} \sum_{k=1}^{n_i} Y_{ik} - n_i^{-1} \left(\sum_{\substack{k=1 \\ k \neq j}}^{n_i} Y_{ik} + [K(n, h)\bar{Y}]_i \right) \\
&= n_i^{-1} Y_{ij} - n_i^{-1} [K(n, h)\bar{Y}]_i \\
&= n_i^{-1} (Y_{ij} - K(n, h)^{(i)}\bar{Y}) .
\end{aligned}$$

Aus dieser Nebenrechnung folgt die Umrechnung von S_3 und S_2 :

$$\begin{aligned}
S_3 &= \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} K(n, h)_{ii}^2 \{n_i^{-1} (Y_{ij} - K(n, h)^{(i)}\bar{Y})\}^2 \\
&= \sum_{i=1}^m n_i^{-3} (n_i - 1) K(n, h)_{ii}^2 \tilde{s}_i^2 , \quad \text{wobei}
\end{aligned}$$

$$\tilde{s}_i^2 := (n_i - 1)^{-1} \sum_{j=1}^{n_i} [Y_{ij} - K(n, h)^{(i)}\bar{Y}]^2 \quad \text{ist .}$$

$$\begin{aligned}
S_2 &= 2 \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - [K(n, h)\bar{Y}]_i\} \{K(n, h)_{ii} n_i^{-1} [Y_{ij} - K(n, h)^{(i)}\bar{Y}]\} \\
&= 2 \sum_{i=1}^m \frac{1}{n_i^2} K(n, h)_{ii} \sum_{j=1}^{n_i} \{Y_{ij} - [K(n, h)\bar{Y}]_i\}^2 \\
&= 2 \sum_{i=1}^m \frac{1}{n_i^2} K(n, h)_{ii} \sum_{j=1}^{n_i} [Y_{ij} - \bar{Y}_i]^2
\end{aligned}$$

$$\begin{aligned}
& +2 \sum_{i=1}^m \frac{1}{n_i^2} K(n, h)_{ii} \sum_{j=1}^{n_i} \{\bar{Y}_i - [K(n, h)\bar{Y}]_i\}^2 \\
& = 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} K(n, h)_{ii} s_i^2 + 2 \sum_{i=1}^m \frac{1}{n_i^2} K(n, h)_{ii} ([I_m - K(n, h)] \bar{Y})_i^2 . \\
S_1 & = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{Y_{ij} - \bar{Y}_i + \bar{Y}_i - [K(n, h)\bar{Y}]_i\}^2 \\
& = \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} \{\bar{Y}_i - [K(n, h)\bar{Y}]_i\}^2 \\
& = \sum_{i=1}^m \frac{n_i - 1}{n_i} \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^m \frac{1}{n_i} n_i \{[I_m - K(n, h)] \bar{Y}\}_i^2 \\
& = \sum_{i=1}^m \frac{n_i - 1}{n_i} s_i^2 + \|[I_m - K(n, h)] \bar{Y}\|^2
\end{aligned}$$

Das Kriterium, das zu minimieren ist, lautet:

$$\begin{aligned}
FCV2(h) & = \|[I_m - K(n, h)] \bar{Y}\|^2 + 2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} K(n, h)_{ii} s_i^2 \\
& \quad + 2 \sum_{i=1}^m \frac{1}{n_i^2} K(n, h)_{ii} \{[I_m - K(n, h)] \bar{Y}\}_i^2 \quad (10.18) \\
& \quad + \sum_{i=1}^m \frac{n_i - 1}{n_i^3} K(n, h)_{ii}^2 s_i^2 .
\end{aligned}$$

$$\begin{aligned}
FCV2'(0) & = -2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} P_{n,ii} s_i^2 + \sum_{i=1}^m \frac{1}{n_i^3} \sum_{j=1}^{n_i} -2 P_{n,ii} (Y_{ij} - \bar{Y}_i)^2 \\
& = -2 \sum_{i=1}^m \frac{n_i - 1}{n_i^2} P_{n,ii} s_i^2 + \sum_{i=1}^m \frac{-2}{n_i^3} P_{n,ii} (n_i - 1) s_i^2 \\
& = -2 \sum_{i=1}^m \frac{P_{n,ii}}{n_i} s_i^2 + 2 \sum_{i=1}^m \frac{P_{n,ii}}{n_i^3} s_i^2
\end{aligned}$$

$$\begin{aligned}
FCV2''(0) &= 2\hat{\omega}_n + 2 \sum_{i=1}^m \frac{2Q_{n,ii}}{n_i} s_i^2 + 2 \sum_{i=1}^m \frac{P_{n,ii}^2}{n_i^2} s_i^2 + 4 \sum_{i=1}^m \frac{1}{n_i} [P_n \bar{Y}]_i^2 \\
&\quad + 2 \sum_{i=1}^m \frac{1}{n_i^2} [P_n \bar{Y}]_i^2 + O_p \left(\max_{1 \leq i \leq m} \{n_i^{-3}\} \right)
\end{aligned}$$

$$FCV2'''(0) = -12\hat{v}_n + O_p \left(\max_{1 \leq i \leq m} \{n_i^{-1}\} \right)$$

$$FCV2^{(4)}(0) = O_p(1).$$

$\hat{h}_{fcv2} \in \arg \min_{h \in H} FCV2(h)$ hat die Gestalt

$$\hat{h}_{fcv2} = A_0 + A_1 + A_2 + o_p(n^{-2}) \quad \text{mit}$$

$$\begin{aligned}
A_0 &= -\frac{FCV2'(0)}{FCV2''(0)} \\
&= \frac{2 \sum_{i=1}^m P_{n,ii} n_i^{-1} s_i^2 + O_p(n^{-3})}{2\hat{\omega}_n + 2 \sum_{i=1}^m 2Q_{n,ii} n_i^{-1} s_i^2 + 4 \sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 + O_p(n^{-2})} \\
&= \frac{\sum_{i=1}^m P_{n,ii} n_i^{-1} s_i^2}{\hat{\omega}_n} - \frac{\left(\sum_{i=1}^m P_{n,ii} n_i^{-1} s_i^2 \right) \left(\sum_{i=1}^m 2Q_{n,ii} n_i^{-1} s_i^2 \right)}{\hat{\omega}_n^2} \\
&\quad - 2 \frac{\left(\sum_{i=1}^m P_{n,ii} n_i^{-1} s_i^2 \right) \left(\sum_{i=1}^m [P_n \bar{Y}]_i^2 n_i^{-1} \right)}{\hat{\omega}_n^2} + O_p(n^{-3})
\end{aligned}$$

$$\begin{aligned}
A_1 &= -\frac{1}{2} FCV2'''(0) [FCV2'(0)]^2 [FCV2''(0)]^{-3} \\
&= \frac{3\hat{v}_n}{\hat{\omega}_n^3} \left(\sum_{i=1}^m n_i^{-2} (n_i - 1) s_i^2 P_{n,ii} \right)^2 + O_p(n^{-3}) \\
&= \frac{3\hat{v}_n}{\hat{\omega}_n^3} \left(\sum_{i=1}^m n_i^{-1} s_i^2 P_{n,ii} \right)^2 + O_p \left(\max_{1 \leq i \leq m} \{n_i^{-1}\} \right) \\
&= \frac{3\lambda_n^2 s^4 \hat{v}_n}{\hat{\omega}_n^3} + O_p \left(n^{-5/2} + \max_{1 \leq i \leq m} \{n_i^{-3}\} \right)
\end{aligned}$$

$$A_2 = O_p(n^{-3}) .$$

Mit den Gleichungen (10.11), (10.12) und (10.16) erhält man

$$\hat{h}_{fcv2} = \hat{h}_{fcv1} - \frac{2 \left(\sum_{i=1}^m n_i^{-1} P_{n,ii} s_i^2 \right) \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right)}{\hat{\omega}_n^2} + o_p(n^{-2}) . \quad (10.19)$$

$$\begin{aligned} E \|\hat{f}_{fcv2} - f\|^2 &= E \left\| \left[I_m - \hat{h}_{fcv2} P_n + \hat{h}_{fcv2}^2 Q_n + O_p(\hat{h}_{fcv2}^3) \right] \bar{Y} - f \right\|^2 \\ &= E \left\| \left[I_m - \hat{h}_{fcv1} P_n + \hat{h}_{fcv1}^2 Q_n + O_p(\hat{h}_{fcv1}^3) \right] \bar{Y} - f + R_f \right\|^2 \\ &= T_1 + T_2 + T_3 , \end{aligned}$$

wobei

$$\begin{aligned} R_f &:= \frac{2 \left(\sum_{i=1}^m n_i^{-1} s_i^2 P_{n,ii} \right) \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right)}{\hat{\omega}_n^2} P_n \bar{Y} \\ &\quad - 2 \hat{h}_{fcv1} \frac{2 \left(\sum_{i=1}^m n_i^{-1} s_i^2 P_{n,ii} \right) \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right)}{\hat{\omega}_n^2} Q_n \bar{Y} + O_p(n^{-4}) . \end{aligned}$$

$$\begin{aligned} T_1 &= E \left\| \left[I_m - \hat{h}_{fcv1} P_n + \hat{h}_{fcv1}^2 Q_n + O_p(\hat{h}_{fcv1}^3) \right] \bar{Y} - f \right\|^2 \\ &= E \left\| \hat{f}_{fcv1} - f \right\|^2 + o(n^{-3}) . \end{aligned}$$

$$T_2 = E \|R_f\|^2 = O(n^{-4}) .$$

$$\begin{aligned} T_3 &= 2E \left\{ \left[\bar{\varepsilon} - \hat{h}_{fcv1} P_n \bar{Y} + \hat{h}_{fcv1}^2 Q_n \bar{Y} + O_p(\hat{h}_{fcv1}^3) \bar{Y} \right] \cdot R_f \right\} \\ &= 2E \left[2 \left(\sum_{i=1}^m n_i^{-1} P_{n,ii} s_i^2 \right) \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right) \hat{\omega}_n^{-2} \varepsilon^T P_n \bar{Y} \right] \\ &\quad - 2E \left[4 \left(\sum_{i=1}^m n_i^{-1} P_{n,ii} s_i^2 \right)^2 \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right) \hat{\omega}_n^{-3} \varepsilon^T Q_n \bar{Y} \right] \end{aligned}$$

$$\begin{aligned}
& - 2E \left[2 \left(\sum_{i=1}^m n_i^{-1} P_{n,ii} s_i^2 \right)^2 \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right) \hat{\omega}_n^{-3} \|P_n \bar{Y}\|^2 \right] + o(n^{-3}) \\
& = T_{31} + T_{32} + T_{33} + o(n^{-3}),
\end{aligned}$$

denn der erste Summand von \hat{h}_{fcv1} ist $\hat{\omega}_n^{-1} \sum_{i=1}^m n_i^{-1} s_i^2 P_{n,ii}$.

$$\begin{aligned}
T_{31} & = 4E \left[\left(\sum_{i=1}^m n_i^{-1} P_{n,ii} s_i^2 \right) \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right) \hat{\omega}_n^{-2} \bar{\varepsilon}^T P_n f \right] \\
& + 4E \left[\left(\sum_{i=1}^m n_i^{-1} P_{n,ii} s_i^2 \right) \left(\sum_{i=1}^m n_i^{-1} [P_n \bar{Y}]_i^2 \right) \hat{\omega}_n^{-2} \bar{\varepsilon}^T P_n \bar{\varepsilon} \right] \\
& = 4E \left[\hat{\omega}_n^{-2} \bar{\varepsilon}^T P_n f \sum_{i,j=1}^m \frac{P_{n,ii} s_i^2}{n_i n_j} [P_n \bar{Y}]_j^2 \right] \\
& + 4E \left[\hat{\omega}_n^{-2} \bar{\varepsilon}^T P_n \bar{\varepsilon} \sum_{i,j=1}^m \frac{P_{n,ii} s_i^2}{n_i n_j} [P_n \bar{Y}]_j^2 \right].
\end{aligned}$$

Nebenrechnungen: Sei e_j die j -te Spalte der Einheitsmatrix I_m .

$$\begin{aligned}
E \left[\frac{\bar{\varepsilon}^T P_n f s_i^2 [P_n \bar{Y}]_j^2}{\hat{\omega}_n^2} \right] & = - \frac{4\sigma^4 f^T P_n^T P_n M_n P_n f}{\omega_n^3} [P_n f]_j^2 \\
& + \frac{2\sigma^4 e_j^T P_n M_n P_n f}{\omega_n^2} [P_n f]_j + O(n^{-3/2});
\end{aligned}$$

$$E \left[\frac{\bar{\varepsilon}^T P_n \bar{\varepsilon} s_i^2 [P_n \bar{Y}]_j^2}{\hat{\omega}_n^2} \right] = \frac{\sigma^4 \lambda_n}{\omega_n^2} [P_n f]_j^2 + o(n^{-1}).$$

Also ist

$$\begin{aligned}
T_{31} & = -16 \frac{\sigma^4 f^T P_n^T P_n M_n P_n f}{\omega_n^3} \sum_{i,j=1}^m \frac{P_{n,ii}}{n_i n_j} [P_n f]_j^2 \\
& + 8 \frac{\sigma^4}{\omega_n^2} \sum_{i,j=1}^m \frac{P_{n,ii} e_j^T P_n M_n P_n f}{n_i n_j} [P_n f]_j + 4 \frac{\sigma^4 \lambda_n}{\omega_n^2} \sum_{i,j=1}^m \frac{P_{n,ii} [P_n f]_j^2}{n_i n_j} + o_p(n^{-3}) \\
& = -16 \frac{\sigma^4 \lambda_n f^T P_n^T P_n M_n P_n f}{\omega_n^3} \sum_{j=1}^m \frac{[P_n f]_j^2}{n_j}
\end{aligned}$$

$$+ 8 \frac{\sigma^4 \lambda_n}{\omega_n^2} \sum_{j=1}^m \frac{e_j^T P_n M_n P_n f}{n_j} [P_n f]_j + 4 \frac{\sigma^4 \lambda_n^2}{\omega_n^2} \sum_{j=1}^m \frac{[P_n f]_j^2}{n_j} + o_p(n^{-3}).$$

$$T_{32} = O(n^{-7/2}).$$

$$\begin{aligned} T_{33} &= -4E \left(\hat{\omega}_n^{-2} \sum_{i,j,k=1}^m \frac{s_i^2 s_j^2 P_{n,ii} P_{n,jj} [P_n \bar{Y}]_k^2}{n_i n_j n_k} \right) \\ &= -4\omega_n^{-2} \sum_{i,j,k=1}^m \frac{\sigma^4 P_{n,ii} P_{n,jj} [P_n f]_k^2}{n_i n_j n_k} + O_p(n^{-7/2}) \\ &= -4\sigma^4 \lambda_n^2 \omega_n^{-2} \sum_{k=1}^m \frac{[P_n f]_k^2}{n_k} + O_p(n^{-7/2}). \end{aligned}$$

Insgesamt ist:

$$E \|\hat{f}_{fcv2} - f\|^2 = E \|\hat{f}_{fcv1} - f\|^2 + T_{fcv} + o(n^{-3}),$$

wobei

$$\begin{aligned} T_{fcv} &= -16 \frac{\sigma^4 \lambda_n f^T P_n^T P_n M_n P_n f}{\omega_n^3} \sum_{j=1}^m \frac{[P_n f]_j^2}{n_j} \\ &\quad + 8 \frac{\sigma^4 \lambda_n}{\omega_n^2} \sum_{j=1}^m \frac{e_j^T P_n M_n P_n f [P_n f]_j}{n_j}. \end{aligned}$$

Es gilt:

$$\sum_{j=1}^m \frac{[P_n f]_j^2}{n_j} = f^T P_n^T M_n P_n f$$

und

$$\sum_{j=1}^m \frac{e_j^T P_n M_n P_n f [P_n f]_j}{n_j} = (P_n M_n P_n f)^T M_n P_n f.$$

Damit ist

$$\begin{aligned} T_{fcv} &= -8 \frac{\sigma^4 \lambda_n}{\omega_n^2} \left(\frac{2f^T P_n^T P_n M_n P_n f f^T P_n^T M_n P_n f}{\omega_n} - f^T P_n^T M_n P_n M_n P_n f \right) \\ &= -8 \frac{\sigma^4 \lambda}{n^3 \omega^2} \left(\frac{2f^T P^T P C P f f^T P^T C P f}{\omega} - f^T P^T C P C P f \right) + O(n^{-4}). \end{aligned}$$

Für $n_1 = \dots = n_m = n$ ist

$$\begin{aligned} T_{fcv} &= -8 \frac{\sigma^4 \lambda_n}{n^2 \omega_n^2} f^T P_n^T P_n P_n f \\ &= -8 \frac{\sigma^4 \lambda_\kappa}{n^3 \omega^2} + O(n^{-4}) \end{aligned}$$

und $-8 \frac{\sigma^4 \lambda_\kappa}{n^3 \omega^2} \leq 0$ für die Nadaraya-Watson Kernschätzung.

Die Korollare 4.2.3 und 4.2.4 folgen direkt aus den Sätzen 4.2.1 bis 4.2.4.

Beweis von Satz 5.0.6

Für die Schätzung (5.3) gilt:

$$\begin{aligned}
 M(h) &= E\|\hat{f}_h - f\|^2 \\
 &= E\|(I_m - h\mathbb{P})\bar{Y} - f\|^2 \\
 &= E\|(I_m - h\mathbb{P})\bar{\varepsilon} - h\mathbb{P}f\|^2 \\
 &= h^2\|\mathbb{P}f\|^2 + \sigma^2 n^{-1} \text{tr}\{(I_m - h\mathbb{P})(I_m - h\mathbb{P})\} \\
 &= h^2\|\mathbb{P}f\|^2 + \sigma^2 n^{-1}(m - 2hl + h^2l), \quad \text{wobei } l := \text{tr}(\mathbb{P}).
 \end{aligned}$$

$$M'(h) = 2h\|\mathbb{P}f\|^2 + \sigma^2 l n^{-1}(h - 1) \quad \text{und}$$

$$M'(h_{opt}) = 0 \Leftrightarrow h_{opt} = \frac{1}{1 + (\sigma^2 l)^{-1} n \|\mathbb{P}f\|^2};$$

$$M''(h) > 0 \quad \forall h \in [0, 1].$$

Es gilt also $h_{opt} = 1$, falls $\|\mathbb{P}f\| = 0$.

$$M(h_{opt}) = \frac{\sigma^2 m}{n} - \frac{\sigma^4 l^2}{n^2(\|\mathbb{P}f\|^2 + \sigma^2 l/n)} \quad \text{und}$$

$$M(h_{opt}) = \frac{\sigma^2}{n}(m - l), \quad \text{falls } \|\mathbb{P}f\| = 0.$$

Beweis von Satz 5.1.1

Aus der Gleichung (5.6) folgt, daß $\hat{h}_{p-i} < 1$. Außerdem folgt aus (5.6) und (5.8), daß

$$\hat{h}_{p-i} = \hat{h}(1 + \hat{h})^{-1} < \hat{h}.$$

Also gilt für einen beliebigen n -dimensionalen Vektor von Beobachtungen, daß

$$\hat{h}_{p-i} < \min(1, \hat{h}) = \hat{h}_{cp}. \quad (10.20)$$

$$\begin{aligned}
 L(h) &= \|\hat{f}_h - f\|^2 \\
 &= \|(I_m - h\mathbb{P})\bar{Y} - f\|^2 \\
 &= \left\| (I_m - \mathbb{P})(\bar{Y} - f) + \mathbb{P}[(1 - h)\bar{Y} - f] \right\|^2
 \end{aligned}$$

$$\begin{aligned}
&= \left\| (I_m - \mathbb{P})(\bar{Y} - f) \right\|^2 + \left\| (1 - h)\mathbb{P}\bar{Y} - \mathbb{P}f \right\|^2 \\
&= \left\| (I_m - \mathbb{P})(\bar{Y} - f) \right\|^2 + (h - 1)^2 \|\mathbb{P}\bar{Y}\|^2 + \|\mathbb{P}f\|^2 \\
&\quad - 2(1 - h)\bar{Y}^T \mathbb{P}f .
\end{aligned}$$

$$L'(h) = 2(h - 1)\|\mathbb{P}\bar{Y}\|^2 + 2\bar{Y}^T \mathbb{P}f \quad (10.21)$$

$$L''(h) = 2\|\mathbb{P}\bar{Y}\|^2 > 0 . \quad (10.22)$$

$$L'(\tilde{h}_0) = 0 \Leftrightarrow \tilde{h}_0 = 1 - \frac{\bar{Y}^T \mathbb{P}f}{\|\mathbb{P}\bar{Y}\|^2} .$$

Falls $\mathbb{P}f = 0$ ist, gilt $\tilde{h}_0 = 1$. Aus der Gleichung (10.22) folgt, daß die Verlustfunktion $L(h)$ streng konvex in h ist.

Diese Aussage zusammen mit der Gleichung (10.20) führen zu

$$L(\hat{h}_{p-i}) > L(\hat{h}_{cp}) .$$

Beweis der Bemerkung 5.1.3

Die Funktion $M(h) = h^2\|\mathbb{P}f\|^2 + \sigma^2 n^{-1}(m - 2hl + h^2l)$ ist konvex auf $[0,1]$. Drei Fälle sind möglich: entweder $\hat{h} < h_{opt}$ oder $h_{opt} < \hat{h} < 1$ oder $\hat{h} > 1$.

Im ersten Fall ist $M(\hat{h}) < M(\hat{h}_{p-i})$, da $h_{opt} = \arg \min_{h \in [0,1]} M(h)$

und $\hat{h}_{p-i} = \hat{h}(1 + \hat{h})^{-1} < \hat{h}$ ist.

Zweiter Fall: $h_{opt} < \hat{h} < 1$. Seien $\epsilon > 0$,

$$\mathcal{N}_\epsilon := \{f \in \mathbb{R}^m : \|\mathbb{P}f\|^2 < \epsilon\}$$

und $\tau := \epsilon/(\epsilon + n^{-1}\sigma^2l)$. Es gilt

$$\begin{aligned}
1 - \tau &= 1 - \epsilon/(\epsilon + n^{-1}\sigma^2l) \\
&= \frac{n^{-1}\sigma^2l}{\epsilon + n^{-1}\sigma^2l} \\
&< \frac{n^{-1}\sigma^2l}{\|\mathbb{P}f\|^2 + n^{-1}\sigma^2l} , \quad \text{falls } f \in \mathcal{N}_\epsilon \\
&= h_{opt}
\end{aligned}$$

d. h. mit einer solchen Konstruktion von τ erhalten wir die Ungleichung

$$1 - \tau < h_{opt} < \hat{h} < 1, \quad \text{von der die Ungleichung} \\ \hat{h} - h_{opt} < \tau \quad \text{folgt.}$$

Eine Taylorentwicklung im Punkt h_{opt} ergibt:

$$\begin{aligned} M(\hat{h}) &= M(h_{opt}) + M'(h_{opt})(\hat{h} - h_{opt}) + O_p((\hat{h} - h_{opt})^2) \\ &= M(h_{opt}) + O_p(\tau^2) \\ &\leq M(\hat{h}_{p-i}) + O_p(\tau^2) \\ &= M(\hat{h}_{p-i}) + O_p(\epsilon^2/(\epsilon + n^{-1}\sigma^2l)^2) \\ &< M(\hat{h}_{p-i}), \quad \text{falls } \epsilon \text{ gen\u00fcgend klein gew\u00e4hlt wird.} \end{aligned}$$

Dritter Fall: $\hat{h} > 1$. In dieser Situation ist $\hat{h}_{cp} = \min(\hat{h}, 1) = 1$.

Im Fall $\|\mathbb{P}f\| = 0$ ist $h_{opt} = 1$ und es gilt

$$M(\hat{h}_{p-i}) > M(h_{opt}) = M(1) = \sigma^2(m-l)/n.$$

Es existiert eine monotone Nullfolge $\{\epsilon_k\}_{k \geq 1}$ positiver reeler Zahlen und ein $k_0 \geq 1$ so, da\u00df

$$\begin{aligned} M(\hat{h}_{p-i}) &> \sigma^2(m-l)/n + \epsilon_k \quad \forall k \geq k_0 \\ &> \sigma^2(m-l)/n + \|\mathbb{P}f\|^2, \quad \text{falls } \|\mathbb{P}f\|^2 < \epsilon_{k_0}. \\ \Rightarrow \exists K_0 \geq k_0 : \inf_{f,n} M(\hat{h}_{p-i}) &> \sigma^2(m-l)/n + \|\mathbb{P}f\|^2, \quad \text{falls } \|\mathbb{P}f\|^2 < \epsilon_{K_0}. \end{aligned}$$

d. h. $\inf_{f,n} M(\hat{h}_{p-i}) > M(1) = M(\hat{h}_{cp})$ f\u00fcr $f \in \mathcal{N}_{K_0}$.

Beweis der Aussage 6.1.1

$$\begin{aligned} M(c) &= E\|\hat{f}_c - f\|^2 & (10.23) \\ &= (1-c)^2 f^T f + c^2 \sigma^2 t_n. \end{aligned}$$

$$\begin{aligned} M'(c) &= 2(c-1)f^T f + 2c\sigma^2 t_n \\ \Rightarrow c_{opt} &:= \arg \min_{c \in (0,1]} M'(c) = 1 - \sigma^2 t_n (f^T f + \sigma^2 t_n)^{-1}. \end{aligned}$$

$$\begin{aligned}
M(c_{opt}) &= \left(\frac{\sigma^2 t_n}{f^T f + \sigma^2 t_n} \right)^2 f^T f + \left(\frac{f^T f}{f^T f + \sigma^2 t_n} \right)^2 \sigma^2 t_n \\
&= \sigma^2 t_n \left[\frac{f^T f \sigma^2 t_n + (f^T f)^2}{(f^T f + \sigma^2 t_n)^2} \right] \\
&= \sigma^2 t_n \left[\frac{(f^T f + \sigma^2 t_n)^2 - f^T f \sigma^2 t_n - \sigma^4 t_n^2}{(f^T f + \sigma^2 t_n)^2} \right] \\
&= \sigma^2 t_n \left[1 - \frac{\sigma^2 t_n}{(f^T f + \sigma^2 t_n)^2} f^T f - \frac{\sigma^4 t_n^2}{(f^T f + \sigma^2 t_n)^2} \right] \\
&= \sigma^2 t_n \left[1 - \frac{\sigma^2}{n} \frac{\text{tr}(C)}{(f^T f + \frac{\sigma^2}{n} \text{tr}(C))^2} f^T f - \frac{\sigma^4}{n^2} \frac{\text{tr}(C)^2}{(f^T f + \frac{\sigma^2}{n} \text{tr}(C))^2} \right] \\
&< \sigma^2 t_n = M(1) \quad \text{für große } n .
\end{aligned}$$

Beweis der Aussage 6.2.1

Es sei $\hat{M}_{cp}(h)$ durch die Gleichung (6.12) gegeben und $\hat{h}_{cp} = \arg \min_{h \in [0,1]} \hat{M}_{cp}(h)$.

Dann folgt (aus (6.12))

$$\hat{h}_{cp} = \frac{\hat{\sigma}^2 t_n}{\|\bar{Y}\|^2}$$

und

$$\begin{aligned}
\hat{h}_{p-i} &= \frac{\hat{\sigma}^2 t_n}{\|\bar{Y}\|^2} - \frac{\hat{\sigma}^4 t_n^2}{\|\bar{Y}\|^4} + O_p(n^{-3}) \\
&= \hat{h}_{cp} - \frac{\hat{\sigma}^4 t_n^2}{\|\bar{Y}\|^4} + O_p(n^{-3}) \quad \text{für große } n .
\end{aligned} \tag{10.24}$$

$$\begin{aligned}
E\|\hat{f}_{p-i} - f\|^2 &= E\left\| \left[I_m - \hat{h}_{p-i} I_m \right] \bar{Y} - f \right\|^2 \\
&= E\left\| \left[I_m - \left(\hat{h}_{cp} - \frac{\hat{\sigma}^4 t_n^2}{\|\bar{Y}\|^4} + O_p(n^{-3}) \right) I_m \right] \bar{Y} - f \right\|^2 + o(n^{-3}) \\
&= E\|\hat{f}_{cp} - f\|^2 + R_{p-i} + o(n^{-3})
\end{aligned}$$

mit

$$R_{p-i} = 2E \left\{ [\bar{\varepsilon} - h_{cp} \bar{Y}]^T \left[\frac{\hat{\sigma}^4 t_n^2}{\|\bar{Y}\|^4} + O_p(n^{-3}) \right] \bar{Y} \right\} + O(n^{-4})$$

$$= 2E \left(\frac{\hat{\sigma}^4 t_n^2 \bar{\varepsilon}^T \bar{Y}}{\|\bar{Y}\|^4} - \frac{\hat{\sigma}^6 t_n^3}{\|\bar{Y}\|^4} \right) + O(n^{-4}) .$$

Es sei $\sigma_0^2 := E(\hat{\sigma}^2)$. Dann ist $\sigma_0^2 = \alpha\sigma^2 + O(n^{-1})$.

$$\begin{aligned} E \left(\frac{\hat{\sigma}^4 t_n^2 \bar{\varepsilon}^T f}{\|\bar{Y}\|^4} \right) &= t_n^2 E \left\{ \frac{2\sigma_0^2}{(f^T f + \sigma^2 t_n)^2} (\hat{\sigma}^2 - \sigma_0^2) (\bar{\varepsilon}^T f - 0) \right. \\ &\quad - \frac{2\sigma_0^4}{(f^T f + \sigma^2 t_n)^3} (\bar{\varepsilon}^T f - 0) (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) \\ &\quad \left. - \frac{2\sigma_0^2 \cdot 2}{(f^T f + \sigma^2 t_n)^3} (\hat{\sigma}^2 - \sigma_0^2) (\bar{\varepsilon}^T f - 0) (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) \right\} + O(n^{-4}) \\ &= -\frac{4t_n^2 \alpha^2 \sigma^6 f^T M_n f}{(f^T f)^3} + O(n^{-4}) \end{aligned}$$

$$\begin{aligned} E \left(\frac{\hat{\sigma}^4 t_n^2 \bar{\varepsilon}^T \bar{\varepsilon}}{\|\bar{Y}\|^4} \right) &= \frac{t_n^2 \sigma_0^4 \sigma^2 t_n}{(f^T f + \sigma^2 t_n)^2} + O(n^{-4}) \\ &= \frac{t_n^3 \alpha^2 \sigma^6}{(f^T f)^2} + O(n^{-4}) \end{aligned}$$

$$\begin{aligned} E \left(\frac{\hat{\sigma}^6 t_n^3}{\|\bar{Y}\|^4} \right) &= \frac{t_n^3 \sigma_0^6}{(f^T f + \sigma^2 t_n)^2} + O(n^{-4}) \\ &= \frac{t_n^3 \alpha^3 \sigma^6}{(f^T f)^2} + O(n^{-4}) . \end{aligned}$$

Also ist

$$R_{p-i} = -\frac{2\alpha^2 \sigma^6 t_n^2}{(f^T f)^2} \left[\frac{4f^T M_n f}{f^T f} - t_n(1 - \alpha) \right] + O(n^{-4}) .$$

Sei $g(\alpha) := -\frac{2\alpha^2 \sigma^6 t_n^2}{(f^T f)^2} \left[\frac{4f^T M_n f}{f^T f} - t_n(1 - \alpha) \right]$. Es gilt:

$$g(\alpha) \leq 0 \quad \text{genau dann, wenn} \quad \alpha \geq \alpha_0 := 1 - \frac{4f^T M_n f}{f^T f t_n} .$$

Damit $\alpha_0 \in (0, 1]$ muß $\frac{4f^T M_n f}{f^T f t_n} < 1$ sein. Eine hinreichende bedingung dafür ist

$$\frac{\max_{1 \leq i \leq m} \{n_i\}}{\min_{1 \leq i \leq m} \{n_i\}} < m/4 , \quad (10.25)$$

denn ist (10.25) gültig, so ist

$$\begin{aligned}
1 &> \frac{4 \max_{1 \leq i \leq m} \{n_i\}}{m \min_{1 \leq i \leq m} \{n_i\}} \\
&\geq \frac{4 \max_{1 \leq i \leq m} \{n_i\}}{m} \frac{f^T M_n f}{f^T f} \quad \forall f \neq 0 \\
&= \frac{4 f^T M_n f}{f^T f \frac{m}{\max_{1 \leq i \leq m} \{n_i\}}} \\
&\geq \frac{4 f^T M_n f}{f^T f t_n} > 0,
\end{aligned}$$

weil $t_n = \text{tr}(M_n) = \sum_{i=1}^m n_i^{-1} \geq m [\max_{1 \leq i \leq m} \{n_i\}]^{-1}$.

$$g'(\alpha) = 0 \quad \text{für} \quad \alpha = \frac{2}{3} \alpha_0.$$

Für $\alpha > \frac{2}{3} \alpha_0$ ist $g'(\alpha) < 0$. Also ist g auf $I_{\alpha_0} := (\alpha_0, 1]$ fallend und es gilt:
 $\arg \min_{\alpha \in I_{\alpha_0}} g(\alpha) = 1$.

$$g(1) = -\frac{8\sigma^6 t^2 f^T C f}{(f^T f)^3} < 0.$$

Die Behauptungen der Aussage 6.2.2 folgen aus den Sätzen des vorigen Kapitels, wobei die Glättungsmatrix $K(n, h) = I_m - hI_m$ zu setzen ist.

Beweis der Aussage 6.2.3

Sei $E_{p-i}^2 := E \left\| \hat{f}_{p-i} - f \right\|^2$. Für große n ist

$$\begin{aligned}
E_{p-i}^2 &= E \left\| (1 - \hat{h}_{p-i}) \bar{Y} - f \right\|^2 \\
&= E \left\| \bar{\varepsilon} - \frac{s^2 t_n}{\|\bar{Y}\|^2} \bar{Y} + \frac{s^4 t_n^2}{\|\bar{Y}\|^4} \bar{Y} \right\|^2 + O(n^{-4}) \\
&= E(\bar{\varepsilon}^T \bar{\varepsilon}) + E \left(\frac{s^4 t_n^2}{\|\bar{Y}\|^2} \right) - 2E \left(\bar{\varepsilon}^T \frac{s^2 t_n}{\|\bar{Y}\|^2} \bar{Y} \right)
\end{aligned}$$

$$\begin{aligned}
& +2E\left(\bar{\varepsilon}^T \frac{s^4 t_n^2}{\|\bar{Y}\|^4} \bar{Y}\right) - 2E\left(\bar{Y}^T \frac{s^6 t_n^3}{\|\bar{Y}\|^6} \bar{Y}\right) + O(n^{-4}) \\
& = T_1 + \dots + T_5 + O(n^{-4}).
\end{aligned}$$

$$T_1 = E(\bar{\varepsilon}^T \bar{\varepsilon}) = \sigma^2 \text{tr}(M_n).$$

Für die Berechnung der Ausdrücke T_2 bis T_5 wenden wir einen Satz über die Taylorentwicklung einer Funktion mehrerer Veränderlichen an (siehe *Hairer and Wanner (1996), Seite 319*).

$$\begin{aligned}
T_2 & = E\left(\frac{s^4 t_n^2}{\|\bar{Y}\|^2}\right) \\
& = t_n^2 E\left\{\frac{\sigma^4}{f^T f + \sigma^2 t_n} + \frac{(s^2 - \sigma^2)^2}{f^T f + \sigma^2 t_n} + \frac{\sigma^4}{(f^T f + \sigma^2 t_n)^3} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n)^2\right\} \\
& \quad + O(n^{-4}) \\
& = t_n^2 E\left\{\frac{\sigma^4}{f^T f + \sigma^2 t_n} + \frac{\text{Var}(s^2)}{f^T f + \sigma^2 t_n} + \frac{4\sigma^6 f^T M_n f}{(f^T f + \sigma^2 t_n)^3}\right\} + O(n^{-4}) \\
& = \frac{\sigma^4 t_n^2}{f^T f} - \frac{\sigma^6 t_n^3}{(f^T f)^2} + \frac{2\sigma^4 t_n^2}{(n-m) f^T f} + \frac{4t_n^2 \sigma^6 f^T M_n f}{(f^T f)^3} + O(n^{-4}) \\
& = \frac{\sigma^4 t^2}{n^2 f^T f} - \frac{\sigma^6 t^3}{n^3 (f^T f)^2} + \frac{2\sigma^4 t^2}{n^2 (n-m) f^T f} + \frac{4t^2 \sigma^6 f^T C f}{n^3 (f^T f)^3} + O(n^{-4})
\end{aligned}$$

$$\begin{aligned}
T_3 & = -2E\left(\bar{\varepsilon}^T \frac{s^2 t_n}{\|\bar{Y}\|^2} \bar{Y}\right) \\
& = -2t_n E\left(\frac{s^2 \bar{\varepsilon}^T f}{\|\bar{Y}\|^2}\right) - 2t_n E\left(\frac{s^2 \bar{\varepsilon}^T \bar{\varepsilon}}{\|\bar{Y}\|^2}\right) \\
& = T_{31} + T_{32}
\end{aligned}$$

$$\begin{aligned}
T_{31} & = -2t_n E\left\{-\frac{\sigma^2}{(f^T f + \sigma^2 t_n)^2} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) (\bar{\varepsilon}^T f - 0)\right. \\
& \quad \left. + \frac{\sigma^2}{(f^T f + \sigma^2 t_n)^3} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n)^2 (\bar{\varepsilon}^T f - 0)\right\}
\end{aligned}$$

$$\begin{aligned}
& - \frac{\sigma^2}{(f^T f + \sigma^2 t_n)^4} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n)^3 (\bar{\varepsilon}^T f - 0) \Big\} \\
= & \frac{2t_n \sigma^2}{(f^T f + \sigma^2 t_n)^2} 2\sigma^2 f^T M_n f - \frac{2t_n \sigma^2}{(f^T f + \sigma^2 t_n)^3} 4 \operatorname{tr} [(f f^T \otimes I_m) \cdot \Psi_{\bar{\varepsilon}}] \\
& + \frac{2t_n \sigma^2}{(f^T f + \sigma^2 t_n)^3} 4 \sigma^2 t_n f^T \sigma^2 M_n f \\
& + \frac{2t_n \sigma^2}{(f^T f + \sigma^2 t_n)^4} 8 \operatorname{tr} [(f f^T \otimes f f^T) \cdot \Psi_{\bar{\varepsilon}}] + O(n^{-4}) \\
= & \frac{4\sigma^4 t f^T C f}{n^2 (f^T f)^2} - \frac{8\sigma^6 t^2 f^T C f}{n^3 (f^T f)^3} - \frac{8\sigma^6 t}{n^3 (f^T f)^3} \operatorname{tr} [(f f^T \otimes I_m) \cdot \Omega] \\
& + \frac{8\sigma^6 t^2 f^T C f}{n^3 (f^T f)^3} + \frac{16\sigma^6 t}{n^3 (f^T f)^4} \operatorname{tr} [(f f^T \otimes f f^T) \cdot \Omega] + O(n^{-4}),
\end{aligned}$$

denn $\lim_{n \rightarrow \infty} n^2 \sigma^{-4} \Psi_{\bar{\varepsilon}} = \Omega$.

$$\begin{aligned}
T_{32} &= -2t_n E \left(\frac{s^2 \bar{\varepsilon}^T \bar{\varepsilon}}{\|\bar{Y}\|^2} \right) \\
&= -2t_n E \left\{ \frac{\sigma^2 \sigma^2 t_n}{f^T f + \sigma^2 t_n} - \frac{\sigma^2}{(f^T f + \sigma^2 t_n)^2} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) (\bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) \right. \\
&\quad - \frac{\sigma^2 t_n}{(f^T f + \sigma^2 t_n)^2} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) (s^2 - \sigma^2) \\
&\quad + \frac{\sigma^2}{(f^T f + \sigma^2 t_n)^3} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n)^2 (\bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n) \\
&\quad \left. + \frac{\sigma^2 \sigma^2 t_n}{(f^T f + \sigma^2 t_n)^3} (2\bar{\varepsilon}^T f + \bar{\varepsilon}^T \bar{\varepsilon} - \sigma^2 t_n)^2 \right\} + O(n^{-4}) \\
&= -\frac{2\sigma^4 t_n^2}{f^T f + \sigma^2 t_n} + \frac{2\sigma^2 t_n}{(f^T f + \sigma^2 t_n)^2} \operatorname{tr} [(I_m \otimes I_m) \cdot \Psi_{\bar{\varepsilon}}] - \frac{2\sigma^6 t_n^3}{(f^T f + \sigma^2 t_n)^2} \\
&\quad - \frac{2\sigma^2 t_n}{(f^T f + \sigma^2 t_n)^3} 4 \operatorname{tr} [(f f^T \otimes I_m) \cdot \Psi_{\bar{\varepsilon}}] + O(n^{-4}) \\
&= -\frac{2\sigma^4 t^2}{n^2 f^T f} + \frac{2\sigma^6 t}{n^3 (f^T f)^2} \operatorname{tr} [\Omega] - \frac{8\sigma^6 t}{n^3 (f^T f)^3} \operatorname{tr} [(f f^T \otimes I_m) \cdot \Omega] + O(n^{-4})
\end{aligned}$$

$$\begin{aligned}
T_4 &= 2E \left(\bar{\varepsilon}^T \frac{s^4 t_n^2}{\|\bar{Y}\|^4} \bar{Y} \right) \\
&= 2t_n^2 E \left(\frac{s^4 \bar{\varepsilon}^T f}{\|\bar{Y}\|^4} + \frac{s^4 \bar{\varepsilon}^T \bar{\varepsilon}}{\|\bar{Y}\|^4} \right) \\
&= 2t_n^2 \left\{ -\frac{4\sigma^4 f^T \sigma^2 M_n f}{(f^T f + \sigma^2 t_n)^3} + \frac{2\sigma^6 t_n^3}{(f^T f + \sigma^2 t_n)^2} \right\} + O(n^{-4}) \\
&= -\frac{8\sigma^6 f^T C f}{n^3 (f^T f)^3} + \frac{2\sigma^6 t^3}{n^3 (f^T f)^2} + O(n^{-4})
\end{aligned}$$

$$\begin{aligned}
T_5 &= -2E \left(\bar{Y}^T \frac{s^6 t_n^3}{\|\bar{Y}\|^6} \bar{Y} \right) \\
&= -2t_n^3 E \left(\frac{s^6}{\|\bar{Y}\|^4} \right) \\
&= -\frac{2\sigma^6 t^3}{n^3 (f^T f)^2} + O(n^{-4}) .
\end{aligned}$$

Die Summe $T_1 + \dots + T_5$ führt zur Behauptung des Satzes.

Beweis von Satz 7.0.1

Aus der Gleichungen (7.5), (3.10) und (3.11) schreibt man

$$M_2(\tilde{h}_\alpha) = M_2(h_{opt}) + g_n(\alpha) \quad \alpha \in [0, 1] , \quad (10.26)$$

wobei

$$g_n(\alpha) = \frac{\sigma^4 \lambda_n}{\omega_n} \left[\lambda_n (\alpha^2 - 2\alpha + 1) + \frac{4\alpha \kappa_n}{\omega_n} \right] \quad \alpha \in [0, 1] . \quad (10.27)$$

Es gilt

$$g'_n(\alpha) = \frac{\sigma^4 \lambda_n}{\omega_n} \left[2\lambda_n \alpha - 2\lambda_n + \frac{4\kappa_n}{\omega_n} \right] .$$

$$g'_n(\alpha) = 0 \Leftrightarrow \alpha = 1 - 2\kappa_n(\lambda_n\omega_n)^{-1} =: \alpha_{n,0}$$

$$g'_n(\alpha) < 0 \Leftrightarrow \alpha < \alpha_{n,0}$$

$$g'_n(\alpha) > 0 \Leftrightarrow \alpha > \alpha_{n,0} .$$

$$g''_n(\alpha) = \frac{2\sigma^4\lambda_n^2}{\omega_n} > 0 \Rightarrow \alpha_{n,0} = \arg \min_{\alpha \in [0,1]} g_n(\alpha) .$$

Für den Schätzer $\tilde{\alpha}_n$ gilt:

$$E(\tilde{\alpha}_n) = \alpha_{n,0} + O(n^{-1}) \quad \text{und} \quad \text{Var}(\tilde{\alpha}_n) = O(n^{-1}) .$$

Für große n ist also

$$\begin{aligned} E \left\| \tilde{f}_{\tilde{\alpha}_n} - f \right\|^2 &= E \left\| \tilde{\varepsilon} - \tilde{h}_{\tilde{\alpha}_n} P_n \bar{Y} + O_p(\tilde{h}_{\tilde{\alpha}_n}) \right\|^2 \\ &= E(\tilde{\varepsilon}^T \tilde{\varepsilon}) + E \left(\tilde{\alpha}_n^2 \frac{s^4 \lambda_n^2}{\hat{\omega}_n^2} \|P_n \bar{Y}\|^2 \right) \\ &\quad - 2E \left(\tilde{\alpha}_n \frac{s^2 \lambda_n}{\hat{\omega}_n} \tilde{\varepsilon}^T P_n \bar{Y} \right) + o(n^{-2}) \\ &= U_1 + U_2 + U_3 . \end{aligned}$$

$$U_1 = \sigma^2 \text{tr}(M_n)$$

$$\begin{aligned} U_2 &= \lambda_n^2 E \left(\frac{\tilde{\alpha}_n^2 s^4}{\hat{\omega}_n} \right) \\ &= \lambda_n^2 \frac{\sigma^4 (\alpha_{n,0} + O(n^{-1}))^2}{\omega_n + \sigma^2 \delta_n} + O(n^{-3}) \\ &= \frac{\sigma^4 \lambda_n^2 \alpha_{n,0}^2}{\omega_n} + O(n^{-3}) \\ &= \frac{\sigma^4 \lambda^2 \alpha_0^2}{n^2 \omega} + O(n^{-3}) , \quad \text{wobei} \quad \alpha_0 := \lim_{n \rightarrow \infty} \alpha_{n,0} = 1 - 2\kappa(\lambda\omega)^{-1} . \end{aligned}$$

$$\begin{aligned} U_3 &= -2\lambda_n \left\{ E \left(\frac{\tilde{\alpha}_n s^2}{\hat{\omega}_n} \tilde{\varepsilon}^T P_n f \right) + E \left(\frac{\tilde{\alpha}_n s^2}{\hat{\omega}_n} \tilde{\varepsilon}^T P_n \tilde{\varepsilon} \right) \right\} \\ &= -2\lambda_n (U_{31} + U_{32}) . \end{aligned}$$

$$\begin{aligned}
U_{31} &= \frac{-(\alpha_{n,0} + O(n^{-1}))\sigma^2}{(\omega_n + \sigma^2\delta_n)^2} E \{ (2\bar{\varepsilon}^T P_n^T P_n f + \bar{\varepsilon}^T P_n^T P_n \bar{\varepsilon} - \sigma^2\delta_n)(\bar{\varepsilon}^T P_n f - 0) \} \\
&\quad + o(n^{-1}) \\
&= \frac{-2\sigma^4\kappa_n \alpha_{n,0}}{\omega_n^2} + o(n^{-1}) \\
&= -2\alpha_0 \frac{\sigma^4\kappa}{n\omega^2} + o(n^{-1}) \\
U_{32} &= \frac{\alpha_{n,0}\sigma^4\lambda_n}{\omega_n + \sigma^2\delta_n} + o(n^{-1}) \\
&= \frac{\alpha_0\sigma^4\lambda}{n\omega} + o(n^{-1}) .
\end{aligned}$$

Insgesamt ist:

$$M_2(\tilde{h}_{\tilde{\alpha}_n}) = \sigma^2 \text{tr}(M_n) + \alpha_0^2 \frac{\sigma^4\lambda^2}{n^2\omega} + 4\alpha_0 \frac{\sigma^4\lambda\kappa}{n^2\omega^2} - 2\alpha_0 \frac{\sigma^4\lambda^2}{n\omega} .$$

Da die Risiken 2. Ordnung von \hat{f}_{p-i} aus der Gleichung (4.6) den Risiken $M_2(\tilde{h}_\alpha)$ mit $\alpha = 1$ entsprechen und da $\arg \min_{\alpha \in (0,1]} g_n(\alpha) = \alpha_{n,0}$, ist

$$M_2(\tilde{h}_{\tilde{\alpha}_n}) < M_2(\hat{h}_{p-i}) .$$

Beweis von Satz 9.0.1

Betrachten wir die Darstellung (9.6). Sei h_{opt} die Lösung von $M'(h) = 0$. Dann gilt:

$$4h_{opt}^3 B_1 - 6h_{opt}^5 B_2 - n^{-1}h_{opt}^{-2}V_1 - n^{-1}V_2 = 0 .$$

Die Multiplikation dieser Gleichung mit h_{opt}^2 ergibt:

$$4h_{opt}^5 B_1 - 6h_{opt}^7 B_2 - n^{-1}V_1 - n^{-1}h_{opt}^2 V_2 = 0 \tag{10.28}$$

und es folgt daraus

$$h_{opt}^5 = \frac{n^{-1}V_1}{4B_1} + \frac{n^{-1}V_2}{4B_1} h_{opt}^2 + \frac{6B_2}{4B_1} h_{opt}^7 . \tag{10.29}$$

Wir erhalten die implizite Gleichung

$$h_{opt} = \left(\frac{n^{-1}V_1}{4B_1} + \frac{n^{-1}V_2}{4B_1} h_{opt}^2 + \frac{6B_2}{4B_1} h_{opt}^7 \right)^{1/5} \tag{10.30}$$

$$= (n^{-1}A_1 + n^{-1}A_2 h_{opt}^2 + A_3 h_{opt}^7)^{1/5}, \quad (10.31)$$

wobei $A_1 := \frac{V_1}{4B_1}$ $A_2 := \frac{V_2}{4B_1}$ $A_3 := \frac{6B_2}{4B_1}$. Der Gleichung (10.31) ergibt

$$\begin{aligned} h_{opt} &= (n^{-1}A_1)^{1/5} \left(1 + \frac{A_2}{A_1} h_{opt}^2 + \frac{A_3}{A_1} n h_{opt}^7 \right)^{1/5} \\ &= (n^{-1}A_1)^{1/5} [1 + x(h_{opt})]^{1/5}, \end{aligned} \quad (10.32)$$

wobei $x(h_{opt}) := \frac{A_2}{A_1} h_{opt}^2 + \frac{A_3}{A_1} n h_{opt}^7$.

Falls $nh^7 = o(1)$ dann ist $x(h_{opt}) = O(n^{-2/5})$. Eine Taylorentwicklung im Punkt $x(h) = 0$ ergibt:

$$[1 + x(h)]^{1/5} = 1 + \frac{1}{5}x(h) - \frac{2}{25}x(h)^2 + o(x(h)^2); \quad (10.33)$$

$$\begin{aligned} \Rightarrow \\ h_{opt} &= (n^{-1}A_1)^{1/5} \left[1 + \frac{1}{5}x(h_{opt}) - \frac{2}{25}x(h_{opt})^2 + o(n^{-4/5}) \right] \\ &= (n^{-1}A_1)^{1/5} \left\{ 1 + \frac{1}{5} \left[\frac{A_2}{A_1} h_{opt}^2 + \frac{A_3}{A_1} n h_{opt}^7 \right] \right. \\ &\quad \left. - \frac{2}{25} \left[\frac{A_2}{A_1} h_{opt}^2 + \frac{A_3}{A_1} n h_{opt}^7 \right]^2 + o(n^{-4/5}) \right\} \\ &= (n^{-1}A_1)^{1/5} \left\{ 1 + \frac{1}{5} \left[\frac{A_2}{A_1} h_{opt}^2 + \frac{A_3}{A_1} n h_{opt}^7 \right] + O(n^{-4/5}) \right\}, \end{aligned}$$

denn

$$\begin{aligned} \frac{2}{25} \left[\frac{A_2}{A_1} h_{opt}^2 + \frac{A_3}{A_1} n h_{opt}^7 \right]^2 &= O(h_{opt}^4 + n h_{opt}^9 + n^2 h_{opt}^{14}) \\ &= O(n^{-4/5}). \end{aligned}$$

Aus (10.33) ist

$$\begin{aligned} h_{opt}^2 &= (n^{-1}A_1)^{2/5} \left(1 + \frac{2}{5}x(h_{opt}) - \frac{3}{25}x(h_{opt})^2 + o(x(h_{opt})^2) \right) \\ &= (n^{-1}A_1)^{2/5} (1 + O(n^{-2/5})) \\ h_{opt}^7 &= (n^{-1}A_1)^{7/5} (1 + O(n^{-2/5})). \end{aligned}$$

Also ist

$$\begin{aligned}
h_{opt} &= (n^{-1}A_1)^{1/5} \left\{ 1 + \frac{1}{5} \frac{A_2}{A_1} h_{opt}^2 + \frac{1}{5} \frac{A_3}{A_1} n h_{opt}^7 + O(n^{-4/5}) \right\} \\
&= (n^{-1}A_1)^{1/5} \left\{ 1 + \frac{A_2}{5A_1} (n^{-1}A_1)^{2/5} (1 + O(n^{-2/5})) \right. \\
&\quad \left. + \frac{A_3}{5A_1} n (n^{-1}A_1)^{7/5} (1 + O(n^{-2/5})) + O(n^{-4/5}) \right\} \\
&= (n^{-1}A_1)^{1/5} + (n^{-1}A_1)^{3/5} \frac{A_2}{5A_1} + \frac{A_3}{5A_1} n (n^{-1}A_1)^{8/5} + O(n^{-1}) \\
&= (n^{-1}A_1)^{1/5} + (n^{-1}A_1)^{3/5} \left(\frac{A_2}{5A_1} + \frac{A_3}{5A_1} \right) + O(n^{-1}) .
\end{aligned}$$

Beweis von Satz 9.0.2

Es seien $\mathbf{m} := (m(X_1), \dots, m(X_n))^T$, $Y := (Y_1, \dots, Y_n)^T$,

$W_h(x) := \text{diag}(K(X_i - x)/h)_{1 \leq i \leq n}$ und $\mathbf{1}_n := (1, \dots, 1)^T$ ein n -dimensionaler Vektor. Es ist

$$\begin{aligned}
\hat{m}_h^{NW}(x) &= (\mathbf{1}_n^T W_h(x) \mathbf{1}_n)^{-1} \mathbf{1}_n^T W_h(x) Y \\
E[\hat{m}_h^{NW}(x) | X_1, \dots, X_n] &= (\mathbf{1}_n^T W_h(x) \mathbf{1}_n)^{-1} \mathbf{1}_n^T W_h(x) \mathbf{m} \\
\text{Bias}[\hat{m}_h^{NW}(x) | X_1, \dots, X_n] &= (\mathbf{1}_n^T W_h(x) \mathbf{1}_n)^{-1} \mathbf{1}_n^T W_h(x) (\mathbf{m} - \mathbf{1}_n m(x)) \\
&= (\mathbf{1}_n^T W_h(x) \mathbf{1}_n)^{-1} \mathbf{1}_n^T W_h(x) R_{n,1}(x) , \quad (10.34)
\end{aligned}$$

$$\text{wobei} \quad R_{n,1}(x) := (m(X_1) - m(x), \dots, m(X_n) - m(x))^T .$$

Die Taylorentwicklung von m im Punkt x ergibt:

$$m(X_j) - m(x) = m'(x)(X_j - x) + \frac{m''(x)}{2} (X_j - x)^2 + o((X_j - x)^2) \quad j = 1, \dots, n. \quad (10.35)$$

$(\mathbf{1}_n^T W_h(x) \mathbf{1}_n) = nh \hat{f}_h(x)$, wobei $\hat{f}_h(x)$ die Kernschätzung der Densität ist. Aus *Silverman (1986) Abschnitt 3.3.1* gilt

$$E(\hat{f}_h(x)) = f(x) + \frac{h^2}{2} \mu_2 f''(x) + \frac{h^4}{4!} \mu_4 f^{(4)}(x) + o(h^4) . \quad (10.36)$$

Nach dem Gesetz der großen Zahlen ist

$$(nh)^{1/2} \left[\hat{f}_h(x) - E(\hat{f}_h(x)) \right] = O_p(1) . \quad (10.37)$$

Eine weitere Taylorentwicklung ist

$$\begin{aligned} \frac{1}{\hat{f}_h(x)} &= \frac{1}{E(\hat{f}_h(x))} - \frac{1}{[E(\hat{f}_h(x))]^2} \left[\hat{f}_h(x) - E(\hat{f}_h(x)) \right] \\ &+ O_p \left(\left[\tilde{f}_h(x) - E(\hat{f}_h(x)) \right]^2 \right), \end{aligned}$$

$$\text{wobei} \quad |\tilde{f}_h(x) - E(\hat{f}_h(x))| < |\hat{f}_h(x) - E(\hat{f}_h(x))|.$$

Mit (10.37) folgt

$$\begin{aligned} \frac{1}{\hat{f}_h(x)} &= \frac{1}{E(\hat{f}_h(x))} + O_p((nh)^{-1/2}) \\ &= \frac{1}{f(x) + \frac{h^2}{2} \mu_2 f''(x) + \frac{h^4}{4!} \mu_4 f^{(4)}(x) + o(h^4)} + O_p((nh)^{-1/2}) \\ &= \frac{1}{f(x)} \left\{ 1 - \frac{h^2}{2} \mu_2 \frac{f''(x)}{f(x)} - \frac{h^4}{4!} \mu_4 \frac{f^{(4)}(x)}{f(x)} \right. \\ &\quad \left. + \frac{h^4}{4} \mu_2^2 \frac{f''(x)^2}{f(x)^2} + o(h^4) \right\} + O_p((nh)^{-1/2}). \end{aligned} \tag{10.38}$$

$$\Rightarrow (\mathbb{1}_n^T W_h(x) \mathbb{1}_n)^{-1} = (nh)^{-1} \frac{1}{\hat{f}_h(x)}. \tag{10.39}$$

Aus (10.35) gilt

$$\mathbb{1}_n^T W_h(x) R_{n,1}(x) = \left\{ m'(x) S_{n,1}(x) + \frac{m''(x)}{2} S_{n,2}(x) + o(S_{n,2}(x)) \right\}$$

$$\text{mit} \quad S_{n,j} := \sum_{i=1}^n (X_i - x)^j K((X_i - x)/h).$$

In *Fan und Huang (1997)* wird es gezeigt, daß :

$$S_{n,1} = nh^2 \left[hf'(x)\mu_2 + \frac{1}{3!} f^{(3)}(x)\mu_4 h^3 + o(h^3) + O_p((nh)^{-1/2}) \right]$$

$$S_{n,2} = nh^3 \left[f(x)\mu_2 + \frac{1}{2} f''(x)\mu_4 h^2 + o(h^2) + O_p((nh)^{-1/2}) \right]$$

$$S_{n,3} = nh^4 (hf'(x)\mu_4 + o(h)).$$

$$o(S_{n,3}) = o(nh^5).$$

Mit der Bezeichnung $b_h^{NW}(x) := \text{Bias}[m_h^{NW}(x)|X_1, \dots, X_n]$ ist

$$\begin{aligned} b_h^{NW}(x) &= [\hat{f}(x)]^{-1} \left\{ m'(x) \left[h^2 f'(x) \mu_2 + \frac{1}{3!} h^4 f^{(3)} \mu_4 + o(h^4) + O_p(n^{-1/2} h^{1/2}) \right] \right. \\ &\quad + \frac{m''(x)}{2} \left[h^2 f(x) \mu_2 + \frac{1}{2} h^4 f''(x) \mu_4 + o(h^4) + O_p(n^{-1/2} h^{3/2}) \right] \\ &\quad \left. + \frac{m'''(x)}{3!} \left[h^4 f'(x) \mu_4 + o(h^4) + O_p(n^{-1/2} h^{5/2}) \right] \right\} \end{aligned}$$

Mit der Gleichung (10.38) folgt:

$$\begin{aligned} b_h^{NW}(x) &= h^2 \mu_2 \left[\frac{m'(x) f'(x)}{f(x)} + \frac{m''(x)}{2} \right] \\ &\quad - h^4 \left\{ \frac{\mu_2 f''(x)}{2 f(x)} \left[\frac{m'(x) f'(x)}{f(x)} + \frac{m''(x)}{2} \right] \right. \\ &\quad \left. - \frac{\mu_4}{f(x)} \left[\frac{m'(x) f'''(x)}{3!} + \frac{m''(x) f''(x)}{4} + \frac{m'''(x) f'(x)}{3!} \right] \right\} \\ &\quad + o(h^4) + O_p(n^{-1/2} h^{1/2}) . \end{aligned}$$

$$\begin{aligned} v_h^{NW}(x) &:= \text{Var}(m_h^{NW}(x)|X_1, \dots, X_n) \\ &= (\mathbb{1}_n^T W_h(x) \mathbb{1}_n)^{-1} \mathbb{1}_n^T W_h(x) \Sigma_Y W_h(x) \mathbb{1}_n (\mathbb{1}_n^T W_h(x) \mathbb{1}_n)^{-1} \\ &= (\mathbb{1}_n^T W_h(x) \mathbb{1}_n)^{-2} R_{n,2}(x) , \end{aligned}$$

wobei

$$\Sigma_Y := \text{diag}(\sigma^2(X_i))_{1 \leq i \leq n} \quad \text{und} \quad R_{n,2}(x) := \sum_{i=1}^n K^2((X_i - x)/h) \sigma^2(X_i) .$$

Es sei $g(x + hz) := \sigma^2(x + hz) f(x + hz)$. Eine Standardberechnung ergibt

$$\begin{aligned} E(R_{n,2}(x)) &= n E \left[(K^2((X_1 - x)/h)) \sigma^2(X_1) \right] \\ &= n \int K^2 \left(\frac{y - x}{h} \right) \sigma^2(y) f(y) dy \\ &= nh \int K^2(z) g(x + hz) dz \end{aligned}$$

$$\begin{aligned}
&= nh \int K^2(z) \left[g(x) + g'(x)hz + \frac{g''(x)}{2}(hz)^2 + o((hz)^2) \right] dz \\
&= nh \left[\nu_0 g(x) + h \nu_1 g'(x) + h^2 \nu_2 \frac{g''(x)}{2} + o(h^2) \right]. \quad (10.40)
\end{aligned}$$

Analog zeigt man, daß $Var(R_{n,2}(x)) = O(nh)$.

Nach dem Gesetz der großen Zahlen gilt deshalb:

$$R_{n,2}(x) = nh \left[\frac{E(R_{n,2}(x))}{nh} + O_p((nh)^{-1/2}) \right]. \quad (10.41)$$

Die Gleichungen (10.38), (10.39), (10.40) und (10.41) ergeben

$$\begin{aligned}
v_h^{NW}(x) &= (nh)^{-1} \frac{\nu_0 \sigma^2(x)}{f(x)} + n^{-1} \frac{g'(x) \nu_1}{f(x)^2} \\
&\quad - n^{-1} h \left[\nu_0 \mu_2 f''(x) \sigma^2(x) - \frac{\nu_2 g''(x)}{2} \right] \\
&\quad + o(n^{-1}h) + O_p((nh)^{-3/2}).
\end{aligned}$$

Literaturverzeichnis

- [1] O.E. Barndorff-Nielsen. *Parametric Statistical Inference and Likelihood*. Springer, New York, Berlin, Heidelberg, 1988.
- [2] O. Bunke. Minimax linear and quadratic estimators in semiparametric regression models. *Statistics*, 33:1–35, 1999.
- [3] O. Bunke, Droge B., and J. Polzehl. Model selection, transformations and variance estimation in nonlinear regression. *Statistics*, 33:197–240, 1999.
- [4] O. Bunke and B. Droge. Bootstrap and cross-validation estimates of the prediction error for linear regression models. *Annals of Statistics*, 12:1400–1424, 1984.
- [5] H. Cramér. *Mathematical Methods of Statistics*. University Press, Princeton, 1946.
- [6] B. Droge. Some comments on cross-validation. In W. Härdle and M.G. Schimek, editors, *Statistical Theory and Computational Aspects of smoothing*, pages 178–199. Physica, Heidelberg, 1996.
- [7] B. Droge. Asymptotic optimality of full cross-validation for selecting linear regression models. *Statistics and Probability Letters*, 44:351–357, 1999.
- [8] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, London., 1993.
- [9] J. Fan and L.S. Huang. Rates of convergence for the pre-asymptotic substitution bandwidth selector. (to appear in statistics and probability letters). Technical report, University of North Carolina, Chaptel Hill, 1997.
- [10] P.J. Green, C. Jennison, and A. Seheult. Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society Ser. B*, 47:299–315, 1985.

- [11] P.J. Green and B.W. Silverman. *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, 1994.
- [12] E. Hairer and G. Wanner. *Analysis by its History*. Springer Verlag, 1996.
- [13] P. Hall, S.J. Sheather, M.C. Jones, and J.S. Marron. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78:263–269, 1991.
- [14] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, New York, 1990.
- [15] W. Härdle, P. Hall, and J.S. Marron. How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, 83:86–101, 1988.
- [16] J.S. Huang. Third-order expansion of mean squared error of medians. *Statistics and Probability Letters*, 42:185–192, 1999.
- [17] K.M.S. Humak. *Statistische Methoden der Modellbildung. (Statistische Inferenz für Kovarianzparameter)*, volume 3. Akademie Verlag, 1984.
- [18] C.M. Hurvich, Simonoff J.S., and C.L. Tsai. Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society Ser. B*, 60:271–293, 1998.
- [19] P.E. Ilouga. Asymptotik höherer ordnung für lineare glättungen in nicht-parametrischen regressionsmodellen. Diplomarbeit, Fachbereich Mathematik, Humboldt Universität zu Berlin, 1994.
- [20] R.I. Jennrich. Asymptotic properties of nonlinear least squares estimators. *Annals of mathematical Statistics* 40, pages 633–643, 1969.
- [21] M.C. Jones, J.S. Marron, and S.J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91:401–407, 1996.
- [22] A. Kielbasinski and H. Schwetlick. *Numerische lineare Algebra*. Deutscher Verlag der Wissenschaften Berlin, 1988.
- [23] D.N. Lawley. A general method for approximating to the distribution of the likelihood ratio criteria. *Biometrika*, 43:295–303, 1956.
- [24] C.L. Mallows. Some comments on c_p . *Technometrics*, 15:661–675, 1973.

- [25] M.H. Neumann. Second order asymptotic risks of smoothed linear estimators in nonparametric regression models. *Statistics*, 23:217–236, 1992.
- [26] J.K. Patel and C.B. Read. *Handbook of the Normal Distribution (Second Edition, revised and expanded)*. Marcel Dekker, 1996.
- [27] D. Ruppert, S.J. Sheather, and M.P. Wand. An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90:1257–1270, 1995.
- [28] B.W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986.
- [29] R. Thrum. Some properties of the noncentral F distribution. Unpublished manuscript, 1999.
- [30] M.P. Wand and M.C. Jones. *Kernel Smoothing*. Chapman & Hall, London, 1995.

BEZEICHNUNGEN

m Anzahl der Versuchspunkte.

n_i Anzahl der Beobachtungen im Versuchspunkt x_i .

$n := \sum_{i=1}^m n_i$ Gesamtzahl der Beobachtungen.

$f_i := f(x_i)$ Wert der unbekanntes Regressionsfunktion im Punkt x_i .

$f := (f_1, \dots, f_m)^T$ Mittelwertvektor.

$\bar{Y}_i := n_i^{-1} \sum_{j=1}^{n_i} Y_{ij}$ Mittelwert der Beob. im Punkt x_i .

$\bar{Y} := (\bar{Y}_1, \dots, \bar{Y}_m)^T$ Vektor der Mittelwerte der Beobachtungen.

$\bar{\varepsilon} := \bar{Y} - f$ Vektor der Mittelwerte der Fehler.

$\hat{f}_h := K(n, h)\bar{Y}$ Schätzung des Mittelwertvektors durch Glättung mit der Bandbreite h .

$s_i^2 := (n_i - 1)^{-1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$

$s^2 := (n - m)^{-1} \sum_{i=1}^m \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$

$\hat{\sigma}^2 := \alpha s^2 + \gamma \bar{Y}^T \Gamma_n \bar{Y}$

$\|x\| := \sqrt{x^T x}$ Euklidische Norm des Vektors x .

$\lambda_{\max}(A)$ Maximaler Eigenwert der Matrix A .

$\|A\| := \sqrt{\lambda_{\max}(A^T A)}$ Spektralnorm der Matrix A .

Es seien $\{A_n\}_{n \geq 1}$ eine Folge reellwertiger Matrizen und $\{b_n\}_{n \geq 1}$ eine Folge reeller positiver Zahlen.

$$A_n = O(b_n) \Leftrightarrow_{Def} \exists C < \infty : \frac{\|A_n\|}{b_n} \leq C \quad \text{für } n \rightarrow \infty.$$

$$A_n = o(b_n) \Leftrightarrow_{Def} \lim_{n \rightarrow \infty} \frac{\|A_n\|}{b_n} = 0.$$

Es seien $\{X_n\}_{n \geq 1}$ eine Folge zufälliger reellwertiger Matrizen und $\{y_n\}_{n \geq 1}$ eine Folge reellwertiger positiver Zufallsgrößen.

$$X_n = O_p(y_n) \Leftrightarrow_{Def} \forall \varepsilon > 0, \exists M, \exists N : P\left(\frac{\|X_n\|}{y_n} > M\right) \leq \varepsilon \quad \forall n \geq N.$$

$$X_n = o_p(y_n) \Leftrightarrow_{Def} \forall \varepsilon > 0, \lim_{n \rightarrow \infty} P\left(\frac{\|X_n\|}{y_n} > \varepsilon\right) = 0.$$

Matrizen und Vektoren

I_m $m \times m$ – dimensionale Einheitsmatrix .

$\mathbb{1}_m$ m – dimensionaler Vektor mit $(\mathbb{1}_m)_i = 1 \quad \forall i \in \{1, \dots, m\}$.

e_j j – te Spalte der Einheitsmatrix .

$K(n, h)$ Glättungsmatrix mit der Bandbreite h .

$M_n := \text{diag}(n_1^{-1}, \dots, n_m^{-1})$.

$c_i := \lim_{n \rightarrow \infty} \frac{n_i}{n} \quad i = 1, \dots, m ; \quad C := \lim_{n \rightarrow \infty} n M_n = \text{diag}(c_1^{-1}, \dots, c_m^{-1})$.

$P_n := (P_{n,ij})_{1 \leq i, j \leq m} ; \quad P := (P_{ij})_{1 \leq i, j \leq m} = \lim_{n \rightarrow \infty} P_n$.

$Q_n := (Q_{n,ij})_{1 \leq i, j \leq m} ; \quad Q := (Q_{ij})_{1 \leq i, j \leq m} = \lim_{n \rightarrow \infty} Q_n$.

$R_n := (R_{n,ij})_{1 \leq i, j \leq m} ; \quad R := (R_{ij})_{1 \leq i, j \leq m} = \lim_{n \rightarrow \infty} R_n$.

$A \otimes B$ Tensorprodukt der Matrix A mit der Matrix B .

$\Psi_Z := E(Z \otimes Z Z^T)$ Matrix der 3. Momente des Zufallsvektors Z .

$\Phi_Z := E(Z Z^T \otimes Z Z^T)$ Matrix der 4. Momente des Zufallsvektors Z .

Spuren

$$\lambda_n := \text{tr}(P_n M_n) \quad ; \quad \lambda := \lim_{n \rightarrow \infty} n \lambda_n = \text{tr}(PC) .$$

$$\mu_n := \text{tr}(Q_n M_n) \quad ; \quad \mu := \lim_{n \rightarrow \infty} n \mu_n = \text{tr}(QC) .$$

$$\delta_n := \text{tr}(P_n^T P_n M_n) \quad ; \quad \delta := \lim_{n \rightarrow \infty} n \delta_n = \text{tr}(P^T PC) .$$

$$\rho_n := \delta_n + 2\mu_n \quad ; \quad \rho := \lim_{n \rightarrow \infty} n \rho_n = \delta + 2\mu .$$

$$t_n := \text{tr}(M_n) \quad ; \quad t := \text{tr}(C) = \lim_{n \rightarrow \infty} n t_n .$$

Quadratische Formen

$$\omega_n := f^T P_n^T P_n f \quad ; \quad \omega := f^T P^T P f \quad ; \quad \hat{\omega}_n := \bar{Y} P_n^T P_n \bar{Y} .$$

$$\nu_n := f^T P_n^T Q_n f \quad ; \quad \nu := f^T P^T Q f \quad ; \quad \hat{\nu}_n := \bar{Y} P_n^T Q_n \bar{Y} .$$

$$\kappa_n := f^T P_n^T P_n M_n P_n f \quad ; \quad \kappa := f^T P^T P C P f = \lim_{n \rightarrow \infty} n \kappa_n .$$

Weitere Bezeichnungen

χ_k^2	Zentrale Chi-quadrat Verteilung mit k Freiheitsgraden .
$\chi_{k,\lambda}^2$	Nichtzentrale Chi-quadrat Verteilung mit k Freiheitsgraden und Nichtzentralitätsparameter λ .
F_{k_1,k_2}	Zentrale F -Verteilung mit k_1 und k_2 Freiheitsgraden .
$F_{k_1,k_2,(\lambda_1,\lambda_2)}$	Doppelnichtzentrale F -Verteilung mit k_1 und k_2 Freiheitsgraden und Nichtzentralitätsparametern λ_1 und λ_2 .
$F_{k_1,k_2,(\lambda_1,\lambda_2)}(\cdot)$	Verteilungsfunktion einer doppelnichtzentralen F -Verteilten Zufallsgröße mit k_1 und k_2 Freiheitsgraden und Nichtzentralitätsparametern λ_1 und λ_2 .

$L(h)$	Verlust der Schätzung \hat{f}_h (Seite 29) .
$M(h)$	Mittlerer quadratischer Fehler der Schätzung \hat{f}_h .
SEP	Quadratischer Vorhersagefehler .
$M_2(h)$	Risiken 2. Ordnung der Schätzung \hat{f}_h (Seite 13) .
$M_3(h)$	Risiken 3. Ordnung der Schätzung \hat{f}_h (Seite 13) .

$$\mu_i := \int z^i K(z) dz \quad ; \quad \nu_i := \int z^i K^2(z) dz \quad , \quad i = 0, 1, 2, \dots .$$

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad \text{Normalkern}$$

$$K(u) = \frac{3}{4} (1 - u^2) \mathbb{1}_{\{|u| \leq 1\}} \quad \text{Epanechnikovkern}$$

Selbständigkeitserklärung

Hiermit erkläre ich, daß ich die vorliegende Arbeit selbständig, ohne unerlaubte Hilfe und nur unter Verwendung der angegebenen Literatur angefertigt habe.

Berlin, den 22.08.2000

Ilouga, Pierre Emmanuel

Lebenslauf

Name : ILOUGA

Vornamen : Pierre Emmanuel

Geboren am : 22 Novembre 1968 in Yaoundé (Kamerun)

Ausbildung

- 1979-1986: Besuch des Gymnasiums "*lycée Leclerc*" in Yaoundé.
- 1986: Abitur ("*Baccalauréat serie D*") (*Mention Bien*) in Yaoundé .
- 1986-1987: Deutsche Sprachausbildung in Glauchau (bei Zwickau).
- 1987-1994: Mathematikstudium an der Humboldt Universität zu Berlin. Diplom-Abschluß 1994 mit der Note "sehr gut".

Diplomarbeit zum Thema: *Asymptotik höherer Ordnung für lineare Glättungen in nichtparametrischen Regressionsmodellen.*

Betreuer: Prof. Dr. habil. O. Bunke.

- Ab Februar 1995: Arbeit an der Dissertation am Institut für Mathematik der Humboldt Universität zu Berlin.
Betreuer: Prof. Dr. habil. O. Bunke.

Praktika

- 1988-1989: Grundpraktikum in numerischer Mathematik am Institut für angewandte Mathematik (1 Semester).
- 1990: Betriebspraktikum am Zentralinstitut für Arbeitsmedizin in Berlin. Betreuer: Dr. D. Breuer. (3 Monate).
- 1991-1992: Fachpraktikum mathematische Statistik (2 Semester).

- 1995: Mitarbeiter am Sonderforschungsbereich 373 (Teilprojekt B1) der Humboldt Universität zu Berlin (11 Monate).

Weitere Aktivitäten

- Sprecher der Sportkommission der VKS (*Vereinigung Kamerunische Studenten*) in Berlin.