Die approbierte Originalversion dieser Diplom-/Masterarbeit ist in der Hauptbibliothek der Technischen Universität Wien aufgestellt und zugänglich.

http://www.ub.tuwien.ac.at

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology.

http://www.ub.tuwien.ac.at/eng

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Diplomarbeit

Räumliche Effekte in hedonischen Preismodellen:

Eine Untersuchung für den Wiener Büroimmobilienmarkt

ausgeführt zum Zwecke der Erlangung des akademischen Grades des Diplom-Ingenieurs unter der Leitung von

A.o.Univ.Prof. Dipl.-Ing. Dr.techn Wolfgang Feilmayr

E 280 Fachbereich für Stadt- und Regionalforschung

eingereicht an der Technischen Universität Wien

Fakultät für Architektur und Raumplanung

von

Jakob Ransmayr, BSc

Matrikelnummer: 0727189

Gross-Dörnbach-Str. 8

4073 Wilhering

Wien, am 31. März 2015 Jakob Ransmayr

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich meine Diplomarbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur genannt habe. Ich versichere, dass ich diese Arbeit bisher weder im In- noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe.

Wien, am 31. März 2015 Jakob Ransmayr

Danksagung

Mein Dank gilt ao.Univ.Prof. Dipl.-Ing. Dr.techn. Wolfgang Feilmayr für die Betreuung der Arbeit, die entgegengebrachte Zeit, das für die Arbeit so wichtige Feedback und nicht zuletzt für die Möglichkeit die Arbeit am Fachbereich für Stadt- und Regionalforschung an der TU Wien verfassen zu können.

Darüber hinaus möchte ich mich bei Dr. Wolfgang Brunauer (Real(e)value Immobilien Bewertungs GmbH) bedanken, welcher mit seinen Anregungen und seinem Interesse an der Thematik geholfen hat im Zuge der Arbeit die richtigen Fragen zu stellen und dadurch die Arbeit bereichern konnte. Für die Einladung zum Workshop „Der Weg zu einem Preisindex für österreichische Gewerbe- und Wohnimmobilien“ und den kurzen Einblick in die Arbeit der Österreichischen Nationalbank möchte ich mich bei Fr. Dr. Karin Wagner bedanken. Die Expertise und Hilfe von Univ.Ass. Dipl.-Ing. Robert Kalasek und Univ.Ass. Dipl.-Ing. Kurt Weninger in Bezug auf die Arbeit mit Geodaten hat in der Phase der Datenerhebung und Modellierung viel Zeit erspart, wofür ich mich herzlich bedanken möchte.

Bedanken möchte ich mich auch bei meinen Freunden und Studienkollegen an der TU Wien und der ETH Zürich für die gute Zeit.

Besonderer Dank gilt meiner Familie, meiner Schwester Juliane Ransmayr für das Korrekturlesen der Arbeit und im speziellen meinen Eltern, Gudrun und Albert Ransmayr, welche mir dieses Studium ermöglicht haben und mir immer die notwendige Unterstützung entgegenbrachten. Vielen Dank!

Kurzfassung

Die Arbeit analysiert den Büromietenmarkt in Wien und geht der Frage nach, welche Eigenschaften von Immobilie und Standort den Mietpreis determinieren. Die Datenbasis umfasst ca. 6000 Angebots- und Transaktionsdaten. Diese wurden mit in GIS kalkulierten Lage- und Nachbarschaftscharakteristika angereichert.

Im Zuge der Arbeit wurden hedonische Preismodelle geschätzt, welche in der Lage sind, die marginalen Preiseffekte von unterschiedlichen Struktur-, Lage- und Nachbarschaftscharakteristika auf die Miete von Büroimmobilien zu bewerten. Es wurden lineare Regressionsmodelle (Basis-Modell) mit räumlichen Regressionsmodellen verglichen. Durch die Anwendung von Spatial Lag Modellen kann der Modell-Fit im Vergleich zu den OLS geschätzten Modellen deutlich gesteigert werden. Dies bedeutet, das traditionelle hedonische Preismodelle, welche räumliche Strukturen nicht berücksichtigen, zu verzerrten Schätzern führen können.

Als Preisdeterminanten zeigen sich in der überwiegenden Zahl der Modelle das Baualter, der Zustand der Immobilie, das Vorhandensein eines Liftes, die Zugehörigkeit zu einem Bürocluster, die Distanz zum Stephansdom sowie die Distanz zum höherrangigen öffentlichen Verkehrsnetz als statistisch signifikant.

Keywords: Hedonische Preismodelle, Räumliche hedonische Regressionsmodelle, Büromieten Wien

Abstract

This thesis analyses the office market in Vienna to answer the question of whether the implicit value of the structural and locational characteristics of real estate in the office sub-market is capitalized in the office rent. The sample contains about 6000 asking and transaction listings of office rental units. Location and neighbourhood characteristics were calculated by using GIS software.

This thesis uses a hedonic modelling approach for estimating the office price determinants. Besides linear hedonic regression models (base model), spatial hedonic regression models were used. Compared to the traditional linear ordinary least squares approach, spatial lag models can significantly increase the overall model fit. While controlling for spatial effects by applying spatial models helps to estimate more efficient parameters.

The age of the office unit, the condition of the property, the presence of a lift, the distance to the CBD and the distance to the public transport system add significant explanatory power to most of the tested models.

Keywords: Hedonic pricing model, spatial hedonic regression, office rent Vienna

Inhaltsverzeichnis

Inhaltsverzeichnis **v**

Abbildungsverzeichnis **vii**

Tabellenverzeichnis **ix**

1 Einleitung **1**
1.1 Motivation und Problemstellung 1
1.2 Forschungsfragen und Ziele der Arbeit 1
1.3 Struktur der Arbeit 2

2 Theoretische Grundlagen **3**
2.1 Büroimmobilienmarkt in Wien 3
2.2 Die hedonische Theorie 6
2.3 Determinanten des Preises 9
2.4 Räumliche Effekte 18

3 Methodik **21**
3.1 Lineare Regressionsmodelle 21
3.2 Räumliche Regressionsmodelle 24
3.3 Explorative (räumliche) Datenanalyse 31

4 Daten **33**
4.1 Datenbasis 33
4.2 Datenaufbereitung und deskriptive Statistik 33

5 Modellbildung **51**
5.1 Methodischer Ablauf 51
5.2 Geschätzte Modelle 52
5.3 Modellvergleich 63

6 Schlussbetrachtung und Ausblick **69**

Literaturverzeichnis **73**

A Verwendete Geodaten **79**

B OLS Diagnostik **81**

Abbildungsverzeichnis

2.1 Räumliche Systemgrenze: Wiener Stadtgebiet, eigene Darstellung. 4

3.1 Räumliche Gewichtungsfunktion, FOTHERINGHAM et al. (2002, S.44) 26
3.2 Adaptive Gewichtungsfunktion, FOTHERINGHAM et al. (2002, S.47) 27

4.1 Boxplot: Mietpreise pro m^2 nach Wiener Bezirke, eigene Darstellung. 35
4.2 Mietpreise pro m^2 (Median) auf einer Bezugsfläche von 300 mal 300 Metern, eigene Darstellung. 36
4.3 Anteil der Mietflächen an allen Mietflächen im Datensatz auf einer Bezugsfläche von 300 mal 300 Metern, eigene Darstellung. 37
4.4 Boxplot: Alter der Immobilien nach Wiener Bezirke, eigene Darstellung. 39
4.5 Boxplot: Mietpreise pro m^2 nach Zustandsklasse, eigene Darstellung. 40
4.6 Nettomiete Gesamt bezogen auf die Fläche der Immobilie, eigene Darstellung. 41
4.7 Mietpreis pro m^2 bezogen auf die Fläche der Immobilie, eigene Darstellung. 42
4.8 Boxplot: Mietpreise pro m^2 nach Angebotsjahr, eigene Darstellung. 42
4.9 Boxplot: Distanz bis zur nächsten ÖV Station unterschieden nach Verkehrsmittel, eigene Darstellung. 45
4.10 Boxplot: Distanz bis zum Stephansplatz und zur nächsten Autobahnauffahrt, eigene Darstellung. 46
4.11 Dichte der Beobachtungen in Bezug zur Distanz zum Stephansplatz und Miete pro m^2 in Bezug zur Distanz zum Stephansplatz mit LOWESS Fit, eigene Darstellung. 48
4.12 GWR Koeffizienten, v.l.n.r. Zeile 1: lokale R^2, Koeffizient Distanz zur S-Bahn; Zeile 2: Koeffizient logarithmierte Größe der Immobilie, Koeffizient Distanz zur Autobahnauffahrt; Zeile 3: Koeffizient logarithmierte Distanz zur U-Bahn, eigene Darstellung. 49
4.13 Density Scatter Plot: Dichte der Beobachtungen im Untersuchungsgebiet, eigene Darstellung. 50
4.14 Ad-hoc Cluster: dunkelrot = Innere Stadt, rot = Innere Bezirke, gelb = Donau City, eigene Darstellung. 50

5.1 Darstellung des methodischen Ablaufs der Datenanalyse, eigene Darstellung. 67

B.1 Modelldiagnostik OLS_1: QQ-Plot der Residuen und Residuenplot, eigene Darstellung. 81
B.2 Modelldiagnostik OLS_2: QQ-Plot der Residuen und Residuenplot, eigene Darstellung. 82
B.3 Modelldiagnostik OLS_3: QQ-Plot der Residuen und Residuenplot, eigene Darstellung. 82

Tabellenverzeichnis

2.1 Überblick - Studien zum Büroimmobilienmarkt 11
2.2 Determinanten des Preises von Büroimmobilien - Erklärende Variablen mit erwarteten Vorzeichen des Zusammenhanges und Stärke des Einflusses 17

4.1 Übersicht der Modellvariablen - Bauliche Charakteristika 36
4.2 5 Punkt Zusammenfassung - Größe der Immobilien 38
4.3 Übersicht der Modellvariablen - Vertragliche Charakteristika . . . 40
4.4 Übersicht der Modellvariablen - Lage- und Nachbarschaftscharakteristika 43
4.5 Ergebnisse GWR 47

5.1 Ergebnisse Modell OLS_1 53
5.2 Ergebnisse Modell OLS_2 55
5.3 Ergebnisse Modell OLS_3 57
5.4 Moran's I Test - Abhängige Variable Mietpreis 59
5.5 Ergebnisse Modell SAR_{err} mit Distanzgewichtung und k-nearest-neighbour (knn=14) 61
5.6 Ergebnisse Modell SAR_{lag} mit Distanzgewichtung und k-nearest-neighbour (knn=14) 63
5.7 Ergebnisse der besten Modelle 65

Kapitel 1

Einleitung

„Human beings are pattern-seeking, story-telling animals.“ (vgl. Leamer, 2008)

1.1 Motivation und Problemstellung

Die Standortwahl von Unternehmen ist von einer Vielzahl harter und weicher Faktoren abhängig. Diese Faktoren werden bei der Wahl des Standortes durch die jeweiligen Unternehmen indirekt bewertet. Diese Bewertung spiegelt sich in den Preisen von Büroimmobilien als implizite Preise der Eigenschaften von Immobilie und Standort wieder (implizite Eigenschaftsmärkte) (vgl. Rosen, 1974). Ein tieferes Verständnis über die Preisdeterminanten von Immobilien bedeutet somit gleichzeitig ein tieferes Verständnis der Standortwahl von Unternehmen (vgl. Caduff, 2013). Dieses Verständnis der Dynamik der Standortwahl und deren Faktoren ist für die räumliche Planung essenziell.
Hedonische Preismodelle stellen einen effizienten Ansatz zur Erklärung des Immobilienpreises sowie zur Generierung von Wissen über Preisdeterminaten dar (vgl. Rosen, 1984). Vermehrt wird aber in Studien darauf hingewiesen, dass einfache lineare Regressionsmodelle die Muster räumlicher Abhängigkeiten und Heterogenität ignorieren und daher zu verzerrten und ineffizienten Schätzwerten führen können (vgl. Löchl, 2010). Das Problem der räumlichen Effekte kann durch die Einführung von räumlichen Regressionsmodellen teilweise gelöst werden (vgl. Anselin, 1988b; LeSage und Pace, 2009; Fotheringham et al., 2002).

1.2 Forschungsfragen und Ziele der Arbeit

Die geografische Variation der Büromieten wird durch einen (unbekannten) räumlichen Prozess generiert (vgl. Gret-Regamey und van Strien, 2013). Ziel der Arbeit ist es diesen räumlichen Prozess, welcher zu im geografischen Raum variierenden Merkmalsausprägung führt, besser zu verstehen. Aufbau-

end auf der oben dargestellten Problemstellung und Zieldefinition ist die Beantwortung folgender Forschungsfragen relevant:

- Welche Variablen können als Determinanten von Büromieten identifiziert werden?
- Welche räumliche Variation weist der Einfluss von ausgewählten Struktur- und Lageparameter auf den Mietpreis im Untersuchungsgebiet auf?
- Kann die Einführung von räumlichen Regressionsmodellen eine Verbesserung der Erklärungsgüte bewirken?

1.3 Struktur der Arbeit

Im Abschnitt 2 werden die theoretischen Grundlagen im Kontext der Forschungsfragen erarbeitet. Im Vordergrund steht die Auseinandersetzung mit dem Wiener Büroimmobilienmarkt, der Theorie der hedonischen Preismodelle, der möglichen Preisdeterminanten für Büroimmobilien und die Frage nach räumlichen Effekten bei Daten mit geografischen Bezug.
Abschnitt 3 stellt die zur Anwendung kommende Methodik vor. In diesem Kapitel werden die mathematischen und statistischen Grundlagen erarbeitet, welche in Abschnitt 5 zur Anwendung kommen. Abschnitt 4 stellt die Datenbasis vor.

Kapitel 2

Theoretische Grundlagen

2.1 Büroimmobilienmarkt in Wien

Dieser Abschnitt schafft die im Zuge der Arbeit notwendige räumliche, zeitliche und sachliche Systemabgrenzung. Die Untersuchungen der Arbeit werden sich auf das Stadtgebiet der österreichischen Bundeshauptstadt beschränken. Zur Verfügung stehen Transaktion- resp. Angebotsdaten für das Büroimmobiliensegment aus den letzten zehn Jahren.

Das Untersuchungsgebiet

In Wien leben verteilt auf 23 Verwaltungsbezirke (vgl. Abbildung 2.1) mit einer Fläche von insgesamt 415 Quadratkilometern (WIEN.GV.AT, 2014a) etwa 1,73 Mio. Menschen (UNDATA, 2012).

Die Agglomeration Wien wuchs seit den den frühen 90er Jahren, nach einer kurzen Phase der Schrumpfung und Deurbanisierung zwischen 1971 und 1981 bedingt durch die internationale Ölkrise und der fortschreitenden Deindustralisierung, wieder stärker. Besonderen Einfluss auf die Entwicklung der Bevölkerungsstruktur nimmt der Fall des Eisernen Vorhangs und die EU Osterweiterung, wie auch die Erreichbarkeitsgewinne durch den Ausbau der Verkehrsinfrastruktur. Bei der Bevölkerungsentwicklung in den letzten Zehn Jahren lassen sich unterschiedliche Phasen unterscheiden. In den frühen 1990er-Jahren kam es zu enormen Bevölkerungswachstum, welcher hauptsächlich durch den Krieg in Exjugoslawien und den Zuzug durch die Reformländer zu begründen ist (PIRHOFER und STIMMER, 2013). In der zweiten Hälfte der 1990er-Jahre kam es wieder zu einem leichten Rückgang der Bevölkerung. Allgemein spielt der Einfluss der internationalen Migration auf die Einwohnerzahlen in der Vienna Region eine große Rolle. Darüber hinaus ist ein anhaltender Trend einer Suburbanisierung zu erkennen. Bei einer Betrachtung der Vienna Region und deren Bevölkerungsentwicklung in den letzten zwanzig Jahren ist eine klare Entwicklung hin zu gut erreichbaren Stadtrandlagen zu erkennen. Bis 2030

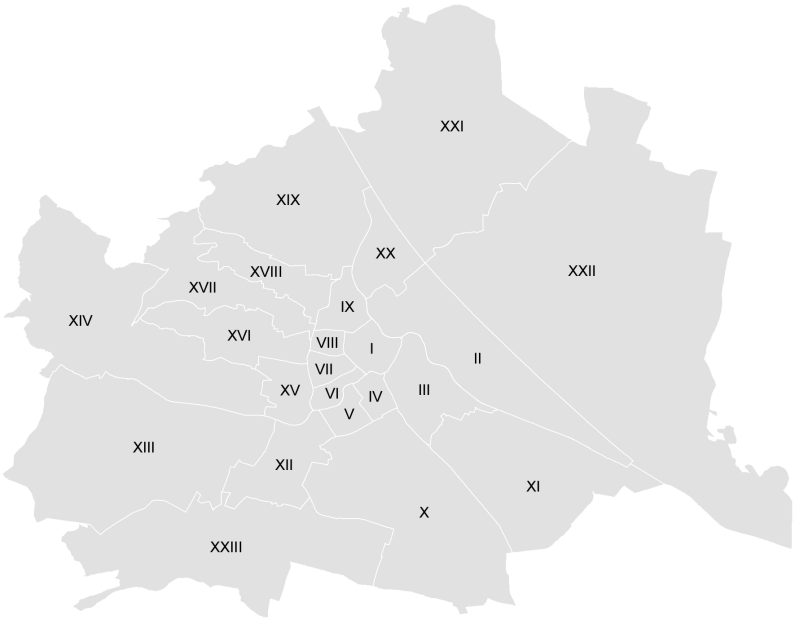

Abbildung 2.1: Räumliche Systemgrenze: Wiener Stadtgebiet, eigene Darstellung.

wird die Bevölkerung auf 2 Millionen anwachsen (ÖIR und Europaforum-Wien, 2014).
Nicht nur in Hinblick auf die Bevölkerungsentwicklung nimmt Wien in Österreich eine besondere Stellung ein. Auch als Arbeitsmarktzentrum ist Wien von großer Bedeutung. Laut Untersuchungen der Statistik Austria beweisen die österreichischen Erwerbstätigen bei der Wahl ihrer Arbeitsstätte große Mobilität. Mehr als 500.000 Erwerbstätige in Österreich verlassen zur Arbeit ihren Wohnstandort. Wien als Bundeshauptstadt verzeichnet den größten Einpendleranteil an allen Städten Österreichs. 28 Prozent oder etwa 247.000 der in Wien beschäftigten Menschen pendeln täglich ein. Wien ist (neben Salzburg) das einzige Bundesland, welches ein positives Pendlersaldo aufweist, d.h. die Zahl der Erwerbstätigen in Wien ist größer als die Anzahl der Beschäftigten (Statistik-Austria, 2014).
Die Betrachtung von Landnutzung und Verkehrsinfrastruktur ist immer integral zu betrachten, was dazu führt, das eine Zusammenhangsanalyse zwischen Büromarkt resp. Bürostandorten im Lichte der verkehrlichen Erreichbarkeit geschehen muss.

Die Büroimmobilie

Nach Bone-Winkel et al. (2008, S.16) sind Immobilien *„[...] Wirtschaftsgüter, die aus unbebauten und bebauten Grundstücken mit dazugehörigen Gebäuden und Außenanlagen bestehen. Sie werden von Menschen im Rahmen*

von physisch-technischer, rechtlicher, wirtschaftlicher und zeitlicher Grenzen für Produktions-, Handels-, Dienstleistungs- und Konsumzwecke genutzt“. Aufbauend auf dieser Definition kann im Zuge der Arbeit eine Abgrenzung des Büroimmobiliensegments vorgenommen werden: Die Büroimmobilien ist als Teil eines Gebäudes zu sehen, welcher vorrangig für Dienstleistungszwecke genutzt wird und auf Grund dieser Nutzungsausrichtung, seiner Charakteristika und sonstiger Rahmenbedingungen auf dem Büroimmobilienmarkt einen bestimmten Mietpreis erzielt.

Der Büroimmobilienmarkt in Wien

Im internationalen Vergleich ist der Büroimmobilienmarkt in Wien vor allem was die Mieten betrifft als stabiler Markt anzusehen, welcher sich durch eine geringe Volatilität auszeichnet (vgl. CBRE, 2014; Bank-Austria, 2014; EHL, 2014). Begründet wird diese Entwicklung unter anderem durch die geringe Neuflächenproduktion im Bürosegment bei gleichzeitig in etwa konstanter Vermietungsleistung (vgl. Bank-Austria, 2014), was wiederum zu einer Leerstandsrate auf einem stabilen Niveau von 4 bis 7 Prozent bei der Betrachtung der letzten 10 Jahre führt (vgl. CBRE, 2014). Die Neuflächenproduktion am Büroimmobilienmarkt lag in den letzten 15 Jahren zwischen 100.000 m^2 und 350.000 m^2 pro Jahr wobei abgesehen von den Boomjahren 2007 und 2008 eine abnehmende Tendenz zu erkennen ist. Nach Daten von CBRE liegt die Fertigstellungen in Wien für das Jahr 2014 bei 114.000 m^2 und somit fast 27 Prozent unter dem Wert von 2013. Im Vergleich zum Jahr 2013 folgt der Take-up von Büroflächen im Jahr 2014 einem ähnlichen Muster wie die Fertigstellungen und sinkt leicht. Bei einer Betrachtung der letzten 10 Jahre ist erkennbar, dass nach dem Boomjahren in 2007 und 2008 ein signifikanter Einbruch der Vermietunsleistungen erkennbar ist und seit dem zwischen 200.000 m^2 pro Jahr und 300.000 m^2 pro Jahr liegt (vgl. CBRE, 2014; Bank-Austria, 2014). Ebenfalls seit dem Boomjahren rückläufig sind die Spitzenrenditen und liegen im Bürosegment für das Jahr 2014 unter 5 Prozent. Der Bestand von Büroflächen wird in Wien auf etwa 10,85 Millionen m^2 geschätzt(vgl. CBRE, 2014).
Nach wie vor sind die wichtigsten Bürozentren die Innere Stadt und die innerhalb des Gürtels liegenden Bezirke (CBD). Darüber hinaus ist die Donau City mit der UNO City und einigen großvolumigen Bürotürmen ein attraktiver Bürostandort (vgl. Bank-Austria, 2014). Dies zeigt sich auch in den erzielbaren Spitzenmieten. Diese liegt für den CBD im vierten Quartal 2014 bei 25,75 Euro pro m^2 und Monat (vgl. CBRE, 2014; Bank-Austria, 2014). Neben den oben genannten Bürostandorte, konnten sich in den letzten Jahren die Lagen im Süden Wiens um den Wienerberg, das Areal um den neuen Wiener Hauptbahnhof sowie die Gebiete um den Praterstern und der Erdbergstraße etablieren.

2.2 Die hedonische Theorie

Das Konzept der hedonischen Preise ist als Methodik zur Erklärung von Preiseffekten einzelner Eigenschaften eines heterogenen Gutes auf den Gesamtpreis, einem Teilgebiet der Wirtschaftswissenschaften, der Ökonometrie zuzurechnen. Das Wesen der Ökonometrie kann durch das Zusammenspiel zwischen ökonomischer Theorie, Mathematik und Statistik beschrieben werden. Ziel der Ökonometrie ist die Erforschung und Durchleuchtung von Phänomenen der realen Welt unter Verwendung von mathematischen und statistischen Methoden, sie gibt ökonomischen Theorien den nötigen empirischen Gehalt und kann zur Theoriendiskriminierung beitragen (vgl. Maddala und Kajal, 2009; Stocker, 2013).

Wenn wir nun die Miete für unsere Wohnung zahlen, so zahlen wir nicht für das Gut "Wohnung" an sich, sondern für z.B. ein bestimmtes Maß an Komfort, Erreichbarkeit, Zentralität oder Ruhe. Im Auswahlprozess um das Gut "Wohnung" spiegelt die Zahlungsbereitschaft für die einzelnen Ausprägungen unseren erwarteten Nutzen wieder und offenbart unsere Präferenzen(vgl. Haase, 2011). Um Präferenzen von Marktteilnehmern zu messen, werden "Stated-Preference Analysen" (direkte Verfahren) oder "Revealed-Preference Analysen" (indirekte Verfahren) angewandt, wobei die hedonische Methode der zweiten Kategorie zuzuordnen ist.

Die hedonische Methode wurde erstmals in den 1920er Jahren zur Kalkulation von Preisen für Agrarflächen in den US Bundesstaaten Minnesota (Haas, 1922) und Iowa (Wallace, 1926) verwendet (vgl. Colwell und Dilmore, 1999), wobei der Begriff "hedonisch" erst später geprägt wurde und eine ökonomisch-theoretische Fundierung der Methode bisweilen fehlte. Breitere Bekanntheit erreichte die Methode und erstmals der Begriff der hedonischen Preise durch die Arbeiten von Court (1939) und Griliches (1961), welche an hedonischen Preisindizes für die Automobilinstudrie arbeiteten (vgl. Goodman, 1998). Die vermutlich wichtigsten Beiträge wurden von Lancaster (1966) und Rosen (1974) erarbeitet. Einen guten Überblick über die historische Entwicklung der Methode der hedonischen Preise gibt Herath und Maier (2010) sowie Malpezzi (2002).

Mit Lancasters neuen Ansätzen zur Konsumtheorie wird mit traditionellen Konzepten gebrochen und die Idee der impliziten, nutzenstiftenden Charakteristika von Gütern wird in der mikroökonomischen Theorie verankert. Lancaster (1966, S.133) schreibt dazu: *„The chief technical novelty lies in breaking away from the traditional approach that goods are the direct objects of utility and, instead, supposing that it is the properties or characteristics of goods from which utility is derived. "* Anwendung findet das Konzept von Lancaster auf unterschiedliche Teilbereiche der Ökonomie, allerdings scheint die

grundsätzliche Eignung des Ansatzes für den Immobilienmarkt offensichtlich zu sein. Im Entscheidungsprozess um eine bestimmte Immobilie an einem bestimmten Standort wird der Mieter seine Präferenzen offenbaren, indem er den einzelnen impliziten Ausprägungen der Immobilie einen bestimmten Nutzwert zuordnet. Die Summe dieser Nutzwerte stellt den Präferenzwert dar, welcher als Maß für den Erfüllungsgrad des erwarteten Nutzens und der Bedürfnisbefriedigung anzusehen ist. Unter den Prämissen der Nutzenmaximierung, individueller Präferenzen und Budgetrestriktionen sowie sonstigen Restriktionen spiegelt der realisierte Mietpreis die Zahlungsbereitschaft für die Untergrenze des Präferenzwertes wieder (vgl. Haase, 2011). Rosen (1974) griff in seiner Arbeit zu „Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition“ unter anderem die Idee von Lancaster (1966) weiter auf und entwickelte die erste mikroökonomisch fundierte Theorie der hedonischen Preise und formulierte wegweisend den hedonischen Ansatz für die Wirtschaftswissenschaften auf der hedonischen Hypothese (Haase, 2011, S.52): Heterogene Güter setzen sich aus einem Bündel an nutzenstiftenden Eigenschaften zusammen, welche messbar und bewertbar sind, Determinanten des Gesamtpreises darstellen, allerdings auf Grund ihres Wesens unteilbar und nur im Bündel handelbar sind.

Die hedonische Preistheorie geht somit davon aus, dass der Preis eines heterogenen Gutes durch die einzelnen Preise seiner Eigenschaften und Charakteristika beschrieben werden kann. Mathematisch kann diese Überlegung durch einen Vektor von k Attributen x abgebildet werden (vgl. Rosen, 1974):

$$x = (x_1, x_2, x_3, ..., x_k) \tag{2.1}$$

Der Preis (Mietertrag) dieses Gutes p ist eine Funktion der Produktattribute.

$$p(x) = p(x_1, x_2, x_3, ..., x_k) \tag{2.2}$$

So könnten etwa das Attribut x_1 für die Größe der Immobilie, das Attribut x_2 für das Alter des Gebäudes oder das Attribut x_3 für ein Maß der Erreichbarkeit stehen. In der englischsprachigen Literatur werden die Attribute oftmals gruppiert. In Folge dieser Abstraktion wird eine Unterteilung in R=“rent“ (Mietertrag), S=“structural characteristics“ (bauliche Charakteristika), N=“neighborhood characteristics “ (Nachbarschaftscharakteristika), L=“location within the market“ (Lagecharakteristika), C=“contract conditions or characteristics“ (vertragliche Charakteristika) und T=“time the rent is observed“ (zeitliche Periode in der die Daten erhoben wurden) vorgenommen (Malpezzi, 2002, S.2), wobei sich der funktionale Zusammenhang als $R = f(S, N, L, C, T)$ darstellt und der Logik in Formel (2.2) folgt. Eine genauere Auseinandersetzung mit Form des Eigenschaftsvektors und den Deter-

minanten des Preises ist in Abschnitt 2.3 dieser Arbeit zu finden. Durch partielles Ableiten der Preisfunktion (2.2) nach deren Eigenschaften x_i errechnen sich die impliziten oder hedonischen Preise für die Attribute der Liegenschaft:

$$p_{x_i}(x_i; x_{-i}) = \frac{\partial p(x)}{\partial x_i}; (i{=}1 \text{ bis } k) \tag{2.3}$$

Die implizite Preisfunktion gibt an, um wie viel der Nachfrager mehr zahlen müsste, um von einem Merkmal x_i eine marginal höhere Ausprägung zu erhalten, wobei alle anderen Merkmale konstant gehalten werden (vgl. Wieser, 2006). In seiner einfachsten Form lässt sich die hedonische Preisfunktion in eine multiple (lineare) Regressionsgleichung überführen. Die Spezifikation der Funktion ist eine der größten Herausforderungen bei der Schätzung von impliziten Preisen. Eine genaue Auseinandersetzung mit den Möglichkeiten der Ableitung der Regressionsgleichung findet in Abschnitt 3 zur Methodik statt. Der große Vorteil der hedonischen Bodenpreisanalyse ist, dass das Verfahren auf effektiven Marktdaten passiert und somit die Bewertung aller Markteilnehmer wiederspiegelt (vgl. ZKB, 2008). Daraus erwächst allerdings die Tatsache, dass die unter speziellen impliziten Marktbedingungen zustande gekommenen hedonischen Preise nicht auf andere Märkte übertragbar sind. Day (2003, S.4) führt dazu ein gutes Beispiel an: *„[. . .], the price paid for waterfront properties in London and Stockholm. Whilst in both cities such properties command considerable premia, the relatively low availability of "Thames – side" properties in London means that they command highly inflated prices compared to those in Stockholm, a city built upon a series of islands."*

Die Anwendung der hedonischen Theorie verlangt nach einigen Modellannahmen. So spielt die klassische Lehrbuchökonomie und die stark mechanistische Sichtweise des Marktes eine wichtige Rolle. Nach der ökonomischen Theorie werden am Markt Angebot und Nachfrage zusammengeführt. Dort wo sich beide treffen, entsteht ein Gleichgewichtspreis der den Markt räumt (vgl. Bannock et al., 2003). Marktteilnehmer handeln rational und nutzenmaximierend, was impliziert, dass Anbieter versuchen werden ihre Gewinne zu maximieren, während Nachfrager versuchen werden ihren Nutzen zu maximieren. Es gibt keine externen Effekte, weder durch die Nutzung noch durch die Produktion eines Gutes. Es gibt auf Produzenten, wie auch auf Konsumentenseite keine Präferenzen bezüglich Anbieter bzw. Nachfrager. Die Eigenschaften des am Markt gehandelten Gutes sind bekannt (vollkommene Transparenz und Information). In der Realität sind solche Modellmärkte nicht anzutreffen. Trotz der oben beschriebenen Restriktionen hat sich die hedonische Methode in unterschiedlichen Bereichen als Standard-Methode durchgesetzt. In der Immobilienwirtschaft kommen hedonische Modelle in der Erstellung von Immobilienpreisindizes sowie im Immobilieninvestment und der Immobilienbewertung (vor allem im angloamerikanischen Raum aber auch in der Schweiz und in

Deutschland) zum Einsatz (vgl. Peter, 2013). In Österreich ist der Ansatz weniger stark verbreitet, was an den hohen Anforderungen an die Datenbasis liegen dürfte. Traditionelle Verfahren zur Bewertung von Immobilien zielen i.d.R. darauf ab, einen Verkehrswert zu schätzen. Die hedonische Methode ist allerdings darüber hinaus fähig, den Einfluss wertbestimmender Attribute zu bestimmen (vgl. Maier, 2011) und somit ein tieferes Verständnis über den Prozess, welcher diese beobachteten Merkmalsausprägungen generiert, zu erzeugen. Im folgenden Abschnitt 2.3 werden nun mögliche preisbestimmende Attribute anhand von bereits durchgeführten Studien diskutiert und Hypothesen über den Einfluss dieser Attribute gebildet.

2.3 Determinanten des Preises

Es stellt sich vor der eigentlichen Modellbildung die Frage, welche Variablen den Preis (oder allgemeiner die Zielvariable) von Büroimmobilien determinieren. Caduff (2013, S.4ff) bringt in seiner Arbeit zum Commercial Real Estate Markt in Singapore einen guten Überblick in Hinblick auf durchgeführte Studien zur hedonischen Bewertung von Büroimmobilien. Dieser Überblick soll auch dieser Arbeit als Stütze dienen, wobei die Erkenntnisse durch neuere Studien ergänzt werden. In einem ersten Schritt werden sehr allgemein die Preisdeterminanten, welche in eine hedonische Regressionsgleichung eingehen können dargestellt. Die darauf folgende Literaturstudie setzt den Schwerpunkt auf die Untersuchung der ins Modell eingegangenen exogenen Variablen und deren Signifikanzniveau. Aus der Literatur werden Hypothesen für die Untersuchungen des Wiener Büroimmobilienmarktes abgeleitet und Erkenntnisse für die Variablenselektion im Modell bezogen.

Literaturüberlick zum Büromarkt

Die wertbestimmenden Charakteristika in hedonischen Preismodellen werden, wie einführend bereits erwähnt, oft gruppiert. Malpezzi (2002, S.2) unterscheidet S= “structural characteristics“ (bauliche Charakteristika), N= “neighborhood characteristics“ (Nachbarschaftscharakteristika), L= “location within the market“ (Lagecharakteristika), C= “contract conditions or characteristics“ (vertragliche Charakteristika) und T= “time the rent is observed“ (zeitliche Periode in der die Daten erhoben wurden). Eine andere Kategorisierung der Charakteristika einer Immobilie bringt die Züricher Kantonalbank in ihrer Studie zum Bodenmarkt in der Schweiz. Hier wird zwischen Charakteristika der Liegenschaft, Merkmale der Mikrolage (Lärmbelastung, Distanz zur Hauptstraße, Hanglage etc.) und Merkmale der Makrolage (Erreichbarkeit, Steuerkraft etc.) unterschieden (vgl. ZKB, 2008). Auch Bökemann und Feilmayr (1994) nehmen eine ähnliche Unterscheidung vor: Sie unterscheiden naturräumliche Eigenschaften, infrastruktubestimmte Lageeigenschaften,

sozialbestimmte Eigenschaften und ökologisch bestimmte Eigenschaften. Diese Gruppierungen kommen vor allem im Wohnimmobilienbereich zum Einsatz und können in leicht adaptierter Form auch auf das Bürosegment angewandt werden. Sie dienen der Literaturstudie als Analyserahmen für die Einordnung von möglichen Preisdeterminanten von Büroimmobilien. Tabelle 2.1 stellt die im Zuge der Arbeit analysierten Studien im Überblick dar. Die Angaben zum $(adj.)R^2$ beziehen sich i.d.R. auf das Basismodell und sollen nur eine grobe Einschätzung der Anpassungsgüte der getesteten Modelle geben.

Tabelle 2.1: Überblick - Studien zum Büroimmobilienmarkt

Autor (Jahr)	Region	Modelle	*Nobs*	$(adj.)R^2$	Zielvariable
Özyurt (2014)	Niederlande	QR/SAR	10000	0.88-0.98	$\ln(P/m^2)$
Caduff (2013)	Singapur	OLS/SAR	1271-7333	0.56-0.98	$\ln(P/m^2)$
Chegut et al. (2013)	London	OLS/SAR	1284	0.54-0.60	ln(P)
Chegut et al. (2013)	Frankfurt	OLS/SAR	242	0.73-0.76	ln(P)
Chegut et al. (2013)	Paris	OLS/SAR	1025	0.40-0.41	ln(P)
Haase (2011)	Zürich	OLS/MLM	105-1232	0.61-0.76	ln(R)
Farooq et al. (2010)	Toronto	OLS/Box-Cox/MLM	1302	0.45-0.47	P/m^2
Chalermpong und Wattana (2009)	Bangkok	OLS/SAR	85	0.53-0.59	P
Jennen und Brounen (2009)	Amsterdam	OLS	1465	0.53-0.56	$\ln(P/m^2)$
Nappi-Choulet und Tristan-Pierre (2009)	Paris	OLS/WLS/STAR	538-2587	0.88-0.94	ln(P)
Fuerst (2007)	New York	OLS/PDM	1073	0.28-0.49	$\ln(P/m^2)$
Kim und Zhang (2005)	Seoul	SAR	731	0.38-0.83	$P/m^2(ap)$
Ustaoglu (2003)	Ankara	OLS	244	0.28-0.40	$\ln(P/m^2)$

Note: $(adj.)R^2$ bezieht sich i.d.R. auf Basismodell;
OLS: Ordinary Least Squares;
GWR: Geographically Weighted Regression;
SAR: Spatial Autoregressive Model;
Kr: Kriging;
WLS: Weighted Least Squares;
QR: Quantil Regression;
MLM: Multilevel Modelling;
PDM: Paneldatenmodell;
STAR: Spatiotemporal Autoregressive Model;
$\ln(P/m^2)$: logarithmierte (Miet)Preis pro m^2
ln(R): logarithmierter Mietertrag pro m^2 und Jahr;
$P(/m^2)$: Mietpreis (pro m^2)
$P/m^2(ap)$: geschätzter Preis

Özyurt (2014) untersucht in seinem Paper den Commercial Real Estate Markt in den Niederlanden und erstellt alternative transaktionsbasierte Indizes auf Quartalsebene für den Zeitraum 2001 bis 2011. Als Datengrundlage für die Analyse wird die Investment Property Databank (IPD), welche für die Niederlande etwa 10.000 Beobachtungen beinhaltet, herangezogen. Die IPD beinhaltet vor allem sehr ausführliche Informationen zu Cash-Flows, Charakteristika der Immobilien sind dagegen nur in geringer Zahl vorhanden, was die Anwendung eines traditionellen hedonischen Preismodells nicht möglich macht. Das Basismodell sowie das Modell mit exogenen räumlichen Lag-Variablen verwenden den logarithmierten Preis in Euro als Zielvariable. Die Spezifikation der Modelle wird in dieser Studie stark von der Datenbasis bestimmt. Im Basismodell dient der gutachterliche Kapitalwert (*lnVal*) der Vorperioden als erklärende Variable. Zusätzlich wird eine Dummy Variable für das jeweilige Segment (Office, Retail, Residential und Industrial) mit ins Modell aufgenommen. Die Beobachtungen werden somit über alle Segmente „gepoolt“ und für jedes Quartal einzeln geschätzt. *lnVal* ist im Modell hoch signifikant, mit einem *adjusted* R^2 von über 0.9. Die Dummies für die unterschiedlichen Segmente sind in 27 der 44 Quartale auf einem 5 Prozent Signifikanzniveau signifikant. Im Spatial Autoregressive Model (SAR) mit exogener Lag-Variable konnte die Erklärungsgüte des Modells verbessert werden. Probleme sieht Özyurt (2014) hinsichtlich der Datenbasis und deren Repräsentativität, da die IPD ausschließlich Objekte von institutionellen Investoren (Versicherungsunternehmen, Pensionsfonds, Real Estate Investment Trusts etc.) beinhaltet und somit nur einen bestimmten Teil des Commercial Real Estate Marktes in den Niederlanden abdeckt. Institutionelle Investoren werden in ihren Entscheidungen über den Kauf und Verkauf von Portfolio Anteilen allerdings mit hoher Wahrscheinlichkeit anders reagieren als private Eigentümer. Es wird in Zukunft wichtig sein, dieses Daten-Gap zu schließen, um repräsentative Aussagen über die Preisentwicklung am Commercial Real Estate Markt machen zu können.

Das Future City Laboratory (FCL) der ETH Zürich in Singapur forscht unter anderem an den Preisdeterminanten von Immobilien. Caduff (2013) untersucht in seiner Arbeit den Commercial Real Estate Markt im Stadtstaat. Für die Untersuchungen wurden Daten von mehreren Immobilienplattformen ausgelesen und strukturiert gesammelt. Aus diesem Grund steht, je nach Segment (Industrial, Office und Retail), ein sehr umfangreicher Datensatz mit Angebots- und Transaktionsdaten zur Verfügung. Neben einer sehr großen Anzahl an baulichen Charakteristika werden auch Angaben zur vertraglichen Situation gemacht (Art und Dauer des Mietverhältnisses und Datum des Abschlusses). Darüber hinaus wurden eine Vielzahl von Nachbarschaftscharakteristika und Lagecharakteristika im Modell aufgenommen und getestet. Neben Clusterindikatoren liegt der Schwerpunkt der Analyse auf verschiedenen Erreichbarkeitsparametern. Hier kommt für den öffentlichen Verkehr (ÖV) unter anderem ein Erreichbarkeitsindex zum Einsatz, welcher mit der *agent based*

transport simulation Software *MATSim* erstellt wurde. Dieser Index gibt für jede Immobilie die Anzahl an Erreichbaren Aktivitäten durch das ÖV Netz wieder. Für das Office Segment wurde der logarithmierte Preis pro m^2 (Kaufpreismodell) und für die Angebotsdaten zusätzlich der logarithmierte Mietpreis pro m^2 (Mietpreismodell) als Zielvariable herangezogen. Die wichtigsten exogenen Variablen in den Kaufpreismodellen sind die Größe der Immobilie, Distanz zur nächsten Autobahnauffahrt, Erreichbarkeit des ÖV, Art des Mietverhältnisses, die Zugehörigkeit zu einem Office Cluster und in wenigen Modellen auch der Anteil an Akademikern im Umfeld. Das Mietpreismodell zeigt allgemein eine schlechtere Performance. Signifikant sind auch hier Größe der Immobilie, Distanz zur nächsten Autobahnauffahrt, Erreichbarkeit des ÖV und nur mehr teilweise die Cluster Indikatoren. Durch die Einführung von räumlichen Modellen konnte der Goodness-of-Fit in fast allen Modellen verbessert werden.

Das Paper von Chegut et al. (2013) untersucht die Auswirkungen von räumlichen Effekten auf die Bewertung von Commercial Real Estate in drei der größten Märkte in Europa: London, Paris und Frankfurt. Als Datengrundlage dienen Transaktionsdaten der Real Capital Analytics (RCA). Als exogene Variablen wird nur die Größe der Immobilie herangezogen. Darüber hinaus werden Dummies für die einzelnen räumlichen Sub-Märkte (Clusterindikatoren) und zeitliche Dummies mit ins Modell aufgenommen, um einen jährlichen Index zu erstellen. Für jeden Markt wurden drei verschiedene Spezifikationen für die räumliche Gewichtungsmatrix erstellt und verglichen, um die Sensitivität der Ergebnisse zu prüfen. Im Gegensatz zu anderen Studien wird hier nur eine geringe räumliche Autokorrelation festgestellt und die räumlichen Modelle zeigen kaum einen verbesserten Goodness-of-Fit gegenüber dem OLS Basismodell.

Haase (2011) gibt in seiner Arbeit zum Büroimmobilienmarkt in Zürich zu Beginn einen Überblick zu bereits durchgeführten Studien und geht dabei vor allem auf die verwendeten exogenen Variablen in den Modellen ein. Sehr allgemein hält er fest, dass vor allem Alter und Größe der Büroimmobilie sowie deren Standort als dominierende Variablen in den hedonischen Modellen zu sehen sind. Aber auch das Mietverhältnis und hier im besonderen die Vertragslänge können einen signifikanten Einfluss auf den Mietvertrag haben. Aus diesem Grund sieht Haase (2011) die Verwendung von realisierten Vertragsmieten als Zielvariable im hedonischen Modell als wünschenswert, da diese die wahre Marktmiete repräsentieren. Die Eignung von Angebotsmieten für Analysen schließt er hinsichtlich der leichteren Datenverfügbarkeit allerdings nicht aus. Getestet werden in dieser Arbeit in einem ersten Schritt drei klassische hedonische Preismodelle mit dem logarithmierten Mietertrag pro m^2 und Jahr als Zielvariable. Als exogene Variablen wurden eine Vielzahl von Parameter getestet. Darunter befinden sich 17 Variablen, die die Standortqualität

der Immobilie (Charakteristika des Grundstückes, sozio-ökonomisches Umfeld und Erreichbarkeiten) beschreiben sollen. Als signifikante Größen wurde die Mikrolage, unterschiedliche Erreichbarkeitsvariablen, die Mietfläche, der Zustand des Gebäudes und die Bauperiode (Gebäudealter) identifiziert. Darüber hinaus zeigten Variablen der Mietvertragsqualität einen signifikanten Einfluss auf den Mietertrag im Modell. Gegen die Erwartungen des Autors war die Leerstandsrate des Gebäudes und deren Einfluss auf den Mietvertrag in keinem der drei Modelle signifikant. In einem zweiten Schritt werden hedonische Mehrebenenmodelle (*multilevel modelling approach*) für die Analyse herangezogen, wodurch der Model-Fit verbessert werden konnte.

Farooq et al. (2010) analysiert den Büroimmobilienmarkt in Toronto (Greater Toronto Area). Besonderes Augenmerk wird auf die Untersuchung von räumlicher Heterogenität und Clusterung der Angebotsmieten von Büroimmobilien gelegt. Die Verwendung von Angebotsmieten wird den Transaktionsmieten vorgezogen, da Angebotsmieten die vom Anbieter, unter Berücksichtigung der Qualität von Standort und Immobilie sowie den herrschenden Marktbedingungen, erwartete Miete wiederspiegelt. Der Autor widerspricht somit in diesem Punkt anderen Studien (vgl. Haase, 2011). Der Zusammenhang zwischen Angebotsmieten pro m^2 und den erklärenden Variablen wird in einem ersten Schritt als lineare Funktion modelliert. Darüber hinaus werden Box-Cox Transformationen sowie hierarchische und additive Modelle getestet. Signifikanten Einfluss auf die Angebotsmieten hatten die Qualität der Gebäude, die Leerstandsrate des Gebäudes, der Anteil der Landnutzungsklassen in der Nachbarschaft, die Erreichbarkeit im ÖV sowie die Nähe zum CBD.

Die Auswirkungen von Verkehrsinfrastrukturinvestitionen auf den Immobilienmarkt sind Analysegegenstand vieler Publikationen. Chalermpong und Wattana (2009) untersucht die Erreichbarkeit von Stationen des schienengebundenen, öffentlichen Verkehrs durch Bürostandorte in Bangkok und deren Einfluss auf die monatlichen Mieten. Die Daten wurden durch Befragungen erhoben, was zu einem Datensatz von 85 Bürogebäuden führte. Im OLS geschätzten Modell mit einem R^2 von 0.556 waren folgende Variablen signifikant: Anzahl Lifte, Dummy für Air Condition, Distanz zu Stationen des schienengebundenen ÖV und Dummy für Bürocluster. Die angewandten räumlichen Modelle zeigen ein ähnliches Verhalten, bei leicht verbesserter Anpassungsgüte.

Die Effekte von Office Cluster auf die Angebotsmiete pro m^2 in der niederländischen Hauptstadt Amsterdam untersucht Jennen und Brounen (2009) in deren 2009 durchgeführten Studie. Es kommen klassische hedonische Modelle mit Log-Linearer Spezifikation zum Einsatz. Es konnte ein hoher signifikanter Einfluss von hoher Bürodichte auf die Angebotsmiete nachgewiesen werden. Neben den räumlichen Clusterindikatoren wurde im Modell auf Gebäudeal-

ter, Gebäudezustand, Sanierung des Gebäudes sowie Distanz zur Autobahn und zu Stationen des öffentlichen Verkehrs kontrolliert. Es zeigte sich, dass die Distanz zur Autobahn einen positiven Zusammenhang mit dem Mietpreis aufweisst. Die Untersuchung zeigte allerdings für den öffentlichen Verkehr den umgekehrten Zusammenhang. Je geringer die Distanz zu Stationen des öffentlichen Verkehrs, desto höher ist der Mietpreis.

Eine Untersuchung der Transaktionspreise für Büroimmobilien in Paris ist in Nappi-Choulet und Tristan-Pierre (2009) zu finden. Der räumliche Fokus wird auf die beiden Business Districts (Central Business District und La Defense District) gelegt. Getestet werden klassische hedonische Modelle, und Modelle die auf die Raum-Zeit Abhängigkeiten kontrollieren (STAR Modelle). Die Variablen Bauperiode und Zustand des Gebäudes zeigen in fast allen Modellen einen signifikanten Einfluss auf den Preis.

Fuerst (2007) analysiert die Preisdeterminanten des New Yorker Büroimmobilienmarktes und stellt sich die Frage, inwieweit diese im zeitlichen Kontext und abhängig von (räumlichen) Submärkten schwanken können. Diese Volatilität von Immobilienmärkten stellt einen Risikofaktor für Investoren dar. Als Datengrundlage dient die *Manhattan Office Property Database*, welche Immobiliendaten auf Gebäudeebene beinhaltet. Es werden zwei unterschiedliche OLS geschätzte hedonische Modelle getestet, wobei die Spezifikation als loglinear Modell die beste Performance zeigt. Die Schätzungen werden für die beiden Modelle für die Jahre 1999 bis 2004 quartalsweise durchgeführt, was teilweise zu niedrigen Anpassungsgüten mit einem R^2 von 0.284 im ersten Quartal 2001 führt. Signifikant waren folgende Variablen: Leerstandsrate des Gebäudes, die Größe, das Gebäudealter, die Anzahl der Stockwerke, die Anzahl an Serviceeinrichtungen im Gebäude, die Anzahl an Bürogebäude in der Nachbarschaft und die Nähe zu Bahnstationen. In einem zweiten Schritt wurden Paneldatenmodelle angewandt, um zusätzliche Informationen über den Zyklus der Teilmärkte zu generieren.

In der Studie von Kim und Zhang (2005) wird der geschätzte Bodenpreis für gewerbliche Nutzungswidmung in Seoul untersucht. Von speziellem Interesse sind die Auswirkungen von der Entfernung zur nächsten Bahnstation auf den Preis. Es wurden acht Modelle geschätzt, wobei die erklärenden Variablen in allen Modellen auf Distanzparameter zu ÖV-Stationen und auf Dummy Variablen für die einzelnen Submärkte beschränkt sind. Darüber hinaus wird ein Index für die Dichte an Finanzinstituten im Bezirk in alle Modelle mit aufgenommen. Auf Grund der geringen Anzahl an Transaktionsdaten im Bereich der gewerblich genutzten Grundstücke werden geschätzte Bodenpreise für die Studie herangezogen, was ein gewisses Risiko mit sich bringt, da diese möglicherweise nicht den Marktpreis abbilden. Durch die Verwendung von räumlichen SAR Modellen kann auch in dieser Studie die Modellgüte in allen

Modellen erhöht werden.

In der Arbeit von Ustaoglu (2003) werden die Büromieten in Ankara analysiert. Ein Datensatz mit 244 Transaktionen stand für die Untersuchungen zur Verfügung. Neben der Miete pro Monat sind die Größe des Gebäudes, die Größe der Immobilie/Mieteinheit und das Gebäudealter vorhanden. Dieser Datensatz wurde im Zuge einer Befragung der Vertragsparteien zusätzlich durch weitere Variablen angereichert. Beide Modelle in der Arbeit wurden mit OLS geschätzt. Modell 1 nimmt neben baulichen Charakteristika auch Lagecharakteristika und vertragliche Charakteristika mit ins Modell auf. Die Lagecharakteristika beschränken sich auf Dummy Variablen für die 5 Submärkte. Modell 2 nimmt nur die baulichen und vertraglichen Charakteristika ins Modell auf. Signifikanten Einfluss hatten folgende Variablen: Cluster- bzw. Submarkt Dummy, die vertragliche Bindung, das Geschoss in dem die Mieteinheit situiert ist, Vorhandensein eines Sicherheitssystems und die Gebäudegröße.

Aufbauend auf der Literaturrecherche werden in Tabelle 2.2 die möglichen abhängigen Variablen mit der erwarteten Richtung des Zusammenhanges und erwarteter Stärke des Einflusses auf den Preis von Büroimmobilien dargestellt.

Hypothesenbildung für den Wiener Büroimmobilienmarkt

Die im Zuge der Literaturrecherche in Abschnitt 2.3 untersuchten Arbeiten zeigen eine große Streuung an verwendeter Methoden, allgemein kann aber ein gewisser Vorteil bei der Verwendung von räumlichen Modellen festgestellt werden. In allen Fällen konnte der Goodness-of-Fit verbessert werden. Für die Untersuchungen des Wiener Büroimmobilienmarkt und den Einfluss der verwendeten Variablen werden folgende Hypothesen aufgestellt:

- Das Gebäudealter und der Mietpreis weisen einen negativen Zusammenhang auf. Dies ist vor allem durch die hohe Sensibilität gegenüber Betriebskosten und durch den Anspruch an moderne Raumkonzepte zu begründen.

- Mit der Größe der Immobilie, die zur Vermietung kommt, sinkt der Mietpreis pro m^2. Hier ist von einem einem Art Skalleneffekt auszugehen, wobei größere Flächen proportional einen geringeren Quadratmeterpreis verzeichnen als kleinere Flächen.

- Das Vorhandensein einer Förderanlage wirkt sich positiv auf den Mietpreis aus.

Tabelle 2.2: Determinanten des Preises von Büroimmobilien - Erklärende Variablen mit erwarteten Vorzeichen des Zusammenhanges und Stärke des Einflusses

Exogene Variable	erw. Zusammenhang
Bauliche Charakteristika	
Gebäudealter	− − −
Größe der Immobilie	+
(Miet)Fläche	++
Zustand	++
Leerstandsrate Gebäude	−
Vorhandensein eines Liftes	+
Parkplatzangebot	+
Zustandsklasse	+/−
Nachbarschaftscharakteristika	
Nähe/Zugehörigkeit Büro-Cluster	+ + +
Akademikeranteil	+
Anteil an Wohnnutzung in der Nachbarschaft	−
Anteil an gewerblicher Nutzung in der Nachbarschaft	+
Lagecharakteristika	
Distanz zur nächsten Bahnstation	− − −
Distanz zur nächsten Bus/Tramstation	−−
Distanz zur nächsten Autobahnauffahrt	++
Distanz zum CBD	− − −
Erreichbarkeitsindikator* ÖV	++
Erreichbarkeitsindikator* IV	++
Vertragliche Charakteristika	
Mietdauer	+
Art des Mietvertrags	+/−
Angebots-/Abschlussjahr	+ + / − −

hoch signifikanter negativer/positiver Zusammenhang: − − −/ + ++
signifikanter negativer/positiver Zusammenhang: − − / + +
geringer bis nicht signifikanter negativer/positiver Zusammenhang: −/+
*Def. Erreichbarkeit lt. Axhausen (2013, S.30)

- Der Zustand der Immobilie steht in einem positiven Zusammenhang mit dem Mietpreis. Je höher die Zustandsklasse, je besser der Zustand und desto höher ist der Mietpreis.

- Die Zugehörigkeit zu einem Bürocluster wird durch einen Zuschlag im Mietpreis honoriert.

- Der Zusammenhang zwischen der Distanz zum nächsten Bürocluster und dem Mietpreis ist negativ.

- Der Zusammenhang zwischen der Distanz zur nächsten Haltestelle des öffentlichen Verkehrs und dem Mietpreis ist negativ.

- Der Zusammenhang zwischen der Distanz zum ersten Bezirk und dem Mietpreis ist negativ.

2.4 Räumliche Effekte

Unsere Umwelt produziert täglich Phänomene wie Regenwolken, Aktienkurse, Todesfälle, Hitzewellen oder eben Miet- und Immobilienpreise, welche, so möchte man glauben, durch einen unbekannten datengenerierenden Prozess erzeugt werden, der tieferen Gesetzmäßigkeiten folgt (vgl. Stocker, 2013). Auf Grund der Komplexität der Mechanismen und deren Kausalitäten gibt es für uns keine Möglichkeiten diese Gesetzmäßigkeiten direkt zu beobachten, nur deren Folgen stehen uns als Daten zur Verfügung. Die Welt kann somit stark vereinfacht als ein „System„ gesehen werden, eine imaginäre Maschine, welche wir in ein theoretisches, mathematisches Modell gießen, um Ursachen und Wirkungen, welche in Realität längst bestehen, greifbar zu machen, mit dem Ziel aus den beobachteten Daten Rückschlüsse auf den unbekannten Prozess zu ziehen (vgl. Smith und Recktenwald, 1999; Stocker, 2013). In der Ökonometrie werden zur Messung von Zusammenhängen sehr oft regressionsanalytische Methoden verwendet.
Haben die beobachteten Daten einen räumlichen Bezug, sind sie mit einer geographische Position unmittelbar verbunden, so spricht man von räumlichen Daten *(„spatial data“)*. Diese besitzen eine Reihe von speziellen Eigenschaften, was die Komplexität der theoretischen Modelle erhöht und einige methodische Probleme mit sich bringt. Haining (2009, S.4) schreibt diesbezüglich: *„Data is the lynch pin in the process of “doing science” and it is essential that methodologies for spatial data analysis are tuned to the properties of spatial data“*. In den Regionalwissenschaften, wie auch in anderen Sozialwissenschaften, sieht man sich mit dem Problem konfrontiert, dass Modelle niemals den gesamten Prozess abbilden können und somit immer ein (oft sehr großer) unerklärter Teil verbleibt (vgl. Fotheringham et al., 2002).
Beispielsweise können eine fehlende Übereinstimmung zwischen dem zu untersuchenden räumlichen Phänomen und der gewählten räumlichen Beobachtungseinheit (vgl. Haining, 2009), die Notwendigkeit unterschiedliche Daten aus unterschiedlichen Quellen mit unterschiedlichen räumlichen Aggregationsniveaus zu kombinieren (vgl. Anselin, 2001), Spillover-Effekte wie auch Messfehler zu räumlichen Effekten *(„spatial effects“)* führen. Unter diesem Terminus werden die Aspekte der räumlichen Heterogenität und der räumlichen Abhängigkeit resp. Autokorrelation zusammengefasst (vgl. Anselin, 1988b; Löchl, 2010), welchen in den nächsten Abschnitten besondere Aufmerksamkeit gewidmet wird.

Räumliche Heterogenität

Räumliche Heterogenität beschreibt die Tatsache, dass Merkmalsausprägungen über den geografischen Raum variieren können. Beispielsweise würde man annehmen, dass die jährliche Niederschlagsmenge in den einzelnen Ländern Europas teilweise stark von einander abweichen. Ein Wert der den durchschnittlichen Niederschlag in Europa repräsentiert, hätte somit auf Grund der länderspezifischen klimatischen Unterschiede eine geringe Aussagekraft und würde für die einzelnen Regionen im speziellen keine nützliche Information beinhalten. FOTHERINGHAM et al. (2002) beschreiben diese Statistiken als „Globale Statistiken“ und zeigen auf, dass diese nur im Falle einer geringen Variation der lokalen Beobachtungen als repräsentativ betrachtet werden können. Je größer die Unterschiede auf lokaler Ebene sind, desto geringer wird die Aussagekraft der globalen Statistik.
Überträgt man diese Ansätze auf den Wiener Immobilienmarkt und stellt sie im Kontext zu den in Abschnitt 2.2 gemachten Überlegungen zur hedonischen Theorie dar, so ergibt sich folgendes Bild: In der Regressionsgleichung der Form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...\beta_n x_n + \varepsilon$, unter Verwendung des kleinsten Quadrate Schätzers (OLS), wobei y die abhängigen Variable Immobilienpreis, β_{1-n} die Regressionskoeffizienten der unabhängigen Variablen x_{1-n} (Größe der Immobilie, Ausstattung, Zustand, Alter, Anteil an Akademikern, Bevölkerungsdichte etc.) und ε den Störterm darstellt, repräsentiert β_1 den durchschnittlichen Zusammenhang zwischen Immobilienpreis und beispielsweise der Bevölkerungsdichte (ceteris paribus) für das gesamte Untersuchungsgebiet. Dieser globale Wert des Regressionskoeffizenten bildet aber möglicherweise lokale Unterschiede im Zusammenhang zwischen Immobilienpreis und Bevölkerungsdichte nicht ausreichend ab (vgl. FOTHERINGHAM et al., 2002). Es wäre rein hypothetisch denkbar, dass beispielsweise der Zusammenhang von Bevölkerungsdichte und Büromiete in Wien, in der inneren Stadt stärker ausgeprägt ist als in der Umgebung der Donaucity (vgl. Simpson Paradoxon). Prozesse, welche in den Regionalwissenschaften von Interesse sind, weisen meist eine räumliche Instabilität auf und werden als „räumlich nicht-stationärer Prozesse“ bezeichnet (vgl. FOTHERINGHAM et al., 2002). Betrachtet man die Literatur zur Wirtschaftsgeografie und zu den Regionalwissenschaften so finden sich in den Theorien eine Vielzahl an Hinweisen darauf, dass Effekte und Beziehungen räumlich variieren. Hier können die Hierarchie der zentralen Orte von Christaller oder das Lagerentenmodell von v. Thünen als prominente Beispiele genannt werden. (vgl. ANSELIN, 1988b)

Räumliche Abhängigkeit und Autokorrelation

TOBLER (1970) formuliert das „First Law of Geography“: *„Everything is related to everything else, but near things are more related than distant things“*. Dieses „First Law of Geography“ impliziert, dass Ereignisse die geografisch nä-

her beieinander liegen, dazu tendieren ähnlicher zu sein, als solche die weiter von einander entfernt sind. Auf Grund der Lage und Distanz von Beobachtungen besteht ein funktionaler Zusammenhang zwischen den Ereignissen am Ort X und allen anderen Orten im Raum. Räumliche Beobachtungen sind somit nicht statistisch unabhängig und neigen dazu räumlich zu korrelieren (vgl. LeSage und Pace, 2009), was eine Herausforderung für die statistischen Modelle mit sich bringt, welche die Struktur der räumlichen Prozesse abbilden sollen. Haining (2009, S.6) merkt an, dass Toblers „First Law of Geography“ offensichtlich eine starke Vereinfachung darstellt und in Bezug auf manche räumlichen Prozesse unzutreffend ist, allerdings einen brauchbaren Aphorismus darstellt. Die Gründe, warum von räumlicher Autokorrelation im Modell ausgegangen werden kann sind vielfältig (vgl. Abschnitt 3.2), wobei oft aus der den Untersuchungen zu Grunde liegenden Theorie von einer *a priori* Autokorrelation ausgegangen wird, resp. diese nicht ausgeschlossen werden kann (vgl. Anselin, 1988b). Nimmt man wieder die oben gefundende hedonische Regressionsgleichung der Form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...\beta_n x_n + \varepsilon$, wobei y einen $n \times 1$ Vektor der abhängigen Variable Immobilienpreis, β_0 den Intercept Term, β_{1-n} die Regressionskoeffizienten der unabhängigen Variablen x_{1-n} (Größe der Immobilie, Ausstattung, Zustand, Alter, Anteil an Akademikern, Bevölkerungsdichte etc.) und ε den Störterm darstellt, so sind bei Vorhandensein von räumlicher Autokorrelation die *off-diagonal* Elemente der Varianz-Kovarianz Matrix zumindest teilweise von Null ($Cov(u_i, u_j) \neq 0$) verschieden und eine OLS Schätzung nicht erwartungstreu.

Räumliche Ökonometrie

„Spatial effects are the essential reason for the existence of a separate field of spatial econometrics“ (Anselin, 1988b, S.11). Die räumliche Ökonometrie beschäftigt sich somit mit Theorien aus der räumlichen Ökonomie und verwandten Disziplinen und entwickelt mathematisch-statistische Methoden um diesen Theorien den nötigen empirischen Gehalt zu geben. Räumliche Heterogenität und räumliche Abhängigkeit resp. Autokorrelation sind als theoretische Konzepte relativ einfach nachzuvollziehen und finden sich auch in vielen regionalwissenschaflichen Theorien wieder. Die Herausforderung liegt in der Berücksichtigung und Einbindung dieser Überlegungen bei der Modellierung. Werden diese räumlichen Effekte ignoriert, so kommt es aus statistischer Sicht zu ineffizienten oder nicht mehr erwartungstreuen Regressionskoeffizienten. Fotheringham et al. (2002, S.10) relativiert allerdings die Möglichkeiten der Einbindung von räumlichen Effekten und die Idee eines „perfekten Modells“ und schreibt dazu: *„We can never be completely confident that our models are correct specifications of reality because of our lack of theoretical understanding of the processes governing human spatial behaviour“*.

Kapitel 3

Methodik

In Kapitel 3 werden die mathematischen und statistischen Methoden erarbeitet, welche in weiterer Folge in der Datenanalyse zur Anwendung kommen. Abschnitt 3.1 setzt sich mit dem klassischen linearen Regressionsmodell auseinander und zeigt, dass der OLS-Schätzer *(„Ordinary Least Squares Estimator “)* unter bestimmten Annahmen erwartungstreu und varianzminimal ist. Der Punkt zur explorativen räumlichen Datenanalyse stellt Techniken vor, die räumliche Muster und Abhängigkeiten identifizieren können und dadurch die für die Erwartungstreue und Effizienz gebrauchten Annahmen des OLS-Schätzers deren Gültigkeit entziehen. Der letzte Abschnitt 3.2 beschäftigt sich mit räumlichen Regressionsmodellen und hier im speziellen mit *„Geographically Weighted Regression Models (GWR)“* nach Fotheringham et al. (2002) und *„Simultaneous Autoregressive Models (SAR)“* nach Anselin (1988b) resp. LeSage und Pace (2009).

Neben den getesteten räumlichen Regressionsmodellen stehen weitere Ansätze zur Verfügung. Ein erster Hinweis auf die unterschiedlichen methodischen Ausrichtungen in Hinblick auf die Modellwahl wird in Abschnitt 2.3 gegeben. Untersuchungen von Brunauer et al. (2013) zeigen beispielsweise einen neuartigen Ansatz zur Vorhersage von Hauspreisen in Österreich mit Hilfe eines *hybriden hierarchischen STAR Modells („Structured Additive Regression Model“).*

3.1 Lineare Regressionsmodelle

Im Abschnitt 2.2 wurde bereits darauf hingewiesen, dass die hedonische Preisfunktion in ihrer einfachsten Form als multiple, lineare Regressionsgleichung folgender Form geschrieben werden kann:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...\beta_n x_n + \varepsilon \tag{3.1}$$

oder in Matrixnotation

$$\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon \tag{3.2}$$

Die Parameter der Grundgesamtheit sind natürlich nicht beobachtbar. Aus diesem Grund wird das oben geschriebene Gleichungssystem als *„Population Regression Function (PRF)“* bezeichnet. Bei der Anwendung auf Stichprobendaten erhält man eine *„Sample Regression Function (SRF)“* der Form:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\beta} + \hat{\varepsilon} \tag{3.3}$$

Die abhängige Variable $\hat{y}$ ist ein nx1 Vektor mit den Preisen der Immobilien, welcher durch seine Eigenschaften in Form der nxk Datenmatrix X beschrieben wird. $\hat{\beta}$ ist der kx1 Vektor der zu schätzenden Koeffizienten und $\hat{\varepsilon}$ stellt den nx1 Residuenvektor dar. Gesucht ist der Vektor $\hat{\beta}$, welcher möglichst nah an dem wahren Koeffizientenvektor der Grundgesamtheit β liegt und somit erwartungstreu ist und deren Stichprobenkennwertverteilung eine möglichst kleine Varianz aufweist. Die Grundidee der OLS-Methode liegt darin, den Vektor $\hat{\varepsilon}$ so zu wählen, dass die Summe der quadrierten Abweichungen in einer Stichprobe so klein wie möglich wird (vgl. Stocker, 2013), was zu folgendem Minimierungsproblem führt $\min_{\hat{\beta}}(\hat{\varepsilon}'\hat{\varepsilon})$. Als Ergebnis erhalten wir den OLS-Punktschätzer (vgl. Wooldridge, 2009).

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'y} \tag{3.4}$$

Dieser OLS-Schätzer ist unter den Gauss-Markov Annahmen BLUE *(„Best Linear Unbiased Estimator“)* (vgl. Stiassny, 2012), d.h. in der Klasse der linearen erwartungstreuen Schätzfunktionen hat der OLS-Schätzer die kleinste Varianz und ist somit effizient *(„Gauss-Markov Theorem“)*. Die Gauss-Markov Annahmen spielen somit eine fundamentale Rolle für die Effizienz des OLS-Schätzers und sollen hier kurz dargestellt werden, wobei eine Unterscheidung in Annahmen, die die Spezifikation der Funktion, die Matrix X der erklärenden Variablen und die Störterme ε betreffen, stattfinden kann (vgl. Wooldridge, 2009).

- Die wahre Beziehung zwischen den erklärenden Variablen x_i und der zu erklärenden Variable y_i ist linear in den Parametern β_i. Darüber hinaus ist das Modell korrekt spezifiziert, was einen „omitted variable bias“ausschließt.

- Die Zahl der Beobachtungen n ist größer als die zu schätzenden Parameter k ($n \geqq k$) und es besteht keine exakte lineare Abhängigkeit zwischen den x Variablen (keine perfekte Multikollinearität).

- Die erklärenden Variablen x_i sind stochastisch unabhängig von ε_i. Die Exogenitätsannahme $E(\varepsilon_i|X) = 0$ ist eine der zentralsten Annahmen und garantiert, dass der OLS-Schätzer erwartungstreu und konsistent ist.
- Die Störterme ε_i sind mit einem Erwartungswert von Null und einer Varianz von σ^2 unabhängig und identisch verteilt. Dies kann als $\varepsilon_i \sim i.i.d.(0, \sigma^2)$ geschrieben werden. Dies impliziert Homoskedastizität und keine Autokorrelation der Störterme. Man schreibt $var(\varepsilon_i) = \sigma^2 \boldsymbol{I}$.

Es wird anschließend noch gezeigt, dass speziell bei der Arbeit mit räumlichen Daten i.d.R. nie alle Annahmen erfüllt werden können.

Das Bestimmtheitsmaß R^2

Das Bestimmtheitsmaß R^2 (oft auch Bestimmtheitskoeffizient) ist ein Maß für die Anpassungsgüte des Modells an die Datenbasis. Der *„Fit“* des Modells ist definiert als Anteil der durch die abhängigen Variablen x_i erklärte Variation der unabhängigen Variable y (vgl. Huib, 2011).

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} \tag{3.5}$$

Es ist erkennbar, dass es sich beim R^2 im wesentlichen um eine Streuungszerlegung handelt, wobei die gesamte Streuung von y in einen erklärten und einen unerklärten Teil zerlegt wird. TSS steht für *„Total Sum Squared“* und ist als die Gesamtstreuung der y_i um deren Mittelwert anzusehen. *„Explained Sum Squared“* wird als ESS geschrieben und ist die Streuung der gefitteten $\hat{y}_i$ um den Mittelwert $\bar{y}$. Die Quadratsumme der Residuen ist als SSR definiert und bedeutet *„Sum of Squared Residuals“* (vgl. Stocker, 2013).

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{3.6}$$

$$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{3.7}$$

$$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} \tag{3.8}$$

Das R^2 ist ein Anteil und liegt aus diesem Grund zwischen 0 und 1. Je höher das R^2 ist, desto höher ist der *„Fit“* des Modells, wobei eine „Überinterpretation“der Werte vermieden werden sollte (vgl. Huib, 2011; Stocker, 2013). Messfehler in y oder eine Fehlspezifikation des Modells kann zu einem hohen R^2 führen, ohne aber die Güte der Anpassung widerzuspiegeln. Darüber hinaus eignet sich das Bestimmtheitsmaß nicht für den Vergleich von Modellen

mit unterschiedlich vielen erklärenden Variablen, da durch die Aufnahme von zusätzlichen x das R^2 unweigerlich erhöht wird. Hier sollte nur mit dem *adjusted* R^2 gearbeitet werden(vgl. HUIB, 2011). In diesem Fall wird um die Anzahl der Freiheitsgrade korrigiert.

3.2 Räumliche Regressionsmodelle

Geographically Weighted Regression Models (GWR)

Das in Abschnitt 3.1 vorgestellte multiple lineare Regressionsmodell der Form (3.3) ist als globales Modell in der Lage globale Parameter zu schätzen, die unter den Gauss-Markov Annahmen erwartungstreu und varianzminimal sind. Diese Parameter spiegeln einen durchschnittlichen Zusammenhang zwischen dem Regressand $\hat{y}$ und den Regressoren $\boldsymbol{X}$ wieder und können mögliche räumliche Unterschiede in den Parametern nicht wiedergeben. Allerdings sind Prozesse, welche wir in den Regionalwissenschaften untersuchen wollen, oftmals räumlich nicht konstant und werden hier als räumlich nicht-stationäre Prozesse bezeichnet. Auf dies wurde bereits in Abschnitt 2.4 hingewiesen. FOTHERINGHAM et al. (2002, S.9) führt als möglichen Grund für räumliche nicht-stationarität folgendes an: *„[...] possible cause of observed spatial non - stationarity in relationships is that, for whatever reasons, some relationships are intrinsically different across space. Perhaps, for example, there are spatial variations in people's attitudes or preferences or there are different administrative, political or other contextual issues that produce different responses to the same stimuli over space.“* Lokale Modelle wie *„Geographically Weighted Regression Models (GWR)“* sind in der Lage räumlich variierende Parameter zu schätzen, was im wesentlichen eine Erweiterung der linearen multiplen Regression um einen Lagebezug bedeutet.

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{3.9}$$

Der Lagebezug wird durch das Koordinatenpaar (u_i, v_i) in die Regressionsgleichung aufgenommen. $\beta_k(u_i, v_i)$ ist somit der k-te Schätzer an einem geografischen Punkt i, wodurch für jeden Punkt im Raum eine Schätzung möglich ist und einem räumlich nicht-stationären datengenerierenden Prozess Rechnung getragen wird. Es ergibt sich daher eine Situation, in der mehr zu schätzende Parameter k als Beobachtungen n vorhanden sind ($n < k$). Daraus folgt eine der wichtigsten Annahmen der GWR, welche besagt, dass die Parameter keiner Zufallsverteilung folgen, sondern eine deterministische Funktion der Lage im Raum sind (vgl. FOTHERINGHAM et al., 2002; LOADER, 1999), was wiederum dazu führt, dass lokale Parameter nicht unverzerrt geschätzt werden können. Ein Ausweg aus dieser Situation wird gefunden, in dem das Modell nicht auf das gesamte Datenset angewandt wird, sondern eine Kalibration auf ein Subset stattfindet. Für die Schätzung an einem beliebigen Punkt i werden

nur die Datenpunkte in dessen „näherer Umgebung“ mit einbezogen, wobei ein bestimmter Grad an räumlicher Kontinuität vorausgesetzt wird (um mit OLS schätzen zu können, vgl. dazu auch Linearitätsannahme). Der Kalibrierungsprozess des Modells kann als ein Trade-Off zwischen Bias (Verzerrung) und Varianz der Parameter gesehen werden. Je größer das lokale Subset ist, desto größer ist die Wahrscheinlichkeit Parameter mit einem hohen Grad an Verzerrung zu schätzen, desto kleiner ist allerdings die Varianz. Demgegenüber steht ein kleines lokales Subset, welches den Bias reduziert, allerdings zu einer hohen Varianz der Schätzer führt. Ein großer Vorteil der GWR ist darin zu sehen, dass sie sich auf das gut bekannte Framework der OLS-Methode stützt, womit die multiple lineare Regression der Form (3.3) ein Spezialfall von (3.9) ist. Der Parameterschätzer folgt in diesem Fall der Logik einer „weighted least square (WLS)“Schätzung(vgl. Fotheringham et al., 2002, S.67).

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\boldsymbol{X'W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X'W}(u_i, v_i)\boldsymbol{y} \tag{3.10}$$

$W(u_i, v_i)$ (verkürzt auch einfach $W(i)$) ist eine Gewichtungsmatrix welche im Grunde Toblers „First Law of Geography“ operationalisiert. Die Diogonalelemente der n x n Matrix $W(i)$ beinhalten die geografische Gewichtung für jede Beobachtung n für den Regressionspunkt i. Punkte die weiter entfernt von i liegen erhalten weniger Einfluss auf die Schätzung, als Punkte die nahe bei i liegen, wodurch auch dem oben beschriebenen Varianz Bias Effekt entgegengewirkt werden kann (vgl. Fotheringham et al., 2002, S.53). Auf die möglichen Ausformungen der Gewichtungsfunktion wird im darauf folgenden Abschnitt eingegangen.

GWR Gewichtungsmatrix und Gewichtungsfunktion

Aus dem oben Beschriebenen ist erkennbar, dass die Gewichtungsmatrix, welche den räumlich funktionalen Zusammenhang zwischen dem Regressionspunkt i und allen anderen Beobachtungspunkten j wieder gibt, einen entscheidenden Einfluss auf die Ergebnisse der GWR nimmt. Die Gewichtungsmatrix $W(i)$ hat, wie bereits angedeutet, folgende Form:

$$W(i) = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & w_{in} \end{pmatrix} \tag{3.11}$$

w_{in} ist das Gewicht des n-ten Datenpunktes bei der Kalibrierung des Modells für den i-ten Standort. Die Gewichtungsmatrix muss daher für jeden Punkt i neu berechnet werden, wobei die einzelnen Gewichte w_{in} die Nähe zum Regressionspunkt i wiedergeben und mit abnehmender Distanz zu i stärker in die Schätzung der Parameter miteinfliesen (vgl. Fotheringham et al., 2002,

S.54). Die Wahl der Gewichtungsfunktion (*kernel*) ist somit entscheidend für die Kalibrierung des Modells. FOTHERINGHAM et al. (2002, S.56) und FISCHER und GETIS (2010, S.463) zeigen mehrere Möglichkeiten für die Spezifizierung der Funktion auf, wobei eine „Bi-quadrat Funktion " der Form

$$w_{ij} = \begin{cases} (1 - \frac{d_{ij}^2}{b^2}) & \text{wenn } d_{ij} < b \\ 0 & \text{wenn } d_{ij} > b \end{cases} \tag{3.12}$$

bzw. eine „Gaussain Funktion " der Form

$$w_{ij} = \begin{cases} \exp\left(-\frac{1}{2}(\frac{d_{ij}}{b})^2 \right) & \text{wenn } d_{ij} < b \\ \text{virtuell } 0 & \text{andernfalls} \end{cases} \tag{3.13}$$

am besten das Problem der Diskontinuität in der Gewichtung abbilden können. In Formel 3.12 und 3.13 ist der Parameter b als Beobachtungsbandbreite (*bandwidth*) definiert, welcher die maximale Ausdehnung der Gewichtungsfunktion bestimmt. Abbildung 3.1 zeigt die Funktionsweise einer räumlichen Gewichtungsfunktion. Die Gewichte w_{ij} werden als Funktion der euklidischen Distanz d_{ij} zwischen dem Regressionspunkt i und allen anderen Datenpunkten j spezifiziert. Die Beobachtungsbandbreite (*bandwidth*) kann als unveränder-

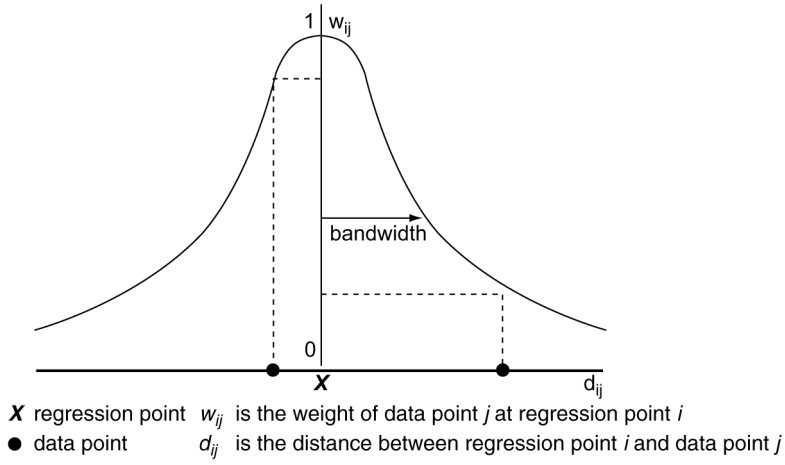

Abbildung 3.1: Räumliche Gewichtungsfunktion, FOTHERINGHAM et al. (2002, S.44)

bar oder als adaptiv angenommen werden. Eine adaptive Beobachtungsbandbreite bedeutet, dass der Gewichtungsfunktion erlaubt wird, räumlich in ihrer Form zu variieren. In Regionen, wo viele Datenpunkte zur Verfügung stehen, wird der lokale Beobachtungsraum verkleinert und in Regionen, wo die Dichte der Datenpunkte niedrig ist, wird der lokale Beobachtungsraum erhöht. Abbildung 3.2 veranschaulicht die Idee.

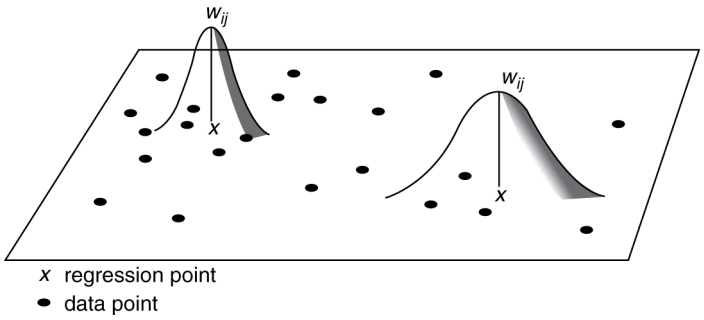

Abbildung 3.2: Adaptive Gewichtungsfunktion, FOTHERINGHAM et al. (2002, S.47)

Beobachtungsbandbreite (*bandwidth*) und Selektionskriterien

In den oben vorgestellten Gewichtungsfunktionen 3.12 und 3.13 befindet sich der unbekannte Parameter b für die Beobachtungsbandbreite. Es wurde bereits auf dessen Definition als maximale Ausdehnung der Gewichtungsfunktion hingewiesen. Der unbekannte Parameter b muss vor der Kalibrierung des GWR Modells aus den Daten geschätzt werden. Für diese Schätzung stehen unterschiedliche Selektionskriterien zur Verfügung, wobei die folgenden drei Zugänge am häufigsten Anwendung finden (vgl. FISCHER und GETIS, 2010, S.465):

- Direkte Festlegung der Bandbreite b durch Bestimmung der „number of nearest neighbours“ (vgl. MCMILLEN, 1996)
- „cross-validation (CV)“ (vgl. FOTHERINGHAM et al., 2002)
- „corrected Akaike Information Criterion (AIC_c)“ (vgl. FOTHERINGHAM et al., 2002)

Cross-Validation (CV) ist ein iterativer Prozess, in dem eine Beobachtungsbandbreite b gesucht wird, welche den „Root Mean Squared Prediction Error (RMSPE)“ minimiert (vgl. FISCHER und GETIS, 2010, S.465), wobei es sich um eine leave-one-out Cross-Validation handelt, da der Punkt i bei der Schätzung von y_i nicht mit eingezogen wird. Das korrigierte Akaike Information Criterion (AIC_c) verfolgt einen ähnlichen Ansatz wie das CV, wobei seine Anwendbarkeit allgemeiner ist und einen Kompromiss zwischen Goodness-of-Fit des Models und der Modellkomplexität (Anzahl der Freiheitsgrade) darstellt (vgl. FOTHERINGHAM et al., 2002, S.61).

Mixed GWR

GWR hat als lokales Modell mit Problemen zu kämpfen und steht auf Grund ihres methodischen Ansatzes teilweise unter Kritik (vgl. FOTHERINGHAM et al.,

2002; Löchl, 2010; Lehner, 2011; Fischer und Getis, 2010; Fuhrer, 2012). Fischer und Getis (2010, S.467) führt dazu an,*„[...] that GWR does not propose a base model for the source of the variation, and is thus more appropriately seen as a heuristic approach.“* Aus diesem Grund wurde die Methode weiterentwickelt. Im Zuge dieser Arbeit soll in aller kürze eine Erweiterung der GWR zur Mixed GWR vorgestellt werden, wobei andere Ansätze ebenfalls vorhanden sind (vgl. Fischer und Getis, 2010, S.472ff). Grundidee der Mixed GWR ist, dass manche Parameter räumlich nicht-stationär sind und andere wiederum nicht über den Beobachtungsraum variieren und somit global geschätzt werden können.

$$y_i = \sum_{j=1}^{k_a} a_j x_{ij}(a) + \sum_{l=1}^{k_b} b_l(u_i, v_i) x_{il}(b) + \varepsilon_i \tag{3.14}$$

In der oben geschriebenen mathematische Notation 3.14 ist für die Beobachtung i, y_i die abhängige Variable, (u_i, v_i) die geographische Information, $\{a_1 \dots a_{k_a}\}$ sind die k_a globalen Koeffizienten und $\{b_1(u,v) \dots b_{k_b(u,v)}\}$ sind die k_b lokalen Koeffizienten. Die unabhängigen Variablen werden in zwei Gruppen unterteilt, je nach dem ob deren Einfluss auf die abhängige Variable y_i global oder lokal geschätzt wird. $\{x_{i1}(a) \dots x_{ik_a}\}$ sind die unabhängigen Variablen, welche zu den globalen Koeffizienten gehören. $\{x_{i1}(b) \dots x_{ik_b}(b)\}$ sind die unabhängigen Variablen, welche zu den lokalen Koeffizienten gehören.

Simultaneous Autoregressive Models (SAR)

Ein populärer Ansatz zur Berücksichtigung von räumlichen Effekten in einer hedonischen Regression ist die Anwendung von *„Simultaneous Autoregressive Models (SAR)“* (vgl. Axhausen und Löchl, 2010; Dreger et al., 2011; Kissling und Carl, 2008; Anselin, 1988b; LeSage und Pace, 2009).
Das einfachste SAR Modell geht davon aus, dass die Response Variable y_i am Standort i nicht nur als Funktion der erklärenden Variablen X am Standort i gesehen werden kann, sondern darüber hinaus durch die Response Variablen der benachbarten Standorte j erklärt wird. Es wird somit das Vorhandensein von räumlicher Autokorrelation unterstellt, welche bei Immobiliendaten, die von Grund auf eine räumliche Dimension aufweisen, nicht *a priori* ausgeschlossen werden kann (vgl. Choumert et al., 2014). Die Nachbarschaftsbeziehungen werden durch eine räumliche Gewichtungsmatrix W modelliert. Die Elemente w_{ij} der Gewichtungsmatrix bilden den Zusammenhang zwischen Standort i und j ab und die Hauptdiagonale besteht aus Nullen. Für die Konstruktion von W stehen unterschiedliche Ansätze zur Verfügung (vgl. Bivand, 2014), welche im unteren Abschnitt noch genauer beleuchtet werden.
In der Literatur sind drei unterschiedliche Ansätze zur Modellierung von räumlicher Autokorrelation am verbreitetsten. Im Zuge dieser Arbeit folgt die Unterscheidung der drei SAR Modelle in ein *„Spatial Simultaneous Autoregressive*

Lag Model (SAR_{lag})“, ein *„Spatial Simultaneous Autoregressive Error Model (SAR_{err})*“ sowie ein *„Spatial Durbin Model (SAR_{mix})*“ (vgl. Kissling und Carl, 2008). Die Modelle werden über Maximum-Liklelihood (ML) geschätzt. Ganz allgemein kann gezeigt werden, das ML Schätzer eine Reihe von angenehmer asymptotischer Eigenschaften haben, welche auch in Gegenwart von räumlichen Lags in Variablen gelten. ML Schätzer sind asymtotisch erwartungstreu, konsistent, asymtotisch effizient und asymtotisch normalverteilt, insofern bestimmte Annahmen, auf die hier im Detail nicht eingegangen werden soll, nicht verletzt sind (vgl. Stocker, 2013, K.18,S.2).
Die Unterscheidung zwischen den drei Ansätzen ruht darauf, wo die räumliche Autokorrelation vermutet wird. Beim SAR_{lag} Modell wird davon ausgegangen, dass die räumliche Autokorrelation durch die Response Variable eingeführt wird und eignet sich somit dafür Spillover Effekte der endogenen Variable im Modell abzubilden(vgl. Dreger et al., 2011). Man könnte somit sagen, dass die Auswirkungen von Immobilienpreise an benachbarten Standorten auf den betrachteten Preis am Standort i mit ins Modell einfließen. Ein SAR_{lag} Modell kann wie folgt geschrieben werden:

$$\boldsymbol{y} = \rho \boldsymbol{W} y + \beta \boldsymbol{X} + \varepsilon \tag{3.15}$$

Wie gewohnt ist β ein Vektor mit Regressionskoeffizienten der exogenen Variablen X und ε stellt einen Errorterm dar. ρ ist der räumliche Autokorrelationskoeffizient und W ist die räumliche Gewichtungsmatrix. Der Term $\rho \boldsymbol{W}$ (Spatial Lag) spiegelt somit die räumliche Autokorrelation der Response Variable y im Sample wieder.
Im Gegensatz zum SAR_{lag} Modell wird beim SAR_{err} Modell die räumliche Korrelation über den Fehlerterm ins Modell eingeführt und es kommt zur Inklusion von räumlich korrelierten Störgrößen. Motiviert wird dieser Ansatz durch die Tatsache, dass die räumliche Autokorrelation nicht durch die ins Modell einfließenden exogenen Variablen erklärt werden kann. Der Grund dafür ist unter anderem in der möglichen Fehlspezifikation des Modells und durch das Weglassen von wichtigen erklärenden Variablen (ommitted variable bias) zu sehen. Das SAR_{err} Modell kann wie folgt geschrieben werden:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u} \tag{3.16}$$

wobei der Störterm u einem räumlich autokorrelierten Prozess folgt und als

$$\boldsymbol{u} = \lambda \boldsymbol{W} u + \varepsilon \tag{3.17}$$

formuliert wird. Hier ist λ der räumliche Autokorrelationskoeffizient und die übrigen Ausdrücke werden wie oben definiert. Der Preis einer Immobilie an

einem betrachteten Standort ist somit im SAR_{err} Modell nicht nur eine Funktion der erklärenden Variablen an diesem Standort sondern auch der „ommited variables“der benachbarten Standorte (vgl. Axhausen und Löchl, 2010).
ρ und λ als räumliche Autokorrelationskoeffizienten spiegeln die Stärke des Einflusses der räumlichen Gewichtungsmatrix W wider. Sehr hohe negative oder positive Werte für ρ und λ repräsentieren starke räumliche Effekte, ist ρ und λ allerdings Null, so sind keine räumlichen Effekte vorhanden (vgl. Fischer und Getis, 2010, 268).
Als dritte Ansatz steht das SAR_{mix} Modell (oft auch SAR_{Durbin} Modell) zur Verfügung.

$$\boldsymbol{y} = \rho \boldsymbol{W} y + \beta \boldsymbol{X} + \boldsymbol{W} \boldsymbol{X} \gamma + \varepsilon \tag{3.18}$$

In diesem Fall wird neben dem Spatial Lag der endogenen Variable y auch ein Spatial Lag der exogenen Variblen mit in die Modellgleichung aufgenommen. Der zusätzliche Term $WX\gamma$ setzt sich aus dem räumlichen Autoregressionskoeffizienten γ und dem Spatial Lag der exogenen Variablen WX zusammen(vgl. Dreger et al., 2011, S.74). Es ist an dieser Stelle wichtig anzumerken, dass beim SAR_{lag} Modell und dem SAR_{mix} Modell, durch die Aufnahme eines Spatial Lags der endogenen Variable in die Modellgleichung, die Interpretation der Schätzkoeffizienten für die erklärenden Variablen erschwert wird (vgl. Dreger et al., 2011; LeSage und Pace, 2009).

Räumliche Gewichtungsmatrix im SAR Modell

Neben der Wahl des richtigen Modells spielt die Spezifikation der räumlichen Gewichtungsmatrix W eine wichtige Rolle bei der Kalibrierung des Modells (vgl. Dreger et al., 2011), da durch sie die Nachbarschaftsbeziehungen mit aufgenommen werden (vgl. Choumert et al., 2014). Eine umfassende Darstellung der möglichen Methoden zur Konstruktion der räumlichen Gewichtungsmatrix ist in Bivand (2014) zu finden. Die einfachste Form der Gewichtungsmatrix ist eine binäre Matrix, welche Standorte entweder als Nachbarn deklariert oder nicht. Diese einfache Form wurde weiterentwickelt und im Zuge der Arbeit werden räumliche Gewichtungsmatrizen getestet, die Nachbarschaftsbeziehungen distanzbasiert oder über die k-nächsten Nachbarn modellieren. Alle Matrizen werden über ihre Zeilensumme standardisiert, um die Interpretation der zu Grunde liegenden Modelle zu erleichtern (vgl. Choumert et al., 2014, S.11).

Beurteilung der Modellgüte im SAR Modell

Für die Bewertung des Goodness-of-Fit eines OLS geschätzten Modells wurde bereits das R^2 bzw. das $adj.R^2$ vorgestellt. Für die Bestimmung der Modellgüte im SAR Modell ist diese Maßzahl allerdings nicht mehr anwendbar. Alternativ wird versucht, den Log-Likelihood zu maximieren sowie das Akaike's

Information Criterion (AIC) und das Schwartz Kriterium (oft auch Bayesian Information Criterion) zu minimieren (vgl. Löchl, 2010, S.54).

3.3 Explorative (räumliche) Datenanalyse

Löchl (2010, S.6) nennt die explorative (räumliche) Datenanalyse als eine der vier Hauptfelder der räumlichen Analyse und definiert diese wie folgt: *„Those techniques collectively known as exploratory data analysis which consists of methods to explore data (and also model outputs) in order to suggest hypotheses or to examine the presence of unusual values in the data set. Often, exploratory data analysis involves the visual display of spatial data generally linked to a map "* Im Zuge der Arbeit wird diese Definition übernommen. Zur Prüfung der Verletzung von Modellannahmen wird in den folgenden Abschnitten auf die Testung von Kollinearität und räumlicher Autokorrelation eingegangen, sowie auf das Problem der Heteroskedastizität hingewiesen. Darüber werden Möglichkeiten der Findung von Custern resp. räumliche Hot-Spots/Cold-Spots diskutiert.

Heteroskedastizität

Die Modelldiagnostik ist ein wesentlicher Bestandteil bei der Datenanalyse, da sie Hinweise auf die Verletzung von gemachten Annahmen geben kann (vgl. dazu Abschnitt 3.1). Homoskedastizität (Annahme der Streuungsgleichheit) liegt vor, sobald die Residualwerte (Störterme) die gleiche Streuung aufweisen (vgl. Huib, 2011, S.136). Ist diese Annahmen verletzt, spricht man von Heteroskedastizität. Die Konsequenz von Heteroskelastizität ist, dass der OLS Schätzer für den Parameter $\hat{\beta}$ zwar erwartungstreu ist (unter der Exogenitätsannahme) allerdings nicht mehr effizient (varianzminimal) und somit nicht BLUE *(„Best Linear Unbiased Estimator")* ist. Streng genommen sind auch die t-Tests nicht mehr zulässig (vgl. Huib, 2011). Im Zuge der Arbeit wird der Breusch-Pagan Test eingesetzt (vgl. Stiassny, 2012, S.79). Heteroskedastizität ist oftmals in der Fehlspezifikation des Modells begründet bzw. wird sie bei räumlichen Daten durch die räumliche Nicht-Stationarität der Schätzparameter verursacht (Erklärende Variablen haben eine starke Aussagekraft in manchen Regionen, aber schwache Aussagekraft in anderen).

Kollinearität

Multikollinearität liegt vor, wenn zwei oder mehr unabhängige Variablen (perfekt) miteinander korreliert sind. Dies führt dazu, dass die betroffene Variable, für die eine lineare Abhängigkeit betreffend den anderen unabhängigen Variablen besteht, für die Erklärung der abhängigen Variable redundant ist (vgl. Huib, 2011; Stocker, 2013). Zur Konsequenz von Multikollinearität führt Stocker (2013, K.10,S.7) an, dass der OLS-Schätzer nach wie vor BLUE

ist, allerdings lassen sich die Einflüsse einzelner unabhängiger Variablen bei Kollinearität schlechter isolieren und die Standardfehler werden zwar richtig aber auch größer geschätzt. Im Zuge der Arbeit wird die Mulitkollinearität mit dem *„Variance Inflation Factor (VIF) “* geprüft:

$$VIF_j = \frac{1}{1 - R^2} \tag{3.19}$$

In der Literatur finden sich unterschiedliche Angaben zur Höhe eines akzeptablen VIF, wobei oft vom Wert 10 als Obergrenze ausgegangen wird.

Räumliche Autokorrelation

Auf das theoretische Konzept der räumlichen Autokorrelation wurde bereits in Abschnitt 2.4 eingegangen. Statistisch kann die Stärke und Signifikanz von räumlicher Autokorrelation mit Hilfe des Morans Index (I) gemessen werden:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{3.20}$$

Wobei n die Anzahl der Beobachtungen ist, x_i stellt die ite Beobachtung dar, $\overline{x}$ das Mittel aller x und w_{ij} gibt die räumliche Gewichtung zwischen dem Punkt i und dem Punkt j wieder.

Identifizierung geografischer Muster

Die Suche nach räumlichen Mustern wird in einer Vielzahl von Publikationen behandelt und kann als eines der fundamentalen Themen der räumlichen Datenanalyse bezeichnet werden. Als gute Beispiele zur Abhandlung von unterschiedlichen Methoden gelten Fischer und Getis (2010), Fischer und Wang (2011) oder im speziellen in Bezug auf die Statistik Software R auch Bivand et al. (2008). Im Zuge der Arbeit wurden bereits die GWR Modelle vorgestellt, welche als eine Weiterentwicklung des Moving-Window Ansatzes als explorative Methode zur Identifizierung von räumlichen Mustern eingesetzt werden können (vgl. Abschnitt 4.2). Zur Findung von Hot-Spots bzw. Cold-Spots stehen auch andere Indikatoren zur Verfügung, wie beispielsweise $Getis - OrdGi*$ (vgl. Fischer und Wang, 2011, S.26). Für die Regressionsmodelle werden Ad-Hoc Cluster getestet, die sich an der Dichte der Beobachtungen und an den in aktuellen Immoblienmarktberichten festgeschriebenen Bürozentren in Wien orientieren.

Kapitel 4

Daten

Dieses Kapitel stellt die zur Modellbildung herangezogenen Daten vor. In Abschnitt 4.1 werden die Bezugsquellen offen gelegt. Der darauf folgende Abschnitt 4.2 zeigt die notwendigen Schritte der Datenaufbereitung auf und beinhaltet deskriptive Statistiken zu den einzelnen im Modell verwendeten Variablen.

4.1 Datenbasis

Dem Fachbereich für Stadt- und Regionalforschung der TU Wien steht ein Datensatz mit Angebotsdaten von Immobilien der Fa. Ediorg für Forschungszwecke zur Verfügung. Für das Büroimmobiliensegment beinhaltet der unbereinigte Datensatz etwa 12.250 Beobachtungen für gesamt Österreich zwischen den Jahren 1999 und 2013. Neben dem Mietpreis werden teilweise bauliche Charakteristika angeführt, wobei vor allem das Baujahr, eine Einschätzung des Zustandes, das Angebotsjahr und die Größe der Immobilie für den Großteil der Beobachtungen aufgelistet ist. Darüber hinaus werden für einige Immobilien Angaben zur Heizungsart, dem Stockwerk oder dem Vorhandensein eines Liftes gemacht. Jeder Angebotsfall im Datensatz wurde georeferenziert und besitzt somit eine räumliche Information.
Die Geodaten werden von Open Government Data Wien (`https://open.wien.at/site/`) bzw. aus dem offenen Datenkatalog auf `data.gv.at` bezogen. Eine Liste mit den Links zu den verwendeten Daten findet sich im Anhang der Arbeit.

4.2 Datenaufbereitung und deskriptive Statistik

In einem ersten Schritt wurden die Beobachtungen im Wiener Stadtgebiet über die Postleitzahl aus dem Angebotsdatensatz für Österreich heraus gelöst. Es verbleiben 12.000 Fälle. Von diesen wurden über die Kriterien geographische Position (xy Koordinaten), die Größe der Immobilie und die Nettomiete

gesamt etwa 1.100 Duplikate ermittelt. Beobachtungen die entweder keine Angaben zur Größe der Immobilie oder zur Nettomiete machen werden ebenfalls ausgeschlossen, da eine Kalkulation der Monatlichen Nettomiete pro m^2 dadurch unmöglich wird. Wie bereits in Abschnitt 2.1 dargestellt wurde, liegen die Spitzenmieten für Büroimmobilien in Wien bei etwa 26 Euro (vgl. EHL, 2014). Etwa 350 Fälle weisen allerdings einen Mietpreis pro m^2 zwischen 30 und 40 Euro auf. Diese Fälle wurden nach eingehender Analyse ebenfalls für die Modellbildung ausgeschlossen, da kein Grund für eine so hohen Mietzins gefunden werden konnte und von einer fehlerhaften Eingabe der Daten auszugehen ist. Es verbleiben 8.162 Fälle. Es liegen offensichtlich einige Fälle direkt am Bezirksmittelpunkt und sind somit falsch georeferenziert. Bei der Berechnung von Lageparametern und Distanzmaßen dürfen diese Beobachtungen im Sample nicht in die Analyse mit einbezogen werden, da dies zu starken Verzerrungen führen würde. Die Angebotsfälle im Datenset, welche keine Adresse aufweisen und am Bezirksmittelpunkt georeferenziert sind, wurden somit aus dem Datensatz entfernt. Es verbleiben 6.236 Fälle.

Angebotsdaten

Der Mietpreis pro m^2 geht ins Modell als abhängige Variable ein. Der Boxplot 4.1 zeigt, die Verteilung der Mietpreise nach Bezirken differenziert. Im Plot sind für fast alle Bezirke eine Reihe von Ausreißern zu beobachten. Nicht unerwartet zeigen die Bezirke 1 und 22 deutlich höhere Mietpreise pro m^2 im Median als die anderen 21 Wiener Gemeindebezirke. Abbildung 4.2 zeigt die Medianwerte auf einer Bezugsebene eines 300 mal 300 Meter Rasters und Abbildung 4.3 den Anteil der Mietflächen an den Gesamtflächen im Datensatz je 300 mal 300 Meter Rasterzelle. Eine Konzentration hoher Mietpreise pro m^2 im ersten Bezirk und in der Donaucity sind auch hier ablesbar, allerdings zeigt sich, dass vereinzelt auch in anderen Bezirken hohe Mieten für Büroimmobilien möglich sind. Dies gilt vor allem für Teile des dritten Bezirkes und für die Region Nord.

Bauliche und vertragliche Charakteristika

Der Rohdatensatz mit 8.162 Fällen macht Angaben zu einer Reihe von baulichen Charakteristika. Aufbauend auf den Erkenntnissen der Literaturrecherche in Abschnitt 2.3 wurden Variablen zum Baujahr der Immobilie, zum Zustand des Gebäudes und zum Vorhandensein eines Liftes in die Analyse aufgenommen. Diese Variablen wurden teilweise umkodiert und neu berechnet. Die Tabelle 4.1 gibt einen Überblick zu den verwendeten baulichen Charakteristika.

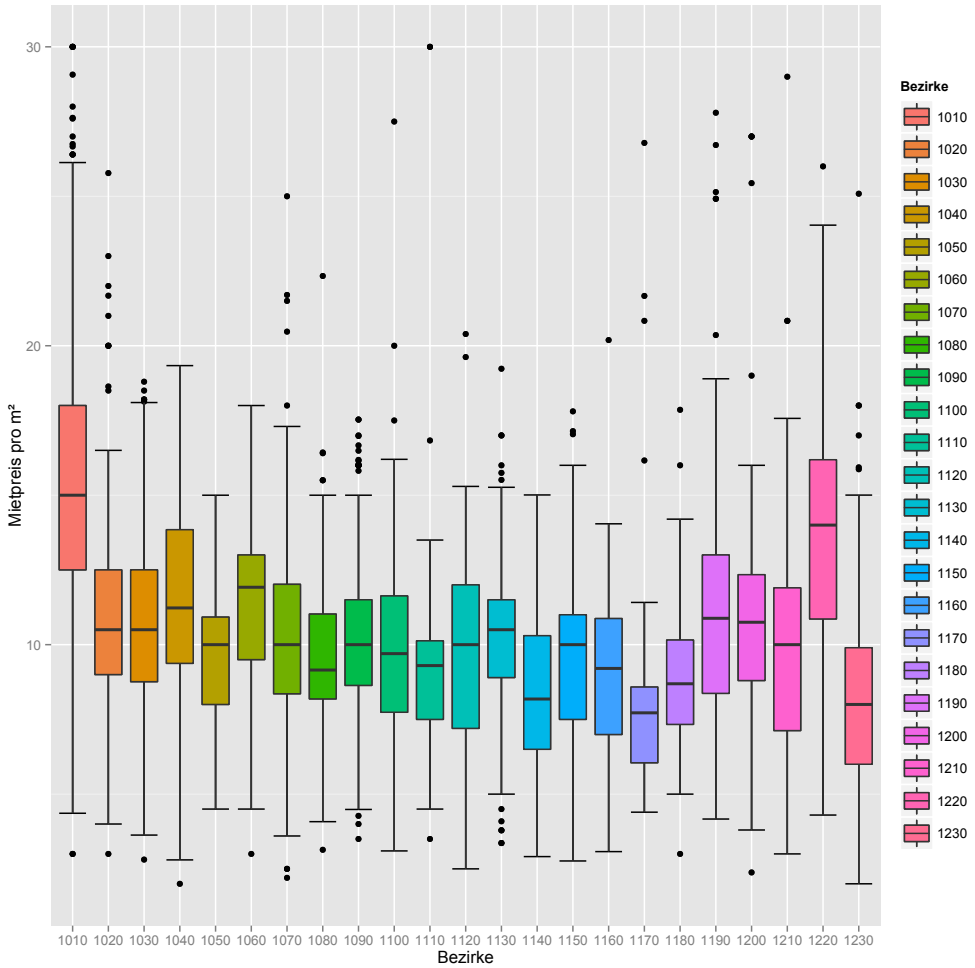

Abbildung 4.1: Boxplot: Mietpreise pro m^2 nach Wiener Bezirke, eigene Darstellung.

Bauperiode und Gebäudealter

Kaum in einem Immobiliensegment ändern sich die Anforderungen an Architektur, Raumlayout und Komfort so rasant über den Lebenszyklus wie bei Büroimmobilien. Oft ist die Anpassung durch Modernisierungsmaßnahmen auf Grund von unflexiblen Grundrissen, baurechtlichen und bautechnischen Gegebenheiten oder aus wirtschaftlichen Gründen nur bedingt möglich. Es kann somit angenommen werden, dass Immobilien mit einem geringen Gebäudealter die Anforderungen an ein modernes Bürogebäude eher erfüllen und in Hinblick auf die technische Gebäudeausrüstung (Konditionierung des Gebäudes, mechanische Be- und Entlüftungsanlagen, moderne brandschutztechnische Anlagen usw.), der thermischen Qualität, der Architektur und der

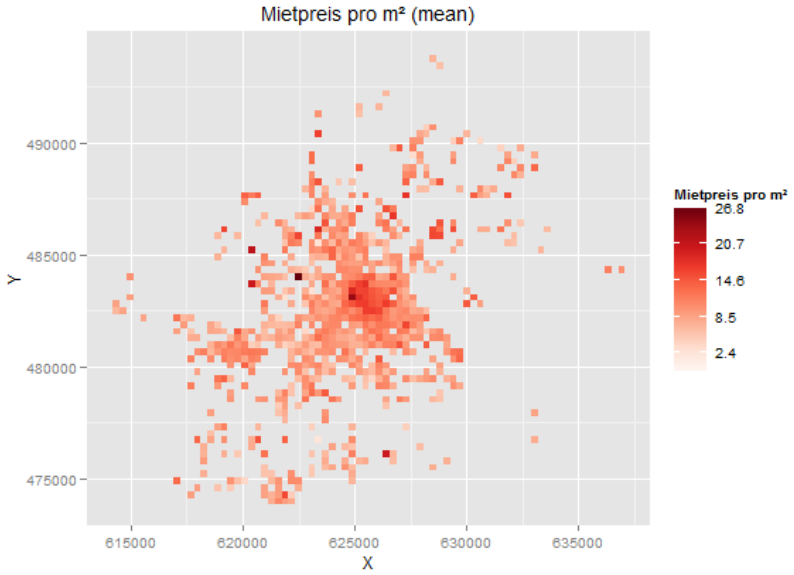

Abbildung 4.2: Mietpreise pro m^2 (Median) auf einer Bezugsfläche von 300 mal 300 Metern, eigene Darstellung.

Tabelle 4.1: Übersicht der Modellvariablen - Bauliche Charakteristika

Variable	Beschreibung	Einheit
Bauliche Charakteristika		
BP1	Bauperiode vor 1849	*D*
BP2	Bauperiode 1849 bis 1918	*D*
BP3	Bauperiode 1919 bis 1945	*D*
BP4	Bauperiode 1946 bis 1976	*D*
BP5	Bauperiode 1977 bis 2000	*D*
BP6	Bauperiode 2001 bis Heute	*D*
BP_KA	keine Angabe zur Bauperiode	*D*
cond_1	sehr guter Zustand	*D*
cond_2	guter Zustand	*D*
cond_3	befriedigender bis schlechter Zustand	*D*
cond_ka	keine Angabe zum Zustand	*D*
floor_area	Größe der Immobilie in m^2	m^2
lift_1	Lift im Gebäude vorhanden	*D*
lift_0	Lift im Gebäude nicht vorhanden	*D*
lift_ka	Keine Angaben zum Lift	*D*

D: Dummy Variable

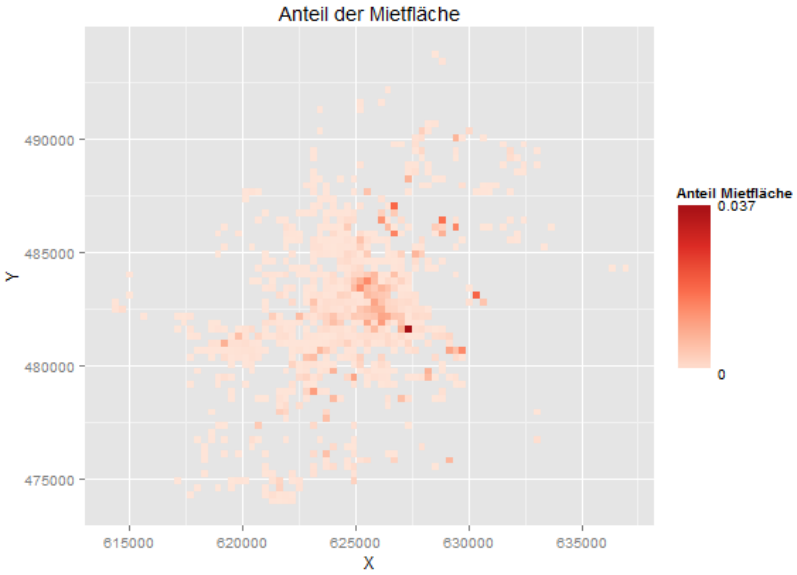

Abbildung 4.3: Anteil der Mietflächen an allen Mietflächen im Datensatz auf einer Bezugsfläche von 300 mal 300 Metern, eigene Darstellung.

Flexibilität der Grundrisslayouts stärker am Markt nachgefragt werden und somit einen höheren Mietpreis generieren.
Das Baujahr ist für etwa 60% der Daten vorhanden, woraufhin das Gebäudealter bis ins Jahr 2014 bestimmt wurde. Der Boxplot 4.4 zeigt die Verteilung der Büroimmobilien im Datensatz bezogen auf das Alter differenziert nach Bezirken. Wie erwartet liegt der Median für das Alter der Gebäude für die Bezirke 1, 4, 6, 7, 8, 9 und 18 über der 100 Jahre Grenze, wobei teilweise hohe Streuungen, Ausreißer und linksschiefe Verteilungen erkennbar sind. Für den zweiten Wiener Gemeindebezirk zeigt der Boxplot einen eher geringen Wert für den Median. Dies ist vermutlich mit neueren Entwicklungen im Nahbereich der Lassalestraße und der Messe Wien zu begründen. Ein ähnliches Bild zeigt der Plot für den dritten Bezirk, in dem zum einen die für die Büronutzung adaptierten Gebäude aus der Gründerzeit eine rechtsschiefe Verteilung erzeugen und die neuen Entwicklungen um Neu Marx und Erdberg den Median nach unten treiben. Für die Bezirke 10 bis 12 und 19 bis 22 liegt der Median für das Alter der Gebäude zwischen 10 und 15 Jahren. In diesen Gebieten liegen die jüngeren Büroentwicklungsgebiete Wiens (Region Süd und Wienerberg, Region Nord und die Donaucity). Für die Analyse werden die Daten über das Gebäudealter in Bauperioden umkodiert. Es kommt zu einer Unterscheidung, welche sich zum einen nach der Bauperiodenskala der Stadt Wien (vgl. WIEN.GV.AT, 2014b)richtet und zum anderen die Angaben der Wohnungs- und Gebäudezählung der Statistik Austria heranzieht (vgl. STATISTIK-AUSTRIA,

2004):

- Bauperiode 1 *(BP_1)*: vor 1849 Hochbarock, Rokoko, Klassizismus und Biedermeier
- Bauperiode 2 *(BP_2)*: 1849 bis 1918 Gründerzeit (Früh-, Hoch- und Spätgründerzeit)
- Bauperiode 3 *(BP_3)*: 1919 bis 1945 Zwischenkriegszeit
- Bauperiode 4 *(BP_4)*: 1946 bis 1976 Nachkriegszeit
- Bauperiode 5 *(BP_5)*: 1977 bis 2000 Postfordismus
- Bauperiode 6 *(BP_6)*: 2001 bis Heute

Zustand

Der Gesamtzustand der Immobilie in Hinblick auf dessen gebäudetechnische und bauliche Qualität wird mit Hilfe von drei Kategorien bewertet. Mit zunehmender Qualität wird angenommen, dass auch der Mietpreis zunimmt. Der Boxplot 4.5 bestätigt diese Annahme.

Größe der Immobilie

Da bei Unternehmen eine allgemeine Tendenz der Zentralisierung an einem Standort beobachtet werden kann (vgl. Haase, 2011), könnte angenommen werden, dass der Mietpreis pro m^2 mit der Größe der Büroeinheit steigt. Der Scatterplot 4.6 zeigt, dass die Nettomiete wie erwartet mit der Größe zunimmt. Die absolute Streuung ist bei großen Flächen höher als bei kleineren, wobei relativ gesehen augenscheinlich kein Unterschied bestehen dürfte. Im Scatterplot 4.7 ist zu sehen, dass die meisten Beobachtungen zwischen 2 und 2500 m^2 liegen. Für die Darstellung im Plot 4.6 und 4.7 wurden auf Grund der besseren Lesbarkeit alle Immobilien mit einer Gesamtflächen über 8500 m^2 ausgeschlossen. Die 5 Punkt Zusammenfassung in Tabelle 4.2 zeigt die große Spannweite bei den Größen der untersuchten Immobilien. Die Spannweite im Datensatz reicht von 2 bis 80.000 m^2.

Tabelle 4.2: 5 Punkt Zusammenfassung - Größe der Immobilien

Min.	1st Qu.	Median	3rd Qu.	Max.
2	128.5	229	428	80000

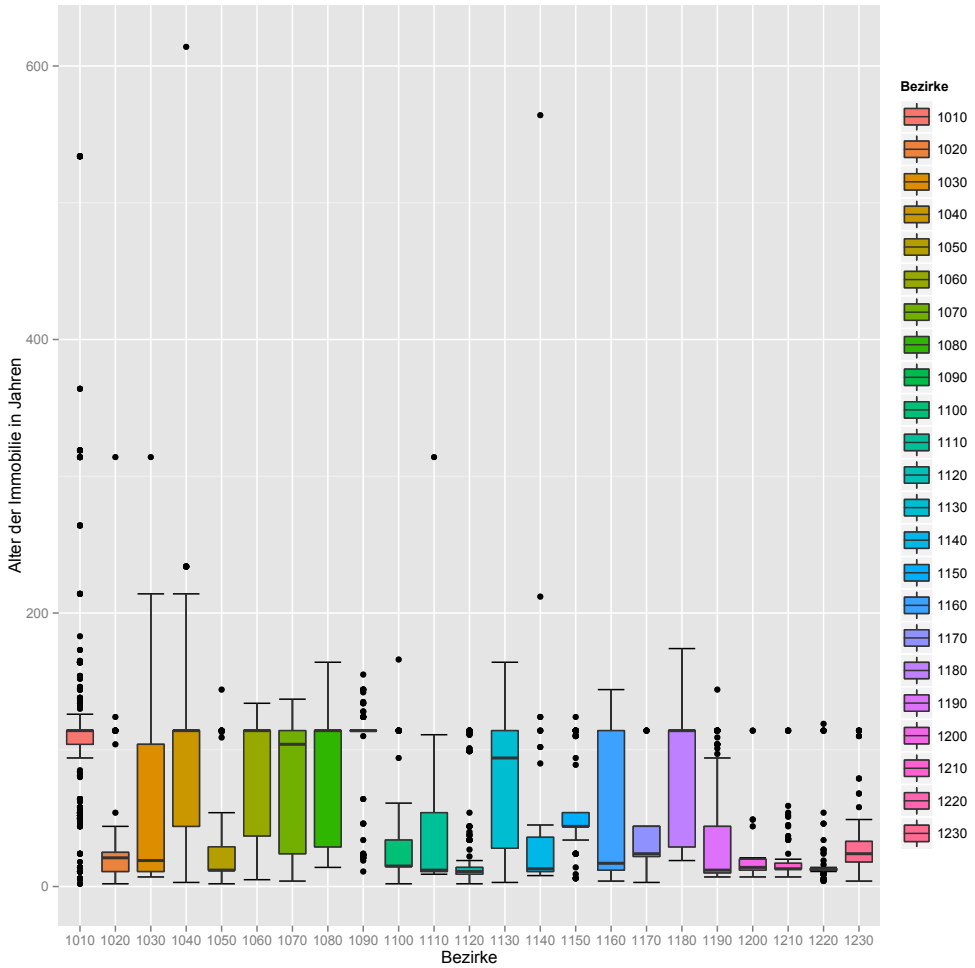

Abbildung 4.4: Boxplot: Alter der Immobilien nach Wiener Bezirke, eigene Darstellung.

Angebotsjahr

Im Zuge der Literaturrecherche konnte gezeigt werden, dass die Art des Mietverhältnisses einen entscheidenden Einfluss auf den Mietpreis hat. Im vorliegenden Datensatz finden sich keine Angaben zur Art und Länge des Mietvertrages, was dazu führt, dass eine wichtige den Mietpreis pro m^2 potentiell beeinflussende Größe nicht in die Modellbildung miteinbezogen werden kann. Zyklische Schwankungen des Immobilienmarktes können mit Hilfe des Angebotsjahres im Modell kontrolliert werden. Die Angebotsjahre werden als Dummy Variablen für die Jahre 2003 bis 2013 kodiert (vgl. dazu Tabelle 4.3) und gehen so in die Analyse mit ein. Mit Hilfe dieser Variablen kann ein Zeit-Dummy-Index konstruiert werden, welcher einfach zu erstellen ist, allerdings

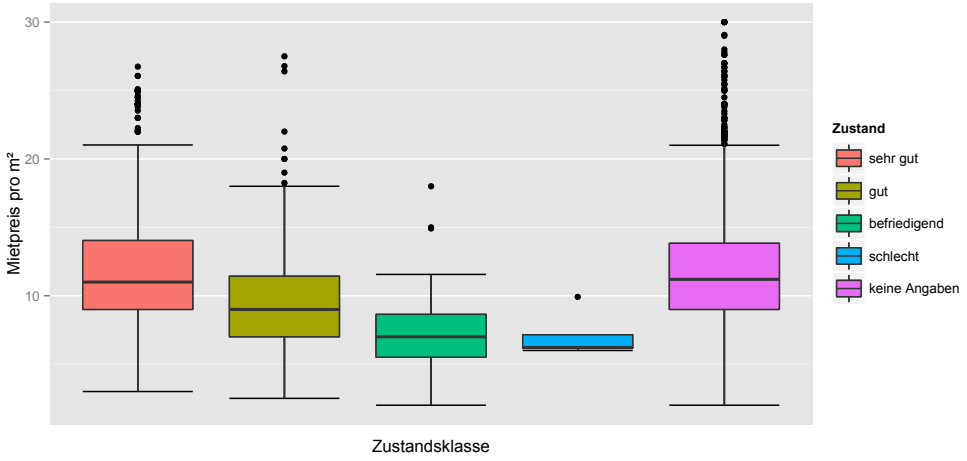

Abbildung 4.5: Boxplot: Mietpreise pro m^2 nach Zustandsklasse, eigene Darstellung.

mit einer Vielzahl von Problemen zu kämpfen hat. Im Zuge einer Verkettung der Dummy Werten ist anzunehmen, dass die Zahlungsbereitschaft für die Qualitäten der Immobilien über den Betrachtungszeitraum konstant ist. Ändert sich der Einfluss der Qualitätsvariablen, so ist das Ergebnis nicht mehr gültig (vgl. Bökemann und Feilmayr, 1994, S.76). Der Boxplot in Abbil-

Tabelle 4.3: Übersicht der Modellvariablen - Vertragliche Charakteristika

Variable	Beschreibung	Einheit
Vertragliche Charakteristika		
AJ2003	Angebots-/Abschlussjahr 2003	*D*
AJ2004	Angebots-/Abschlussjahr 2004	*D*
AJ2005	Angebots-/Abschlussjahr 2005	*D*
AJ2006	Angebots-/Abschlussjahr 2006	*D*
AJ2007	Angebots-/Abschlussjahr 2007	*D*
AJ2008	Angebots-/Abschlussjahr 2008	*D*
AJ2009	Angebots-/Abschlussjahr 2009	*D*
AJ2010	Angebots-/Abschlussjahr 2010	*D*
AJ2011	Angebots-/Abschlussjahr 2011	*D*
AJ2012	Angebots-/Abschlussjahr 2012	*D*
AJ2013	Angebots-/Abschlussjahr 2013	*D*

D: Dummy Variable

dung 4.8 zeigt den Mietpreis pro m^2 differenziert nach den Angebotsjahren. Es ist in den einzelnen Jahren keine Abhängigkeit des Mietpreises pro m^2 vom

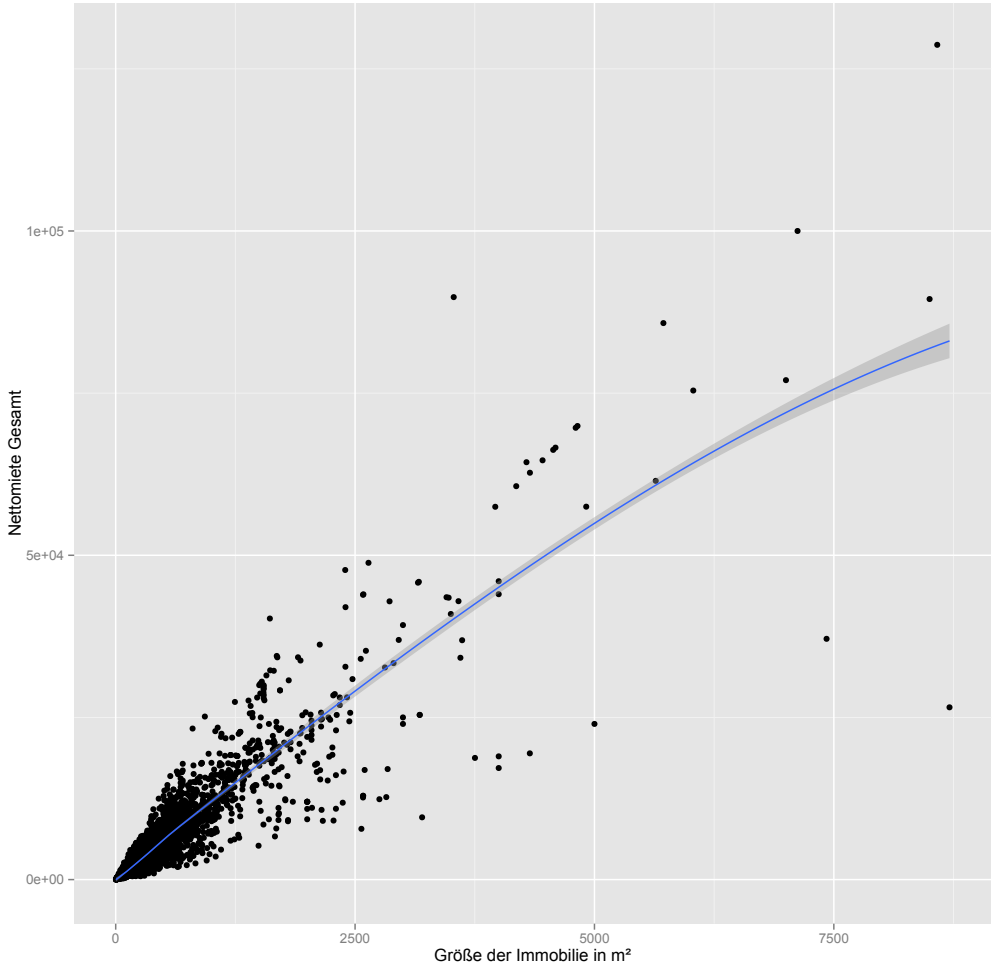

Abbildung 4.6: Nettomiete Gesamt bezogen auf die Fläche der Immobilie, eigene Darstellung.

Angebotsjahr zu erkennen. Ab dem Jahr 2009 nimmt die Zahl der Ausreißer nach Oben zu.

Lage- und Nachbarschaftscharakteristika

Tabelle 4.4 stellt die analysierten Lage- und Nachbarschaftscharakteristika vor.

Lagecharakteristika

Im Zuge der Arbeit wurden einfache Lageparamter mit Hilfe des *ModelBuilders* des GIS Softwarepakets *ArcGIS 10.2.2* berechnet. Datenbasis bilden die

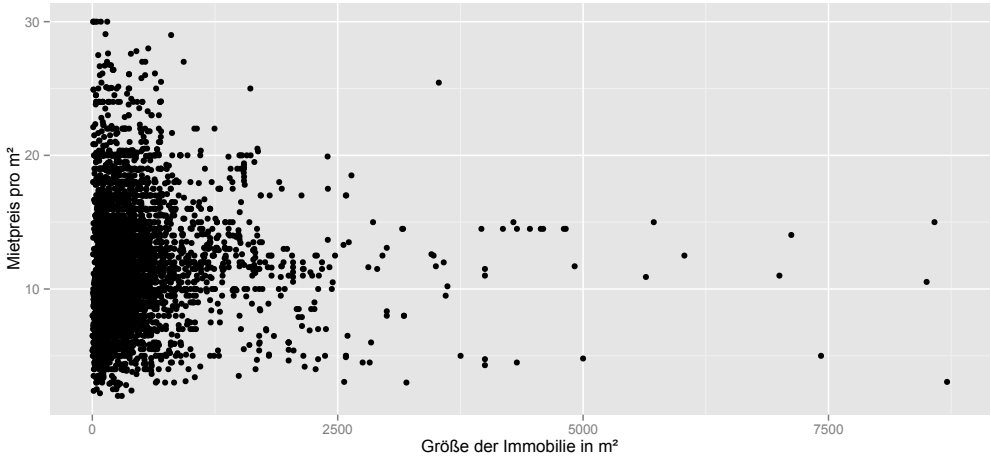

Abbildung 4.7: Mietpreis pro m^2 bezogen auf die Fläche der Immobilie, eigene Darstellung.

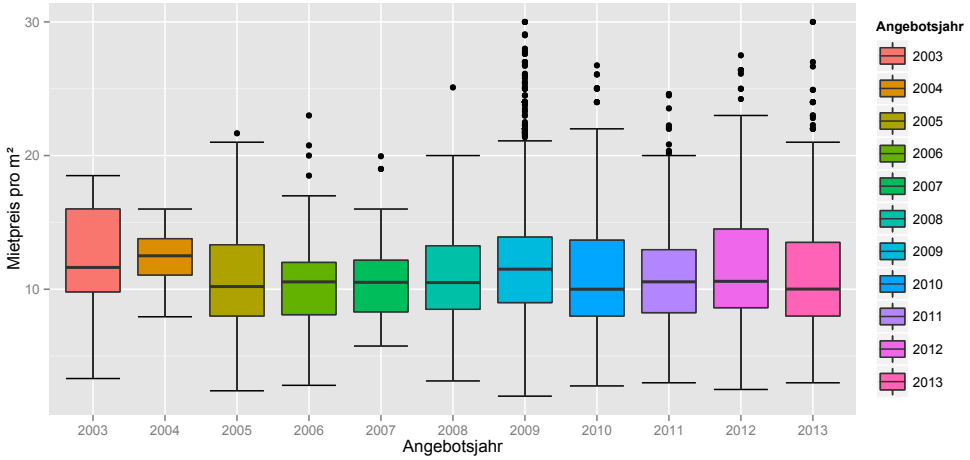

Abbildung 4.8: Boxplot: Mietpreise pro m^2 nach Angebotsjahr, eigene Darstellung.

auf Open Government Data Wien (`https://open.wien.at/site/`) bzw. auf `data.gv.at` frei zur Verfügung gestellten Geodaten. Diese Daten sind auf dem World Geodetic System 1984 (WGS84) referenziert. Die Angebotsdaten der Fa. Ediorg liegen auf einer alten MGI Lambert Austria Projektion. Im Zuge der Arbeit wurden die Angebotsdaten sowie die genutzten Geodaten auf die

Tabelle 4.4: Übersicht der Modellvariablen - Lage- und Nachbarschaftscharakteristika

Variable	Beschreibung	Einheit
Nachbarschaft		
DIST_CLUST	Nähe zum nächsten Büro-Cluster	m
city	Zugehörigkeit zum Cluster City	D
bezinnen	Zugehörigkeit zum Cluster Innere Bezirke	D
dc	Zugehörigkeit zum Cluster Donau City	D
Lage		
DIST_UBAHN	Distanz zur nächsten U-Bahnstation	m
DIST_SBAHN	Distanz zur nächsten S-Bahnstation	m
DIST_BUSTR	Distanz zur nächsten Bus oder Tram Station	m
DIST_WLB	Distanz zur nächsten WLB-Station	m
DIST_CBD	Distanz bis zum Stephansplatz	m
DIST_HIGHW	Distanz zur nächsten Autobahnauffahrt	m

D: Dummy Variable; m: Meter; n: nominal skalierte Variable

MGI Austria Lambert Projektion [1] umgerechnet. Die Transformation wurde mit der Open Source Software *QGIS Desktop 2.4.0* durchgeführt.

Die extern bezogenen Geodaten mussten teilweise bearbeitet werden. Der Datensatz der Wiener Linie und des Verkehrsverbundes Ost-Region beinhaltet Informationen zu Haltepunkten der S- und U-Bahn, Tram und Bus sowie zur Wiener Lokalbahn (WLB). Dieser Datensatz wurde nach Verkehrsmitteln selektiert und gesplittet. Darauf aufbauend wurde für jedes Verkehrsmittel eine Berechnung der euklidischen Distanz zwischen Haltepunkt und Büroimmobilie durchgeführt. Die gleiche Berechnungssystematik wird auch für die Entfernung zum Central Business District (CDB), welcher mit dem Stephansdom als zentralem Punkt definiert wurde, angewandt. Für die Entfernung zum höherrangigen Straßennetz (Schnellstraßen und Autobahn) wurde ein Straßengraf der ASFINAG herangezogen. Der frei verfügbare Datensatz enthält allerdings keine Informationen zu den Auf- und Abfahrten, weshalb diese händisch georeferenziert werden mussten. Die Ergebnisse der Analyse werden im Boxplot 4.9 und 4.10 dargestellt.

Die durchschnittliche Distanz zur nächsten U-Bahnstation liegt bei 525 Metern. Auf Grund von vielen Ausreißern liegt der Median allerdings deutlich unterhalb diesem Wert bei etwa 340 Metern. Der Boxplot für die Entfernung

[1] $+proj = lcc + lat_1 = 49 + lat_2 = 46 + lat_0 = 47.5 + lon_0 = 13.33333333333333 + x_0 = 400000 + y_0 = 400000 + ellps = bessel + towgs84 = 577.326, 90.129, 463.919, 5.137, 1.474, 5.297, 2.4232 + units = m + no_defs$

zur nächsten S-Bahnstation zeigt, dass die durchschnittliche Entfernung zum nächsten Haltepunkt deutlich über der der U-Bahn liegt, was eine weniger starke Netzabdeckung durch die Schnellbahn wiederspiegelt. Auch hier sind einige Ausreißer erkennbar. Die Verteilung ist mit einem Schiefkoeffizienten von 0.88 nur leicht rechtsschief. Die stärkste Netzabdeckung zeigt nicht unerwartet die Verkehrsmittel Bus und Tram. Der Mittelwert für die Entfernung Büroimmobilie zum nächsten Haltepunkt liegt bei 13 Metern. Die Wiener Lokalbahn (WLB) hat mit nur einer Linie einen sehr hohen mittleren Wert für die Entfernung Büroimmobilie und nächstgelegener Haltepunkt. Der Mittelwert liegt bei über 2000 Metern und somit 4 mal so hoch wie bei der U-Bahn. Die mittlere Distanz bis zum Stephansplatz liegt bei etwa 3000 Metern, mit einigen Ausreißern nach oben. Abbildung 4.11 zeigt in diesem Zusammenhang gut auf, dass eine große Anzahl an Beobachtungen innerhalb von 2.5 km vom Stephansplatz angesiedelt sind. Ein zweiter Peak ist bei 5 km zu erkennen, was auf eine polyzentrische Struktur der Bürostandorte hinweist. Diese Überlegungen decken sich auch mit dem Bild im Plot Mietpreis pro m^2 gegen die Distanz zum Stephansplatz in Abbildung 4.11. Die Mietpreise sind nahe dem Stephansplatz (CBD) durchschnittlich am höchsten und fallen mit der Distanz ab. Bei einer Entfernung von 5 km vom CBD ist allerdings eine leichte Steigerung der Mietpreise erkennbar.

Nachbarschaftscharakteristika

GWR Modelle werden verwendet, um die nicht-Stationarität des Einfluss einzelner Variablen im Raum zu testen und somit der Missspezifikation der geschätzten Modelle entgegenzuwirken. In dieser Arbeit soll eine Identifikation von Clustern und räumlichen Mustern durch die Verwendung dieser Modelle unterstützt werden, wobei aus den Ergebnissen ein Trend hinsichtlich räumlicher Heterogenität abgelesen werden soll, allerdings keine „rasterzellengenaue“Analyse der Ergebnisse sinnvoll ist und auch nicht angestrebt wird (vgl. Fuhrer, 2012). Die Ergebnisse aus der GWR Analyse werden mit Hilfe von Ordinary Kriging räumlich auf das gesamte Stadtgebiet interpoliert.

Es wurden mehrere GWR Modelle geschätzt, wobei unter anderem in der Spezifikation der räumlichen Gewichtungsfunktion unterschieden wurde. Die Modelle wurden über das AIC verglichen. Als Selektionskriterium für die Bandwidth wurde leave-one-out Cross-Validation verwendet. Der beste Model-Fit konnte durch eine fixe Bandwidth mit Gaussain Gewichtungsfunktion erreicht werden. Abbildung 4.12 stellt die Ergebnisse der GWR Analyse für das Modell dar. Bei der Betrachtung des lokalen R^2 ist in den inneren Bezirken und im Westgürtel eine gewisse Korrelation zwischen der Dichte der Beobachtungspunkte und der Höhe des Bestimmtheitsmaßes zu erkennen. Fotheringham (2000) weißt darauf hin, dass durch GWR auch lokale Parameter für $\hat{\boldsymbol{\beta}}(u_i, v_i)$ geschätzt werden, welche statistisch nicht signifikant sind und die

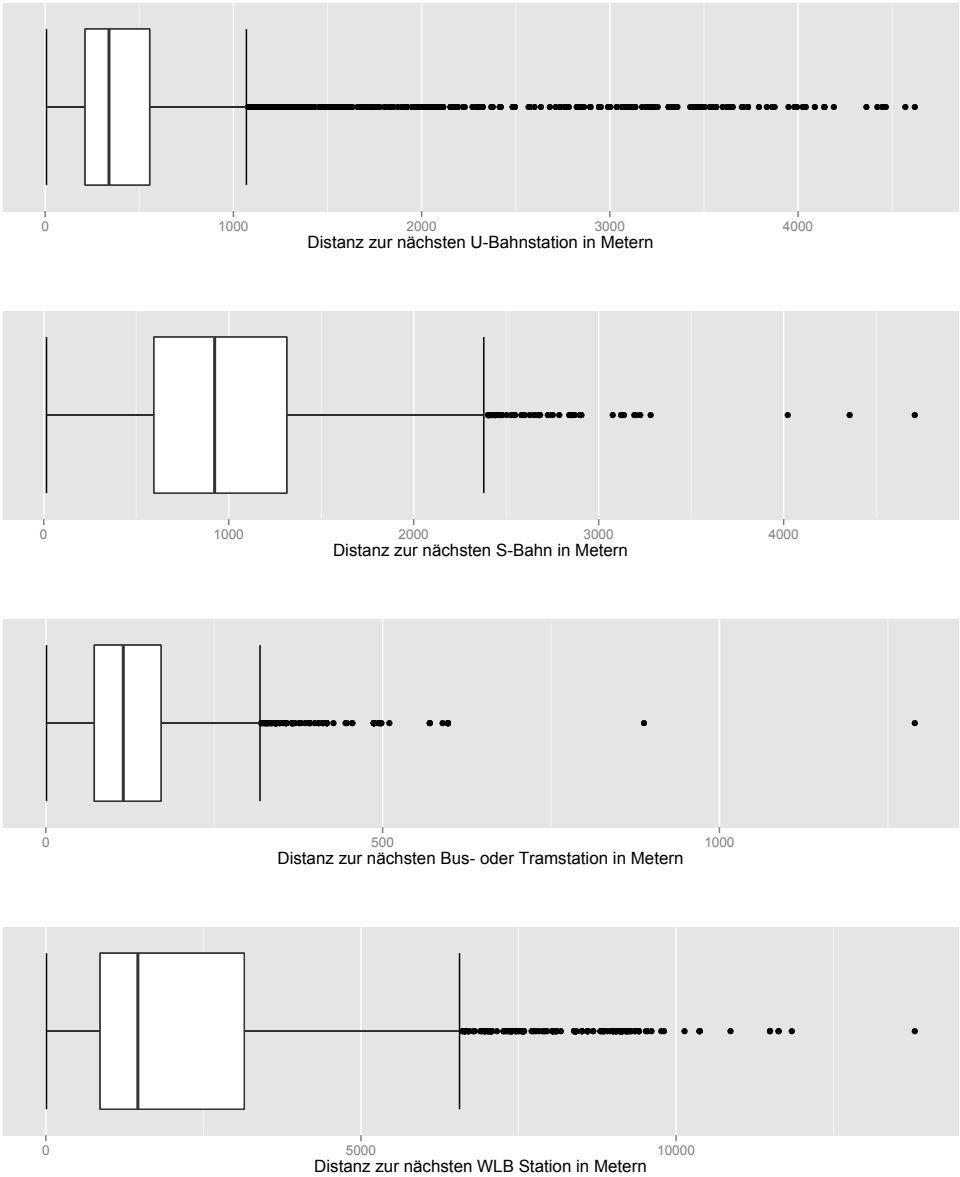

Abbildung 4.9: Boxplot: Distanz bis zur nächsten ÖV Station unterschieden nach Verkehrsmittel, eigene Darstellung.

Aussage über räumliche Trends dadurch verzerren können.

Die Tabelle 4.5 stellt die Ergebnisse des GWR Modells dar. Es wird der Median der geschätzten Koeffizienten, die Standardabweichung (SD) sowie die Minimum- und Maximumwerte gegenübergestellt. Hohe Standardabweichun-

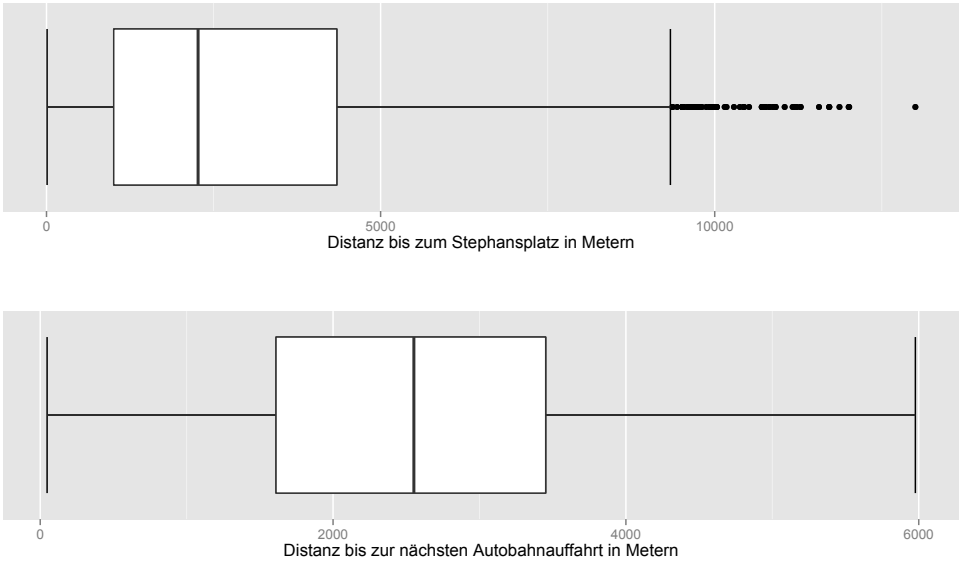

Abbildung 4.10: Boxplot: Distanz bis zum Stephansplatz und zur nächsten Autobahnauffahrt, eigene Darstellung.

gen für die geschätzten Koeffizienten weisen auf eine stärkere räumliche Variation hin (vgl. Lehner, 2011). Für alle lokal geschätzten Koeffizienten kann gezeigt werden, dass das Vorzeichen und somit der geschätzte Zusammenhang zwischen Mietpreis und unabhängiger Variablen lokal variieren kann. Besonders stark ist dies der Fall bei der logarithmierten Distanz zum Stephansplatz (CBD). Hier liegen die geschätzten Koeffizienten zwischen -20.723 und +7.635. Ebenfalls starke Unterschiede in den lokalen Schätzern wurden für die logarithmierte Distanz zur Wiener Lokalbahn festgestellt. Der kleinste Wert für den Beta-Schätzer ist -8.523, der Maximumwert liegt bei 22.026.

Es wird durch die GWR Analyse ein erster Hinweis darauf gegeben, dass der Einfluss von ausgewählten Struktur- und Lageparameter im Untersuchungsgebiet schwanken kann. Diese Tatsache ist bei der Modellbildung zu berücksichtigen, da durch die Missspezifikation unter anderem Heteroskedastizität eingeführt wird.

Ad-hoc Cluster wurden zur Testung in den Regressionsmodellen festgelegt. Diese orientieren sich an der Dichte der Beobachtungen im Untersuchungsgebiet (vgl. Abbildung 4.13), sowie an den in den aktuellen Immobilienmarktberichten teilweise festgeschriebenen Bürocluster (vgl. CBRE, 2014; Bank-Austria, 2014; EHL, 2014). Zusätzlich wird mit Hilfe des *ModelBuilders* des GIS Softwarepakets *ArcGIS 10.2.2* die Distanz zum jeweils nächsten Cluster berechnet und als erklärende Variable ins Modell mit aufgenommen. Die

Tabelle 4.5: Ergebnisse *GWR*

Endogene Variable:	*log(rent_sqm)*			N = 2498
Exogene Variable	Coeff. Median	SD	Coeff. Min.	Coeff. Max.
Intercept	3.662	1.379	-11.559	14.775
Bauliche Char.				
BP1	-0.084	0.170	-3.454	0.304
BP2	-0.204	0.091	-0.830	0.300
BP3	-0.205	0.144	-1.130	0.445
BP4	-0.250	0.088	-0.844	0.410
BP5	-0.130	0.095	-0.635	0.607
BP_KA	-0.140	0.060	-0.744	0.211
log(floor_area)	0.039	0.043	-0.372	0.073
cond_2	-0.155	0.040	-0.402	0.034
cond_3	-0.353	0.157	-1.446	0.735
cond_ka	0.004	0.052	-0.262	0.542
lift_0	-0.090	0.037	-0.296	0.269
lift_ka	-0.106	0.046	-0.260	0.465
Lagecharakteristika				
log(DIST_UBAHN)	-0.007	0.038	-0.173	0.257
DIST_SBAHN	7.251e-05	4.623e-05	-8.223e-05	3.204e-04
DIST_BUSTR	-1.102e-04	2.820e-4	-0.001	0.001
log(DIST_WLB)	0.000167	0.829	-8.523	22.026
log(DIST_CBD)	-0.177	0.772	-20.723	7.635
DIST_HIGHW	-4.665e-06	3.605e-05	-1.528e-04	1.700e-04
Overall Model Fit				
$Adj.R^2$ Mean	0.397			

Bandwidth = 1532 Meter

Abbildung 4.14 zeigt die drei definierten Bürocluster. Die Zugehörigkeit zum Cluster findet jeweils als Dummycodierung statt (vgl. Tabelle 4.4).

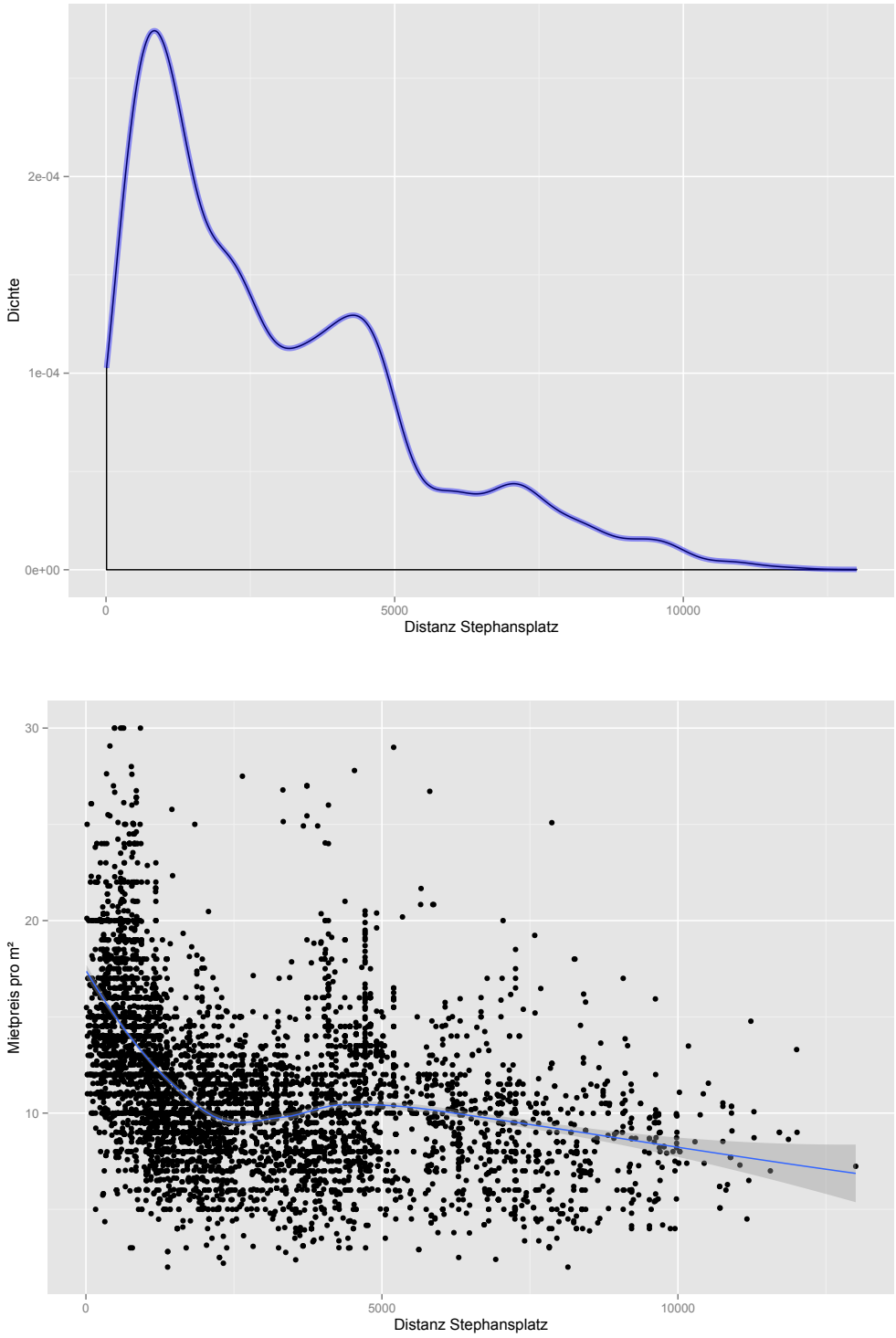

Abbildung 4.11: Dichte der Beobachtungen in Bezug zur Distanz zum Stephansplatz und Miete pro m^2 in Bezug zur Distanz zum Stephansplatz mit LOWESS Fit, eigene Darstellung.

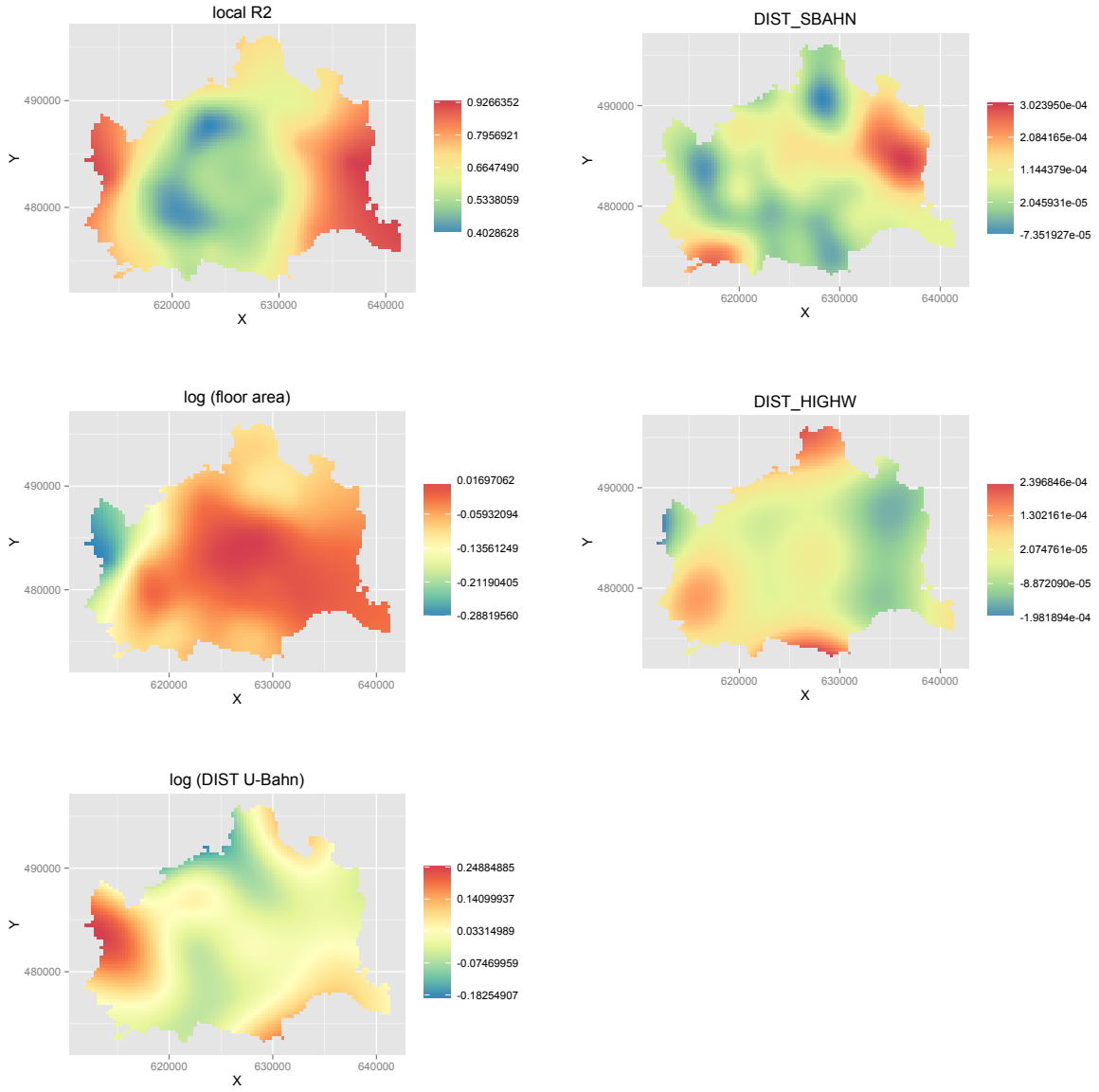

Abbildung 4.12: GWR Koeffizienten, v.l.n.r. Zeile 1: lokale R^2, Koeffizient Distanz zur S-Bahn; Zeile 2: Koeffizient logarithmierte Größe der Immobilie, Koeffizient Distanz zur Autobahnauffahrt; Zeile 3: Koeffizient logarithmierte Distanz zur U-Bahn, eigene Darstellung.

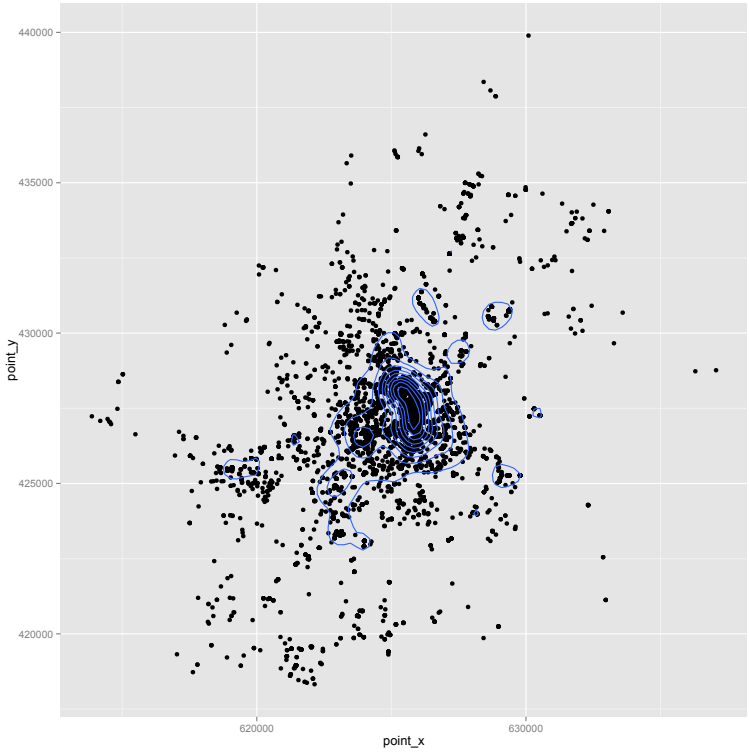

Abbildung 4.13: Density Scatter Plot: Dichte der Beobachtungen im Untersuchungsgebiet, eigene Darstellung.

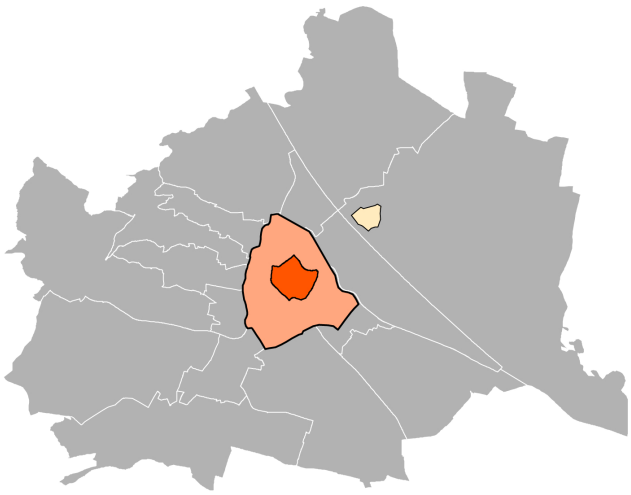

Abbildung 4.14: Ad-hoc Cluster: dunkelrot = Innere Stadt, rot = Innere Bezirke, gelb = Donau City, eigene Darstellung.

Kapitel 5

Modellbildung

5.1 Methodischer Ablauf

Die Abbildung 5.1 stellt den methodischen Ablauf der Analyse dar. Im für die Modellbildung herangezogenen Datensatz wurden Angebotsdaten, bauliche Charakteristika sowie Nachbarschafts- und Lagecharakteristika zusammengeführt. Die GWR Analyse wird im Zuge der Modellbildung als exploratives Tool verwendet und soll Hinweise und Hilfestellungen bei der Findung von räumlichen Clustern geben. Der Datensatz mit 6.236 Beobachtungen wird für die OLS und für die SAR Modelle in zwei Subsets aufgesplittet. Die randomisierte Auswahl an Fällen im Trainingsset beinhaltet 80 Prozent aller Beobachtungen. Mit diesem Datensatz werden die Modelle geschätzt. Die restlichen 20 Prozent der Beobachtungen im Validierungsset werden ausschließlich für die Bewertung der Modellgüte und den Vergleich der Modelle untereinander herangezogen. Für die Angebotsfälle im Validierungsset werden die Mietpreise pro m^2 geschätzt und der Root Mean Square Error (RMSE) errechnet. Der RMSE stellt die Wurzel der mittleren quadratischen Abweichung eines Schätzers vom tatsächlichen zu schätzenden Wert im Datenset dar. Um so kleiner dieser Wert, desto kleiner ist der Bias (Verzerrung bezogen auf den Erwartungswert) und die Varianz des Schätzers (Streuung der Schätzfunktion). Darüber hinaus werden die Modelle über das AIC verglichen. Je niedriger der Wert, desto besser ist der Fit des Modells.
Wie in Abschnitt 2 und 3 dargestellt, kann die Spezifikation der funktionalen Form einer hedonischen Regressionsgleichung vielfältig sein. In dieser Arbeit werden nur semi-log und log-log Spezifikationen verwendet, da hier die Möglichkeit einer einfachen, direkten ökonomischen Interpretation gegeben ist (Haase, 2011). Bei log-log Modellen ist die erklärende Variable y und eine oder mehrere abhängige Variablen logarithmiert. Die geschätzten Koeffizienten können als Elastizitäten interpretiert werden. Geht man von einem einfachen Modell mit folgender Spezifikation aus, $\log(y) = \beta_0 + \beta_1 \log(x) + \varepsilon$, so bedeutet eine Erhöhung um 1 Prozent in x eine Zunahme von y um β_1

Prozent. Bei semi-log Modellen (log-level model) geht nur die erklärende Variable y logarithmiert ins Modell ein. Die geschätzten Koeffizienten können als Semi-Elastizitäten interpretiert werden. Geht man von einem einfachen Modell mit folgender Spezifikation aus, $\log(y) = \beta_0 + \beta_1(x) + \varepsilon$, so bedeutet eine Erhöhung von x um eine Einheit eine Zunahme von y um $100 \times \beta_1$ Prozent.

Alle Modelle werden mit Hilfe der Statistik-Software R [1] geschätzt.

5.2 Geschätzte Modelle

Modell OLS

Die mittels OLS geschätzten Modelle weißen alle als endogene Variable den logarithmierten Mietpreis pro m^2 auf. Zur Variablen Selektion im Zuge der Regressionsanalyse wird in einigen Arbeiten eine *„stepwise selection“* durchgeführt (vgl. Lehner, 2011; Haase, 2011). Die Statistik-Software R ermöglicht dies ebenfalls, wobei diese Methode für die Arbeit nicht angewandt wurde. Die OLS geschätzten Modelle werden schrittweise erweitert, und auf Kollinearität, Heteroskedastizität und räumliche Autokorrelation getestet. Das Modell OLS_1 schließt nur bauliche Charakteristika ein. Im Modell OLS_2 wird das Modell um die Lagecharakteristika erweitert. Die Nachbarschaftscharakteristika werden im Modell OLS_3 mit aufgenommen. Das OLS Modell mit dem besten Goodness-of-Fit wird für die Schätzung der SAR Modelle herangezogen.

Modell OLS_1

Im Modell OLS_1 werden nur die baulichen Charakteristika in die Regressionsgleichung mit aufgenommen. Die Ergebnisse werden in der Tabelle 5.1 dargestellt. Das geschätzte Modell ist mit einer F-Statistik von 51.14 und einem p-value von 2.2e-16 bei einem Signifikanzniveau von 0.05 statistisch signifikant. Das $adj.R^2$ liegt bei 0.108. Es können etwa 11 Prozent der Variation in y durch das Modell erklärt werden. Das Akaike Information Criterion (AIC) nimmt einen Wert von 3671.552 an. Mit Hilfe der geschätzten Koeffizienten wurde der logarithmierte Mietpreis für die Beobachtungen im Validierungsset ermittelt und der Root Mean Square Error (RMSE) kalkuliert. Dieser liegt bei 0.356.

Die logarithmierte Größe der Immobilie ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Das positive Vorzeichen des geschätzten Koeffizient weißt auf einen positiven Zusammenhang zwischen Mietpreis und Größe der Immobilie hin. Erhöht sich die Größe der Immobilie um 1 Prozent, so steigt der Mietpreis um 0.047 Prozent. Die für die Bauperioden definierten

[1] http://www.r-project.org/

Tabelle 5.1: Ergebnisse Modell OLS_1

Endogene Variable:	*log(rent_sqm)*				N = 4989
Exogene Variable	Coeff.	SE	t-value	Sign.	VIF
Intercept	2.294	0.036	63.434	***	
Bauliche Char.					
BP1	0.210	0.047	4.476	***	1.106
BP2	-0.044	0.020	-2.152	*	2.351
BP3	-0.176	0.059	-2.982	**	1.073
BP4	-0.106	0.026	-4.016	***	1.385
BP5	-0.231	0.021	-10.884	***	1.655
BP_KA	-0.071	0.016	-4.371	***	2.679
log(floor_area)	0.047	0.005	8.618	***	1.136
cond_2	-0.180	0.020	-8.979	***	1.205
cond_3	-0.432	0.051	-8.490	***	1.030
cond_ka	-0.007	0.018	-0.387		3.195
lift_0	-0.168	0.015	-11.169	***	1.609
lift_ka	-0.079	0.019	-4.090	***	3.815
Overall Model Fit					
$Adj.R^2$	0.108				
AIC	3671.552				
$RMSE$	0.356				
$Moran's\ I$	0.138				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Dummy-Variablen sind alle auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die einzelnen Kategorien für die Bauperioden zeigen wie erwartet einen negativen Zusammenhang zwischen Mietpreis und Gebäudealter bezogen auf die Referenzkategorie *(BP_6)*: 2001 bis Heute. Ausnahme bildet die Kategorie *(BP_1)*: vor 1849 Hochbarock, Rokoko, Klassizismus und Biedermeier, hier wird ein positiver Zusammenhang geschätzt. Für Büroimmobilien ohne Lift wird ein Abschlag von 15 Prozent bezogen auf Büroimmobilien mit Lift durch das Modell geschätzt. Die Dummy Variable ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die definierten Zustandskategorien sind auf einem Signifikanzniveau von 0.05 statistisch signifikant. Bezogen auf die Referenzkategorie sehr guter Zustand der Immobilie werden Abschläge für Immobilien mit einem guten sowie mit einem befriedigenden bis schlechten Zustand geschätzt. Für Büroimmobilien mit einem guten Zustand wird ein Abschlag von 16 Prozent bezogen auf die Referenzkategorie erwartet.

Im Zuge der Modelldiagnostik zeigt die grafische Überprüfung des QQ-Plots keine Verletzung der Annahme der Normalverteilung der Residuen. Leich-

te Abweichungen von der Normalverteilungsgerade sind zu erkennen, allerdings bei großen Stichproben zu vernachlässigen (vgl. Haase, 2011). Die Cooks Distance gibt keine Hinweise auf extreme Beobachtungen. Die grafischen Überprüfung des Residuenplots lässt die Vermutung zu, dass die Varianz der Fehlerterme nicht konstant ist und Heteroskedastizität vorliegt. Der Breusch-Pagan Test gegen Heteroskedastizität ist auf einem Signifikanzniveau von 0.05 signifikant, was mit dem Vorhandensein von räumlicher Abhängigkeit zusammenhängen könnte. Der Moran's I Test ist auf einem Signifikanzniveau von 0.05 signifikant. Die Nullhypothese, dass keine Autokorrelation in den Residuen besteht kann verworfen werden. Der Variance Inflation Factor (VIF) weißt nicht auf Kollinearität hin.

Modell OLS_2

Im Modell OLS_2 werden neben den baulichen Charakteristika auch die Lagecharakteristika in die Regressionsgleichung mit aufgenommen.

Das geschätzte Modell ist mit einer F-Statistik von 140.5 und einem p-value von 2.2e-16 bei einem Signifikanzniveau von 0.05 statistisch signifikant. Das $adj.R^2$ liegt bei 0.335. Es können etwa 34 Prozent der Variation in y durch das Modell erklärt werden. Das Akaike Information Criterion (AIC) nimmt einen Wert von 2239.285 an. Mit Hilfe der geschätzten Koeffizienten wurde der logarithmierte Mietpreis für die Beobachtungen im Validierungsset ermittelt und der Root Mean Square Error (RMSE) kalkuliert. Dieser liegt bei 0.304.

Die logarithmierte Größe der Immobilie ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Das positive Vorzeichen des geschätzten Koeffizient weißt auf einen positiven Zusammenhang zwischen Mietpreis und Größe der Immobilie hin. Erhöht sich die Größe der Immobilie um 1 Prozent, so steigt der Mietpreis um 0.048 Prozent. Die für die Bauperioden definierten Dummy-Variablen sind alle auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die einzelnen Kategorien für die Bauperioden zeigen wie erwartet einen negativen Zusammenhang zwischen Mietpreis und Gebäudealter bezogen auf die Referenzkategorie *(BP_6)*: 2001 bis Heute. Selbst die Kategorie *(BP_1)*: vor 1849 Hochbarock, Rokoko, Klassizismus und Biedermeier folgt diesem negativen Zusammenhang. Für Büroimmobilien ohne Lift wird ein Abschlag von 5,6 Prozent bezogen auf Büroimmobilien mit Lift durch das Modell geschätzt. Die Dummy Variable ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die definierten Zustandskategorien sind auf einem Signifikanzniveau von 0.05 statistisch signifikant. Bezogen auf die Referenzkategorie sehr guter Zustand der Immobilie werden Abschläge für Immobilien mit einem guten sowie mit einem befriedigenden bis schlechten Zustand geschätzt. Für Büroimmobilien mit einem guten Zustand wird ein Abschlag von 14 Prozent bezogen auf die Referenzkategorie erwartet.

Tabelle 5.2: Ergebnisse Modell OLS_2

Endogene Variable:	*log(rent_sqm)*				N = 4989
Exogene Variable	Coeff.	SE	t-value	Sign.	VIF
Intercept	3.769	0.057	65.672	***	
Bauliche Char.					
BP1	-0.130	0.043	-3.038	**	1.181
BP2	-0.237	0.0183	-12.938	***	2.581
BP3	-0.289	0.050	-5.767	***	1.090
BP4	-0.262	0.024	-10.934	***	1.437
BP5	-0.198	0.019	-10.626	***	1.679
BP_KA	-0.197	0.015	-13.508	***	2.893
log(floor_area)	0.048	0.005	10.194	***	1.157
cond_2	-0.153	0.018	-8.616	***	1.222
cond_3	-0.371	0.044	-8.494	***	1.035
cond_ka	-0.017	0.016	-1.043		3.250
lift_0	-0.058	0.014	-4.232	***	1.733
lift_ka	-0.044	0.017	-2.584	**	3.941
Lagecharakteristika					
log(DIST_UBAHN)	-0.021	0.006	-3.728	***	1.342
DIST_SBAHN	0.0001	8.204e-06	10.625	***	1.047
DIST_BUSTR	-0.0001	0.0001	-1.557		1.088
log(DIST_WLB)	0.015	0.005	3.091	**	1.337
log(DIST_CBD)	-0.189	0.006	-33.193	***	1.792
DIST_HIGHW	-1.381e-05	3.720e-06	-3.713	***	1.197
Overall Model Fit					
$Adj.R^2$	0.335				
AIC	2239.285				
$RMSE$	0.304				
$Moran's\ I$	0.002				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Von den 6 Lageparameter sind 5 auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die Distanz zur nächsten Bus- oder Tramhaltestelle kann im Modell nicht signifikant geschätzt werden. Verringert sich die Distanz zur nächsten U-Bahnstation um 1 Prozent, so steigt der Mietpreis um 0.021 Prozent. Der Zusammenhang zwischen der Distanz zur nächsten S-Bahnstation und dem Mietpreis wird im Modell als positiver Zusammenhang geschätzt, wobei die marginalen Effekte sehr klein sind. Ein ähnliches Bild zeigen die geschätzten Koeffizienten für die Distanz zur nächsten Station der Wiener-Lokalbahn, auch hier prognostiziert das Modell einen positiven Zusammenhang. Den Erwartungen entsprechen die geschätzten Koeffizienten für die Di-

stanz zur nächsten Autobahnauffahrt und der Distanz zum CBD. Verringert sich die Distanz zum CBD um 1 Prozent, so steigt der Mietpreis um 0.189 Prozent.

Im Zuge der Modelldiagnostik zeigt die grafische Überprüfung des QQ-Plots keine Verletzung der Annahme der Normalverteilung der Residuen. Leichte Abweichungen von der Normalverteilungsgerade sind zu erkennen, allerdings bei großen Stichproben zu vernachlässigen (vgl. Haase, 2011). Die Cooks Distance gibt keine Hinweise auf extreme Beobachtungen. Die grafische Überprüfung des Residuenplots lässt in diesem Fall die Vermutung zu, dass die Varianz der Fehlerterme konstant ist und somit keine Heteroskedastizität vorliegt. Der Breusch-Pagan Test gegen Heteroskedastizität ist allerdings auch hier auf einem Signifikanzniveau von 0.05 signifikant, was wieder mit dem Vorhandensein von räumlicher Abhängigkeit zusammenhängen könnte. Der Moran's I Test ist auf einem Signifikanzniveau von 0.05 signifikant. Die Nullhypothese, dass keine Autokorrelation in den Residuen besteht kann verworfen werden. Der Variance Inflation Factor (VIF) weißt nicht auf Kollinearität hin.

Modell OLS_3

Im Modell OLS_3 werden neben den baulichen Charakteristika und den Lagecharakteristika auch Nachbarschaftscharakteristika in die Regressionsgleichung mit aufgenommen. Das geschätzte Modell ist mit einer F-Statistik von 136.2 und einem p-value von 2.2e-16 bei einem Signifikanzniveau von 0.05 statistisch signifikant. Das $adj.R^2$ liegt bei 0.374. Es können etwa 37 Prozent der Variation in y durch das Modell erklärt werden. Das Akaike Information Criterion (AIC) nimmt einen Wert von 1943.309 an. Mit Hilfe der geschätzten Koeffizienten wurde der logarithmierte Mietpreis für die Beobachtungen im Validierungsset ermittelt und der Root Mean Square Error (RMSE) kalkuliert. Dieser liegt bei 0.294.

Die logarithmierte Größe der Immobilie ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Das positive Vorzeichen des geschätzten Koeffizient weißt auf einen positiven Zusammenhang zwischen Mietpreis und Größe der Immobilie hin. Erhöht sich die Größe der Immobilie um 1 Prozent, so steigt der Mietpreis um 0.046 Prozent. Die für die Bauperioden definierten Dummy-Variablen sind alle auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die einzelnen Kategorien für die Bauperioden zeigen wie erwartet einen negativen Zusammenhang zwischen Mietpreis und Gebäudealter bezogen auf die Referenzkategorie *(BP_6)*: 2001 bis Heute. Selbst die Kategorie *(BP_1)*: vor 1849 Hochbarock, Rokoko, Klassizismus und Biedermeier folgt diesem negativen Zusammenhang. Für Büroimmobilien ohne Lift wird ein Abschlag von 4,5 Prozent bezogen auf Büroimmobilien mit Lift durch das Modell geschätzt. Die Dummy Variable ist auf einem Signifikanzniveau von 0.05 statistisch signifi-

Tabelle 5.3: Ergebnisse Modell OLS_3

Endogene Variable:	*log(rent_sqm)*				N = 4989
Exogene Variable	Coeff.	SE	t-value	Sign.	VIF
Intercept	2.727	0.103	26.394	***	
Bauliche Char.					
BP1	-0.138	0.042	-3.331	***	1.189
BP2	-0.235	0.018	-13.070	***	2.638
BP3	-0.299	0.049	-6.127	***	1.095
BP4	-0.245	0.023	-10.455	***	1.457
BP5	-0.178	0.018	-9.690	***	1.730
BP_KA	-0.180	0.014	-12.530	***	2.981
log(floor_area)	0.046	0.005	10.032	***	1.160
cond_2	-0.157	0.017	-8.918	***	1.222
cond_3	-0.348	0.043	-8.203	***	1.038
cond_ka	-0.033	0.016	-2.092	*	3.264
lift_0	-0.046	0.013	-3.478	***	1.746
lift_ka	-0.029	0.017	-1.752	.	3.958
Nachbarschaftschar.					
DIST_CLUST	-8.192e-06	4.670e-06	-1.754	.	2.984
city	0.373	0.026	14.272	***	7.159
bezinnen	0.082	0.015	5.390	***	2.856
dc	0.273	0.036	7.621	***	1.652
Lagecharakteristika					
log(DIST_UBAHN)	-0.019	0.006	-3.472	***	1.405
DIST_SBAHN	1.817e-05	9.655e-06	1.882	.	1.540
DIST_BUSTR	-2.031e-05	5.267e-05	-0.386		1.108
log(DIST_WLB)	0.007	0.005	1.454		1.461
log(DIST_CBD)	-0.055	0.012	-4.718	***	7.850
DIST_HIGHW	-8.546e-06	3.845e-06	-2.222	*	1.358
Overall Model Fit					
$Adj.R^2$	0.374				
AIC	1943.309				
$RMSE$	0.294				
$Moran's\ I$	0.017				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

kant. Die definierten Zustandskategorien sind auf einem Signifikanzniveau von 0.05 statistisch signifikant. Bezogen auf die Referenzkategorie sehr guter Zustand der Immobilie werden Abschläge für Immobilien mit einem guten sowie mit einem befriedigenden bis schlechten Zustand geschätzt. Für Büroimmobilien mit einem guten Zustand wird ein Abschlag von 14,5 Prozent bezogen

auf die Referenzkategorie erwartet.

Von den 6 Lageparameter sind nur mehr 3 auf einem Signifikanzniveau von 0.05 statistisch signifikant. Für die Distanz zur nächsten Bus- oder Tramhaltestelle, für die Distanz zur nächsten S-Bahnstation sowie für die Distanz zur nächsten Station der Wiener Lokalbahn schätzt das Modell keinen signifikanten Einfluss auf den Mietpreis. Verringert sich die Distanz zur nächsten U-Bahnstation um 1 Prozent, so steigt der Mietpreis um 0.019 Prozent. Den Erwartungen entsprechen die geschätzten Koeffizienten für die Distanz zur nächsten Autobahnauffahrt und der Distanz zum CBD. Der Zusammenhang im Modell ist negativ. Verringert sich die Distanz zum CBD um 1 Prozent, so steigt der Mietpreis um 0.055 Prozent.

Die Clusterparameter sind auf einem Signifikanzniveau von 0.05 statistisch signifikant. Für die Entfernung zum nächsten Cluster wird ein negativer Zusammenhang zum Mietpreis im Modell geschätzt, allerdings liegt der p-value über 0.05. Büroimmobilien in der City erfahren einen Zuschlag auf den Mietpreis bezogen auf Immobilien ohne Zugehörigkeit zu einem Cluster von etwa 45 Prozent. Das Modell prognostiziert für Büroimmobilien in den Inneren Bezirken eine Prämie von etwa 8,5 Porzent bezogen auf Immobilien ohne Zugehörigkeit zu einem Cluster. Ist die Büroimmobilie teil des Büroclusters Donau City, so wird ein marginaler Effekt auf den Mietpreis von etwa 31 Prozent bezogen auf Immobilien ohne Zugehörigkeit zu einem Cluster geschätzt.

Im Zuge der Modelldiagnostik zeigt die grafische Überprüfung des QQ-Plots keine Verletzung der Annahme der Normalverteilung der Residuen. Leichte Abweichungen von der Normalverteilungsgerade sind zu erkennen, allerdings bei großen Stichproben zu vernachlässigen (vgl. Haase, 2011). Die Cooks Distance gibt keine Hinweise auf extreme Beobachtungen. Die grafische Überprüfung des Residuenplots lässt in diesem Fall die Vermutung zu, dass die Varianz der Fehlerterme konstant ist und somit keine Heteroskedastizität vorliegt. Der Breusch-Pagan Test gegen Heteroskedastizität ist allerdings auch hier auf einem Signifikanzniveau von 0.05 signifikant, was wieder mit dem Vorhandensein von räumlicher Abhängigkeit zusammenhängen könnte. Der Moran's I Test ist auf einem Signifikanzniveau von 0.05 signifikant. Die Nullhypothese, dass keine Autokorrelation in den Residuen besteht kann verworfen werden. Der Variance Inflation Factor (VIF) weißt nicht auf Kollinearität hin.

Modell SAR

Das OLS_3 Modell zeigt den kleinsten Root Mean Square Error (RMSE) und den kleinsten AIC und wird für die Schätzung der räumlichen Modelle herangezogen. Der Moran's I für die abhängige Variable wurde berechnet. Dieser ist auf einem Signifikanzniveau von 0.05 statistisch signifikant (vgl. Tabelle

5.4). Die Nullhypothese, dass keine Autokorrelation in den Mietpreisen besteht kann verworfen werden.

Tabelle 5.4: Moran's I Test - Abhängige Variable Mietpreis

Moran's I Statistik	Expectations	Variance	p-value
0.1657	-0.0002	1.209388e-06	$<$ 2.2e-16

Um die Art der räumlichen Autokorrelation zu eruieren und die Wahl des SAR Modells zu unterstützen empfiehlt es sich Lagrange Multiplier Tests (LM-Tests) durchzuführen (vgl. Chalermpong und Wattana, 2009; Lehner, 2011; Caduff, 2013; Choumert et al., 2014). In Anselin (1988a) findet sich eine ausführliche Abhandlung dieser Tests für das Anwendungsfeld der räumlichen Statistik. Lagrange Multiplier Tests geben einen Hinweis darauf, ob der datengenerierende Prozess einem spatial error oder einem spatial lag SAR folgt (vgl. Bivand et al., 2008). Die Wahl des Models wird aufgrund des LM-Tests mit der höchsten Signifikanz gefällt. Beide Lagrange Mulitplier Tests, LMerr und LMlag, sind für das OLS_3 Modell auf einem Signifikanzniveau von 0.05 statistisch signifikant. In diesem Fall werden robuste LM-Tests (RLM-Tests) ausgeführt. Die Teststatistik für den RLMerr Test zeigt eine höhere Signifikanz, was auf einen datengenerierenden Prozess hinweist, welcher am wahrscheinlichsten durch ein SAR_{err} Modell beschrieben werden kann. Wie in Abschnitt 3.2 dieser Arbeit bereits dargestellt, kann dies durch die Fehlspezifikation des OLS Modells bzw. durch einen omitted variable bias motiviert sein.

In weiterer Folge werden SAR_{err} und SAR_{lag} Modelle geschätzt. Es wird hinsichtlich der räumlichen Nachbarschaftsbeziehungen und der Definition der Nachbarn unterschieden. Zum einen werden die Nachbarschaftsbeziehungen in den Modellen durch eine Distanzgewichtung modelliert. Zum anderen werden k-nearest-neighbour Beziehungen in der Nachbarschaftsmatrix getestet. Die Modelle werden mit 8, 10, 12 und 14 k-nearest neighbours geschätzt, wobei die besten Ergebnisse mit knn=14 Nachbarn erzielt werden. Der im Abschnitt 4 deskriptiv analysierte Datensatz enthält eine Vielzahl von räumlichen Duplikaten, was bei der Modellierung über k-nearest-neighbour Nachbarschaftsbeziehungen zu Problemen führen kann. Aus diesem Grund werden die geografischen Positionen der Datenpunkte (lat,lon) *„gejittert“* (vgl. Stahel und Maechler, 2015). Die Vektoren mit den latitude und longitude Koordinaten beinhalten die räumliche Information für jeden Datenpunkt. Diesen Vektoren wird ein Noise zugegeben, welcher die Datenpunkte im Umkreis von 50m zur ursprünglichen Position zufällig verteilt. Es wird vermutet, dass dadurch eine gewisse Verzerrung ins Modell eingeführt wird, welche aber keine entscheiden-

den Einschränkungen für die Schätzung mit sich bringt. Diese Prämisse könnte dadurch geprüft werden, indem das Jittering der Daten mit unterschiedliche Trainingssets wiederholt wird und die geschätzten Koeffizienten jeweils verglichen werden. Dies wurde allerdings im Zuge der Arbeit auf Grund der rechen- und zeitintensiven Vorgangsweise nicht durchgeführt.

Modell SAR_{err} und SAR_{lag}

Eine direkte Interpretation der geschätzten Koeffizienten wie im *OLS* Modell ist auf Grund der Aufnahme des Spatial Lags ins Modell nicht mehr möglich (vgl. Dreger et al., 2011).

Die Tabelle 5.5 stellt die Ergebnisse für die SAR_{err} Modelle dar. Das Modell mit dem besten Modell-Fit verwendet eine k-nearest-neighbour Gewichtungsmatrix mit 14 Nachbarn. Es wird ein AIC von 1175.3 erreicht. Das SAR_{err} Modell mit Distanzgewichtung und einem Schwellenwert für die Distanz von 2.3 km erreicht einen geringeren Modell-Fit von 1876.1 für das AIC.

Für beide SAR_{err} Modelle ist der räumliche Autokorrelationskoeffizient Lambda (λ) auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die für die Bauperioden definierten Dummy-Variablen sind alle auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die einzelnen Kategorien für die Bauperioden zeigen wie erwartet einen negativen Zusammenhang zwischen Mietpreis und Gebäudealter bezogen auf die Referenzkategorie *(BP_6)*: 2001 bis Heute. Die logarithmierte Größe der Immobilie ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Das positive Vorzeichen des geschätzten Koeffizient weißt auf einen positiven Zusammenhang zwischen Mietpreis und Größe der Immobilie hin. Die Dummy-Variablen für die Zustandsklassen sind auf einem Signifikanzniveau von 0.05 statistisch signifikant. Für Büroimmobilien in der Zustandsklasse „guter Zustand“und „befriedigender bis schlechter Zustand“wurde ein Abschlag auf den Mietpreis bezogen auf Büroimmobilien mit der Zustandsklasse „sehr guter Zustand“geschätzt. Auch für das Nichtvorhandensein eines Liftes wird ein Abschlag auf den Mietpreis in beiden Modellen berechnet.

Bei den Nachbarschaftscharakteristika sind die Dummy-Variablen für die Bürocluster in beiden SAR_{err} Modellen auf einem Signifikanzniveau von 0.05 statistisch signifikant. Für Büroimmobilien in einem der drei Cluster wird ein Zuschlag auf den Mietpreis bezogen auf Immobilien ohne Cluster-Zugehörigkeit geschätzt. Die Distanz zum nächsten Cluster ist statistisch nicht signifikant.

Für das SAR_{err} Modell mit k-nearest neighbour Gewichtungsmatrix ist für die 6 Lagecharakteristika nur mehr die logarithmierte Distanz zum Stephansplatz auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die

logarithmierte Distanz zur nächsten U-Bahn und die Distanz zur nächsten S-Bahn ist nur im SAR_{err} Modell mit Distanzgewichtung auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die Vorzeichen weisen auf einen negativen Zusammenhang zwischen Distanz und Mietpreis hin.

Tabelle 5.5: Ergebnisse Modell SAR_{err} mit Distanzgewichtung und k-nearest-neighbour (knn=14)

Endogene Variable:	*log(rent_sqm)*			N = 4989
	SAR_{err} (Dist.=)		SAR_{err} (knn=14)	
Exogene Variable	Coeff.	Sign.	Coeff.	Sign.
Intercept	2.7225	***	2.872	***
Lambda(λ)	0.702	***	0.631	***
Bauliche Char.				
BP1	-0.136	***	-0.168	***
BP2	-0.230	***	-0.255	***
BP3	-0.300	***	-0.319	***
BP4	-0.240	***	-0.247	***
BP5	-0.160	***	-0.166	***
BP_KA	-0.170	***	-0.179	***
log(floor_area)	0.047	***	0.039	***
cond_2	-0.154	***	-0.127	***
cond_3	-0.351	***	-0.307	***
cond_ka	-0.026	.	-0.018	
lift_0	-0.050	***	-0.050	***
lift_ka	-0.035	*	-0.048	*
Nachbarschaftschar.				
DIST_CLUST	-5.708e-06		3.892e-06	
city	0.360	***	0.317	***
bezinnen	0.077	***	0.095	**
dc	0.278	***	0.315	***
Lagecharakteristika				
log(DIST_UBAHN)	-0.012	*	-0.006	
DIST_SBAHN	3.556e-05	**	2.413e-05	
DIST_BUSTR	1.298e-06		-9.627e-05	
log(DIST_WLB)	-0.012	.	0.006	
log(DIST_CBD)	-0.048	***	-0.082	***
DIST_HIGHW	-4.745e-06		2.644e-06	
Overall Model Fit				
AIC	1876.1		1175.3	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Die Tabelle 5.6 stellt die Ergebnisse für die SAR_{lag} Modelle dar. Das Modell mit dem besten Modell-Fit verwendet eine k-nearest-neighbour Gewichtungsmatrix mit 14 Nachbarn. Es wird ein AIC von 1189.0 erreicht. Das SAR_{lag} Modell mit Distanzgewichtung und einem Schwellenwert für die Distanz von 2.3 km erreicht einen geringeren Modell-Fit von 1903.6 für das AIC.

Für die geschätzten Koeffizienten der SAR_{lag} Modelle zeigt sich ein sehr ähnliches Bild wie bei den Koeffizienten der SAR_{err} Modellen. Für beide SAR_{lag} Modelle ist der räumliche Autokorrelationskoeffizient Rho (ρ) auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die für die Bauperioden definierten Dummy-Variablen sind alle auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die einzelnen Kategorien für die Bauperioden zeigen wie erwartet einen negativen Zusammenhang zwischen Mietpreis und Gebäudealter bezogen auf die Referenzkategorie *(BP_6)*: 2001 bis Heute. Die logarithmierte Größe der Immobilie ist auf einem Signifikanzniveau von 0.05 statistisch signifikant. Das positive Vorzeichen des geschätzten Koeffizient weißt auf einen positiven Zusammenhang zwischen Mietpreis und Größe der Immobilie hin. Die Dummy-Variablen für die Zustandsklassen sind auf einem Signifikanzniveau von 0.05 statistisch signifikant. Für Büroimmobilien in der Zustandsklasse „guter Zustand“und „befriedigender bis schlechter Zustand“wurde ein Abschlag auf den Mietpreis bezogen auf Büroimmobilien mit der Zustandsklasse „sehr guter Zustand“geschätzt. Auch für das Nichtvorhandensein eines Liftes wird ein Abschlag auf den Mietpreis in beiden Modellen berechnet.

Bei den Nachbarschaftscharakteristika sind Unterschiede zu den SAR_{err} Modellen erkennbar. Im Modell mit Distanzgewichtung ist die Dummy-Variable für den Cluster der inneren Bezirke statistisch nicht signifikant. Bei der Verwendung einer k-nearest-neighbour Gewichtungsmatrix sind die Dummy-Variablen für die Bürocluster in beiden SAR_{err} Modellen auf einem Signifikanzniveau von 0.05 wieder statistisch signifikant.

Für das SAR_{lag} Modell mit k-nearest neighbour Gewichtungsmatrix ist für die 6 Lagecharakteristika nur mehr die logarithmierte Distanz zum Stephansplatz auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die logarithmierte Distanz zur nächsten U-Bahn ist nur im SAR_{lag} Modell mit Distanzgewichtung auf einem Signifikanzniveau von 0.05 statistisch signifikant. Die Vorzeichen weisen auf einen negativen Zusammenhang zwischen Distanz und Mietpreis hin.

Der Breusch-Pagan Test gegen Heteroskedastizität ist sowohl für die SAR_{err} Modelle als auch für die SAR_{lag} Modelle auf einem Signifikanzniveau von 0.05 signifikant.

Tabelle 5.6: Ergebnisse Modell SAR_{lag} mit Distanzgewichtung und k-nearest-neighbour (knn=14)

Endogene Variable:	*log(rent_sqm)*			N = 4989
	SAR_{lag} (Dist.)		SAR_{lag} (knn=14)	
Exogene Variable	Coeff.	Sign.	Coeff.	Sign.
Intercept	1.660	***	1.032	***
Rho(ρ)	0.400	***	0.594	***
Bauliche Char.				
BP1	-0.145	***	-0.097	**
BP2	-0.245	***	-0.176	***
BP3	-0.307	***	-0.236	***
BP4	-0.250	***	-0.163	***
BP5	-0.178	***	-0.104	***
BP_KA	-0.185	***	-0.112	***
log(floor_area)	0.046	***	0.036	***
cond_2	-0.154	***	-0.139	***
cond_3	-0.351	***	-0.312	***
cond_ka	-0.030	.	-0.021	
lift_0	-0.047	***	-0.043	***
lift_ka	-0.032	.	-0.043	**
Nachbarschaftschar.				
DIST_CLUST	5.175e-06		2.259e-06	
city	0.326	***	0.180	***
bezinnen	0.024		0.051	**
dc	0.236	***	0.103	**
Lagecharakteristika				
log(DIST_UBAHN)	-0.014	*	-0.005	
DIST_SBAHN	1.721e-05	.	6.502e-06	
DIST_BUSTR	-1.218e-05		-5.284e-05	
log(DIST_WLB)	-0.003		0.005	
log(DIST_CBD)	-0.003	**	-0.022	*
DIST_HIGHW	-3.311e-07		3.458e-06	
Overall Model Fit				
AIC	1903.6		1189.0	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.3 Modellvergleich

Die Tabelle 5.7 stellt die Ergebnisse der besten Modelle nochmal zusammenfassend dar. Das SAR_{err} Modell mit k-nearest neighbour Gewichtungsmatrix und 14 Nachbarn erreicht das kleinste AIC und ist somit das Modell mit dem besten Modell-Fit.

Die Richtung des Zusammenhanges zwischen der Response Variable Mietpreis und den unabhängigen Variablen ist, insofern diese auf einem Signifikanzniveau von 0.05 statistisch signifikant sind, für alle drei Modelle gleich. Es sind Parallelen im Vergleich zur untersuchten Literatur in Abschnitt 2.3 erkennbar. Im Hinblick auf die gemachten Hypothesen wurden folgende Erkenntnisse gemacht:

- **Das Gebäudealter und der Mietpreis weisen einen negativen Zusammenhang auf. Dies ist vor allem durch die hohe Sensibilität gegenüber Betriebskosten und durch den Anspruch an moderne Raumkonzepte zu begründen.**
  Für drei Modelle kann diese Hypothese bestätigt werden. Die definierten Dummy-Variablen für die Bauperiode weisen ein negatives Vorzeichen auf. Im Vergleich zur Referenzkategorie *(BP_6)*: 2001 bis Heute sind Abschläge auf den Mietpreis zu erwarten.

- **Mit der Größe der Immobilie, die zur Vermietung kommt, sinkt der Mietpreis pro m^2. Hier ist von einem einem Art Skalleneffekt auszugehen, wobei größere Flächen proportional einen geringeren Quadratmeterpreis verzeichnen als kleinere Flächen.**
  Diese Hypothese muss für alle drei Modelle verworfen werden. Die Größe der Immobilie und der Mietpreis pro m^2 weisen einen positiven Zusammenhang auf. Steigt die Mietfläche, so steigt auch der Mietpreis pro m^2. Dieses Ergebnis widerspricht dem Großteil der im Zuge der Literaturanalyse untersuchten Studien (Özyurt, 2014; Caduff, 2013; Haase, 2011).

- **Das Vorhandensein einer Förderanlage wirkt sich positiv auf den Mietpreis aus.**
  Diese Hypothese kann bestätigt werden. In allen drei Modellen ist die Dummy-Variable für das Nichtvorhandensein eines Liftes statistisch signifikant (0.05) und weist ein negatives Vorzeichen auf. Büroimmobilien mit einer Förderanlage erfahren einen Zuschlag auf den Mietpreis.

- **Der Zustand der Immobilie steht in einem positiven Zusammenhang mit dem Mietpreis. Je höher die Zustandsklasse, je besser der Zustand und desto höher ist der Mietpreis.**
  Diese Hypothese kann bestätigt werden. Für Büroimmoblien mit einem sehr guten Zustand wird in allen drei Modellen ein Zuschlag auf den Mietpreis geschätzt.

- **Die Zugehörigkeit zu einem Bürocluster wird durch einen Zuschlag im Mietpreis honoriert.**
  Diese Hypothese kann bestätigt werden. Die definierten Dummy-Variablen

für die drei Ad-hoc Cluster sind in allen Modellen auf einem Signifikanzniveau von 0.05 statistisch signifikant. Büroimmobilien in der inneren Stadt, den inneren Bezirken und in der Donau City sind auf Grund ihrer Lage und der Zugehörigkeit zu einem Cluster mit einem Mietpreisaufschlag zu honorieren.

Tabelle 5.7: Ergebnisse der besten Modelle

Endogene Variable:	*log(rent_sqm)*					N = 4989
	OLS_3		SAR_{lag} *knn=14*		SAR_{err} *knn=14*	
Exogene Variable	Coeff.	Sign.	Coeff.	Sign.	Coeff.	Sign.
Intercept	2.727	***	1.032	***	2.872	***
Rho/Lambda(ρ/λ)	-	-	0.594	***	0.631	***
Bauliche Char.						
BP1	-0.138	***	-0.097	**	-0.168	***
BP2	-0.235	***	-0.176	***	-0.255	***
BP3	-0.299	***	-0.236	***	-0.319	***
BP4	-0.245	***	-0.163	***	-0.247	***
BP5	-0.178	***	-0.104	***	-0.166	***
BP_KA	-0.180	***	-0.112	***	-0.179	***
log(floor_area)	0.046	***	0.036	***	0.039	***
cond_2	-0.157	***	-0.139	***	-0.127	***
cond_3	-0.348	***	-0.312	***	-0.307	***
cond_ka	-0.033	*	-0.021		-0.018	
lift_0	-0.046	***	-0.043	***	-0.050	***
lift_ka	-0.029	.	-0.043	**	-0.048	*
Nachbarschaftschar.						
DIST_CLUST	-8.192e-06	.	2.259e-06		3.892e-06	
city	0.373	***	0.180	***	0.317	***
bezinnen	0.082	***	0.051	**	0.095	**
dc	0.273	***	0.103	**	0.315	***
Lagecharakteristika						
log(DIST_UBAHN)	-0.019	***	-0.005		-0.006	
DIST_SBAHN	1.817e-05	.	6.502e-06		2.413e-05	
DIST_BUSTR	-2.031e-05		-5.284e-05		-9.627e-05	
log(DIST_WLB)	0.007		0.005		0.006	
log(DIST_CBD)	-0.055	***	-0.022	*	-0.082	***
DIST_HIGHW	-8.546e-06	*	3.458e-06		2.644e-06	
Overall Model Fit						
AIC	1943.3		1189.0		1175.3	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- **Der Zusammenhang zwischen der Distanz zum nächsten Bürocluster und dem Mietpreis ist negativ.**
  Diese Hypothese muss für alle drei Modelle verworfen werden. Die Distanz zum nächsten Cluster ist in keinem der geschätzten Modelle statistisch signifikant.

- **Der Zusammenhang zwischen der Distanz zur nächsten Haltestelle des öffentlichen Verkehrs und dem Mietpreis ist negativ.**
  Diese Hypothese kann nur teilweise angenommen bzw. verworfen werden. Die logarithmierte Distanz zur nächsten U-Bahnstation ist im OLS_3 Modell auf einem Signifikanzniveau von 0.05 statistisch signifikant. In den Spatial Lag Modellen können die Lagecharakteristika, welche die Nähe zu Haltestellen des öffentlichen Verkehrs abbilden sollen nicht mehr signifikant geschätzt werden.

- **Der Zusammenhang zwischen der Distanz zum ersten Bezirk und dem Mietpreis ist negativ.**
  Diese Hypothese kann bestätigt werden. Die logarithmierte Distanz zum Stephansdom (CBD) ist in allen Modellen statistisch signifikant. Der Koeffizient weißt in allen Modellen ein negatives Vorzeichen auf und weißt somit auf einen negativen Zusammenhang zwischen Response Variable und unabhängiger Variable hin. Je weiter die Büroimmobilie vom CBD entfernt liegt, desto geringer wird der Mietpreis geschätzt.

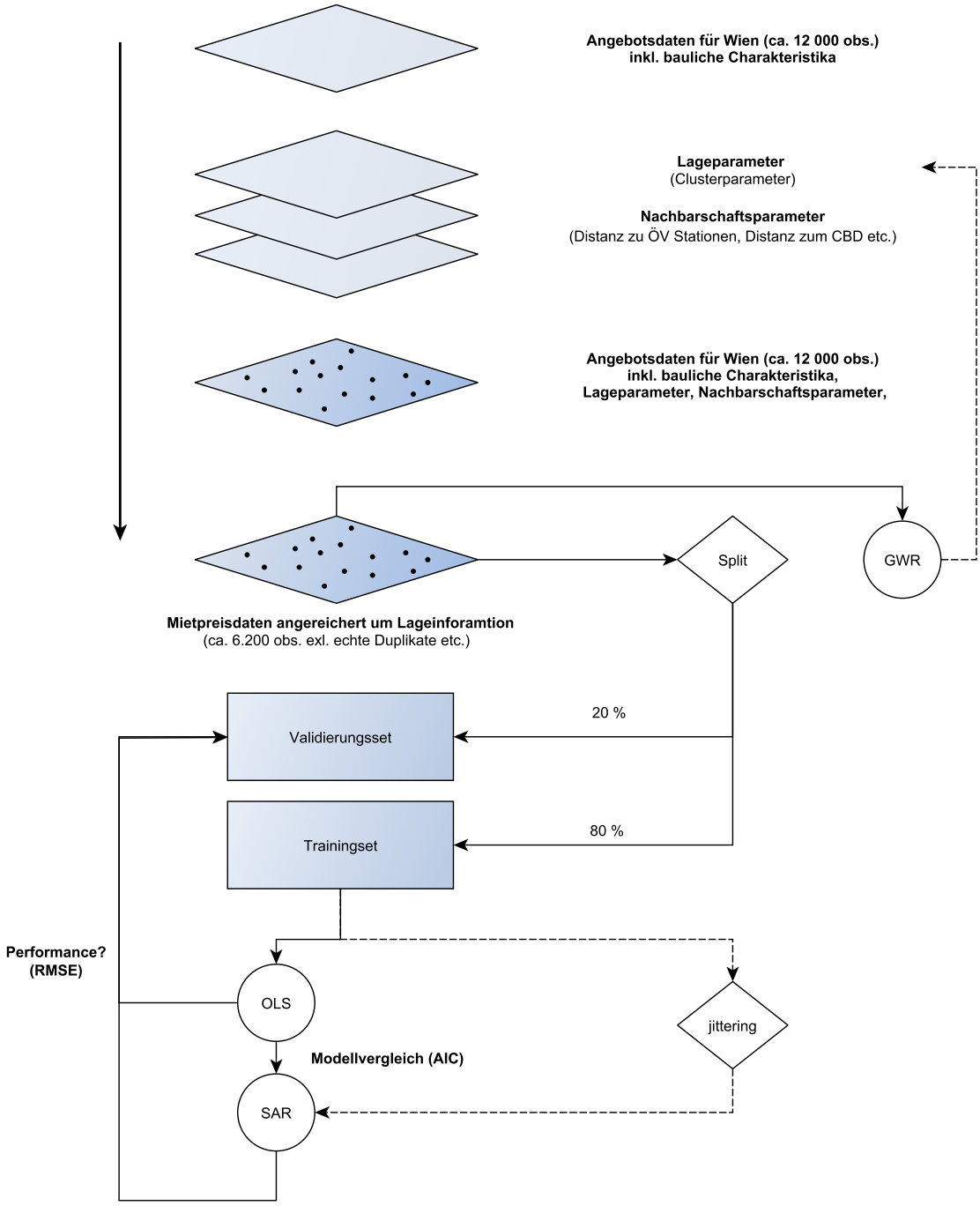

Abbildung 5.1: Darstellung des methodischen Ablaufs der Datenanalyse, eigene Darstellung.

Kapitel 6

Schlussbetrachtung und Ausblick

Um den unbekannten räumlichen Prozess, welcher zur geografischen Variation von Büromieten in Wien führt, besser zu verstehen, wurden hedonische Preismodelle mit dem logarithmierten Mietpreis als Response Variable geschätzt. Hedonische Preismodelle sind in der Lage, die marginalen Preiseffekte von unterschiedlichen Struktur-, Lage- und Nachbarschaftscharakteristika auf die Miete von Büroimmobilien zu bewerten und sind somit ein effizienter Ansatz zur Erklärung des Mietpreises auf Grund der Eigenschaften einer Immobilie. Eine Besonderheit die sich bei der Analyse von Immobiliendaten ergibt, ist, dass diese Daten nicht nur Träger von Attributen sind, sondern auch eine geografische Position und somit einen Raumbezug haben. Diese räumliche Information spielt bei der Analyse eine entscheidende Rolle. Die Eingangs dieser Arbeit gestellten Forschungsfragen können wie folgt beantwortet werden:

- **Welche Variablen können als Determinanten von Büromieten identifiziert werden?**
  Aus Sicht der baulichen Charakteristika wurden folgende Determinanten für den Mietpreis gefunden: Die Bauperiode und das damit implizierte Baualter einer Immobilie wirkt sich auf den Mietpreis negativ aus. Im Vergleich zu einer Büroimmobilie aus den Jahren 2001 bis Heute wird für ältere Immobilien im Durchschnitt ein geringerer Mietpreis lukriert (ceteris paribus). Der Zustand spielt, wie vermutet, eine entscheidende Rolle für den Mietpreis. Bezogen auf die Zustandskategorie „sehr guter Zustand“sind Abschläge für schlechtere Zustandskategorien zu erwarten. Das Vorhandensein einer Förderanlage führt zu einem Zuschlag auf den Mietpreis. Bei den untersuchten Nachbarschaftscharakteristika wird gezeigt, dass die Zugehörigkeit zu einem Bürocluster einen positiven Effekt auf den Mietpreis hat. Die Distanz zum CBD (Stephansdom) und vereinzelt auch die Distanz zum höherrangigen öffentlichen Verkehrs-

netz (U-Bahn und S-Bahn) wurden als Determinanten von Büromieten identifiziert.

- **Welche räumliche Variation weist der Einfluss von ausgewählten Struktur- und Lageparameter auf den Mietpreis im Untersuchungsgebiet auf?**
  Es konnte gezeigt werden, dass der Einfluss einzelner Parameter im Stadtgebiet varrieren kann und marginale Effekte auf den Mietpreis räumlich Heterogen sein können. Die im Zuge der GWR Analyse geschätzten Koeffizienten mit hohen Standardabweichungen weisen auf ein nicht-stationäres Verhalten hin.

- **Kann die Einführung von räumlichen Regressionsmodellen eine Verbesserung der Erklärungsgüte bewirken?**
  Durch die Anwendung von Spatial Lag Modellen kann der Modell-Fit im Vergleich zu den geschätzten OLS Modellen deutlich gesteigert werden. Dies bedeutet, das traditionelle hedonische Preismodelle, welche räumliche Strukturen nicht berücksichtigen, zu verzerrten Schätzern führen können.

Für zukünftige Forschungen scheint die Aufnahme von Landnutzungskategorien in die hedonischen Preismodelle sinnvoll. Dieser Ansatz wurde in Studien bereits getestet (vgl. zum Beispiel Farooq et al., 2010). Die generierten Entfernungsmaße zu den Haltestellen des öffentlichen Verkehrs und den Autobahnauffahrten waren im Modell nur teilweise statistisch signifikant. Hier sind weiter Bemühungen anzustellen, um die einfachen Entfernungsparameter durch komplexere Erreichbarkeitsmodelle im Individualverkehr und im öffentlichen Verkehr zu ersetzen und dadurch die Erklärungsgüte der Modelle weiter zu verbessern. Die Zugehörigkeit zu einem Bürocluster ist als Determinante des Mietpreises von Büroimmoblien zu sehen. Aus diesem Grund ist auf die Findung von räumlichen Clustern besonderes Augenmerk zu legen. Für die Segmentierung des Büromietenmarktes in Teilmärkte stehen diffizile Methoden zur Verfügung.

Wenn Unternehmen bei der Wahl ihres Standortes die Eigenschaften von Standort und Immobilie indirekt bewerten und sich diese Bewertung in den impliziten Preisen der Eigenschaften von Standort und Immobilie widerspiegeln, so bedeutet ein tieferes Verständnis dieser Preisdeterminanten, ein tieferes Verständnis der Standortwahl von Unternehmen, was eine wichtige Grundlage für die räumliche Planung darstellt. Nach Scholl (vgl. 2013, S.24) *„geht es in der Raumplanung und Raumentwicklung nicht vordringlich um die Beschreibung von Verhältnissen und Situationen, sondern darum raumbedeutsame Probleme und Konflikte zu klären und zu lösen sowie Vorstellungen über die wünschbare zukünftige Entwicklung zu entwickeln“*. Es wird der Raumplanung somit eine proaktive Rolle zugeschrieben, die es ermöglicht, Rahmenbedingungen so zu

setzen, dass räumliche Entwicklungen in die aus gesellschaftlicher Sicht gewünschte Richtung gesteuert werden können. Um dies Möglich zu machen ist ein tieferes Verständnis der komplexen räumlichen Zusammenhänge essentiell.

Literaturverzeichnis

Anselin, Luc (1988a). *Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity.* Geographical Analysis, VOL:20(1):1–17.

Anselin, Luc (1988b). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, Boston.

Anselin, Luc (2001). *Spatial Effects in Econometric Practice in Environmental and Resource Economics.* American Journal of Agricultural Exonomics, VOL:83(3):705–710.

Axhausen, Kay W. (2013). *Verkehrssysteme und Raumnutzung.* Vorlesungsuntergen LVA Verkehrssysteme ETH Zürich.

Axhausen, Kay W. und M. Löchl (2010). *Modeling hedonic residential rents for land use and transport simulation while considering spatial effects.* The Journal of Transport and Land Use, VOL:3(2):39–63.

Bank-Austria (2014). *Real Estate Country Facts - Wie stabil ist der österreichische Immobilienmarkt.* Online. Abgerufen am 02.09.2014 unter `http://www.bankaustria.at/files/RealEstate_Oesterreich_deutsch.pdf`.

Bannock, Graham, R. Baxter und D. Evan (2003). *Dictionary of Economics.* Bloomberg Press, Princton, 4 Aufl.

Bivand, Roger (2014). *Spatial dependence: weighting schemes, statistics and models.* Online. Abgerufen am 03.08.2014 unter `http://cran.r-project.org/web/packages/spdep/spdep.pdf`.

Bivand, Roger S., E. J. Pebesma und V. Gomez-Rubio (2008). *Applied Spatial Data Analysis with R.* Springer Verlag, New York.

Bökemann, Dieter und W. Feilmayr (1994). *Kleinräumige Analyse der Wiener Grundstückspreise.* Seminarbericht Gesellschaft für Regionalforschung, VOL:35:22–46.

Bone-Winkel, Stephan, K.-W. Schulte und C. Foeke (2008). *Immobilienökonomie Bd.1 Betriebswirtschaftliche Grundlagen*. Oldenbourg Wissenschaftsverlag, München.

Brunauer, Wolfgang, S. Lang und W. Feilmayr (2013). *Hybrid multilevel STAR models for hedonic house prices*. Jahrbuch für Regionalwissenschaft, VOL:33(2):151–172.

Caduff, Luzian (2013). *Spatial hedonic regression modelling of commercial and office space in Singapore*. Master Thesis, ETH Zurich.

CBRE (2014). *Marketview Wien Büro*. Online. Abgerufen am 02.09.2014 unter `http://www.cbre.at/at_de/research/vienna_reports`.

Chalermpong, Saksith und K. Wattana (2009). *Rent Capitalization of Access to Rail Transit Stations: Spatial Hedonic Models of Office Rent in Bangkok*. Journal of the Eastern Asia Society for Transportation Studies, VOL:8:926–940.

Chegut, Andrea, P. Eichholtz, P. Rodrigues und R. Weerts (2013). *Spatial Dependence in Commercial Real Estate*. Online. Abgerufen am 06.08.2014 unter `http://rri.wvu.edu/wp-content/uploads/2013/07/Fullpaper_2.B.1.pdf`.

Choumert, Johanna, S. Herath und G. Maier (2014). *The value of the greenbelt in Vienna: a spatial hedonic analysis*. Working Papers 201402, CERDI.

Colwell, Peter und G. Dilmore (1999). *Who was first? An examination of an early hedonic study*. Land Economics, VOL:75(4):620–626.

Court, Andrew (1939). *Hedonic price indexes with automotive examples*. The Dynamics of Automotive Demand, S. 98–119.

Day, Brett (2003). *Submarket identification in property markets: A hedonic housing price model for Glasgow*. CSERGE Working Paper EDM, No. 03-09.

Dreger, Christian, R. Kosfeld und M. Türck (2011). *Empirische Regionalforschung heute - Festschrift für Prof. Dr. Hans-Friedrich Eckey*. Gabler Verlag, Wiesbaden.

EHL (2014). *Büromarktbericht Wien*. Online. Abgerufen am 02.09.2014 unter `http://www.ehl.at/de/download?research=61`.

Farooq, Bilal, E. Miller und M. Haider (2010). *Hedonic Analysis of Office Space Rent*. Transportation Research Record: Journal of the Transportation Research Board, VOL:2174:118–127.

Fischer, Manfred und A. Getis (2010). *Handbook of Applied Spatial Analysis - Software Tools, Methodes and Applications.* Springer Verlag, New York.

Fischer, Manfred und J. Wang (2011). *Spatial Data Analysis - Models, Methodes and Techniques.* Springer Verlag, New York.

Fotheringham, Stevard (2000). *Quantitative Geography - A Perspectives on Spatial Data Analysis.* Sage Publications.

Fotheringham, Stewart, C. Brunsdon und M. Charlton (2002). *Geographically Weighted Regression - The Analysis of Spatially Varying Relationships.* John Wiley and Sons Ltd, West Sussex, 1 Aufl.

Fuerst, Franz (2007). *Office Rent Determinants: A Hedonic Panel Analysis.* Online. Abgerufen am 06.08.2014 unter `http://mpra.ub.uni-muenchen.de/11445/`.

Fuhrer, Raphael (2012). *A Hedonic Rental Price Model for the Canton Zurich.* Master Thesis, ETH Zurich.

Goodman, Allen (1998). *Andrew Court and the Invention of Hedonic Price Analysis.* Journal of Urban Economics, VOL:44(2):291–298.

Gret-Regamey, Adrienne und M. van Strien (2013). *Multicriteria Decision Analysis.* ETH Zurich - Script for course 103-0307-00L.

Griliches, Zvi (1961). *Hedonic Price Indexes for Automobiles: An Econometric of Quality Change.* The Price Statistics of the Federal Goverment. NBER, ISBN: 0-87014-072-8 Aufl. 173-196.

Haas, George Casper (1922). *A statistical analysis of farm sales in blue earth county, Minnesota, as a basis for farm land appraisal.* Master Thesis, University of Minnesota.

Haase, Ronny (2011). *Ertragspotenziale - Hednonische Mietpreismodellierungen am Beispiel von Büroimmobilien.* Doktorarbeit, ETH Zürich.

Haining, Robert (2009). *The Special Nature of Spatial Data.* The Sage Handbook of Spatial Analysis. Sage Publications Inc. Kapitel 2.

Herath, Shanaka und G. Maier (2010). *The Hedonic Price Method in Real Estate and Housing Market Research: A Review of the Literature.* SRE-Discussion 2010/03.

Huib, Ernste (2011). *Angewandte Statistik in Geografie und Umweltwissenschaften.* vdf Hochschulverlag AG an der ETH Zürich, Zürich, 1 Aufl.

Jennen, Maarten und D. Brounen (2009). *The Effect of Clustering on Office Rents: Evidence from the Amsterdam Market*. Real Estate Economics, VOL:37:185–208.

Kim, Jin und M. Zhang (2005). *Determining Transits Impact on Seoul Commercial Land Values: An Application of Spatial Econometrics*. International Real Estate Review, VOL:8:1–26.

Kissling, Daniel und G. Carl (2008). *Spatial autocorrelation and the selection of simultaneous autoregressive models*. Global Ecology and Biogeography, VOL:17:59–71.

Lancaster, Kelvin J. (1966). *Andrew Court and the Invention of Hedonic Price Analysis*. Journal of Political Economy, VOL:74(2):132–157.

Leamer, Edward (2008). *Macroeconomic Patterns and Stories*. Springer Verlag, Berlin.

Lehner, Manuel (2011). *Modelling housing prices in Singapore applying spatial hedonic regression*. Master Thesis, ETH Zurich.

LeSage, James und R. K. Pace (2009). *Indroduction to Spatial Econometrics*. CRC Press, New York.

Loader, Clive (1999). *Local Regression and Likelihood*. Springer Verlag, New York.

Löchl, Michael (2010). *Application of spatial analysis methods for understanding geographic variation of prices, demand and market success*. Dissertation ETH Zürich.

Maddala, Gangadharrao, Soundalyarao und L. Kajal (2009). *Indroduction to Econometrics*. Wiley, 4th Aufl.

Maier, Gunther (2011). *Immobilienbewertung - theoretische Konzepte und praktische Anwendungen*. SRE-Discussion 2011/03.

Malpezzi, Stephen (2002). *Hedonic Pricing Models: A Selective and Applied Review*. Housing Economics and Public Policy, S. 67–89.

McMillen, Daniel (1996). *One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach*. Journal of Urban Economics, VOL:40:100–124.

Nappi-Choulet, Ingrid und M. Tristan-Pierre (2009). *A Spatiotemporal Autoregressive Price Index for the Paris Office Property Market*. Real Estate Economics, VOL:37:305–340.

ÖIR und Europaforum-Wien (2014). *STEP2025 - Stadtentwicklungsplan Wien*. Online. Abgerufen am 30.11.2013 unter `https://www.wien.gv.at/stadtentwicklung/studien/b008379.html`.

Özyurt, Selin (2014). *Spatial Dependence in Commercial Property Prices - Micro Evidence from the Netherlands*. Online. Abgerufen am 06.08.2014 unter `http://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1627.pdf`.

Peter, Markus (2013). *Hedonische Immobilienbewertung von Dachgeschosswohnungen mittels „Geographically Weighted Regression“*. Diplomarbeit, Technische Universität Wien.

Pirhofer, Gottfried und K. Stimmer (2013). *Pläne für Wien - Theorie und Praxis der Wiener Stadtentwicklung von 1945 bis 2005*. Online. Abgerufen am 30.11.2013 unter `https://www.wien.gv.at/stadtentwicklung/grundlagen/planungsgeschichte.html`.

Rosen, K.T. (1974). *Hedonic prices and implicit markets: product differentiation in price competition*. Journal of Political Economy, VOL:82(1):34–55.

Rosen, K.T. (1984). *Toward a Model of the Office Building Sector*. Real Estate Economics, VOL:12(3):261–269.

Scholl, Bernd (2013). *Raumentwicklung 1: Einführung in die Aufgabe Raumplanung und Raumentwicklung*. Online. Abgerufen am 06.12.2013 unter `http://www.irl.ethz.ch/re/education/index`.

Smith, Adam und H. C. Recktenwald (1999). *Der Wohlstand der Nation: Eine Untersuchung seiner Natur und Ursache*. Deutscher Taschenbuch Verlag, München, 8 Aufl.

Stahel, Werner und M. Maechler (2015). *Jitter (Add Noise) to Numbers*. Online. Abgerufen am 02.02.2015 unter `http://stat.ethz.ch/R-manual/R-devel/library/base/html/jitter.html`.

Statistik-Austria (2004). *Gebäude- und Wohnungszählung 2001*. Online. Abgerufen am 02.09.2014 unter `http://www.statistik.at/web_de/statistiken/wohnen_und_gebaeude/`.

Statistik-Austria (2014). *Abgestimmte Erwerbsstatistik 2009*. Online. Abgerufen am 30.11.2013 unter `http://www.statistik.at/web_de/presse/060078`.

Stiassny, Alfred (2012). *Basic Econometric Theory and Methods I*. Vienna University of Economics and Business.

Stocker, Herbert (2013). *Einführung in die angewandte Ökonometrie*. Online. Abgerufen am 09.03.2014 unter `http://www.hsto.info/econometrics`.

Tobler, Waldo (1970). *A Computer Movie Simulating Urban Growth in the Detroit Region.* Economic Geography, VOL:46:234–240.

UNData (2012). *City population by sex, city and city type.* Online. Abgerufen am 30.11.2014 unter `http://data.un.org/Data.aspx?d=POP&f=tableCode3A240`.

Ustaoglu, Eda (2003). *Hedonic Price Analysis of Office Rents: A Case Study of the Office Market in Ankara.* Master Thesis, The Graduate School of Social Sciences of Middle East Technical University - The Department of Economics.

Wallace, Henry (1926). *Comparative farmland values in Iowa.* Journal of Land and Public Utility Economics, VOL:2:385– 392.

wien.gv.at (2014a). *Stadtgebiet - Statistiken.* Online. Abgerufen am 30.11.2014 unter `https://www.wien.gv.at/statistik/lebensraum/stadtgebiet/`.

wien.gv.at (2014b). *Wien Kulturgut: Bauperioden.* Online. Abgerufen am 19.09.2014 unter `http://www.wien.gv.at/kultur/kulturgut/architektur/bauperioden.html`.

Wieser, Robert (2006). *Wirkungen der U-Bahn auf den Bodenmarkt.* Working-Paper Nr.: 1/2006.

Wooldridge, Jeffrey M. (2009). *Indroductory Econometrics.* South Western Cengage Learning, Michigan State University, 2 Aufl.

ZKB (2008). *Wertvoller Boden: Die Funktionsweise des Bodenmarktes im Kanton Zürich.* Züricher Kantonalbank und Statistisches Amt des Kantons Zürich.

Anhang A

Verwendete Geodaten

OpenStreetMap Data Austria:
http://download.geofabrik.de/europe/austria.html
Baublöcke für Wien:
https://open.wien.at/site/datenkatalog/?search-term=baubl%C3%B6cke&connection=and#showresults
Straßengraph ASFINAG:
https://www.data.gv.at/katalog/dataset/db389f75-56c3-4d61-9bb7-1f1c675edeaf
ÖV Linien und Haltestellen:
https://open.wien.at/site/datenkatalog/?search-term=Haltestellen&connection=and#showresults

Anhang B

OLS Diagnostik

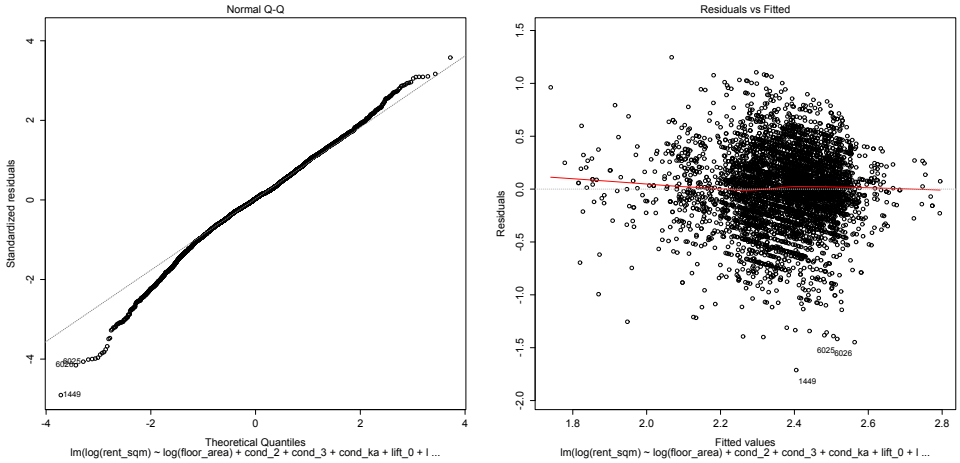

Abbildung B.1: Modelldiagnostik OLS_1: QQ-Plot der Residuen und Residuenplot, eigene Darstellung.

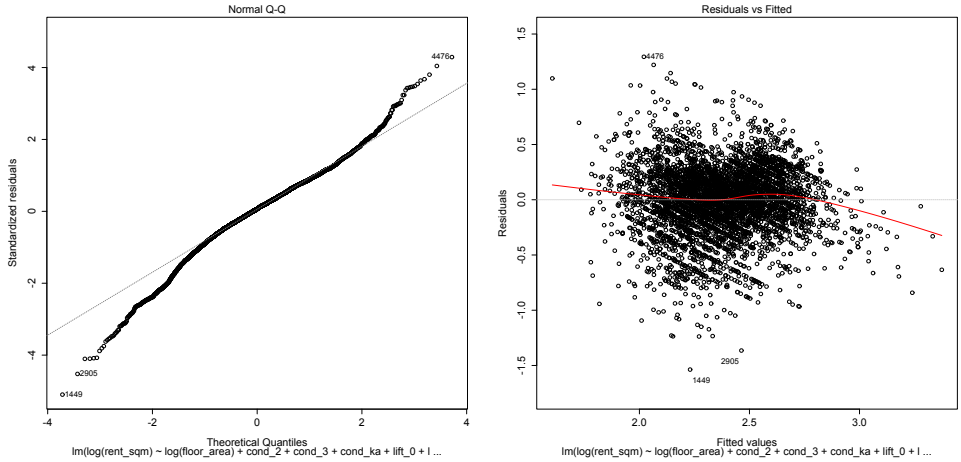

Abbildung B.2: Modelldiagnostik OLS_2: QQ-Plot der Residuen und Residuenplot, eigene Darstellung.

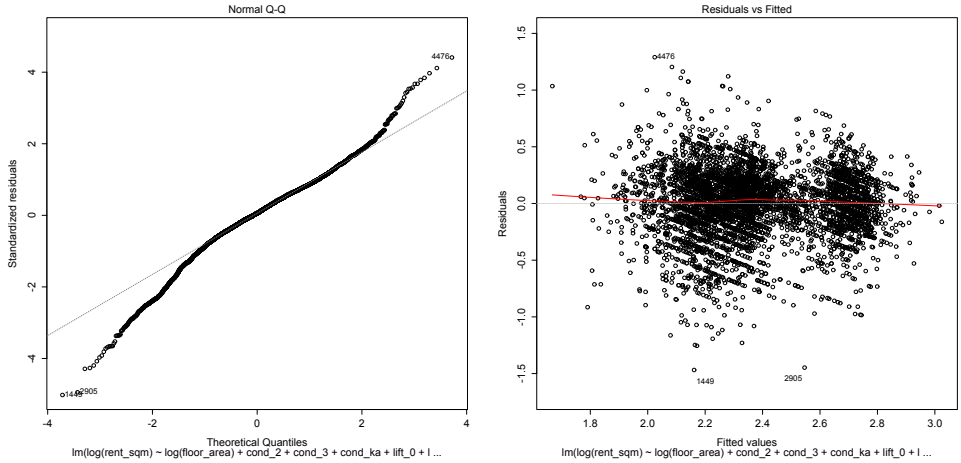

Abbildung B.3: Modelldiagnostik OLS_3: QQ-Plot der Residuen und Residuenplot, eigene Darstellung.