

Erstellung von Mietspiegeln:
Ein semiparametrisches Regressionsmodell
zur Schätzung der ortsüblichen
Vergleichsmiete

Dissertation

zur Erlangung der Würde eines
Doktors der Wirtschafts- und Sozialwissenschaften,
vorgelegt der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität Freiburg in der Schweiz
von

María Josefa Redondo González
aus Spanien

Genehmigt von der Wirtschafts- und Sozialwissenschaftlichen Fakultät
am 12. November 2002 auf Antrag von
Herrn Prof. Dr. Hans Wolfgang Brachinger (erster Referent) und
Herrn Prof. Dr. Reiner Wolff (zweiter Referent)

Freiburg i. Ue. 2002

Mit der Annahme einer Dissertation beabsichtigt die Wirtschafts- und Sozialwissenschaftliche Fakultät der Universität Freiburg nicht, zu den darin enthaltenen wissenschaftlichen Meinungen des Verfassers Stellung zu nehmen (Fakultätsbeschluss vom 23. Januar 1990).

Vorwort

Ich möchte mich vor allem bei Björn Schmolck für seine einmalige Unterstützung, Geduld und Ermutigung bedanken. Die langen und oft angeregten Diskussionen waren eine unermessliche Quelle der Motivation.

Meinem Doktorvater Herrn Professor Dr. Hans Wolfgang Brachinger danke ich dafür, daß er mir die Freiheit gegeben hat, meine Forschungsinteressen zu verfolgen. Seine Kommentare zur Verbesserung dieser Arbeit habe ich sehr geschätzt. Ebenfalls möchte ich mich bei Herrn Professor Dr. Reiner Wolff für die Übernahme des Korreferats zu dieser Arbeit bedanken. Seine Kommentare, insbesondere über die Mikroökonomie, haben zur Verbesserung dieser Arbeit beigetragen. Nicht weniger danke ich allen meinen Kollegen vom Seminar für Statistik der Universität Fribourg für unsere gemeinsame Arbeit und ihre Unterstützung. Insbesondere denke ich an Paul-André Monney, Uwe Steinhauser und Ali Vakili, die meine Fragen betreffend Mathematik und Programmierung beantwortet haben. Herrn Roland Alles vom Institut Wohnen und Umwelt (Darmstadt) danke ich dafür, daß er den Datensatz für die Erstellung des Mietspiegels für die Stadt Darmstadt (D), auf welchen meine Arbeit basiert, zur Verfügung stellte. Außerdem möchte ich meinen Dank an das Studienzentrum Gerzensee, seine Professoren und Studenten, aussprechen. Dabei möchte ich mich bei Herrn Professor James Powell und insbesondere bei Herrn Professor Bo Honoré für ihr Interesse an meiner Arbeit und die Diskussionen bedanken.

Dankbar bin ich Claudio Kasperl und Kathrin Schmolck für die sprachliche Unterstützung. Ja, die deutsche Sprache: Sie hatte ich anfangs unterschätzt!

Nicht zuletzt danke ich ganz herzlich meinen Freunden und meiner Familie, die mich jederzeit unterstützt und Verständnis für meine "Abwesenheit" gezeigt haben. Ich freue

mich darauf, wieder Zeit für sie zu haben. Ebenfalls möchte ich meinen neuen Arbeitskollegen für ihre Ermunterung und ihr Verständnis danken, die zum letzten Schwung beigetragen haben.

Zürich, im Dezember 2002

María Josefa Redondo González

Inhaltsverzeichnis

Abbildungsverzeichnis	viii
Tabellenverzeichnis	x
Einleitung	xi

I Institutionelle und theoretische Grundlagen von Mietspiegeln

1

1 Mietspiegel und Wohnungsmarkt: eine Einführung	3
1.1 Einleitung	3
1.2 Charakterisierung von Wohnungsmärkten	5
1.3 Mietspiegel	7
1.3.1 Zweck und Aufgaben	7
1.3.2 Gesetzliche Grundlage in Deutschland	7
1.3.3 Vor- und Nachteile	9
1.3.4 Methoden zur Erstellung von Mietspiegeln	11
1.4 Schweizerischer Wohnungsmarkt und Mietspiegel	16
2 Regressionsmietspiegel und die hedonische Hypothese	21
2.1 Einleitung	21
2.2 Theoretische Grundlagen von Regressionsmietspiegeln	22
2.2.1 Die hedonische Hypothese	22
2.2.2 Mikroökonomische Fundierung der hedonischen Hypothese . . .	24
2.3 Implikationen der hedonischen Hypothese für Regressionsmietspiegel . .	33

II Semiparametrischer Modellansatz für Regressionsmiet-	37
3 Modellansätze für Regressionsmiet Spiegel	39
3.1 Einleitung	39
3.2 Modellansätze für Regressionsmiet Spiegel	40
3.3 Ein semiparametrischer Modellansatz für Regressionsmiet Spiegel	43
3.4 Semiparametrische Regressionsmodelle: eine Einführung	44
4 Schätzung von additiven semiparametrischen Modellen	47
4.1 Einleitung	47
4.2 Nichtparametrische Kernschätzung	48
4.2.1 Nichtparametrische Schätzung von Dichtefunktionen	48
4.2.2 Kernfunktionen	50
4.2.3 Die Wahl der Bandbreite	51
4.2.4 Asymptotische Eigenschaften des Nadaraya-Watson-Kernschätzers	53
4.2.5 Konfidenzintervalle für die bedingten Erwartungswertfunktionen	54
4.3 Schätzung von additiven semiparametrischen Regressionsmodellen . . .	60
4.3.1 Die asymptotische Verteilung des semiparametrischen Schätzers	61
5 Schätzung des semiparametrischen Regressionsmiet Spiegels	65
5.1 Einleitung	65
5.2 Definition von Mietpreisdeterminanten	66
5.3 Datenreduktion und Bildung von komplexen Faktoren	72
5.3.1 Einleitung in die explorative Faktorenanalyse	73
5.3.2 Anwendung von Faktorenanalyse	82
5.4 Schätzung des semiparametrischen Modells	96
5.4.1 Problemstellung	96
5.4.2 Transformation des Modells	97
5.4.3 Nichtparametrische Schätzergebnisse für die auf die Wohnfläche bedingten Erwartungswerte	98

5.4.4	Konfidenzintervalle für den geschätzten Erwartungswert der Nettomiete	99
5.4.5	Schätzung des parametrischen Teils des Modells	102
5.4.6	Asymptotische Konfidenzintervalle und Tests für den semiparametrischen Schätzer	103
5.5	Konfidenzintervalle für Mietpreise	106
5.5.1	Konfidenzintervalle für die Grundmiete	108
5.5.2	Konfidenzintervalle für die Nettomiete	109
Zusammenfassung und Schlußfolgerungen		111
A Überblick über die Faktoren für die Mietpreisdeterminanten		115
B Kernschätzungen		119
C Asymptotische Verteilung semiparametrischer Schätzer		121
Literaturverzeichnis		127
Namenverzeichnis		133
Index		135

Abbildungsverzeichnis

2.1	Das lineare Modell für Charakteristika: Beispiel aus dem Wohnungsmarkt.	31
2.2	Beispiel einer vergessenen explikativen Variable.	35
4.1	Schema des Bootstrap-Prozesses zur Schätzung eines Konfidenzintervalls für eine Funktion $\hat{m}(x)$	56
5.1	Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Er- wartungswertfunktion für die Nettomiete	98
5.2	Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Er- wartungswertfunktion für die Nettomiete und Beobachtungen für die Nettomiete.	100
5.3	Approximierte asymptotische 95%-Konfidenzintervalle für den auf die Wohnfläche bedingten Erwartungswert der Nettomiete.	101
5.4	Vergleich zwischen Bootstrap 95%-Konfidenzintervallen mittels der BC_a -Methode und approximierten asymptotischen 95%- Konfidenzintervallen für den auf die Wohnfläche bedingten Er- wartungswert der Nettomiete	102
B.1	Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Er- wartungswertfunktion für <i>F6AUS</i>	119

B.2 Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Erwartungswertfunktion für die Variable <i>Baualter</i>	120
---	-----

Tabellenverzeichnis

1.1	Ortsübliche Vergleichsmieten in DM/qm für die Stadt Darmstadt (1977)	14
1.2	Vergleich zwischen Tabellen- und Regressionsmethode.	17
5.1	Anzahl Variablen und Faktoren vor und nach Faktorenanalyse.	83
5.2	Faktorenanalyse der Gebäudeart-Variablen.	84
5.3	Hauptfaktorenanalyse der Lage-Variablen.	87
5.4	Hauptfaktorenanalyse der Beschaffenheits-Variablen.	88
5.5	Hauptfaktorenanalyse der Sondereinrichtungs-Variablen.	90
5.6	Hauptfaktorenanalyse der Ausstattungs-Variablen	94
5.7	Schätzergebnisse des parametrischen Teils des semiparametrischen Re- gressionsmodells.	104
5.8	Statistisch signifikante Mietpreisdeterminanten und entsprechende im- plizite Preise des semiparametrischen Modells für die Nettomiete. . . .	107
5.9	Konfidenzintervalle für die Grundmiete mittels der Methode des wilden Bootstraps	109
5.10	Konfidenzintervalle für die Nettomiete mittels der Methode des wilden Bootstraps	110

Einleitung

Mietspiegel spielen in Deutschland seit Jahren eine wichtige ökonomische und rechtliche Rolle für den Wohnungsmarkt und werden häufig als Grundlage für Mietpreiserhöhungen herangezogen. Die gegenwärtige Diskussion im Rahmen der Revision des Schweizerischen Mietrechts deutet darauf hin, daß Mietspiegel in der Zukunft möglicherweise auch in der Schweiz eine wichtige Rolle spielen werden.

Unter dem Begriff “Mietspiegel” versteht man eine statistische Darstellung der durchschnittlich tatsächlich bezahlten Mietpreise für bestimmte Wohnungstypen in einer Stadt/Gemeinde. Der Wohnungstyp wird durch spezifische Wohnungscharakteristika wie z.B. Größe, Alter oder Ausstattung einer Wohnung festgelegt. Die in einem Mietspiegel dargestellten durchschnittlichen Mietpreise sind Schätzungen, die anhand einer Stichprobe aus der Population aller Wohnungen der Stadt/Gemeinde bestimmt werden. In der Praxis wird zwischen sogenannten Tabellenmietspiegeln und sogenannten Regressionsmietspiegeln unterschieden. Regressionsmietspiegel sind dadurch gekennzeichnet, daß der Mietpreis mit Hilfe eines Regressionsansatzes erklärt wird, in den verschiedene Wohnungscharakteristika als erklärende Variablen eingehen. Die in der Praxis erstellten Regressionsmietspiegel unterscheiden sich in der Form des Regressionsansatzes.

In dieser Arbeit wird ein neues allgemeineres Regressionsmodells eines Regressionsmietspiegels entwickelt und geschätzt. Es handelt sich dabei um ein additiv-multiplikatives semiparametrisches Regressionsmodell, in dem der Mietpreis als das Produkt der Grundmiete und des Nettomietfaktors spezifiziert wird. Die Grundmiete ist dabei

derjenige Teil des Mietpreises, der von der Wohnfläche abhängt, während der Einfluß aller anderen Mietpreisdeterminanten durch den Nettomietfaktor modelliert wird. Bei solchen Spezifikationen wirkt der Einfluß jeder Mietpreisdeterminante additiv, und gleichzeitig werden die Interaktionen zwischen der Wohnfläche und den restlichen Mietpreisdeterminanten berücksichtigt. Darüber hinaus wird der Zusammenhang zwischen dem Mietpreis und der Wohnfläche in solchen Modellen nichtparametrisch modelliert, während der Einfluß der restlichen Mietpreisdeterminanten parametrisch spezifiziert wird (deshalb werden solche Modelle auch als “semiparametrische Modelle” bezeichnet). Diese Spezifikation erlaubt eine große Flexibilität bei der Modellierung des Zusammenhangs zwischen der Wohnfläche und dem Mietpreis. Darüber hinaus hat dieses Modell nicht die a priori Restriktionen der üblichen Regressionsmietspiegel und stellt somit eine Verallgemeinerung dar.

Das in dieser Arbeit vorgeschlagene semiparametrische Regressionsmodell wird auf der Grundlage eines Datensatzes aus der Stadt Darmstadt (DE) geschätzt. Die Darmstadt-Daten wurden explizit 1994/95 für einen Mietspiegel der Stadt erhoben. Da die Schätzmethoden für nichtlineare semiparametrische Regressionsmodelle in der Praxis wenig bekannt sind, werden sie im Rahmen des Anwendungsbeispiels diskutiert.

Diese Arbeit umfaßt zwei Teile: Der erste Teil befaßt sich mit den institutionellen und theoretischen Grundlagen von Mietspiegeln, während im zweiten Teil ein spezifisches Regressionsmodell, das semiparametrische Regressionsmodell, zur Erstellung von Mietspiegeln diskutiert wird.

Kapitel 1 beschreibt das Konzept eines Mietspiegels und seine Rolle im Rahmen des Mietrechts in Deutschland. Die gesetzlichen Grundlagen für Mietspiegel in Deutschland zunächst werden erörtert. Die Aufgaben eines Mietspiegels sowie seine Vor- und Nachteile werden ebenfalls diskutiert. Zur Erstellung von Mietspiegeln stehen grundsätzlich zwei Methoden zur Verfügung: die Tabellen- und die Regressionsmethode. Beide werden in diesem Kapitel betrachtet. Aufgrund der Vorteile der Regressionsmethode wird diese bei der Erstellung von Mietspiegeln in der Praxis in Deutschland häufig vorgezogen. Dieses Kapitel schließt mit einem Überblick über die Diskussion der gegenwärtigen

Mietrevision in der Schweiz.

Kapitel 2 befaßt sich mit einer der zwei Methoden zur Erstellung von Mietspiegeln, dem Regressionsmietspiegel. Der Regressionsmietspiegel geht von der Annahme aus, daß der Mietpreis einer Wohnung von verschiedenen Charakteristika dieser Wohnung abhängt. Diese Annahme basiert auf der hedonischen Hypothese, welche besagt, daß heterogene Güter durch ihre Charakteristika gekennzeichnet sind und es diese Charakteristika sind, die den Konsumenten einen Nutzen stiften.

Sowohl die Definition der hedonischen Hypothese als auch zwei mikroökonomische Modelle werden eingeführt. Dabei handelt es sich um das Standardmodell der Produktionstheorie des Haushalts und das lineare Modell für Charakteristika, auf welchen die hedonische Hypothese basiert. Die Eigenschaften dieser Modelle erlauben es, bestimmte Aussagen für die Erstellung von Regressionsmietspiegeln zu machen, insbesondere für die Spezifikation des Modellansatzes und die Variablen, die in einer Analyse berücksichtigt werden sollten.

Im zweiten Teil der Arbeit wird ein neues allgemeineres Modell eines Mietspiegels entwickelt. Dieses Modell wird zuerst in Kapitel 3 eingeführt, nachdem die in Deutschland üblichen Modellansätze für Regressionsmietspiegel kurz vorgestellt wurden. In der Literatur werden verschiedene Regressionsmietspiegel je nach Art des Regressionsansatzes unterschieden. Eine erste Klasse von Modellen ist dadurch gekennzeichnet, daß der Zusammenhang zwischen dem Mietpreis und der Wohnfläche explizit parametrisch spezifiziert wird. Im Gegensatz dazu gehört das hier vorgeschlagene Regressionsmodell zu einer zweiten Klasse von Modellen, den sogenannten semiparametrischen Regressionsmodellen, bei denen dieser Zusammenhang nichtparametrisch modelliert wird, während der Einfluß der restlichen Mietpreisdeterminanten weiterhin parametrisch modelliert wird. Semiparametrische Regressionsmodelle sind in der Praxis bisher wenig bekannt und stellen eine Alternative zur Modellierung eines Mietspiegels dar. Zu einem besseren Verständnis des vorgeschlagenen semiparametrischen Regressionsmodells werden die Grundlagen semiparametrischer Regressionsmodelle dargestellt. Dabei wird ein rein additives semiparametrisches Modell betrachtet.

In Kapitel 4 wird auf die Schätzung von semiparametrischen Regressionsmodellen eingegangen. Insbesondere werden Schätzmethoden für den einfacheren Fall des additiven semiparametrischen Regressionsmodells diskutiert. Zur Schätzung der nichtparametrischen Komponente des Modells wird der Nadaraya-Watson-Kernschätzer herangezogen. Darüber hinaus werden Simulationstechniken zur Schätzung von Konfidenzintervallen erläutert. In Anschluß daran wird die Schätzung der parametrischen Komponente des Regressionsmodells betrachtet.

Kapitel 5 befaßt sich mit einem Anwendungsbeispiel des in dieser Arbeit vorgeschlagenen semiparametrischen Regressionsmodells für Mietpreise. Konkret wird auf die Schätzung des Regressionsmodells anhand von Daten eingegangen, die 1994/95 für den Mietspiegel der Stadt Darmstadt (D) erhoben wurden. Dieser Datensatz enthält eine Vielfalt von Variablen, die auf den Mietpreis wirken. Aufgrund der großen Zahl von erklärenden Variablen, die ähnliche Information enthalten, werden die Daten mit Hilfe der Faktorenanalyse reduziert. Danach wird ausführlich die Qualität der Schätzung betrachtet, wobei die asymptotische Verteilung des semiparametrischen Schätzers im Detail im Anhang C hergeleitet wird. Diese asymptotische Verteilung ist die Basis für Tests auf Signifikanz der Parameter des semiparametrischen Modells. Schließlich werden Konfidenzintervalle sowohl für die Grundmiete als auch für die Nettomiete ermittelt, welche für die Erstellung von Mietspiegeln relevant sind.

Teil I

Institutionelle und theoretische Grundlagen von Mietspiegeln

Kapitel 1

Mietspiegel und Wohnungsmarkt: eine Einführung

1.1 Einleitung

In diesem Einführungskapitel soll das Konzept eines Mietspiegels diskutiert werden. Unter einem Mietspiegel versteht man eine statistische Darstellung der durchschnittlich tatsächlich gezahlten Mietpreise für bestimmte Wohnungstypen in einer Stadt/Gemeinde. Der Wohnungstyp wird durch spezifische Wohnungscharakteristika wie z.B. Größe, Alter oder Ausstattung einer Wohnung festgelegt. Die in einem Mietspiegel dargestellten durchschnittlichen Mietpreise sind Schätzungen, die anhand einer Stichprobe aus der Population aller Wohnungen der Stadt/Gemeinde bestimmt werden. In Deutschland spielt der Mietspiegel eine wichtige ökonomische und rechtliche Rolle für den Wohnungsmarkt und wird häufig als Argument für Mietpreiserhöhungen herangezogen. Die gegenwärtige Diskussion in der Schweiz deutet darauf hin, daß der Mietspiegel in der Zukunft möglicherweise auch hier eine wichtige Rolle spielen wird. Deshalb ist es sinnvoll, die jahrelangen Erfahrungen mit Mietspiegeln in Deutschland näher zu betrachten. Insbesondere lohnt es sich, auf die Methoden zur Erstellung von Mietspiegeln ausführlich einzugehen, da zum einen die geschätzten Durchschnittsmie-

ten von der Methodik abhängen. Zum anderen beeinflusst sie, inwieweit eine Miete bei Rechtsstreitigkeiten als mißbräuchlich definiert werden kann.

In einem “Land von Mietern” wie der Schweiz (mit einem Mieteranteil von über 60% im Jahre 1997) ist der Mietwohnungsmarkt von hoher Relevanz. Wegen ihrer weitreichenden Effekte finden Änderungen auf diesem Markt große Beachtung. Einige Änderungen zur Mietpreisbildung sind im Vorschlag des Schweizerischen Bundesrats vom 15.09.1999 vorgeschlagen worden.

Diese Reform des Mietrechts würde das bisherige Regime der Kostenmiete aufheben, das durch die Koppelung der Mieten an den Hypothekarzins und andere Kosten gekennzeichnet ist. Die Fixierung der Miete auf den Hypothekarzins basierte auf dem politischen Wunsch nach einer Gewährleistung angemessener Renditen für den Eigentümer zur Förderung der Wohnungsbauinvestitionen. Mit der Revision des Mietrechts ist die seit der Einführung der Kostenmiete im Jahre 1990 ausgebrochene Debatte über Markt- versus Kostenmiete an einen Wendepunkt gelangt. Dies bedeutet aber nicht, daß die Marktmiete – bei der sich die Mietpreise aus dem freien Spiel von Angebot und Nachfrage ergeben – eingeführt wird. Die Empfehlung des Schweizerischen Bundesrates zur Mietzinsanpassung beinhaltet die Koppelung der Mieten an den Landesindex der Konsumentenpreise und an eine “ortsübliche Vergleichsmiete”. Die ortsübliche Vergleichsmiete wird dabei gemäß dem Schweizerischen Bundesamt für Wohnungswesen durch die “statistisch erhobenen Durchschnittsmietzinse”, definiert (siehe BUNDESAMT FÜR WOHNUNGSWESEN (1999), S. 18). In Deutschland versteht man unter dem Begriff “ortsübliche Vergleichsmiete” die in einer Stadt/Gemeinde durchschnittlich tatsächlich gezahlten Entgelte für nicht preisgebundenen Wohnraum vergleichbarer Art, Größe, Ausstattung, Beschaffenheit und Lage. Die durch die Vergleichsmiete (oder durch den Mietspiegel) gelieferte Information ist ein wertvolles Instrument zum Erreichen erhöhter Markttransparenz.

Nach einer kurzen Beschreibung der Charakteristika des Wohnungsmarkts, der sich aus den Mietwohnungs- und Eigentumswohnungsteilmärkten zusammensetzt, werden im diesen Kapitel zunächst die Grundlagen von Mietspiegeln dargestellt: die Definition

des Begriffs, der Ziele, Aufgaben und gesetzlichen Grundlagen von Mietspiegeln in Deutschland, sowie ein Überblick über die Methoden zu ihrer Erstellung. Dieses Kapitel endet mit einer Einführung in den Schweizerischen Wohnungsmarkt und einer Skizze der für dieses Thema relevanten Aspekte der gegenwärtigen Mietrevision, die den Weg für die Einführung eines Mietspiegels in der Schweiz ebnen könnte.

1.2 Charakterisierung von Wohnungsmärkten

Um die Rolle des Mietspiegels zu verstehen, ist eine Beschreibung der Merkmale von Wohnungsmärkten hilfreich. Diese Merkmale können wie folgt beschrieben werden:

- Das Gut Wohnung ist ein heterogenes, dauerhaftes, nicht teilbares und immobiles Gut mit verschiedenen Charakteristika (Größe, Baujahr, usw.) und unterschiedlichen Funktionen. Das Gut Wohnung kann als Dienstleistung für den Wohnungsnutzer und Eigentümer oder als Kapitalinvestition für den Vermieter betrachtet werden (vgl. RONNING (1995)). Aufgrund der Klassifikation eines Wohnungsmarktes nach Charakteristika und nach Funktionen ergibt sich eine Vielfalt von Teilmärkten.
- Wohnungsmärkte sind in vielen Ländern stark reguliert. Auf Wohnungsmärkten sind die Agenten abhängig von subjektiven Beurteilungen. So existiert z.B. in Ländern wie Deutschland oder der Schweiz im “Gesellschaftsgewissen” ein Recht auf Wohnraum, auch wenn es nicht einklagbar ist. Solche subjektiven Beurteilungen werden als Argument herangezogen, eine starke Regulierung des Wohnungsmarktes zu verlangen. Ein anderes Argument für eine Reglementierung des Marktes könnte das Schutzbedürfnis des Mieters darstellen.
- Die Transaktionskosten auf dem Wohnungsmarkt sind relativ groß. Transaktionskosten sind diejenigen Kosten, die für Mieter und Vermieter bei der Kündigung eines bestehenden Mietvertrages und/oder dem Abschluß eines neuen Mietverhältnisses entstehen. Dazu gehören etwa Suchkosten (Inserate, Vermittlungs-

kosten usw.), Einzugs- bzw. Umzugskosten, Reinigungskosten und Reparaturkosten. LAMBELET/ZIMMERMANN (1991), S. 82 schließen unter diesem Begriff auch die psychologischen Kosten ein, welche sowohl auf der Seite der Mieter als auch auf der der Vermieter bei einem Wohnungs- bzw. Mieterwechsel entstehen.

Eine Folge der hohen Transaktionskosten ist laut JAEGER/BÜHLER (1995) eine niedrige Preiselastizität der Nachfrage für den Mieter und eine sehr niedrige Preiselastizität des Angebotes auf der Vermieterseite solange die Höhe der Transaktionskosten Preiserhöhungen übersteigt (siehe JAEGER/BÜHLER (1995), S. 44 ff. für eine ausführliche Analyse der Rolle dieser Kosten auf dem Mietwohnungsmarkt).

- Die Information über Preise und Mengen auf dem Wohnungsmarkt ist asymmetrisch bezüglich der Mieter, die einen beschränkten oder zumindest einen relativ teuren Zugang zu Information haben, da sie sich an Zwischenhändler und Immobilienverwaltungen wenden müssen. Dadurch wird das Suchen einer Wohnung erschwert. Die Veröffentlichung von Mietspiegeln soll diese Art von Intransparenz vermindern.

Diese Charakteristika des Wohnungsmarktes sind sicher nicht erschöpfend: Die Verbindung zwischen diesem Markt und den Kapitalmärkten als Finanzierungsquelle, die negativen externen Effekte dieses Gutes auf die Landschaft, die Rolle der Spekulation und des Steuerwesens auf diesem Markt, usw. sind weitere Elemente, die einen Einfluß auf das Funktionieren des Marktes ausüben (siehe LAMBELET/ZIMMERMANN (1991)).

Als Hauptquellen von Intransparenzen auf dem Wohnungsmarkt kann man die Heterogenität des Gutes Wohnung und die asymmetrische Information hervorheben. Diese können durch Mietspiegel teilweise überwunden werden. In den nächsten Abschnitten werden die Gründe dafür diskutiert.

1.3 Mietspiegel

1.3.1 Zweck und Aufgaben

Das wesentliche Ziel eines Mietspiegels ist, die Transparenz auf dem Wohnungsmarkt zu erhöhen. Gleichzeitig soll er die Informationskosten für die Einzelakteure möglichst niedrig halten. Insbesondere soll ein Mietspiegel helfen, Streit zwischen Mietvertragsparteien zu vermeiden und bei Mietstreitfällen in gerichtlichen Verfahren Information über das Mietpreisniveau zu liefern.

Mietspiegel können zu verschiedenen Zwecken herangezogen werden. Dazu gehören etwa (siehe WULLKOPF (1996) S. 455):

- Hilfsmittel beim Aushandeln von Mieterhöhungen,
- Entscheidungsgrundlage bei Gerichtsverfahren in Zusammenhang mit Mietstreitigkeiten,
- gesetzliche Definition von “überhöhten” Mietpreisen,
- Bemessung des “Nutzungswertes” von Wohnungen bei familiengerichtlichen Streitigkeiten oder Räumungsklagen,
- Basis für die Bestimmung des “geldwerten Vorteils” in Finanzgerichtsverfahren,
- Bemessung des Steuerwertes bei Vermögens- und Erbschaftssteuerfragen.

Zu weiteren Details über die Funktionen von Mietspiegeln im deutschen Rechtssystem siehe etwa RIPS (1996) und KOFNER (1999).

1.3.2 Gesetzliche Grundlage in Deutschland

Das Konzept des Mietspiegels wurde 1971 in das deutsche Mietrecht eingeführt. Die Erstellung von Mietspiegeln hat sich aber erst mit dem Gesetz zur Regelung der Miethöhe

(MHG) von 1974 verbreitet, in dem der Mietspiegel eine von drei im Gesetz vorgesehenen Verfahrensweisen ist, um ortsübliche Vergleichsmieten zu bestimmen. Die drei unter §2 Abs. 2 des MHGs vorgesehenen Methoden zum Nachweis von ortsüblichen Vergleichsmieten sind die folgenden:

1. die Benennung von mindestens drei Vergleichswohnungen,
2. das Vorlegen eines Sachverständigengutachtens,
3. die Verwendung des Mietspiegels der betreffenden Stadt (falls dieser zur Verfügung steht).

Von den drei Alternativen steht der Mietspiegel in Deutschland als juristisches Beweismittel für die Bestimmung von Vergleichsmieten im Vordergrund und wird in der Gerichtspraxis am häufigsten benutzt (vgl. RONNING (1996)).

Andere Elemente des deutschen Mietrechts spielen neben dem Mietspiegel eine wichtige Rolle zur Bestimmung von Vergleichsmieten und zur Regelung von Mieterhöhungen, wie z.B. die sogenannten *Kappungsgrenzen* (Obergrenzen), die in das 2. Wohnraumkündigungsschutzgesetz (wovon das MHG ein Bestandteil ist) eingefügt wurden. Die Kappungsgrenzen begrenzen die Erhöhung von Mietzinsen und sind durch die folgenden drei Bedingungen definiert, die bei Mietzinserhöhungen erfüllt werden müssen:

1. Die bisherige Miete darf seit einem Jahr nicht verändert worden sein.
2. Die neue Miete darf die ortsübliche Miete für vergleichbare Wohnungen nicht übersteigen. Außerdem berücksichtigt die ortsübliche Miete nur die üblichen Entgelte, die in den letzten vier Jahren vereinbart oder geändert wurden. Diese Bedingung ist für Mietspiegel auch relevant, da sie die Mieten erwähnt, die zur Bestimmung von ortsüblichen Vergleichsmiete zulässig sind.

3. Die Miete darf sich innerhalb eines Zeitraums von drei Jahren (Einbeziehungsfrist) nicht um mehr als 30% erhöht haben¹.

Zusätzlich sieht das Gesetz einige Ausnahmen zu dieser letzten Maximalerhöhung um 30% vor, wie z.B. eine Reduktion der Maximalerhöhung auf bis zu 20% für Wohnungen, die vor 1981 gebaut wurden und deren gegenwärtiger Quadratmeterpreis höher als 8 DM ist (vgl. §2 Abs. 1 des VIERTEN MIETRECHTSÄNDERUNGSGESETZES (1993) für weitere Ausnahmen). Die Kappungsgrenzen sollen einen Schutz für die Mieter gewähren, indem sehr hohe Mietpreissteigerungen verhindert werden. Anhand der Maximalerhöhung von 30% und der Einbeziehungsfrist von 3 Jahren regulieren die Kappungsgrenzen auch die Geschwindigkeit des Anpassungsprozesses niedriger Mieten an die ortsübliche Miete.

1.3.3 Vor- und Nachteile

Wie RONNING (1996) erwähnt, sind Mietspiegel das in Deutschland am häufigsten verwendete Verfahren zur Bestimmung ortsüblicher Vergleichsmieten. Hauptsächlich liefern sie Schätzungen für die durchschnittlichen Mieten einer Gemeinde, und je nach Methode, auch Schätzungen für die Miethöhe von Einzelwohnungen (vgl. die Hinweise des BUNDESMINISTERIUMS FÜR RAUMORDNUNG (1980), S. 166). Gegenüber der Benennung von drei Vergleichswohnungen und dem Vorlegen eines Sachverständigen-gutachtens hat der Mietspiegel folgende Vorteile (vgl. RONNING (1995)):

Der Mietspiegel

1. ist transparenter, weil er auf einer nachvollziehbaren Datenbasis beruht,
2. ist für Laien leicht zugänglich und einfach zu benutzen,

¹Diese zunächst bis zum 1.9.1998 geltende Regelung wurde im Verlauf des Jahres 1999 bis zum 31.12.2002 verlängert. Aufgrund der Revision des Mietrechts vom 19. Juni 2001 wird eine einheitliche Kappungsgrenze von 20% in drei Jahren eingeführt (siehe MIETRECHTSREFORMGESETZ (2001)).

3. ist für Privatpersonen (Mieter und Vermieter) eine kostengünstige Informationsquelle,
4. liefert symmetrische Informationen für jeden Agenten des Wohnungsmarktes. Dabei sollte die Stadt/Gemeinde für die Erstellung des Mietspiegels verantwortlich sein, damit die Sonderinteressen der Einzelparteien (Mieter und Vermieter) den Mietspiegel möglichst nicht verzerren.

RONNING (1995) und WULLKOPF (1996) stellen einige Nachteile von Mietspiegeln fest (zur methodologischen Kritik siehe auch KRÄMER (1992a)) und zwar:

1. Die Definition der Grundgesamtheit oder Population der relevanten Wohnungen beeinflusst die Höhe der ortsüblichen Vergleichsmiete. Unter §2 Abs. 2 legt das deutsche Miethöhegesetz eine Einbeziehungsfrist von vier Jahren für die Erstellung von Mietspiegeln fest: Hierfür werden nur die Mieten von nicht-preisgebundenem Wohnraum vergleichbarer Art, Größe, Ausstattung, Beschaffenheit und Lage berücksichtigt, deren Höhe in den letzten vier Jahren neu festgelegt oder geändert worden ist. Damit wird die Population oder Gruppe von relevanten Wohnungen eingegrenzt. Bei steigenden Preisen ist die Durchschnittsmiete bei einer kürzeren Einbeziehungsfrist höher, da die Durchschnittsmiete für die Population der aktuellsten Mietanpassungen und Neuabschlüsse von Mietverträgen höher ist als die einer Population, die ältere Mietverträge einschließt. Diese Verschiebung überträgt sich auf die ortsübliche Vergleichsmiete, die desto höher wird, je kürzer die gemäß Gesetz festgelegte Einbeziehungsfrist ist: Eine Einbeziehungsfrist von 4 Jahren führt bei steigenden Mietpreisen zu einer niedrigeren Vergleichsmiete als eine Frist von 2 Jahren.
2. Im Vergleich zu den anderen Methoden zum Nachweis von ortsüblichen Vergleichsmieten ist der Mietspiegel kostenintensiver, da er eine aufwendige statistische Erhebung erfordert.
3. Ein Mietspiegel muß fortgeschrieben werden, normalerweise alle 2 Jahre, um informativ über die aktuelle Marktsituation zu sein.

4. Die Berücksichtigung von Einflußfaktoren zur Bestimmung der ortsüblichen Vergleichsmiete ist wegen normativer Vorschriften in Deutschland (§2 Abs. 2 des MHGs) auf bestimmte objektive Merkmale beschränkt. Andere Faktoren wie z.B. die Vertragsdauer haben in der Praxis ebenfalls einen Einfluß auf die Miethöhe, werden aber nicht berücksichtigt, da sie als “subjektiv” betrachtet werden, d.h. sie beziehen sich spezifisch auf den Mieter statt auf den physischen Gegenstand. So kann die Nichtberücksichtigung mietpreisrelevanter Einflußfaktoren bei der Bestimmung der ortsüblichen Vergleichsmiete zu Verzerrungen führen und somit die Aussagekraft eines Mietspiegels beeinträchtigen.
5. Eventuell erweist sich die Erhebung einer “repräsentativen” Stichprobe für alle Wohnungstypen als schwierig. Repräsentativ heißt, daß von einer Stichprobe von Mietpreisbeobachtungen über einen bestimmten Wohnungstyp auf die Grundgesamtheit dieses Wohnungstyps geschlossen werden kann.

1.3.4 Methoden zur Erstellung von Mietspiegeln

Das Hauptziel von Mietspiegeln ist, eine erhöhte Transparenz auf dem Wohnungsmarkt zu gewährleisten. Dies wird anhand der bei Mietspiegeln ermittelten Information über die ortsüblichen Mieten von verschiedenen Wohnungstypen erreicht. In diesem Sinn muss die Typisierung von Wohnungen als ein Zwischenschritt bei der Erstellung eines Mietspiegels betrachtet werden. Diese Typisierung wird anhand eines Mietbegriffs und anhand von Einflußfaktoren oder Mietpreisdeterminanten, die den jeweiligen Wohnungstyp charakterisieren, durchgeführt.

Die einem Mietspiegel zugrunde liegende Definition von Mietzins ist heutzutage die sogenannte *monatliche Nettomiete* (vgl. etwa die Hinweise des BUNDESMINISTERIUMS FÜR RAUMORDNUNG (1980), S. 167). Unter diesem Begriff versteht man die monatliche Gesamtmietzahlung ohne die Nebenkosten. Zu den Nebenkosten gehören die Vorauszahlungen für Heizung und Warmwasser und die anfallenden Nebenkosten für Müllabfuhr, Hausbeleuchtung, Wasser/Abwasser usw. (für eine Aufzählung möglicher Nebenkosten siehe etwa ALLES (1990), Anhang 7). In der Praxis ist es oft schwierig,

die Beträge für die verschiedenen Nebenkosten auszuweisen, weil diese in individuellen Mietverträgen nicht immer explizit angegeben werden. In solchen Fällen müssen sie geschätzt und von der Gesamtmiete subtrahiert werden. Hierauf soll an dieser Stelle nicht näher eingegangen werden. Für eine detaillierte Information über die Schätzung der Betriebskosten sei auf ALLES (1990), S. 6 ff. hingewiesen.

Bei der Erstellung des Mietspiegels sollten Merkmale der Wohnungen berücksichtigt werden, die einen Einfluß auf den Mietpreis ausüben. Der deutsche Gesetzgeber schlägt im §2 Absatz 1 Nr. 2 des MHGs die Charakteristika vor, die den jeweiligen Wohnungstyp kennzeichnen und zur Bestimmung der Vergleichsmiete relevant sind: Art, Größe, Ausstattung, Beschaffenheit und Lage. Diese Charakteristika können als Grundmerkmale betrachtet werden. Andere Merkmale sind aber nicht ausgeschlossen, falls sie einen Einfluß auf den Mietpreis haben. Dieses Thema wird im Abschnitt 5.2 des Kapitels 5 vertieft, wo die Einflußvariablen auf den Mietpreis definiert und praktische Überlegungen zu einem konkreten Fall angestellt werden.

Jeder Wohnungstyp wird durch eine bestimmte Konstellation von Merkmalen charakterisiert, die ihrerseits den Erwartungswert des Wohnungstyps bestimmt. Für die Schätzung der aktuellen ortsüblichen Miete für jeden Wohnungstyp werden üblicherweise zwei statistische Verfahren verwendet: die Tabellen- und die Regressionsmethode. Je nach gewählter Methode zur Darstellung der ortsüblichen Mieten ergibt sich der sogenannte *Tabellenmietspiegel* oder der *Regressionsmietspiegel*. Der Tabellenmietspiegel ist schon seit längerem bekannt. In den letzten Jahren hat aber der Regressionsmietspiegel in Deutschland an Bedeutung gewonnen. Nach der Erläuterung von Tabellen- und Regressionsmietspiegel erfolgt ein Vergleich zwischen den beiden Methoden in Tabelle 1.2.

Tabellenmietspiegel

Ein Tabellenmietspiegel besteht aus einer Tabelle, in der jede Zelle einem bestimmten Wohnungstyp entspricht. Die Miete jedes Wohnungstyps (oder jeder Wohnungsklasse) ist durch verschiedene Einflußfaktoren gekennzeichnet. Damit diese Darstellung möglich ist, wird angenommen, daß die Einflußfaktoren als kategoriale Variablen be-

trachtet werden können. Für quantitative Einflußvariablen wie z.B. die Wohnfläche werden die Daten in Klassen eingeteilt. Aufgrund dieser Kategorisierung, sind bei dem Tabellenmietspiegel nur Aussagen über die a priori gebildeten Klassen (Wohnungstypen) möglich.

Ein Tabellenmietspiegel ist in Tabelle 1.1 abgebildet. Es handelt sich hier um einen Tabellenmietspiegel für die Stadt Darmstadt (Deutschland), der aufgrund einer repräsentativen Erhebung im März 1977 ermittelt wurde (vgl. MAGISTRAT DER STADT DARMSTADT (1977)). Drei Grundmerkmale sind hier berücksichtigt: Baujahr des Gebäudes mit 6 Klassen, Wohnungsgröße und Ausstattung mit jeweils 3 Klassen. Mit so wenigen Ausprägungen der betrachteten Merkmale bleibt der Mietspiegel überschaubar. Leere Zellen bedeuten, daß zur Bestimmung der Durchschnittsmiete zu wenig Beobachtungen über Wohnungen mit diesen Charakteristika vorlagen, um eine sinnvolle Aussage über die Grundgesamtheit dieses Wohnungstyps machen zu können (Problem des Stichprobenumfangs).

In Hinblick auf eine übersichtliche Tabelle und auf einen hinreichenden Stichprobenumfang pro Zelle (siehe dazu Punkt 1 über Nachteile von Mietspiegeln in Abschnitt 1.3.3) soll die Klassenbildung so “sparsam wie möglich” gehalten werden. Tatsächlich muß man einen Kompromiß finden: Die Anzahl der Klassen sollte hinreichend gross sein, um die Homogenität und eine gewisse Ähnlichkeit der Stichprobe jeder Klasse in Bezug auf die jeweilige Grundgesamtheit zu gewährleisten. Die Anzahl der Klassen sollte allerdings auch nicht zu gross sein, um leere Zellen (oder solche Zellen mit sehr wenigen Beobachtungen) zu vermeiden. Als Faustregel haben OBERHOFER/SCHMIDT (1993) eine Besetzung von mindestens 30 Wohnungen pro Zelle oder Wohnungsklasse vorgeschlagen. Nach der Festlegung der Klassen wird die Durchschnittsmiete für jede Klasse oder jeden Wohnungstyp bestimmt, in der Praxis oft als Durchschnittsmiete pro Quadratmeter. Am häufigsten wird das arithmetische Mittel oder der Median als Durchschnitt verwendet.

Wie RONNING (1995) erwähnt, ist der Tabellenmietspiegel relativ einfach zu erstellen und für Laien auch einfach zu verstehen; sie ist wegen der beschränkten Anzahl von

Baujahr	Größe	Ausstattung*		
		A	B	C
bis 1918	über 90 qm	2.48	3.28	3.90
	61-90 qm	2.90	3.68	4.15
	bis 60 qm	3.15	4.03	4.55
1919-1948	über 90 qm	-	3.05	4.40
	61-90 qm	2.59	3.64	3.98
	bis 60 qm	2.94	3.67	4.28
1949-60	über 90 qm	-	-	4.33
	61-90 qm	-	3.72	4.80
	bis 60 qm	-	4.29	5.95
1961-68	über 90 qm	-	-	5.08
	61-90 qm	-	-	5.03
	bis 60 qm	-	-	6.06
1969-72	über 90 qm	-	-	4.84
	61-90 qm	-	-	5.45
	bis 60 qm	-	-	6.07
1973-1976	über 90 qm	-	-	5.22
	61-90 qm	-	-	5.42
	bis 60 qm	-	-	6.72

* A: ohne Bad bzw. Dusche und ohne Sammel- bzw. Etagenheizung

B: mit Bad bzw. Dusche oder mit Sammel- bzw. Etagenheizung

C: mit Bad bzw. Dusche und mit Sammel- bzw. Etagenheizung

Tabelle 1.1. Ortsübliche Vergleichsmieten in DM/qm für die Stadt Darmstadt, Stand: März 1977. Quelle: MAGISTRAT DER STADT DARMSTADT (1977).

Einflußfaktoren relativ grob und wegen der Notwendigkeit eines großen Stichprobenumfangs auch relativ “teuer” im Vergleich zur Regressionsmethode (weil jede Zelle oder jeder Wohnungstyp eine Mindestbesetzung braucht).

Ein Vergleich der Tabellen- und Regressionsmethode befindet sich in Tabelle 1.2². Diese Tabelle gibt einen Überblick über die Unterschiede zwischen beiden Methoden bezüglich Modellhypothesen, Stichprobenumfang, Datenanalyse und Ergebnisdarstellung. In Bezug auf die Ergebnissdarstellung werden sogenannte “Mietspannen” betrachtet. Bei Mietspiegeln ist nicht nur die Durchschnittsmiete von Interesse, sondern auch eine Spannbreite, in der die üblichen Mieten liegen. Diese Darstellung wird in Deutschland als Mietspanne genannt. Das deutsche Bundesministerium für Raumordnung schlägt vor, die ortsüblichen Mieten durch Mietspannen darzustellen, so daß zumindest zwei Drittel aller erfaßten Mieten der betreffenden Merkmalskombination innerhalb der Spannen liegen.

Regressionsmietspiegel

Beim Regressionsmietspiegel wird wie beim Tabellenmietspiegel davon ausgegangen, daß die Nettomiete durch verschiedene Einflußfaktoren erklärt ist. Der Zusammenhang zwischen der Nettomiete und den Einflußfaktoren wird durch einen Regressionsansatz modelliert.

Ursprünglich wurde beim Regressionsmietspiegel ein rein linearer Ansatz gewählt. Gemäß diesem Ansatz setzt sich der Mietpreis additiv aus einem sogenannten Grundbetrag, der unabhängig von spezifischen Charakteristika ist, und aus den für die jeweiligen Wohnungscharakteristika entsprechenden Beträgen zusammen. Das Problem bei diesem Ansatz besteht darin, daß die zwischen verschiedenen Einflußvariablen bestehenden Interaktionen nicht automatisch berücksichtigt werden: Der Einfluß der Variable Ausstattung auf den Mietpreis z.B. hängt von der Lage oder Wohnfläche der Wohnung ab. In der Praxis werden hauptsächlich die Interaktionen verschiedener Einflußfaktoren mit der Wohnfläche berücksichtigt, da die Wohnfläche sich als der wichtigste Einflußfaktor

²Diese Tabelle basiert teilweise auf CLAR (1992), wo auch eine detailliertere Aufzählung der Charakteristika der Tabellenmethode zu finden ist.

des Mietpreises ergeben hat. Wegen der Nichtberücksichtigung von Interaktionen ergaben sich beim Münchner Mietspiegel 1990/91 sogar negative Mietpreise. So ist zum Beispiel zu erwarten, daß der Mietpreis zwischen einer gut ausgestatteten Wohnung und einer schlecht ausgestatteten Wohnung mit dem Merkmal “Lage der Wohnung” in der Stadt variiert. Diese Interaktion wird mit einem additiven Modell nicht berücksichtigt. Additiv-multiplikative Modelle hingegen sind geeignet, solche Interaktionen zu erfassen. In späteren Kapiteln werden zwei Varianten solcher Modelle ausführlich diskutiert.

Bei der Regressionsmethode läßt sich im Gegensatz zu der Tabellenmethode eine Prognose über den Mietpreis verschiedener Merkmalskonstellationen machen, die nicht explizit in der Stichprobe berücksichtigt wurden. Da die Regressionsmethode eine Schätzung der impliziten Preise jedes in dem Modell spezifizierten Merkmalen liefert, kann der Mietpreis irgendwelcher Merkmalskonstellation geschätzt werden.

Man kann zeigen, daß die Tabellenmethode als Spezialfall der Regressionsmethode betrachtet werden kann (vgl. REDONDO (1996)). Eine Kombination aus dem Regressions- und Tabellenmietspiegel ist in derjenigen Literatur zu finden, in der die Grundmiete (welche allein von der Wohnfläche bestimmt ist) mittels der Tabellenmethode und der Einfluß der übrigen Mietpreisdeterminanten mittels der Regressionsmethode geschätzt werden (einführend siehe etwa SCHLITTEGEN/UHLIG (1997)).

Die Darstellung eines Regressionsmietspiegels besteht wie beim Tabellenmietspiegel aus einer Tabelle oder mehreren Tabellen. Die Regressionsmethode zur Erstellung von Mietspiegeln wird im zweiten Teil dieser Arbeit ausführlich betrachtet.

1.4 Schweizerischer Wohnungsmarkt und die mögliche Rolle von Mietspiegeln

Das aktuelle Mietregime stammt aus der letzten Revision des Mietrechts von 1990. Damit wurde ein Kostenmieteregime errichtet, das durch die Koppelung der Miete an

Tabellenmethode		Regressionsmethode
<i>I) Modellhypothesen</i>		
• Variablentyp	Abhängige Variable: Nettomiete (kardinal). Unabhängige Variablen: kardinal (aber kategorisiert)/qualitativ.	Abhängige Variable: Nettomiete (kardinal). Unabhängige Variablen: kardinal/qualitativ.
• Zusammenhang	Keine Spezifikation des Zusammenhangs zwischen den Variablen unterstellt. Interaktionen zwischen den Variablen automatisch berücksichtigt.	Linearer Zusammenhang zwischen der Nettomiete und der Wohnfläche. Der Einfluß der restlichen Mietpreisdeterminanten kann additiv oder multiplikativ modelliert werden. Interaktionen müssen explizit modelliert werden.
• Anzahl Charakteristika	In Hinblick auf Übersichtlichkeit, beschränkte Anzahl der Merkmale.	Größere Anzahl von Charakteristika möglich.
• Prognose	Aussage nur für die berücksichtigten Wohnungstypen möglich.	Aussage über alle Wohnungstypen möglich.
<i>II) Stichprobe</i>		
• Stichprobenumfang	Sehr großer Stichprobenumfang nötig, da jede Einzelzelle der Tabelle ausreichend besetzt werden muß.	Kleiner als bei der Tabellenmethode. Die Berücksichtigung von Interaktionen könnte eine Erhöhung des Umfangs benötigen.
<i>III) Datenanalyse und Ergebnisdarstellung</i>		
• Ab- und Zuschläge	Problem der Heterogenität innerhalb eines Wohnungstyps: Einfügen von Ab- und Zuschlägen.	Bereits im Regressionsmodell als implizite Preise berücksichtigt.
• Durchschnittsmiete	Geschätzt, als arithmetisches Mittel oder Median.	Auf die Charakteristika bedingt geschätzter Durchschnittswert.
• Mietspannen	Für jede Zelle: Quantilabstände.	Unter bestimmten Annahmen: Konfidenzintervalle.
• Mißbräuchliche Miete	Der maximale Mietwert innerhalb einer Wohnungskategorie.	Obere Grenze des Konfidenzintervalls.

Tabelle 1.2. Vergleich zwischen Tabellen- und Regressionsmethode.

Kosten wie Bau-, Boden-, Kapital-, Unterhalts- und Reparaturkosten charakterisiert ist, und deren Erhöhung proportionale Aufschläge auf den Mietpreis verursacht. Die Hypothekarzinsen beeinflussen dabei die Entwicklung der Mietzinsen besonders stark (siehe BUNDESAMT FÜR WOHNUNGSWESEN (1999), S. 9). Dieses Regime steht in Kontrast zu einem Marktmieteregime, in dem die Miete "jenen (hypothetischen) Mietpreis bezeichnet, der in einem nicht-regulierten, funktionsfähigen Markt unter wirksamen Wettbewerbsbedingungen zu einer Markträumung führt" (JAEGER/BÜHLER (1995), S. 52-53). Als negative Effekte des Kostenmieteregimes können Preisverzerrungen auf dem Wohnungsmarkt erwähnt werden. Außerdem hat die Koppelung der Mietpreiserhöhung an den Hypothekarzins in der Praxis eine asymmetrische Preis-anpassung verursacht, da höhere Zinsen sich schnell auf den Mietpreis niederschlagen, während tiefere Hypothekarzinsen bei bestehenden Mietverhältnissen kaum überwältzt werden. Dieser Mechanismus bewirkt eine gewisse Trägheit der Mietpreise, die nur mit Verzögerung auf ungünstige konjunkturelle Phasen reagieren (SCHIPS/MÜLLER (1991)).

In Hinblick auf die Revision des Mietrechts wurde das Kostenmieteregime in den letzten Jahren zur Diskussion gestellt. Prinzipiell stimmen die verschiedenen Akteure in der Lockerung der Koppelung zwischen dem Mietpreis und den Hypothekarzinsen überein. Trotzdem sind sie sich nicht darüber einig, auf welche Art und mit welcher Geschwindigkeit die Entkoppelung vorgenommen werden soll.

Mit der Initiative "Ja zu fairen Mieten" des Schweizerischen Mieterinnen- und Mieterverbands vom 14. März 1997 wurde vorgeschlagen, daß Mietzinsanpassungen nicht wie bisher an den Hypothekarzinssatz eines Jahres gekoppelt werden, sondern an den Verlauf eines geglätteten Zinssatzes, d.h. an einen über 5 Jahre hinweg berechneten Mittelwert. Dieses Vorgehen hat zum Ziel, Schwankungen der Mieten zu vermindern und eine höhere Stabilität der Mietpreise zu bewirken. Außerdem erwarten die Anhänger dieser Initiative, durch dieses Verfahren das Ungleichgewicht bei der Überwälzung von Zinsermäßigungen vs. Zinserhöhungen mit stabileren (und tieferen) Mieten zu kompensieren.

Die kontroverse Diskussion über die Regulierung des Wohnungsmarktes und insbe-

sondere über den Dualismus Marktmiete - Kostenmiete nimmt im Jahr 1999 mit dem aktuellen Gegenvorschlag des Bundesrats eine neue und "marktnähere" Richtung. Neue Elemente darin sind die Aufhebung der Koppelung des Mietpreises an den Hypothekarzins, die Indexierung der Miete durch den Landesindex der Konsumentenpreise und die Einführung der Vergleichsmiete.

Im Vergleich zu der Initiative des Schweizerischen Mieterinnen- und Mieterverbands geht der Gegenvorschlag des Bundesrates vom 15. September 1999 bei der Berücksichtigung von marktnäheren Elementen weiter. Dies betrifft die Einführung des Konzepts der Vergleichsmiete, das die Situation des Mietmarktes in Betracht zieht: Das Konzept der Vergleichsmiete basiert auf den ortsüblichen Entgelten in einer Stadt/Gemeinde und wird anhand einer statistischen Erhebung bestimmt.

Zur Mietzinsanpassung sieht der Bundesrat zunächst die *Indexierung* oder die Bindung der Miete an den Landesindex der Konsumentenpreise vor: Bis zu 80% der Teuerung sollen auf die Mieten überwältzt werden dürfen. Diese Methode soll zu einer verbesserten Transparenz in den Wohnungsmarkt führen - gegenüber dem Hypothekarzins ist der Landesindex der Konsumenten "eine einfache, transparente, breit kommunizierte volkswirtschaftliche Größe" (siehe BUNDESAMT FÜR WOHNUNGSWESEN (1999), S. 23). Darüber hinaus soll sie zu einer Mietpreisentwicklung führen, die starke Mietpreisbewegungen vermindert.

Des Weiteren wurde der Begriff *Vergleichsmiete* in die öffentliche Diskussion eingeführt. Es handelt sich um einen Begriff, der in Deutschland schon seit Jahren verwendet wird. In einer Übergangsphase bis zur Erstellung von Mietspiegeln ist die Berechnung eines Durchschnittswerts aus drei Vergleichswohnungen zur provisorischen Bestimmung der Vergleichsmiete vorgesehen. Im Vorschlag des Bundesrats ist eine Anpassung der Miete an die übliche Vergleichsmiete nur alle vier Jahre geplant. Außerdem darf die Mieterhöhung nicht größer als 20% sein. Der Schweizerische Gesetzgeber schlägt vor, daß ein mißbräuchlicher Mietzins anhand der Vergleichsmiete bestimmt wird. Dabei ist der mißbräuchliche Mietzins als derjenige definiert, der von der Durchschnittsmiete mehr als 15% nach oben abweicht (vgl. BUNDESAMT FÜR WOHNUNGSWESEN (1999),

S. 36). Der im Gegenvorschlag verwendete Begriff “statistischer Durchschnitt” für die Durchschnittsmiete wird nicht näher präzisiert. Somit ist unklar welcher Durchschnitt gemeint ist wie z.B. das arithmetische Mittel oder der Median, welche im allgemeinen zu unterschiedlichen Ergebnissen führen. Dies gilt auch für die Berechnung der mißbräuchlichen Mietpreise, die ja auf der Durchschnittsmiete basiert.

Im Rahmen der Mietrevision stehen heutzutage zwei weitere Vorschläge zur Diskussion. Der Vorschlag von der Mieterseite und den Westschweizern Vermietern, der sogenannte “accord romand”, wurde am 5. Juni 2002 im Nationalrat diskutiert. Das Konzept des “accord romand” für die Revision des Mietrechts basiert auf dem geltenden Mietrecht, sieht aber die Koppelung der Miete an den Landesindex der Konsumentenpreise statt an den Hypothekarzins vor. Außerdem dürfen die Mieten während der Mietdauer einmal im Jahr zu 80% an den Teuerungsindex angepaßt werden. Zudem erfolgt die Überprüfung auf Mißbräuchlichkeit der Mieten anhand einer Renditeberechnung und nicht anhand von Vergleichsmieten, wie der Gegenvorschlag des Bundesrates vorsieht. Dieser Vorschlag wurde vom Nationalrat abgelehnt.

Zurzeit wird ein anderes Konzept für die Revision, das “Cina-Konzept”, diskutiert. Dieser Vorschlag sieht die Festsetzung einer Anfangsmiete vor, die direkt den Vertragsparteien überlassen wird. Die Miete kann zu einem späteren Zeitpunkt durch die Anpassung an die Teuerung (Überwälzung zu 100%) erhöht werden. Analog wie beim Gegenvorschlag des Parlaments wird die Vergleichsmiete zur Mißbrauchskontrolle vorgesehen. Das Cina-Konzept beinhaltet auch, daß die Mieter sowohl bei der Anfangsmiete wie nach Ablauf von fünf Jahren eine Überprüfung auf Mißbräuchlichkeit des Mietzinses verlangen können. An der Sitzung des Nationalrats von 5. Juni 2002 wurde beschlossen, daß die Überprüfung erst nach fünf Jahren verlangt werden kann (siehe NEUER ZÜRCHER ZEITUNG vom 6.6.2002 S. 11 für weitere Details). Damit ist jedoch die Diskussion über die Mietrevision noch nicht beendet.

Kapitel 2

Regressionsmietpiegel und die hedonische Hypothese

2.1 Einleitung

Das Konzept des Regressionsmietpiegels geht von der Annahme aus, daß der Mietpreis einer Wohnung von verschiedenen Charakteristika der Wohnung abhängt. Diese Annahme basiert auf der hedonischen Hypothese, welche besagt, daß heterogene Güter durch ihre Charakteristika gekennzeichnet sind und es die Charakteristika sind, die den Konsumenten einen Nutzen stiften. Nach dieser Hypothese kann der Preis eines Gutes durch die Preise seiner Merkmale mittels eines hedonischen Modells beschrieben werden.

Der Zusammenhang zwischen Regressionsmietpiegel und hedonischer Methode wird in diesem Abschnitt ausführlich behandelt. Das Konzept des Regressionsmietpiegels wird mikroökonomisch fundiert, d.h. der funktionale Zusammenhang zwischen dem Mietpreis und den Charakteristika eines Mietobjekts im Rahmen eines Regressionsmietpiegels wird durch die hedonische Hypothese begründet.

Im folgenden wird zunächst die Definition der hedonischen Hypothese als theoretisches

Basis für den Regressionsmietspiegel erläutert. Die hedonische Hypothese selbst wird durch zwei mikroökonomische Modelle fundiert. Diese sind das Standardmodell der Produktionstheorie des Haushalts und das lineare Modell für Charakteristika. Die Eigenschaften dieser Modelle erlauben es, bestimmte Aussagen für die Erstellung von Regressionsmietspiegeln zu machen, insbesondere über die Spezifikation des Modellansatzes und die Variablen, die in einer Analyse berücksichtigt werden sollten.

2.2 Theoretische Grundlagen von Regressionsmietspiegeln

2.2.1 Die hedonische Hypothese

Die hedonische Hypothese hat ihren Ausgangspunkt in den 30er Jahren mit der Arbeit von COURT (1939) über den Effekt von Preisänderungen von Autos auf ihren Gesamtumsatz. Zur Quantifizierung von Preisänderungen, die sich aufgrund der Änderungen der Qualität von Gütern ergeben, hat er die hedonische Preismethode entwickelt, in der der Preis eines Automodells durch die verschiedenen Eigenschaften des Autos erklärt wird. GRILICHES (1961) entwickelte diese Methodologie zur Qualitätsbereinigung von Preisen weiter. Die theoretische Grundlage der modernen hedonischen Methode basiert aber auf den Arbeiten von LANCASTER (1966) und ROSEN (1974).

Die *hedonische Hypothese* besagt, daß heterogene Güter Bündel oder Kombinationen von Eigenschaften sind, wobei diese Eigenschaften gerade das Gut kennzeichnen. Dies bedeutet, daß ein heterogenes Gut wie z.B. das Gut Wohnung, aus homogenen Einheiten (den Charakteristika) besteht, die ihren Gegenstand kennzeichnen (wie z.B. für das Gut Wohnung: Wohnfläche, Baujahr, usw.). In der traditionellen Konsumtheorie wird angenommen, daß Güter dem Konsumenten einen Nutzen bringen. Somit werden die Güter als Argumente oder erklärende Faktoren in der Nutzenfunktion berücksichtigt. Gemäß der hedonischen Hypothese sind die Charakteristika und nicht die Güter die Argumente der Nutzenfunktion des Konsumenten.

In der Literatur finden sich verschiedene Definitionen von Charakteristika. Lancaster z.B. definiert Charakteristika als die wesentlichen Eigenschaften eines Einzelgutes, die es von den anderen Gütern unterscheiden (vgl. LANCASTER (1991), S. 11). Triplett bezeichnet die Charakteristika als diejenigen “homogenen ökonomischen Variablen”, die die Güter in verschiedene Gruppen oder Varianten einteilen. Der Preis jeder Variante stellt die Gesamtbewertung der jeweiligen Charakteristika dar (vgl. TRIPLETT (1986), S. 37). Mit “ökonomischen Variablen” bezeichnet Triplett die Charakteristika, die das Verhalten der Agenten beeinflussen.

Die hedonische Hypothese unterstellt einen funktionalen Zusammenhang f zwischen dem Preis p eines Gutes und seinen K Charakteristika $\mathbf{z} = (z_1, \dots, z_K)'$, d.h.

$$p = f(\mathbf{z}), \quad (2.1)$$

wobei z_k , $k = 1, \dots, K$, die Menge der k -ten Charakteristik bezeichnet und angenommen wird, daß die Funktion $f(\cdot)$ differenzierbar ist. Die Funktion $f(\cdot)$ wird *hedonische Funktion* genannt. Der sogenannte *hedonische* oder *implizite* Preis ist durch

$$\frac{\partial p}{\partial z_k}(\mathbf{z}) = \frac{\partial f}{\partial z_k}(\mathbf{z}), \quad k = 1, \dots, K$$

definiert (siehe etwa BRACHINGER (2002)). Für jedes Einzelgut mißt der hedonische Preis β_k den Effekt einer zusätzlichen Einheit der Charakteristik z_k auf seinen Preis p . Wie TRIPLETT (1986), S. 38, erwähnt, tragen die impliziten Preise diese Bezeichnung aufgrund der Unmöglichkeit ihrer direkten Beobachtung auf dem Markt, im Gegensatz zum Preis des Gutes.

Bevor die praktischen Implikationen dieser Hypothese für den Mietspiegel betrachtet werden, ist es sinnvoll, die theoretischen Grundlagen der hedonischen Hypothese darzustellen. Die hedonische Hypothese wird mit der in der mikroökonomischen Theorie bekannten “Produktionstheorie des Haushalts” und mit dem “linearen Modell für Charakteristika” verbunden. Diese beiden mikroökonomischen Fundierungen der hedonischen Hypothese werden im nächsten Abschnitt behandelt.

2.2.2 Mikroökonomische Fundierung der hedonischen Hypothese

Die theoretische Fundierung der hedonischen Hypothese beruht auf zwei mikroökonomischen Modellen: dem Standardmodell der Produktionstheorie des Haushalts und dem linearen Modell für Charakteristika. Beide Modelle wurden von Lancaster eingeführt (siehe etwa LANCASTER (1966)). Zudem basieren beide auf der Annahme der “neuen” Konsumtheorie, die besagt, daß der Nutzen eines Haushalts direkt von den Charakteristika der konsumierten Güter abhängt und nicht von den Gütern selbst wie in der “traditionellen” Konsumtheorie.

Die hedonische Funktion ergibt sich bei beiden Modellen aus der optimalen Lösung eines Optimierungsproblems: Beim Modell der Produktionstheorie des Haushalts wird dieser Zusammenhang aus einem Kostenminimierungsproblem für den Haushalt erreicht. Beim linearen Modell für Charakteristika stammt die hedonische Funktion direkt aus einem Nutzenmaximierungsproblem unter einer zusätzlichen Annahme über die Produktionsfunktion. Häufig wird die hedonische Methode mit dem Standardmodell der Produktionstheorie des Haushalts verbunden. Dagegen nehmen anderen Autoren wie z.B. DEATON/MÜLLBAUER (1980), S. 257, das lineare Modell für Charakteristika als Basis für die hedonische Hypothese. Allerdings ist das lineare Modell für Charakteristika restriktiver, weil seine Ergebnisse nur unter der Annahme einer linearen Produktionsfunktion gültig sind. In den weiteren Abschnitten werden beide Modelle vorgestellt.

2.2.2.1 Produktionstheorie des Haushalts

Das Standardmodell der Produktionstheorie des Haushalts basiert auf den Arbeiten von GORMAN (1956), BECKER (1965) und LANCASTER (1966). Dieses Modell wurde im Rahmen der Konsumtheorie entwickelt, um das Verhalten der Konsumenten bei der Auswahl zwischen Gütern zu analysieren. Prinzipiell wird unterstellt, daß Haushalte einen Nutzen aus den Charakteristika erhalten, die ausgehend von den auf dem Markt

vorhandenen Güter und mit Hilfe des Faktors Arbeit produziert werden. Weiter wird unterstellt, daß die Güter die gleichen Charakteristika aufweisen, aber in unterschiedlichen Mengen.

LANCASTER (1966), S. 277 spricht von "einer neuen Einstellung der Konsumtheorie", in der die Charakteristika und nicht die Güter einen direkten Nutzen an die Haushalte (Entscheidungseinheiten) stiften. Unter dieser neuen Perspektive wird die Nutzenfunktion des Haushalts direkt von den Charakteristika des Gutes abhängen, statt von Gütern, wie es bei der traditionellen Konsumtheorie üblich war.

In der Produktionstheorie des Haushalts wird der Haushalt gleichzeitig als Firma und als Konsument betrachtet. Einerseits ist der Haushalt eine Firma, die mittels des Inputfaktors Arbeit die auf dem Markt gehandelten Gütern transformiert und andere Güter (Output) herstellt. Diese hergestellten Güter sind die Charakteristika. Der Haushalt als Firma hat das Ziel, die Kosten dieser Herstellung zu minimieren. Andererseits kann der Haushalt als Konsument betrachtet werden, welchem diese Charakteristika einen direkten Nutzen hervorrufen. Deswegen werden die Charakteristika als Argumente der Nutzenfunktion im Rahmen des Nutzenmaximierungsproblems des Haushalts weiter berücksichtigt.

Betrachtet werden n Güter. Sei q_i eine beliebige Menge des Gutes i , $i = 1, \dots, n$, mit $q_i \geq 0$, das auf dem Markt zum Preis $p_i > 0$ gehandelt wird. Sei l ($l \geq 0$) die Menge des Faktors Arbeit zum Lohnsatz w . Dieser Faktor wird im Transformationsprozeß verwendet, um aus den Gütermengen $\mathbf{q} = (q_1, \dots, q_n)'$ ein gegebenes Produktionsniveau $\mathbf{z} = (z_1, \dots, z_K)'$ von K Charakteristika herzustellen. Das Optimierungsproblem für den Haushalt als Firma besteht darin, diejenige Kombination von n Gütern und Arbeit aus der Menge $\{(\mathbf{q}, l) \in \mathbb{R}_+^{n+1} : f(\mathbf{q}, l) = \mathbf{z}\}$ zu bestimmen, die seine Kosten unter der Nebenbedingung eines festgelegten Produktionsniveaus \mathbf{z} von Charakteristika minimiert. Formal ausgedrückt lautet dieses Optimierungsproblem:

$$C(\mathbf{p}, w, \mathbf{z}) := \underset{\mathbf{q} \geq 0, l \geq 0}{\text{Min}} \{ \mathbf{p}'\mathbf{q} + w l \mid \mathbf{f}(\mathbf{q}, l) = \mathbf{z} \}, \quad (2.2)$$

wobei die Funktion $\mathbf{f}(\mathbf{q}, l)$ eine K -dimensionale Produktionsfunktion mit konvexer Technologie ist. Damit existiert eine eindeutige Innere-Lösung des Optimierungsproblems (siehe MAS-COLELL (1995), S. 137 ff.). Die Produktionsfunktion gibt die maximalen Mengen von Charakteristika \mathbf{z} an, die mit den $n + 1$ Inputs (n Gütern und Arbeit) hergestellt werden können.

Bei der Lösung des Minimierungsproblems (2.2) ergeben sich die auf die K Charakteristika \mathbf{z} (Produktionsniveau) bedingten Nachfragekorrespondenzen $\mathbf{q}^*(\mathbf{p}, \mathbf{z})$ und $l^*(w, \mathbf{z})$ für die Inputfaktoren Güter und Arbeit. Nach dem Einsetzen dieser Nachfragekorrespondenzen in die Zielfunktion des Optimierungsproblems (2.2) ergibt sich der optimale Wert des Kostenminimierungsproblems, der durch die Kostenfunktion $C(\mathbf{p}, w, \mathbf{z}) = \mathbf{p}'\mathbf{q}^* + w l^*$ bezeichnet ist. Diese Funktion besitzt die folgenden Eigenschaften: Sie ist homogen vom Grade Eins in \mathbf{p} und w , (schwach) konkav in \mathbf{p} und w und nichtabnehmend in \mathbf{z} (für einen Beweis dieser Eigenschaften wird auf MAS-COLELL ET AL. (1995), Kapitel 5, bewiesen).

Unter der zusätzlichen Annahme der Monotonie der Produktionsfunktion $\mathbf{f}(\cdot)$ ist die Kostenfunktion $C(\mathbf{p}, w, \mathbf{z})$ differenzierbar¹. Somit werden die *kurzfristigen Grenzkosten* (KGK) durch

$$\beta_k(\mathbf{p}, w, \mathbf{z}) := \frac{\partial C}{\partial z_k}(\mathbf{p}, w, \mathbf{z}) \quad \forall k \quad (2.3)$$

definiert. Ihre Interpretation ist zentral für die hedonische Methode: Die KGK β_k geben an, um wieviel sich die Kosten ändern, falls eine zusätzliche Einheit der Charakteristik z_k hergestellt wird. Der für die hedonische Methode relevante Zusammenhang zwischen dem Preis eines auf dem Markt gehandelten Gutes, den Charakteristika des Gutes und ihren impliziten Preisen ist im Rahmen dieses Modells durch den folgenden Zusammenhang zwischen der weiter oben definierten Kostenfunktion $C(\mathbf{p}, w, \mathbf{z})$, den bei dem Haushalt hergestellten Charakteristika z_k und den kurzfristigen Grenzkosten

¹Im Fall von Technologien mit einem Einzeloutput ist die Differenzierbarkeit der Kostenfunktion durch den Lehrsatz von Shepard gewährleistet. Darüber hinaus kann man zeigen, daß dieser Lehrsatz auch für Technologien mit mehreren Outputs erfüllt ist (siehe Proposition 5.C.2, vi) in MAS-COLELL ET AL. (1995), S. 141).

β_k ($\forall k = 1, \dots, K$) gekennzeichnet:

$$C(\mathbf{p}, w, \mathbf{z}) = \sum_{k=1}^K \beta_k(\mathbf{p}, w, \mathbf{z}) \cdot z_k \quad (2.4)$$

Zum Beweis der Gleichung (2.4) wird von der Definition (2.3) ausgegangen. Dabei wird angenommen, daß die Produktionsfunktion homogen vom Grade Eins in Bezug auf \mathbf{q} und l sei (sie besitzt konstante Skalenerträge), daher ist die Kostenfunktion $C(\mathbf{p}, w, \mathbf{z})$ homogen vom Grade Eins in \mathbf{z} (siehe Proposition 5.C.2, viii) in MAS-COLELL ET AL. (1995), S. 141). Aufgrund der Eigenschaften (Differenzierbarkeit und Homogenität vom Grade Eins) der Kostenfunktion $C(\cdot)$ kann der Theorem von Euler angewendet werden (siehe Theorem M.B.2 in MAS-COLELL ET AL. (1995), S. 929). Daher kann die Kostenfunktion folgendermaßen dargestellt werden:

$$\sum_{k=1}^K \frac{\partial C}{\partial z_k}(\mathbf{p}, w, \mathbf{z}) \cdot z_k = 1 \cdot C(\mathbf{p}, w, \mathbf{z}),$$

oder wegen (2.3)

$$\sum_{k=1}^K \beta_k(\mathbf{p}, w, \mathbf{z}) \cdot z_k = C(\mathbf{p}, w, \mathbf{z}),$$

wobei angenommen wird, daß die Parameter β_k annähernd konstant sind. Diese letzte Gleichung entspricht gerade dem für die hedonische Methode relevanten Zusammenhang, der durch die Gleichung (2.4) gegeben ist.

Die anhand von den n Gütern und dem Faktor Arbeit hergestellten Charakteristika erbringen dem Haushalt einen direkten Nutzen. Zudem ist der Haushalt durch seine Budgetrestriktion $x = C(\mathbf{p}, w, \mathbf{z})$ beschränkt, d.h. unter der Annahme der Budgetausschöpfung gleichen seine Kosten das Gesamteinkommen x aus (optimale Lösung des vorigen Optimierungsproblems des Haushalts als Firma). Für den Haushalt als Konsument besteht das Optimierungsproblem darin, die Kombination von Charakteristika zu wählen, die unter der Nebenbedingung seiner Budgetrestriktion seinen Nutzen maximiert. Das Nutzenmaximierungsproblem des Haushalts lautet dann:

$$\begin{aligned}
& \max_{\mathbf{z} \geq \mathbf{0}} u(\mathbf{z}) \\
& \text{u.d.Nb. } x = C(\mathbf{p}, \omega, \mathbf{z}) = \sum_{k=1}^K \beta_k(\mathbf{p}, w, \mathbf{z}) \cdot z_k,
\end{aligned} \tag{2.5}$$

wobei $u(\cdot)$ eine ordinale Nutzenfunktion ist, die die Präferenzen des Haushalts darstellt (vgl. etwa MAS-COLELL ET AL. (1995), S. 9 für eine formale Definition für Nutzenfunktionen). In der Problemlösung von (2.5) ergeben sich die Nachfragefunktionen $z_k = z_k(x, \boldsymbol{\beta})$, mit $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$, für die Charakteristika. Die Existenz einer Lösung ist unter den Annahmen gewährleistet, daß die Nutzenfunktion stetig ist und die Menge von erreichbaren Kombinationen von Charakteristika abgeschlossen und beschränkt ist. Darüber hinaus ist die Lösung des Optimierungsproblems, unter der Annahme von streng konvexen Präferenzen des Konsumenten, eindeutig.

Die hedonische Hypothese wurde auch aufgrund eines anderen mikroökonomischen Modells hergeleitet, das gemäß DEATON/MÜLLBAUER (1980), S. 257 näher bei der Grundidee der hedonischen Hypothese liegt als das Standardmodell der Produktionstheorie des Haushalts. Es handelt sich um das lineare Modell für Charakteristika, das im nächsten Abschnitt beschrieben wird.

2.2.2.2 Das lineare Modell für Charakteristika

Das lineare Modell für Charakteristika (engl. “linear characteristics model”) basiert auf den Arbeiten von STIGLER (1945) zur Analyse von Ernährungsdiäten und auf der von DANTZIG (1963) entwickelten Simplex-Methode der linearen Programmierung. LANCASTER (1966) und GORMAN (1956) haben diese Ideen in die Konsumtheorie hereingebracht.

Dieses Modell hat als Basis eine bestimmte Klasse von Produktionsfunktionen. Im allgemeinen gibt eine Produktionsfunktion $f(q_1, \dots, q_n)$ die Pareto-effizienten Kombinationen von Outputs z_1, \dots, z_K an, die durch die Inputfaktoren q_1, \dots, q_n erzeugt werden können. Beim linearen Modell für Charakteristika wird von einer bestimmten Klasse von Produktionsfunktionen ausgegangen: Die K Charakteristika \mathbf{z} ergeben sich

aus einer linearen Transformation der auf dem Markt gehandelten n Güter, nämlich

$$\mathbf{z} = \mathbf{Z}\mathbf{q}, \quad (2.6)$$

mit $\mathbf{z} = (z_1, \dots, z_K)'$ und $\mathbf{q} = (q_1, \dots, q_n)'$. Das (k, i) -te Element z_{ki} der Transformationsmatrix \mathbf{Z} , $k = 1, \dots, K$ und $i = 1, \dots, n$, (bei LANCASTER (1991), S. 17 auch Verbrauchstechnologie genannt) mißt die Menge des k -ten Charakteristikums in einer Einheit des i -ten gehandelten Gutes. Die Addition der in jedem Gut enthaltenen Mengen der k -ten Charakteristik gibt die Gesamtmenge dieses Charakteristikums, das alle n auf dem Markt gehandelten Güter liefern, nämlich

$$z_k = \sum_{i=1}^n z_{ki} q_i. \quad (2.7)$$

Dies schließt nur den folgenden Fall ein: “the act of consumption (that is, of extracting characteristics from goods) is linearly combinable, in the sense that the characteristics obtained from any collection of goods is the sum of the characteristics contained in the specified quantities of the individual goods” (LANCASTER (1991), S. 105).

Ähnlich zum Standardmodell der Produktionstheorie des Haushalts ist das Ziel des Haushalts die Maximierung seines Nutzens unter der Nebenbedingung, daß seine Kosten $\mathbf{p}'\mathbf{q}$ sein Gesamteinkommen x ausschöpfen (unter der Annahme der Budgetausschöpfung), wobei \mathbf{q} der n -dimensionale Vektor von auf dem Markt gehandelten Inputfaktoren ist, deren Preise durch den Vektor $\mathbf{p} = (p_1, \dots, p_n)'$ gekennzeichnet sind. Bei gegebener, in (2.6) definierter Produktionsfunktion, ist das Optimierungsproblem für den Haushalt durch

$$\begin{aligned} \max_{\mathbf{q} \geq 0} u(\mathbf{z}) &= \max u(\mathbf{Z}\mathbf{q}) \\ \text{u.d.NB. } x &= \mathbf{p}'\mathbf{q} \end{aligned} \quad (2.8)$$

gekennzeichnet. Die Lagrange-Funktion des Maximierungsproblems kann folgendermaßen geschrieben werden:

$$\mathcal{L} = u(\mathbf{Z}\mathbf{q}) + \lambda(x - \mathbf{p}'\mathbf{q}) \quad (2.9)$$

Die optimale Lösung des Problems ist durch

$$\begin{aligned}
\sum_{k=1}^K \frac{\partial u}{\partial z_k} \frac{\partial z_k}{\partial q_1} &= \lambda p_1, \\
\sum_{k=1}^K \frac{\partial u}{\partial z_k} \frac{\partial z_k}{\partial q_2} &= \lambda p_2, \\
&\vdots \\
\sum_{k=1}^K \frac{\partial u}{\partial z_k} \frac{\partial z_k}{\partial q_n} &= \lambda p_n,
\end{aligned}$$

charakterisiert, wobei der Multiplikator der Lagrange-Funktion λ als der marginale Nutzen des Einkommens interpretiert werden kann. Aus der Produktionsfunktion (2.7) folgt $\frac{\partial z_k}{\partial q_i^*} = z_{ki}$ und nach der Division durch λ ($\lambda \neq 0$) auf beiden Seiten der Gleichungen ergibt sich

$$\sum_{k=1}^K \frac{\partial u}{\partial z_k} \frac{1}{\lambda} z_{ki} = p_i \quad \forall i = 1, \dots, n,$$

wobei $1/\lambda$ als die “marginalen Kosten des Nutzens” interpretiert werden kann² (DEATON/MÜLLBAUER (1980), S. 250). Die Multiplikation der marginalen Kosten des Nutzens mit dem marginalen Nutzen des k -ten Charakteristikums ergibt die marginalen Kosten dieses Charakteristikums. Dies ist gerade die Definition der Parameter β_k , $k = 1, \dots, K$, die in der Produktionstheorie des Haushalts weiter oben präsentiert wurde. Dies ergibt den folgenden Zusammenhang:

$$p_i = \sum_{k=1}^K \beta_k z_{ki}, \tag{2.10}$$

mit $\beta_k := \frac{\partial u}{\partial z_k} \frac{1}{\lambda}$, in dem der Preis der i -ten Variante des gehandelten Gutes in den KKG

²Die Bezeichnung von $1/\lambda$ als die marginalen Kosten des Nutzens kann mit Hilfe des Umhüllungstheorems ersehen werden. Im Rahmen eines Optimierungsproblems behandelt das Umhüllungstheorem die Effekte im Optimum von Parameteränderungen auf die Werte einer Zielfunktion. Betrachtet wird die “Value”-Funktion (2.9) $\mathcal{L}^*(x) = u(\mathbf{Z}\mathbf{q}^*(x)) + \lambda^*(x)(x - \mathbf{p}'\mathbf{q}^*(x))$. Darüber hinaus ist die Budgetrestriktion im Optimum bindend, damit ist $\lambda^*(x) > 0$. Anhand des Umhüllungstheorems kann man schreiben: $\frac{\partial \mathcal{L}^*(x)}{\partial x} = \frac{\partial u^*(x)}{\partial x} = \lambda^*(x)$. Daher ergibt sich λ als der marginale Nutzen der Kosten. Somit kann die Inverse $\frac{1}{\lambda}$ als der marginale Kosten des Nutzens interpretiert werden.

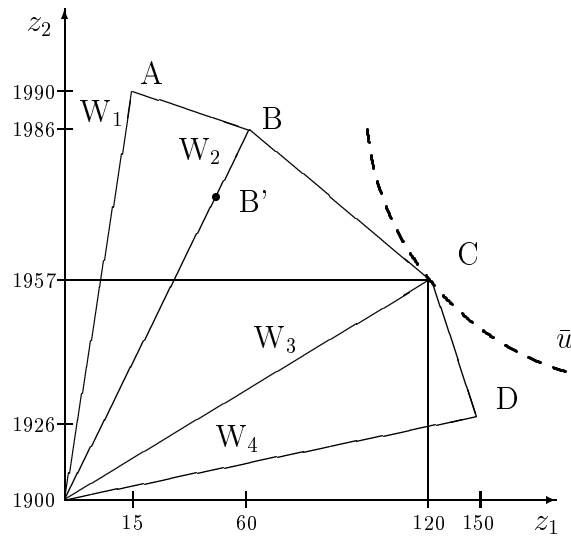


Abbildung 2.1. Das lineare Modell für Charakteristika: Beispiel aus dem Wohnungsmarkt.

(oder “impliziten” Preisen unter der hedonischen Methodologie) der Charakteristika zerlegt ist.

Zum Schluß wird das lineare Modell für Charakteristika graphisch dargestellt. Abbildung 2.1 zeigt ein Beispiel für den Wohnungsmarkt. Betrachtet werden zwei Charakteristika: $z_1 :=$ Wohnfläche der Wohnung in qm (je größer eine Wohnung ist, desto besser) und $z_2 :=$ Baujahr (je neuer eine Wohnung ist, desto höher ist der Nutzen, den sie stiftet). Die auf dem Markt gehandelten Güter, die der Haushalt in seiner Nutzenfunktion berücksichtigt, sind vier Varianten des Einzelgutes Wohnung: W_1 (durch eine kleine Wohnfläche gekennzeichnet), W_2 , W_3 und W_4 (durch ein frühes Baujahr charakterisiert). Sei z_{ki} , $k = 1, 2$ und $i = 1, \dots, 4$, die Menge des k -ten Charakteristikums in der i -ten Wohnung. Jede Variante W_i , $i = 1, \dots, 4$ wird bei den Kombinationen von Charakteristika z_1 und z_2 definiert, die durch den linearen Zusammenhang $z_1 = (\frac{z_{1i}}{z_{2i}})z_2$ gekennzeichnet sind. Dieser lineare Zusammenhang zwischen den Charakteristika ergibt sich aus der besonderen Produktionsfunktion, die dem Modell zugrundeliegt (siehe Gleichung (2.6)). In diesem Beispiel lautet die lineare Transformation der auf dem Markt gehandelten Wohnungen $\mathbf{z} = \mathbf{Z}\mathbf{W}$, mit $\mathbf{W} = (W_1, W_2, W_3, W_4)'$. Das (k, i) -te Element der Transformationmatrix \mathbf{Z} ist z_{ki} , die Menge des k -ten Charakteristikums in

der i -ten Wohnung. Explizit lautet die lineare Transformation (2.6) in diesem einfachen Spezialfall:

$$\begin{aligned} z_1 &= z_{11}W_1 + z_{12}W_2 + z_{13}W_3 + z_{14}W_4 \\ z_2 &= z_{21}W_1 + z_{22}W_2 + z_{23}W_3 + z_{24}W_4. \end{aligned}$$

Betrachtet man nur die Wohnungsvariante W_1 , sind die Charakteristika im Verhältnis z_{11}/z_{21} bestimmt, d.h. $z_1 = (\frac{z_{11}}{z_{21}})z_2$. Die Gerade W_1 in der Abbildung 2.1 stellt diesen linearen Zusammenhang dar. Die übrigen Geraden W_2 bis zu W_4 kennzeichnen die Zusammenhänge $z_1 = (\frac{z_{12}}{z_{22}})z_2$, $z_1 = (\frac{z_{13}}{z_{23}})z_2$ und $z_1 = (\frac{z_{14}}{z_{24}})z_2$.

Die Fläche unter dem Streckenzug \overline{ABCD} besteht aus denjenigen Kombinationen von Charakteristika, die unter einer gegebenen Budgetrestriktion für den Haushalt erreichbar sind. Die Kurve \overline{ABCD} , auch “Charakteristika-Grenze” genannt (engl. “characteristics frontier”), bezeichnet die effizienten Kombinationen von Charakteristika. Darunter versteht man erwünschte Kombinationen von Charakteristika, die mit minimalen Kosten unter einer gegebenen Budgetrestriktion erreichbar sind. Man kann zeigen, daß die Charakteristika-Grenze (schwach) konkav sein kann und wird nur anhand von Änderungen des Einkommens linear und proportional verschoben. Die spezielle Form der Charakteristika-Grenze in Abbildung 2.1 entspricht einer realistischeren Kurve mit verschiedenen Steigungen, wobei die relativen Preise (die die verschiedenen Steigungen bestimmen) der Charakteristika je nach Wert des Definitionsraums der Charakteristika variieren. Nur Preisänderungen der Charakteristika ändern die Form der Charakteristika-Grenze (siehe LANCASTER (1991), S. 20, und DEATON/MUELLBAUER (1980) für zusätzliche Beispiele solcher Funktionen im Rahmen der Konsum- und Produktionstheorien).

Die Punkte A, B, C und D der Charakteristika-Grenze sind Ecklösungen, bei denen das ganze Budget nur für eine Wohnungsvariante ausgegeben wird. Wenn die Lösung des Optimierungsproblems des linearen Modells für Charakteristika der Punkt C ist, wo die Indifferenzkurve \bar{u} tangential die Charakteristika-Grenze berührt, wird das ganze Budget nur für die Wohnungsvariante W_3 zum Preis $p_3 = \beta_1 z_{13} + \beta_2 z_{23}$ ausgegeben

und zwar mit den Mengen von Charakteristika: 1957 für z_2 und 120 m^2 für z_1 zu den impliziten Preisen β_1 bzw. β_2 . Falls der Preis der Wohnungsvariante W_2 steigt und somit der Punkt B nach innen auf B' verschoben wird (jetzt sind unter derselben Budgetrestriktion kleinere Mengen von Charakteristika erreichbar), dann wäre diese Wohnungsvariante nicht mehr effizient.

2.3 Implikationen der hedonischen Hypothese für Regressionsmietspiegel

In diesem Abschnitt werden die Implikationen der hedonischen Hypothese für Regressionsmietspiegel dargestellt. Einige dieser Implikationen haben auch Konsequenzen für die Schätzung der impliziten Preise, die ausführlich im statistischen Teil der Arbeit behandelt wird.

Eine explizite Spezifikation der in (2.1) definiert hedonischen Funktion p wirft zwei Fragen auf: Die erste betrifft die Auswahl der funktionalen Form der hedonischen Preisfunktion. Dabei stellt sich insbesondere die Frage, ob die Spezifikation linear oder nichtlinear sein sollte.

Eine zweite Entscheidungsfrage bezieht sich auf die Selektion von erklärenden Variablen. Die umfaßt die Fragen nach der ökonomischen Relevanz der Variablen, welche die Charakteristika kennzeichnen, und nach der Anzahl von explikativen Variablen, die im Modell berücksichtigt werden sollen.

Funktionale Form des hedonischen Preisansatzes

Die erste direkte Implikation der hedonischen Theorie, die wir hier betrachten werden, betrifft die funktionale Form für den Zusammenhang zwischen dem Preis eines Gutes und seinen Charakteristika, die die hedonische Hypothese unterstellt. Der Zusammenhang zwischen den Charakteristika und den Gütern muß keineswegs linear sein. Wie in Gleichung (2.6) spezifiziert, ist der lineare Ansatz, der auch von Lanca-

ster verwendet wurde, lediglich zur Vereinfachung der Darstellung des Optimierungsproblems (2.8) gedacht. Laut Lancaster können auch allgemeinere Spezifikationen für die Zusammenhänge zwischen den Charakteristika und den Gütern betrachtet werden (siehe LANCASTER (1966), S. 135 und die dort vorgeschlagene Literatur über die Problematik und Lösung von solchen allgemeinen Modellen).

Eine spätere Arbeit von ROSEN (1974) im Kontext mehrfacher Gleichgewichte macht ähnliche Aussagen für unteilbare Güter: Bei einem angenommenen Kontinuum unterscheidbarer Güter und vollkommener Konkurrenz auf den Märkten ist die Funktion $\mathbf{p} = f(\mathbf{z})$ zwischen den Preisen der Güter und den Charakteristika linear. Im Fall von segmentierten Märkten muß die Funktion $f(\cdot)$ nicht unbedingt linear in \mathbf{z} sein (siehe ROSEN (1974), S. 37). Nichtlineare Ansätze können den Zusammenhang zwischen dem Preis und den Charakteristika besser beschreiben. Beim Immobilienmarkt liegt eine Segmentierung des Marktes vor (siehe etwa DALE-JOHNSON (1982)) und somit sollten im allgemeinen nichtlineare Spezifikationen betrachtet werden.

Auch wenn lineare Ansätze in der Empirie gelegentlich verwendet wurden (siehe etwa ARGUEA/HSIAO (1993)), haben sich im Immobilienbereich nichtlineare Spezifikationen als geeigneter erwiesen und durchgesetzt (siehe etwa GOODMAN (1978) und MORESI (1989)).

Auswahl der Charakteristika

Die in Abschnitt 2.2, S. 23, vorgestellten Definitionen von Charakteristika von Lancaster und Triplett helfen in der Praxis nicht viel, wenn entschieden werden muß, welche Charakteristika berücksichtigt werden sollen. Aus diesen Definitionen folgt, daß alle *ökonomisch relevante Charakteristika* berücksichtigt werden sollen. Auch aus statistischer Sicht sollten alle relevanten Charakteristika in die Schätzung einbezogen werden. Die Nichtberücksichtigung von relevanten Charakteristika führt im allgemeinen zu verzerrten Schätzern der impliziten Preise für die betrachteten Variablen, d.h. daß der Erwartungswert der Schätzer nicht den wahren Werten entspricht. Die Konsequenzen auf die geschätzten impliziten Preise können gravierend sein. Man betrachte z.B. das vereinfachte Beispiel, in dem der Mietpreis von der Größe des Gartens abhängig ist.

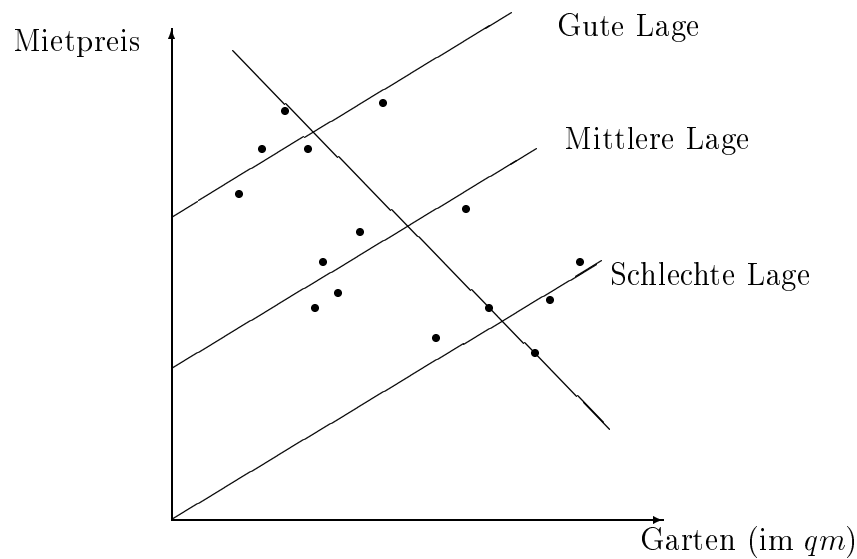


Abbildung 2.2. Beispiel einer vergessenen erklärenden Variable.

Hier ist zu erwarten, daß ein größerer Garten einen höheren Mietpreis bewirkt. Da solche Ausstattungen, wie ein privater Garten, im allgemeinen in guten Stadtvierteln zu finden sind, könnte das Weglassen einer relevanten Variable, die die Lage charakterisiert, einen negativen Einfluß auf den geschätzten impliziten Preis der Variable Garten ausüben. Dieser negative Effekt läßt sich graphisch durch Abbildung 2.2 erläutern. Ohne die Berücksichtigung der Variable Lage würde der geschätzte Koeffizient der Variable Garten ein negatives Vorzeichen aufweisen. Der Einfluß dieser Variable sollte jedoch positiv sein. Wenn die Variable Lage berücksichtigt wird, wird der Koeffizient der Variable Garten richtig geschätzt. Dieses Thema soll aber hier nicht ausführlicher betrachtet werden. Für eine Einführung in die Problematik der Nichtberücksichtigung relevanter Variablen siehe etwa GREENE (1993), Kapitel 8.

Im Zusammenhang mit der Auswahl von Charakteristika sollte eine weitere Anforderung erwähnt werden: Nur *objektiv gemessene Charakteristika*, deren Bewertung nicht von der Präferenzordnung der Konsumenten abhängig sind, sollten bei der Erstellung von Mietspiegeln berücksichtigt werden. In anderen Worten: Es sollten nur Charakteristika berücksichtigt werden, deren Inhalt von jedem Konsument gleich "beurteilt" werden kann.

Die direkten Implikationen der hedonischen Theorie, die Modell-Spezifikation und die Auswahl der Variablen haben wichtige Auswirkungen auf die Ergebnisse eines Mietspiegels. Trotzdem wurde diese zeitintensive Arbeit bei früheren Mietspiegeln sehr oft vernachlässigt, obwohl bekannte statistische Techniken zur Verfügung stehen, um diese Probleme adäquat zu berücksichtigen (siehe Kritiken etwa in KRÄMER (1992a, 1992b, 1993) oder RONNING (1993)).

Teil II

Semiparametrischer Modellansatz für Regressionsmietspiegel

Kapitel 3

Modellansätze für Regressionsmietspiegel: ein semiparametrisches Regressionsmodell

3.1 Einleitung

In der Praxis des Regressionsmietspiegels wurden verschiedene Modellansätze angewendet. Dieses Kapitels präsentiert zuerst drei Modellansätze im Abschnitt 3.2, die häufig den in Deutschland erstellten Mietspiegeln zugrunde liegen. Die verschiedenen Modellansätze unterscheiden sich insbesondere in der Spezifikation des Zusammenhangs zwischen dem Mietpreis und der Wohnfläche.

Zuerst wird das lineare Modell vorgestellt, das ursprünglich bei der früheren Regressionsmietspiegeln angewendet wurde. Nach methodologischer Kritik an dem linearen Modell wurden zwei allgemeinere Modelle in der Literatur vorgeschlagen: das Modell von Oberhofer, auch als Regensburger Modell bekannt, und das Modell von Fahrmeir. Diese beiden Modelle wurden in Deutschland positiv beurteilt und ihre Anwendung hat

sich verbreitet. Das Modell von Fahrmeir ist durch seine Flexibilität charakterisiert, die u.a. eine mögliche Nichtlinearität zwischen den Variablen des Modells - insbesondere zwischen dem Mietpreis und der Wohnfläche - spezifizieren kann. Die Nichtberücksichtigung der Nichtlinearität im Ansatz für den Mietpreis wurde in früheren Mietspiegeln kritisiert (siehe etwa KRÄMER (1992b) S. 14 ff. und (1993) S. 16). Im Gegensatz zu anderen Mietspiegeln wie z.B. für Frankfurt und München (vgl. etwa OBERHOFER (1994) und LANDESHAUPTSTADT MÜNCHEN (1994)) ist der Zusammenhang zwischen dem Mietpreis und der Wohnfläche für Darmstadt (1994/95) eher linear. Somit zeigt sich, daß die Form dieses Zusammenhangs vom konkreten Fall abhängig ist und es keine allgemein gültige explizite Spezifikation für den Zusammenhang zwischen der Nettomiete und der Wohnfläche vorgeschlagen werden kann.

Anschließend wird ein neuer Ansatz zur Modellierung von Mietpreisen vorgeschlagen. Das sogenannte semiparametrische Regressionsmodell ist durch eine erweiterte Flexibilität bei der Modellierung des Zusammenhangs zwischen dem Mietpreis und der Wohnfläche gekennzeichnet. Dieses Modell wird im Abschnitt 3.3 vorgestellt. Auf seiner Schätzung anhand eines bestimmten Datensatzes wird in einem weiteren Kapitel eingegangen.

Zum Schluß und zu einem besseren Verständnis unseres vorgeschlagenen semiparametrischen Regressionsmodells wird die Grundlage semiparametrischer Regressionsmodelle eingeführt. Zur Vereinfachung wird dabei nur der Spezialfall eines rein additiven Modells betrachtet. Auf das konkrete Thema der Schätzung solcher semiparametrischen Regressionsmodelle wird im nächsten Kapitel eingegangen.

3.2 Modellansätze für Regressionsmietspiegel

In diesem Abschnitt werden drei bekannte Ansätze zur Spezifikation des Zusammenhangs zwischen dem Mietpreis und den Charakteristika von Wohnungen vorgestellt.

Bei früheren Regressionsmietspiegeln, wie z.B. demjenigen von der Stadt Kiel 1992,

wurde davon ausgegangen, daß zwischen der abhängigen Variablen Mietpreis und den unabhängigen Variablen, den Mietpreisdeterminanten, ein linearer Zusammenhang besteht. Dieser Zusammenhang läßt sich durch das folgende Regressionsmodell

$$M = \alpha + \delta \cdot WFL + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_k Z_k + \varepsilon, \quad (3.1)$$

beschreiben, wobei M den Mietpreis kennzeichnet. Darüber hinaus bezeichnen WFL die Variable Wohnfläche und Z_1, Z_2, \dots, Z_k weitere k Mietpreisdeterminanten. Die Parameter α , δ und β_1 bis β_k sind die Koeffizienten des Modells und der Term ε ist eine zufällige Störvariable.

In empirischen Untersuchungen zur Modellierung des Mietpreises hat sich gezeigt, daß das lineare Modell nur bedingt geeignet zur Erstellung von Mietspiegeln ist. Daher wurde seine Anwendung in Mietspiegeln, wie beim oben erwähnten Mietspiegel von Kiel' 92 oder dem Münchner Mietspiegel (1991) wegen statistisch-methodischer Mängel für unbrauchbar erklärt (vgl. etwa KRÄMER (1993)). Die Kritikpunkte für die Ablehnung von Mietspiegeln, die auf dem linearen Modells basierten, waren u.a. sowohl die Vernachlässigung von erklärenden Variablen, die relevant für die Bestimmung des Mietpreises waren, als auch die Nichtberücksichtigung von Interaktionen zwischen Einflußfaktoren.

Das nächste verwendete Modell, ein sogenanntes additiv-multiplikatives Modell, wurde zum erstenmal von Aigner, Oberhofer und Schmidt (vgl. AIGNER/OBERHOFER/SCHMIDT (1993)) bei der Erstellung des Regensburger Mietspiegels (1992) verwendet (deswegen auch als "Regensburger Modell" bekannt).

Das Modell basiert auf der empirischen Feststellung, daß die Wohnfläche die wichtigste Einflußgröße der Miete ist (vgl. etwa AIGNER/OBERHOFER/SCHMIDT (1993), S. 17) und alle anderen Mietpreisdeterminanten von der Wohnfläche abhängig sind. Diese werden deshalb als Zu- und Abschläge zur Grundmiete, die nur von der Wohnfläche abhängig ist, modelliert.

Die im Modell betrachtete Definition von Miete ist die "Nettomiete", d.h. die monat-

liche Miete ohne Nebenkosten (Heizung/Warmwasser, Aufzug, Grundsteuer, Müllabfuhr,...) und ohne Mietzinszuschläge wie etwa für eine Garage. Die Nettomiete (NM) ergibt sich aus der Multiplikation zweier Faktoren: der Grundmiete (GM) und dem Nettomietfaktor (NMF). Die Grund- oder Basismiete ist der Teil der Nettomiete, die nur von der Wohnfläche abhängig ist. Dieser Zusammenhang zwischen Nettomiete und Wohnfläche wurde im Regensburger Modell additiv beschrieben. Der sogenannte Nettomietfaktor (NMF) ist der Teil der Nettomiete, der von allen Mietpreis-determinanten außer der Wohnfläche abhängt. Die Spezifikation für den Nettomietfaktor ist ebenfalls additiv. Das Regensburger Modell lautet also:

$$NM = \underbrace{(\alpha + \delta \cdot WFL)}_{GM} \cdot \underbrace{(1 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_k Z_k)}_{NMF} + \varepsilon \quad (3.2)$$

wobei WFL die Wohnfläche kennzeichnet und Z_1, Z_2, \dots, Z_k die k übrigen Mietpreis-determinanten. Der Term ε bezeichnet die Störvariable des Modells.

Mit diesem Modell werden Schwächen des linearen Modells (3.1) überwunden, wie z.B. die fehlende Abhängigkeit zwischen den Einflußfaktoren. Dies wird im Regensburger Model durch die Interaktionen zwischen der Wohnfläche und den restlichen Regressoren vermieden und somit als eine Verbesserung angesehen (vgl. RONNING (1995)).

Beim Münchner Mietspiegel '94 (vgl. FAHRMEIR ET AL. (1994)) wurde der ursprüngliche Ansatz (3.2) modifiziert, um eine nichtlineare Spezifikation des Zusammenhangs zwischen der Nettomiete und der Wohnfläche zu ermöglichen. Der Grund dafür war, eine verbesserte Anpassung der Spezifikation an die Daten zu ermöglichen, da die niedrigen Mieten überschätzt wurden, falls ein linearer Ansatz für die Grundmiete verwendet wurde.

Das Modell von Fahrmeir lautet:

$$NM = \underbrace{(\alpha + f(WFL))}_{BNM} \cdot \underbrace{(1 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_k Z_k)}_{NMF} + \varepsilon \quad (3.3)$$

wobei $f(WFL)$ eine nichtlineare Funktion der Wohnfläche ist. Die Variablen Z_1, Z_2, \dots, Z_k bezeichnen k weitere Mietpreisdeterminanten und der Term ε ist die Störvariable des Modells. Für die Daten des Münchner Mietspiegels '94 erwies sich eine quadratische-lineare Spezifikation als geeignetste (siehe FAHRMEIR ET AL. (1994), S. 7 ff. für Details). Die Grundmiete wird bei Fahrmeir mit *BNM* Basis-Nettomiete bezeichnet. Der wichtige Beitrag von Fahrmeir besteht darin, ein nichtlineares Modell, welches den Zusammenhang zwischen der Nettomiete und der Wohnfläche beschreibt, zu spezifizieren und zu schätzen.

3.3 Ein semiparametrischer Modellansatz für Regressionsmietspiegel

Das Modell von Fahrmeir ist bezüglich der Spezifikation des Zusammenhangs zwischen Nettomiete und Wohnfläche flexibler als das Modell von Oberhofer.

Aufgrund der in der Praxis gezeigten Vielfältigkeit des Zusammenhangs zwischen dem Mietpreis und der Wohnfläche schlagen wir in dieser Arbeit ein viertes Modell zur Beschreibung des Zusammenhangs zwischen dem Mietpreis, der Wohnfläche und den restlichen Mietpreisdeterminanten vor, das einen noch höheren Flexibilitätsgrad in der Spezifikation der Grundmiete aufweist. Im Gegensatz zu Fahrmeir verwenden wir eine nichtparametrische Spezifikation für die Grundmiete. Das semiparametrische Regressionsmodell stellt eine Verallgemeinerung aller bekannten Modelle dar, d.h. alle anderen Modelle (lineares Modell, Modelle von Oberhofer und von Fahrmeir) können als Spezialfälle interpretiert werden.

Analog zu den Modellen von Oberhofer und Fahrmeir berücksichtigt dieses Modell explizit die Interaktionen zwischen der Wohnfläche und den restlichen Mietpreisdeterminanten anhand eines multiplikativen Ansatzes. Die funktionale Form der Grundmiete wird aber nicht explizit spezifiziert, d.h. wir schlagen das Modell

$$NM = \underbrace{q(WFL)}_{GM} \cdot \underbrace{(1 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_k Z_k)}_{NMF} + \varepsilon \quad (3.4)$$

vor, wobei $q(WFL)$ eine beliebig Funktion der Wohnfläche bezeichnet. Die Variablen Z_1, Z_2, \dots, Z_k bezeichnen k weitere Mietpreisdeterminanten. Der Term ε ist die Störvariable des Modells, der alle unsystematischen Einflüsse beinhaltet. Dieser Modellansatz besteht aus zwei Komponenten: Die Grundmiete wird nichtparametrisch modelliert aber der Zusammenhang zwischen der Nettomiete und den üblichen Regressoren (Nettomietfaktor) bleibt linear. Die zwei Komponenten des Modells sind getrennt, und ihr Zusammenhang ist multiplikativ. Ein derartiges additiv-multiplikatives semiparametrisches Modell bildet eine Erweiterung des bekannten additiven semiparametrischen Modells von Fahrmeir. Bevor die Schätzung des semiparametrischen Modells (3.4) behandelt wird, werden wir den Spezialfall des additiven Ansatzes in einer Einführung in die Theorie der semiparametrischen Regressionsmodelle betrachten.

3.4 Semiparametrische Regressionsmodelle: eine Einführung

In diesem Abschnitt werden die sogenannten *semiparametrischen Regressionsmodelle* vorgestellt. Darunter versteht man Modelle, die parametrische und nichtparametrische Komponenten umfassen. Man spricht von der additiven semiparametrischen Regression mit erklärenden Variablen $X_1, \dots, X_p, Z_1, \dots, Z_k$ und Beobachtungen $y_i, x_{1i}, \dots, x_{pi}, z_{1i}, \dots, z_{ki}$ wenn ein Modell der folgenden Form vorliegt

$$Y_i = q(x_{1i}, \dots, x_{pi}) + \mathbf{z}_i \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (3.5)$$

wobei Y_i die i -te Beobachtung der erklärten Variablen ist. Jeder Fehlerterm ε_i ist unabhängig und identisch verteilt. Darüber hinaus gilt für ε_i :

$$E(\varepsilon_i | X_1 = x_{1i}, \dots, X_p = x_{pi}, Z_1 = z_{1i}, \dots, Z_k = z_{ki}) = 0 \quad \forall i = 1, \dots, n.$$

Der Einfluß der systematischen erklärenden Variablen ist anhand zweier unterschiedlicher Spezifikationen modelliert worden: Eine nichtparametrische Komponente besteht in einer unbekannten reellen und stetigen Funktion $q(x_{1i}, \dots, x_{pi})$ von p erklärenden Variablen X_1, \dots, X_p . Die parametrische Komponente umfaßt anhand einer linearen Spezifikation den Einfluß der übrigen erklärenden Variablen. Die i -te Beobachtung der übrigen k erklärenden Variablen ist durch den $(1 \times k)$ -Vektor $\mathbf{z}_i = (z_{1i}, \dots, z_{ki})$ gekennzeichnet. Der Parametervektor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ ist ein $(k \times 1)$ -Vektor von unbekannten Regressionsparametern. Die parametrische Komponente ist in diesen Parametern linear. Deshalb wird dieses Modell auch *partielles lineares Modell* genannt (vgl. HÄRDLE (1990), S. 232).

Das Modell (3.5) kann umgeformt werden, damit es nicht mehr direkt von der unbekannten Funktion $q(\mathbf{X})$ abhängt. Gegeben $\mathbf{Y} = (y_1, \dots, y_n)'$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$, wobei $\mathbf{x}_i = (x_{1i}, \dots, x_{pi})$ für $i = 1, \dots, n$, $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)'$, mit $\mathbf{z}_i = (z_{1i}, \dots, z_{ki})$, lautet der auf \mathbf{X} bedingte Erwartungswert von \mathbf{Y} :

$$E(\mathbf{Y} | \mathbf{X} = \mathbf{x}) = q(\mathbf{X}) + E(\mathbf{Z} | \mathbf{X} = \mathbf{x})\boldsymbol{\beta},$$

wobei der Term $E(\mathbf{Z} | \mathbf{X} = \mathbf{x})$ den bedingten Erwartungswerte der erklärenden Variablen \mathbf{Z} gegeben $\mathbf{X} = \mathbf{x}$ bezeichnet. Die Subtraktion des auf \mathbf{X} bedingten Erwartungswerts von \mathbf{Y} auf beiden Seiten der Gleichung (3.5) ergibt:

$$\mathbf{Y} - E(\mathbf{Y} | \mathbf{X} = \mathbf{x}) = (\mathbf{Z} - E(\mathbf{Z} | \mathbf{X} = \mathbf{x}))\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (3.6)$$

Der nächste Schritt betrifft die Schätzung solcher semiparametrischen Regressionsmodelle wie (3.6). Dafür werden zuerst die beiden bedingten Erwartungswerte geschätzt. Dabei eignen sich nichtparametrische Schätztechniken, die insbesondere zur Schätzung von Dichtefunktionen verbreitet sind. Nach dem Einsetzen der Schätzer für $E(\mathbf{Y} | \mathbf{X} =$

\mathbf{x}) und für $E(\mathbf{Z}|\mathbf{X} = \mathbf{x})$ in Gleichung (3.6) kann der Parametervektor $\boldsymbol{\beta}$ z.B. durch die Kleinste-Quadrate-Methode geschätzt werden. Dieses Schätzproblem semiparametrischer Regressionsmodelle wird im nächsten Kapitel näher betrachtet.

Kapitel 4

Methoden zur Schätzung von additiven semiparametrischen Regressionsmodellen

4.1 Einleitung

Zu einem besseren Verständnis der Schätzung semiparametrischer Regressionsmodelle im allgemeinen werden die Grundlagen der nichtparametrischen Kernschätzung behandelt. Darüber hinaus werden die Eigenschaften des nichtparametrischen Nadaraya-Watson-Kernschätzers diskutiert und verschiedene Konfidenzintervalle vorgeschlagen. Dabei werden u.a. Simulationsmethoden zur Schätzung von Konfidenzintervallen beschrieben. In Kapitel 5 werden diese Methoden zur Schätzung des im letzten Kapitel vorgeschlagenen semiparametrischen Modellansatzes angewendet.

Zunächst wird die Schätzung von semiparametrischen additiven Modellen betrachtet. Darüber hinaus werden die asymptotischen Eigenschaften des semiparametrischen Schätzers diskutiert. Diese Beschreibung der nicht- und semiparametrischen Schätztechniken aufgrund eines additiven Modells bildet die Basis für die kompliziertere Schätzung des additiv-multiplikativen Regressionsmodells. Dies folgt im nächsten Ka-

pitel.

4.2 Nichtparametrische Kernschätzung

4.2.1 Nichtparametrische Schätzung von Dichtefunktionen

Die bedingten Erwartungswerte $E(\mathbf{Y}|\mathbf{X} = \mathbf{x})$ und $E(\mathbf{Z}|\mathbf{X} = \mathbf{x})$ in Gleichung (3.6) können durch eine nichtparametrische Methode geschätzt werden. Aus der Gruppe der nichtparametrischen Methoden zur Dichteschätzung wird üblicherweise der Kernschätzer verwendet (siehe SILVERMAN (1986), S. 34).

Sei $E(Y|X = x)$ eine beliebige bedingte Erwartungswertfunktion, in der (Y, X) ein zweidimensionaler stetiger Zufallsvektor ist, mit den Realisierungen (y_i, x_i) , $i = 1, \dots, n$. Diese bedingte Erwartungswertfunktion lautet

$$m(x) := E(Y|X = x) = \int y f(y|x) dy = \frac{\int y f(x, y) dy}{f(x)}, \quad (4.1)$$

wobei $f(y|x)$ die bedingte Dichtefunktion von Y bei gegebenem $X = x$ ist, $f(x, y)$ die gemeinsame Dichtefunktion der Variablen X und Y und $f(x)$ die Randdichtefunktion der Variable X ist. Die Funktion $m(x)$ soll stetig und differenzierbar sein.

Um die bedingte Erwartungswertfunktion zu schätzen, müssen Schätzer für die unbekannte gemeinsame Dichtefunktion und die unbekannten Randdichtefunktionen ermittelt werden. Dafür können sogenannte Kernschätzer benutzt werden.

Bei gegebenen n Realisierungen x_1, \dots, x_n der stetigen Zufallsvariable X , ist der *Kern-dichteschätzer* für $f(x)$ mit Kernfunktion K im univariaten Fall durch

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R} \quad (4.2)$$

definiert, wobei n der Stichprobenumfang ist. Der Term $K(\frac{x - x_i}{h})$ wird häufig auch

mit $K_h(x - x_i)$ bezeichnet. Der Parameter h wird *Bandbreite* genannt und bestimmt die Breite des Intervalls um den Punkt x herum, in dem die sogenannte Kernfunktion K bewertet wird.

Für den multivariaten Fall ist der Kerndichteschätzer folgendermaßen definiert:

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^p} \sum_{i=1}^n K \left(\frac{(\mathbf{x} - \mathbf{x}_i) \mathbf{S}^{-1} (\mathbf{x} - \mathbf{x}_i)'}{h^2} \right), \quad (4.3)$$

wobei \mathbf{x} ein $1 \times p$ -Vektor von Werten der Variablen X_1, \dots, X_p ist. Die Matrix \mathbf{S} ist eine positive definite Matrix (vgl. FAHRMEIR/HAMERLE/TUTZ (1996) zur Auswahl dieser Matrix).

Als Spezialfall des multivariaten Kerndichteschätzers wird in unserem Fall die gemeinsame Dichtefunktion $f(y, x)$ durch

$$\begin{aligned} \hat{f}_h(x, y) &= \frac{1}{nh^2} \sum_{i=1}^n \tilde{K} \left(\frac{x - x_i}{h}, \frac{y - y_i}{h} \right) \\ &= \frac{1}{nh^2} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right) K \left(\frac{y - y_i}{h} \right) \end{aligned} \quad (4.4)$$

geschätzt, wobei unterstellt wird, daß die Variablen x und y (stochastisch) unabhängig sind ($\tilde{K}(x, y) = K(x) \cdot K(y)$).

Schlußendlich wird die bedingte Erwartungswertfunktion (4.1) durch

$$\hat{m}_h(x) := \hat{E}(Y|X = x) = \frac{\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right) y_i}{\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x - x_i}{h} \right)}, \quad i = 1, \dots, n, \quad (4.5)$$

geschätzt. Dies ist der sogenannte *Nadaraya-Watson-Kernschätzer*, der von den Daten und von zwei weiteren Größen abhängt: der Kernfunktion K und der Bandbreite h . Die Definition dieser zwei Größen und ihre Wirkungen auf den Kernschätzer werden in den nächsten beiden Abschnitten behandelt.

4.2.2 Kernfunktionen

Der *Kern* K ist eine stetige, beschränkte reelle Funktion, die symmetrisch ist

$$K(u) = K(-u),$$

und die folgende Normierungseigenschaft besitzen soll:

$$\int K(u) du = 1,$$

In vielen Anwendungen handelt es sich um eine Dichtefunktion. Für solche Fälle soll die Kernfunktion auch die Eigenschaft der Nichtnegativität erfüllen.

Die Kernfunktion spezifiziert beim nichtparametrischen Kerndichteschätzer die Größe und die Gewichte für diejenigen Datenpunkte, die die Nachbarschaft des Punktes x bilden, so daß die näher bei x liegenden Datenpunkte stärker gewichtet sind als die Datenpunkte, die weiter von x entfernt sind. In diesem Sinn kann die Kernfunktion als eine Gewichtungsfunktion interpretiert werden.

Die am häufigsten verwendeten Kernfunktionen mit den oben erwähnten Eigenschaften sind der Rechteckkern, der Epanechnikov-Kern, der Bisquare- (Biweight- oder Quartic-) Kern und der Gauss-Kern (vgl. etwa FAHRMEIR/HAMERLE/TUTZ (1996), Kapitel 3, für eine kurze Beschreibung dieser Kerne oder SILVERMAN (1986) für eine ausführlichere).

Im Prinzip ist die Wahl zwischen den verschiedenen Kernen nicht entscheidend für die Schätzung (SILVERMAN (1986), S. 43). Darüber hinaus bemerken FAHRMEIR ET AL. (1997), S. 101, daß für einen größeren Datenumfang n die Kerndichteschätzer mit Epanechnikov-, Bisquare- und Gauss-Kern approximativ identisch sind. In unserer Arbeit wird der Gauss-Kern verwendet, der der Standardnormaldichtefunktion entspricht. Dieser lautet für den univariaten Fall

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right), \quad u \in \mathbb{R}$$

und für den multivariaten Fall

$$K(\mathbf{u}) = \frac{1}{(2\pi)^{p/2} h^p |\mathbf{S}|^{1/2}} \exp\left(-\frac{\mathbf{u}' \mathbf{S}^{-1} \mathbf{u}}{2h^2}\right), \quad \mathbf{u} \in \mathbb{R}^p$$

wobei \mathbf{u} ein p -dimensionaler Vektor von Variablen ist und \mathbf{S} eine positive definite Matrix.

4.2.3 Die Wahl der Bandbreite

Im vorherigen Abschnitt 4.2.1 haben wir gesehen, daß der Kernelschätzer einer beliebigen Dichtefunktion $f(\cdot)$ von den Daten, der Kernfunktion K und der Bandbreite h abhängt. Die Wahl der Bandbreite h beeinflußt die Schätzung stärker als die ausgewählte Kernfunktion. Die Bandbreite h ist ein Parameter, der die Glättung des Kerns bestimmt: Für ein großes h ist die Schätzung der Dichtefunktion sehr glatt. Im Gegensatz dazu, falls der Parameter h klein ist, paßt sich die geschätzte Funktion stärker den Beobachtungen an. (Zur Illustration siehe z.B. Abbildung 5.1.)

Eine andere Wirkung der Bandbreite h betrifft die asymptotische Verteilung des Kernelschätzers in Hinblick auf die Größen Varianz und Verzerrung: Die Varianz des Kernelschätzers sinkt mit dem Wert von h . Im Gegensatz dazu nimmt die Verzerrung mit dem Wert von h zu (siehe die asymptotische Verteilung (4.8) im nächsten Abschnitt). Daraus ergibt sich ein “Trade-off” zwischen der Verzerrung und der Varianz (vgl. SILVERMAN (1986), S. 38ff. für Details). Ein mögliches Kriterium für die Bestimmung der optimalen Bandbreite gibt SILVERMAN (1986), S. 47 an: Zur Schätzung einer beliebigen Dichtefunktion $f(x)$ ist die optimale Bandbreite h_{opt} derjenige Wert von h , der den mittleren integrierten quadratischen Fehler (MIQF) (engl. “mean integrated square error”) minimiert. (Für Details vgl. SILVERMAN (1986).) Für den Gauss-Kern ergibt dieses Kriterium die optimale Bandbreite h , die durch

$$h_{opt} = 1.06 \min(\sigma, R/1.34) \cdot n^{-\frac{1}{5}} \quad (4.6)$$

definiert ist (vgl. SILVERMAN (1986), Kapitel 3). Die Standardabweichung σ und der In-

terquantilabstand R der Dichtefunktion $f(x)$ sind üblicherweise unbekannt und müssen geschätzt werden. Die Schätzer sind bei n Urwerten x_1, \dots, x_n durch

$$\begin{aligned}\hat{\sigma}^2 &:= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\ \hat{R} &:= x_{[0.75]} - x_{[0.25]}\end{aligned}$$

definiert, wobei \bar{x} das arithmetische Mittel der Werte x_1, \dots, x_n ist. Die Größen $x_{[0.75]}$ und $x_{[0.25]}$ bezeichnen das 75%- bzw. 25%-Quantil der Verteilung der Daten. Die geschätzte optimale Bandbreite \hat{h}_{opt} ergibt sich nach dem Einsetzen von σ und R durch $\hat{\sigma}$ bzw. \hat{R} .

Dieses Vorgehen kann in der Praxis “überglättete” Schätzungen liefern, weil die Normalverteilung relativ glatt verläuft. Um dieses Problem zu überwinden, wurde in MÜLLER/FAN (1995) die Anwendung der folgenden Familie

$$\{\hat{f}_h(x), h = 1.4^j \cdot \hat{h}_{opt}, j = -3, -2, -1, 0, 1, 2\} \quad (4.7)$$

von über- und unterglätteten Schätzungen der Dichtefunktion $f(x)$ vorgeschlagen. Anhand von sechs verschiedenen geschätzten Bandbreiten wird die Dichtefunktion $f_h(x)$ entsprechend geschätzt. Die Sensibilität der geschätzten Dichtefunktion in Bezug auf die verschiedenen Bandbreiten kann auf eine einfache Art überprüft werden: Aufgrund des visuellen Vergleichs der verschiedenen geschätzten Kurven wird eine Bandbreite gewählt, die die Kernfunktion nicht überglättet.

Es bestehen noch andere Methoden zur Wahl der Bandbreite, wie die verschiedenen Versionen von Kreuzvalidierung (Kleinste-Quadrate, Likelihood), die Test-Graph-Methode oder die interne Schätzung der Unregelmäßigkeit der Dichtefunktion. Die Berechnungen dieser verschiedenen Methoden sind aber wesentlich intensiver. Eine Beschreibung dieser Methoden findet sich SILVERMAN (1986), S. 43 ff.

4.2.4 Asymptotische Eigenschaften des Nadaraya-Watson-Kernschätzers

Härdle und Linton (HÄRDLE/LINTON (1994), S. 2309) haben gezeigt, daß der Nadaraya-Watson-Kernschätzer unter bestimmten Regularitätsbedingungen konsistent und asymptotisch normalverteilt ist. Für den Nachweis der Konsistenz wird angenommen, daß für den Kern $K(\cdot)$ gilt: $\int |K(u)| du \leq \infty$ und $\lim_{|u| \rightarrow \infty} uK(u) = 0$. Darüber hinaus wird angenommen, daß die Funktionen $m(x)$, $f(x)$ und $\sigma^2(x)$ stetig sind und $f(x) > 0$. Unter der weiteren Annahme, daß für $n \rightarrow \infty$, $h = h(n) \rightarrow 0$ und $nh \rightarrow \infty$ ist der Nadaraya-Watson-Kernschätzer $\hat{m}_h(x)$ konsistent, d.h. es gilt

$$\hat{m}_h(x) \xrightarrow{p} m(x).$$

Für den Nachweis der asymptotischen Normalität wird darüber hinaus angenommen (Theorem 2 von HÄRDLE/LINTON (1994), S. 2309 oder Theorem 4.2.1 von HÄRDLE (1990), S. 99), daß für irgendein $\eta > 0$ gilt $\int |K(u)|^{2+\eta} du < \infty$ und daß die Funktionen $m(x)$ und $f(x)$ zweimal stetig differenzierbar sind. Außerdem wird vorausgesetzt, daß $E(|Y|^{2+\eta} | x)$ existiert und stetig ist. Weiter wird angenommen, daß $\lim h^5 n < \infty$. Dann ist der Schätzer $\hat{m}_h(x)$ asymptotisch normalverteilt, d.h. es gilt

$$\sqrt{nh} \left\{ \hat{m}_h(x) - m(x) - h^2 B_{nw}(x) \right\} \xrightarrow{d} N(0, V_x), \quad (4.8)$$

mit

$$\begin{aligned} B_{nw}(x) &:= \frac{1}{2} \mu_2(K) \left[m''(x) + 2m'(x) \frac{f'(x)}{f(x)} \right] \\ V_x &:= \frac{1}{f(x)} \sigma^2(x) v_2(K), \end{aligned}$$

wobei $f(x)$ die Randdichtefunktion der Variablen X ist und $f'(x)$ ihre erste Ableitung an der Stelle x . Die Funktionen $m'(x)$ und $m''(x)$ sind die erste bzw. zweite Ableitung der bedingten Erwartungswertfunktion $m(x)$ an der Stelle x . Die Terme $\mu_2(K)$ und $v_2(K)$ hängen von der Kernfunktion K ab und sind durch

$$\mu_2(K) := \int v^2 K(v) dv \quad \text{und} \quad v_2(K) := \int K(v)^2 dv \quad (4.9)$$

definiert (siehe etwa HÄRDLE/LINTON (1994), S. 2310).

4.2.5 Konfidenzintervalle für die bedingten Erwartungswertfunktionen

Eine der oben erwähnten Eigenschaften des Nadaraya-Watson-Kernschätzers, seine asymptotische Normalität, wird als Grundlage für die Schätzung von asymptotischen Konfidenzintervallen bei bedingten Erwartungswertfunktionen verwendet.

In diesem Abschnitt werden zwei Alternativen zur Schätzung von Konfidenzintervallen betrachtet:

- Approximierte asymptotische punktweise Konfidenzintervalle, die die Verzerrung des Schätzers nicht berücksichtigen, und
- die Schätzung von Konfidenzintervallen mittels Bootstrap. Dabei werden Vorgehensweisen vorgeschlagen, die die auftretende Verzerrung berücksichtigen.

4.2.5.1 Approximierte asymptotische punktweise Konfidenzintervalle

Im vorigen Abschnitt wurde auf die Verzerrung des Nadaraya-Watson-Kernschätzers hingewiesen (siehe Gleichung (4.8)). Der Term $h^2 B_{nw}(x)$ bezeichnet die Verzerrung des Schätzers, die eine Funktion der ersten und zweiten Ableitung der Funktion $m(x)$ und der ersten Ableitung der Randdichtefunktion $f(x)$ ist, was ihre Berechnung kompliziert macht. Wenn die Bandbreite klein genug gewählt wird, ist die Verzerrung relativ klein und kann ignoriert werden (siehe HÄRDLE/LINTON (1994), S. 2315) für Details).

Ein approximiertes asymptotisches punktweises $(1 - \alpha) \cdot 100\%$ -Konfidenzintervall für den Nadaraya-Watson-Kernschätzer $\hat{m}_h(x)$ ist durch

$$[\hat{m}_h(x) - z_{1-\alpha/2} \cdot \hat{s}(x) ; \hat{m}_h(x) + z_{1-\alpha/2} \cdot \hat{s}(x)] \quad (4.10)$$

definiert, wobei $z_{1-\alpha/2}$ das $(1 - \alpha/2) \cdot 100\%$ -Quantil der standardisierten Normalverteilung ist und $\hat{s}^2(x)$ ein konsistenter Schätzer der asymptotischen Varianz V_x , wie z.B. (vgl. HÄRDLE/LINTON (1994), S. 2317):

$$\hat{s}^2(x) := \frac{v_2(K) \hat{\sigma}_h^2(x)}{nh \hat{f}_h(x)}, \quad (4.11)$$

mit

$$\begin{aligned} \hat{\sigma}_h^2(x) &:= \frac{1}{n} \sum_{i=1}^n W_{hi}(x) \hat{\varepsilon}_i^2 \\ \hat{\varepsilon}_i &:= Y_i - \hat{m}_h(x_i) \\ W_{hi}(x) &:= \frac{K_h(x - x_i)}{\hat{f}_h(x)} \\ v_2(K) &:= \int K(v)^2 dv. \end{aligned}$$

Das in (4.10) definierte Intervall ist nur für jeden Einzelpunkt der Folge $\{x_i\}_{i=1}^n$ bestimmt.

4.2.5.2 Konstruktion von Konfidenzintervallen mit Hilfe des Bootstraps

Die approximierten asymptotischen punktweise Konfidenzintervalle für den Nadaraya-Watson-Kernschätzer sind häufig sehr breit und nicht exakt, d.h. sie überdecken den wahren Wert des Parameters mit einer vom festgelegten Konfidenzniveau $(1 - \alpha)$ verschiedenen Wahrscheinlichkeit. Eine in der Praxis oft zuverlässigere Methode zur Konstruktion von Konfidenzintervallen ist der Bootstrap.

Naiver Bootstrap

Die Methode des *Bootstraps* ist eine auf Daten basierende Simulationsmethode zur statistischen Inferenz, die mit Hilfe von Computern relativ einfach zu implementieren ist. Die Grundidee dieser Methode lautet: Aus einer gegebenen Stichprobe werden möglichst viele Bootstrap-Stichproben generiert, aus denen jeweils eine für das betrachtete Problem geeignete Prüfgröße berechnet wird. Eine Bootstrap-Stichprobe bezeichnet dabei eine aus der Urstichprobe generierte Stichprobe gleichen Umfangs n ,

die aus der Urstichprobe dadurch generiert wird, daß dieser nach der Methode mit Zurücklegen n Elemente entnommen werden. Die zahlreichen Bootstrap-Stichproben liefern auf diese Art eine empirische Verteilung der Prüfgröße, die zur Inferenz benutzt wird. In unserem Fall wird die Bootstrap-Methode durchgeführt, um die Verteilung des

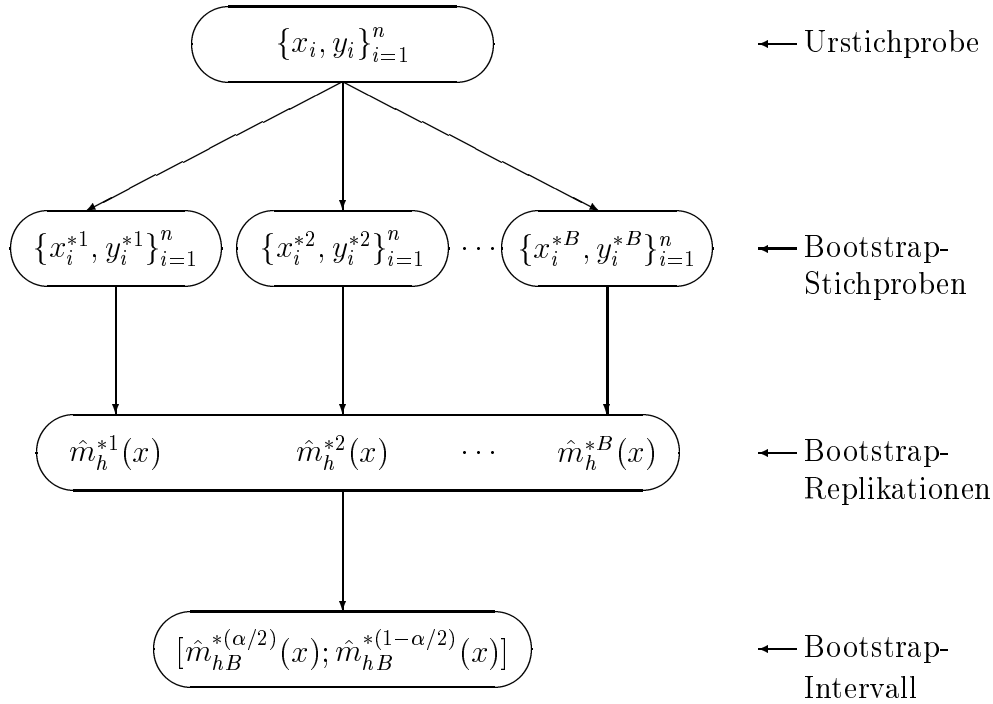


Abbildung 4.1. Schema des Bootstrap-Prozesses zur Schätzung eines Konfidenzintervalls für eine Funktion $\hat{m}(x)$.

Nadaraya-Watson-Kernschätzers $\hat{m}_h(x)$ zu approximieren, so daß die unteren und oberen Grenzen ($\hat{m}_u(x)$ bzw. $\hat{m}_o(x)$) eines Konfidenzintervalls für $\hat{m}_h(x)$ durch die Quantile der empirisch generierten Verteilung für $\hat{m}_h(x)$ gekennzeichnet werden. Dieses Vorgehen ist unter dem Namen *Quantil-Intervall* (engl. “percentile interval”) bekannt (siehe EFRON/TIBSHIRANI (1998), Kapitel 13). Der Algorithmus dieses Verfahrens kann in drei Schritten dargestellt werden (vergleiche das Schema in Abbildung 4.1):

1. *Schritt:* Anhand einer Stichprobe $\{x_i, y_i\}_{i=1}^n$ des zweidimensionalen Zufallsvektors (X, Y) werden B unabhängige Bootstrap-Stichproben gleichen Umfangs

n folgendermaßen generiert: Jedes Element der jeweiligen Bootstrap-Stichprobe wird bei einer Entnahme mit Zurücklegen aus den Originaldaten ausgewählt.

2. *Schritt:* Jede Bootstrap-Stichprobe liefert ein sogenanntes Bootstrap-Replikat $\hat{m}_h^{*b}(x)$, $b = 1, \dots, B$, d.h. einen Wert des Nadaraya-Watson-Kernschätzers, der auf dieselbe Art wie $\hat{m}_h(x)$ anhand der Daten berechnet wurde.

3. *Schritt:* Seien $\hat{m}_{hB}^{*(\alpha/2)}(x)$ und $\hat{m}_{hB}^{*(1-\alpha/2)}(x)$ das $(\alpha/2) \cdot 100\%$ - bzw. das $(1 - \alpha/2) \cdot 100\%$ -Quantil der empirischen Verteilung von $\hat{m}_h^*(x)$. Das approximierte $(1 - \alpha) \cdot 100\%$ -Konfidenzintervall wird durch

$$[\hat{m}_u(x) ; \hat{m}_o(x)] \approx [\hat{m}_{hB}^{*(\alpha/2)}(x) ; \hat{m}_{hB}^{*(1-\alpha/2)}(x)]$$

geschätzt. Im idealen Fall mit unendlichen Replikationen gilt:

$$[\hat{m}_u(x) ; \hat{m}_o(x)] = [\hat{m}_h^{*(\alpha/2)}(x) ; \hat{m}_h^{*(1-\alpha/2)}(x)],$$

wobei $\hat{m}_h^{*(\alpha/2)}(x)$ und $\hat{m}_h^{*(1-\alpha/2)}(x)$ das $(\alpha/2) \cdot 100\%$ -Quantil bzw. das $(1 - \alpha/2) \cdot 100\%$ -Quantil der Bootstrap-Verteilung $\hat{m}_h^*(x)$ sind.

Dieses Vorgehen wird aufgrund seiner Generierung von Bootstrap-Stichproben “naiver” Bootstrap genannt (siehe etwa Algorithmus 4.2.2 von HÄRDLE (1990), S. 104).

Da der Nadaraya-Watson-Kernschätzer verzerrt ist, sollte bei der Schätzung von Konfidenzintervallen mittels Bootstrap die Verzerrung berücksichtigt werden. Die Konvergenzgeschwindigkeit der nach der “Quantil”-Methode geschätzten Bootstrap-Intervalle kann ebenfalls noch verbessert werden. Bei der Quantil-Methode geht der Fehler 1. Art α der Nichtdeckung des wahren Wertes der Funktion $m_h(x)$ gegen Null mit einer Rate von n^{-1} statt einer von $n^{-1/2}$ (siehe EFRON/TIBSHIRANI (1998), S. 187). Bei der Schätzung von Intervallen kann diese Genauigkeit mittels der *BC_a-Methode* (engl. “bias-corrected and accelerated”) erhöht werden¹. Diese Methode fügt zwei Parameter bei der Schätzung der Intervallgrenzen ein: die Korrektur z_0 der Verzerrung und den sogenannten Beschleunigungsparameter a , die beide geschätzt werden müssen.

¹Trotzdem ist die Genauigkeit der Deckung bei dieser Methode anhand kleiner Stichproben nicht unbedingt präzise (vgl. EFRON/TIBSHIRANI (1998), S. 178).

Eine Schätzung des Parameters z_0 ist durch die Proportion von Bootstrap-Replikationen $\hat{m}_h^{*b}(x)$, $b = 1, \dots, B$, die kleiner als die ursprüngliche Schätzung $\hat{m}_h(x)$ sind, gegeben:

$$\hat{z}_0 := \Phi^{-1} \left(\frac{\#\{\hat{m}_h^{*b}(x) < \hat{m}_h(x)\}}{B} \right), \quad (4.12)$$

wobei Φ^{-1} die Inverse der Standardnormalverteilungsfunktion bezeichnet. Das Zeichnen $\#$ bezeichnet die Anzahl der Werte $\hat{m}_h^{*b}(x)$ aus der Bootstrap-Replikationen, die kleiner als die ursprüngliche Schätzung $\hat{m}_h(x)$ sind (siehe EFRON/TIBSHIRANI (1998), S. 186).

Der Beschleunigungsparameter a bezeichnet die Wachstumsrate der Standardabweichung des Schätzers $\hat{m}_h(x)$ in Bezug auf den wahren Wert $m(x)$ und wird durch

$$\hat{a} := \frac{\sum_{i=1}^n (\hat{m}_{h(\cdot)}(x) - \hat{m}_{h(i)}(x))^3}{6 \{\sum_{i=1}^n (\hat{m}_{h(\cdot)}(x) - \hat{m}_{h(i)}(x))^2\}^{3/2}} \quad (4.13)$$

geschätzt, mit $\hat{m}_{h(\cdot)}(x) := \frac{1}{n} \sum_{i=1}^n \hat{m}_{h(i)}(x)$, wobei $\hat{m}_{h(i)}(x)$ die Schätzung der Funktion $m(x)$ ist, die anhand der Stichprobe ohne die Beobachtung i berechnet wurde (siehe EFRON/TIBSHIRANI (1998), S. 186).

Das $(1 - \alpha) \cdot 100\%$ BC_a-Konfidenzintervall ist durch

$$[\hat{m}_h^{*(\alpha_1)}; \hat{m}_h^{*(\alpha_2)}] \quad (4.14)$$

gekennzeichnet, mit

$$\begin{aligned} \alpha_1 &= \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})} \right) \\ \alpha_2 &= \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})} \right), \end{aligned}$$

wobei Φ die Standardnormalverteilungsfunktion ist und $z_{\alpha/2}$ und $z_{1-\alpha/2}$ ihre $(\alpha/2) \cdot 100\%$ - bzw. $(1 - \alpha/2) \cdot 100\%$ -Quantile sind (vgl. EFRON/TIBSHIRANI (1998), S. 185)).

Die Quantile-Methode ist ein Spezialfall der BC_a-Methode: Falls die Parameter \hat{a} und \hat{z}_0 Null sind, liefert die BC_a-Methode die gleichen Konfidenzintervalle wie die Quantile-Methode.

Wilder Bootstrap

Eine andere Variante der Bootstrap-Technik ist der sogenannte “wilde” Bootstrap (siehe etwa HÄRDLE (1990)). Dieser berücksichtigt ebenfalls die Verzerrung der Schätzer, benötigt aber keine zusätzliche Schätzung von Parametern und ist somit einfacher anzuwenden. Aus diesem Grund wurde der wilde Bootstrap hier vorgezogen.

Betrachtet wird ein Regressionsmodell mit Ansatz $Y_i = m(X_i) + \varepsilon_i$, $i = 1, \dots, n$, wobei die ε_i identisch und unabhängig verteilt sind. Gesucht ist eine Schätzung der Verteilung von Y_i anhand der Bootstrap-Beobachtungen Y_i^* , wobei jede i -te Bootstrap-Beobachtung durch $Y_i^* := \hat{m}_h(X_i) + \varepsilon_i^*$ definiert ist. Der Term $\hat{m}_h(X_i)$ bezeichnet den Nadaraya-Watson-Kernschätzer für die unbekannte Funktion $m(X_i)$. Der Term ε_i^* ist das sogenannte Bootstrap-Residuum, das aus einer Zweipunkt-Verteilung \hat{F}_i generiert wird. Die *Zweipunkt-Verteilung* \hat{F}_i ist als eine Konvexkombination von zwei Diracschen Sprungfunktionen definiert, nämlich

$$\hat{F}_i := \gamma \delta_a + (1 - \gamma) \delta_b, \quad (4.15)$$

wobei beide Diracschen Sprungfunktionen δ_a und δ_b diejenigen Verteilungsfunktionen sind, welche den Beobachtungen a und b jeweils die Wahrscheinlichkeitsmasse 1 zuweisen. Die Beobachtungen a und b sind beim wilden Bootstrap wie folgt definiert: $a = \hat{\varepsilon}_i(1 - \sqrt{5})/2$ und $b = \hat{\varepsilon}_i(1 + \sqrt{5})/2$. Darüber hinaus bezeichnen die Parameter γ und $1 - \gamma$ die entsprechenden Wahrscheinlichkeitsmassen der zwei Ereignisse δ_a und δ_b . Beim wilden Bootstrap hat γ den Wert $(5 + \sqrt{5})/10$. Dieser Wert ebenso wie die Werte von a und b sind so ausgewählt, daß die ersten drei Momente der Verteilung von ε_i^* mit den empirischen Momenten der Residuen $\hat{\varepsilon}_i$ des geschätzten ursprünglichen Modells für Y_i überstimmen (siehe HÄRDLE/MARRON (1991) für Details), d.h.

$$E(\varepsilon_i^*) = 0 \quad E(\varepsilon_i^{*2}) = \hat{\varepsilon}_i^2 \quad \text{und} \quad E(\varepsilon_i^{*3}) = \hat{\varepsilon}_i^3. \quad (4.16)$$

²Die Werte a und b sind die Lösungen einer quadratischen Gleichung, die die Grundlage der “Goldenen Sektion” stellt. Die positive Lösung $\frac{1+\sqrt{5}}{2}$ ist “goldene Nummer” genannt. Siehe etwa WALSER/HILTON (2001) für eine Einführung in die Golden Sektion und ihre Anwendungen.

Auf diese Art werden n Bootstrap-Residuen ε_i^* generiert, die zusätzlich mit n Schätzungen der unbekannten Funktion $m(X_i)$ eine Stichprobe von Bootstrap-Beobachtungen Y_i^* , $i = 1, \dots, n$ definieren. Dieses Verfahren wird B -Mal wiederholt, so daß eine Anzahl B von Bootstrap-Stichproben $\{Y_i^{*b}\}_{i=1}^n$, $b = 1, \dots, B$ zur Schätzung zur Verfügung stehen.

“Naiver” und “wilder” Bootstrap unterscheiden sich in der Art, wie die Bootstrap-Residuen ε_i^* , $i = 1, \dots, n$, generiert werden. Beim “naiven” Bootstrap ist jedes Bootstrap-Residuum zufällig aus der Verteilung der Residuen des geschätzten Modells generiert. Im Gegensatz dazu ist jedes Bootstrap-Residuum ε_i^* beim “wilden” Bootstrap aus einem Einzelresiduum $\hat{\varepsilon}_i$ des geschätzten ursprünglichen Modells bestimmt (daher die Bezeichnung “wild”).

Im allgemeinen wird ein überglätteter Schätzer der unbekannten Funktion $m(X)$ für die Konstruktion der Bootstrap-Beobachtungen Y_i^* benötigt. Unter dem “überglätteten” Schätzer versteht man denjenigen Schätzer, der auf einer größeren Bandbreite g als der optimalen Bandbreite h basiert, und deswegen eine glattere geschätzte Funktion liefert. Somit weisen die Verteilung und die mittels Bootstrap generierte Verteilung des Schätzers die gleiche Verzerrung auf (siehe HÄRDLE (1990), S. 107 für Details). Anhand der Bootstrap-Daten $\{(X_i, Y_i^*)\}_{i=1}^n$ wird das Vorgehen wie beim “naiven” Bootstrap durchgeführt.

4.3 Schätzung von additiven semiparametrischen Regressionsmodellen

In diesem Abschnitt werden wir die Schätzung des Parametervektors β des additiven semiparametrischen Modells (3.5) näher betrachten. Dabei wird unterstellt, daß die unbekannte Funktion $q(\cdot)$ in (3.5) nur von einem Regressor X abhängt. Zur Schätzung solcher additiven semiparametrischen Modellen ist der sogenannte *Algorithmus von Speckman* einfach zu verwenden (siehe in GREEN/SILVERMAN (1994), S. 85-86).

Für die bedingten Erwartungswertfunktionen von \mathbf{Y} bzw. \mathbf{Z} seien die Schätzer $\hat{E}(\mathbf{Y}|X = x) = \mathbf{W}_h \mathbf{Y}$ und $\hat{E}(\mathbf{Z}|X = x) = \mathbf{W}_h \mathbf{Z}$ gegeben, z.B. die sogenannten Nadaraya-Watson-Schätzer (siehe Abschnitt 4.2). Hier ist \mathbf{W}_h eine “Gewichtungsmatrix” mit (j, i) -tem Element:

$$W_{hji}(X = x_j) := \frac{K_h(x_j - x_i)}{\hat{f}_h(x_j)},$$

für $j = 1, \dots, k$ und $i = 1, \dots, n$.

Nach dem Einsetzen dieser Schätzer lautet die Gleichung (3.6):

$$\mathbf{Y} - \mathbf{W}_h \mathbf{Y} = (\mathbf{Z} - \mathbf{W}_h \mathbf{Z})\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Mit $\mathbf{Y}^* := (\mathbf{I} - \mathbf{W}_h)\mathbf{Y}$ und $\mathbf{Z}^* := (\mathbf{I} - \mathbf{W}_h)\mathbf{Z}$ ergibt sich:

$$\mathbf{Y}^* = \mathbf{Z}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (4.17)$$

Somit lautet der semiparametrische Schätzer $\hat{\boldsymbol{\beta}}$ für den Parametervektor $\boldsymbol{\beta}$ nach der gewöhnlichen Kleinste-Quadrate-Methode:

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}^{*'} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*'} \mathbf{Y}^*. \quad (4.18)$$

4.3.1 Die asymptotische Verteilung des semiparametrischen Schätzers

ROBINSON (1988) hat gezeigt, daß der semiparametrische Schätzer $\hat{\boldsymbol{\beta}}$ für den additiven Fall unter bestimmten Regularitätsbedingungen asymptotisch normalverteilt ist. Diese Regularitätsbedingungen implizieren zunächst, daß die Kovarianzmatrix $\Phi := E[\{\mathbf{Z}|X - E(\mathbf{Z}|X)\}\{\mathbf{Z}|X - E(\mathbf{Z}|X)\}']$ des bedingten Zufallsvektors $\mathbf{Z}|X$ positiv definit ist. Diese Eigenschaft ist notwendig und hinreichend für die asymptotische Normalität von $\hat{\boldsymbol{\beta}}$, d.h. für

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma^2 \Phi^{-1}), \quad (4.19)$$

wobei $\sigma^2 \Phi^{-1}$ die Varianz-Kovarianz-Matrix des semiparametrischen Schätzers $\hat{\beta}$ ist, mit $\Phi := E[\{\mathbf{Z}|X - E(\mathbf{Z}|X)\}\{\mathbf{Z}|X - E(\mathbf{Z}|X)\}']$ und $\sigma^2 = \text{Var}(\epsilon)$.

Außerdem konvergiert der semiparametrische Schätzer $\hat{\beta}$ in Wahrscheinlichkeit schneller als reine nichtparametrische Schätzer.

Die in (4.19) angegebene Varianz-Kovarianz-Matrix $\text{Var}(\hat{\beta}) := \sigma^2 \Phi^{-1}$ des Schätzers kann durch einen konsistenten Schätzer $\hat{\sigma}_h^2 (\mathbf{Z}^{*'} \mathbf{Z}^*)^{-1}$ geschätzt werden, wobei $\hat{\sigma}_h^2$ die folgende Schätzung der Residualvarianz σ^2 bezeichnet:

$$\hat{\sigma}_h^2 := \frac{\mathbf{d}' \mathbf{d}}{sp(\mathbf{W}_h + \{\mathbf{Z}^{*'} \mathbf{Z}^*\}^{-1} \{\mathbf{Z}^{*'} (\mathbf{I} - \mathbf{W}_h) \mathbf{Z}^*\})}, \quad (4.20)$$

mit $\mathbf{d} := \mathbf{Y} - \hat{q}(x) - \mathbf{Z} \hat{\beta}$, wobei $\hat{q}(x)$ eine nichtparametrische Schätzung der unbekannten Funktion $q(x)$ im Ansatz (3.5) ist, die sich nach dem Einsetzen des Schätzers $\hat{\beta}$ in (3.5) und der Anwendung des Kernschätzers auf die Residuen $y_i - \mathbf{z}_i \hat{\beta}$ ergibt:

$$\hat{q}(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - x_i) (y_i - \mathbf{z}_i \hat{\beta})}{\hat{f}_h(x)}. \quad (4.21)$$

Im Nenner von (4.20) bezeichnet sp die Spur der Hat-Matrix vom Ansatz (4.17)³, d.h. $\mathbf{H} := \mathbf{W}_h + \{\mathbf{Z}^{*'} \mathbf{Z}^*\}^{-1} \{\mathbf{Z}^{*'} (\mathbf{I} - \mathbf{W}_h) \mathbf{Z}^*\}$, mit $sp(\mathbf{H}) = \text{Rang}(\mathbf{H})$, die den Freiheitsgraden der Störvariablen ϵ_i entspricht⁴. Die Matrix \mathbf{I} ist die $(n \times n)$ -Identitätsmatrix.

Dies gilt für den einfacheren additiven Fall. Im Gegensatz dazu ist unser Modell (3.4) multiplikativ. Deswegen muß ein anderer Weg gefunden werden, um die asymptotische Verteilung des semiparametrischen Schätzers herzuleiten. Dieser Weg wird im nächsten Kapitel besprochen, wobei anhand eines Datensatzes der Stadt Darmstadt

³In Matrizen-Notation ergibt sich die Hat-Matrix anhand der prognostizierten Werte $\hat{\mathbf{Y}}$ wie folgt: $\hat{\mathbf{Y}} = \hat{q}(x) + \mathbf{Z} \hat{\beta} = \mathbf{W}_h (\mathbf{Y} - \mathbf{Z} \hat{\beta}) + \mathbf{Z} \hat{\beta} = \mathbf{W}_h \mathbf{Y} + (\mathbf{I} - \mathbf{W}_h) \mathbf{Z} (\mathbf{Z}^{*'} \mathbf{Z}^*)^{-1} \mathbf{Z}^* (\mathbf{I} - \mathbf{W}_h) \mathbf{Y} = [\mathbf{W}_h + \mathbf{Z}^* (\mathbf{Z}^{*'} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*'} (\mathbf{I} - \mathbf{W}_h)] \mathbf{Y} = \mathbf{H} \mathbf{Y}$, da $\hat{\beta} = (\mathbf{Z}^{*'} \mathbf{Z}^*)^{-1} \mathbf{Z}^{*'} \mathbf{Y}^*$ (vgl. GREEN/SILVERMAN (1994), S. 86).

⁴Diese Anwendung im Rahmen von Kernschätzungen folgt das von GREEN/SILVERMAN (1994), S. 39, vorgeschlagene Vorgehen zur Schätzung der Residualvarianz eines Regressionsmodells. Dabei wird eine nichtspezifizierte Funktion betrachtet, die mittels der “spline smoothing”-Technik geschätzt wurde.

(DE) 1994/95 auf die Schätzung des Modells eingegangen wird.

Kapitel 5

Schätzung des semiparametrischen Modells für Regressionsmietspiegel

5.1 Einleitung

In diesem Kapitel wird auf die Schätzung des in Kapitel 3 vorgeschlagenen semiparametrischen Modells für Regressionsmietspiegel eingegangen. Nach einer geeigneten Transformation des Modells werden zuerst die bedingten Erwartungswerte der Nettomiete und der Mietpreisdeterminanten, die aufgrund der Transformation im Modell auftreten, mit Hilfe des nichtparametrischen Nadaraya-Watson-Kernschätzers geschätzt. Diese geschätzten Erwartungswertfunktionen werden folgend in das transformierte Modell eingesetzt, so daß der parametrische Teil des Modells mittels der nichtlineare Kleinst-Quadrat-Methode geschätzt werden kann.

Da im Rahmen von semiparametrischen Regressionsmodellen ein solches additiv-multiplikatives Modell wenig bekannt ist, wird die Modellschätzung hier ausführlich betrachtet. Darüber hinaus werden Signifikanztests durchgeführt und Konfidenzintervalle sowohl für die Grundmiete als auch für die Nettomiete ermittelt.

Die Schätzung basiert auf den Daten, die 1994/95 für den Mietspiegel der Stadt Darm-

stadt erhoben wurden: Es liegen Beobachtungen über 1192 Wohnungen hinsichtlich des Mietpreises und ca. 100 relevanter Wohnungscharakteristika vor.

Bevor die Schätzung des semiparametrischen Regressionsmodells für den Mietpreis diskutiert wird, werden in diesem Kapitel zuerst die relevanten Mietpreisdeterminanten oder erklärenden Variablen definiert.

Aufgrund der Vielfalt von erklärenden Variablen, die teilweise die gleiche Information beinhalten, wird vor der Schätzung des Regressionsmodells eine Faktorenanalyse durchgeführt. Damit wird die Anzahl von Variablen auf weniger voneinander unabhängige Faktoren reduziert. Somit wird das Problem redundanter Information überwunden, welches bei älteren Mietspiegeln vernachlässigt wurde. Deswegen wurde die Gültigkeit einiger Mietspiegel wie z.B. des Kieler Mietspiegels 1992 in Frage gestellt (siehe etwa KRÄMER (1993)).

5.2 Definition von Mietpreisdeterminanten

In diesem Abschnitt werden die wichtigsten Merkmale aufgeführt, die als erklärende Variablen in der hier durchgeführten Regressionsanalyse zur Schätzung eines Mietspiegels berücksichtigt werden. Die folgende einleitende Beschreibung basiert auf Hinweisen des Deutschen Gesetzgebers, der eine Gliederung für die Merkmale vorschlägt. Die in dieser Arbeit betrachteten Merkmale werden gleich gegliedert.

Die Grundstruktur eines Mietspiegels ist in Deutschland durch §2 Abs. 1 Nr. 2 des Miethöhegesetzes (MHG) von 1974 vorgegeben. Wie in der “Fortschreibung der Hinweise für die Aufstellung von Mietspiegeln” des Bundesministerium für Raumordnung, Bauwesen und Städtebau erwähnt (vgl. BUNDESMINISTERIUM FÜR RAUMORDNUNG (1980), S. 167), sind die dort gegebenen Hinweise nur als Leitlinien zu verstehen. Der Mietspiegel sollte nach den folgenden Kriterien, die die Eigenschaften der Wohnungen einschließen, gegliedert werden:

1. **Größe:** Dieses Merkmal mißt die Quadratmeterzahl der Wohnfläche einer Woh-

nung, d.h. “die Grundfläche der Räume, die ausschließlich zur Wohnung gehören”.

2. **Art:** Dieses Vergleichsmerkmal bezieht sich insbesondere auf die Gebäudeart (Einfamilienhäuser, Zweifamilienhäuser, Hochhäuser, sonstige Mehrfamilienhäuser).
3. **Ausstattung:** Dieses Merkmal umfaßt eine Vielzahl von Variablen (z.B. Heizungsart, Küche, Trennung von Bad und Toilette, Aufzug, Balkon, Art der Fußböden, usw.). Da dieses Merkmal sehr heterogen ist, schlägt das Deutsche Bundesministerium die folgende Operationalisierung der Ausstattungs-Variablen vor:

“... Zur Wahrung der Übersichtlichkeit der Mietspiegel und zur Vermeidung übermäßiger Erhebungskosten ist eine Beschränkung auf die wichtigsten Ausstattungsmerkmale notwendig. Es wird empfohlen, das Merkmal “Ausstattung” in der Tabelle [eines Mietspiegels] in mindestens folgender Aufteilung zu berücksichtigen:

- Wohnungen mit Bad oder Duschaum und mit Sammelheizung
- Wohnungen mit Bad oder Duschaum oder Sammelheizung
- Wohnungen ohne Bad, Duschaum und Sammelheizung”.

4. **Beschaffenheit:** Dieses Merkmal charakterisiert Bauweise, Zuschnitt und baulichen Zustand (Erhaltungszustand des Gebäudes bzw. der Wohnung). Dieses Merkmal ist schwierig zu operationalisieren. Deshalb wird es häufig durch das Merkmal Baujahr ersetzt, denn Bauweise und Zuschnitt hängen oft vom Baujahr ab. Das deutsche Bundesministerium für Raumordnung, Bauwesen und Städtebau (DBRBS) empfiehlt z.B. die folgenden Altersgruppierungen:

Baujahr (Jahr der Fertigstellung)

	bis	1948
1949	bis	1960
1961	bis	1971
1972	bis	heute

Hierzu macht das DBRBS zwei Anmerkungen: Erstens können jüngere neugebaute Wohnungen in engere Baualtersgruppen zusammengefaßt werden, damit zu große Mietspannen vermieden werden können. Zweitens kann eine renovierte Wohnung einer jüngeren Baualtersklasse zugeordnet werden.

5. **Lage:** Die wichtigen Merkmale der Lagequalität sind diejenigen, die die Umgebung der Wohnung betreffen, wie z.B. ruhige Lage, verkehrsgünstige Lage, Nähe zu Geschäftszentren, geringe Immissionen, starker Verkehr, Beeinträchtigung des Wohnens durch Handwerks- und Gewerbebetriebe, das Fehlen von Frei- und Grünflächen.

Im folgend werden die Variablen aufgelistet, die jedes Merkmal kennzeichnen und als erklärende Variablen in den weiteren Analysen berücksichtigt werden.

Gebäudeart-Variablen

Aufzug: das Gebäude hat einen Aufzug (dummy).

Awgruen: die Aussenwände sind begrünt (dummy).

Bstgtren: das Gebäude ist nur durch einen Bürgersteig von der Straße getrennt (dummy).

Geb120s: das Gebäude ist freistehend oder einseitig oder zweiseitig angebaut (kategorial).

Gebhoehe: Anzahl der Geschosse / Etagen des Gebäudes (ordinal).

Hinthaus: das Gebäude ist ein Hinterhaus (dummy).

Laubgang: die Wohnung ist über einen Laubengang (Etagen-Aussenflur) zugänglich (dummy).

Stuffrei: die Wohnung ist ohne Überwindung von Treppenstufen zugänglich (dummy).

Gebtyp: das Gebäude ist ein Ein-, Zwei-, Drei-, Mehrfamilienhaus (3 Kategorien: Ein-, Zwei-, Drei- und Mehrfamilienhaus).

Whgzahl: Anzahl der Wohnungen bzw. Wohnungstüren auf dieser Etage (ordinal).

Wohnungsart-Variablen

Wohnungstyp: Art der Wohnung: Einliegerwohnung, komplette Etagenwohnung, Maisonnette-Wohnung, Penthouse-Wohnung, Teilwohnung oder sonstiges Einzelzimmer (6 Kategorien)

Baualter: Jahr der Fertigstellung des Gebäudes (3 Kategorien).

Beschaffenheits-Variablen

Woeb: Die Wohnräume sind durch kleine Stufen/Absätze miteinander verbunden (keine Maisonnette-Wohnung) (dummy).

Qluft0: Es ist keine Querlüftung möglich (dummy).

Gefraum: Die Wohnung hat gefangene Räume (dummy).

Wdiele: Die Wohnung hat eine Wohndiele (Flur mit Fenster) (dummy).

Spkam: Die Wohnung hat eine Speisekammer (dummy).

Hwraum: Es gibt einen Hauswirtschaftsraum (dummy).

Wrhoehe0: Die Höhe der Wohnräume ist niedriger als 2 Meter 40 (dummy).

Tschwel: Es gibt Türschwellen im Türrahmen (dummy).

Tbreite: Alle Türöffnungen sind mindestens 90cm breit (dummy).

Wflkwr: Die Wohnfläche des kleinsten Wohnraums in qm (kardinal).

Wflgwr: Die Wohnfläche des größten Wohnraums in qm (kardinal).

Whglaut: Es sind Lauf- oder sonstige Geräusche in der Wohnung zu hören (dummy).

Balkz2: Es gibt zwei oder mehr Balkone, Terrassen usw. (dummy).

Balkb2: Die Breite des größten Balkons (oder der Terrasse) ist gleich oder größer als 2 m (dummy).

Balkbrn: Die Blickrichtung des Balkons (oder der Terrasse) liegt zu Straße, Eisenbahn und Gewerbetrieben (dummy).

Erzust: Der Erhaltungszustand des Gebäudes innen und/oder außen (9 Kategorien).

Lage-Variablen

Lage der Wohnung innerhalb des Gebäudes:

Etage: In welchem Geschoß liegt die Wohnung? (8 Kategorien: Untergeschoß, 1. Etage/Erdgeschoß/Parterre, 2. Etage, 3. Etage, 4. Etage, 5. Etage, 6. Etage, Dachgeschoß).

Aussicht: Unverbaute Fernsicht (dummy).

Lage des Gebäudes innerhalb der Umgebung/Quartiers:

Wald: Direkt am Wald/an Naherholungsgebiet gelegen oder Wald/Naherholungsgebiet in fußläufiger Entfernung (dummy).

Park: Direkt am Park/an öffentlicher Grünanlage gelegen oder Park/öffentliche Grünanlage in fußläufiger Entfernung (dummy).

Bus: Direkt an Straßenbahn-/Bushaltestelle oder Straßenbahn-/Bushaltestelle in fußläufiger Entfernung(dummy).

Geschaeft: Vorhandensein von Geschäfte des kurzfristigen Bedarfs (z.B. Lebensmittelgeschäft, Bäckerei, Metzgerei), des mittelfristigen Bedarfs (z.B. Apotheke, Schuhe, Bekleidung) oder des langfristigen Bedarfs (z.B. Elektrofachgeschäft, Möbel) in fußläufiger Entfernung (dummy).

Licht0: Unzureichende nächtliche Beleuchtung des Wohnviertels (dummy).

Temp30: Wohnung liegt in einer Tempo-30-Zone (dummy).

Kfzpark0: Unzureichende Parkmöglichkeiten in Wohnungsnähe (dummy).

Zebra0: Unzureichende sichere Straßen-Überquerungs-möglichkeiten im Wohnviertel (dummy).

Bstgpark: Die Bürgersteige im Wohnviertel sind häufig durch parkende Autos eingeengt (dummy).

Vkla: Liegt die Wohnung an einer... (4 Kategorien: stark befahrenen Durchgangsstraße, Straße mit normaler Verkehrsbelastung mit Verkehrsspitzen, Straße mit normaler Verkehrsbelastung ohne Verkehrsspitzen, Anlieger- bzw. Wohnstraße oder Sackgasse).

Indlärm: Lärm bei der Industrie verursacht (3 Kategorien: Es gibt keine Industrie- und Gewerbebelastung, keinen Betrieb in unmittelbarer Nachbarschaft, aber Belastung vorhanden und belastender Betrieb in unmittelbarer Nachbarschaft).

Sondereinrichtungs-Variablen

Waschma: Waschmaschine vom Vermieter gestellt (dummy).

Waetrock: Wäschetrockner vom Vermieter gestellt (dummy).

Waschkue: Waschküche nutzbar (dummy).

Troraum: Wäschetrockenraum nutzbar (evtl. auch auf dem Dachboden/Speicher) (dummy).

Trockplz: Wäschetrockenplatz nutzbar (dummy).

Radkel: Fahrradkeller/Abstellraum nutzbar (dummy).

Hobraum: Hobbyraum/Werkraum nutzbar (dummy).

Partraum: Partyraum/Keller nutzbar (dummy).

Sauna: Sauna nutzbar (dummy).

Schwibad: Schwimmbad nutzbar (dummy).

Spielplz: Kinderspielplatz nutzbar (dummy).

Grillplz: Grillplatz/Sitzgruppe nutzbar (dummy).

Keller: Keller/Kelleranteil (dummy).

Speicher: Speicher/Dachboden (dummy).

Garagetyp: KFZ-Einzelgarage, KFZ-Doppelgarage, KFZ-Sammelgarage oder Reservierter KFZ-Abstellplatz (4 Kategorien).

Garten: Gartenbenutzung in Mietvertrag eingeschlossen (3 Kategorien: alleinige Benutzung, Benutzung mit anderen Hausbewohnern, keine Benutzung).

Ausstattungs-Variablen

44 Variablen beschreiben die Ausstattung der Wohnungen. Der interessierte Leser wird auf die komplexen Faktoren auf den Seiten 92 ff. hingewiesen, die einen deutlichen Überblick geben.

5.3 Datenreduktion und Bildung von komplexen Mietpreisdeterminanten

Für die Schätzung eines Regressionsmietspiegels muß entschieden werden, welche Regressoren in das Modell aufzunehmen sind. Aufgrund der vorhandenen Daten aus Darmstadt 1994/95 steht eine Vielfalt von Variablen zur Verfügung, welche die Wohnungseigenschaften betreffen. Um diese umfangreiche Zahl von Variablen zu reduzieren, die teilweise die gleiche Information beeinhalteten, wird vor der Schätzung des Regressionsmodells, eine Faktorenanalyse durchgeführt. Dabei werden voneinander unabhängige komplexe Variablen (die Faktoren) konstruiert.

Die Faktorenanalyse (FA) wird hier angewendet, um die Dimensionalität der Daten zu verringern: Die große Menge von beobachtbaren Variablen in dem Datensatz aus Darmstadt (118 wie weiter oben aufgelistet wurde) wird damit auf möglichst wenige hypothetische Variablen (Faktoren) reduziert. In diesem Sinn handelt es sich um ein deskriptives Verfahren zur Datenreduktion. Darüber hinaus hilft die FA bei der Analyse von komplexen Merkmalen. Man betrachtet z.B. das komplexe Merkmal Ausstattung. Dieses hat eine Vielzahl von unterschiedlichen Ausstattungsmöglichkeiten:

z.B. bezüglich der Küche, der verschiedenen Zimmer, von Bad und Toilette, Einrichtungselemente wie Fußböden, Verglasung, Heizungsart usw. Wenn alle Variablen, die einen spezifischen Gegenstand betreffen (z.B. die Küche) als Gruppe analysiert werden können und zu einem bestimmten Faktor zusammengefaßt werden, kann der Einfluß bestimmter Merkmale deutlicher identifiziert werden.

Im nächsten Abschnitt wird die Faktorenanalyse vorgestellt. Zunächst wird das Modell der Faktorenanalyse definiert und kurz beschrieben. Anschließend werden zwei Schätzverfahren für die Parameter des Modells angegeben. Dann wird die Rotation der Faktoren und ihre Rolle bei der Interpretation der Faktoren behandelt. Mit der Schätzung der Faktorenwerte wird die Einführung in die Faktorenanalyse abgeschlossen. Im letzten Abschnitt werden die Ergebnisse der für die verschiedene Mietpreismerkmale durchgeführten Faktorenanalyse präsentiert. Die auf diese Art geschätzten Faktoren werden in der Regressionsanalyse für den Mietpreis als erklärende Regressoren verwendet.

5.3.1 Einleitung in die explorative Faktorenanalyse

Das Ziel der FA ist, die Zusammenhänge zwischen Variablen durch wenige unbeobachtbare Größen, den sogenannten Faktoren, zu beschreiben (vgl. BRACHINGER/OST (1996), S. 639 ff.). Bei der FA wird angenommen, daß p beobachtbare Zufallsvariablen X_i , $i = 1, \dots, p$, als Linearkombinationen von k ($k < p$) unbeobachtbaren Zufallsvariablen F_j , $j = 1, \dots, k$, und p spezifischen Variablen ε_i , $i = 1, \dots, p$ dargestellt werden können. Die unbeobachtbaren Zufallsvariablen F_j werden auch *gemeinsame Faktoren* und die spezifischen Variablen ε_i *Einzelrestfaktoren* genannt. Das Modell lautet

$$\begin{aligned} X_1 &= \mu_1 + l_{11}F_1 + l_{12}F_2 + \dots + l_{1k}F_k + \varepsilon_1 \\ X_2 &= \mu_2 + l_{21}F_1 + l_{22}F_2 + \dots + l_{2k}F_k + \varepsilon_2 \\ &\vdots \\ X_p &= \mu_p + l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pk}F_k + \varepsilon_p, \end{aligned}$$

wobei die Koeffizienten l_{ij} , $i = 1, \dots, p$ und $j = 1, \dots, k$, *Ladungen* sind, mit der die

i -te Variable auf den j -ten Faktor lädt. Die Konstante μ_i bezeichnet den Erwartungswert der Variable X_i . In matrizieller Notation wird das Modell der FA folgendermaßen dargestellt:

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L} \mathbf{f} + \boldsymbol{\varepsilon}, \quad (5.1)$$

wobei \mathbf{X} ein $(p \times 1)$ -Vektor von Zufallsvariablen ist, $\boldsymbol{\mu}$ ist ein $(p \times 1)$ -Vektor von Erwartungswerten von \mathbf{X} , \mathbf{L} ist die Ladungsmatrix, \mathbf{f} ist ein $(k \times 1)$ -Vektor aus den gemeinsamen Faktoren und $\boldsymbol{\varepsilon}$ bezeichnet einen $(p \times 1)$ -Vektor von Einzelrestfaktoren.

Falls für jede Zufallsvariable X_i n Realisierungen beobachtbar sind, spricht man vom *Modell für die Daten*:

$$\mathbf{X} - \mathbf{M} = \mathbf{F}\mathbf{L}' + \mathbf{E}, \quad (5.2)$$

mit $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)' = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p)$, $\mathbf{X}_i = (X_{1i}, \dots, X_{ni})'$ und $\mathbf{M} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n)'$, wobei jede Zeile der Matrix \mathbf{M} einen konstanten Vektor von Erwartungswerten bezeichnet. Die restlichen Matrizen sind folgendermaßen definiert: $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n)' = (\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_k) \in \mathbb{R}^{n \times k}$ und $\mathbf{E} = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \dots, \boldsymbol{\varepsilon}_n)' = (\mathbf{E}_1, \mathbf{E}_2, \dots, \mathbf{E}_p) \in \mathbb{R}^{n \times p}$.

Die folgenden Annahmen werden zusätzlich eingefügt:

$$(A1) \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}_p$$

$$(A2) \quad \text{Var}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Psi} = \text{diag}\{\psi_1^2, \dots, \psi_p^2\}.$$

Die Einzelrestfaktoren besitzen den Erwartungswert Null, sind untereinander unkorreliert und haben die Varianzen $\psi_1^2, \dots, \psi_p^2$. Die $(p \times p)$ -Matrix $\boldsymbol{\Psi}$ wird *Einzelrestvarianzmatrix* genannt.

$$(A3) \quad E(\mathbf{f}) = \mathbf{0}_k$$

$$(A4) \quad \text{Var}(\mathbf{f}) = E(\mathbf{f}\mathbf{f}') = \mathbf{I}_k.$$

Es wird angenommen, daß die Faktoren standardisiert sind, d.h. sie besitzen Erwartungswert Null und Varianz 1. Darüber hinaus sind die Faktoren zusätzlich untereinander unkorreliert.

$$(A5) \quad \text{Cov}(\mathbf{f}, \boldsymbol{\varepsilon}) = \mathbf{0}_{k \times p},$$

was bedeutet, daß die gemeinsamen Faktoren \mathbf{f} mit den Einzelrestfaktoren $\boldsymbol{\varepsilon}$ nicht korreliert sind.

Sei $\boldsymbol{\Sigma}$ die Kovarianzmatrix der beobachtbaren Variablen. Aufgrund der Annahmen (A1) bis (A5) erhält man:

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}. \quad (5.3)$$

Diese Gleichung ist unter dem Namen *Grundgleichung* oder *Fundamentaltheorem der FA* bekannt. Für den Beweis und eine detaillierte Diskussion über die Identifizierbarkeit der Lösung der Grundgleichung nach \mathbf{L} und $\boldsymbol{\Psi}$ wird auf BRACHINGER/OST (1996), S. 643-648 verwiesen. Nach dieser Gleichung gibt es folgende Zerlegung für die Varianz σ_{ii} der Variablen X_i , $i = 1, \dots, p$:

$$\sigma_{ii} = h_i^2 + \psi_i^2, \quad (5.4)$$

mit $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{ik}^2$. Diese Summe der Quadrate der Ladungen der i -ten Variablen auf die k gemeinsamen Faktoren ist der Anteil an der Varianz σ_{ii} , der von den gemeinsamen Faktoren erklärt wird. Dieser Anteil heißt *Kommunalität*. Aufgrund der Modellannahmen (A1) bis (A5) folgt, daß die Ladungsmatrix \mathbf{L} die Kovarianzmatrix zwischen den beobachtbaren Variablen \mathbf{x} und den Faktoren \mathbf{f} darstellt, nämlich

$$\text{Cov}(\mathbf{x}, \mathbf{f}) = \mathbf{L}. \quad (5.5)$$

In der Praxis ist die Kovarianzmatrix $\boldsymbol{\Sigma}$ nicht bekannt und muß geschätzt werden. Als Schätzer wird die empirische Kovarianzmatrix \mathbf{S} verwendet, die folgendermaßen definiert ist: $\mathbf{S} := \frac{1}{n} \mathbf{X}' \mathbf{J}_n \mathbf{X}$, mit $\mathbf{J}_n := \mathbf{I}_n - n^{-1} \mathbf{i}_n \mathbf{i}_n'$ und \mathbf{i}_n ist ein $(n \times 1)$ -Vektor, dessen

Elemente den Wert 1 annehmen. Dann soll Schätzer $\hat{\mathbf{L}}$ und $\hat{\mathbf{\Psi}}$ für \mathbf{L} und $\mathbf{\Psi}$ bestimmt werden, so daß

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\mathbf{\Psi}} \quad (5.6)$$

möglichst eine gute Anpassung an \mathbf{S} ist (vgl. BRACHINGER/OST (1996)).

Falls die Datenvariablen standardisiert sind, wird die empirische Korrelationsmatrix $\mathbf{R} := \mathbf{S}_{\text{diag}}^{-1/2} \mathbf{S} \mathbf{S}_{\text{diag}}^{-1/2}$ anstatt der empirischen Kovarianzmatrix verwendet, mit $\mathbf{S}_{\text{diag}} = \text{diag}(\mathbf{S})$.

5.3.1.1 Schätzverfahren

Um die Ladungsmatrix \mathbf{L} , die Einzelrestvarianzmatrix $\mathbf{\Psi}$ und die Anzahl der Faktoren k des faktorenanalytischen Modells zu schätzen, stehen verschiedene Verfahren zur Verfügung: Am häufigsten werden die Hauptkomponentenanalyse (HKA), die die Hauptkomponentenmethode (HKM) und die Hauptfaktorenanalyse (HFA) umfaßt, und die Maximum-Likelihood-Methode (MLM) verwendet. Hier wird nur auf die Hauptkomponentenanalyse eingegangen.

PEARSON (1901) und HOTELLING (1933) entwickelten eine empirische Methode, die sogenannte *Hauptkomponentenanalyse*, die auf die Reduktion von untereinander korrelierten beobachtbaren Variablen zielt, so daß möglichst viel der Gesamtvarianz dieser Variablen durch wenige untereinander unkorrelierte Variablen (die sogenannten Hauptkomponenten) erklärt wird. Die wichtigste Eigenschaft dieser Methode ist, daß die Komponenten nacheinander den maximalen Anteil der Varianz der Gruppe von Variablen erklären: Die erste Hauptkomponente ist diejenige Linearkombination der Datenvariablen, die den größten Anteil der Varianz erklärt; die zweite Hauptkomponente ist eine andere Linearkombination der Variablen, die den zweitgrößten Anteil der Varianz erklärt, usw., so daß die gesamte Varianz durch die Hauptkomponenten erklärt wird. Die Hauptkomponenten sind untereinander unkorreliert. Weil das Ziel der Analyse häufig in der Datenreduktion liegt, werden in der Praxis nicht alle Haupt-

komponenten einbezogen, auch wenn damit nicht 100% der Varianz der ursprünglichen Variablen erklärt wird.

Die *Hauptkomponentenmethode* ist eine rein deskriptive Methode zur Bestimmung des in (5.2) angegebenen Modells für die Daten. Sei \mathbf{Z} die $(n \times p)$ -dimensionale Matrix der standardisierten Datenvektoren $\mathbf{Z}_1, \dots, \mathbf{Z}_p \in \mathbb{R}^n$, mit $\mathbf{Z}_i := \mathbf{X}_i - \boldsymbol{\mu}_i$ für alle $i = 1, \dots, p$. Das Ziel der Methode besteht darin, anhand der empirischen Korrelationsmatrix \mathbf{R} der standardisierten Daten \mathbf{Z} , die die Zusammenhänge zwischen den Variablen kennzeichnet, die ersten k ($k < p$) Faktoren $\mathbf{F}_1, \dots, \mathbf{F}_k$ oder Hauptkomponenten auszuwählen, während die restlichen $(p - k)$ Faktoren vernachlässigt werden.

Aus einer vollständigen Hauptkomponentenmethode ergibt sich die folgende Transformation der Datenmatrix \mathbf{Z}

$$\mathbf{Z} = \mathbf{F}\mathbf{L}' \quad \text{und} \quad \mathbf{R} = \mathbf{L}\mathbf{L}',$$

mit $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_p)$, $\mathbf{L} = (\mathbf{L}_1, \dots, \mathbf{L}_p)$ und $\mathbf{L} = \mathbf{T}\boldsymbol{\Lambda}^{1/2}$, wobei $\mathbf{T} = (\mathbf{t}_1, \dots, \mathbf{t}_p)$, und \mathbf{t}_i der normierte Eigenvektor ist, der mit dem i -ten größten Eigenwert λ_i ($\lambda_i \geq 0$) der empirischen Korrelationsmatrix \mathbf{R} verbunden ist. Für $\boldsymbol{\Lambda}$ gilt: $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_p\}$.

Von der Gesamtvarianz $p = \text{tr}(\mathbf{R})$ der standardisierten Daten erklärt der i -te Faktor den Anteil λ_i ¹.

Statt allen Hauptkomponenten p werden nur die ersten k Hauptkomponenten ausgewählt. Somit ist die Modellschätzung durch

$$\mathbf{Z} = \mathbf{F}^{(k)}\mathbf{L}^{(k)'} + \mathbf{E}^{(k)} \tag{5.7}$$

gegeben, mit $\mathbf{F}^{(k)} = (\mathbf{F}_1, \dots, \mathbf{F}_k)$, $\mathbf{L}^{(k)} = (\mathbf{l}_1, \dots, \mathbf{l}_k)$ und $\mathbf{E}^{(k)} = \mathbf{F}_{k+1}\mathbf{l}_{k+1}' + \dots + \mathbf{F}_p\mathbf{l}_p'$.

Die geschätzten Einzelrestvarianzen sind durch die Diagonalelemente der Matrix $\mathbf{E}^{(k)'}\mathbf{E}^{(k)} = \mathbf{R} - \mathbf{L}^{(k)}\mathbf{L}^{(k)'}$ gegeben.

¹Wegen $\text{tr}[(\mathbf{F}_i\mathbf{l}_i')'(\mathbf{F}_i\mathbf{l}_i')] = \text{tr}(\mathbf{l}_i\mathbf{l}_i') = \sum_{j=1}^p l_{ji}^2 = \lambda_i$.

Die Anzahl von Hauptkomponenten, die als Faktoren für das Modell extrahiert werden, werden häufig aufgrund eines der folgenden drei Kriterien bestimmt:

1. *Eigenwertkriterium*: Nach diesem Kriterium werden alle Hauptkomponenten extrahiert, die den Eigenwert $\lambda_i \geq 1$ besitzen, um Komponenten zu erhalten, die mindestens die Varianz 1 der ursprünglichen Variablen erklären.
2. *Prozentsatzkriterium*: Nach diesem Kriterium werden so viele Komponenten einbezogen, bis ein bestimmter subjektiv festgelegter Prozentsatz der Gesamtvarianz erklärt wird.
3. *Signifikanztest der Hauptkomponenten* nach BARTLETT (1951): Unter der Normalitätsannahme der Datenvariablen wird die Nullhypothese H_0 : *k Faktoren genügen* gegen die alternative Hypothese H_1 : *mehr als k Faktoren werden benötigt* getestet. Dafür lautet die Teststatistik für $0 < k < p - 1$:

$$T_k = (N - 1)[- \ln |\mathbf{R}| + \ln(\lambda_1 \cdot \dots \cdot \lambda_k) + (p - k) \ln \lambda] \quad (5.8)$$

mit $\lambda = (p - \lambda_1 - \dots - \lambda_k)/(p - k)$. Die Statistik T_k ist approximativ χ^2 -verteilt mit d_k Freiheitsgraden, wobei $d_k = \frac{1}{2}(p - k + 2)(p - k - 1)$.

Die Einzelrestfaktoren des Modells der FA spielen keine Rolle bei der HKM. Außerdem ist die Hauptkomponentenmethode lediglich eine reine Approximierungsmethode für die empirische Korrelationsmatrix \mathbf{R} (siehe BRACHINGER/OST (1996), S. 668).

Die zweite Methode zur Schätzung des faktorenanalytischen Modells ist die von THOMSON (1934) (vgl. ANDERSON (1984), S. 568) entwickelte Methode der *Hauptfaktorenanalyse* (HFA).

Sei die aufgrund der Grundgleichung (5.3) des faktorenanalytischen Modells (5.1) entsprechende Zerlegung der empirischen Korrelationsmatrix

$$\mathbf{R} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}, \quad (5.9)$$

wobei $\mathbf{\Psi} = \text{diag}\{\psi_1^2, \dots, \psi_p^2\}$ eine Diagonalmatrix ist, deren Hauptdiagonaleelemente die empirischen Einzelrestvarianzen sind. Bei der HFA werden die folgenden Schritte durchgeführt:

1. Anhand einer Schätzung \hat{h}_i^2 der Kommunalitäten werden die Einzelrestvarianzen ψ_i^2 bestimmt. Dabei wird die Definition $h_i^2 := 1 - \psi_i^2$ verwendet, welche auf die Zerlegung $\sigma_{ii} = h_i^2 + \psi_i^2$ für die Varianz σ_{ii} der standardisierten Daten basiert. Wie vorher gesehen wurde, ergibt sich diese Zerlegung aus die Grundgleichung (5.3) des faktorenanalytischen Modells auf der Seite 74. Darüber hinaus wird berücksichtigt, daß die Varianz standardisierter Daten Eins ist. Dann erhält man die *reduzierte Korrelationsmatrix* \mathbf{R}_h

$$\mathbf{R}_h = \mathbf{R} - \mathbf{\Psi}. \quad (5.10)$$

2. Aufgrund der Eigenwerte $\lambda_1 \geq \dots \geq \lambda_p$ und den assoziierten normierten Eigenvektoren $\mathbf{t}_1, \dots, \mathbf{t}_p$ von \mathbf{R}_h erhält man $\mathbf{L} = \mathbf{T}\mathbf{\Lambda}^{1/2}$, mit $\mathbf{T} = (\mathbf{t}_1, \dots, \mathbf{t}_p)$ und $\mathbf{\Lambda} = \text{diag}\{\lambda_1 \dots \lambda_p\}$. Dann gilt $\mathbf{R} = \mathbf{L}\mathbf{L}'$ und die Hauptfaktoren $\mathbf{F} = (\mathbf{F}_1, \dots, \mathbf{F}_p)$ berechnen sich aus $\hat{\mathbf{F}} = \mathbf{Z}\mathbf{R}_h^{-1}\hat{\mathbf{L}}$.

Wie bei der HKM wird nur eine kleine Anzahl k ($k < p$) von Hauptfaktoren genommen. Die Modellschätzung für die ersten k Hauptfaktoren lautet dann

$$\mathbf{Z} = \mathbf{F}^{(k)}\mathbf{L}^{(k)'} + \mathbf{E}^{(k)}. \quad (5.11)$$

Die empirische Korrelationsmatrix ist durch $\mathbf{R} = \mathbf{L}^{(k)}\mathbf{L}^{(k)'} + \mathbf{L}^{(p-k)}\mathbf{L}^{(p-k)'} + \mathbf{\Psi}$ gegeben. Auch wenn die Matrix \mathbf{R} im allgemeinen nicht diagonal ist, läßt sie sich mit dieser Methode besser als bei der HKM schätzen, weil die Matrix $\mathbf{L}^{(p-k)}\mathbf{L}^{(p-k)'} + \mathbf{\Psi}$ näher bei einer Diagonalmatrix ist (in BRACHINGER/OST (1996), S. 672).

5.3.1.2 Rotation und Interpretation

Das Ziel der Rotation besteht in der Suche nach Faktoren, die eine möglichst direkte, inhaltliche Interpretation erlauben (vgl. BRACHINGER/OST (1996)). Eine möglichst klare Interpretation der Faktoren ist von Vorteil, falls eine Erklärung der Einflüsse auf die explikativen Variablen \mathbf{X} benötigt wird.

Die in der Grundgleichung (5.3) des faktorenanalytischen Modells (5.1) vorgestellte Ladungsmatrix \mathbf{L} ist nur eindeutig bis auf eine orthogonale Transformation bzw. Rotation. Dies stellt das theoretische Argument für eine orthogonale Rotation der geschätzten Ladungsmatrix $\mathbf{L}^{(k)}$ dar. Die Rotation sollte so durchgeführt werden, daß die geschätzte Ladungsmatrix $\mathbf{L}^{(k)}$ nach der Rotation besser interpretiert werden kann.

Eine orthogonale Transformation ist nicht anderes als eine orthogonale Rotation des Koordinatenkreuzes (der Faktoren in der FA), so daß sich die Position jeder Koordinatenachse in die Richtung einer bestimmten Gruppe von Faktorladungen orientiert. Gesucht wird die Ladungsmatrix $\tilde{\mathbf{L}}$ nach einer orthogonalen Rotation,

$$\tilde{\mathbf{L}} = \mathbf{L}\mathbf{T}, \quad (5.12)$$

wobei \mathbf{T} eine orthonormale Transformationsmatrix ist, mit $\mathbf{T}'\mathbf{T} = \mathbf{I}$.

Von allen möglichen Rotationen werden diejenigen vorgezogen, die die Struktur der Ladungen einfacher machen, d.h. daß möglichst jede Variable hoch auf einem Faktor einlädt und kleinere Ladungen in den restlichen Faktoren besitzt. Es gibt eine Vielzahl von Kriterien, die eine angemessene Struktur der Ladungsmatrix verursachen. Ein Überblick findet sich etwa bei HARMAN (1976), Teil II, BRACHINGER/OST (1996), S. 679 f. und GORSUCH (1983), Kapitel 9–10. Hier wird nur auf eine der orthogonalen Methoden eingegangen, so daß die neuen Faktoren orthogonal zueinander werden: die Varimax-Methode. Die von KAISER (1958) entwickelte *Varimax-Methode* besteht darin, die orthogonale Transformation $\mathbf{T} \in \mathbb{R}^{k \times k}$ auszuwählen, die die Varianz der Ladungsquadrate maximiert. Diese Maximierung führt dazu, daß die entsprechenden Ladungswerte zu einem bestimmten Faktor sehr hoch (bis zu 1) oder sehr klein (bis zu 0) werden. Sei $\tilde{\mathbf{L}} := (\tilde{l}_{ij})$ die Ladungsmatrix der transformierten Faktoren. Das Kriterium lautet

$$\text{Max}_{\tilde{l}_{ij}} V = \text{Max}_{\tilde{l}_{ij}} \frac{1}{p} \sum_{j=1}^k \left(\sum_{i=1}^p \tilde{l}_{ij}^2 - m_i \right)^2, \quad (5.13)$$

wobei $m_i := \frac{1}{p} \sum_{i=1}^p \tilde{l}_{ij}^2$ der Mittelwert der Ladungsquadrate des j -ten Faktors ist. Die Maximierung wird iterativ durchgeführt (siehe die eben erwähnte Literatur für Details).

Die Interpretation der rotierten Faktoren basiert auf der transformierten Ladungsmatrix $\tilde{\mathbf{L}}$, deren Elemente \tilde{l}_{ij} Schätzungen der für die Kovarianzen der Faktoren mit den beobachtbaren Variablen (oder die Korrelationen im Fall von standardisierten beobachtbaren Variablen) stellen. Da Korrelationen erst ab eine bestimmte Größe “signifikant” sind, werden nur zur Interpretation der Faktoren diejenigen Ladungen gezogen, die eine vorgegebene Grenze von z.B. 0.5 überschreiten. Diese zu einem Faktor gehörigen Hochladungen bestimmen die Variablen, die gemeinsam durch diesen Faktor beschrieben sind. Anhand der Bedeutung dieser Gruppe von Variablen kann der Faktor interpretiert und benannt werden.

5.3.1.3 Berechnung der Faktorenwerte

Nach der Bestimmung der Erwartungswerte $\boldsymbol{\mu}$ der beobachtbaren Variablen, der Anzahl der Faktoren k , der Ladungsmatrix \mathbf{L} und der Matrix $\boldsymbol{\Psi}$ der Einzelrestvarianzen für das faktorenanalytischen Modells

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L} \mathbf{f} + \boldsymbol{\varepsilon}, \quad (5.14)$$

sind die Faktoren \mathbf{f} zu schätzen. Zur Schätzung der Faktorenwerte stehen zwei Methoden zur Verfügung: Eine Methode ist unter dem Namen *Gewichtete Kleinste-Quadrate-Methode* bekannt. Diese Methode wurde von BARLETT (1937, 1938a, 1938b) vorgestellt. Die zweite Methode ist die *Regressionsmethode* von THOMSON (1951), in der die geschätzten Faktorenwerte die Ergebnisse einer durchgeführten Regression der Faktoren auf die Datenvariablen sind (eine ausführliche Beschreibung beider Methoden findet sich in BRACHINGER/OST (1996)).

Bei beiden Methoden wird angenommen, daß die geschätzten Ladungen \tilde{l}_{ij} , $i = 1, \dots, p$ und $j = 1, \dots, k$, und die geschätzten Einzelrestvarianzen $\hat{\psi}_i^2$ wahre Werte und nicht Schätzungen sind. Weiter wird davon ausgegangen, daß die Faktorenwerte als Linearkombinationen der Datenvariablen dargestellt werden können (vgl. JOHNSON/WICHERN (1992)).

Nach einer HKM mit einer Anzahl gemeinsamer Faktoren k und nach Rotation wird die Matrix $\tilde{\mathbf{F}}^{(k)}$ von Faktorenwerten durch

$$\tilde{\mathbf{F}}^{(k)} = \mathbf{Z}\mathbf{R}^{-1}\tilde{\mathbf{L}}^{(k)}\mathbf{\Psi} \quad (5.15)$$

geschätzt, wobei \mathbf{Z} die $(n \times p)$ -dimensionale Matrix der standardisierten Datenvektoren $\mathbf{Z}_1, \dots, \mathbf{Z}_p \in \mathbb{R}^n$ ist, \mathbf{R} ist die empirische Korrelationsmatrix der Daten und $\mathbf{\Psi} := \text{cov}(\mathbf{f}) = \mathbf{T}^{-1}\mathbf{T}^{-1'}$ die Kovarianzmatrix der orthogonalen Faktoren. Analog dazu wird für den Fall einer HFA die reduzierte Korrelationsmatrix \mathbf{R}_h in (5.10) statt der Korrelationsmatrix \mathbf{R} betrachtet.

5.3.2 Anwendung von Faktorenanalyse auf die Darmstadt-Daten

In diesem Abschnitt wird auf die Anwendung der Faktorenanalyse auf Daten aus Darmstadt 1994/95 eingegangen. Aufgrund der großen Anzahl von Variablen im Datensatz, die ähnliche Information beinhalten und daher eine hohe Korrelation aufweisen, erweist sich eine Reduktion der Anzahl dieser Variablen auf wenige voneinander unabhängigen hypothetischen Faktoren als vorteilhaft. Somit wird das Multikollinearitätsproblem vermieden.

Tabelle 5.1 faßt die Anzahl ursprünglicher Variablen im Datensatz aus Darmstadt 1994/95 für jedes Wohnungsmerkmal und die nach der Faktorenanalyse resultierenden Faktoren zusammen. Nähere Ausführungen dazu finden sich in den folgenden Abschnitten.

Wohnungsmerkmal	Anzahl Variablen	Anzahl Faktoren
<i>Gebäudeart</i>	10	4
<i>Beschaffenheit</i>	16	6
<i>Lage</i>	13	5
<i>Ausstattung</i>	44	15
<i>Sondereinrichtung</i>	16	6

Tabelle 5.1. Anzahl Variablen und Faktoren vor und nach Faktorenanalyse.

5.3.2.1 Ergebnisse der Faktorenanalyse

Faktorenanalyse der Gebäudeart-Variablen

Um die Anzahl der die Gebäudeart betreffenden Variablen zu reduzieren, wurde eine Faktorenanalyse mit den beiden Methoden Hauptkomponentenmethode (HKM) und Hauptfaktorenanalyse (HFA) durchgeführt². Beide Ergebnisse sind in Tabelle 5.2 dargestellt, wo jeweils die Elemente der Ladungsmatrix $\tilde{\mathbf{L}}$ nach einer Rotation mit dem Varimax-Kriterium angegeben sind. Nach dem Prozentsatzkriterium wurden vier Faktoren extrahiert, um einen erklärten Anteil der Gesamtvarianz der Variablen von mindestens 50% zu erreichen. Diese Entscheidung wird durch die Ergebnisse des Chi-Square-Tests nach BARLETT (1951) unterstützt: mit einer Wahrscheinlichkeit (p-Wert) von 0.56 konnte die Nullhypothese $H_0 : 4 \text{ Faktoren genügen}$ gegenüber der alternativen Hypothese $H_1 : \text{mehr als } 4 \text{ Faktoren werden benötigt}$ nicht abgelehnt werden. Wenn die Nullhypothese lautet $H_0 : 3 \text{ Faktoren genügen}$, wird sie abgelehnt.

Als Hochladungen wurden die nach der HKM geschätzten Elemente der Ladungsmatrix definiert, die größer als 0.5 waren. Die Elemente der auf die HFA basierenden reduzierten Korrelationsmatrix \mathbf{R}_h zeigen kleinere Werte als diejenigen der auf eine

²Zum Vergleich wurden auch Faktorenanalysen nach der Maximum-Likelihood-Methode durchgeführt (für eine Einführung in die Maximum-Likelihood-Methode siehe etwa JOHNSON/WICHERN (1992), S. 411-416 und ANDERSON (1984), S. 557–563 für eine theoretische Beschreibung). Die Verteilungsannahmen dieser Methode und die möglichen Konvergenzprobleme (siehe etwa SAS INSTITUTE INC. (1990)) haben die Methode für unsere Analysen ungeeignet gemacht.

Rotierte geschätzte Ladungen					
Variable	HKM				\hat{h}_i^2
	F_1	F_2	F_3	F_4	
<i>Gebhoehe</i>	<u>0.8099</u>	0.1441	0.0097	−0.1972	0.7157
<i>Hinthauss</i>	−0.0822	<u>−0.60881</u>	−0.0409	0.0655	0.3834
<i>Bstgtren</i>	−0.1102	<u>0.7696</u>	0.1606	0.1066	0.6415
<i>Awgruen</i>	0.1172	0.1391	−0.1819	<u>0.7675</u>	0.6552
<i>Aufzug</i>	<u>0.7763</u>	−0.00178	0.4068	0.0201	0.7685
<i>Stuffrei</i>	<u>0.7516</u>	−0.0778	−0.0859	0.0890	0.5862
<i>Laubgang</i>	−0.3690	0.0274	<u>0.6637</u>	−0.0711	0.4538
<i>Gebtyp</i>	0.1389	0.2058	−0.0417	<u>−0.5839</u>	0.4043
<i>Whgzahl</i>	0.2771	0.0647	<u>0.7127</u>	−0.0245	0.5895
<i>Geb120s</i>	0.0682	<u>0.5590</u>	−0.3155	−0.3539	0.5419

Variable	HFA				\hat{h}_i^2
	F_1	F_2	F_3	F_4	
<i>Gebhoehe</i>	<u>0.6852</u>	0.0786	0.1571	0.1695	0.5291
<i>Hinthauss</i>	−0.0729	−0.0334	<u>−0.3372</u>	−0.0761	0.1253
<i>Bstgtren</i>	−0.0568	0.0704	<u>0.3880</u>	0.1604	0.1624
<i>Awgruen</i>	0.0364	−0.0913	−0.0612	−0.1517	0.0364
<i>Aufzug</i>	<u>0.6724</u>	<u>0.4153</u>	0.0139	−0.0295	0.6256
<i>Stuffrei</i>	<u>0.5451</u>	0.0118	−0.0323	−0.0564	0.3015
<i>Laubgang</i>	0.0121	0.2438	0.0014	0.0507	0.0621
<i>Gebtyp</i>	0.0918	0.0364	0.1831	<u>0.2470</u>	0.1043
<i>Whgzahl</i>	0.2114	<u>0.5001</u>	0.0441	0.0281	0.2976
<i>Geb120s</i>	0.0426	−0.1498	<u>0.3720</u>	<u>0.2602</u>	0.2304

Tabelle 5.2. Faktorenanalyse der Gebäudeart-Variablen.

HKM basierenden Korrelationsmatrix \mathbf{R} an. Damit sind die Eigenwerte und die Ladungen auch kleiner. Deswegen wurde die Grenze für Hochladungen bei der HFA auf 0.3 heruntergesetzt.

Mittels der HFA und nach der Rotation der geschätzten Ladungsmatrix können die folgenden Faktoren bestimmen werden:

1. Der Faktor 1 oder *Art des Gebäudes bezüglich der Höhe* besitzt hohe Ladungen in den Variablen *Gebhoehe*, *Aufzug* und *Stuffrei*, deshalb wurden diese Variablen durch diesen Faktor zusammengefaßt.
2. Faktor 2 lädt in den Variablen *Aufzug* und *Whgzahl* hoch und kennzeichnet den *Zusammenhang zwischen dem Vorhandensein eines Aufzugs und der Anzahl der Wohnungen in jedem Stock*.
3. Faktor 3 lädt in den Variablen *Hinthaus*, *Bstgtren* und *Geb120s* und erklärt die Lage des Gebäudes bezüglich der anderen Gebäude. Deshalb wurde dieser Faktor *Nähe des Gebäudes zu anderen Gebäuden* genannt.
4. Faktor 4 lädt in den Variablen *Geb120s* und *Gebtyp* hoch. Dieser Faktor faßt die zwei Variablen zusammen, die die *Art des Gebäudes* betreffen.

Weitere Faktorenanalysen nach der HKM und nach der Maximum-Likelihood-Methode anhand zweier verschiedenen Unterstichproben (aus der Einteilung aller Beobachtungen der Gesamtstichprobe in zwei Stichproben gleiches Umfangs) zeigen, daß die Faktoren 1 und 2 stabil sind; dies gilt aber nicht für die beiden anderen Faktoren. Diese Nichtstabilität der Faktoren wirkt sich negativ auf die Interpretation der Faktoren aus: Kleine Änderungen in der Varianz der Faktoren verursachen Änderungen in den geschätzten Ladungen, die die Interpretation der Faktoren schwerer macht. Für eine ausführliche Analyse über die Stabilität von Faktoren (oder Hauptkomponenten) wird etwa auf JOLLIFFE (1986) und KRZANOWSKI (1984) verwiesen.

Faktorenanalyse der Lage-Variablen

Bei der Faktorenanalyse wurden nach dem Eigenwertkriterium fünf Faktoren beibehalten, um über 50% der Gesamtvarianz der Variablen zu erklären. In Tabelle 5.3 werden die Ergebnisse durch die HFA nach der Varimax-Rotation dargestellt, die wie folgt interpretiert werden können:

1. Der Faktor 1 lädt auf die Variablen *Kfzpark0*, *Indlärm* und *Wald* mit über 0.25. Aufgrund von verschiedenen Vorzeichen zeigt der Faktor einen Kontrast zwischen Lärm der Industrie und Ruhe und wurde als *Lärmbelastung* bezeichnet.
2. Der Faktor 2 besitzt hohe Ladungen in den Variablen *Temp30* und *Vklaerm* und betrifft den Verkehr, deshalb wurde er als *Verkehrsbelastung* gekennzeichnet.
3. Der Faktor 3 umfaßt die Variablen *Park*, *Bus* und *Geschaeft* und wurde als *Infrastruktur der Stadt* bezeichnet.
4. Faktor 4 zeigt den Zusammenhang zwischen der Aussicht der Wohnung und der Etage, in der die Wohnung sich befindet, weil er in die Variablen *Etage* und *Aussicht* mit hohen Ladungen lädt. Er wurde *Aussicht der Wohnung* genannt.
5. Faktor 5 umfaßt die Variablen *Zebra0* und *Bstgpark* und betrifft die *Sicherheitsbedingungen für die Passanten bezüglich des Verkehrs*.

Die Ergebnisse nach der HKM führen zu sehr ähnlichen Ergebnissen.

Faktorenanalyse der Beschaffenheits-Variablen

Nach dem Eigenwertkriterium erhält man sechs Faktoren, um einen Anteil von ca. 50% der Gesamtvarianz der Variablen zu erklären. Die Ergebnisse der Faktorenanalyse nach der Hauptfaktorenanalyse und nach einer Varimax-Rotation sind in Tabelle 5.4 dargestellt:

1. Faktor 1 lädt mit hohen Ladungen (über 0.23) in den Variablen *Qluft0*, *Gefraum* und *Wflkwr*. Dieser Faktor betrifft *kleine Räume und ihre Besonderheiten*, wie

Variable	Rotierte geschätzte Ladungen					\hat{h}_i^2
	F_1	F_2	F_3	F_4	F_5	
<i>Etage</i>	0.0720	-0.0544	0.0376	<u>0.3704</u>	-0.0290	0.1476
<i>Aussicht</i>	-0.1195	0.0517	-0.0944	<u>0.3841</u>	0.0400	0.1750
<i>Licht0</i>	-0.0328	0.0536	-0.1058	0.0009	0.1820	0.0483
<i>Temp30</i>	-0.0743	<u>0.4391</u>	0.1025	0.0103	0.0419	0.2107
<i>Kfzpark0</i>	<u>0.4854</u>	-0.1257	0.2348	-0.0508	0.0178	0.3094
<i>Zebra0</i>	0.1092	-0.2281	0.1185	0.0136	<u>0.3447</u>	0.1833
<i>Bstgpark</i>	0.2391	-0.0184	0.0723	-0.0400	<u>0.2557</u>	0.1297
<i>Vklaerm</i>	-0.2246	<u>0.4875</u>	0.0208	-0.0594	-0.1816	0.3296
<i>Indlaerm</i>	<u>0.2995</u>	-0.0884	-0.0467	0.0395	0.1128	0.1140
<i>Wald</i>	<u>-0.4005</u>	0.2673	-0.1272	0.0572	0.1409	0.2712
<i>Park</i>	0.1067	0.0259	<u>0.3866</u>	0.0061	-0.0990	0.1713
<i>Bus</i>	-0.0754	0.0779	<u>0.3502</u>	-0.0532	-0.0062	0.1373
<i>Geschaef</i>	0.1413	0.0003	<u>0.4177</u>	0.0013	0.0277	0.1952

Tabelle 5.3. Hauptfaktorenanalyse der Lage-Variablen.

Variable	Rotierte geschätzte Ladungen						\hat{h}_i^2
	F_1	F_2	F_3	F_4	F_5	F_6	
<i>Woeb</i>	-0.0262	-0.0114	0.0124	0.0183	0.0209	<u>0.2281</u>	0.0538
<i>Qluft0</i>	<u>0.3353</u>	-0.0992	-0.1468	<u>-0.2469</u>	0.0062	-0.0499	0.2073
<i>Gefraum</i>	<u>-0.3303</u>	-0.0068	-0.0445	0.0199	-0.0606	0.0593	0.1187
<i>Wdiele</i>	-0.0316	0.1196	<u>0.3437</u>	-0.0036	0.0276	0.0758	0.1400
<i>Spkam</i>	-0.0973	0.0023	0.1550	0.2080	-0.0946	0.1364	0.1043
<i>Hwraum</i>	-0.0917	0.1765	0.0379	0.0174	0.0091	0.0464	0.0435
<i>Wrhoehe0</i>	-0.0206	-0.0099	-0.0034	<u>-0.2757</u>	-0.0868	-0.0179	0.0844
<i>Tschwel</i>	-0.1448	<u>-0.2478</u>	0.00006	0.0893	<u>-0.2527</u>	0.0431	0.1561
<i>Tbreite</i>	0.0027	0.0592	<u>0.2564</u>	0.0093	-0.0737	-0.0300	0.0756
<i>Wflkur</i>	<u>0.5490</u>	0.0535	-0.0619	0.0735	-0.0190	0.0208	0.3143
<i>Wflgur</i>	0.2280	<u>0.3491</u>	0.2224	<u>0.2914</u>	0.1346	0.1947	0.3642
<i>Whglaut</i>	-0.0920	-0.2095	-0.0942	0.1654	0.0443	0.0096	0.0907
<i>Balkz2</i>	-0.0361	0.0640	<u>0.2733</u>	0.1088	0.1982	0.1705	0.1603
<i>Balkb2</i>	-0.0141	0.2151	0.1693	0.1094	0.1105	<u>0.2294</u>	0.1519
<i>Balkbrn</i>	-0.0140	0.0417	-0.0336	0.1015	<u>0.2727</u>	0.0570	0.0910
<i>Erzust</i>	-0.0256	<u>0.3691</u>	0.1232	0.0508	0.0886	-0.0617	0.1663

Tabelle 5.4. Hauptfaktorenanalyse der Beschaffenheits-Variablen.

z.B., daß alle Fenster auf einer Seite des Zimmers stehen oder daß es sich um ein Durchgangszimmer handelt.

2. Faktor 2 besitzt hohe Ladungen in den Variablen *Tschwel*, *Wflgwr* und *Erzust*, was einen *Kontrast zwischen dem Vorhandensein von Türschwellen und großen Räumen* (beide mit negativem Vorzeichen) *und dem Gesamterhaltungszustand des Gebäudes* zeigt.
3. Faktor 3 umfaßt so unterschiedliche Variablen wie *Wdiele*, *Tbreite* und *Balkz2* und wurde als *Zusatzkomfort* bezeichnet.
4. Faktor 4 zeigt einen *Kontrast zwischen niedrigen Räumen (Höhe $\leq 2.40m$) und Luftdurchgang* (beide mit negativem Vorzeichen) *und großen Räumen*, weil er hohe Ladungen in den Variablen *Qluft0*, *Wrhoehe0* und *Wflgwr* besitzt.
5. Faktor 5 lädt mit hohen Ladungen in die Variablen *Tschwel* und *Balkbrn*. Diese letzte Variable steht im Gegensatz zur Variable *Tschwel*, was eine mögliche Abhängigkeit in der Bauart zeigen könnte. Dieser Faktor wurde als *Balkonorien-tierung und Vorhandensein von Türschwellen* bezeichnet.
6. Faktor 6 besitzt hohe Ladungen in die Variablen *Woeb* und *Balkb2* und wurde mit *Zusammenhang zwischen den durch kleine Stufen/Absätze miteinander verbundenen Wohnräume und Tiefe/Breite des Balkons* gekennzeichnet.

Wenn man dies mit den Ergebnisse einer HKM vergleicht, sind die Faktoren 2 und 4 nicht gleich, was gegen die Stabilität dieser Faktoren spricht.

Faktorenanalyse der Sondereinrichtungs-Variablen

Für die Variablen, die dieses Merkmal kennzeichnen, wurde eine Hauptfaktorenanalyse durchgeführt. In Tabelle 5.5 sind die Ergebnisse nach einer Varimax-Rotation dargestellt.

Nach dem Eigenwertkriterium wurden sechs Faktoren extrahiert, die mehr als 50% der Gesamtvarianz der Variablen erklären.

Variable	Rotierte geschätzte Ladungen						\hat{h}_i^2
	F_1	F_2	F_3	F_4	F_5	F_6	
<i>Waschma</i>	<u>0.7557</u>	0.1135	0.0409	0.0778	0.1165	0.0658	0.7766
<i>Waetrock</i>	<u>0.7076</u>	0.0155	0.0079	-0.0120	0.1446	0.0519	0.7123
<i>Waschkue</i>	<u>0.4205</u>	-0.0249	0.0847	<u>0.3667</u>	0.0494	0.0885	0.5993
<i>Troraum</i>	0.1496	0.0214	0.0656	<u>0.3705</u>	0.0096	<u>0.3744</u>	0.5570
<i>Trockplz</i>	0.0441	0.0691	0.2201	0.2696	-0.0191	0.1849	0.6509
<i>Radkel</i>	0.0713	0.0270	0.0586	0.0957	0.0737	<u>0.3634</u>	0.6399
<i>Hobraum</i>	0.0966	0.0369	0.2412	0.1402	<u>0.3979</u>	-0.0916	0.6368
<i>Partraum</i>	0.1559	0.1408	0.0481	-0.0388	<u>0.5415</u>	0.0896	0.7596
<i>Sauna</i>	0.0734	<u>0.5895</u>	-0.0914	-0.0931	0.1856	0.1092	0.8075
<i>Schwibad</i>	0.0158	<u>0.5660</u>	0.1244	0.0641	0.0071	-0.0750	0.7796
<i>Spielplz</i>	0.0224	-0.0252	<u>0.2611</u>	0.0675	-0.0858	<u>0.3366</u>	0.6342
<i>Grillplz</i>	0.0697	0.0124	<u>0.4495</u>	-0.0061	0.2322	0.0399	0.6613
<i>Keller</i>	-0.0291	-0.0955	0.0093	<u>0.2688</u>	-0.0493	0.2029	0.8098
<i>Speicher</i>	-0.1285	-0.0012	0.1260	0.2057	0.0644	0.1612	0.6093
<i>Garten</i>	-0.0149	0.0346	<u>0.4293</u>	0.1109	0.0397	0.1581	0.5523
<i>Garage</i>	0.0506	0.0118	0.0220	<u>0.2882</u>	0.0295	0.0176	0.7963

Tabelle 5.5. Hauptfaktorenanalyse der Sondereinrichtungs-Variablen.

Gemäß der nach der Varimax-Rotation geschätzten Ladungsmatrix wurden die folgenden Faktoren bestimmt:

1. Faktor 1 lädt in den Variablen *Waschma*, *Waetrock* und *Waschkue* höher als 0.26. Diese Variablen können unter den Begriff *Waschmöglichkeiten* zusammengefaßt werden.
2. Faktor 2 lädt in den Variablen *Sauna* und *Schwibad*. Dieser Faktor wurde als *Wellness-Möglichkeiten* bezeichnet.
3. Der Faktor 3 besitzt hohe Ladungen in den Variablen *Spielplz*, *Grillplz* und *Garten*, deshalb wurden diese Variablen durch diesen Faktor zusammengefaßt. Dieser Faktor wurde mit *Garten/Gartennutzung* bezeichnet.
4. Faktor 4 lädt in den Variablen *Waschkue*, *Troraum*, *Keller* und *Garage*. Dieser Faktor wurde als *zusätzliche Infrastruktur im Haus* bezeichnet.
5. Faktor 5 umfaßt die Variablen *Hobraum* und *Partraum* mit hohen Ladungen und bestimmt *spezielle zusätzliche Räume*.
6. Faktor 6 besitzt hohe Ladungen in den Variablen *Troraum*, *Radkel* und *Spielplz* und wurde *Benutzung von Räumen in unteren Geschossen* genannt.

Diese Ergebnisse sind nach einer HKM sehr ähnlich.

Faktorenanalyse der Ausstattungs-Variablen

Dies stellt eine sehr heterogene Gruppe von Variablen dar, wobei Merkmale der Heizung und Warmwasserversorgung, von Bad und Toilette, Küche und sonstiger Ausstattung zusammengefaßt sind. Die Schwierigkeit bei dieser Gruppe ergibt sich aus ihrer Heterogenität und besteht darin, eine plausible Interpretation der Faktoren zu finden, weil diese Faktoren meistens verschiedene Gegenstände umfassen, d.h. daß die Faktoren Variablen umschließen, die z.B. die Küche und auch das Bad betreffen. Trotz den möglichen Interpretationsschwierigkeiten erweist sich eine Faktorenanalyse als geeignet, um die hohe Dimension der Daten zu reduzieren.

Nach dem Eigenwertkriterium wurden fünfzehn Faktoren extrahiert, die 58% der Gesamtvarianz der originalen Variablen erklären. Die Faktorenanalyse wurde nach der Hauptfaktorenanalyse durchgeführt und ihre Ergebnisse sind in Tabelle 5.6 dargestellt. Eine Interpretation der Ladungsmatrix führt aufgrund ihrer Hochladungen (denjenigen, die einen Wert von 0.38 oder höher betragen) zu den folgenden Faktoren:

1. Faktor 1 lädt mit hohen Ladungen in den Variablen *Ebaukues*, *Mikherd*, *Ggkomul*, *Dunstabz*, *Kuehlsch*, *Eisschra* und *Spuelma* und wurde als *Küchenausstattung* bezeichnet.
2. Faktor 2 besitzt hohe Ladungen in den Variablen *Bad0fen*, *Baddusch*, *Kuetyp* und *Schrakue*. Dieser Faktor faßt die *Art der Küche und des Bades* zusammen.
3. Faktor 3 lädt in den Variablen *Badl*, *Badb* und *Badb3* hoch und wurde *Grösse des Badezimmers* genannt.
4. Faktor 4 umfaßt die Variablen *Ebaukues*, *Elgaherd* und *Kuehlsch* und bezeichnet das *Vorhandensein eines Kühlschranks, modernen Einbauküchenschränken und einem Elektrogasherd in der Küche*.
5. Faktor 5 besitzt auch hohe Ladungen in den Variablen *Kuenutz*, *Moeb* und *Hratt*, was einen *Zusammenhang zwischen dem Vorhandensein von Hausrat/Geschirr, der Kochgelegenheit und dem Möblierungstyp der Wohnung* konstituiert.
6. Faktor 6 zeigt einen Kontrast zwischen der Variablen *Waermmes* und den Variablen *Hzaust0* und *Fbhz*. Dieser Faktor kann mit dem Begriff *Fußbodenheizung und Messung der Wärme* gekennzeichnet werden.
7. Faktor 7 umfaßt nur zwei Variablen: *Antenne* und *Kabel* und wurde *Fernsehmöglichkeiten* genannt.
8. Faktor 8 lädt mit hohen Ladungen in den Variablen *Badkach* und *Kuewaka*. Diese Variablen wurden unter dem Begriff *Kachelung von Bad und Küche* zusammengefaßt.

9. Faktor 9 faßt die Variablen *Baddusch* und *Badduab* zusammen und zeigt das *Vorhandensein Bad- und Duschmöglichkeiten*.
10. Faktor 10 lädt nur in einer Variable über 0.38, *Toisep*, und wurde *Vorhandensein einer Toilette separat von Bad* genannt.
11. Faktor 11 besitzt nur eine hohe Ladung in der Variable *Bklschra* und wurde als *begehrter Kleiderschrank* bezeichnet.
12. Faktor 12 lädt hoch in der Variable *Toi2x* und wurde als *Vorhandensein einer zweiten Toilette* benannt.
13. Faktor 13 lädt in den Variablen *Hztot* (unter 0.37) und *Ww*, die unter dem Begriff *Warmwasserversorgung und Beheizungsmöglichkeit in allen Zimmern* zusammengefaßt werden können.
14. Höchsten Ladungen des Faktors 14 entsprechen den Variablen *Elrolad* und *Spuele*, sind aber weit unter dem festgelegten Niveau von 0.37. Man kann diesen Faktor *Zusammenhang zwischen der Art der Spüle und dem Vorhandensein elektrischer Rolläden* nennen. Die Bedeutung dieses Zusammenhangs erscheint eher fraglich.
15. Faktor 15 besitzt auch nur eine Hochladung in der Variable *Hratv*, deshalb wurde als *Wohnung ist vollständig mit Hausrat/Geschirr ausgerüstet* bezeichnet.

Aus den 44 Variablen, die die Ausstattung der Wohnung kennzeichnen, wurden 15 unabhängige Faktoren extrahiert, die über 50% der Varianz der ursprünglichen Variablen erklären. Diese Reduktion ist insbesondere vorteilhaft bei der umfangreichen Anzahl von ähnlichen Variablen, die die Küche und das Bad betreffen.

Schlußbemerkungen zur Anwendung von Faktorenanalyse

In diesem Abschnitt wurde mittels der Faktorenanalyse auf die Reduktion der erklärenden Variablen oder Mietpreisdeterminanten eingegangen: Aus 99 Variablen, die häufig dieselbe Information enthalten haben, konnten 36 Faktoren isoliert werden, die untereinander unkorreliert sind.

Variable	Rotierte geschätzte Ladungen							F_8
	F_1	F_2	F_3	F_4	F_5	F_6	F_7	
HZTYPVM	-0.0680	-0.0199	-0.0136	0.0603	-0.1053	0.0414	0.2332	-0.1607
WAERMMES	-0.0715	0.0428	-0.0609	-0.0372	-0.0802	<u>-0.4631</u>	0.0751	0.2356
HZAUST0	-0.0236	0.0609	0.0304	0.0227	-0.0503	<u>0.5579</u>	0.0435	-0.0537
HZTOT	-0.0284	-0.0876	0.0797	-0.0631	-0.0043	0.0074	0.0804	0.1092
WWIN	-0.0490	-0.2358	-0.0487	-0.1477	-0.0351	0.2480	-0.0647	-0.1640
BAD0WABE	-0.0229	0.0695	-0.0278	-0.0657	0.3124	-0.0244	0.0186	0.0055
Bad0FEN	-0.0247	<u>0.4534</u>	-0.0466	0.0064	0.0098	-0.0252	-0.0501	0.1091
BADKACH	0.0139	0.0464	0.0037	-0.0120	0.0002	-0.0030	-0.0576	<u>0.3855</u>
BADDUSCH	0.0197	<u>0.4218</u>	-0.1991	-0.0032	0.2069	0.0809	-0.0322	0.0344
BADDUAB	0.0686	0.0601	0.0320	0.0521	-0.0038	-0.0224	0.0437	0.1075
BAD2WABE	0.0941	-0.0664	0.2548	-0.0719	0.0386	-0.0245	0.0221	0.1182
TOISEP	0.0593	-0.0751	-0.0614	-0.0237	0.0944	-0.0591	-0.0208	-0.0315
BADL	0.1353	-0.3674	<u>0.4350</u>	0.0003	-0.0527	0.0584	0.0732	-0.0941
BADB	0.0369	-0.0845	<u>0.7260</u>	0.0263	-0.0731	-0.0310	0.0518	0.0768
TOI2X	0.1213	-0.1569	0.1780	0.0234	-0.0763	0.0991	0.0596	0.1051
KUETYP	0.0054	<u>-0.5468</u>	0.1083	0.0018	-0.1505	-0.0167	0.0643	0.0546
SCHRAKUE	-0.0076	<u>0.6030</u>	-0.0650	-0.1156	0.0247	0.1170	-0.0979	-0.0442
EBAUKUES	<u>0.5109</u>	-0.0340	-0.0390	<u>0.4567</u>	0.1124	0.0805	0.0824	0.1084
ELGAHERD	0.3095	-0.1100	-0.0167	<u>0.6405</u>	0.1822	-0.0038	0.0177	0.0178
MIKHERD	<u>0.5148</u>	-0.0033	0.1640	-0.0195	0.1079	0.0261	0.0033	0.0614
GKKOMUL	<u>0.7318</u>	-0.0159	0.0534	-0.0385	-0.0124	-0.0447	-0.0313	-0.0399
DUNSTABZ	<u>0.6621</u>	0.0268	0.0214	0.2683	0.0048	0.0444	0.0782	0.1630
KUEHLSCH	<u>0.4547</u>	0.0085	-0.0006	<u>0.5419</u>	0.3127	-0.0053	-0.0519	0.0061
EISSCHRA	<u>0.6619</u>	-0.0410	0.0124	0.1432	-0.0154	-0.0091	-0.0202	0.0239
SPUELMA	<u>0.7759</u>	0.0120	0.1020	0.0796	0.0437	0.0519	0.0613	0.0386
KUENUTZ	0.0201	-0.0206	-0.0319	0.1379	<u>0.4191</u>	-0.0248	-0.0298	-0.1375
MOEB	0.0946	0.1130	-0.0368	0.1844	<u>0.4078</u>	0.0753	-0.0444	0.0354
ROLAD	-0.0065	-0.1440	0.0169	-0.0002	0.0274	0.0697	0.0839	0.1768
ELROLAD	0.0987	0.0393	0.0404	0.0822	-0.0493	-0.0459	0.0002	-0.0580
SPRECHAN	0.0340	0.2764	0.0738	-0.0566	-0.0731	0.0584	0.1226	0.2919
ANTENNE	-0.0486	0.0968	-0.0278	0.0239	-0.1055	-0.0366	<u>-0.5637</u>	0.0514
KABEL	0.0047	-0.0291	0.0064	0.0250	-0.1208	-0.0037	<u>0.6023</u>	0.1104
BKLSCHRA	0.1363	0.0034	-0.0315	0.0423	-0.0288	-0.0296	-0.0190	-0.0975
HOSTWADE	0.0102	-0.0692	0.0647	-0.0105	0.0237	-0.0206	0.0090	0.0783
PARKKACH	0.1037	-0.1276	0.0826	-0.0374	-0.0229	0.1303	0.0065	0.1435
FBHZ	0.0310	0.0583	-0.0387	-0.0323	-0.0287	<u>0.4466</u>	0.0719	0.2009
WW	0.0467	0.0153	-0.0225	0.0800	-0.0233	-0.0191	-0.0283	0.0749
BADB3	0.1172	-0.0441	<u>0.6665</u>	-0.0273	-0.0124	0.0576	-0.0321	0.0009
HRATT	0.0676	0.0650	-0.0148	0.0647	<u>0.4795</u>	-0.0227	-0.0069	0.0300
HRATV	0.1204	-0.0138	-0.0082	0.0306	-0.0040	-0.0429	-0.0036	0.0535
INSTP	-0.0157	-0.0898	0.0226	0.0211	-0.0245	0.1316	-0.0088	-0.2341
SPUELE	0.0890	0.1080	0.0545	-0.0603	0.0467	0.1559	0.0622	0.0634
KUEWAKA	0.1162	-0.0225	0.0213	0.1067	0.0083	-0.0920	0.0458	<u>0.3750</u>
ISOL	0.0604	0.0210	0.0221	0.0139	-0.1313	-0.0591	0.1568	0.1985

Tabelle 5.6. Hauptfaktorenanalyse der Ausstattungs-Variablen (Fortsetzung auf der nächste Seite).

Variable	Rotierte geschätzte Ladungen							\hat{h}_i^2
	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	
HZTYPVM	-0.0343	0.2246	-0.0182	0.0802	0.1191	-0.1585	-0.0038	0.1996
WAERMMES	-0.0486	-0.0748	-0.0057	-0.0167	-0.0755	0.0314	0.0126	0.3092
HZAUSTO	-0.0701	0.0042	-0.0294	-0.0056	-0.0380	-0.0630	0.0197	0.3359
HZTOT	-0.0595	-0.0082	-0.0173	0.0318	<u>0.3391</u>	0.1170	0.0483	0.1732
WWIN	0.0592	0.1006	-0.0398	-0.1580	-0.1644	-0.0035	-0.1040	0.2541
BAD0WABE	-0.0799	0.1267	0.0079	-0.0152	-0.1834	0.0859	0.0390	0.1743
BAD0FEN	-0.0184	0.0922	-0.1303	0.0124	-0.0785	-0.0516	0.0525	0.2610
BADKACH	0.0645	-0.0678	-0.0315	0.0653	0.0624	0.0072	0.0268	0.1731
BADDUSCH	<u>0.3910</u>	-0.0562	-0.0128	-0.1544	0.0094	-0.0255	-0.0516	0.4531
BADDUAB	<u>0.4355</u>	0.1130	0.0881	0.0251	0.0158	0.0159	0.0552	0.2404
BAD2WABE	0.0737	0.0816	0.2884	0.2545	0.0190	0.0158	-0.0755	0.2662
TOISEP	0.0359	<u>0.4373</u>	-0.0133	0.0053	-0.0957	0.0008	-0.0229	0.2298
BADL	0.0188	0.0033	0.2144	0.0256	0.0901	0.0157	0.0148	0.4185
BADB	-0.0278	-0.0876	0.0494	0.1653	0.0589	0.0456	0.0078	0.5951
TOI2X	-0.0018	-0.0211	0.0764	<u>0.4625</u>	0.0184	0.0467	0.0616	0.3283
KUETYP	0.0161	0.1775	0.0013	0.1329	0.0339	-0.0607	0.0739	0.4005
SCHRAKUE	0.1533	-0.2223	0.0885	-0.0583	0.0054	0.1784	-0.0455	0.5251
EBAUKUES	0.1146	0.0352	0.1033	-0.1099	0.0563	0.2305	0.0964	0.6127
ELGAHERD	-0.0465	-0.0073	-0.0166	0.0715	0.0628	0.0182	0.0049	0.5641
MIKHERD	0.0570	-0.0242	-0.1500	0.0036	-0.0152	-0.1354	-0.0522	0.3561
GKKOMUL	0.0279	-0.0049	0.0975	0.1257	-0.0023	-0.0375	0.1607	0.5982
DUNSTABZ	-0.0061	-0.0238	0.1793	-0.0950	-0.0106	0.1643	-0.0113	0.6153
KUEHLSCH	0.1626	-0.0091	0.0353	-0.0618	0.0122	0.1166	0.1082	0.6580
EISSCHRA	0.0565	0.0235	0.0523	0.0971	0.0194	0.0969	0.0602	0.4910
SPUELMA	-0.0046	0.0671	0.1189	0.0340	0.0465	0.1062	0.0888	0.6698
KUENUTZ	0.1396	0.0749	-0.0873	0.0400	-0.0020	-0.0634	-0.1053	0.2664
MOEB	-0.0391	-0.0684	0.1288	-0.1465	-0.0267	0.0083	0.2238	0.3275
ROLAD	0.0112	0.1949	0.1337	0.0930	0.1287	-0.0034	0.0924	0.1547
ELROLAD	0.0178	-0.0200	-0.0367	0.0285	0.0262	<u>0.2551</u>	0.0164	0.0965
SPRECHAN	0.1180	0.0036	0.0053	0.0646	0.0968	0.0047	0.0428	0.2245
ANTENNE	-0.0321	-0.0413	0.0157	-0.0248	0.0173	-0.0199	0.0290	0.3511
KABEL	0.0205	-0.0454	0.0133	0.0084	0.0502	0.0236	0.0312	0.3978
BKLSCHRA	-0.0262	-0.0426	<u>0.3593</u>	0.0306	-0.0350	-0.0347	0.0129	0.1681
HOSTWADE	0.1560	0.0208	0.2820	0.0054	0.0304	0.0064	-0.0021	0.1217
PARKKACH	0.1540	-0.0072	-0.1100	0.0275	0.0279	0.1251	0.1579	0.1514
FBHZ	0.0850	-0.1048	-0.0309	0.0910	-0.0107	0.0800	-0.0687	0.2931
WW	0.0642	-0.0694	0.0125	-0.0094	<u>0.4435</u>	-0.0355	-0.0234	0.2243
BADB3	0.0412	-0.0055	-0.0465	-0.0073	-0.0243	0.0274	-0.0045	0.4704
HRATT	0.0173	-0.0146	0.0077	-0.0050	0.0474	-0.0547	-0.0276	0.2512
HRATV	0.0316	-0.0112	-0.0020	0.0250	0.0089	0.0150	<u>0.3537</u>	0.1476
INSTP	0.1235	0.3000	-0.0096	-0.1129	-0.0933	-0.0819	-0.0436	0.2174
SPUELE	0.2718	-0.1304	0.1105	0.0228	0.0179	<u>0.2503</u>	0.0121	0.2273
KUEWAKA	0.0443	0.0358	0.0151	-0.0380	0.1309	-0.0719	0.0613	0.2082
ISOL	0.0639	-0.0882	0.0641	0.0202	-0.0771	-0.1155	-0.0393	0.1267

Die Hauptschwierigkeit bei der Anwendung der Faktorenanalyse bestand in unserem Fall darin, sinnvoll interpretierbare Faktoren aus der Analyse zu gewinnen. Auch nach der Rotation der Faktoren, die ihre Interpretation verbessern sollte, ist die Interpretation keine einfache Aufgabe, insbesondere mit einer derart großen Vielfalt von unterschiedlichen Variablen wie im Fall der Ausstattungskategorie. Darüber hinaus bleibt die Interpretation der Faktoren subjektiv und abhängig von der Rotationsmethode.

5.4 Schätzung des semiparametrischen Regressionsmodells für die Nettomiete

5.4.1 Problemstellung

Das additiv-multiplikative Modell (3.4) kann anhand eines multiplikativen semiparametrischen Ansatzes mit additiver Störvariable folgendermaßen dargestellt werden:

$$NM_i = q(WFL_i) \cdot (1 + \beta_1 z_{1i} + \dots + \beta_k z_{ki}) + \varepsilon_i, \quad (5.16)$$

wobei für die Störvariable ε_i die folgende Annahme gilt: $E(\varepsilon_i | WFL_i) = 0$, für $i = 1, \dots, n$.

In diesem Abschnitt wird auf die Schätzung dieses spezifischen Modells für die Nettomiete eingegangen. Nach einer geeigneten Transformation des Modells werden zuerst die bedingten Erwartungswerte der Nettomiete und der Mietpreisdeterminanten, die aufgrund der Transformation im Modell auftreten, mit Hilfe des nichtparametrischen Nadaraya-Watson-Kernschätzers geschätzt. Konfidenzintervalle für die geschätzten Erwartungswerte werden auch ermittelt. Folgend wird die Schätzung des parametrischen Teils des Modells betrachtet.

5.4.2 Transformation des Modells

Ebenso wie beim rein additiven Modell (3.5) in Kapitel 3.4 wird das additiv-multiplikative Modell (5.16) geeignet transformiert, so daß die unbekannte Funktion $q(WFL)$ keine Rolle bei der Schätzung der Nettomiete spielt. Im Gegensatz zu dem additiven Modell werden die bedingten Erwartungswerte $E(NM_i|WFL_i)$ und $E(\mathbf{z}_i|WFL_i)$ nicht durch Subtraktion in den Ansatz eingehen, sondern durch Division (vgl. ROBINSON (1988), S. 947). Das Modell lautet dann

$$\frac{NM_i}{E(NM_i|WFL_i)} = \frac{f(\mathbf{z}_i, \boldsymbol{\beta})}{E(f(\mathbf{z}_i, \boldsymbol{\beta})|WFL_i)} + \epsilon_i, \quad (5.17)$$

mit

$$\begin{aligned} f(\mathbf{z}_i, \boldsymbol{\beta}) &:= 1 + \mathbf{z}_i \boldsymbol{\beta} \\ \mathbf{z}_i &= (z_{1i}, \dots, z_{ki}) \\ \boldsymbol{\beta} &= (\beta_1, \dots, \beta_k)'. \end{aligned}$$

Für die Störvariable ϵ_i gilt $E(\epsilon_i) = 0$. Nach dem Einsetzen von nichtparametrischen Schätzern für $E(NM_i|WFL_i)$ und $E(\mathbf{z}_i|WFL_i)$ ³ ergibt sich

$$NM_i^* = \frac{1 + \mathbf{z}_i \boldsymbol{\beta}}{1 + \hat{E}(\mathbf{z}_i|WFL_i) \boldsymbol{\beta}} + \epsilon_i, \quad (5.18)$$

mit $NM_i^* := \frac{NM_i}{\hat{E}(NM_i|WFL_i)}$. Dieses Modell ist nichtlinear in den Parametern $\boldsymbol{\beta}$. Nach dem Einsetzen von nichtparametrischen Schätzungen $\hat{E}(NM_i|WFL_i)$ und $\hat{E}(\mathbf{z}_i|WFL_i)$, $i = 1, \dots, n$, können diese Parameter mittels der nichtlinearen Kleinst-Quadrat-Methode geschätzt werden. Bevor wir uns hiermit auseinandersetzen, wird im nächsten Abschnitt die nichtparametrische Schätzung der Funktionen $E(NM_i|WFL_i)$ und $E(\mathbf{z}_i|WFL_i)$ näher betrachtet.

³Wegen $E(f(\mathbf{z}_i, \boldsymbol{\beta})|WFL_i) = E((1 + \mathbf{z}_i \boldsymbol{\beta})|WFL_i) = 1 + E(\mathbf{z}_i|WFL_i) \boldsymbol{\beta}$.

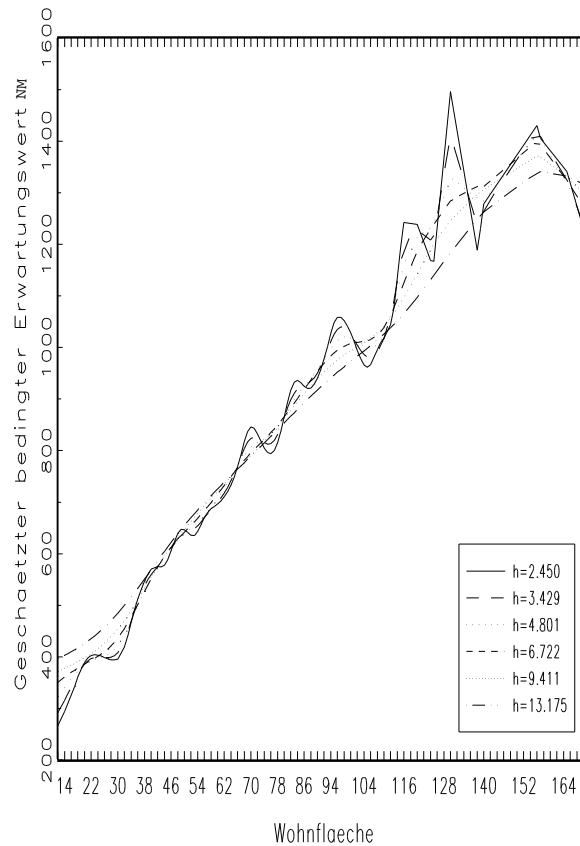


Abbildung 5.1. Die mittels des Nadaraya-Watson-Kernschätzers geschätzte bedingte Erwartungswertfunktion für die Nettomiete (mit verschiedenen Bandbreiten).

5.4.3 Nichtparametrische Schätzergebnisse für die auf die Wohnfläche bedingten Erwartungswerte

Die oben genannten bedingten Erwartungswerte wurden nichtparametrisch mittels des Nadaraya-Watson-Kernschätzers geschätzt. Als Kernfunktion wurde der Gauss-Kern verwendet. Für diesen Kern kann die optimale Bandbreite h durch die auf Seite 51 eingeführte Faustregel von Silverman (4.6) bestimmt werden. Für unsere Daten ergibt sich eine geschätzte optimale Bandbreite $\hat{h} = 6.7$.

Wie in MÜLLER/FAN (1995) vorgeschlagen, wurde eine Familie (4.7) von Schätzungen durchgeführt, um eine mögliche Überglättung der zu schätzenden Funktion zu prüfen. Abbildung 5.1 zeigt den auf die Wohnfläche bedingten geschätzten Erwartungswert der Nettomiete für jeden Wert der Wohnfläche.

Wie aus der Abbildung hervorgeht, ergibt der Nadaraya-Watson-Kernschätzer mit der geschätzten optimalen Bandbreite $\hat{h} = 6.7$ eine Schätzung des auf die Wohnfläche bedingten Erwartungswertes der Nettomiete, die weder sehr glatt (dann könnten wichtige Details verloren werden) noch sehr rauh (dann könnten Details hervortreten, die sich nur aufgrund der Zufälligkeit der Stichprobe ergeben haben) ist und deswegen vorgezogen wird. Als Beispiel einer übergeglätteten Funktion, die im Bereich der mittleren Wohnungen approximativ linear verläuft und damit die Details verbirgt, kann die Schätzung mit $h = 13.2$ betrachtet werden. (Exemplarisch sind die Schätzungen der bedingten Erwartungswerte von zwei anderen Faktoren im Anhang B dargestellt.) Sowohl die ausgewählten geschätzten Erwartungswerte der Nettomiete als auch die Beobachtungen der Variablen Nettomiete und Wohnfläche sind in Abbildung 5.2 dargestellt. Für alle Wohnungen, die kleiner als 156 qm sind, ist die geschätzte Erwartungswertfunktion der Nettomiete eine in der Wohnfläche zunehmende Funktion. Allerdings fällt die Funktion in der Wohnfläche für Wohnungen, die größer als 156 qm sind. Die möglichen Erklärungen für diesen Verlauf sind der additive Effekt anderer erklärenden Variablen als der Wohnfläche und der wenige Beobachtungen in der Stichprobe für diese großen Wohnungen.

5.4.4 Konfidenzintervalle für den geschätzten Erwartungswert der Nettomiete

Trotz des unbekannten Zusammenhangs zwischen der Nettomiete und der Wohnfläche ist es möglich, Konfidenzintervalle für die Nettomiete zu konstruieren. Dabei wurde für jeden Wert der geschätzten Erwartungswertfunktion der Nettomiete ein wie in (4.10) definiert approximiertes asymptotisches punktweises 95%-Konfidenzintervall für den Nadaraya-Watson-Kernschätzer ermittelt. Diese werden in Abbildung 5.3 gezeigt: Ap-

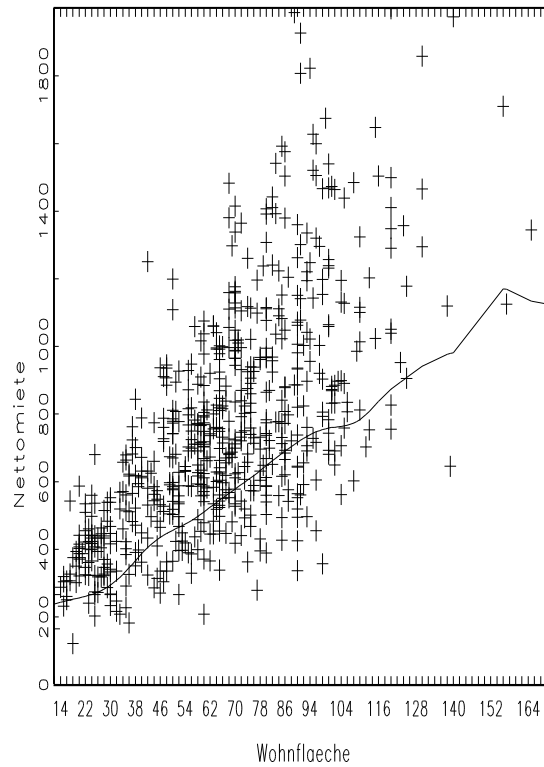


Abbildung 5.2. Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Erwartungswertfunktion für die Nettomiete und Beobachtungen für die Nettomiete.

proximierte asymptotische 95%-Konfidenzintervalle für den auf die Wohnfläche bedingten Erwartungswert der Nettomiete sind zunächst für 8 Werte der Wohnfläche dargestellt (siehe Abbildung 5.3.1). Damit soll betont werden, daß es sich um diskrete Konfidenzintervalle und nicht um stetige Konfidenzbänder handelt. In Abbildung 5.3.2 werden die Konfidenzintervalle für Wohnflächen, die kleiner als oder gleich 156 qm sind, dargestellt.

Abbildung 5.4 zeigt den Vergleich dieser asymptotischen Konfidenzintervalle mit denjenigen aus der BC_a -Methode (siehe Seiten 57 ff. für Details der Methode). Ein beachtlicher Unterschied ist zu merken. Die asymptotischen Konfidenzintervalle sind über

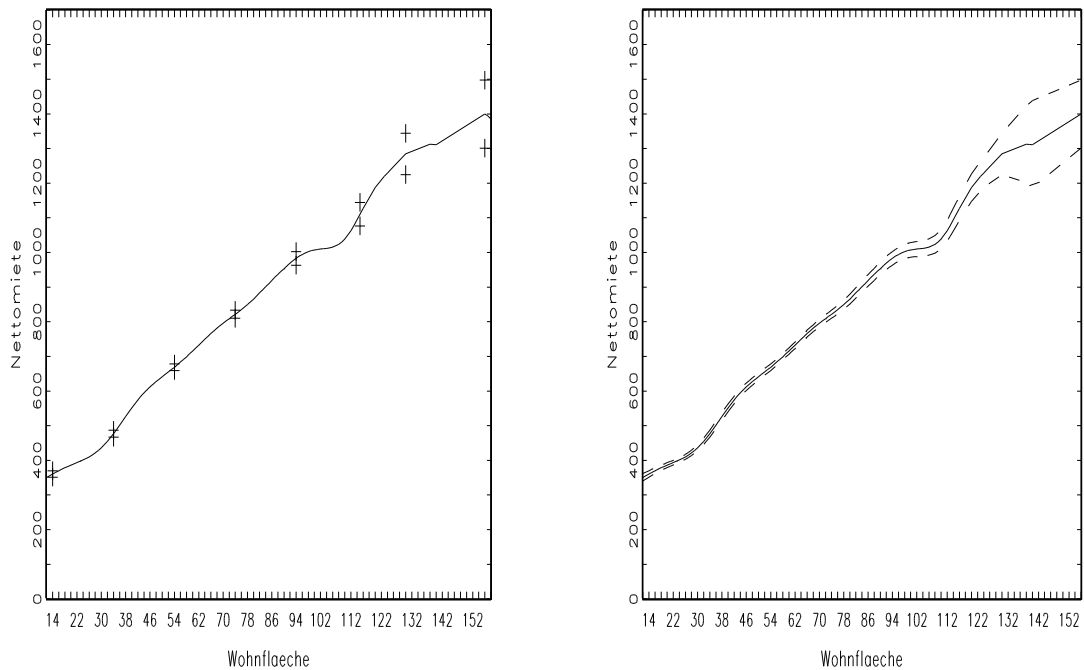


Abbildung 5.3. Approximierte asymptotische 95%-Konfidenzintervalle für den auf die Wohnfläche bedingten Erwartungswert der Nettomiete.

den ganzen Bereich der Wohnfläche schmäler. Darüber hinaus wird der Unterschied viel größer im Bereich der größeren Wohnungen, wo nur wenige Beobachtungen für die Nettomiete vorliegen. Außerdem ergab sich keinen substantiellen Unterschied aus dem Vergleich zwischen den durch die Quantil-Methode und den durch die BC_a -Methode geschätzten Konfidenzintervalle für die kleinen und mittleren Wohnungen.

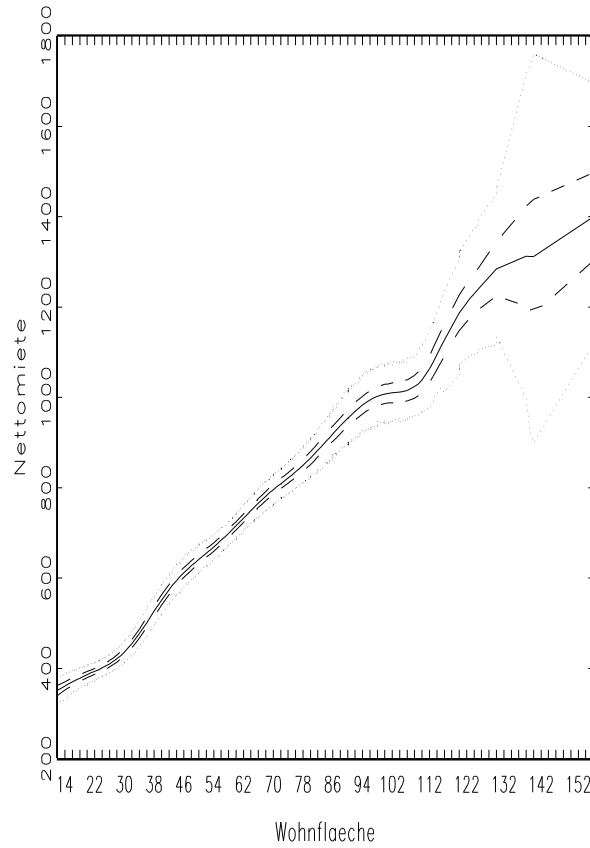


Abbildung 5.4. Vergleich zwischen Bootstrap 95%-Konfidenzintervallen mittels der BC_a -Methode (punktierte Linie) und approximierten asymptotischen 95%-Konfidenzintervallen (gestrichelte Linie) für den auf die Wohnfläche bedingten Erwartungswert der Nettomiete.

5.4.5 Schätzung des parametrischen Teils des Modells

In vorigen Abschnitte wurde auf den ersten Schritt zur Schätzung des Modells (5.18)

$$\frac{NM_i}{E(NM_i|WFL_i)} = \frac{1 + \mathbf{z}_i\boldsymbol{\beta}}{1 + E(\mathbf{z}_i|WFL_i)\boldsymbol{\beta}} + \epsilon_i,$$

eingegangen, nämlich die Schätzung der auf die Wohnfläche bedingten Erwartungswertfunktionen. Darüber hinaus wurden Konfidenzintervalle für $E(NM_i|WFL_i)$ geschätzt. In diesem Abschnitt wird der Einfluß der weiteren, über die Wohnfläche hinausgehenden Mietpreisdeterminanten geschätzt. Das Modell (5.18) ist nichtlinear in den Parametern β . Diese Parameter können durch die nichtlineare Kleinste-Quadrate-Methode (NLKQ-Methode) geschätzt werden. Der NLKQ-Schätzer $\hat{\beta}$ ist derjenige Wert von möglichen Werten des Parametervektors β , der die folgende Funktion minimiert

$$S_n(\mathbf{b}) = \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{z}_i, \mathbf{b}))^2.$$

In unserem Fall:

$$\hat{\beta} = \arg \min_{\mathbf{b}} \frac{1}{n} \sum_{i=1}^n \left(NM_i^* - \frac{1 + \mathbf{z}_i \mathbf{b}}{1 + \hat{E}(\mathbf{z}_i|WFL_i)\mathbf{b}} \right)^2,$$

mit $NM_i^* := \frac{NM_i}{\hat{E}(NM_i|WFL_i)}$, wie bereits oben definiert wurde.

Mit Hilfe eines iterativen numerischen Optimierungsalgorithmus wurden die NLKQ-Schätzungen bestimmt, und zwar mit dem Marquardt-Levenberg-Algorithmus (siehe etwa JUDGE ET AL (1985), S. 963 ff. für eine detaillierte Beschreibung). Nach 12 Iterationen konvergierten die Schätzungen (Konvergenzkriterium: Die relative Veränderung zwischen sukzessiven Summen der Quadrate der Residuen ist mindestens 10^{-8}). Die Werte sind in Tabelle 5.7 wiedergegeben. Bei Wiederholungen des Schätzverfahrens mit anderen Anfangswerten zeigte sich, daß der Algorithmus gegenüber der Wahl der Anfangswerte robust ist.

5.4.6 Asymptotische Konfidenzintervalle und Tests für den semiparametrischen Schätzer

In diesem Abschnitt werden sowohl Konfidenzintervalle für den Parametervektor β des parametrischen Teils unseres additiv-multiplikativen Modells konstruiert als auch

Variable	$\hat{\beta}_i$	$\sqrt{\widehat{\text{var}}(\hat{\beta}_i)}$	Z-Prüfgröße	p-Wert	95%-KI	
<i>F1GE</i>	-0.0273	0.0225	-1.2136	0.8876	-0.0713	0.0168
<i>F2GE</i>	0.0710	0.0357	1.9901*	0.0233	0.0011	0.1409
<i>F3GE</i>	0.0573	0.0494	1.1591**	0.1232	-0.0396	0.1541
<i>F4GE</i>	-0.0500	0.0719	-0.6955	0.7566	-0.1911	0.0910
<i>F1LA</i>	-0.0524	0.0333	-1.5737	0.9422	-0.1176	0.0129
<i>F2LA</i>	0.0165	0.0280	0.5865	0.2788	-0.0385	0.0714
<i>F3LA</i>	0.0251	0.0279	0.8976	0.1847	-0.0297	0.0799
<i>F4LA</i>	-0.0289	0.0356	-0.8103	0.7911	-0.0987	0.0410
<i>F5LA</i>	-0.0585	0.0377	-1.5506	0.9395	-0.1324	0.0154
<i>F1BE</i>	0.0061	0.0317	0.1921	0.4238	-0.0560	0.0682
<i>F2BE</i>	0.0151	0.0458	0.3287	0.3712	-0.0748	0.1049
<i>F3BE</i>	-0.0194	0.0435	-0.4469	0.6725	-0.1046	0.0657
<i>F4BE</i>	-0.0411	0.0454	-0.9045	0.8171	-0.1301	0.0479
<i>F5BE</i>	0.0946	0.0434	2.1794*	0.0147	0.0095	0.1797
<i>F6BE</i>	-0.0196	0.0611	-0.3210	0.6259	-0.1393	0.1001
<i>F1SON</i>	0.0386	0.0221	1.7440*	0.0406	-0.0048	0.0819
<i>F2SON</i>	-0.1292	0.1881	-0.6866	0.7538	-0.4978	0.2395
<i>F3SON</i>	-0.0289	0.0339	-0.8528	0.8031	-0.0954	0.0376
<i>F4SON</i>	-0.0180	0.0350	-0.5135	0.6962	-0.0866	0.0507
<i>F5SON</i>	-0.0369	0.0357	-1.0340	0.8494	-0.1069	0.0331
<i>F6SON</i>	-0.0951	0.0393	-2.4203	0.9922	-0.1721	-0.0181
<i>F1AUS</i>	0.0436	0.0323	1.3487**	0.0887	-0.0197	0.1069
<i>F2AUS</i>	0.0327	0.0354	0.9254	0.1774	-0.0366	0.1020
<i>F3AUS</i>	0.0248	0.0213	1.1628**	0.1225	-0.0170	0.0665
<i>F4AUS</i>	-0.0354	0.0230	-1.5395	0.9382	-0.0805	0.0097
<i>F5AUS</i>	-0.0138	0.0376	-0.3661	0.6429	-0.0875	0.0600
<i>F6AUS</i>	0.0290	0.0207	1.4016**	0.0805	-0.0116	0.0696
<i>F7AUS</i>	0.0246	0.0218	1.1240**	0.1305	-0.0183	0.0674
<i>F8AUS</i>	0.0936	0.0242	3.8682*	0.0001	0.0462	0.1410
<i>F9AUS</i>	0.0914	0.0392	2.3342*	0.0098	0.0147	0.1682
<i>F10AUS</i>	0.0275	0.0427	0.6436	0.2599	-0.0562	0.1111
<i>F11AUS</i>	0.0174	0.0302	0.5774	0.2818	-0.0417	0.0765
<i>F12AUS</i>	-0.0366	0.0510	-0.7183	0.7637	-0.1365	0.0633
<i>F13AUS</i>	0.0622	0.0349	1.7828*	0.0373	-0.0062	0.1306
<i>F14AUS</i>	0.0415	0.0350	1.1857**	0.1179	-0.0271	0.1102
<i>F15AUS</i>	0.0290	0.0287	1.0132	0.1555	-0.0271	0.0852
<i>Baualter</i>	0.1886	0.1681	1.1221**	0.1309	-0.1408	0.5180
<i>Wohnungstyp</i>	0.0167	0.0458	0.3651	0.3575	-0.0730	0.1065

* Koeffizient signifikant bei 5%

** Koeffizient signifikant bei approx. 10%

Tabelle 5.7. Schätzergebnisse des parametrischen Teils des semiparametrischen Regressionsmodells.

Hypothesentests durchgeführt. Dafür wird von dem im Abschnitt 5.4.5 ermittelten Schätzer $\hat{\beta}$ ausgegangen. Ein derartiger Schätzer wird als “semiparametrischer” Schätzer bezeichnet, da es sich um einen Schätzer eines Parameters handelt, der von nichtparametrischen Schätzungen abhängt.

Die asymptotische Verteilung des semiparametrischen Schätzers lautet:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_\beta^{-1} \mathbf{\Omega} \mathbf{G}_\beta^{-1'}), \quad (5.19)$$

wobei $\frac{1}{n} \mathbf{G}_\beta^{-1} \mathbf{\Omega} \mathbf{G}_\beta^{-1'}$ die asymptotische Varianz-Kovarianz-Matrix des semiparametrischen Schätzers ist. (Siehe Anhang C für die Herleitung.)

Dann ist ein 95%-Konfidenzintervall für β_k , $k = 1, \dots, 38$, durch

$$\left[\hat{\beta}_k \pm 1.96 \cdot \sqrt{\widehat{var}(\hat{\beta})_{kk}} \right]$$

gegeben, wobei $\widehat{var}(\hat{\beta})_{kk}$ das k -te Hauptdiagonalelement der geschätzten asymptotischen Varianz-Kovarianz-Matrix $n^{-1} \hat{\mathbf{G}}_\beta^{-1} \hat{\mathbf{\Omega}} \hat{\mathbf{G}}_\beta^{-1'}$ des semiparametrischen Schätzers $\hat{\beta}$ ist. Die entsprechenden asymptotischen Tests auf Signifikanz der Einzelkoeffizienten basieren auf der folgenden Prüfgröße

$$Z_k := \frac{\hat{\beta}_k}{\sqrt{\widehat{var}(\hat{\beta})_{kk}}} \sim N(0, 1).$$

Die Ergebnisse der Tests und 95%-Konfidenzintervalle für jeden Koeffizient sind in den letzten drei Spalten der Tabelle 5.7 aufgeführt.

Eine Anzahl von 22 individuell nichtsignifikanten geschätzten Parametern wurde anhand eines Wald-Tests auf ihre gemeinsame Signifikanz getestet. Die Nullhypothese der gemeinsamen Nichtsignifikanz konnte nicht abgelehnt werden. Die Wald-Prüfgröße ergab 32.31; die Quantilwerte einer χ^2 -Verteilung mit 22 Freiheitsgraden zu 1% und 5% betragen 40.29 bzw. 33.93. In Anbetracht der Testergebnisse und der ähnlichen Schätzwerte sowohl beim vollständigen wie beim reduzierten Modell wurden nur die 16

restlichen signifikanten Mietpreisdeterminanten beibehalten. Als Fazit sind diese Ergebnisse (Schätzungen) und die Bezeichnung dieser signifikanten Mietpreisdeterminanten in Tabelle 5.8 dargestellt. Aus den Werten der impliziten Preise läßt sich schließen, daß die Variablen *Baualter*, *F5BE*, *F8AUS*, *F9AUS* und *F2GE* den größten Einfluß auf den Mietpreis haben (d.h. das Vorhandensein eines Aufzugs, die Ausrichtung des Balkons und Ausstattungs-Variablen wie das Vorhandensein einer abgetrennten Dusche, eines gekachelten Bads bzw. einer gekachelten Küche). Die Variable *Baualter* hat den größten impliziten Preis. Andere Ausstattungs-Variablen wie die Art der Kochgelegenheit, das Vorhandensein von Warmwasserversorgung und Heizungsmöglichkeiten, als auch Einrichtungselemente wie elektrische Rolladen haben auch einen großen Einfluß auf die Mietpreis. Was die Lage betrifft, sind sowohl das Vorhandensein von Parks und Geschäften in der Nähe als auch gute Verkehrsverbindungen für den Mietpreis von Bedeutung.

5.5 Konfidenzintervalle für Mietpreise

Zur Illustration werden die Ergebnisse des additiv-multiplikativen Modells für die Nettomiete für bestimmte Typen von Wohnungen zusammengefaßt. Dafür werden Intervallschätzungen für die Grundmiete und die Nettomiete ermittelt, um die Variabilität der Mietpreise wegen der Zufälligkeit der Stichprobe aus Darmstadt zu berücksichtigen. Im Rahmen von additiv-multiplikativen Modellen sind die zwei folgenden Alternativen zur Berechnung von Konfidenzintervallen geeignet: erstens approximative asymptotische Konfidenzintervalle und zweitens mittels Bootstrap geschätzte Konfidenzintervalle. Die Bootstrap-Technik ist einfach zu implementieren und wird hier angewendet. Es wurde die Variante des “wildten” Bootstraps verwendet, die im Abschnitt 4.2.5.2 beschrieben wurde. Der Vorteil dieser Variante besteht darin, daß die Verzerrung der Schätzer des Modells berücksichtigt werden, ohne weitere Schätzung von zusätzlichen Parametern.

<i>Mietpreis-</i> <i>determin.</i>	<i>implizite</i> <i>Preise</i>	<i>Beschreibung</i>
<i>F2GE</i>	0.053	Zusammenhang zwischen dem Vorhandensein eines Aufzugs und der Anzahl der Wohnungen in jedem Stockwerk
<i>F3GE</i>	0.019	Nähe des Gebäudes zu anderen Gebäuden
<i>F3LA</i>	0.032	Infrastruktur der Stadt
<i>F5BE</i>	0.058	Balkonorientierung gegenüber dem Vorhandensein von Türschwellen
<i>F1SON</i>	0.026	Waschmöglichkeiten
<i>F1AUS</i>	0.038	Übliche Kücheneinrichtungen und Küchenmaschinen
<i>F2AUS</i>	0.043	Art der Kochgelegenheit und das Vorhandensein eines einfachen Bads
<i>F3AUS</i>	0.013	Länge und Breite des Badezimmers
<i>F6AUS</i>	0.039	Fußbodenheizung und Messung der Wärme
<i>F7AUS</i>	0.014	Fernsehübertragung
<i>F8AUS</i>	0.097	Gekacheltes Bad und gekachelte Küche
<i>F9AUS</i>	0.110	Vorhandensein nur einer festen abgetrennten Duschwanne
<i>F13AUS</i>	0.040	Warmwasserversorgung und Heizmöglichkeit in allen Zimmern
<i>F14AUS</i>	0.045	Zusammenhang zwischen der Art der Spüle und dem Vorhandensein elektrischer Rolläden
<i>F15AUS</i>	0.035	Wohnung ist vollständig mit Hausrat/Geschirr ausgerüstet
<i>Baualter</i>	0.19	

Tabelle 5.8. Statistisch signifikante Mietpreisdeterminanten und entsprechende implizite Preise des semiparametrischen Modells für die Nettomiete.

5.5.1 Konfidenzintervalle für die Grundmiete

In diesem Abschnitt werden Konfidenzintervalle für die Grundmiete vorgestellt. In unserem semiparametrischen Modell ist die Grundmiete folgendermaßen definiert: $GM_i := q(WFL_i)$, die anhand eines Schätzers für die unbekannte Funktion $q(\cdot)$ nach der Schätzung der zwei auf die Wohnfläche bedingten Erwartungswertfunktionen $E(NM|WFL_i)$ und $E(\mathbf{z}|WFL_i)$ und des Parametervektors $\boldsymbol{\beta}$ geschätzt werden kann (siehe ROBINSON (1988), S. 936 für den additiven Fall), nämlich durch

$$\hat{q}(WFL_i) = \frac{\hat{E}(NM_i|WFL_i)}{1 + \hat{E}(\mathbf{z}|WFL_i)\hat{\boldsymbol{\beta}}} \quad i = 1, \dots, n. \quad (5.20)$$

Die Gleichung (5.20) ergibt sich aus dem auf die Wohnfläche bedingten Erwartungswert des Modells (5.16), nämlich

$$E(NM_i|WFL_i) = q(WFL_i) \cdot E(1 + \mathbf{z}|WFL_i)\boldsymbol{\beta}), \quad (5.21)$$

nach dem Ersetzen der bedingten Erwartungswertfunktionen und der Parameter $\boldsymbol{\beta}$ durch ihre entsprechenden nichtparametrischen bzw. semiparametrischen Schätzungen.

Diese Schätzung wurde anhand der Darmstadt-Stichprobe für 1000 Bootstrap-Stichproben ermittelt, so daß eine mittels Bootstrap generierte Verteilung der geschätzten Grundmiete zugrunde liegt. Weiterhin wurden 95%-Konfidenzintervalle für die Grundmiete ermittelt. Die Konfidenzintervalle sind zur Illustration für ausgewählte Wohnungen unterschiedlicher Größe in Tabelle 5.9 dargestellt.

Die dritte Spalte der Tabelle 5.9 zeigt die Prognosewerte für die Grundmiete jeder Wohnungskategorie. Diese Prognosewerte sind das arithmetische Mittel aller Prognosewerte für die Grundmiete aus den 1000 Bootstrap-Stichproben für jede betrachtete Wohnungskategorie.

In der vierten Spalte der Tabelle folgen die 95%-Konfidenzintervalle für die Grundmiete. Zum Vergleich der Intervalle kann das folgende Maß der Streuung verwendet werden:

<i>Grundmiete</i>				
Bezeichnung Wohnung	Wohnfläche	Prognosewert	95%-KI	<i>HQA</i>
sehr klein	14 m^2	289	[255 ; 325]	35.0
klein	37 m^2	380	[331 ; 430]	49.5
mittel	85 m^2	667	[583 ; 751]	84.0
groß	140 m^2	920	[722 ; 1106]	192.0

Tabelle 5.9. Mittels der Methode des wilden Bootstraps Konfidenzintervalle einiger ausgewählter Wohnungen und Prognosewerte für die Grundmiete.

der halbe Interquartilabstand (*HQA*), der folgendermaßen definiert ist:

$$HQA = \frac{Q_o - Q_u}{2}, \quad (5.22)$$

wobei Q_o und Q_u das obere bzw. untere Quartil der mittels des Bootstraps simulierten Verteilung der Grundmiete bezeichnen. Diese Werte entsprechen dem oberen bzw. unteren Wert eines Konfidenzintervalls für die Grundmiete.

Angesichts der Ergebnisse kann man feststellen, daß die durch *HQA* definierte Länge der Intervalle mit der Wohnungsgröße zunimmt. Bei der Schätzung des semiparametrischen Modells führt das Vorliegen von wenigen Beobachtungen für einen bestimmten Typ von Wohnungen, wie der Fall für die großen Wohnungen in der Darmstadt-Stichprobe ist, zu größeren Konfidenzintervallen.

5.5.2 Konfidenzintervalle für die Nettomiete

Die Konfidenzintervalle für die Nettomiete basieren auf dem geschätzten Regressionsmodell

$$\widehat{NM}_i = \widehat{GM}_i \cdot (1 + \mathbf{z}_i \hat{\boldsymbol{\beta}}) \quad i = 1, \dots, n, \quad (5.23)$$

wobei \widehat{GM}_i wie in Abschnitt 5.5.1 definiert ist.

Die Ergebnisse für die Nettomiete sind in Tabelle 5.10 abgebildet. Für die Nettomiete

<i>Nettomiete</i>				
Bezeichnung Wohnung	Wohnfläche	Prognosewert	95%-KI	HQA
sehr klein	14 m^2	391	[338 ; 445]	53.5
klein	37 m^2	711	[650 ; 774]	62.0
mittel	85 m^2	889	[670 ; 1087]	208.5
groß	140 m^2	1666	[1320 ; 1988]	334.0

Tabelle 5.10. Mittels der Methode des wilden Bootstraps Konfidenzintervalle einiger ausgewählter Wohnungen und Prognosewerte für die Nettomiete.

stellen wir ebenfalls fest, daß die Länge der Konfidenzintervalle mit der Größe der Wohnung zunimmt. Der halbe Interquartilabstand bei der mittleren Wohnung ist erheblich größer als bei der sehr kleinen und kleinen Wohnung. Zu weiteren Beispielen der Anwendung von solchen Techniken zur Schätzung von Konfidenzintervallen im Rahmen von semiparametrischen Regressionsmodelle siehe auch ANGLIN/GENÇAY (1996).

Zusammenfassung und Schlußfolgerungen

Der erste Teil dieser Arbeit befaßte sich mit den institutionellen und theoretischen Grundlagen eines Regressionsmietspiegels, während im zweiten Teil ausführlich auf die Entwicklung eines neuen allgemeineren Modells eines Mietspiegels eingegangen wurde. Dieses Modell hat nicht die a priori Restriktionen der anderen Modelle und stellt somit eine Verallgemeinerung dar, d.h. die anderen Modelle können als Spezialfälle interpretiert werden.

Das erste Kapitel beschrieb das Konzept eines Mietspiegels und seine Rolle im Rahmen des Mietrechts in Deutschland. Die gesetzlichen Grundlagen für Mietspiegel in Deutschland wurden zunächst erörtert. Die Aufgaben eines Mietspiegels sowie seine Vor- und Nachteile wurden ebenfalls diskutiert. Darüber hinaus wurden zwei Methoden zur Erstellung von Mietspiegeln, die Tabellen- und die Regressionsmethode, betrachtet. Dieses Kapitel wurde mit einem Überblick über die Diskussion der gegenwärtigen Mietrevision in der Schweiz abgeschlossen.

Das zweite Kapitel, "Hedonische Methode und Regressionsmietspiegel", befaßte sich mit der theoretischen Fundierung von Mietspiegeln. Zuerst wurde die hedonische Hypothese eingeführt, welche besagt, daß heterogene Güter durch ihre Charakteristika gekennzeichnet sind und es die Charakteristika sind, die den Konsumenten einen Nutzen stiften. Zwei mikroökonomische Modelle folgten, nämlich das Standardmodell der Produktionstheorie des Haushalts und das lineare Modell für Charakteristika. Auf diesen

Modellen beruht die hedonische Hypothese. Anschließend wurde das lineare Modell für Charakteristika für den Wohnungsmarkt anhand eines Beispiels illustriert. Die Eigenschaften dieser Modelle erlauben, bestimmte Aussagen für die Erstellung von Regressionsmietspiegeln zu machen, insbesondere über die Spezifikation des Modellansatzes und die Variablen, die in einer Analyse berücksichtigt werden sollen.

In einem zweiten Teil der Arbeit wurde ausführlich auf die Entwicklung eines allgemeineren Modells eines Regressionsmietspiegels eingegangen. Dabei wurde ein additiv-multiplikatives semiparametrisches Regressionsmodell für die Nettomiete vorgeschlagen. Dieses Modell wurde in Kapitel 3 eingeführt, nachdem die in Deutschland üblichen Modellansätze für Regressionsmietspiegel kurz vorgestellt wurden. Im allgemeinen gibt es verschiedene Ansätze zur Spezifikation eines Regressionsmietspiegels. Eine erste Klasse von Modellen ist dadurch gekennzeichnet, daß der Zusammenhang zwischen dem Mietpreis und der Wohnfläche explizit parametrisch spezifiziert wird. Im Gegensatz dazu gehört das in dieser Arbeit vorgeschlagene Regressionsmodell zu einer zweiten Klasse von Modellen, den sogenannten semiparametrischen Regressionsmodellen, bei denen dieser Zusammenhang nichtparametrisch modelliert wird, während der Einfluß der restlichen Mietpreisdeterminanten weiterhin parametrisch modelliert wird. Semiparametrische Regressionsmodelle sind in der Praxis bisher wenig bekannt und können als eine Alternative zur Modellierung eines Mietspiegels betrachtet werden.

Zu einem besseren Verständnis des semiparametrischen Regressionsmodells wurde die Grundlage semiparametrischer Regressionsmodelle ebenfalls in diesem Kapitel eingeführt, wobei ein rein additives semiparametrisches Modell betrachtet wurde.

In Kapitel 4 wurden Schätzmethoden für semiparametrische Regressionsmodelle eingeführt. Insbesondere wurden Schätzmethoden für den einfacheren Fall des additiven semiparametrischen Regressionsmodells diskutiert. Zur Schätzung der nichtparametrischen Komponente des Modells betrachteten wir den Nadaraya-Watson-Kernschätzer. Darüber hinaus wurden verschiedene Simulationstechniken zur Schätzung von Konfidenzintervallen erläutert. In Anschluß daran wurde die Schätzmethode für die parametrische Komponente des Regressionsmodells beschrieben.

Kapitel 5 befaßte sich mit der Schätzung des in dieser Arbeit vorgeschlagenen semiparametrischen Regressionsmodells für Mietpreise. Konkret wurde auf die Schätzung des Regressionsmodells anhand der Daten, die 1994/95 für den Mietspiegel der Stadt Darmstadt (D) erhoben wurden, eingegangen. Dieser Datensatz enthält eine Vielfalt von Variablen, die den Mietpreis bestimmen.

Aufgrund der großen Zahl (ca. 100) von möglichen erklärenden Variablen oder Mietpreisdeterminanten wurde zunächst auf eine Reduktion der Daten eingegangen, wobei komplexe Faktoren durch die Methode der Faktorenanalyse konstruiert wurden, die Variablen mit ähnlicher Information umfaßten. Somit wurde die Anzahl der Variablen auf 36 unkorrelierte Faktoren reduziert. Bei unserer Anwendung der Faktorenanalyse bestand die Hauptschwierigkeit darin, sinnvoll interpretierbare Faktoren aus der Analyse zu gewinnen. Auch nach der Rotation der Faktoren, die ihre Interpretation verbessern sollte, war die Interpretation keine einfache Aufgabe, insbesondere mit einer derart großen Vielfalt von unterschiedlichen Variablen wie im Fall der Ausstattungskategorie.

Im Anschluß daran wurde die Schätzung des semiparametrischen Modells anhand der Daten für den Mietpreis und die Faktoren durchgeführt. Um auf die Signifikanz der Koeffizienten der semiparametrischen Komponente des Modells zu testen, wurde die asymptotische Verteilung des semiparametrischen Schätzers im Detail im Anhang C hergeleitet. Aufgrund der Ergebnisse wurden 16 von 38 Regressoren als signifikant beibehalten. In Anbetracht der Schätzergebnisse kann geschlossen werden, daß die Variablen, die die Ausstattung der Wohnung (z.B. das Vorhandensein einer abgetrennten Dusche) und der Art des Gebäudes (z.B. das Vorhandensein eines Aufzugs) beschreiben, einen wichtigen Einfluß auf den Mietpreis ausüben. Dabei hatte die Variable Baualter den größten Einfluß auf den Mietpreis. Was die Lage betrifft, waren nur das Vorhandensein von Parks und Geschäften in der Nähe und von guten Verkehrsverbindungen für den Mietpreis von Bedeutung.

Konfidenzintervalle sowohl für die Grundmiete als auch für die Nettomiete wurden für vier Wohnungskategorien exemplarisch ermittelt. Diese sind für die Erstellung von Regressionsmietspiegeln relevant und stellen eine statistisch fundierte Methode dar,

Mietspannen und nicht nur einen einzelnen Durchschnittspreis für einen bestimmten Wohnungstyp zu zeigen.

Anhang A

Überblick über die Faktoren für die Mietpreisdeterminanten

Die vier folgenden Faktoren betreffen die *Art des Gebäudes*, nämlich

- F1GE* Art des Gebäudes bezüglich der Höhe,
- F2GE* Zusammenhang zwischen dem Vorhandensein eines Aufzugs
und der Anzahl der Wohnungen in jedem Stock,
- F3GE* Nähe des Gebäudes zu anderen Gebäuden,
- F4GE* Art des Gebäudes.

Die *Lage der Wohnung* ist durch die folgenden fünf Faktoren gekennzeichnet:

- F1LA* Kontrast zwischen Lärm der Industrie und Ruhe,
- F2LA* Verkehrsbelastung,
- F3LA* Infrastruktur der Stadt,
- F4LA* Aussicht der Wohnung,
- F5LA* Sicherheitsbedingungen für die Passanten
bezüglich des Verkehrs.

Bezüglich der *Beschaffenheit der Wohnung* berücksichtigen wir sechs Faktoren:

- F1BE* Kleine Räume und ihre Besonderheiten,
F2BE Kontrast zwischen dem Vorhandensein von Türschwellen und großen Räumen und dem Gesamterhaltungszustand des Gebäudes,
F3BE Zusammenhang zwischen dem Vorhandensein von Wohndiele, Balkonen und der Breite der Türöffnungen,
F4BE Kontrast zwischen niedrigen Räumen (Höhe $\leq 2.40\text{m}$), Luftdurchgang und großen Räumen,
F5BE Balkonorientierung gegenüber dem Vorhandensein von Türschwellen,
F6BE Zusammenhang zwischen den durch kleine Stufen/Absätze miteinander verbundenen Wohnräume und Tiefe/Breite des Balkons.

Das Merkmal *Sondereinrichtung* umfaßt sechs Faktoren, nämlich

- F1SON* Waschmöglichkeiten,
F2SON Luxuriöse Wassereinrichtungen,
F3SON Garten/Gartennutzung,
F4SON Zusätzliche Infrastruktur im Haus,
F5SON Spezielle zusätzliche Räume,
F6SON Benutzung von Räumen in unteren Geschossen.

Die letzte Gruppe von Faktoren betrifft das Merkmal, das unter der Bezeichnung *Ausstattung* definiert wurde. Die Faktoren sind die folgenden:

- F1AUS* Küchenmaschinen und übliche KÜcheinrichtungen,
- F2AUS* Art der Kochgelegenheit und das Vorhandensein eines
Bads nur mit einer Duschwanne und ohne Fenster,
- F3AUS* Länge und Breite des Badezimmers,
- F4AUS* Vorhandensein von Kühlschrank, modernen Einbau-
küchenschränken und einem Elektrogasherd in der Küche,
- F5AUS* Zusammenhang zwischen dem Vorhandensein von
Hausrat/Geschirr, Kochgelegenheit und
dem Möblierungstyp der Wohnung
- F6AUS* Fußbodenheizung und Messung der Wärme,
- F7AUS* Fernsehübertragung,
- F8AUS* Kachelung von Bad und Küche,
- F9AUS* Vorhandensein von nur einer festen
abgetrennten Duschwanne,
- F10AUS* Vorhandensein einer Toilette separat von Bad,
- F11AUS* Begehbarer Kleiderschrank,
- F12AUS* Vorhandensein einer zweiten Toilette,
- F13AUS* Warmwasserversorgung und Beheizungsmöglichkeit
in allen Zimmern,
- F14AUS* Zusammenhang zwischen der Art der Spüle und dem
Vorhandensein elektrischer Rolläden,
- F15AUS* Wohnung ist vollständig mit Hausrat/Geschirr
ausgerüstet.

Anhang B

Kernschätzungen

In diesem Abschnitt werden zwei Beispiele von mittels des Nadayara-Watson-Kernschätzers geschätzten bedingten Erwartungswertfunktionen exemplarisch dargestellt.

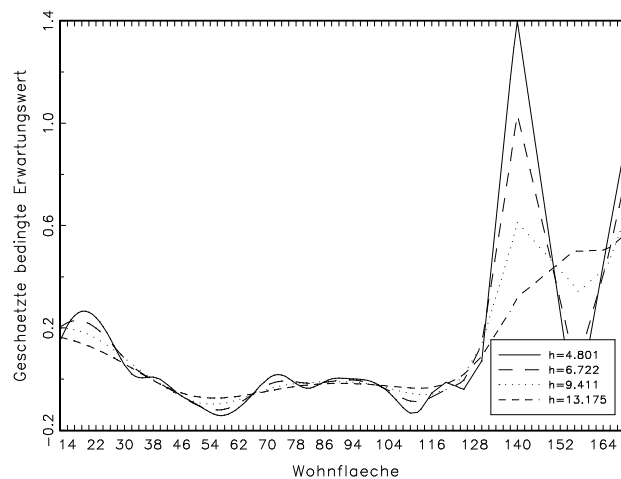


Abbildung B.1. Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Erwartungswertfunktion für den Faktor *F6AUS: Fußbodenheizung und Messung der Wärme* (mit verschiedenen Bandbreiten).

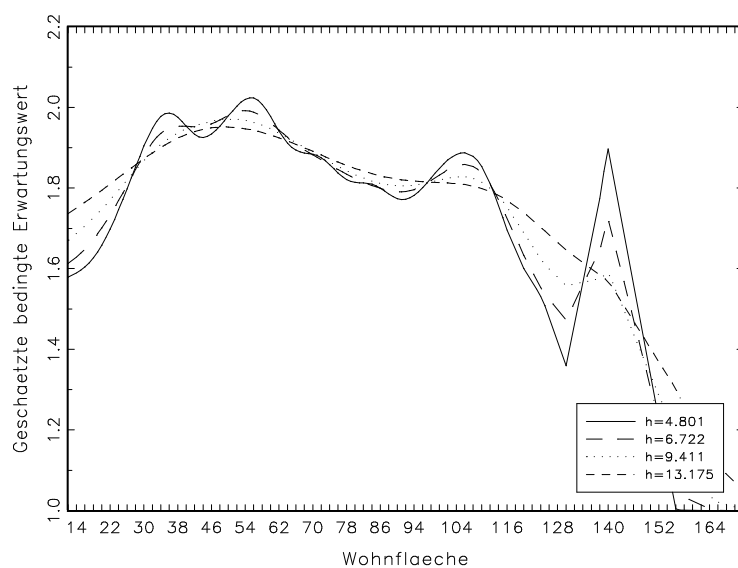


Abbildung B.2. Die mittels des Nadayara-Watson-Kernschätzers geschätzte bedingte Erwartungswertfunktion für die Variable *Baualter* (mit verschiedenen Bandbreiten).

Anhang C

Herleitung der asymptotischen Verteilung der Schätzer des semiparametrischen Regressionsmietspiegels

Ein Schätzer für die asymptotische Varianz-Kovarianz-Matrix des Schätzers $\hat{\beta}$ für die Koeffizienten des parametrischen Teils des additiv-multiplikativen Modells (5.16) wird in diesem Abschnitt hergeleitet. So ein Schätzer wird als “semiparametrischer” Schätzer bezeichnet, da es sich um ein Schätzer eines Parameters handelt, der von nichtparametrischen Schätzungen abhängt. Dafür wird der semiparametrische Schätzer als Schätzer einer bestimmten Art angesehen.

Hierzu wurde der Schätzer als ein Schätzer der Verallgemeinerten-Momente-Methode (engl. “generalized method of moments” (GMM)) betrachtet, d.h. als Schätzer, der sich als Lösung eines Systems sogenannter Moment-Bedingungen ergibt. Solche Moment-Bedingungen bezeichnen Annahmen eines empirischen Modells, die erfüllt werden sollen, wie z.B. für die Zufallsvariable Y soll die folgende Bedingung $E(Y) = \mu$ oder $E(Y - \mu) = 0$ erfüllt werden. Der GMM-Schätzer $\hat{\mu}$ für den unbekannten Parameter

μ ist derjenige, der die entsprechende empirische Kenngröße $\frac{1}{n} \sum_i (y_i - \hat{\mu}) = 0$ oder empirische Momente-Bedingung erfüllt, wobei $\{y_i\}_{i=1}^n$ n Beobachtungen der Variable Y sind.

Im letzten Abschnitt wurde der NLKQ-Schätzer des Parameters β vorgestellt. Dieser Schätzer ist ein Spezialfall einer größeren Klasse von Schätzern, welche sich als Lösung eines Systems von Moment-Bedingungen oder “Orthogonalitätsbedingungen” ergeben. Im Rahmen des nichtlinearen Modells (5.18) wird angenommen, daß jeder Regressor Z_j mit der Störvariablen ϵ_i unkorreliert ist, d.h. daß die Moment-Bedingungen

$$E(Z_j \epsilon_i) = 0, \quad \forall j = 1, \dots, 38$$

gelten. Dabei werden so viele Moment-Bedingungen wie zu schätzenden Parameter erfordert (“just-identified” GMM Spezifikation). Die entsprechenden empirischen Momente sind durch $g_j = \frac{1}{n} \sum_i z_{ji} e_i = 0$ gegeben, wobei e_i das i -te Residuum bezeichnet und z_{ji} die i -te Beobachtung des j -ten Regressors ist. Weiter wird

$$\begin{aligned} h_j(\mathbf{w}_i, \beta, \hat{\mathbf{m}}_h) &= z_{ji} \cdot e_i \\ &= z_{ji} \cdot \left(\frac{NM_i}{\hat{E}(NM_i|WFL_i)} - \frac{1 + \mathbf{z}_i \beta}{1 + \hat{E}(\mathbf{z}_i|WFL_i) \beta} \right), \end{aligned}$$

definiert, wobei $\hat{\mathbf{m}}_h$ die mittels des Nadaraya-Watson-Kernschätzers geschätzten bedingten Erwartungswertfunktionen $\hat{E}(NM_i|WFL_i)$ und $\hat{E}(z_{ji}|WFL_i)$, $j = 1, \dots, 38$ bezeichnet, mit $\mathbf{h}(\mathbf{w}_i, \beta, \hat{\mathbf{m}}_h) = (h_1, \dots, h_{38})'$. Jede Funktion $h_j(\cdot)$ ist eine Funktion der Daten $\mathbf{w}_i = (NM_i, z_i)$, für $i = 1, \dots, n$, der Parameter β und der Funktionen $\hat{\mathbf{m}}_h$. Dann kann der empirische Moment g_j wie folgt spezifiziert

$$g_j(\mathbf{w}_i, \beta, \hat{\mathbf{m}}_h) = \frac{1}{n} \sum_{i=1}^n h_j(\mathbf{w}_i, \beta, \hat{\mathbf{m}}_h) \quad (\text{C.1})$$

werden.

Sei $\mathbf{g}(\mathbf{w}_i, \beta, \hat{\mathbf{m}}_h) = (g_1, \dots, g_{38})'$ ein Funktionenvektor. Der semiparametrische Schätzer für β wird als der Schätzer definiert, der die folgenden empirischen Moment-Bedingungen erfüllt:

$$\mathbf{g}(\mathbf{w}_i, \boldsymbol{\beta}, \hat{\mathbf{m}}_h) = \mathbf{0}. \quad (\text{C.2})$$

Unter bestimmten Annahmen, die die Konsistenz und die asymptotische Normalverteilung des Schätzers gewährleisten (siehe Details über die Annahmen 5.1 bis 5.6 in NEWBY (1994b), oder vgl. Theorem 8.12 von NEWBY/MCFADDEN (1994), S. 2210) lautet die asymptotische Verteilung eines semiparametrischen Schätzers, der auf einer Kernschätzung basiert, folgendermaßen:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \mathbf{G}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\Omega} \mathbf{G}_{\boldsymbol{\beta}}^{-1'}), \quad (\text{C.3})$$

mit

$$\begin{aligned} \mathbf{G}_{\boldsymbol{\beta}} &= E[\partial \mathbf{h}(\mathbf{w}_i, \boldsymbol{\beta}, \hat{\mathbf{m}}_h) / \partial \boldsymbol{\beta}] \\ \boldsymbol{\Omega} &= E[(\partial \mathbf{h}(\mathbf{w}_i, \boldsymbol{\beta}, \hat{\mathbf{m}}_h) / \partial \boldsymbol{\beta})(\partial \mathbf{h}(\mathbf{w}_i, \boldsymbol{\beta}, \hat{\mathbf{m}}_h) / \partial \boldsymbol{\beta})']. \end{aligned}$$

(Siehe NEWBY/MCFADDEN (1994), S. 2194 ff., für die Herleitung der asymptotischen Verteilung eines allgemeinen semiparametrischen Schätzers.)

Die asymptotische Varianz kann durch das Einfügen von konsistenten Schätzern für die Matrizen $\mathbf{G}_{\boldsymbol{\beta}}^{-1}$ und $\boldsymbol{\Omega}$ geschätzt werden. Ein konsistenter Schätzer für $\mathbf{G}_{\boldsymbol{\beta}}'$ ist anhand der Delta-Methode¹ durch

$$\begin{aligned} \hat{\mathbf{G}}'_{\boldsymbol{\beta}} &:= \left. \frac{\partial \mathbf{g}(\mathbf{w}_i, \boldsymbol{\beta}, \hat{\mathbf{m}}_h)}{\partial \boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\ &= \frac{1}{n} \sum_{i=1}^n \left. \frac{\partial \mathbf{h}(\mathbf{w}_i, \boldsymbol{\beta}, \hat{\mathbf{m}}_h)}{\partial \boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \end{aligned}$$

gegeben. Nach dem Einsetzen der Funktionen $\mathbf{h}(\cdot)$ und nach der Differenzierung ergibt sich eine (38×38) -Matrix $\hat{\mathbf{G}}'_{\boldsymbol{\beta}}$ mit (j, k) -tem Element:

¹Die Delta-Methode wird angewendet, um die asymptotische Varianz-Kovarianz-Matrix einer Menge von Funktionen von Schätzern zu schätzen. Für Details siehe etwa GREENE (1993) S. 297.

$$\hat{\mathbf{G}}'_{\beta jk} = \frac{1}{n} \sum_{i=1}^n \frac{-z_{ij}z_{ik}[1 + \hat{E}(\mathbf{z}_i|WFL_i)\hat{\boldsymbol{\beta}}] + z_{ij}(1 + \mathbf{z}_i\hat{\boldsymbol{\beta}})\hat{E}(z_{ik}|WFL_i)}{[1 + \hat{E}(\mathbf{z}_i|WFL_i)\hat{\boldsymbol{\beta}}]^2}, \quad (\text{C.4})$$

für $j, k = 1, \dots, 38$.

Wie NEWBY (1994a), S. 237 erwähnt, ist ein konsistenter Schätzer für $\boldsymbol{\Omega}$ schwierig zu finden, da der Effekt der geschätzten unbekannten Funktionen $\hat{\mathbf{m}}_h$ berücksichtigt werden muß. Eine Möglichkeit zur Konstruktion der asymptotischen Varianz des semiparametrischen Schätzers besteht darin, zuerst den Einfluß jeder Beobachtung i , $i = 1, \dots, n$, im Kernschätzer $\hat{\mathbf{m}}_h$ auf den Funktionsvektor $\mathbf{g}(\mathbf{w}_i, \hat{\boldsymbol{\beta}}, \hat{\mathbf{m}}_h)$ zu schätzen. Folgend wird diese Schätzung bei der Schätzung der Varianz $\boldsymbol{\Omega}$ berücksichtigt.

Der Einfluß der i -ten Beobachtung im Kernschätzer auf den Funktionenvektor $\mathbf{h}(\cdot)$, der für die asymptotische Konvergenz in der Verteilung des semiparametrischen Schätzers notwendig ist, kann laut NEWBY/MCFADDEN (1994), S. 2211, folgendermaßen geschätzt werden:

$$\hat{\delta}_i := \nabla_{\xi} \left[\frac{1}{n} \sum_{m=1}^n h_j \left\{ \mathbf{w}_m, \hat{\boldsymbol{\beta}}, \hat{\mathbf{m}}_h + \xi NM_i K_h(WFL_m - WFL_i) \right\} \right] \Bigg|_{\xi=0}, \quad (\text{C.5})$$

für einen beliebigen Skalar ξ , wobei ∇_{ξ} der Gradient der Funktionen $\mathbf{h}(\cdot)$ ist, d.h. die vektorielle Funktion $\mathbf{h}(\cdot)$ wird nach ξ differenziert. Nach numerischer Differenzierung lautet das j -te Element des (38×1) -Schätzvektors $\hat{\delta}_i$:

$$\begin{aligned} \hat{\delta}_{ij} = & \frac{1}{n} \sum_{m=1}^n \left(\frac{z_{mj} \cdot NM_m \cdot NM_i \cdot K_h(WFL_m - WFL_i)}{\hat{E}^2(NM_m|WFL_m)} + \right. \\ & \left. \frac{z_{mj}(1 + \mathbf{z}_m\hat{\boldsymbol{\beta}})K_h(WFL_m - WFL_i)NM_i\hat{\beta}_j}{[1 + \hat{E}(\mathbf{z}_m|WFL_m)]^2} \right), \end{aligned}$$

wobei $K_h(\cdot)$ die Gauss-Kernfunktion ist.

Die Varianz $\boldsymbol{\Omega}$ des semiparametrischen Schätzers kann dann nach der Berücksichtigung von $\hat{\delta}_i$ mit dem Term $\mathbf{h}(\mathbf{w}_i, \hat{\boldsymbol{\beta}}, \hat{\mathbf{m}}_h)$ in der empirischen Varianz geschätzt werden. Dann lautet der konsistente Schätzer für $\boldsymbol{\Omega}$:

$$\hat{\Omega} := \frac{1}{n} \sum_{i=1}^n \hat{\psi}_i \hat{\psi}_i', \quad (\text{C.6})$$

wobei das i -te Element des (38×1) -Vektors $\hat{\psi}_i$ durch

$$\hat{\psi}_i := \mathbf{h}(\mathbf{w}_i, \hat{\beta}, \hat{\mathbf{m}}_h) + \hat{\delta}_i - \sum_{m=1}^n \frac{\hat{\delta}_m}{n}, \quad (\text{C.7})$$

definiert (vgl. NEWBY (1994a), S. 238) ist.

Literaturverzeichnis

- AIGNER, K. / OBERHOFER, W. / SCHMIDT, B. (1993): Eine neue Methode zur Erstellung eines Mietspiegels am Beispiel der Stadt Regensburg; *Wohnungswirtschaft und Mietrecht*, **1/2**, 16–21.
- ALLES, R. (1990): *Die ortsüblichen Vergleichsmieten in der Landeshauptstadt München. Gutachten zur Erstellung eines Mietspiegels nach §2 Miethöhegesetz*; Darmstadt: Institut Wohnen und Umwelt.
- ANDERSON, T. W. (1984): *An introduction to multivariate statistical analysis*; 2. Auflage, New York: Wiley and Sons.
- ANGLIN, P. M. / GENÇAY, R. (1996): Semiparametric Estimation of a Hedonic Price Function; *Journal of Applied Econometrics*, **11**, 633–648.
- ARGUEA, N. M. / HSIAO, CH. (1993): Econometric issues of estimating hedonic price functions; *Journal of Econometrics*, **56**, 243–267.
- BARLETT, M. S. (1937): The statistical conception of mental factors; *Brit. J. Psychol.*, **28**, 97–104.
- BARLETT, M. S. (1938a): Methods of estimating mental factors; *Nature London*, **141**, 609–610.
- BARLETT, M. S. (1938b): Mutivariate analysis; *J. R. Statist. Soc., Suppl. 9B*, 176–179.
- BARLETT, M. S. (1951): The effect of standardisation on an approximation in factor analysis; *Biometrika*, **38**, 337–344.
- BECKER, G. S. (1965): A theory of the allocation of time; *Economic Journal*, **75**, 493–517.
- BRACHINGER, H. W. / OST, F. (1996): *Modelle mit latenten Variablen: Faktorenanalyse*; in: FAHRMEIR, L. / HAMERLE, A. / TUTZ, G. (Hrsg.) (1996): *Multivariate statistische*

- Verfahren*; 2. Auflage, Berlin und New York: Walter de Gruyter, 639–766.
- BRACHINGER, H. W. (2002): *Statistical Theory of Hedonic Price Indices*, Working Paper (wird erscheinen in: BRACHINGER, H. W. / DIEWERT, E.: *Hedonic Methods in Price Statistics: Theory and Practice*; Heidelberg: Springer).
- BUNDESAMT FÜR WOHNUNGSWESEN (1999): *Neues Mietrecht - Grundlagen für die Botschaft*; Bern: Bundesamt für Wohnungswesen.
- BUNDESMINISTERIUM FÜR RAUMORDNUNG, BAUWESEN UND STÄDTEBAU (Hrsg.) (1980): Fortschreibung der Hinweise für die Aufstellung von Mietspiegeln, *Wohnungswirtschaft und Mietrecht*, **8**, 165–171.
- CLAR, M. (1992): Tabellen- versus Regressionsmethode bei der Mietspiegelerstellung - An-dante?; *Wohnungswirtschaft und Mietrecht*, **12**, 662–666.
- COURT, A. T. (1939): *Hedonic price indexes with automotive examples*; in: *The dynamics of automobile demand*; New York: General Motors.
- DALE-JOHNSON, D. (1982): An alternative approach to housing market segmentation using hedonic price data; *Journal of Urban Economics*, **11**, 311–332.
- DANTZIG, G. D. (1963): *Linear Programming and extensions*; Princeton: Princeton University Press.
- DEATON, A. / MUELLBAUER, J. (1980): *Economics and consumer behavior*; Cambridge, New York: Cambridge University Press.
- EFRON, B. / TIBSHIRANI, R. J. (1998): *An introduction to the Bootstrap*; Boca Raton, New York: Chapman & Hall/CRC.
- FAHRMEIR, L. / GIEGER, C. / MATHES, H. / SCHNEEWEISS, H. (1994): *Teil B: Statistische Analyse der Nettomieten*; in: *Gutachten zur Erstellung des Mietspiegels für München 1994*, München: Landeshauptstadt München Sozialreferat - Amt für Wohnungswesen.
- FAHRMEIR, L. / HAMERLE, A. / TUTZ, G. (Hrsg.) (1996): *Multivariate statistische Verfahren*; 2. Auflage, Berlin und New York: Walter de Gruyter.
- FAHRMEIR, L. / KÜNSTLER, R. / PIGEOT, I. / TUTZ, G. (1997): *Statistik-Der Weg zur Datenanalyse*; Berlin, Heidelberg: Springer.
- GORMAN, W. M. (1956): *A possible procedure for analysing quality differentials in the egg market*; Ames: Iowa State College, mimeo.

- GOODMAN, A. C. (1978): Hedonic Prices, Price Indices and Housing Markets; *Journal of Urban Economics*, **5**, 471–484.
- GORSUCH, R. L. (1983): *Factor Analysis*; 2. Auflage, Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- GREEN, P. J. / SILVERMAN, B. W. (1994): *Nonparametric Regression and Generalized Linear Models*; London: Chapman and Hall.
- GREENE, W. H. (1993): *Econometric Analysis*; 2. Auflage, New Jersey: Prentice-Hall.
- GRILICHES, Z. (1961): *Hedonic price indexes for automobiles: an econometric analysis of quality change*; in: GRILICHES, Z. (Hrsg.) (1971): *Price indexes and quality change: Studies in new methods of measurement*; Cambridge, Mass.: Harvard University Press.
- HARMAN, H. H. (1976): *Modern Factor Analysis*; 3. Auflage, Chicago: The University of Chicago Press.
- HÄRDLE, W. (1990): *Applied Nonparametric Regression*; New York, Cambridge: Cambridge University Press.
- HÄRDLE, W. / MARRON, J. S. (1991): Bootstrap simultaneous error bars for nonparametric regression; *Annals of Statistics*, **19**, 778–796.
- HÄRDLE, W. / LINTON, O. (1994): *Applied Nonparametric Methods*; in: ENGLE, R. F. / MCFADDEN, D. L. (Hrsg.) (1994): *Handbook of Econometrics*; Vol. IV, Amsterdam, New York: North-Holland, Kapitel 38.
- HOTELLING, H. (1933): Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, **24**, 417–441, 498–520.
- JAEGER, F. / BÜHLER, S. (1995): *Marktmiete. Schweizer Wohnungsmieten zwischen Politik und Markt*; Chur, Zürich: Verlag Rüegger.
- JOHNSON, R. A. / WICHERN, D. W. (1992): *Applied Multivariate Statistical Analysis*; 3. Auflage, Englewood Cliffs, New York: Prentice-Hall.
- JOLLIFFE, I. T. (1986): *Principal Component Analysis*; New York: Springer.
- JUDGE, G. G. / GRIFFITHS, W. E. / CARTER HILL, R. / LÜTKEPOHL, H. / LEE, T.-C. (1985): *The Theory and Practice of Econometrics*; 2. Auflage, New York: John Wiley and Sons.

- KAISER, H. F. (1958): The varimax criterion for analytic rotation in factor analysis, *Psychometrika*, **23**, 187–200.
- KOFNER, S. (1999): *Der Vergleichsmietenbegriff des Gesetzes zur Regelung der Miethöhe*; Hochschule für Technik, Wirtschaft und Sozialwesen Zittau/Görlitz.
- KRÄMER, W. (1992a): Pro und Contra die Erstellung von Mietspiegeln mittels Regressionsanalyse, *Wohnungswirtschaft und Mietrecht*, **4**, 172–175.
- KRÄMER, W. (1992b): *Gutachten zur statistischen Methodik des Mietspiegels für München '91 -Stand April 1991- (Fortschreibung des Mietspiegels '90)*; Dortmund.
- KRÄMER, W. (1993): *Gutachten zur statistischen Fundierung des Mietspiegels der Landeshauptstadt Kiel*; Dortmund.
- KRZANOWSKI, W. J. (1984): Sensitivity of Principal Components; *Journal of the Royal Statistical Society, Series B*, **46**, 558–563.
- LAMBELET, J.-C. / ZIMMERMANN, CH. (1991): *Droit au logement ou économie de marché ? Une analyse de l'immobilier en Suisse*; Lausanne: Editions Payot Lausanne.
- LANCASTER, K. J. (1966): A new approach to consumer theory; *Journal of Political Economy*, **74**, 132–157.
- LANCASTER, K. J. (1991): *Modern Consumer Theory*; Aldershot: Edward Elgar Publishing Limited.
- LANDESHAUPTSTADT MÜNCHEN SOZIALREFERAT - AMT FÜR WOHNUNGSWESEN (Hrsg.) (1994): *Gutachten zur Erstellung des Mietspiegels für München 1994*; München.
- MAGISTRAT DER STADT DARMSTADT (1977): *Mietspiegel für Darmstadt. Ortsübliche Vergleichsmieten für nicht preisgebundenen Wohnraum auf der Grundlage der Repräsentativerhebung vom März 1977*; Darmstadt.
- MAS-COLELL, A. / WHINSTON, M. D. / GREEN, J. R. (1995): *Microeconomic Theory*; Oxford, New York: Oxford University Press.
- MIETHÖHEGESETZ ODER GESETZ ZUR REGELUNG DER MIETHÖHE vom 18.12.1974.
- MIETRECHTSREFORMGESETZ ODER GESETZ ZUR NEUGLIEDERUNG, VEREINFACHUNG UND REFORM DES MIETRECHTS vom 19. Juni 2001, in BUNDESGESETZBLATT (2001), Teil I, Nr. 28; Berlin: Deutscher Bundestag.

- MORESI, S. (1989): *Hedonic prices and functional specification: An empirical analysis of the urban housing market*, MIT Econometric Research Paper; Massachusetts: MIT.
- MÜLLER, M. / FAN, J. (1995): *Density and Regression Smoothing*; in HÄRDLE, W. / KLINKE, S. / TURLACH, B. A. (Hrsg.) (1995): *XploRe: An Interactive Statistical Computing Environment*; New York, Berlin: Springer-Verlag, Kapitel 5.
- NEUE ZÜRCHER ZEITUNG (NZZ) (6. Juni 2002): *Mehr Markt im Mietrecht*, 11.
- NEWBY, W. K. / MCFADDEN, D. (1994): *Large Sample Estimation and Hypothesis Testing*; in ENGLE, R. F. / MCFADDEN, D. L. (Hrsg.) (1994): *Handbook of Econometrics*; Vol. IV, Amsterdam, New York: North-Holland, Kapitel 36.
- NEWBY, W. K. (1994a): Kernel Estimation of Partial Means and a General Variance Estimator; *Econometric Theory*, **10**, 233–253.
- NEWBY, W. K. (1994b): The Asymptotic Variance of Semiparametric Estimators; *Econometrica*, **62**, 1349–1382.
- OBERHOFER, W. / SCHMIDT, B. (1993): Das Mietspiegelproblem - eine unendliche Geschichte?; *Wohnungswirtschaft und Mietrecht*, **10**, 585–588.
- OBERHOFER, W. (1994): *Die Methodik der Mietspiegelerstellung. Arbeitsbericht zur Erstellung des Frankfurter Mietspiegels 1994*; Teil 2, Regensburg.
- PEARSON, K. (1901): On lines and planes of closest fit to a system of points in space; *Phil. Mag.*, **2**, 6th series, 557–572.
- REDONDO, M. J. (1996): *Varianzanalyse: eine Anwendung zur Klassenbildung im Rahmen eines Tabellenmietspiegels*, unveröffentlicht.
- RIPS, F.-G. (1996): Vereinfachung und Reformen im Mietrecht. Teil II: Gesetzliche Regelungen über Mietspiegel; *Wohnungswirtschaft und Mietrecht*, **5**, 255–258.
- ROBINSON, P. M. (1988): Root- N -Consistent Semiparametric Regression; *Econometrica*, **56**, N. 4, 931–954.
- RONNING, G. (1993): *Gutachten im Zusammenhang mit Auseinandersetzungen über die Aussagefähigkeit des Münchner Mietspiegels*; Tübingen.
- RONNING, G. (1995): Ökonometrie für jedermann: Der Mietspiegel als Regressionsmodell; *IAW Mitteilungen*, **3**, 4–10.

- RONNING, G. (1996): Mietspiegel in der Diskussion – Ein Überblick; *Wohnungsmonitor Baden-Württemberg*, **3**, 16–25.
- ROSEN, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition; *Journal of Political Economy*, **82**, 34–55.
- SAS INSTITUTE INC. (1990): *SAS/STAT User's Guide*; Vol. I, Version 6, 4. Auflage, Cary, NC: SAS Institute Inc..
- SCHIPS, B. / MÜLLER, E. (1991): *Der Schweizer Wohnungsmarkt: Ist-Zustand und alternative Lösungsvorschläge*, St. Gallen: Forschungsstelle für empirische Wirtschaftsforschung.
- SILVERMAN, B. W. (1986): *Density Estimation*; London, New York: Chapman and Hall.
- STIGLER, G. J. (1945): The cost of subsistence, *Journal of Farm Economics*, **27**, 303–314.
- THOMSON, G. H. (1934): Hotelling's method modified to give Spearman's "g", *Journal of Educational Psychology*, **25**, 366–374.
- THOMSON, G. H. (1951): *The factorial analysis of human ability*; London: London University Press.
- TRIPLETT, J. E. (1986): The economic interpretation of hedonic methods, *Survey of Current Business*, 36–40.
- VIERTES MIETRECHTSÄNDERUNGSGESETZ ODER GESETZ ZUR ÄNDERUNG MIETRECHTLICHER VORSCHRIFTEN vom 21. Juli 1993.
- WALSER, H. / HILTON, P. (2001): *The Golden Section*; Washington: The Mathematical Association of America.
- WULLKOPF, U. (1996): Bedeutungen des Mietspiegels, *Wohnungswirtschaft und Mietrecht*, **8**, 455–458.

Namenverzeichnis

- Anderson, T. W., 78
- Aigner, K., 41
- Alles, R., 11, 12
- Anderson, T. W., 83
- Anglin, P. M., 110
- Arguea, N. M., 34
- Bühler, S., 6, 18
- Barlett, M. S., 81, 83
- Becker, G. S., 24
- Brachinger, H. W., 23, 73, 75, 76, 78–81
- Bundesamt für Wohnungswesen, 4
- Bundesministerium für Raumordnung,
66
- Clar, M., 15
- Court, A. T., 22
- Dale-Johnson, D., 34
- Dantzig, G. D., 28
- Deaton, A., 24, 28, 30, 32
- Efron, B., 56–58
- Fahrmeir, L., 42, 43, 49, 50
- Fan, J., 52, 99
- Gençay, R., 110
- Goodman, A. C., 34
- Gorman, W. M., 24, 28
- Gorsuch, R. L., 80
- Green, P. J., 60, 62
- Greene, W. H., 35, 123
- Griliches, Z., 22
- Härdle, W., 45, 53–55, 57, 59, 60
- Hamerle, A., 49, 50
- Harman, H. H., 80
- Hotelling, H., 76
- Hsiao, Ch., 34
- Jaeger, F., 6, 18
- Johnson, R. A., 81, 83
- Jolliffe, I. T., 85
- Judge, G. G., 103
- Kaiser, H. F., 80
- Kofner, S., 7
- Krämer, W., 10, 36, 40, 41, 66
- Krzanowski, W. J., 85
- Lambelet, J.-C., 6
- Lancaster, K. J., 22–25, 28, 29, 32, 34
- Landeshauptstadt München, 40
- Linton, O., 53–55
- Müllbauer, J., 24, 28, 30
- Müller, E., 18

Müller, M., 52, 99

Marron, J. S., 59

Mas-Colell, A., 26–28

McFadden, D., 123, 124

Moresi, S., 34

Muellbauer, J., 32

Newey, W. K., 123–125

Oberhofer, W., 13, 40, 41

Ost, F., 73, 75, 76, 78–81

Pearson, K., 76

Redondo, M. J., 16

Rips, F.-G., 7

Robinson, P. M., 61, 97, 108

Ronning, G., 5, 8–10, 13, 36, 42

Rosen, S., 22, 34

SAS Institute, 83

Schips, B., 18

Schlittgen, R., 16

Schmidt, B., 13, 41

Silverman, B. W., 48, 50–52, 60, 62

Stigler, G. J., 28

Thomson, G. H., 78, 81

Tibshirani, R. J., 56–58

Triplett, J. E., 23

Tutz, G., 49, 50

Uhlig, S., 16

Walser, H., 59

Wichern, D. W., 81, 83

Wullkopf, U., 7, 10

Zimmermann, Ch., 6

Index

- Bandbreite, 49, 51
 - optimale, 51
 - Schätzung, 52
 - Silverman's Faustregel, 51
- Basismiete, 42
- Bootstrap, 55
 - BC_a -Methode, 57
 - Bootstrap-Replikat, 57
 - Bootstrap-Residuum, 59
 - Bootstrap-Stichprobe, 55
 - naiver, 57
 - Quantil-Intervall-Methode, 56, 58
 - wilder, 59, 106
- Charakteristika-Grenze, 32
- Dichtefunktion
 - gemeinsame, 49
 - Interquantilabstand, 52
 - Schätzung, 49
 - Standardabweichung, 51
- Dichteschätzung, 48
- Diracsche Sprungfunktion, 59
- Einbeziehungsfrist, 9, 10
- Erwartungswert
 - bedingte, 45
- Fahrmeir Modell, 42
- Faktorenanalyse, 72
 - Einzelrestfaktoren, 73
 - Einzelrestvarianzmatrix, 74
 - Fundamentaltheorem, 75
 - gemeinsame Faktoren, 73
 - Grundgleichung, 75
 - Ladungen, 73
 - orthogonale Transformation, 80
 - Regressionsmethode, 81
 - Varimax-Methode, 80
- Grundmiete, 42
- halbe Interquartilabstand, 109
- Hauptfaktorenanalyse, 78
- Hauptkomponentenanalyse, 76
- Hauptkomponentenmethode, 77
 - Eigenwertkriterium, 78
 - Prozentsatzkriterium, 78
 - Signifikanztest, 78
- hedonische Funktion, 23
- hedonische Hypothese, 22
- hedonische/implizite Preis, 23
- Kappungsgrenzen, 8
- Kern, 50
 - Gauss-Kern, 50

- Kerndichteschätzer, 48
 - asymptotische Verteilung, 51
- Kernschätzer, 51
- Kleinste-Quadrate-Methode
 - gewichtete, 81
 - nichtlineare, 103
- Komunalität, 75
- Kostenmiete, 4, 16
- kurzfristige Grenzkosten, 26
- lineare Modell für Charakteristika, 28
- Marktmiete, 4, 18
- Mietspannen, 15
- Mietspiegel, 3
 - Nachteile, 10
 - Regressionsmietspiegel, 12, 15
 - relevante Wohnungen, 8, 10
 - Tabellenmietspiegel, 12
 - Vorteile, 9
 - Zweck, 7
- Nadaraya-Watson-Kernschätzer, 49
 - asymptotische Normalität, 53
 - Konfidenzintervalle
 - approx. asymptotische punktweise, 54
 - mittels Bootstraps, 55
 - Konsistenz, 53
 - Verzerrung, 54
- Nettomiete, 11, 41
- Nettomietfaktor, 42
- Produktionsfunktion, 28
- Produktionstheorie des Haushalts, 24
- Regensburger Modell, 41
- Regressionsmodelle
 - additiv-multiplikative semiparametrische, 96
 - additive semiparametrische, 60
 - partielle lineare, 45
 - semiparametrische, 44
- semiparametrischer Schätzer, 61, 105
 - asymptotische Normalität, 61
 - asymptotische Verteilung, 105
 - Konsistenz, 62
 - Varianz-Kovarianz-Matrix, 62
- Speckman's Algorithmus, 60
- Transformationmatrix, 29
- Vergleichsmiete, 19
 - ortsübliche, 4
- Wohnungsmärkte, 5
- Zweipunkt-Verteilung, 59